

SCOPE-FL: Seleção de Clientes por Ordem de Entropia no Aprendizado Federado

Isaque O. Silva¹, Carlos Vitelli¹, Iago Medeiros¹

¹*Universidade Federal do Pará (UFPA)*

isaque.silva@tucurui.ufpa.br

carlos.vitelli@itec.ufpa.br

iagomedeiros@ufpa.br

Resumo

A crescente utilização de dispositivos conectados exige novos métodos para lidar com a quantidade e privacidade dos dados compartilhados. Federated Learning (FL) surge como uma solução, permitindo o treinamento de modelos sem compartilhar dados diretamente, preservando a privacidade dos clientes. No entanto, nem todos os clientes são igualmente úteis para o aprimoramento de modelos globais, tornando necessária uma seleção eficiente de clientes. O SCOPE-FL propõe um mecanismo dinâmico de seleção de clientes, atribuindo pesos à entropia dos dados e ao tamanho do dataset, para garantir uma contribuição mais eficiente para o modelo global. Isso é feito calculando uma pontuação de relevância para cada cliente, com base nesses fatores, e ajustando os pesos atribuídos a cada cliente. O SCOPE-FL usa o método FedAvg para agregar modelos locais, priorizando clientes com dados mais relevantes. Após simulações utilizando o dataset MNIST, o SCOPE-FL superou métodos tradicionais, mostrando uma taxa de acurácia superior a 60% após 12 rodadas, alcançando até 80% em 22 rodadas.

Abstract

The increasing use of connected devices requires new methods to handle the quantity and privacy of shared data. Federated Learning (FL) emerges as a solution, enabling model training without directly sharing data, preserving the clients' privacy. However, not all clients are equally useful for improving global models, making efficient client selection necessary. SCOPE-FL proposes a dynamic client selection mechanism, assigning weights to data entropy and dataset size to ensure a more efficient contribution to the global model. This is done by calculating a relevance score for each client based on these factors and adjusting the weights assigned to each client. SCOPE-FL uses the FedAvg method to aggregate local models, prioritizing clients with more relevant data. Tested with MNIST, SCOPE-FL outperformed traditional methods, showing an accuracy rate of over 60% after 12 rounds, reaching up to 80% after 22 rounds.

1 Introdução

O avanço da Internet das Coisas (IoT), da computação ubíqua e de sistemas distribuídos vem transformando radicalmente a forma como dados são gerados, processados e

utilizados. Estima-se que, até 2030, mais de 500 bilhões de dispositivos estarão conectados, formando uma rede heterogênea de sensores, dispositivos móveis, wearables, veículos inteligentes e sistemas industriais.

Esse fenômeno facilita a integração de ataques direcionados aos dados e invasões que comprometem a segurança, resultando na perda de informações sensíveis dos usuários. Quando os dados precisam ser transmitidos constantemente entre dispositivos e servidores centrais, a superfície de ataque aumenta consideravelmente, abrindo espaço para espionagem, vazamentos e manipulação de informações.

O crescimento acelerado da Internet das Coisas (IoT), que conecta milhões de dispositivos heterogêneos, intensifica esse cenário, pois amplia o número de pontos vulneráveis e torna a proteção da privacidade ainda mais desafiadora. Ao mesmo tempo, os modelos de aprendizado de máquina dependem de grandes volumes de dados para alcançar bons resultados, mas os métodos tradicionais exigem que essas informações sejam centralizadas em um servidor. Essa prática implica custos elevados de comunicação e expõe os usuários a riscos adicionais de privacidade [AbdulRahman et al. 2020].

O desafio, portanto, não está apenas em treinar modelos com precisão, mas em equilibrar desempenho, eficiência e segurança. Esse contexto abre espaço para novas abordagens, como o Aprendizado Federado (Federated Learning - FL), que possibilita o treinamento distribuído sem a necessidade de transferir dados brutos, reduzindo a exposição de informações e preservando a privacidade dos usuários. Diante disso, o obstáculo não está apenas em obter precisão nos modelos, mas também em equilibrar desempenho, eficiência e segurança. Esse cenário abre espaço para novas abordagens, como o Aprendizado Federado, que permite treinar modelos distribuídos sem transferir os dados brutos, preservando a privacidade dos usuários.

