

FedWS: Uma Nova Abordagem para Aprendizado Federado usando Dados Heterogêneos

Flávio Vieira¹, Carlos Alberto V. Campos¹

¹Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro -- RJ -- Brasil

flaviovieira@edu.unirio.br, beto@uniriotec.br

Resumo. Drones suportam diversos tipos de aplicações envolvendo Sistemas Inteligentes de Transporte que visam entregar serviços seguros, sustentáveis e autônomos. Para isso, o Aprendizado Federado (Federated Learning - FL) provê aprendizado de máquina distribuído e privacidade para esses serviços. Entretanto, FL pode ser afetado negativamente pela heterogeneidade dos dados dos clientes. Neste artigo, é proposta uma técnica de FL para dados heterogêneos que suaviza localmente os pesos das camadas convolucionais de uma rede neural, com o objetivo de melhorar o resultado das rodadas de aprendizado. Os resultados mostram que a técnica proposta obteve uma acurácia com 3% à 6% superior e uma redução no custo de comunicação de 25% à 50% comparado as outras técnicas em tarefas de classificação de imagens no dataset EuroSAT.

Abstract. Drones support several applications involving Intelligent Transportation Systems that aim to deliver safe, sustainable, and autonomous services. Federated Learning (FL) provides secure, distributed machine learning for these services but is negatively affected by the heterogeneity of client data. In this paper, we propose an FL technique for heterogeneous data that smoothes the weights of the convolutional layers of a neural network locally, improving the result of FL rounds. The results show that our proposal achieved 3% to 6% higher accuracy and communication cost reduction of 25% to 50% compared to other techniques in image classification tasks in the EuroSAT dataset.

1. Introdução

Drones ou *Unmanned Aerial Vehicles* (UAVs) são poderosos equipamentos que com suas capacidades de comunicação com outros dispositivos, agilidade e versatilidade, se tornaram parte de várias aplicações em Sistemas Inteligentes de Transportes (ITS)[Butler et al. 2020]. Essas características permitem seu emprego em tarefas de mobilidade aérea envolvendo entrega de encomendas, serviços de emergência, vigilância ou monitoramento, e ainda, sistemas de transporte aéreo no perímetro urbano, de forma segura, sustentável, acessível, autônoma e inteligente. Para isso, necessidades e requisitos específicos de inteligência e segurança são de vital importância para o funcionamento dos UAVs [Cohen et al. 2021].

O Aprendizado Federado (FL) se apresenta como uma opção para a utilização em UAVs agregando valor aos serviços de mobilidade aérea, provendo aprendizado de

máquina de forma distribuída [McMahan et al. 2017], utilizando a capacidade de processamento dos dispositivos de borda e UAVs, além da proximidade destes com os dados a serem analisados. O resultado do aprendizado local do FL é repassado para o servidor de agregação, no sentido de construir um modelo global de aprendizado e sem a necessidade de transferência de dados brutos entre as partes.

Contudo, a utilização desta técnica tem suas próprias limitações e peculiaridades. Pois, os clientes que neste caso são UAVs, podem ter acesso a dados distribuídos de forma heterogênea, levando a uma baixa acurácia ou estagnação no aprendizado de máquina, o que pode afetar a qualidade e segurança dos serviços que se utilizam desta técnica [Li et al. 2021].

Segundo Zhu et al. (2021), atualmente, os efeitos colaterais causados pela heterogeneidade dos dados é um dos grandes desafios em FL, e assim, é proposta uma taxonomia dos tipos de técnicas para endereçar problema da heterogeneidade. Dentro das classes de técnicas que se utilizam de estratégias via transformação dos dados, as opções apresentadas são as técnicas de *Group Normalization* (GN) [Wu e He 2018] e *Batch Normalization* (BN) [Ioffe e Szegedy 2015]. Porém, analisando sobre o escopo de FL em ITS, o uso destas técnicas tem suas desvantagens. BN normalmente é afetada pela agregação realizada a nível do servidor central, resultando em valores que não representam corretamente os pesos e vieses globais. GN não é afetada por este tipo de situação, mas não consegue boa performance em valores de *batch size* elevados, cujo tamanho influencia de forma linear na quantidade de memória necessária para o processamento do treinamento, limitando o uso de dispositivos móveis, pois normalmente possuem memória limitada [Asad et al. 2021]. Em resumo, não há uma técnica que possibilite se sair melhor em distribuições com dados heterogêneos e, ao mesmo tempo, aproveitar da melhor forma a capacidade de processamento dos dispositivos móveis, sejam os que possuem maior memória disponível ou os mais reduzidos.

