**A Critical Assesment of: Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing**

**\*\*Gemini 2.0 Flash Thinking\*\***

**Section 1 - Introduction**

Federated Learning (FL) has emerged as a promising paradigm for training machine learning models on decentralized data, addressing growing concerns about data privacy. By enabling collaborative model training without direct data sharing, FL facilitates the utilization of vast datasets residing on edge devices, such as smartphones and IoT devices. This approach is particularly relevant in scenarios where data privacy is paramount, such as healthcare and personalized services. However, FL faces significant challenges, particularly in handling non-Independent and Identically Distributed (non-IID) data, which is common in real-world edge computing environments. Furthermore, the traditional focus on global model accuracy in FL often overlooks the critical aspect of individual user performance. This essay presents a critical analysis of the research paper "Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing" by Mills, Hu, and Min, which proposes a novel approach, Multi-Task Federated Learning (MTFL), to address these challenges and enhance personalized learning in federated settings.

**Section 2 - Research Questions and Motivations**

The central research question addressed by Mills, Hu, and Min is: How can Federated Learning be enhanced to effectively handle non-IID data and improve user-level model accuracy (User Accuracy - UA), while also preserving user data privacy and maintaining computational efficiency? This overarching question stems from several key motivations:

Firstly, the problem of non-IID data in FL is a significant impediment to convergence and model performance. Data heterogeneity across edge devices is inherent, and traditional FL algorithms often struggle to generalize effectively in such scenarios. This necessitates the development of techniques that are robust to data distribution variations.

Secondly, the authors highlight the limitation of focusing solely on global model accuracy. In many practical applications of FL, such as personalized recommendation systems, the performance of the model for individual users is more critical than the average performance across all users. Therefore, there is a need to shift the focus towards improving UA.

Thirdly, data privacy remains a core tenet of FL. While FL inherently aims to protect privacy by avoiding direct data sharing, further enhancements are desirable to minimize information leakage and strengthen privacy guarantees.

Finally, computational and storage efficiency are crucial considerations for deploying FL on resource-constrained edge devices. Personalized FL approaches should ideally minimize additional computational overhead and storage requirements on these devices. Existing personalized FL methods often introduce significant complexity, motivating the need for simpler and more efficient solutions.

These motivations collectively underscore the need for a novel FL approach that addresses the limitations of existing methods in terms of personalization, non-IID data handling, privacy, and efficiency.

**Section 3 - Contributions**

The paper makes several significant contributions to the field of Federated Learning, primarily centered around the introduction and evaluation of the Multi-Task Federated Learning (MTFL) algorithm.

* The MTFL Algorithm: The core contribution is the proposal of MTFL, a novel algorithm that integrates Multi-Task Learning principles into the Federated Learning framework. MTFL leverages private Batch Normalization (BN) layers as "patches" within a shared global Deep Neural Network (DNN) model. These private BN layers are personalized to each client's local data, enabling the model to adapt to individual data distributions while still benefiting from federated aggregation. This approach is compatible with existing iterative FL algorithms, such as FedAvg.
* User Accuracy (UA) Metric: The authors introduce and advocate for the User Accuracy (UA) metric as a more relevant performance indicator for FL in personalized settings. UA, defined as the average accuracy across individual client test sets, directly reflects the performance experienced by each user, contrasting with the traditional global model accuracy measured on a centralized IID test set. This shift in metric emphasizes the practical relevance of personalized learning in FL.
* Analysis of Private BN Layers: The paper provides a detailed analysis of the impact of private BN layers on model training and inference. Through theoretical arguments and empirical evidence, the authors demonstrate how private BN layers help maintain neuron activation distributions closer to pre-aggregation states, mitigating performance degradation after federated averaging. Furthermore, they investigate the effect of keeping different components of BN layers (statistics and/or trainable parameters) private, providing valuable insights into the trade-offs between personalization, convergence, and privacy.
* Empirical Validation: The paper presents extensive experimental results on MNIST and CIFAR10 datasets under various non-IID data distributions, client participation rates, and optimization strategies. These experiments convincingly demonstrate that MTFL, particularly when combined with FedAvg-Adam, significantly outperforms standard FL (FedAvg) and other personalized FL algorithms (Per-FedAvg, pFedMe) in terms of UA and convergence speed. The results highlight the effectiveness of MTFL in addressing the challenges of non-IID data and achieving personalized learning in federated settings.
* Edge-Computing Testbed Evaluation: Beyond simulations, the authors validate MTFL on a realistic edge-computing testbed using Raspberry Pi devices. This practical evaluation demonstrates that the performance benefits of MTFL, in terms of UA and convergence speed, outweigh the computational overhead introduced by the algorithm in a real-world deployment scenario.

**Section 4 - Research Methodology and Methods**

**Section 4.1 - Research Methodology**

The research methodology employed in the paper is primarily empirical and experimental, focused on developing and rigorously evaluating the proposed MTFL algorithm. The authors follow a structured approach:

1. Algorithm Design and Formulation: They begin by designing the MTFL algorithm, clearly articulating its core components, including the use of private BN layers and its compatibility with iterative FL frameworks. They mathematically formulate the objective function for MTFL, contrasting it with the traditional FL objective.
2. Implementation and Experimental Setup: They implement MTFL and baseline FL algorithms (FedAvg, FedAdam, FedAvg-Adam, Per-FedAvg, pFedMe) in a simulation environment. They carefully design experimental scenarios using MNIST and CIFAR10 datasets, focusing on non-IID data partitioning to mimic real-world federated settings. They systematically vary parameters such as the number of clients, client participation rates, and optimization strategies to comprehensively evaluate the performance of MTFL under different conditions.
3. Performance Evaluation and Metric Selection: The authors introduce and utilize the User Accuracy (UA) metric as the primary performance indicator, aligning with the focus on personalized learning. They measure and compare the UA achieved by MTFL and baseline algorithms, along with the number of communication rounds required to reach target UA levels. They also analyze training and testing accuracy curves to understand the dynamics of model learning and generalization.
4. Statistical Analysis and Result Interpretation: Experiments are repeated over multiple trials with different random seeds to ensure the robustness and statistical significance of the findings. Results are presented using tables and figures, with clear interpretations and comparisons between algorithms. Statistical measures like confidence intervals are used to quantify the variability of results.
5. Real-World Validation: To further validate the practical applicability of MTFL, the authors conduct experiments on a physical edge-computing testbed, demonstrating its performance and feasibility in a more realistic deployment environment.