O Aprendizado Federado (Federated Learning – FL) [McMahan et al. 2017] surge como um paradigma capaz de responder a esses desafios. Em vez de centralizar os dados em um servidor, o FL permite que os modelos sejam treinados localmente nos dispositivos clientes, que compartilham apenas parâmetros ou gradientes com um servidor agregador. Essa proposta reduz a necessidade de armazenamento centralizado de dados para o treinamento de modelos de aprendizado de máquina, contribuindo para a preservação da privacidade dos clientes. Além disso, essa abordagem favorece a generalização, pois o aprendizado é desenvolvido diretamente nos dispositivos de borda, locais onde os dados são gerados, diminuindo significativamente a quantidade de informações que precisam ser enviadas a um servidor central.

No entanto, no cenário de FL, nem todos os clientes são necessariamente ideais para o desenvolvimento das aplicações. Em muitos casos, os dados locais podem ser pequenos, enviesados ou pouco diversificados, o que limita sua contribuição para o modelo global. Nessas situações, é mais eficiente priorizar clientes que forneçam informações mais relevantes, equilibrando quantidade e diversidade, de modo a garantir uma convergência mais estável e resultados mais robustos.

O papel central do FL na melhoria da acurácia durante o treinamento de aprendizado de máquina está no foco de generalização do modelo global, de forma que qualquer novo participante possa utilizar o modelo global atualizado [Pires et al. 2020]. Portanto, ao escolher os clientes, variando com base na quantidade do seu *dataset* ou no desempenho do seu treinamento em relação à perda do modelo global, temos argumentos que são interdependentes para aprimorar a arquitetura.

Utilizar menos clientes pode reduzir a necessidade de mais iterações, permitindo que o modelo se estabilize mais rapidamente. No entanto, essa prática também traz contras:

ao diminuir a quantidade de participantes, há risco de perder diversidade de dados, o que pode gerar vieses no modelo global e comprometer sua capacidade de generalização. Assim, a seleção de clientes deve buscar um equilíbrio entre eficiência e representatividade.

Neste artigo, buscamos trabalhar com a seleção de clientes de forma mais genérica, a fim de evitar exclusões desnecessárias e garantir uma relativa generalização do modelo. Ou seja, o SCOPE-FL auxiliará o servidor a escolher os clientes sem necessariamente excluir sua participação, mas atribuindo pesos para que sempre haja a possibilidade de o cliente ser selecionado. Essa abordagem contribui tanto para a melhoria da acurácia quanto para a generalização, pois, ao atribuir pesos à Entropia dos dados do cliente e relacionar um peso secundário ao tamanho do seu *dataset*, garantimos uma especialização ao treinar os clientes de forma genérica e asseguramos uma estrutura mais palpável desse cenário.

O SCOPE-FL também apresenta grande potencial de aplicação em cenários reais, como na saúde, onde hospitais podem treinar modelos de diagnóstico sem expor dados sensíveis; em veículos autônomos, possibilitando que frotas compartilhem aprendizado de sensores sem transmitir dados brutos; e em sistemas de IoT e cidades inteligentes, em que sensores urbanos colaboram em previsões de tráfego ou consumo de energia de forma descentralizada. Setores financeiros também podem se beneficiar ao treinar modelos de detecção de fraude preservando a confidencialidade dos clientes. Nessas situações, a entropia contribui para maior diversidade de amostras e a relevância adaptativa do SCOPE-FL promove balanceamento eficiente, acelerando a convergência e ampliando a representatividade do modelo global.