Motivados pela necessidade de avançar no tratamento de casos de dados heterogêneos em FL e levando também em consideração os requisitos de dispositivos móveis relativos a capacidade e no contexto de ITS, é apresentado, no presente trabalho, uma proposta de solução considerando esses fatores.

Nesse artigo é proposto o *Federated Learning Algorithm with Weight Standardization on Convolutional Layers* (FedWS), que reduz os efeitos negativos causados por dados heterogêneos em FL, melhorando a acurácia do treinamento federado, reduzindo o custo de comunicação federada e apresentando uma convergência maior mesmo nessas situações. Ele se utiliza da técnica de normalização dos pesos *Weight Standardization* (WS) [Qiao et al. 2019] aplicado às camadas convolucionais no treinamento local do FL, suavizando o comportamento dos gradientes e erros (*losses*) no contexto federado, visando reduzir a divergência entre os pesos notados nos casos com datasets heterogêneos. Além disso, foi adotada em conjunto a técnica de GN de forma a proporcionar uma convergência maior no treinamento e reduzir o custo de comunicação.

Assim, neste trabalho destacam-se as seguintes contribuições:

- Propor a técnica FedWS que avança no estado da arte do tratamento dos impactos negativos das distribuições de dados heterogêneos sobre FL, aplicando a técnica de normalização dos pesos, para funcionar em conjunto com a agregação fede-

rada e alcançando resultados expressivos de desempenho que serão mostrados nos resultados dos experimentos.

- Comprovar empiricamente que o uso de FedWS consegue suavizar o comportamento de FL em cenários com diferentes graus de heterogeneidade dos dados.
- Mostrar através de experimentos a superioridade da técnica comparada as demais técnicas de normalização no contexto de classificação de imagens processadas por UAVs.
- Analisar os benefícios da redução do custo de comunicação no FL em UAVs que FedWS proporciona, permitindo uma melhoria nos serviços atendidos.

Para evidenciar a utilidade do FedWS em aplicações móveis foram conduzidos experimentos de classificação de imagens usando o dataset EuroSAT [Helber et al. 2019].

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3, os fundamentos teóricos envolvidos no estudo são descritos. A Seção 4 apresenta o método proposto. A Seção 5 descreve os experimentos realizados e analisa os resultados obtidos. A Seção 6 apresenta a conclusão e proposta de trabalhos futuros.

2. Trabalhos Relacionados

Existem diversos estudos que tentam tratar o problema dos dados heterogêneos considerando a utilização de técnicas de normalização locais em aprendizado federado. FedBN [Li et al. 2021] propõe uso de BN sem levar os parâmetros específicos desta técnica para a agregação, mas de forma limitada em amplitude de uso. Fed2 [Yu et al. 2021] se propõe a utilizar uma técnica baseada em alocação de características com utilização de GN como camada de normalização. No entanto, os testes com dados heterogêneos consideraram somente a presença ou não de classes nos clientes, considerando somente 4 distribuições sem uma análise mais detalhada dos resultados sobre as distribuições. Em FedNorm [Du et al. 2022] é desenvolvido um estudo comparativo entre as técnicas de BN, GN e *Layer Normalization* (LN), apresentando esta última como a técnica que possui empiricamente os melhores resultados de acurácia e convergência. Contudo, alguns resultados mostraram que a técnica de GN possuía acurácia bem próxima a da LN, e sua convergência não foi comparada no estudo. Em resumo, não foi apresentada uma justificativa forte para não se utilizar o GN se os resultados foram tão próximos. Todos esses trabalhos também não consideraram o custo de comunicação no Aprendizado Federado como um dos fatores para análise.

Assim, diferentemente dos trabalhos anteriores, nossa proposta visa avançar o treinamento em FL com dados heterogêneos se utilizando de uma métrica que consegue graduar a heterogeneidade de forma a comparar o seu impacto em diferentes níveis via experimentos e análises, considerando a característica dos dados e seu efeito na acurácia e no custo de comunicação. Além disso, as características dos dados e seu efeito na precisão e custos de comunicação serão considerados, visando avaliar os impactos do seu uso em imagens de UAVs.

3. Fundamentação Teórica

3.1. Aprendizado Profundo de Máquina

O aprendizado de máquina é uma técnica que visa submeter dados de entrada à máquina de forma que aprenda um padrão com os dados e consiga produzir como saída, um con-

junto de valores que representam o que foi observado naquela etapa de aprendizado [LeCun et al. 2015]. Nesta operação, que é chamada de treinamento, estes valores são comparados com o gabarito dos dados, de forma que seja calculada a diferença entre o que máquina aprendeu e o que realmente o dado significa, chamada de erro.