**Section 4.2 - Research Methods**

The core research method revolves around the development and comparative analysis of machine learning algorithms, specifically within the Federated Learning domain. Key methods employed include:

* Algorithm Design: The central method is the design of the MTFL algorithm itself. This involves conceptualizing the use of private BN layers as patches, integrating them into the FedAvg framework, and defining the training and aggregation procedures.
* Empirical Comparison of Algorithms: The research heavily relies on the empirical comparison of MTFL with several baseline algorithms:
  + FL (FedAvg): Serves as the fundamental baseline for standard Federated Learning without personalization.
  + FedAdam & FedAvg-Adam: Evaluates the impact of adaptive optimization strategies within both standard FL and MTFL.
  + Per-FedAvg & pFedMe: Represent state-of-the-art personalized FL algorithms, allowing for a direct comparison of MTFL's personalization capabilities.
* Performance Metric Evaluation: The authors introduce and utilize User Accuracy (UA) as a novel evaluation metric, contrasting it with the traditional global model accuracy. This methodological choice reflects a deliberate shift towards evaluating personalized performance in FL.
* Statistical Significance Testing (Implicit): While not explicitly stated as formal statistical significance tests, the use of multiple trials and averaging of results across random seeds implicitly addresses the need to assess the statistical robustness of observed performance differences. The inclusion of confidence intervals in figures further strengthens the reliability of the empirical comparisons.
* Edge-Computing Testbed Experimentation: Moving beyond simulation, the use of a Raspberry Pi-based testbed provides a crucial method for validating the practical feasibility and performance of MTFL in a real-world edge-computing environment. This method helps assess the algorithm's behavior under resource constraints and network communication overheads.

**Section 5 - Critique/Evaluation**

The research paper by Mills, Hu, and Min presents a compelling and well-executed study on personalized Federated Learning. The proposed MTFL algorithm is innovative, theoretically sound, and empirically validated across diverse scenarios. However, like any research, it also has certain limitations and areas for further exploration.

Strengths:

* Novelty and Significance: MTFL is a genuinely novel approach that effectively addresses the critical challenges of non-IID data and personalized learning in FL. The use of private BN layers as patches is a clever and efficient way to achieve personalization without significantly increasing computational overhead or communication costs. The shift towards the UA metric is a valuable contribution to the field, promoting a more user-centric evaluation of FL performance.
* Comprehensive Empirical Validation: The paper provides extensive and rigorous empirical validation of MTFL across various datasets, experimental settings, and optimization strategies. The results convincingly demonstrate the superiority of MTFL over baseline algorithms in terms of UA and convergence speed. The inclusion of both simulation-based experiments and a real-world testbed evaluation strengthens the robustness and practical relevance of the findings.
* Theoretical Justification: The authors provide a clear theoretical rationale for the effectiveness of private BN layers in MTFL, explaining how they help maintain neuron activation distributions and mitigate performance degradation due to federated averaging. This theoretical grounding adds depth and credibility to the empirical results.
* Clarity and Presentation: The paper is well-written, clearly structured, and effectively presents complex concepts and experimental results. The use of figures and tables is effective in visualizing performance comparisons and highlighting key findings.

Weaknesses and Areas for Future Work:

* Limited Patch Types: The paper primarily focuses on BN layers as private patches. While BN layers are shown to be effective, exploring other types of private layers or patch mechanisms could further enhance personalization and performance. For example, investigating private attention mechanisms or lightweight convolutional layers as patches could be a potential avenue for future research.
* Hyperparameter Sensitivity: While the authors mention tuning hyperparameters, the paper could benefit from a more detailed analysis of the hyperparameter sensitivity of MTFL, particularly regarding the learning rate and other optimization parameters. Understanding how MTFL's performance varies with different hyperparameter settings would be valuable for practical deployment.
* Communication Cost Analysis: Although the paper acknowledges the increased communication cost of FedAvg-Adam within MTFL, a more detailed analysis of the communication overhead for different MTFL configurations and patch sizes would be beneficial. Quantifying the communication-performance trade-off would be crucial for optimizing MTFL for bandwidth-constrained edge environments.
* Broader Application Scenarios: The experimental evaluation is primarily focused on image classification tasks (MNIST and CIFAR10). Extending the evaluation of MTFL to other application domains, such as natural language processing or time-series analysis, would further demonstrate its generalizability and applicability.
* Privacy Guarantees: While MTFL enhances privacy by keeping BN layers private, a more formal analysis of the privacy guarantees offered by MTFL would be a valuable extension. Investigating differential privacy or other privacy-enhancing techniques in conjunction with MTFL could further strengthen its privacy-preserving capabilities.

**References**

* [Mills, J., Hu, J. and Min, G., 2021. Multi-task federated learning for personalised deep neural networks in edge computing. *IEEE Transactions on Parallel and Distributed Systems*, *33*(3), pp.630-641.] (This is based on the provided assignment list citation style, you should adjust to Harvard referencing style (Bath) as requested and include other references you have used in your critical analysis if any).
* Bonawitz, K. et al. (2019). “Towards federated learning at scale: System design.” *Proc. Syst. Mach. Learn. Conf.*
* Dinh, C. T., Tran, N., and Nguyen, J. (2020). “Personalized federated learning with moreau envelopes.” *Proc. Conf. Neural Inf. Process. Syst.*, vol. 33, pp. 21394-21405.
* Fallah, A., Mokhtari, A., and Ozdaglar, A. (2020). “Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach.” *Proc. Conf. Neural Inf. Process. Syst.*, vol. 33, pp. 3557-3568.
* McMahan, B. et al. (2017). “Communication-efficient learning of deep networks from decentralized data.” *Proc. 20th Int. Conf. Artif. Intell. Stat.*, pp. 1273-1282.
* Smith, V., Chiang, C.-K., Sanjabi, M., and Talwalkar, A. (2017). “Federated multi-task learning.” *Proc. Conf. Neural Inf. Process. Syst.*, pp. 4427-4437.

**\*\*Deepseek R1\*\***

**Presenting a Research Paper: "Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing"**

1. Introduction

Federated Learning (FL) has emerged as a privacy-preserving paradigm for training machine learning models across decentralized devices without sharing raw data. However, traditional FL frameworks face challenges such as non-Independent and Identically Distributed (non-IID) data across clients and a focus on global model accuracy rather than individual user performance. This paper, "Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing" by Mills et al. (2022), addresses these limitations by proposing a Multi-Task FL (MTFL) algorithm. MTFL introduces personalization through private Batch Normalization (BN) layers and combines this with adaptive optimization strategies to enhance both convergence speed and user-specific model accuracy. The work is motivated by the growing need for edge computing solutions that balance privacy, efficiency, and tailored performance in real-world applications like healthcare and content recommendation.