2 Fundamentação Teórica

O Aprendizado Federado (Federated Learning – FL) é um paradigma de aprendizado de máquina distribuído que busca responder a dois desafios centrais da era da computação ubíqua: a explosão de dados gerados em dispositivos de borda e a necessidade de preservar a privacidade. Diferente do aprendizado centralizado, onde todos os dados são transferidos para um servidor, no FL as informações permanecem localmente nos clientes, e apenas os parâmetros do modelo são compartilhados [AbdulRahman et al. 2020]. O ciclo básico ocorre em rodadas sucessivas: o servidor envia o modelo inicial, os clientes treinam localmente com seus dados, devolvem as atualizações, e o servidor as agrega para atualizar o modelo global. Esse processo se repete até que o modelo atinja convergência. Entre as vantagens do FL destacam-se a preservação da privacidade, já que dados sensíveis não saem dos dispositivos; a escalabilidade, permitindo treinar em milhares de clientes; e a redução do custo de comunicação, uma vez que apenas parâmetros são transmitidos.

Apesar de suas vantagens, o FL enfrenta limitações significativas. A presença de dados não-IID (non-independent and identically distributed) dificulta a convergência do modelo global, pois cada cliente possui distribuições locais distintas, comprometendo a consistência das atualizações agregadas [Seo et al. 2025]. Além disso, a heterogeneidade de recursos entre os clientes, como diferenças de hardware, conectividade e capacidade de processamento, afeta diretamente a participação e o desempenho. Outro obstáculo importante é o desbalanceamento de dados, quando poucos clientes concentram a maior parte dos exemplos, gerando viés e reduzindo a representatividade do modelo global. O algoritmo clássico, o Federated Averaging (FedAvg), combina parâmetros por meio de uma média ponderada em função do tamanho dos datasets. Embora eficiente em cenários homogêneos, sua performance degrada em ambientes heterogêneos, o que motiva a busca

por novas estratégias de seleção e agregação de clientes.

Nesse contexto, métodos como o SCOPE-FL exploram métricas como a entropia de Shannon, que quantifica a diversidade de uma distribuição de probabilidade. No aprendizado federado, a entropia mede a variedade de classes nos dados de cada cliente, funcionando como um indicativo da riqueza estatística de sua contribuição. Entretanto, os desafios não são apenas técnicos: do ponto de vista ético, o FL enfrenta o dilema entre equidade e eficiência. Clientes com maior poder computacional ou datasets mais ricos tendem a dominar as atualizações, marginalizando dispositivos menos potentes ou populações sub-representadas. Esse viés pode levar a modelos enviesados, comprometendo a imparcialidade das decisões automatizadas. [Mozaffari and Houmansadr 2022] destacam esse problema ao propor o E2FL, que busca equilibrar igualdade e equidade entre clientes, mitigando distorções causadas por distribuições desiguais de dados e capacidades.

Além disso, aspectos legais e regulatórios, como a Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral sobre a Proteção de Dados (GDPR) na União Europeia, impõem obrigações adicionais de transparência, auditoria e responsabilidade no tratamento de dados. A LGPD exige consentimento explícito e responsabilização no uso de dados pessoais [Instituto Nacional do Seguro Social (INSS) 2025], enquanto o GDPR estabelece princípios como minimização de dados e direito ao esquecimento [European Union 2016]. Assim, métodos como o SCOPE-FL devem ser avaliados não apenas em termos de desempenho técnico, mas também de conformidade legal e ética. Nesse sentido, iniciativas de boas práticas em ciência de dados, como as discutidas em [McKinsey & Company 2022], reforçam a importância de alinhar o desenvolvimento de modelos inteligentes aos valores humanos e à justiça social, garantindo que avanços em eficiência não comprometam direitos fundamentais dos usuários.

3 Trabalhos Relacionados

Souza *et al.* usa uma abordagem de diferenciação entre os clientes de acordo com seus desempenhos estruturais. Dessa forma, a utilização de recursos pode ser reduzida exponencialmente, diminuindo tanto a quantidade de clientes que treinam por rodada quanto a variabilidade da generalização, ao variar os clientes selecionados. Esta abordagem proporciona uma maior integração entre os clientes, tornando-os menos repetidos e utilizando um cenário mais dinâmico e menos estocástico [de Souza et al. 2023].

