

Avaliação do Algoritmo HETAFF em Novos Cenários de Aprendizado Federado

K. C. M. Santos F. M. Roberto L. F. Bittencourt

Relatório Técnico - IC-PFG-25-37

Projeto Final de Graduação

2025 - Novembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.
O conteúdo deste relatório é de única responsabilidade dos autores.

Avaliação do Algoritmo HETA AFF em Novos Cenários de Aprendizado Federado

Kevin Caio Marques dos Santos* Filipe Maciel Roberto* Luiz Felipe Bittencourt*

Resumo

Este trabalho apresenta uma avaliação ampliada do algoritmo HETA AFF, uma estratégia adaptativa de ajuste do *fit fraction* em aprendizado federado. Enquanto os estudos originais do HETA AFF concentraram-se em apenas alguns cenários e conjuntos de dados, este projeto expande substancialmente o escopo da análise, analisando o comportamento do algoritmo em diferentes tarefas, arquiteturas de rede e políticas de seleção de clientes. A avaliação inclui novos contextos, além da classificação de imagem, como reconhecimento de áudio e predição de texto, além de comparar o HETA AFF às estratégias FedAvg, Critical-FL e AFF em múltiplos níveis de heterogeneidade estatística, definidos via particionamento Dirichlet e variados perfis de clientes. Além disso, são analisadas métricas de custo, como volume de comunicação e energia consumida, com o objetivo de fornecer uma visão mais completa da relação entre desempenho e eficiência operacional nas diferentes configurações experimentais.

1 Introdução

O Aprendizado Federado (*Federated Learning* – FL) surge como uma alternativa ao aprendizado de máquina tradicional ao permitir o treinamento de modelos sem a necessidade de centralização dos dados. Nesse paradigma, os dados permanecem localmente nos dispositivos clientes e apenas os parâmetros atualizados dos modelos são enviados ao servidor central, o qual realiza a agregação e redistribuição do modelo global [2]. Essa abordagem reduz os riscos relacionados à privacidade de dados, assunto altamente discutido recentemente e apoiado pelas novas legislações, GDPR [3] e PIPL [4], e possibilita a aplicação de técnicas de inteligência artificial em domínios sensíveis, como saúde [6], finanças [5] e internet das coisas [7].

Um dos principais desafios do FL está na heterogeneidade estatística e sistêmica entre os clientes, que pode impactar significativamente a convergência e o desempenho do modelo global [8]. Diversos estudos têm buscado estratégias de seleção de clientes que considerem essa heterogeneidade, ajustando de forma dinâmica a fração de participantes em cada rodada de treinamento [1], também conhecida como *fit fraction*.

Uma das abordagens de seleção de clientes é o FedAvg [2], que define uma fração fixa e constante de clientes para todo o treinamento. Nesse algoritmo, o servidor realiza a agregação das atualizações locais por meio de uma média ponderada pelo número de amostras de cada participante. Apesar de sua simplicidade, obtém bons resultados em cenários IID (*Independent and Identically Distributed*), que são quando os dados tem bastante homogeneidade estatística, contudo, o uso de uma taxa de participação estática pode ser ineficiente em cenários não-IID, nos quais a distribuição dos dados e a capacidade computacional dos clientes variam significativamente. Por isso, para tornar o processo mais adaptativo, surgiram abordagens dinâmicas de seleção de clientes, como por exemplo

*Instituto de Computação, Universidade Estadual de Campinas, 13081-970 Campinas, SP

o Critical-FL [9] que identifica períodos críticos de aprendizado comparando a variação entre gradientes de rodadas sucessivas e ajusta o número de clientes de acordo com essa medida, de modo a concentrar esforços nas fases mais instáveis do treinamento. Em seguida, o AFF (*Adaptive Fit Fraction*) [10] propõe uma adaptação mais sofisticada, introduzindo uma janela de observação sobre o desempenho do modelo global para inferir uma tendência de evolução (*slope*). A partir dessa tendência, o algoritmo aumenta ou reduz a fração de clientes nos blocos subsequentes de rodadas, buscando equilibrar custo de comunicação e desempenho.