2. Research Questions and Motivations

The paper investigates three core research questions:

1. How can personalized models improve individual User model Accuracy (UA) in FL under non-IID data?  
   Traditional FL optimizes a global model, which may underperform on individual clients due to data heterogeneity. The authors argue that personalization is critical for applications like personalized recommendations, where local data distributions vary significantly.
2. How can convergence speed in FL be accelerated despite non-IID data?  
   Non-IID data slows convergence as client models diverge during local training. The paper explores adaptive optimization techniques (e.g., FedAvg-Adam) to mitigate this.
3. How to balance personalization with the benefits of federated aggregation?  
   Fully independent local models lose the advantages of collaborative learning, while overly globalized models ignore user-specific patterns. MTFL strikes this balance by keeping only lightweight BN layers private.

Motivations stem from practical challenges in FL deployments:

* Privacy concerns: Users resist uploading sensitive data, necessitating decentralized training.
* Edge computing demands: Low-latency processing requires efficient, localized models.
* Limitations of existing FL: Global models often fail to address user-specific needs, especially with non-IID data.

3. Contributions

The paper makes five key contributions:

1. MTFL Algorithm: Integrates Multi-Task Learning (MTL) into FL by privatizing BN layers. Clients retain personalized BN parameters (γ, β) and statistics (μ, σ), enabling tailored models while sharing other layers.
2. User model Accuracy (UA): Introduces UA as a performance metric, reflecting real-world objectives where improving individual client accuracy matters more than global metrics.
3. FedAvg-Adam Optimization: Proposes a distributed Adam optimization strategy that accelerates convergence by averaging client-specific momentum and variance parameters.
4. Empirical Validation: Demonstrates MTFL reduces communication rounds by up to 5× with FedAvg and 3× further with FedAvg-Adam on MNIST and CIFAR10 datasets.
5. Edge Computing Testbed: Validates practicality using Raspberry Pi clients, showing MTFL’s overheads are outweighed by its UA and convergence benefits.

4. Research Methodology and Methods

4.1 Research Methodology

The MTFL framework operates in iterative client-server rounds:

1. Client Selection: A subset of clients participates per round.
2. Model Download: Clients receive the global model and augment it with private BN layers.
3. Local Training: Clients train personalized models using their data.
4. Parameter Upload: Non-private layers (e.g., convolutional weights) are sent to the server.
5. Aggregation: The server averages uploaded parameters to update the global model.

4.2 Research Methods

* Datasets and Models: Experiments use MNIST and CIFAR10 with non-IID splits. Models include a 2-layer Fully Connected Network (2NN) and a Convolutional Neural Network (CNN).
* Optimization Strategies: Compared FedAvg, FedAdam, and FedAvg-Adam. FedAvg-Adam applies Adam on clients and averages optimizer states, reducing rounds needed for convergence.
* BN Layer Analysis: Evaluated keeping BN parameters (γ, β), statistics (μ, σ), or both private. Results showed private γ/β outperformed private μ/σ, as mismatched statistics harmed generalization.

5. Critique/Evaluation

Strengths

* Effective Personalization: Private BN layers account for local data distributions, improving UA without significant communication overhead (BN parameters are <1% of model size).
* Novel Metric: UA aligns with real-world objectives better than global accuracy, especially in personalized applications.
* Adaptive Optimization: FedAvg-Adam’s distributed momentum averaging accelerates convergence, addressing a key FL bottleneck.
* Practical Validation: The edge-computing testbed demonstrates feasibility in resource-constrained environments.

Weaknesses

* Limited Layer Exploration: The paper focuses solely on BN layers for personalization. Other lightweight layers (e.g., depthwise convolutions) could offer additional benefits.
* Scalability Concerns: Experiments used up to 400 clients, but real-world FL may involve millions. The impact of extreme scalability on BN personalization remains unaddressed.
* Assumption of Client Participation: The framework assumes clients willingly participate in rounds, ignoring potential dropout or adversarial behavior.

Comparison to Existing Work

MTFL outperforms personalized FL baselines like Per-FedAvg and pFedMe in UA and convergence speed (Fig. 5). Unlike these methods, MTFL requires fewer hyperparameters and computational resources, making it more deployable. However, Per-FedAvg’s meta-learning approach might generalize better to unseen tasks, a trade-off not explored here.

Future Work

* Extending MTFL to decentralized Peer-to-Peer FL.
* Exploring privacy-accuracy trade-offs when more layers are personalized.
* Testing on larger-scale datasets (e.g., ImageNet) and heterogeneous client hardware.

6. Conclusion

Mills et al. present a compelling solution to FL’s personalization and convergence challenges. By integrating BN-based Multi-Task Learning with adaptive optimization, MTFL achieves state-of-the-art UA while maintaining privacy and efficiency. The paper’s empirical rigor and practical validation strengthen its contributions, though scalability and layer diversity warrant further investigation. This work advances FL’s applicability in edge computing, offering a blueprint for future research in personalized, privacy-preserving machine learning.

References

Mills, J., Hu, J., & Min, G. (2022). Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing. IEEE Transactions on Parallel and Distributed Systems, 33(3), 630–641. <https://doi.org/10.1109/TPDS.2021.3098467>

**\*\*GPT 4o\*\***

Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing

1. Introduction

Federated Learning (FL) is a decentralized approach to training machine learning models without sharing raw data between users. This technique is particularly useful in privacy-sensitive applications such as healthcare, finance, and mobile-based predictive modeling. However, traditional FL faces challenges, including non-IID (Independent and Identically Distributed) data, slow convergence rates, communication overhead, and inefficient personalization. The research paper by Mills et al. proposes a novel Multi-Task Federated Learning (MTFL) algorithm that addresses these issues by incorporating private Batch Normalization (BN) layers, improving User Model Accuracy (UA), and optimizing training efficiency

FL is widely adopted in applications where privacy is crucial, such as medical imaging analysis, where patient data should not be shared across institutions. Similarly, FL is used in smart devices, such as personalized mobile keyboards and recommendation systems. Despite these advantages, FL suffers from the challenge of non-uniform data distribution among clients, leading to suboptimal global models. This paper aims to improve the effectiveness of FL by personalizing models while maintaining efficiency

2. Research Questions and Motivations

The primary research questions investigated in the paper are:

1. How can personalized Federated Learning improve individual client model accuracy while maintaining privacy?
2. How do non-IID data distributions impact Federated Learning, and can Multi-Task Learning mitigate these effects?
3. What optimization strategies enhance the convergence speed and efficiency of FL?