Deng *et al.* utiliza uma abordagem mais relativista e estática em relação à seleção de clientes. Enquanto o aprendizado funciona normalmente, o servidor verifica quais clientes apresentam maior variabilidade na qualidade. Esse fator pode variar tanto em função da quantidade de rótulos quanto do tamanho total dos dados, funcionando como um método de ranqueamento para a seleção de clientes [Deng et al. 2021].

Esses trabalhos têm como principal foco a seleção de clientes, e ambos se preparam para selecionar com base na escolha estrutural dos clientes, a fim de realizar a decisão de prioridade a clientes a serem escolhidos para realizar a próxima etapa de treino. Dessa forma, este trabalho foca em tornar o cenário dinâmico, utilizando os dois métodos para selecionar os clientes de maneira não estocástica.

Outros trabalhos tentam buscar soluções dinâmicas. [Cho et al. 2020] introduziram o UCB-CS, que emprega técnicas de multi-armed bandits para equilibrar justiça e eficiência na seleção. O método reduz o custo de comunicação ao evitar a coleta frequente de métricas locais, mantendo boa convergência mesmo com dados não-IID. [Zhang et al. 2023] apresentaram o FedALA, que realiza agregação adaptativa considerando caracterís-

ticas personalizadas dos clientes. A abordagem melhora a generalização local ao ajustar dinamicamente a influência do modelo global sobre cada cliente. [Condori Bustincio et al. 2024] propuseram o EntropicFL, utilizando entropia e divergência entre modelos locais como critério principal de seleção. O algoritmo prioriza clientes com maior diversidade informacional, promovendo representatividade e acelerando a convergência do modelo global. O SCOPE-FL diferencia-se ao adotar uma métrica híbrida que combina volume de dados e entropia de Shannon. Ao invés de excluir clientes, ele ajusta dinamicamente seus pesos, garantindo maior equilíbrio entre eficiência e diversidade estatística, aspecto pouco explorado em trabalhos anteriores.

4 Metodologia: SCOPE-FL

Esta seção introduz o algoritmo do SCOPE-FL (*Selection of Clients by Order of Entropy in Federated Learning*), o qual melhora a seleção de clientes em relação ao estado da arte no processo de FL ao considerar os valores de entropia que aquele modelo gera e a quantidade de dados de cada cliente. O algoritmo utiliza como base os clientes com maior quantidade de dados para ranqueá-los e, em seguida, verifica os valores de peso, mensurando que, quanto maior o peso, mais preferível é a escolha desse cliente a cada interação a partir da entropia de Shannon de seus dados [Orlandi et al. 2023]. Esta seção descreve o modelo do sistema e os detalhes operacionais do SCOPE-FL.

4.1 Formulação Geral

Considere um conjunto de n clientes $U = \{u_1, u_2, \dots, u_n\}$, cada um possuindo um dataset local D_i . Em cada rodada de treinamento, os clientes selecionados atualizam seus modelos locais θ_i , e o servidor central agrega os parâmetros para atualizar o modelo global θ_{global} . A atualização segue a formulação do FedAvg:

$$\theta_{global} = \frac{1}{n} \sum_{i=1}^n w_i \theta_i \quad (1)$$

em que w_i representa o peso associado ao cliente i , definido inicialmente como a fração do tamanho do seu dataset em relação ao total:

$$w_i = \frac{|D_i|}{\sum_{i=1}^n |D_i|} \quad (2)$$

Esse esquema assegura que clientes com maior volume de dados exerçam maior influência sobre o modelo global, mas não leva em conta a qualidade ou diversidade intrínseca das amostras.

4.2 Incorporação da Entropia de Dados

Para capturar a diversidade informacional de cada cliente, o SCOPE-FL integra a entropia de Shannon [Shannon 1948] como métrica de relevância estatística. A entropia é definida como:

$$H(X) = - \sum_x P(x) \log P(x) \quad (3)$$

onde $P(x)$ corresponde à probabilidade de ocorrência da classe x dentro do dataset do cliente. Valores mais altos de entropia indicam maior diversidade de classes, o que implica maior potencial de contribuição para o modelo global.

