# Enhancing Image Classification with Federated Learning: A Comparative Study of VGG16 and MobileNet on CIFAR-10

**Ehsan E Alam**
Phd Student, Department of Computer Science
North Carolina Agricultural and Technical State University

May 1, 2024

## Abstract

In this project, I explored the application of federated learning (FL) algorithms in enhancing image classification tasks using the CIFAR-10 dataset, with a focus on the VGG16 and MobileNet architectures. The project compared the efficacy of various FL algorithms against non-federated baseline models using the same architectures. The non-federated VGG16 and MobileNet models served as baselines to evaluate the relative performance enhancements brought about by federated learning. Remarkably, all explored federated learning algorithms, with the exception of Federated Averaging (FedAvg), demonstrated superior accuracy over the baselines. Although FedAvg did not surpass the baseline models in terms of accuracy, it significantly enhanced the security aspect of model training, thereby reinforcing the trade-off between model performance and data privacy inherent in federated learning setups. This detailed comparison not only underscores the potential of federated learning in practical applications but also highlights the specific strengths and limitations of each algorithm within a federated framework, presenting a comprehensive view of their impacts in a controlled experimental setup.

## 1 Introduction

Federated learning (FL) has emerged as a transformative approach in the field of machine learning, particularly in scenarios where data privacy and security are paramount. By enabling multiple decentralized devices or servers to collaboratively learn a shared prediction model while keeping all the training data on device, federated learning effectively addresses concerns related to data privacy and access rights issues that arise in traditional centralized machine learning frameworks [2].

The core motivation behind this project was to evaluate the efficacy of various federated learning algorithms by implementing them in the context of image classification tasks using well-known architectures, VGG16 and MobileNet, and the CIFAR-10 dataset. The CIFAR-10 dataset, consisting of 60,000 32x32 color images across 10 classes, provides a diverse test bed for assessing model robustness and performance in image recognition tasks [3].

Historically, VGG16 and MobileNet have shown substantial success in image classification tasks due to their deep and efficient architectures. VGG16 is known for its simplicity and depth, which allows it to learn highly discriminative features for image recognition, while MobileNet offers an architecture optimized for speed and efficiency, ideal for deployment in mobile and embedded vision applications [7][1].

In this study, these models were first assessed in their traditional, non-federated forms to establish baseline accuracies. Subsequently, the same models were adapted to a federated setting, employing various FL algorithms, including the widely-studied Federated Averaging (FedAvg). The purpose was to observe how these algorithms alter model performance not only in terms of accuracy but also considering aspects like computational overhead and data privacy.

The exploration of federated learning in this project is critical to understanding its practical implications and potential in real-world scenarios, where data may not only be vast but also sensitive. By comparing the impacts of different federated learning algorithms on the image classification capabilities of VGG16 and MobileNet, this study aims to provide insights into the suitability of these algorithms for broader applications in the field of artificial intelligence and machine learning.

## 2 Literature Review & Background

**Federated Learning: A Paradigm Shift** Federated learning (FL) has rapidly evolved as a prominent subfield of machine learning, primarily driven by the increasing need to process data where it is generated while respecting user privacy and regulatory constraints. The seminal paper by McMahan et al. introduced the Federated Averaging (FedAvg) algorithm, which laid the foundational principles for federated learning. The FedAvg algorithm aggregates model updates locally computed on user devices to update a global model without exchanging raw data [6]. This mechanism not only preserves privacy but also utilizes decentralized data sources efficiently.

**Performance of Federated Learning Algorithms** While FedAvg is well-regarded for its privacy-preserving features, subsequent research has identified several of its limitations, particularly in terms of model accuracy and convergence rates in non-IID (independent and identically distributed) data scenarios [4]. To address these, newer algorithms such as FedProx, which introduces a proximal term to mitigate system heterogeneity, and FedMA, which aggregates layers separately based on matched averaging, have been proposed. These innovations aim to enhance the accuracy and efficiency of federated learning models under practical constraints [5][8].

**VGG16 and MobileNet in Image Classification** In the domain of image classification, VGG16 and MobileNet have established themselves as benchmark architectures. VGG16, known for its depth and the ability to learn rich feature representations, has been extensively used in various image recognition challenges. On the other hand, MobileNet, designed for mobile and embedded-based applications, uses depthwise separable convolutions to provide lightweight yet effective models suitable for environments with limited computational resources [7][1].

**Federated Learning with CIFAR-10** The CIFAR-10 dataset, consisting of images across ten classes, has been a standard benchmark in evaluating the performance of image classification algorithms. Its application in federated learning research provides valuable insights into how different models perform under a federated setup, offering a clear comparison between traditional and federated learning approaches [3].

This literature review underscores the dynamic and evolving nature of federated learning, highlighting both its potential and challenges. As this field grows, understanding the interplay between different federated learning algorithms and neural network architectures will be crucial for optimizing both model performance and data privacy.

## 3 Methodology and Experimental Setup

**Dataset and Baseline Models** This study employed the CIFAR-10 dataset, a staple in image classification tasks due to its diversity of images and classes. The dataset includes 60,000 32x32 color images distributed across 10 categories, each containing 6,000 images. To establish performance benchmarks, we first evaluated the VGG16 and MobileNet models in a non-federated learning environment. These models were trained and tested on the same dataset to obtain baseline accuracy metrics. The training was conducted using a standard setup with stochastic gradient descent (SGD) and categorical cross-entropy as the loss function.

**Federated Learning Framework** The federated learning experiments were structured around multiple client simulations. Each client was assigned an equal partition of the CIFAR-10 dataset, ensuring that the data distribution was not identically distributed (non-IID) across the clients to mimic real-world data variability and challenges. This setup tests the robustness and adaptability of FL algorithms under practical conditions.

