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

**Introduction**

Federated Learning (FL) aims to collaboratively train machine learning models while keeping sensitive user data on local devices. This decentralised method both enhances privacy and reduces communication overhead. However, non-Independent and Identically Distributed (non-IID) data remains a significant challenge, as a single global model generally struggles to capture the diverse data characteristics present across users. This limitation is particularly problematic in applications such as personalised smartphone keyboards or edge-based medical diagnostics, where each client’s data distribution is unique.

In their 2021 paper, Mills, Hu, and Min propose a Multi-Task Federated Learning (MTFL) approach that enhances user-specific model performance through localised Batch Normalisation (BN) layers. They also introduce a User Accuracy (UA) metric that evaluates performance on a per-client basis, thereby offering a clearer view of the benefits of personalisation. This essay critically examines the paper’s research questions, contributions, methodology, and limitations, situating these findings within the broader context of personalised FL.

**Research Questions and Motivations**

The paper addresses two key research questions:

1. How should FL handle non-IID data to improve each client’s accuracy?
2. How can FL incorporate personalisation while retaining the advantages of collaboration?

Standard FL aggregates parameters to form a global model. However, this generally results in local data disparities which can lead to suboptimal performance for individual clients. Consequently, any personalisation strategy must enhance client-specific accuracy without sacrificing the benefits of collaborative training or significantly increasing communication costs.

These research questions are driven by practical considerations. For example, smartphones require tailored solutions that still safeguard user privacy, while edge health diagnostics must contend with distinct data patterns and strict regulatory constraints.

Contributions

The paper makes four significant contributions:

1. Multi-Task Federated Learning (MTFL) Algorithm -  
   The authors propose a system where a global deep neural network (DNN) architecture is shared, but the BN parameters (mean and variance) are maintained locally. This design enables personalisation without compromising the collaborative FL framework.
2. User Accuracy (UA) Metric -  
   Unlike global accuracy metrics, the UA metric evaluates performance on a per-client basis before averaging, thereby revealing the true impact of personalisation on individual clients.
3. Adaptive Optimisation Strategies -   
   The paper examines momentum-based updates, such as FedAvg-Adam, which can accelerate convergence when dealing with skewed or imbalanced data—a particularly beneficial approach for resource-constrained devices.
4. Empirical and Theoretical Validation -  
   The authors validate their model using both standard benchmarks (MNIST and CIFAR10) and Raspberry Pis, demonstrating that localised BN personalisation is computationally feasible even on low-power hardware.

Research Methodology and Methods

Research Methodology

Their authors employ a typical FL pipeline where each client updates model parameters locally and a central server aggregates these updates periodically. The key innovation is the exclusion of BN parameters from aggregation. This is where each client retains its BN statistics to better capture localised data characteristics. They formally adjust the FL objective to treat BN parameters as private variables while sharing and averaging the remaining network layers.

To evaluate their approach, the authors introduce the UA metric alongside global accuracy, ensuring that improvements or regressions on individual devices are clearly reflected.

Research Methods

* Datasets and Non-IID Splits -  
  The authors partition MNIST and CIFAR10 to create skewed data distributions. For example, concentrating specific classes on particular clients or introducing heavy imbalances. This setup mimics the heterogeneity found in real-world scenarios.
* Algorithm Comparisons -  
  They compare the standard FedAvg with a momentum-based FedAvg-Adam, both with and without the personalised BN layers. This allows them to assess whether advanced optimisers can mitigate the effects of non-IID data.
* Edge Deployment -  
  A notable strength of the study is the validation on physical devices (Raspberry Pis). Metrics such as training time, energy consumption, and communication bandwidth were recorded, demonstrating that local BN personalisation is viable even under resource constraints.

By benchmarking their method against standard FL techniques, the authors confirm that isolating BN parameters yields improved stability and personalised performance without a significant communication penalty.

Critique and Evaluation

Strengths

1. Strong Personalisation -  
   Localising BN parameters is a relatively small component of the model, this provides a lightweight means of capturing individual data distributions without overhauling the entire FL process.
2. User-Centric Evaluation -  
   The introduction of the UA metric ensures that the approach is evaluated on real-world applicability. By focusing on each users performance. This study highlights the practical benefits of personalisation.
3. Accelerated Convergence -  
   The use of momentum-based updates (FedAvg-Adam) helps reduce the number of training rounds, which is an important factor for devices with limited power and bandwidth.
4. Real-World Validation -  
   Testing on Raspberry Pis strengthens the paper’s claims by showing that the method is not only theoretically sound but also practical for edge deployments.