4.3 Pontuação de Relevância

A pontuação de relevância de cada cliente i é calculada combinando duas dimensões principais: o tamanho do dataset local (D_i) e a diversidade de classes, medida pela entropia de Shannon (H_i). A fórmula utilizada é dada por:

$$R_i = \alpha \cdot \frac{D_i}{\max(D)} + \beta \cdot \frac{H_i}{\max(H)}$$

Os hiperparâmetros α e β determinam a importância relativa entre a quantidade de dados disponíveis em cada cliente e a diversidade estatística do conjunto de clientes. A configuração inicial adotou $\alpha = \beta$, atribuindo pesos iguais às duas dimensões. Essa escolha segue práticas recorrentes na literatura, funcionando como um ponto de partida neutro que evita vieses artificiais e permite que o processo de seleção de clientes considere de maneira equilibrada tanto a representatividade dos dados quanto a heterogeneidade da rede.

Para avaliar a sensibilidade do método, foram realizados testes variando os valores de α e β . Os resultados indicaram que o desempenho do SCOPE-FL permaneceu consistente, evidenciando a robustez do algoritmo mesmo frente a diferentes configurações. Na prática, os hiperparâmetros podem ser ajustados de acordo com os objetivos do sistema: aumentar α para priorizar clientes com maior volume de dados ou elevar β quando a diversidade entre os clientes for essencial para reduzir enviesamentos e melhorar a generalização do modelo global. Essa adaptabilidade permite que o SCOPE-FL seja aplicado em diferentes cenários de Aprendizado Federado, mantendo ganhos de desempenho sem necessidade de ajustes finos e garantindo flexibilidade para lidar com redes heterogêneas e dados de distribuição variada.

4.4 Normalização por Faixa de Dados

No cálculo da relevância, datasets muito pequenos poderiam gerar distorções na fórmula, pois poderiam apresentar entropia alta apenas por acaso, mesmo sendo pouco representativos. Para resolver isso, adotamos uma normalização por faixa, considerando apenas os 40% maiores datasets.

Na prática, isso significa que os clientes com datasets grandes servem como ponto de comparação. Clientes menores continuam participando, mas com peso reduzido, evitando que informações irrelevantes ou ruidosas tenham impacto exagerado no modelo global. Em outras palavras, clientes com pouquíssimos dados não desaparecem, mas também não puxam o modelo para direções erradas.

O corte de 40% foi definido de forma empírica: cortes menores (20%) reduziram a diversidade, enquanto cortes maiores (60%) aumentaram a instabilidade. Assim, os 40% se mostraram um equilíbrio adequado entre representatividade e diversidade. Esse valor pode ser ajustado em outros cenários conforme a aplicação.

4.5 Fluxo Operacional

O processo de treinamento do **SCOPE-FL** pode ser resumido em quatro etapas principais, representadas no ciclo da Figura 1:

1. **Seleção de clientes:** um subconjunto $C \subseteq U$ é escolhido para a rodada de treinamento.
2. **Treinamento local:** cada cliente $u_i \in C$ atualiza seu modelo local θ_i com base em D_i .
3. **Cálculo da relevância:** para cada cliente, calcula-se a entropia H_i e a pontuação de relevância R_i .
4. **Agregação ponderada:** os modelos locais são combinados proporcionalmente a R_i , gerando o novo modelo global θ_{global} .