Posteriormente a primeira etapa de treinamento, a máquina ajusta seus parâmetros internos, chamados pesos (*weights*) de forma a reduzir o erro. O componente que realiza esse treinamento são as redes neurais, formada por várias camadas compostas por neurônios que recebem como entrada parâmetros e pesos w , e produzem uma saída y que deve ser comparada com um limite b chamado de viés.

$$y_i = \sum_{i=1}^n w_i x_i + b \quad (1)$$

A cada etapa de treinamento os valores de y devem se aproximar cada vez mais do valor do gabarito, de forma a maximizar o acerto medido pela métrica acurácia. Esta métrica representa o quão perto do valor real, chegou a previsão da máquina.

3.2. Aprendizado Federado

O aprendizado federado é uma técnica distribuída de aprendizado de máquina que foi idealizada por McMahan et al. (2017). Esta técnica prevê um cenário no qual os clientes com capacidade de processamento possuem acesso a dados distribuídos e comunicação com um servidor central para o envio de informações para esses clientes, como mostrado na Figura 1.

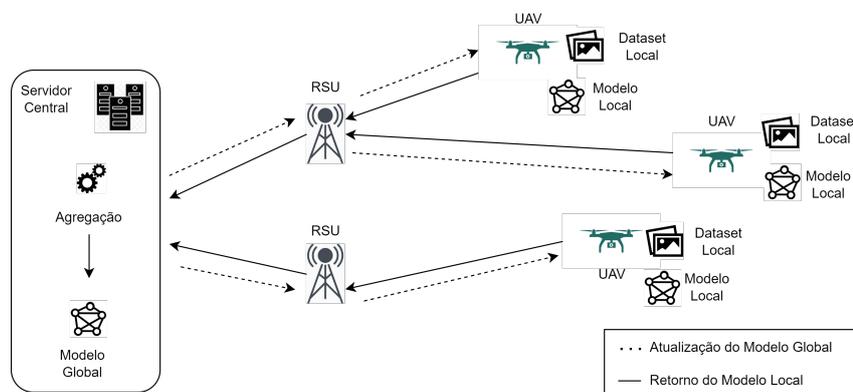


Figura 1. Treinamento em Aprendizado Federado

O treinamento federado se inicia com o envio, pelo servidor central, do modelo global para os clientes, que é composto pelos pesos e vieses atuais aprendidos pela máquina. Inicialmente, esse modelo irá com valores aleatórios de pesos e vieses, que a partir deste envio, os clientes realizarão tarefas locais de aprendizado de máquina sobre seus dados locais, gerando eles mesmos um modelo local correspondente justamente ao que cada cliente aprendeu em seu estágio de treinamento. Esses valores são então retornados para o servidor central que realizará a tarefa de agregar os pesos e vieses que cada cliente enviou. Caso o resultado esperado seja atingido o treinamento para, caso contrário uma nova rodada se inicia com o servidor central enviando o novo modelo global para os

clientes. Na proposta de McMahan et al. (2017) essa agregação é a média dos pesos e vieses dos clientes e por isso, essa técnica é chamada de FedAvg. No contexto de ITS, o treinamento seria realizado por UAVs que enviariam os modelos locais via *Roadside Unit* (RSU) para um servidor central realizar a agregação e gerar o modelo global.

3.3. Distribuição Heterogênea de Dados e non-IID

Distribuições heterogêneas de dados normalmente são comuns nas amostras utilizadas pelos clientes em FL [Ma et al. 2022] e o termo *not independent and identically distributed* (non-IID) denota um tipo muito especial de distribuição com heterogeneidade estatística que deve seguir vários comportamentos para ser confirmada.

No contexto deste artigo, será tratada a heterogeneidade dos dados em que os clientes (UAVs) de uma tarefa de FL possuem acesso a dados locais com desequilíbrio de representatividade entre suas classes. Conforme exemplificado na Figura 2, neste tipo de distribuição, cada cliente tem acesso a dados com somente parte das classes presentes no conjunto de dados disponíveis para todos os clientes e resultando em um desbalanceamento.



Figura 2. Ilustração de distribuição heterogênea em aprendizado federado

3.4. Custo de Comunicação em Aprendizado Federado

Um dos pontos mais críticos para o FL é o custo de comunicação que consiste na quantidade de dados transmitidos durante o processo de FL entre os clientes e o servidor. O *overhead* de comunicação pode ser calculado com a Equação 2 [Liu et al. 2021], onde T é o número total de rodadas (rounds) de aprendizado federado, multiplicado por 2 representando a ida e o retorno da informação, K é o número de clientes e ω^* é a quantidade de pesos do modelo local.

$$W = 2T(K.\omega^*) \quad (2)$$

Assim, a quantidade de dados trafegados neste contexto é fortemente influenciada pela convergência do algoritmo de FL, que pode demorar muito a atingir a acurácia desejada gerando valores muito grandes de T , pela quantidade de clientes K e pela complexidade da rede neural envolvida no treinamento local, que acaba definindo a quantidade de pesos ω^* trafegados.