A partir do AFF, foi proposto o HETA AFF (*Heterogeneity-Aware Adaptive Fit Fraction*) [1], que incorpora de forma explícita a heterogeneidade entre os modelos locais como variável de controle na política de ajuste do *fit fraction*. Essa heterogeneidade é quantificada por meio do CKA (*Centered Kernel Alignment*), que mede a similaridade entre matrizes de representação aprendidas pelos modelos dos clientes, sendo os valores próximos de zero indicam redes altamente divergentes, enquanto que valores próximos de um refletem elevada similaridade interna. Com isso, ao incorporar a heterogeneidade observada como fator moderador na taxa de ajuste da participação de clientes, o HETA AFF evita que as decisões de adaptação se baseiem apenas em métricas globais, como acurácia ou perda, que não capturam diretamente o grau de divergência entre modelos locais. Dessa forma, o HETA AFF introduz um mecanismo adicional de controle sensível ao contexto não-IID, ajustando o tamanho do conjunto de clientes de modo mais conservador quando a diversidade entre modelos locais é elevada e permitindo ajustes mais agressivos quando há maior coesão entre os clientes.

Embora tanto o AFF quanto o HETA AFF tenham demonstrado, em seus estudos originais, resultados superiores, às estratégias conhecidas, quanto ao desempenho preditivo em cenários IID e não-IID, suas avaliações concentraram-se principalmente na acurácia do modelo, em alguns cenários específicos. Desse modo, permanece desconhecido o comportamento do algoritmo em contextos mais variados ou sob diferentes condições operacionais. Assim sendo, o presente trabalho tem como objetivo expandir a avaliação do algoritmo em novos cenários experimentais, envolvendo diferentes arquiteturas de rede, conjuntos de dados e políticas de seleção de clientes, além de incorporar novas métricas relacionadas aos custos, como consumo energético e volume de comunicação, de modo a fornecer uma avaliação mais abrangente do equilíbrio entre desempenho e eficiência proporcionado pela solução HETA AFF em diferentes cenários.

O restante do trabalho está dividido na seguinte estrutura: Seção II apresenta a fundamentação teórica e a motivação. Seção III descreve a metodologia utilizada nas análises. Seção IV apresenta os experimentos realizados. Seção V discorre sobre os resultados obtidos e a Seção VI finaliza apresentando as conclusões obtidas.

2 Fundamentação Teórica

Esta seção será dividida em três partes principais, começando com uma introdução de como o aprendizado federado funciona, passando pelo conceito de seleção de clientes e finalizando com a variação da participação dos clientes no treinamento.

2.1 Aprendizado Federado

O Aprendizado Federado, também denominado *Federated Optimization* [2], é uma variação da otimização distribuída na qual o objetivo global é minimizar uma função de perda composta pelas distribuições locais dos clientes. Formalmente, o problema pode ser expresso como:

$$\min_{w \in \mathbb{R}^d} f(w), \quad \text{onde} \quad f(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w) \quad (1)$$

em que K representa o número total de clientes, n_k é o número de amostras pertencentes ao cliente k ,

$$n = \sum_{k=1}^K n_k \quad (2)$$

é o total de exemplos no sistema, e

$$F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} \ell(x_i, y_i; w) \quad (3)$$

corresponde à função de perda local do cliente k , definida sobre seu conjunto de dados P_k . Nesse contexto, cada cliente mantém seus dados privados e contribui para a minimização de $f(w)$, por meio de atualizações locais, que são subsequentemente agregadas pelo servidor central.

Nesse sentido, temos o cenário IID (*Independent and Identically Distributed*) quando os dados são distribuídos aleatoriamente, de modo que cada cliente representa a população global, contudo, no aprendizado federado real, os dados são geralmente não-IID e desbalanceados, nos quais o número de amostras varia significativamente entre os participantes, e isso ocorre pois refletem o uso individual de cada dispositivo ou aplicação, resultando em quantidades e distribuições distintas entre clientes [2]. Por conta desse desbalanceamento e diversidade estatística, a convergência global do modelo é prejudicada e vieses locais podem ser gerados no modelo agregado.