***Relation to Other Approaches***

*The focus on local BN parameters aligns with similar strategies, such as FedBN (Li et al., 2021), which also exploits localised statistics to stabilise learning under non-IID conditions. In contrast, FedDWA (Liu et al., 2023) addresses data skew by dynamically adjusting client weights, while Zhong et al. (2023) highlight the limitations of BN under extreme heterogeneity, suggesting alternative normalisation methods. These complementary approaches illustrate the ongoing effort to balance global collaboration with client-specific adaptations in FL. Mills et al.’s MTFL approach stands out by minimally altering the standard FL pipeline, making it both practical and accessible for real-world edge applications.*

**Conclusion**

The author effectively demonstrate that localising Batch Normalisation parameters can significantly enhance personalised performance in Federated Learning, especially when dealing with non-IID data. By isolating BN parameters on each client, the model gains the flexibility to adapt to unique data distributions without compromising the benefits of federated aggregation. The introduction of the User Accuracy metric further emphasises the need for per-client performance evaluation rather than relying solely on a global average.

While the approach shows promise, evidenced by improved accuracy, faster convergence, and successful edge deployment, the paper leaves open questions regarding scalability, systematic hyperparameter tuning, communication efficiency, and security. Future research that integrates dynamic weighting, alternative normalisation techniques, or robust privacy measures could build on these findings to further refine personalised FL.

**References**

* Mills, J., Hu, J. and Min, G., 2021. Multi-task federated learning for personalised deep neural networks in edge computing. - <https://arxiv.org/abs/2007.09236>
* Zhong, Q., Zhao, T., & Fu, Y. (2023). On the Sensitivity of Batch Normalization in Heterogeneous Federated Learning.
* Li, X., Jiang, M., Zhang, X., Kamp, M., & Dou, Q. (2021). FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. International Conference on Learning Representations (ICLR).
* Liu, F., Yang, J., Wang, G., & Zhou, Z. (2023). FedDWA: Dynamic Weight Adjustment for Communication-Efficient Federated Learning.

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

**Introduction**

Federated Learning (FL) offers a promising way to train machine learning models collaboratively without forcing users to share raw data. By maintaining local ownership of sensitive information, FL alleviates privacy concerns and reduces communication overhead. Yet a major challenge arises when each client (e.g., a smartphone or edge device) has its own non-Independent and Identically Distributed (non-IID) data. A single global model often struggles to account for these heterogeneous data patterns. Personalising the model to each user has therefore become a central focus of federated learning research.

A representative example can be seen in smartphone keyboard predictions: each device’s vocabulary or style is unique, making a single model suboptimal for many users. Similarly, in medical diagnostics, patient data distribution might vary widely by region or by device. These scenarios highlight why an FL framework that emphasizes both collaboration and personalisation is increasingly in demand.

Mills, Hu, and Min (2021) propose a Multi-Task Federated Learning (MTFL) approach designed to address such needs. They incorporate local Batch Normalisation (BN) layers to boost client-specific performance and introduce a “User Accuracy” (UA) metric that foregrounds how well each individual client does. This paper critically examines the rationale, contributions, and potential pitfalls of that approach, as well as areas where more comprehensive solutions might emerge.

**Research Questions and Motivations**

Mills et al. (2021) investigate two pivotal questions: (1) how FL can handle non-IID data effectively so that each client’s accuracy improves, and (2) how to incorporate personalisation while still retaining the general advantages of federated collaboration. Typically, FL aggregates parameters from local training into one global model. But because every user has its own distribution, the global model often lacks the nuance to excel in each location. Consequently, the authors emphasize methods that enhance local accuracy while minimizing communication overhead.

Their motivations reflect pressing real-world concerns. Privacy regulations and resource constraints (e.g., battery life and bandwidth) demand a lightweight, secure solution. In parallel, industries that rely on personalization—like text prediction, healthcare, and recommendation systems—must ensure that FL can adapt to many diverse environments. This blend of performance, privacy, and practicality shapes the goals of the study.

**Contributions**

The paper’s chief contribution is a Multi-Task Federated Learning (MTFL) architecture in which all clients share a global Deep Neural Network (DNN), but maintain local BN parameters (means and variances) that are not aggregated. This strategy provides a “personalisation lever” without expanding the model’s core architecture. Complementing MTFL is a User Accuracy (UA) metric that averages client accuracies after each local evaluation, instead of relying on a global accuracy. By emphasizing per-client performance, the authors can more directly evaluate personalisation’s impact.

Additionally, they explore adaptive momentum-based optimization (FedAvg-Adam) to tackle skewed or imbalanced data. FL is susceptible to slow convergence, especially if client data vary substantially. By integrating momentum updates, the authors aim to reduce the number of training rounds needed, which can be critical in real-world edge scenarios with limited battery or bandwidth. Lastly, empirical tests on MNIST, CIFAR10, and physical Raspberry Pi devices confirm that local BN personalisation is computationally feasible even on modest hardware.