3.5. Impacto da Divergência dos Pesos sobre FL

Segundo Zhao et al. (2018), a divergência dos pesos entre o aprendizado federado e o aprendizado convencional *Stochastic Gradient Descent* (SGD) influencia a acurácia em casos de distribuições de dados heterogêneos. Considerando os pesos $\omega_t^{(c)}$ $\omega_t^{(k)}$ como resultantes respectivamente do treinamento SGD e o treinamento federado, em uma situação de dados heterogêneos, a cada etapa de treinamento federado a divergência entre eles aumenta, pois a distância entre os pesos dos clientes k aumenta de forma cumulativa e maior que em cenários mais homogêneos.

Uma forma de medir o impacto de diferentes níveis de distribuições de dados heterogêneos na acurácia do treinamento de FL é a utilização da métrica *Earth Mover's Distance* (EMD) que é calculada como a relação entre as distribuições p de cada cliente k em relação a distribuição geral [Zhao et al. 2018]. Valores maiores de EMD representam distribuições heterogêneas de dados que geram um maior impacto na acurácia do FL e valores menores geram os menores impactos. Considerando o custo de comunicação em FL, o aumento da divergência e das medidas de EMD pode causar desde um aumento de etapas T de treinamento federado para se chegar a acurácia desejada até a resultados de acurácia baixos, afetando diretamente o custo de comunicação W , inviabilizando a operação de FL.

4. Método Proposto

Nesta seção é detalhado o método proposto que é chamado Algoritmo de Aprendizado de Máquina com Normalização de Pesos nas Camadas Convolucionais (*Federated Learning Algorithm with Weight Standardization on Convolutional Layers* - FedWS).

Considerando o efeito negativo que as distribuições heterogêneas possuem na divergência dos pesos visto na Seção 3.5, tem-se que o mesmo é cumulativo a cada rodada de treinamento e afeta primeiramente o treinamento local dos clientes e, posteriormente, a agregação. O treinamento local em FL se dá conforme mostrado na Equação 3, onde ω são os pesos locais, η é a taxa de aprendizado, ∇ é o gradiente descendente e $l(w; b)$ é o erro da etapa de treinamento local. O gradiente é responsável por informar ao treinamento em que direção e qual o tamanho da correção que será utilizada na próxima etapa, de forma a minimizar os erros e maximizar a acurácia. Em casos de distribuições heterogêneas para FL, o gradiente direciona notadamente o treinamento local a uma situação de divergência de pesos, influenciando em como os pesos se comportarão em etapas posteriores ao treinamento.

$$\omega \leftarrow \omega - \eta \nabla l(w; b) \quad (3)$$

Uma das formas de melhorar as decisões do gradiente nesta situação é normalizar a sua variação. No contexto de aprendizado profundo, *Weight Standardization* (WS) é uma técnica que consegue suavizar a variação dos gradientes e erros nas camadas convolucionais de uma CNN [Qiao et al. 2019]. Isso permite que as decisões do otimizador em direção ao gradiente não gerem resultados tão drásticos nos erros e gradientes, permitindo uma convergência mais suave no treinamento. Esta técnica atua diretamente nos pesos de saída da camada convolucional, aplicando sobre os pesos uma normalização se utilizando de média (μ) e desvio-padrão (σ) do conjunto de pesos conforme a Equação 4.

$$\omega^{ws} = \frac{\omega - \mu\omega}{\sigma\omega} \quad (4)$$

O objetivo do FedWS é se utilizar da normalização proposta por WS nas camadas convolucionais, suavizando os valores dos erros e gradiente no treinamento local, para que a agregação reflita essa característica no modelo global do aprendizado federado e, consequentemente, reduzindo os saltos dos pesos nas fazes seguintes, que acabam agravando as situações de divergência de pesos descritas na Seção 3.5. A partir deste uso, o treinamento local do FL nas camadas convolucionais ficaria como mostra a Equação 5.

$$\omega \leftarrow \omega - \eta \nabla l(\omega^{ws}; b) \quad (5)$$

Na agregação dos pesos dos clientes no servidor para as camadas convolucionais a aplicação de ω^{ws} é mostrada na Equação 6, onde n é o conjunto total de amostras do dataset, n_k é o conjunto de amostras para o cliente e ω^{wsk}_{t+1} são os pesos que os clientes enviaram provenientes do treinamento local.

$$\omega^{ws}_{t+1} = \sum_{k=1}^K \frac{n_k}{n} \omega^{wsk}_{t+1} \quad (6)$$

A normalização realizada na Equação 4 é aplicada diretamente sobre os pesos da camada convolucional, e não faz parte dos parâmetros que são alvo de aprendizado do treinamento local. Portanto, ela não sofre de anomalias quando ocorre a agregação dos parâmetros e pesos no nível do modelo global, ao contrário do que ocorre com BN.