The motivation behind this research arises from the limitations of standard FL approaches, which often result in suboptimal performance for individual users due to the averaging of model parameters across diverse datasets. Personalization in FL can improve local model accuracy, but existing approaches either require excessive computation, suffer from high latency, or compromise privacy. The proposed MTFL method aims to achieve better convergence speed and personalization while maintaining privacy

3. Contributions

The paper makes several significant contributions to the field of Federated Learning:

* Introduction of MTFL: A novel Multi-Task Federated Learning framework that integrates private BN layers for personalization.
* Enhanced Optimization Strategy: Proposal of FedAvg-Adam, an adaptive optimization approach that accelerates convergence compared to traditional FedAvg.
* Evaluation of User Model Accuracy (UA): Demonstration of UA as a more effective metric for assessing the performance of personalized FL models.
* Empirical Validation: Extensive experiments using MNIST and CIFAR10 datasets, showing that MTFL achieves significantly faster convergence and higher UA compared to traditional FL approaches.
* Implementation on Edge Computing Testbed: Deployment on a Raspberry Pi-based testbed to validate the real-world applicability of MTFL

4. Research Methodology and Methods

4.1 Research Methodology

The study follows an empirical approach, where the proposed MTFL method is evaluated through simulations and real-world edge computing experiments. The methodology includes:

* Theoretical formulation of MTFL, incorporating private BN layers to allow local model adaptation.
* Comparison of different optimization strategies, including FedAvg, FedAdam, and the proposed FedAvg-Adam.
* Evaluation of performance metrics such as convergence speed, UA, and computational efficiency.
* Simulation experiments on standard FL datasets (MNIST and CIFAR10) to analyze performance under different conditions.
* Real-world validation on an edge computing testbed with Raspberry Pi devices

4.2 Research Methods

The study uses the following methods to evaluate MTFL:

* FL Optimization Strategies: Three optimization strategies—FedAvg, FedAdam, and FedAvg-Adam—are tested. FedAvg is the baseline, while FedAvg-Adam introduces adaptive optimization for improved performance.
* Dataset Partitioning: To simulate non-IID conditions, datasets are split into shards and assigned to clients in a non-uniform manner, ensuring realistic FL scenarios.
* Model Architecture: Two neural network architectures are used: a simple 2-layer fully connected network for MNIST and a CNN for CIFAR10.
* Performance Metrics: The main metric is UA, which measures individual model accuracy on local client data, providing better insights into personalization benefits compared to traditional global model accuracy.
* Hardware Considerations: The real-world testbed includes Raspberry Pi devices to simulate the computational constraints of edge environments

5. Critique and Evaluation

Strengths

1. Personalization and Privacy: The introduction of private BN layers effectively enhances local model accuracy while preserving privacy, a major advantage over standard FL approaches.
2. Efficient Optimization: The use of FedAvg-Adam significantly reduces the number of communication rounds required for convergence, making FL more practical for real-world applications.
3. Comprehensive Evaluation: The paper provides extensive experimental validation, both in simulations and on an edge computing testbed, strengthening the credibility of its findings.
4. Scalability and Adaptability: The MTFL approach can be integrated with various FL frameworks and optimization strategies, making it highly adaptable.
5. Computational Efficiency: Compared to other personalized FL methods, MTFL achieves faster convergence with reduced communication overhead

Limitations

1. Computational Overhead: While MTFL improves convergence speed, it introduces additional computational complexity due to the need for private BN layers and optimization parameter sharing.
2. Applicability to Larger Models: The study focuses on relatively simple models (e.g., CNNs for MNIST and CIFAR10). It remains unclear how well MTFL scales to large-scale deep learning models.
3. Limited Real-World Testing: Although tested on Raspberry Pi devices, further evaluation on diverse edge computing infrastructures is necessary to fully assess its deployment feasibility.
4. Security Concerns: While MTFL maintains privacy, it does not explicitly address adversarial attacks, which could be a crucial factor in real-world applications

Future Work

The paper suggests several directions for future research:

* Exploring Alternative Personalization Mechanisms: Investigating other methods for local model adaptation beyond BN layers.
* Scalability to Complex Models: Applying MTFL to more complex deep learning models and real-world applications such as medical imaging and autonomous systems.
* Reducing Communication Costs: Developing strategies to minimize the additional communication overhead introduced by MTFL.
* Decentralized MTFL: Investigating how MTFL could be applied in fully decentralized FL settings without a central coordinating server.
* Security Enhancements: Exploring adversarial defenses and differential privacy techniques to further secure FL implementations

6. Conclusion

The research paper presents a novel Multi-Task Federated Learning (MTFL) approach that significantly enhances personalization and convergence speed in Federated Learning. By introducing private BN layers and leveraging an adaptive optimization strategy (FedAvg-Adam), MTFL effectively addresses the challenges of non-IID data and inefficient training in FL. Experimental results demonstrate that MTFL outperforms traditional FL approaches in terms of User Model Accuracy and convergence speed. However, challenges related to computational overhead, security, and real-world deployment need further investigation. Overall, this study represents a valuable contribution to the evolving field of personalized Federated Learning

**\*\*o1 Pro\*\***

**Introduction**

Federated learning (FL) has emerged as a compelling paradigm for training machine learning models across distributed devices—often referred to as clients—without transferring their private local data to a central server (McMahan et al., 2017). This distributed approach is particularly relevant in multi-access edge computing environments, where substantial amounts of data are generated on mobile devices or Internet-of-Things (IoT) endpoints and cannot be centrally stored due to privacy or bandwidth constraints. Nonetheless, the growing adoption of FL faces persistent challenges arising from non-IID (non-independent and identically distributed) data, heterogeneous device capabilities, and the need to personalise model performance to each client’s local context.

In their paper, *Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing*, Mills et al. (2021) tackle these challenges by proposing a method—Multi-Task Federated Learning (MTFL)—that layers personalisation and multi-task learning on top of standard iterative federated optimisation. They demonstrate that integrating private Batch Normalisation (BN) “patches” into the shared model significantly improves both the convergence speed and local user accuracy when compared to conventional FL methods.