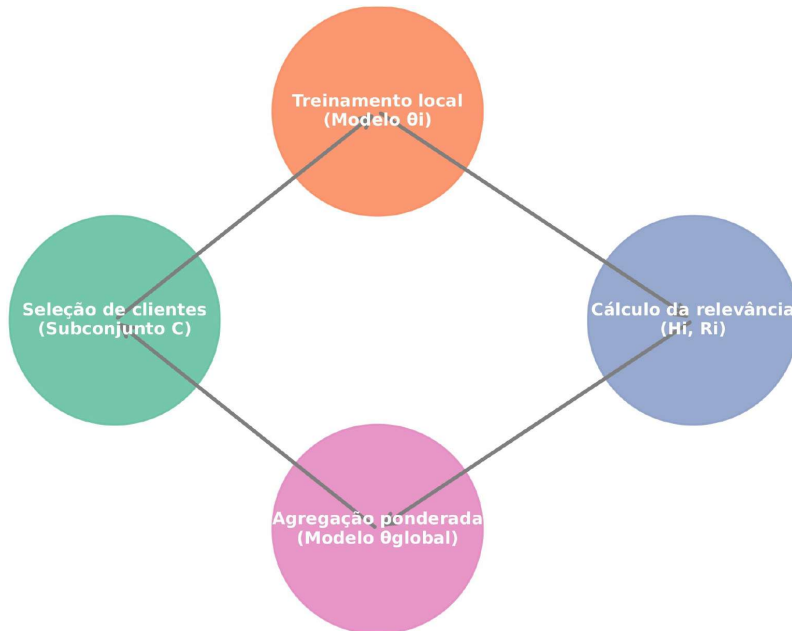


Figura 1: Fluxo operacional do SCOPE-FL.

Esse fluxo (seleção de clientes, treinamento local, cálculo da relevância e agregação ponderada) garante que o processo de aprendizado federado seja ao mesmo tempo **eficiente**, ao considerar datasets maiores, e **estatisticamente representativo**, ao valorizar a diversidade por meio da entropia.

5 Avaliação

Nesta seção, são apresentados os resultados da comparação entre o estado da arte e o SCOPE-FL, considerando os indicadores de perda, acurácia e tempo, visando melhorar a performance da simulação de FL. O estudo foi realizado utilizando o framework PFLlib [Zhang et al. 2023] em um servidor, com o conjunto de dados público MNIST. A arquitetura do modelo é uma *Convolutional Neural Network* (CNN) com duas camadas

convolucionais de filtros 5x5, seguidas de max-pooling 2x2 após cada camada. Foram utilizados 100 clientes em 50 rodadas, com 10% deles selecionados para o treino.

A Figura 2 apresenta a medição de acurácia para três abordagens comparadas: a Perda baseada no trabalho de Souza *et al.* [de Souza et al. 2023], a abordagem Aleatória e a proposta deste trabalho, o SCOPE-FL. Neste cenário, com apenas 10% dos clientes selecionados, observa-se que, com poucos ajustes e um número mínimo de clientes, o SCOPE-FL atinge mais de 60% de acurácia em 12 rodadas e continua progredindo até cerca de 80% de acurácia após 22 rodadas. Isso demonstra que o SCOPE-FL possui um bom desempenho, alcançando resultados elevados com poucas interações. Em contraste, o método de Perda apresenta uma progressão mais lenta, atingindo uma convergência máxima média de 64%. O método Aleatória, por sua vez, apresenta resultados semelhantes aos do SCOPE-FL, mas com uma convergência menos suave, alcançando apenas 56% de acurácia após 12 rodadas, embora ainda progrida para valores próximos aos do SCOPE-FL nas últimas rodadas.