Além disso, no contexto do aprendizado federado horizontal, é comum que o sistema seja massivamente distribuído, sendo esperado que o número de clientes participantes em uma otimização seja muito maior que a média de dados por clientes [2], e por operar sob restrições severas de comunicação. Diferentemente dos ambientes de *data center*, nos quais a limitação principal é o custo computacional, no FL o gargalo principal é o custo de transmissão de parâmetros entre clientes e servidor. Mesmo assim, o próprio treinamento local também demanda otimizações, especialmente em dispositivos móveis com restrições de energia, processamento e conectividade. Por conta disso, existem muitos estudos que buscam reduzir tanto o volume de dados transmitidos quanto o custo computacional envolvido em cada rodada federada. Dentre as abordagens propostas, destacam-se técnicas de compressão e quantização de modelos, que diminuem consideravelmente o tamanho das mensagens transmitidas [18]; métodos de esparsificação de gradientes, como a transmissão apenas dos top-k coeficientes mais relevantes, reduzindo consideravelmente o custo de comunicação por rodada [19]; além de estratégias de seleção de clientes que levam em consideração o consumo energético, priorizando dispositivos com maior disponibilidade de recursos ou menor custo de participação [20].

Outro desafio decorre da natureza intermitente dos dispositivos clientes, que podem estar desconectados, apresentar falhas de comunicação ou dispor de recursos computacionais limitados. Assim, apenas uma parcela dos clientes é selecionada a cada rodada, não apenas por eficiência, mas também por disponibilidade prática [2]. Como resultado, o aprendizado federado busca equilibrar três dimensões fundamentais: (i) heterogeneidade estatística e sistêmica entre os clientes, (ii) restrições de comunicação e conectividade, e (iii) escalabilidade em ambientes com alto número de participantes e dados dinâmicos [2].

2.2 Seleção de Clientes

A seleção de clientes constitui um componente essencial no aprendizado federado, responsável por determinar, a cada rodada de treinamento, quais dispositivos locais participarão do processo de atualização do modelo global [16]. Como o número total de clientes disponíveis no aprendizado federado horizontal é geralmente muito superior à quantidade que pode ser efetivamente envolvida em cada iteração, a escolha desse subconjunto influencia diretamente o desempenho, a estabilidade e o custo de comunicação do sistema federado.

A forma mais simples de seleção é o sorteio aleatório de uma fração fixa de clientes, abordagem adotada pelo algoritmo FedAvg [2]. Embora eficiente do ponto de vista da implementação, essa estratégia não considera diferenças entre os participantes, podendo resultar em amostras pouco representativas em cenários com forte heterogeneidade estatística ou restrições de recursos. Por esse motivo, diversos estudos têm proposto métodos de seleção mais sofisticados, que levam em conta características individuais dos clientes ou o estado atual do treinamento [16].

Entre as principais abordagens, destacam-se aquelas baseadas em utilidade, que priorizam clientes com maior potencial de contribuição para o aprendizado global, considerando fatores como perda local, gradientes ou diversidade de dados. Como exemplo, temos o Oort, que combina fatores estatísticos (como perda local) com fatores sistêmicos (como velocidade de processamento) para selecionar clientes com maior impacto esperado na convergência [17]. Outras estratégias são baseadas em recursos, nas quais são selecionados clientes de acordo com sua disponibilidade computacional, capacidade de comunicação ou energia, de modo a maximizar a eficiência do treinamento [21]. Há ainda os métodos baseados em diversidade, como o Div-FL, que buscam garantir que o conjunto selecionado represente um campo mais amplo das distribuições locais, reduzindo redundâncias, mitigando o efeito da heterogeneidade estatística [16] e evitando que clientes com distribuições de dados muito distintas dominem o processo de agregação.