**Research Questions and Motivations (Pt1)**

Mills et al. (2021) begin by foregrounding two primary research questions. First, **how can federated learning approaches adapt to non-IID data distributions**, which are typical in real-world applications where each client’s dataset may capture unique trends or label distributions? Second, **how can FL be personalised so that each client’s model attains higher local accuracy** rather than focusing purely on the global model’s performance?

The paper’s motivation stems from well-documented empirical findings that non-IID data impedes the convergence and final accuracy of standard FL algorithms. Moreover, in many real-world scenarios—ranging from healthcare analytics to personalised recommendations—improving the “per-client” or “per-user” accuracy is more relevant than optimising a single global model for all participants (Fallah et al., 2020; Dinh et al., 2020). As such, Mills et al. seek to design a method that not only alleviates the detrimental effects of non-IID data but also **personalises** the final models in a way that preserves client privacy and minimises additional computation or memory overhead

**2. Significant Contributions (Pt2)**

To address these questions, Mills et al. (2021) offer four main contributions:

1. **Multi-Task Federated Learning (MTFL) Algorithm**:  
   They propose augmenting standard federated architectures with private BN “patches,” effectively treating each client’s dataset as a unique but related task. By keeping certain BN parameters private—such as the trainable BN scale and shift factors—clients gain personalised representations without having to store or upload an entire personalised model.
2. **User Model Accuracy (UA) Metric**:  
   A novel metric, “average User model Accuracy” (UA), is introduced to better reflect the genuine objective in many FL deployments: optimising how well each user’s local model performs. This departure from measuring only a global test accuracy is vital in highlighting benefits for clients whose local data differ considerably from the global average.
3. **Investigation of BN Patch Layer Effects**:  
   The authors systematically examine which BN parameters—tracked statistics (mean, variance) or learned scale and shift—should be private to yield the best convergence–performance trade-off. Their analysis reveals that storing private trainable BN parameters (scale and shift) often provides the most pronounced gain in accuracy and convergence speed.
4. **Integration with Adaptive Optimisation (FedAvg-Adam)**:  
   Building on existing iterative methods (e.g., FedAvg), they show that combining MTFL with an adaptive optimisation strategy that distributes the Adam update across clients (FedAvg-Adam) accelerates learning. The results demonstrate fewer communication rounds are needed to attain target UAs, especially in challenging, non-IID settings.

These contributions are significant for machine learning and edge computing because they not only **highlight a path toward scalable personalisation** (via lightweight BN patching) but also **improve FL’s suitability for practical deployments** where data distribution skew and user-centric performance metrics are crucial.

**3. Research Methodology and Methods (Pt3)**

**3.1 Methodology**

Mills et al. (2021) situate their work within the broader iterative FL framework, where a central server and multiple clients iteratively share and update model parameters. The **core innovation** lies in treating each client’s local learning as a form of multi-task learning. Specifically, each client’s model becomes a composite of globally shared parameters and localised “patch” layers. Because different clients operate in unique data regimes (non-IID data), the localised patches enable each client’s model to adapt to that client’s training distribution.

They frame FL as an optimisation problem over all client objectives, summing each client’s local loss. Personalisation is introduced by allowing a subset of parameters—namely BN statistics or trainable BN factors—to remain private. This approach delivers two major advantages:

1. **Better Local Accuracy**: Private BN parameters directly adjust the intermediate activation distributions to match local data.
2. **Increased Privacy**: The global server sees fewer parameters from each client, thus reducing the risk of inferring private data distributions.

**3.2 Methods and Experimental Design**

Their experiments use the well-known **MNIST** (handwritten digits) and **CIFAR10** (object classification) datasets. Each dataset is partitioned in a non-IID manner by assigning shards of data to clients—thereby replicating the typical heterogeneous data environment. The authors then:

* Compare **FL (FedAvg)** (McMahan et al., 2017) against **MTFL (FedAvg)**, **MTFL (FedAdam)**, and **MTFL (FedAvg-Adam)**.
* Evaluate **User model Accuracy (UA)**, i.e., the average classification accuracy on each client’s local test data, as well as the required **communication rounds** to reach certain accuracy thresholds.
* Analyse different patch layer strategies:
  + Keeping BN trainable parameters (scale, shift) private,
  + Keeping BN statistics (running mean, variance) private,
  + Keeping both sets of BN parameters private.
* Investigate robustness to noisy client data, thus testing how personalised patches might shield non-noisy clients from corrupted global updates.

On top of simulation-based experiments, the paper presents a real-world edge computing testbed with Raspberry Pi devices to measure the practical runtime overhead and confirm that the speed-up in convergence compensates for any extra communication cost introduced by Adam-based optimisation.

**4. Critical Assessment (Pt4)**

**4.1 Strengths and Novel Insights**

One of the paper’s key strengths is its **direct focus on personalisation** and the explicit introduction of User model Accuracy (UA) as a metric. In many application domains, it is indeed more relevant to optimise for each client’s individual performance than to maximise the global model’s test accuracy. By tackling **non-IID data**, the authors produce valuable insights that bring FL closer to deployment in settings like personalised mobile services and healthcare.

Additionally, the use of **BN patch layers** is elegant and computationally lightweight. Previous works on personalised FL, such as Per-FedAvg (Fallah et al., 2020) or pFedMe (Dinh et al., 2020), often require extra fine-tuning steps or store larger local models. By contrast, BN patches only keep a minimal subset of parameters private, thus striking an **excellent balance** between personalisation capability, communication overhead, and memory footprint.

The authors also contribute a new perspective on **distributed adaptive optimisation**, i.e., FedAvg-Adam. Many previous FL studies concentrate on either standard SGD or server-only adaptive updates (Reddi et al., 2021). Mills et al.’s demonstration that distributed Adam on clients can substantially reduce the number of communication rounds is an important insight for large-scale FL applications where round-efficiency is paramount.