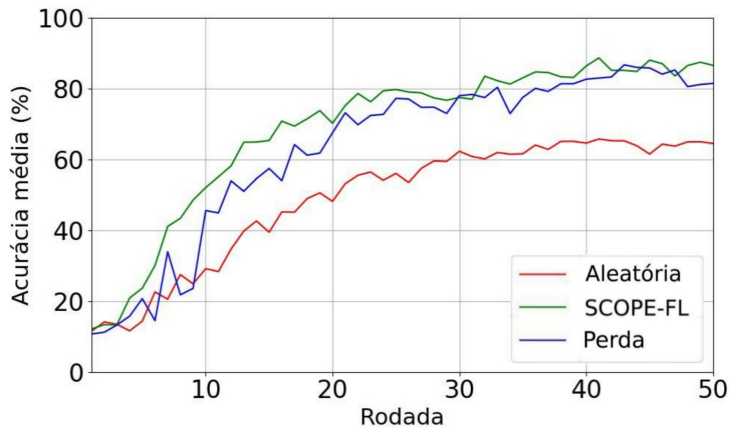


Figura 2: Medição de Acurácia

A Figura 3 mostra a medição de perda dos modelos, evidenciando como a seleção de clientes com melhores preferências de *dataset* impacta o aprendizado. Com mais dados e maior importância deles, os modelos nos clientes conseguem aprender de forma mais eficaz. O SCOPE-FL apresentou os melhores resultados, alcançando valores de perda abaixo de 1 após cerca de 15 rodadas. Em contrapartida, o método de Perda não teve o mesmo desempenho, não conseguindo reduzir a perda abaixo de 1, mostrando um gargalo de aprendizado por volta da 30^a rodada. O método aleatório, embora com desempenho semelhante ao do SCOPE-FL, não superou o modelo proposto neste artigo, alcançando resultados melhores apenas após cerca de 40 rodadas, quando ambos já estavam em fase de convergência.

Embora o SCOPE-FL tenha apresentado resultados promissores, algumas limitações devem ser reconhecidas: em termos de escalabilidade restrita, os testes envolveram até 100 clientes; em cenários reais, esse número pode chegar a milhares.

Além disso, algumas métricas de recursos computacionais não foram consideradas, tais como energia, latência ou custo de comunicação. Por fim, não foram explorados datasets não-IID extremos, que poderiam evidenciar cenários de maior heterogeneidade.

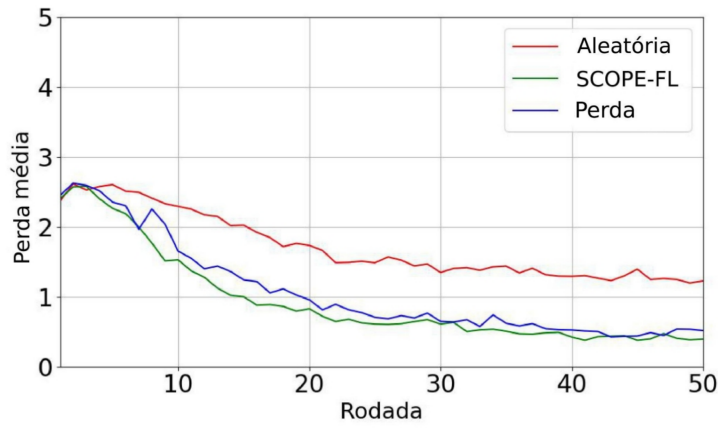


Figura 3: Medição de Perda

6 Conclusão

Este trabalho apresentou o *SCOPE-FL*, uma abordagem inovadora que combina entropia de Shannon e tamanho relativo dos datasets como critério de seleção e ponderação de clientes em aprendizado federado. O *SCOPE-FL* demonstrou ser eficaz na melhoria da precisão e eficiência do treinamento, superando métodos tradicionais, como Perda e Aleatória, especialmente em termos de acurácia e velocidade de convergência. As simulações realizadas com o MNIST mostraram que o *SCOPE-FL* obteve mais de 60% de acurácia após 12 rodadas.

A estratégia proposta mostrou ganhos consistentes em acurácia, velocidade de convergência e robustez frente a métodos tradicionais como seleção aleatória e baseada em perda local. Os resultados sugerem que o *SCOPE-FL* é particularmente útil em cenários de grande heterogeneidade, típicos de sistemas IoT, cidades inteligentes e redes de dispositivos de borda. Ao valorizar simultaneamente diversidade e volume, o método promove maior generalização sem excluir rigidamente clientes.

Como trabalho futuro, a implementação de métodos adaptativos de seleção de clientes com base em características dinâmicas de dados e o aprimoramento da abordagem *SCOPE-FL* podem aumentar ainda mais a robustez e a escalabilidade do FL. A exploração de diferentes cenários de dados não independentes e idênticos (non-IID) pode expandir o alcance e a aplicabilidade da metodologia, proporcionando uma solução mais eficiente e privada para sistemas de FL em dispositivos de borda. Assim, utilizando métodos como a distribuição patológica, em que a distribuição é ajustada para cenários mais desafiadores, implica no uso de *datasets* ainda mais complexos, aumentando a dificuldade do modelo.