De modo geral, os algoritmos de seleção buscam encontrar um compromisso entre desempenho do modelo e eficiência sistêmica. Estratégias mais seletivas tendem a acelerar a convergência, mas podem aumentar a desigualdade de participação entre os clientes. Por outro lado, políticas puramente aleatórias promovem equidade, mas podem exigir maior número de rodadas para atingir a mesma acurácia. Assim, a seleção de clientes permanece um dos principais desafios do aprendizado federado, sendo objeto de estudo contínuo para equilibrar custo, representatividade e desempenho em diferentes cenários de heterogeneidade [15].

2.3 Variação da participação

A variação da participação de clientes no aprendizado federado é um mecanismo para equilibrar custo e desempenho em sistemas descentralizados. Essa variação consiste em ajustar, ao longo do treinamento, a fração de clientes selecionados em cada rodada, buscando um compromisso entre diversidade estatística, estabilidade do aprendizado e eficiência de comunicação. Diversas propostas foram apresentadas com esse propósito, cada uma fundamentada em diferentes critérios de decisão e hipóteses sobre o comportamento do modelo global. Tais propostas serão abordadas nos parágrafos seguintes.

O FedAvg adota uma fração fixa e constante de clientes em todas as rodadas, representando a abordagem mais simples e amplamente utilizada [2], sendo utilizada como base para comparação com outras propostas. Seu objetivo é fornecer uma base estável de comparação, embora não considere as variações de desempenho ou heterogeneidade entre os clientes, o que pode levar a desperdício de recursos em cenários com alta dispersão estatística.

Partindo para abordagens dinâmicas, tem-se o Critical-FL [9], o qual introduz uma adaptação

do número de clientes fundamentada na observação de períodos críticos do aprendizado. O método compara a variação entre gradientes de rodadas consecutivas e ajusta a quantidade de clientes conforme a instabilidade detectada: quando há grande variação, mais clientes são incluídos para estabilizar o processo; quando o modelo apresenta convergência estável, a fração é reduzida, economizando comunicação.

Nesse mesmo sentido, tem-se o *Adaptive Fit Fraction* (AFF) [10] que propõe uma estratégia mais refinada, utilizando uma janela de observação sobre o desempenho global do modelo para calcular a tendência de evolução (*slope*). Com essa tendência, tem-se um indicador para verificar se o modelo está melhorando ou não o seu desempenho, guiando o aumento ou a redução gradual da participação de clientes. Com isso, o objetivo do AFF é ajustar dinamicamente a taxa de participação com base na trajetória do aprendizado, reduzindo custos sem comprometer a acurácia.

Por fim, o *Heterogeneity-Aware Adaptive Fit Fraction* (HETA AFF) [1] estende o AFF ao incorporar explicitamente a heterogeneidade entre os modelos locais como variável de decisão. Essa heterogeneidade é mensurada pelo alinhamento das representações internas dos modelos, por meio do *Centered Kernel Alignment* (CKA), o qual indica o grau de similaridade entre os clientes. O HETA AFF utiliza essa medida para modular o ajuste do *fit fraction*: quando a heterogeneidade é alta, as variações são aplicadas de forma mais conservadora; quando há maior coesão entre clientes, a adaptação é mais ampla. Dessa forma, o algoritmo busca equilibrar desempenho e custo considerando tanto a dinâmica global do aprendizado quanto às diferenças locais entre os participantes.

Em síntese, as propostas diferem essencialmente nas fontes de informação utilizadas para ajustar a fração de clientes — enquanto FedAvg mantém a participação constante, Critical-FL baseia-se na variação dos gradientes, AFF na tendência de desempenho e HETA AFF na tendência e na heterogeneidade dos modelos. Essa diversidade de estratégias reflete a evolução das tentativas de tornar o aprendizado federado mais eficiente e adaptável aos cenários não-IID e às restrições sistêmicas típicas dessa abordagem.