**4.2 Limitations and Potential Improvements**

Despite these strengths, some limitations remain:

1. **Limited Exploration of Larger, More Complex Models**:  
   The evaluation focuses on relatively small architectures for MNIST and CIFAR10. While this is standard in FL research, future work could explore larger, more modern architectures where BN usage is more widespread (e.g., ResNet or transformers). Such studies could yield deeper insights into how patch layers function in large-scale networks and whether the overhead remains manageable.
2. **Early Training Phase Challenges**:  
   The authors note that the advantage of BN patches grows as training stabilises. During early rounds, when gradients are large, local parameters (including BN) diverge significantly across clients. This divergence might still slow early progress. Strategies that adaptively switch patch parameters on or off at different stages of training—or use partial mixing of BN parameters—might address the initially slow collaboration.
3. **Security and Robustness Against Malicious Clients**:  
   While the authors do test resilience to noisy data, a thorough security analysis under adversarial conditions could be beneficial (e.g., how BN patching interacts with model poisoning attacks or Byzantine behaviours). FL often operates in open ecosystems where malicious clients might attempt to degrade global performance or glean information about others.
4. **Applicability Beyond Convolutional/BN Layers**:  
   Although BN is a natural choice for image tasks, some domains or architectures may rely on alternative normalisation techniques (e.g., LayerNorm in transformers). Investigating whether the concept of “private patches” generalises to other normalisation or parameter blocks would broaden the approach’s applicability.

**4.3 Future Research Directions**

Building on the above, future work could:

* Evaluate **BN patch personalisation** on large-scale, real-world tasks involving millions of clients with resource constraints typical of mobile networks.
* Investigate dynamic patch layer selection, so that local layers adapt to data drift or different user contexts over time.
* Explore advanced defence mechanisms if clients act maliciously, possibly merging personalised patches with robust aggregation (Konstantinov and Lampert, 2019) or differential privacy.
* Generalise the private patch concept to other **normalisation** approaches or partially freeze patches to further reduce overhead.

**References (Harvard – University of Bath style)**

* Dinh, C.T., Tran, N. and Nguyen, J., 2020. Personalised federated learning with Moreau envelopes. *Advances in Neural Information Processing Systems*, 33, pp.21394–21405.
* Fallah, A., Mokhtari, A. and Ozdaglar, A., 2020. Personalised federated learning with theoretical guarantees: A model-agnostic meta-learning approach. *Advances in Neural Information Processing Systems*, 33, pp.3557–3568.
* Konstantinov, N. and Lampert, C., 2019. Robust learning from untrusted sources. *Proceedings of the 36th International Conference on Machine Learning*, pp.3488–3498.
* McMahan, H.B., Moore, E., Ramage, D. and y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp.1273–1282.
* Mills, J., Hu, J. and Min, G., 2021. Multi-task federated learning for personalised deep neural networks in edge computing. *IEEE Transactions on Parallel and Distributed Systems*, 33(3), pp.630–641.
* Reddi, S.J. et al., 2021. Adaptive federated optimization. *International Conference on Learning Representations*. Available at: https://openreview.net/pdf?id=LkFG3lB13U5 [Accessed 2 February 2025].

**\*\*o3 Mini High\*\***

**Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing: A Critical Analysis**

**Introduction**

Federated Learning (FL) has emerged as a promising paradigm for training deep neural networks (DNNs) on decentralized data while preserving user privacy (McMahan et al., 2017). However, practical challenges such as non-Independent and Identically Distributed (non-IID) data and the need for personalised models have driven further innovation. In their paper, Mills, Hu and Min (2021) propose a Multi-Task Federated Learning (MTFL) algorithm that incorporates private “patch” layers—in particular, Batch Normalisation (BN) layers—to enable personalised deep models on edge devices. This essay critically examines the paper’s research questions and motivations, its significant contributions, the methodologies employed, and finally provides an informed critique supported by additional literature.

**Research Questions and Motivations**

Mills et al. (2021) address two interrelated research questions:

1. **How can federated learning be adapted to improve individual user model accuracy (UA) when client data are non-IID?**
2. **Can the introduction of private model components (via patch layers) enhance both convergence speed and privacy in edge computing environments?**

The motivation behind these questions stems from real-world applications—such as personalised content recommendation and mobile keyboard prediction—where a global model may underperform on individual clients (Hard et al., 2018). Traditional FL methods, like Federated Averaging (FedAvg) (McMahan et al., 2017), focus on global accuracy measured on centralised IID test-sets; however, they can be suboptimal for personalised tasks due to divergence in local data distributions. By proposing a method that emphasises user-specific performance (UA) and reduces the harmful impact of non-IID data, the authors seek to address critical challenges in both model performance and privacy.

**Significant Contributions**

The paper makes several notable contributions:

* **MTFL Algorithm with Private Patch Layers:** The authors introduce a novel FL approach that integrates Multi-Task Learning (MTL) principles by incorporating private BN layers into the global DNN architecture. This “patching” not only tailors models to individual data distributions but also reduces the risk of privacy leakage by keeping critical parameters local.
* **User Model Accuracy (UA) as a Metric:** Recognising that global model accuracy may not reflect client-specific performance, the authors propose UA, measured on local test sets, as a more appropriate performance metric for personalised FL.
* **FedAvg-Adam Optimisation Strategy:** Beyond using standard FedAvg, the paper evaluates a distributed version of the Adam optimiser (FedAvg-Adam), demonstrating that adaptive optimisation can further accelerate convergence in the FL setting.
* **Empirical Validation:** Through extensive experiments on MNIST and CIFAR10 and a realistic edge-computing testbed (using Raspberry Pi devices), the authors demonstrate that MTFL can reduce the number of communication rounds required to achieve target UA by significant margins compared to traditional FL and other personalised FL methods.

These contributions are significant in that they not only propose a method for improving personalised performance in FL but also address the computational and communication challenges inherent in edge computing environments.

**Research Methodology and Methods**

The methodology of Mills et al. (2021) builds on the standard iterative FL framework and introduces two key innovations:

1. **Private BN Patch Layers:**  
   The central idea is to treat BN layers as “patches” that remain private on each client. By decoupling these layers from the federated aggregation process, each client can retain personal statistics and trainable parameters (g and b) that better adapt to local data. The paper provides both theoretical insight and empirical evidence (e.g., Fig. 3 and Eqns. 6–10) to explain how these private patches help maintain consistency in neuron activations and enhance UA.
2. **Federated Optimisation Strategies:**  
   The paper evaluates several optimisation strategies, notably:
   * **FedAvg:** where clients perform local SGD and the server aggregates weights using a weighted average.
   * **FedAdam:** which uses Adam on the server by treating the difference between the previous and current global model as a pseudo-gradient.
   * **FedAvg-Adam:** a distributed variant in which clients employ Adam optimisers locally, and the server aggregates both model weights and Adam’s moment estimates.