Outra condição que impacta o resultado do FL é a distribuição não uniforme dos pesos (covariance shift) nas camadas do treinamento local. Nesse caso, FedWS utiliza-se de uma camada de normalização (GN) posterior a aplicação da técnica da Equação 4, aumentando a convergência no treinamento.

5. Experimentos

Como forma de comprovar as melhorias propostas por FedWS, experimentos de classificação de imagens, em cenários onde os dispositivos encontram uma situação de níveis diferentes de distribuições de dados heterogêneos foram realizados. O dataset utilizado foi o EuroSAT [Causa et al. 2023] que possui um conjunto de 27.000 imagens capturadas via satélite, de dimensões de 64X64 pixels e divididos em 10 classes, que representam diferentes tipos de terrenos (plantações permanentes e temporárias, rodovias, rios, florestas, pastos, lagos, vegetação diversa, áreas industriais e residenciais).

Primeiramente são mostradas a configuração dos experimentos e do dataset. Em seguida, são apresentados e analisados os resultados da classificação de imagens e do custo de comunicação em FL.

5.1. Configuração e Dataset

A máquina utilizada para o treinamento federado foi um notebook com processador Core i5, 16 GB RAM de memória e placa gráfica Nvidia GTX 1650 com 4 GB de memória e suporte a CUDA. O Experimento utilizou um *framework*¹ que foi alterado para permitir

¹<https://github.com/c-gabri/Federated-Learning-PyTorch>

o uso da nova técnica (FedWS) e realização dos experimentos, utilizando-se de ambiente Python 3.8 e Pytorch 1.19.

O dataset EuroSAT foi configurado com a proporção de 80% para treino e 20% para validação e testes. Sua distribuição dos dados heterogêneos foi realizada por meio de cinco arranjos de dados representando valores de EMD 0, 0,37, 0,74, 1,08 e 1,41 [Zhao et al. 2018], onde os clientes tem desbalanceamento de classes em seus dados.

Os treinamentos federados foram configurados com número de rodadas $T = 200$, taxa de aprendizado $\mu = 0,01$, tamanho de *batch size* do treinamento local $B = 50$, 5 épocas de treinamento local e número de clientes $K = 10$. Seu modelo local é uma CNN com a arquitetura definida por [McMahan et al. 2017]. O tamanho de B foi escolhido de forma a representar um valor grande considerando que o treinamento local será em um dispositivo UAV e que WS possui resultados comprovadamente bons em tamanhos de B muito pequenos [Qiao et al. 2019].

Os cenários considerados foram: Sem Normalização (SN), onde não foi aplicada nenhuma técnica de normalização no modelo, três outros que representam as técnicas utilizadas pelos trabalhos relacionados: FedBN(BN) [Li et al. 2021], Fed2(GN) [Yu et al. 2021] e FedNorm (LN) [Du et al. 2022] e, finalmente, um cenário com o uso do FedWS, executados nas 5 distribuições do dataset distintas de EMD previstas.

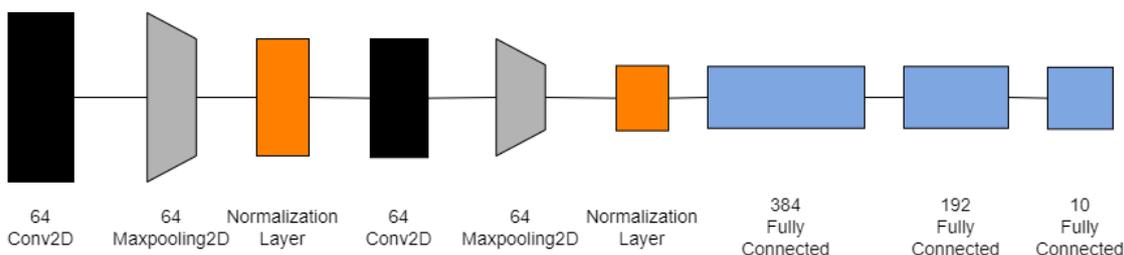


Figura 3. Arquitetura da CNN utilizada nos experimentos

5.2. Classificação de Imagens em Dados Heterogêneos

Aqui são abordados os resultados na tarefa de classificação de imagens no dataset EuroSAT. A Tabela 1 mostra que FedWS conseguiu melhores resultados que as demais técnicas com uma margem de 3% a 6% de acurácia nos diferentes cenários de EMD selecionados.