3 Metodologia

A metodologia adotada neste projeto fundamenta-se na extensão do estudo original do algoritmo HETA AFF, com adaptações destinadas a explorar novos cenários e arquiteturas de rede. O HETA AFF e os demais algoritmos de variação do *fit fraction* foram avaliados em condições experimentais ampliadas, incorporando modelos de redes neurais convolucionais bastante utilizadas em dispositivos móveis, como a ShuffleNet [15], bem como arquiteturas específicas para diferentes cenários. Para a tarefa de classificação de comandos de voz, como o conjunto de dados *Speech Commands*, utilizouse a Kwt3 [22], projetada para classificação de áudio, já para a tarefa de predição de caracteres em sentenças, adotou-se uma arquitetura Lstm [23], utilizando-a no conjunto de dados *Shakespeare*. A inclusão desses modelos reflete a preocupação em aproximar os experimentos de cenários práticos, com domínios distintos, e com características computacionais e estatísticas próprias.

Além das mudanças de arquitetura, os experimentos também passaram a incluir novos contextos de seleção de clientes, com isso, os algoritmos de adaptação do *fit fraction* foram avaliados sob políticas mais avançadas de seleção de clientes, não somente seleção aleatória. Foram utilizadas as estratégias Oort e Div-FL, expostos anteriormente, e que propõem políticas de amostragem de clientes orientadas por utilidade e diversidade, respectivamente. Esses métodos permitem avaliar o HETA AFF sob políticas de amostragem diversas, variadas e sensíveis à heterogeneidade dos dados e ao custo sistêmico do treinamento. Com essa ampliação de escopo busca-se compreender a robustez do HETA AFF e sua capacidade de adaptação quando inserido em ambientes mais realistas e variados.

As implementações foram desenvolvidas utilizando o *framework* Flower [11] em conjunto com PyTorch [13], considerando simulações de múltiplos clientes, diferentes níveis de heterogeneidade induzidos pelo particionamento Dirichlet - uma técnica que distribui os dados entre os clientes de forma controlada, permitindo ajustar o quão desbalanceadas e não-IID serão as partições a partir de um parâmetro α [13] - e o controle de *seeds* para garantir reprodutibilidade. De modo geral, os experimentos seguiram uma sequência estruturada que inclui preparação e particionamento dos dados, inicialização dos clientes, execução dos algoritmos de variação da participação, coleta das métricas e análise comparativa. Esta seção descreve a visão geral dos experimentos, os detalhes numéricos e específicos de cada cenário e configuração serão apresentados na seção seguinte.

Por fim, além da acurácia, incorporaram-se métricas relacionadas ao custo sistêmico do treinamento federado, como volume de comunicação entre clientes e servidor e estimativas de consumo energético. A utilização de tais métricas visa expandir a avaliação do estudo original, exposto anteriormente, do HETAFF, de modo a oferecer uma avaliação mais completa do equilíbrio entre desempenho da solução e eficiência operacional, em contextos mais variados.

4 Experimentos

Os experimentos propostos neste trabalho foram organizados em três grupos principais, cada um estruturado para investigar aspectos de comportamento distintos do algoritmo HETAFF e das demais estratégias de variação da participação de clientes. Em todos os cenários, os modelos foram treinados por 150 rodadas, considerando um conjunto inicial de 10 clientes ativos e um total de 100 clientes disponíveis no ambiente federado. A cada rodada, o número de clientes selecionados permaneceu limitado entre 75% da quantidade inicial de clientes e 75% da quantidade total de clientes disponíveis na rodada, definidos para garantir consistência experimental entre as estratégias avaliadas.

Além disso, foram utilizados perfis computacionais reais, representando dispositivos frequentemente encontrados em aplicações práticas de aprendizado federado, incluindo Jetson AGX Orin, Jetson Orin Nano, LattePanda Delta 3 e Orange Pi 5B. Esses perfis foram empregados para estimar custos de execução mais realistas, tais como tempo de processamento e consumo energético, reforçando o caráter aplicado dos experimentos.

Em todos os cenários, avaliou-se o desempenho das quatro estratégias de ajuste do *fit fraction*: FedAvg, Critical-FL, AFF e HETAFF, além de serem executados cinco vezes, cada um com um valor de *seed* diferente, de modo a garantir robustez estatística e reduzir impacto de flutuações aleatórias nas análises. A presente seção será dividida em três subseções principais, cada uma descrevendo o cenário em questão.