The authors detail the algorithm (Algorithm 1) and present tables comparing communication rounds required to reach target UA under various configurations. This comprehensive experimental design underscores how MTFL, particularly when combined with FedAvg-Adam, can significantly reduce communication rounds compared to traditional FL.

The experimental setup is rigorous: datasets (MNIST and CIFAR10) are partitioned in a non-IID manner to simulate realistic client heterogeneity. Moreover, experiments also account for noisy client data, reinforcing the robustness of MTFL. By comparing against state-of-the-art personalised FL approaches (e.g., Per-FedAvg and pFedMe), the authors demonstrate that MTFL not only improves accuracy but also reduces computational and hyperparameter tuning overhead.

**Critical Assessment**

Mills et al. (2021) present an innovative approach with several strengths:

* **Novelty and Practical Relevance:** The integration of private BN patch layers within FL is a creative solution to the non-IID challenge and the need for personalisation. This is particularly relevant for real-world edge computing scenarios where data heterogeneity and privacy are paramount.
* **Empirical Rigor:** The extensive experimental validation—spanning simulated environments and an MEC testbed—lends credibility to the claims. The use of UA as a performance metric is a welcome shift toward more realistic evaluations.
* **Efficient Convergence:** The demonstrated improvements in convergence speed (up to a 5× reduction with FedAvg and an additional 3× with FedAvg-Adam) are significant. Reducing communication rounds is critical in environments with limited bandwidth and low-powered devices.

However, there are several areas where the work could be improved:

* **Scalability and Generalisation:** While the experiments on MNIST and CIFAR10 provide initial validation, these datasets are relatively small and well-studied. Future work could evaluate MTFL on larger, more complex datasets (e.g., ImageNet) and diverse architectures to better understand its scalability and generalisation properties.
* **Communication Overhead:** Although the authors note that the additional communication cost of FedAvg-Adam is offset by faster convergence, the trade-off between increased per-round communication and overall system efficiency warrants further investigation, especially in bandwidth-constrained settings.
* **Limited Scope of Private Layers:** The focus on BN layers is well justified given their small parameter footprint. Nonetheless, the exploration of other types of private layers or hybrid approaches could further enhance model personalisation while balancing the trade-offs between local and global performance.
* **Theoretical Analysis:** While the paper provides insights into the behaviour of BN layers (e.g., the effect on activation distributions), a more formal theoretical analysis—perhaps with convergence proofs or guarantees under varying degrees of non-IID data—would strengthen the contribution.

In comparison with other personalised FL methods (e.g., Per-FedAvg and pFedMe), MTFL stands out for its simplicity and reduced hyperparameter burden. However, the increased variance between client models and the need for additional infrastructure (e.g., managing private BN parameters) might pose challenges in large-scale deployments. Further research should explore these trade-offs in real-world settings with diverse and highly heterogeneous client populations.

**Conclusion**

Mills, Hu and Min (2021) make a significant contribution to the field of federated learning by proposing the MTFL algorithm, which effectively integrates multi-task learning principles via private BN layers to achieve personalised models in edge computing environments. By introducing the concept of User model Accuracy (UA) and leveraging an adaptive optimisation strategy (FedAvg-Adam), the paper addresses critical challenges associated with non-IID data and model convergence. Despite some limitations—such as the need for further scalability studies and a deeper theoretical analysis—the proposed approach represents a promising direction for personalised federated learning and is well supported by empirical evidence.

**References**

Hard, A., Rao, K., Mathews, R., Beaufays, F., Augenstein, S., Eichner, H., ... & Ramage, D. (2018) ‘Federated learning for mobile keyboard prediction’, *arXiv preprint arXiv:1811.03604*.

Mills, J., Hu, J. and Min, G. (2021) ‘Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing’, *IEEE Transactions on Parallel and Distributed Systems*, 33(3), pp. 630–641.

McMahan, B., Moore, E., Ramage, D., Hampson, S. and Arcas, B.A.Y. (2017) ‘Communication-efficient learning of deep networks from decentralized data’, in *Proc. 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282.

**O3 Mini High 5 essay merge**

**Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing: A Critical Analysis**

**1. Introduction**

Federated Learning (FL) has emerged as a paradigm shift in machine learning, enabling the training of models across decentralized devices while ensuring that raw data remains local. This distributed approach is especially vital in privacy-sensitive applications such as healthcare, finance, and personalized mobile services. Despite its advantages, conventional FL often struggles with non-Independent and Identically Distributed (non-IID) data, which can significantly degrade the performance of the aggregated global model. Moreover, a global focus tends to overlook individual user needs, making personalization an imperative yet challenging task.

In response to these challenges, Mills, Hu, and Min (2021) propose a Multi-Task Federated Learning (MTFL) algorithm that integrates private Batch Normalisation (BN) layers into the global deep neural network (DNN). This approach enables each client to retain personalized adaptations that account for local data peculiarities. The paper also introduces the User Accuracy (UA) metric—a measure more aligned with real-world objectives where individual performance is paramount. This essay critically examines the paper’s research questions, contributions, methodologies, and limitations, while also suggesting avenues for future research.

**2. Research Questions and Motivations**

The paper by Mills et al. (2021) addresses two central research questions:

1. **How can FL be adapted to overcome the challenges posed by non-IID data distributions and improve the accuracy of individual client models?**
2. **How can personalized learning be incorporated into FL frameworks without forfeiting the benefits of federated aggregation?**

These questions arise from the need to reconcile the tension between global model performance and client-specific accuracy. In traditional FL settings, a single aggregated model may underperform for individual users whose local data diverges significantly from the global distribution. This limitation is particularly evident in applications like personalized recommendations and mobile keyboard predictions, where individual user patterns are unique and critical. Additionally, the need to minimize data leakage and communication overhead in edge computing environments further motivates the search for efficient personalization strategies.

The paper’s focus on leveraging private BN layers as “patches” for personalization is both innovative and practical, as these layers represent a minimal increase in parameter overhead while offering substantial gains in local adaptation.