Tabela 1. Resultados de acurácia (%) em classificação EuroSAT

EMD	0	0,37	0,71	1,08	1,41
SN	75,7 (0,2)	74,6 (0,3)	73,3 (2,1)	69,8 (0,8)	61,7 (1,1)
BN	82,4 (0,2)	79,6 (0,7)	79,5 (0,7)	75,0 (0,3)	69,7 (3,1)
GN	81,1 (0,3)	80,4 (0,1)	79,4 (0,9)	74,0 (0,1)	69,2 (0,8)
LN	80,4 (0,2)	79,9 (0,3)	78,6 (1,1)	73,3 (0,3)	68,8 (0,1)
FedWS	85,5 (0,6)	86,0 (0,6)	85,0 (0,8)	80,7 (0,3)	76,9 (0,1)

Desvio padrão entre parênteses

Interessante salientar, que as outras técnicas se revezaram como a segunda melhor colocada nos resultados, variando nos níveis de EMD. Isso mostra que FedWS possui uma estabilidade maior nos resultados, independente da distribuição utilizada. Outro ponto

importante é que a diferença favorável a FedWS se iniciou com valor de 3% para o EMD 0 e aumentou para o patamar de 6% se mantendo nesse linear mesmo em EMD 1,41, que representa a distribuição com o maior impacto na acurácia, o que corrobora para o entendimento de sua maior estabilidade. Considerando a diferença entre o maior valor de EMD (1,41) e menor valor (0) para cada técnica, FedWS também se destacou, pois seu resultado foi de 8,6% enquanto as outras técnicas ficaram entre 11,6% e 14%.

Outro ponto importante a ser observado é a convergência das técnicas. A Figura 4 mostra o detalhamento da utilização das técnicas de normalização em FL sobre o dataset EuroSAT com EMD 0,71 e 1,41 a partir da rodada 20. No primeiro caso, FedWS conseguiu chegar a acurácia de 75% em 60 rodadas de treino enquanto que os demais só conseguiram alcançar esse nível a partir da rodada 140. Com EMD 1,41, FedWS conseguiu valores acima de 70% em 80 rodadas e o restante somente chegou perto desta acurácia em 200 rodadas.

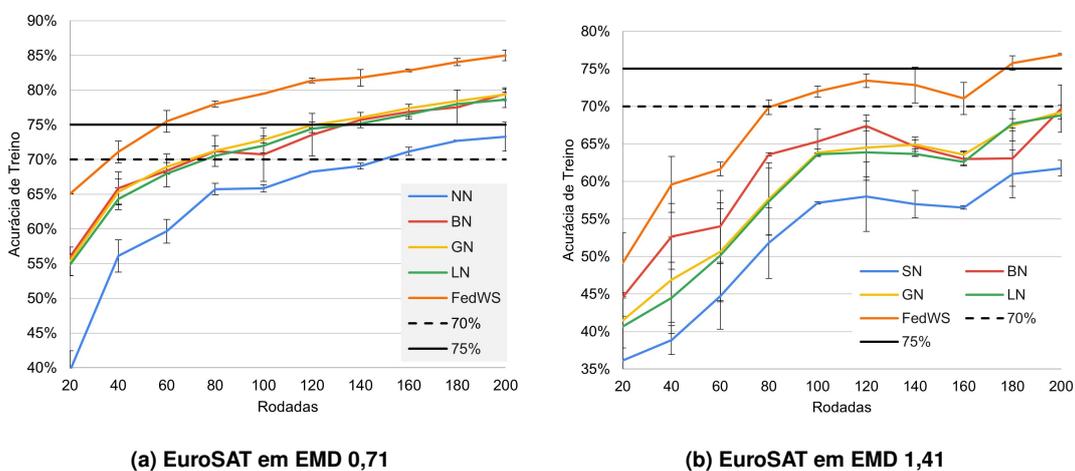


Figura 4. Análise de convergência

A relação entre diferença da acurácia de treino e teste pode indicar tendências a *overfitting* no aprendizado de máquina, situações em que o resultado do treino superou em muito o teste, sinalizando um modelo que não consegue generalizar bem em situações reais de uso.