Além disso, como possibilidades de trabalhos futuros, este trabalho pode se estender para tentar uma abordagem de aprendizado federado personalizado ou tentar a integração de métricas de energia e latência em sua análise. Esses caminhos indicam o potencial do *SCOPE-FL* como contribuição relevante para o avanço do aprendizado federado em cenários distribuídos e sensíveis à privacidade.

Referências

[AbdulRahman et al. 2020] AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., and Guizani, M. (2020). A survey on federated learning: The journey from

- centralized to distributed on-site learning and beyond. *IEEE Internet of Things Journal*, 8(7):5476–5497.
- [Cho et al. 2020] Cho, Y. J., Wang, J., Joshi, G., and Poor, H. V. (2020). Client selection in federated learning via submodular maximization. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8866–8870. IEEE.
- [Condori Bustincio et al. 2024] Condori Bustincio, R. W. C., de Souza, A. M., da Costa, J. B. D., and Bittencourt, L. F. (2024). Entropicfl: Efficient federated learning via data entropy and model divergence. In *Proceedings of the 16th IEEE/ACM International Conference on Utility and Cloud Computing (UCC)*.
- [de Souza et al. 2023] de Souza, A. M., Bittencourt, L. F., Cerqueira, E., Loureiro, A. A., and Villas, L. A. (2023). Dispositivos, eu escolho vocês: Seleção de clientes adaptativa para comunicação eficiente em aprendizado federado. In *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, pages 1–14. SBC.
- [Deng et al. 2021] Deng, Y., Lyu, F., Ren, J., Wu, H., Zhou, Y., Zhang, Y., and Shen, X. (2021). Auction: Automated and quality-aware client selection framework for efficient federated learning. *IEEE Transactions on Parallel and Distributed Systems*, 33(8):1996–2009.
- [European Union 2016] European Union (2016). Regulation (eu) 2016/679 of the european parliament and of the council. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>. Acessado em: 29 ago. 2025.
- [Instituto Nacional do Seguro Social (INSS) 2025] Instituto Nacional do Seguro Social (INSS) (2025). Lei geral de proteção de dados pessoais (lgpd). <https://www.gov.br/inss/pt-br/aceso-a-informacao/lei-geral-de-protecao-de-dados-pessoais>. Acessado em: 29 ago. 2025.
- [McKinsey & Company 2022] McKinsey & Company (2022). Data ethics: What it means and what it takes. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/data-ethics-what-it-means-and-what-it-takes>. Acessado em: 29 ago. 2025.
- [McMahan et al. 2017] 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. PMLR.
- [Mozaffari and Houmansadr 2022] Mozaffari, H. and Houmansadr, A. (2022). E2fl: Equal and equitable federated learning. *arXiv preprint arXiv:2205.10454*.
- [Orlandi et al. 2023] Orlandi, F. C., Dos Anjos, J. C., Santana, J. F. d. P., Leithardt, V. R., and Geyer, C. F. (2023). Entropy to mitigate non-iid data problem on federated learning for the edge intelligence environment. *IEEE Access*.
- [Pires et al. 2020] Pires, I. M., Marques, G., Garcia, N. M., Flórez-Revuelta, F., Canavaro Teixeira, M., Zdravevski, E., Spinsante, S., and Coimbra, M. (2020). Pattern recognition techniques for the identification of activities of daily living using a mobile device accelerometer. *Electronics*, 9(3):509.

- [Seo et al. 2025] Seo, J., Catak, F. O., and Rong, C. (2025). Understanding federated learning from iid to non-iid dataset: An experimental study. *arXiv preprint arXiv:2502.00182*.
- [Shannon 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.
- [Zhang et al. 2023] Zhang, J., Hua, Y., Wang, H., Song, T., Xue, Z., Ma, R., and Guan, H. (2023). Fedala: Adaptive local aggregation for personalized federated learning. In *AAAI Conference on Artificial Intelligence*, volume 37, pages 11237–11244.