**3. Contributions**

Mills et al. (2021) make several significant contributions to the field of Federated Learning:

* **MTFL Algorithm:** The core innovation is the introduction of a Multi-Task Federated Learning algorithm that integrates private BN layers into the shared global DNN. This design allows each client to fine-tune aspects of the model (via BN parameters) to their specific data distributions, effectively transforming each local training task into a personalized learning problem.
* **User Accuracy (UA) Metric:** Recognizing that global accuracy does not fully capture user-specific performance, the authors propose the UA metric. UA, calculated as the average accuracy on each client’s local test data, provides a more realistic evaluation of model performance in personalized applications.
* **Adaptive Optimisation Strategy:** The paper explores variants of the traditional FedAvg algorithm, including adaptive methods like FedAvg-Adam, which leverage momentum and variance parameters to accelerate convergence. These strategies demonstrate reduced communication rounds and faster achievement of target UA levels.
* **Empirical Validation:** Extensive experiments are conducted on benchmark datasets (MNIST and CIFAR10) under non-IID conditions. In addition, the authors validate their approach on a real-world edge computing testbed using Raspberry Pi devices, underscoring the practical feasibility of MTFL.
* **Theoretical Insights:** Detailed analysis of the private BN layers elucidates how decoupling certain parameters from the federated averaging process preserves local activation distributions and enhances convergence, thereby bridging the gap between global and personalized performance.

**4. Research Methodology and Methods**

**4.1 Research Methodology**

The study employs an empirical methodology anchored in the standard iterative FL framework. The researchers first develop a mathematical formulation for MTFL, highlighting the integration of private BN layers that are excluded from the federated aggregation process. This theoretical foundation is followed by the design of a comprehensive experimental protocol that simulates non-IID conditions typical in real-world FL deployments.

Key aspects of the methodology include:

* **Algorithm Design:** Formulation of the MTFL algorithm that combines shared and private model components.
* **Implementation:** Integration of MTFL with existing FL frameworks, such as FedAvg, and its adaptive variants.
* **Simulation:** Creation of controlled experiments using MNIST and CIFAR10 datasets partitioned in a non-IID manner to mimic diverse client distributions.
* **Real-World Testing:** Deployment on a Raspberry Pi-based edge computing testbed to assess performance under practical resource constraints.

**4.2 Research Methods**

The experimental methods used in the study include:

* **Comparative Analysis:** The performance of MTFL is compared against traditional FL approaches (e.g., FedAvg) and state-of-the-art personalized FL methods (e.g., Per-FedAvg, pFedMe). This comparison focuses on key metrics such as UA and convergence speed.
* **Metric Evaluation:** The adoption of UA over traditional global accuracy metrics shifts the evaluation towards individual client performance.
* **Adaptive Optimisation:** Testing different optimisation strategies (FedAvg, FedAdam, and FedAvg-Adam) highlights the impact of adaptive methods in reducing communication rounds and accelerating convergence.
* **Edge Computing Validation:** Experiments on physical devices validate that the MTFL approach is not only theoretically sound but also practically viable in environments with limited computational and communication resources.

**5. Critique and Evaluation**

**Strengths:**

* **Innovation in Personalization:** The use of private BN layers to enable personalized learning is both novel and efficient. This approach minimizes extra computational and communication overhead, making it particularly suitable for resource-constrained edge devices.
* **User-Centric Metric:** The introduction of the UA metric provides a more nuanced understanding of model performance, emphasizing the importance of client-specific accuracy over global averages.
* **Robust Empirical Validation:** The combination of simulation experiments and real-world tests on Raspberry Pi devices lends strong credibility to the proposed method, demonstrating both its theoretical and practical merits.
* **Adaptive Convergence:** The exploration of FedAvg-Adam as an adaptive optimisation strategy effectively addresses the slow convergence issue commonly encountered in FL with non-IID data.

**Weaknesses:**

* **Scalability Concerns:** While the experiments on MNIST and CIFAR10 are promising, the study’s applicability to larger, more complex datasets and model architectures (e.g., ResNet or transformer-based models) remains unproven.
* **Hyperparameter Sensitivity:** The paper briefly mentions hyperparameter tuning but does not provide an in-depth analysis of how sensitive the MTFL algorithm is to these parameters. Such an analysis is crucial for practical deployment.
* **Communication Overhead:** Although adaptive methods reduce the number of rounds required for convergence, the potential increase in per-round communication costs—especially in bandwidth-limited settings—warrants further investigation.
* **Security and Robustness:** The work could be strengthened by a formal analysis of security aspects, including the resilience of the MTFL approach against adversarial attacks and malicious client behavior.

**Future Research Directions:**

* **Broader Model Evaluation:** Future studies should extend the evaluation of MTFL to more complex and larger-scale models, as well as to diverse application domains beyond image classification.
* **Alternative Personalization Mechanisms:** Exploration of additional lightweight personalization strategies—such as alternative normalization techniques or attention mechanisms—could further enhance FL performance.
* **In-depth Parameter Sensitivity Analysis:** A systematic study of hyperparameter sensitivity would provide valuable insights for deploying MTFL in varied real-world scenarios.
* **Security Enhancements:** Integrating differential privacy techniques or robust aggregation methods could fortify the approach against potential security threats in open FL ecosystems.

**6. Conclusion**

Mills et al. (2021) provide a significant advancement in the field of Federated Learning by introducing the Multi-Task Federated Learning (MTFL) algorithm, which effectively integrates personalized BN patch layers within a global model. By shifting the evaluation metric to User Accuracy (UA) and adopting adaptive optimization strategies like FedAvg-Adam, the paper addresses the dual challenges of non-IID data and personalization in edge computing environments. While certain limitations—such as scalability, hyperparameter sensitivity, and security—remain to be fully addressed, the proposed method represents a promising step toward more effective and personalized FL implementations.

The insights from this study not only contribute to the theoretical understanding of personalized federated learning but also offer practical solutions for real-world deployments. As FL continues to evolve, further research into scalable, secure, and efficient personalization techniques will be essential for realizing its full potential across diverse applications.

**References**

Dinh, C.T., Tran, N. and Nguyen, J., 2020. Personalised federated learning with Moreau envelopes. *Advances in Neural Information Processing Systems*, 33, pp.21394–21405.

Fallah, A., Mokhtari, A. and Ozdaglar, A., 2020. Personalised federated learning with theoretical guarantees: A model-agnostic meta-learning approach. *Advances in Neural Information Processing Systems*, 33, pp.3557–3568.

Hard, A. et al., 2018. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*.

McMahan, B. et al., 2017. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp.1273–1282.

Mills, J., Hu, J. and Min, G., 2021. Multi-task federated learning for personalised deep neural networks in edge computing. *IEEE Transactions on Parallel and Distributed Systems*, 33(3), pp.630–641.

Reddi, S.J. et al., 2021. Adaptive federated optimization. In: *International Conference on Learning Representations*. Available at: <https://openreview.net/pdf?id=LkFG3lB13U5> [Accessed 2 February 2025].