**Methodology**

Mills et al. follow a typical iterative FL pipeline, in which each client performs local training on its own data before sending model updates to a central server for aggregation. Their main innovation excludes BN parameters—local means and variances—from this global averaging step, letting each device preserve its environment-specific features. Thus, the BN layers are effectively “private,” while the rest of the network layers undergo federated updates as usual. This arrangement allows for personalisation without requiring entirely separate models for each client.

To measure efficacy, the authors introduce UA (User Accuracy), comparing local performance on a client’s data distribution before aggregating client accuracies across the network. This diverges from a more conventional global-accuracy benchmark, which might overlook how poorly a single model could perform on heavily skewed data.

They evaluate their system using MNIST and CIFAR10, splitting both data sets into non-IID partitions where each client sees either a subset of classes or a heavily imbalanced label distribution. They then compare standard FedAvg (no momentum, BN layers fully federated) with FedAvg plus local BN parameters, as well as an adaptive version (FedAvg-Adam) that includes momentum. Timing experiments on Raspberry Pis capture the real-world complexity of training with limited power and bandwidth.

**Critique, Limitations, and Future Directions**

**Strengths and Personalisation Value**  
The MTFL design demonstrates that isolating BN parameters can serve as an effective and lightweight approach to local customization. By focusing personalisation on BN layers (which comprise only a small fraction of total parameters), the overhead remains low. The authors’ introduction of UA also underscores the real goal—ensuring consistent performance across diverse user sets—rather than just inflating a global accuracy number.

**Experimental Breadth and Real-World Validation**  
Unlike purely theoretical or simulation-based studies, Mills et al. validate their model on physical Raspberry Pi devices. This practical element adds credibility to claims that local BN personalisation is not just a neat concept but is deployable within real-world edge constraints. Moreover, they demonstrate how momentum-based updates (Adam) can alleviate some of the difficulties posed by highly skewed local data, potentially speeding up convergence.

**Scalability to Complex Tasks**  
One limitation lies in focusing on moderately sized datasets (MNIST, CIFAR10). While helpful benchmarks, they do not fully represent the breadth of real-world tasks, which can include high-resolution imaging or large-vocabulary language models. Performance might shift when dealing with architectures like ResNet or Transformers. Exploring if group normalisation or layer normalisation scales better under FL conditions could further refine the approach.

**Hyperparameter Tuning**  
The authors recognize that local momentum, BN momentum, and learning rates must be handled carefully, but they do not present a fully systematic method for tuning them. This omission slightly hampers reproducibility and might deter adoption in production environments. Incorporating automated search or ablation studies could yield detailed best practices for future practitioners.

**Communication and Bandwidth Considerations**  
Because the authors show that local BN statistics remain small, they do not deeply investigate bandwidth overhead. Yet in large-scale FL scenarios, especially with thousands or millions of devices, even small transmissions accumulate. More advanced compression or communication-adaptive strategies (e.g., sending updates less frequently or only transmitting critical gradients) could enhance scalability without sacrificing local personalization benefits.

**Security and Robustness**  
FL faces unique threats like poisoning and inference attacks, but this paper only briefly touches on adversarial risk. Local BN parameters may actually reduce exposure of personal data patterns, but malicious clients could still degrade system performance by uploading bogus updates. Integrating secure aggregation or differential privacy could offer robust defenses. Future iterations of MTFL might incorporate these techniques, safeguarding local BN layers and reinforcing trust in a broader range of deployments.

**Relation to Other Approaches**

Recent methods also grapple with non-IID data in federated environments. FedBN (Li et al., 2021) similarly focuses on local BN to stabilize learning, echoing the idea of personalizing internal feature statistics. FedDWA (Liu et al., 2023) tackles data skew by adaptively adjusting client weights during aggregation, while Zhong et al. (2023) show that batch normalization may underperform in extremely heterogeneous settings, suggesting alternative strategies. The Mills et al. MTFL method stands out for its minimal disruption to standard FedAvg and its complementary usage of momentum-based optimizers, highlighting a pragmatic balance of simplicity and efficacy.

**Conclusion**

By localizing BN parameters, Mills et al. (2021) present a compelling route toward personalisation in federated learning. Users with non-IID data distributions benefit from client-specific adaptations, yet the advantages of collaborative training remain intact. The new User Accuracy metric appropriately reflects personalized performance, a key objective in real-world deployments.

The authors’ implementation on Raspberry Pis and their use of momentum-based FL hint at broad applicability for resource-constrained settings. Nonetheless, questions remain: how effectively does the approach scale to far larger datasets or modern DNN architectures? Will dynamic communication strategies or advanced privacy safeguards become essential in truly massive FL ecosystems? Exploring such avenues can further refine MTFL and help federated learning achieve robust, personalized, and secure solutions in practice.