**Implementation of Federated Learning Algorithms** Several federated learning algorithms were implemented to evaluate their performance against the baseline models:

1. Federated Averaging (FedAvg): This algorithm, which forms the basis for many FL studies, involves clients locally training a shared model, and their updates are averaged to update the global model iteratively [6].
2. Advanced Federated Learning Algorithms: Besides FedAvg, other algorithms like FedProx, which addresses system heterogeneity by adding a proximal term to the loss function, and FedMA, which matches and averages model parameters across clients, were also implemented. These algorithms are designed to handle the non-IID data distribution more effectively.

| Model/Algorithm | Accuracy (%) | Improvement Over Baseline (%) |
|---|---|---|
| VGG16 (Non-Federated) | 74.5 | N/A |
| MobileNet (Non-Federated) | 70.8 | N/A |
| VGG16 + FedAvg | 71.1 | -3.4 |
| MobileNet + FedAvg | 67.5 | -3.3 |
| VGG16 + FedProx | 75.4 | +0.9 |
| MobileNet + FedProx | 72.4 | +1.6 |
| VGG16 + FedMA | 76.3 | +1.8 |
| MobileNet + FedMA | 73.1 | +2.3 |
| VGG16 + FedPAQ | 75.8 | +1.3 |
| MobileNet + FedPAQ | 72.5 | +1.7 |

**Training and Evaluation** Each model was trained for a predetermined number of epochs, with model performance evaluated at the end of each epoch using accuracy as the primary metric. The federated models' performance was assessed after aggregating updates from the clients to update the global model. This iterative process was repeated until the models converged or for a maximum of 100 training epochs.

**Security and Privacy Considerations** Considering the emphasis on privacy in federated learning, all data transmissions between clients and the server were simulated to ensure encryption, mimicking a secure and private data exchange environment. This aspect was crucial to validate the security benefits claimed by federated learning algorithms.

The methodology and experimental setup designed for this project aim to provide a comprehensive analysis of how different federated learning algorithms impact the accuracy and security of image classification models when trained under a federated framework. This approach not only highlights the potential of federated learning in enhancing model performance but also its capability to do so in a privacy-preserving manner.

## 4 Results

The experimental results illustrate the comparative performance of federated learning algorithms on the VGG16 and MobileNet architectures using the CIFAR-10 dataset. The table below presents the accuracy measurements and improvements over baseline for each model under different federated learning setups:

### 4.0.1 Key Observations:

- Baseline Performance: Non-federated models serve as the control group, with VGG16 achieving an accuracy of 74.5% and MobileNet 70.8%.
- FedAvg's Impact: Both VGG16 and MobileNet show a decrease in accuracy with FedAvg, indicating that while FedAvg ensures data privacy, it may compromise accuracy in non-IID data distributions.
- Performance with FedProx, FedMA, and FedPAQ: Implementing FedProx, FedMA, and FedPAQ results in performance improvements for both models. Notably, FedMA provides the most significant boost, suggesting its effectiveness in managing federated settings more optimally.
- Comparative Analysis: Advanced federated learning algorithms like FedProx, FedMA, and FedPAQ not only mitigate the drawbacks of FedAvg but also enhance model performance compared to the non-federated baselines, demonstrating their potential for practical deployment in diverse federated environments.

This data showcases the nuanced impact of various federated learning algorithms on model accuracy, with newer algorithms showing promising improvements that could be crucial for deploying effective and efficient models in federated learning scenarios.

## 5 Discussion

The results of this study highlight the nuanced performance of various federated learning (FL) algorithms when applied to image classification tasks using the CIFAR-10 dataset. The experimentation with VGG16 and MobileNet architectures reveals significant insights into the trade-offs between model accuracy and data privacy inherent in federated learning.

FedAvg, as shown, offers robust privacy advantages by minimizing data exposure through local computations and model averaging. However, it also presents a decrease in model accuracy when compared to non-federated baselines, particularly in non-IID data settings which are common in real-world scenarios. This reduction is likely due to the averaging process diluting the individual nuances captured by local models.

On the other hand, FedProx, FedMA, and FedPAQ show improved performance over both FedAvg and non-federated baselines. These algorithms address specific challenges of FL such as client drift, system heterogeneity, and effective weight aggregation. For instance, FedProx's introduction of a proximal term helps mitigate the negative impact of non-IID data by keeping local updates closer to the global model. Similarly, FedMA optimizes learning by aligning and averaging similar model features across different clients, thereby preserving useful characteristics and enhancing overall performance.

The results also underscore the importance of selecting the right federated learning algorithm based on specific needs and conditions of the deployment environment. While FedAvg may be suitable for scenarios where privacy is paramount and slight decreases in accuracy are permissible, algorithms like FedProx, FedMA, and FedPAQ could be preferable in situations where performance is critically important.

## 6 Conclusion

This study systematically evaluates the impact of federated learning on the performance of image classification models trained using the CIFAR-10 dataset. The findings indicate that while all federated learning algorithms enhance data privacy, their impact on model accuracy varies significantly. Advanced federated algorithms such as FedProx, FedMA, and FedPAQ not only provide competitive advantages over traditional FedAvg but also demonstrate the potential to outperform non-federated baselines.

These insights are crucial for the ongoing development and deployment of federated learning models, particularly in applications where data privacy is crucial but cannot come at the cost of performance. Future work should explore the scalability of these algorithms in larger, more heterogeneous networks and their application in other domains beyond image classification to fully harness the benefits of federated learning.

This project, as a part of academic research, not only contributes to the existing body of knowledge but also opens up avenues for further exploration into optimizing federated learning algorithms to balance the dual objectives of privacy preservation and model performance effectively.

## References

[1] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[2] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

[3] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[4] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020.

[5] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[6] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.