Tabela 2. Relação percentual treino e teste de classificação EuroSAT (EMD 0,79)

Descrição	SN	BN	GN	LN	FedWS
Média	5,3%	5,5%	2,3%	3,1%	2,1%
Desvio-padrão	1,5%	1,6%	1,1%	1,1%	1,1%

A Tabela 2 mostra a relação percentual entre as acurácias de treino e teste para um EMD 0,79, com o objetivo de caracterizar como as diferentes técnicas se comportam nesse cenário. FedWS possui a menor relação média bem próximo a GN, enquanto que SN, BN e LN apresentaram valores maiores podendo indicar um certo nível de *overfitting*.

Um dos objetivos do FedWS é suavizar a curva de erro do treinamento FL e ao observarmos os resultados da Figura 5, fica claro que nossa proposta se confirmou, possuindo a curva mais suavizada e com valores menores em comparações as outras técnicas

tomando como amostra as distribuições com EMD 0,74 e 1,41. A suavização do erro impactou positivamente em situações de heterogeneidade de dados, indicando que o gradiente trabalhou de forma suave, evitando grandes saltos no processo de aprendizado e reforçando os resultados de acurácia superiores de FedWS mostrados na Tabela 1.

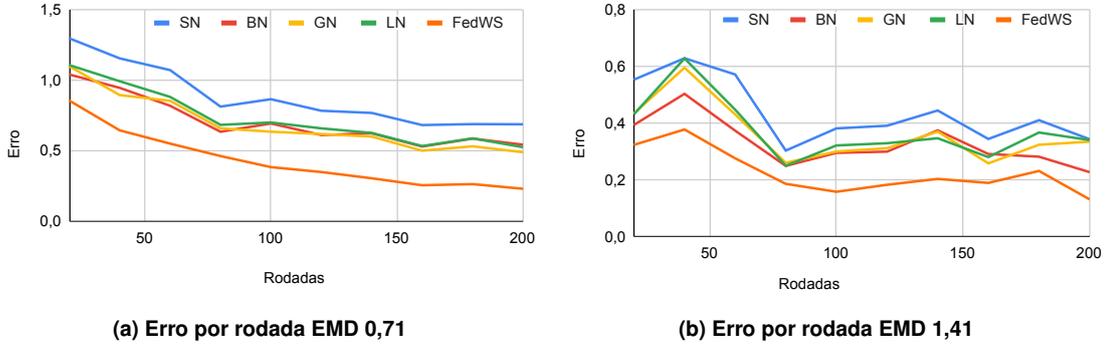


Figura 5. Erros EuroSAT

5.3. Custo de Comunicação de Aprendizado Federado

Considerando os custos de comunicação definido na Seção 3.4 como W , pode-se calcular o valor baseado nos valores das rodadas de treinamento T , do número de clientes K e dos pesos dos modelos locais ω . Partindo do modelo da Figura 3, a quantidade em bytes que o cliente retorna como modelo local, medidas em experimento para todos os casos é de 4,27MB. Isso mostra que FedWS não aumentou a complexidade em tamanho do modelo de forma a penalizar o tamanho do modelo local transferido para o servidor central. Com isso, como existem $K = 10$ clientes nos experimentos, o valor de W é de 42,7MB $\times T$. Ou seja, a técnica que apresentar maior convergência normalmente necessitará de menos rodadas T para chegar a um determinado valor de acurácia e terá um valor menor do custo W .

As Tabelas 3 e 4 mostram respectivamente os resultados do custo W para as técnicas envolvidas nos experimentos, com EMDs distintos visando atingir 70% e 75% de acurácia respectivamente. Esse valores alvo foram escolhidos baseado no média e valor máximos aproximados dos resultados das técnicas com EMD 1,41, mostrados na Tabela 1 e também destacados na Figura 4.

Os resultados mostram que FedWS consegue atingir acurácia de 70% e 75% gastando muito menos rodadas que os restante das técnicas observadas. Na faixa de 70% FedWS conseguiu redução de custos entre 25% a 50% e na faixa de 75% entre 42% a 50%, comparado as outras técnicas. Em alguns casos de EMD maiores somente ele conseguiu atingir a acurácia exigida. Isso resulta em redução de custos para aplicações em UAVs, considerando gastos de energia, banda de transmissão e tempo de transmissão, permitindo também que os clientes sejam liberados em menos tempo de suas tarefas.

6. Conclusão

Neste trabalho, é proposto FedWS, uma abordagem que permite reduzir o custo do aprendizado federado (FL) em situações que os dados estão distribuídos em diferentes graus de heterogeneidade. A normalização dos pesos das camadas convolucionais nos treinamentos locais permitiu que os erros ficassem suavizados no nível do agrupamento federado,

Tabela 3. Custo de comunicação (MB) visando alcançar 70% de acurácia

EMD	0	0,37	0,71	1,08	1,41
SN	5.124	5.978	6.832	NA	NA
BN	1.708	1.708	2.562	3.416	8.540
GN	4.270	1.708	2.562	5.124	8.540
LN	2.562	2.562	3.416	6.832	NA
FedWS	854	854	1.708	2.562	4.270

Tabela 4. Custo de comunicação (MB) visando alcançar 75% de acurácia

EMD	0	0,37	0,71	1,08	1,41
SN	7.686	8.540	NA	NA	NA
BN	3.416	4.270	5.978	NA	NA
GN	2.562	5.978	5.978	NA	NA
LN	4.270	5.978	5.978	NA	NA
FedWS	2.562	2.562	2.562	3.416	7.686

proporcionando ao otimizador condições de obter melhores resultados a cada rodada de treinamento, mesmo em situações de distribuições de dados heterogêneos. Além disso, o uso em conjunto da técnica de normalização GN permitiu uma maior convergência do treinamento devido a normalização dos pesos posteriormente as camadas convolucionais da rede neural. Os resultados experimentais nos mostraram que o FedWS conseguiu reduzir o impacto da distribuição heterogênea, resultando em valores de acurácia na classificação de imagens de 3% a 6% superiores a todas as técnicas comparadas em diferentes graus de heterogeneidade dos dados. O FedWS também apresentou uma menor diferença entre a acurácia de treino e teste, gerando modelos mais propensos a generalização e conseguiu uma convergência muito superior, com resultados de acurácia em menos rodadas que os demais, além de suavizar os erros no aprendizado federado no nível de agregação, mesmo em distribuições com alto nível de heterogeneidade. Finalmente, nossa proposta foi capaz de reduzir o custo de aprendizado federado de 25% a 50% se comparado as demais técnicas.

Como trabalhos futuros, pretende-se aprofundar o estudo da técnica FedWS envolvendo situações em que os clientes sejam heterogêneos, permitindo quantidades diferentes de épocas nos treinamentos locais e analisar o seu efeito no custo de comunicação, diferentemente, do que foi mostrado neste artigo. Além disso, pretende-se aplicar o FedWS na análise de outros datasets de imagens.

Referências

- Asad, M., Moustafa, A., Ito, T., and Aslam, M. (2021). Evaluating the communication efficiency in federated learning algorithms. In *24th IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 552–557.
- Butler, L., Yigitcanlar, T., and Paz, A. (2020). Smart urban mobility innovations: A comprehensive review and evaluation. *IEEE ACCESS*, 8:196034–196049.
- Causa, F., Franzone, A., and Fasano, G. (2023). Strategic and tactical path planning for urban air mobility: Overview and application to real-world use cases. *Drones*, 7(1):11.

- Cohen, A. P., Shaheen, S. A., and Farrar, E. M. (2021). Urban air mobility: History, ecosystem, market potential, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(9):6074–6087.
- Du, Z., Sun, J., Li, A., Chen, P.-Y., Zhang, J., Li, H. H., and Chen, Y. (2022). Rethinking normalization methods in federated learning. In *3rd International Workshop on Distributed Machine Learning*, pages 16–22.
- Helber, P., Bischke, B., Dengel, A., and Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, pages 448–456.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Li, X., Jiang, M., Zhang, X., Kamp, M., and Dou, Q. (2021). Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*.
- Liu, S., Yu, J., Deng, X., and Wan, S. (2021). Fedcpf: An efficient-communication federated learning approach for vehicular edge computing in 6g communication networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):1616–1629.
- Ma, X., Zhu, J., Lin, Z., Chen, S., and Qin, Y. (2022). A state-of-the-art survey on solving non-iid data in federated learning. *Future Generation Computer Systems*, 135:244–258.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282.
- Qiao, S., Wang, H., Liu, C., Shen, W., and Yuille, A. (2019). Micro-batch training with batch-channel normalization and weight standardization. *arXiv preprint arXiv:1903.10520*.
- Wu, Y. and He, K. (2018). Group normalization. In *European Conference on Computer Vision (ECCV)*, pages 3–19.
- Yu, F., Zhang, W., Qin, Z., Xu, Z., Wang, D., Liu, C., Tian, Z., and Chen, X. (2021). Fed2: Feature-aligned federated learning. In *27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2066–2074.
- Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., and Chandra, V. (2018). Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*.
- Zhu, H., Xu, J., Liu, S., and Jin, Y. (2021). Federated learning on non-iid data: A survey. *Neurocomputing*, 465:371–390.