4.1 Cenário 1

Neste primeiro cenário, o objetivo principal era avaliar o desempenho das estratégias de variação de *fit fraction* sob uma política de seleção aleatória de clientes, que é a abordagem mais simples e tradicional. Nesse cenário, os clientes participantes de cada rodada foram amostrados, dentro da fração de seleção, de forma aleatória, sem qualquer critério adicional de utilidade, disponibilidade ou diversidade. A fim de examinar o desempenho dos algoritmos frente a diferentes casos de heterogeneidade estatística, foram utilizados dois valores do parâmetro α de Dirichlet: 0,1 e 1; de modo a representar cenários altamente não-IID e próximos de IID. Esse cenário foi desenhado com características para ser utilizado como linha base de comparação.

Com isso, para cada combinação de α e algoritmo de ajuste de *fit fraction*, cinco execuções independentes foram realizadas, de modo a criar um conjunto de experimentos que permitem avaliar

o desempenho de cada algoritmo em um cenário neutro de seleção e impactado exclusivamente pela heterogeneidade estatística. Tais configurações do experimento podem ser vistas na Figura 1 abaixo.

Estratégias de Ajuste de Fit Fraction	Estratégias de Seleção de Clientes	Arquitetura da Rede Neural	Valores de α de Dirichlet	Valores de Seeds
FedAvg Constant	Random	ShuffleNet	0,1 1	1, 2, 3, 4, 5
Critical-FI				
Aff				
Hetaaff				

Figura 1: Variáveis utilizadas no primeiro cenário experimental. Todas as combinações entre os valores apresentados foram executadas, embora não estejam explicitadas individualmente por simplicidade

4.2 Cenário 2

Neste segundo cenário, o objetivo foi avaliar o desempenho das estratégias de variação do *fit fraction* quando combinadas com políticas de seleção de clientes mais sofisticadas, especificamente Oort e Div-FL. Diferentemente da seleção aleatória, essas políticas introduzem critérios estruturados: o Oort prioriza clientes com maior utilidade estimada para o aprendizado, enquanto o Div-FL busca maximizar a diversidade das distribuições locais. Desse modo, a seleção deixa de ser neutra e passa a refletir vieses explícitos orientados por utilidade ou diversidade. Para investigar a robustez das estratégias de *fit fraction* sob esses mecanismos de seleção, foram avaliados os mesmos dois níveis de heterogeneidade estatística gerados pelos valores α de Dirichlet (0,1 e 1), representando cenários altamente não-IID, e próximos de IID. Nesse conjunto de experimentos, utilizou-se uma única arquitetura de rede, a ShuffleNet, escolhida por seu bom equilíbrio entre desempenho e eficiência computacional.

Assim como no cenário anterior, cada combinação entre α , política de seleção e algoritmo de *fit fraction* foi executada cinco vezes com *seeds* distintas, porém agora acrescentando duas estratégias de seleção de clientes diferentes, permitindo analisar o comportamento dos algoritmos em condições nas quais a seleção dos clientes não é aleatória, mas orientada por critérios específicos que afetam diretamente a dinâmica de convergência. Tais configurações do experimento podem ser vistas na Figura 2 abaixo.

Estratégias de Ajuste de Fit Fraction	Estratégias de Seleção de Clientes	Arquitetura da Rede Neural	Valores de α de Dirichlet	Valores de Seeds
FedAvg Constant				
Critical-FI	Div-FL		0,1	
Aff	Oort	ShuffleNet	1	1, 2, 3, 4, 5
Hetaaff				

Figura 2: Variáveis consideradas no segundo cenário experimental. Foram realizados experimentos para todas as combinações possíveis entre os valores apresentados. A principal diferença entre os dois cenários está na escolha das estratégias de seleção de clientes enquanto o primeiro utiliza seleção aleatória, o segundo considera duas políticas de seleção (Oort e Div-FL)

4.3 Cenário 3

Neste terceiro e último cenário, o objetivo era investigar as estratégias em domínios além da classificação de imagens, ampliando o escopo da avaliação do HETAFF. Nesse cenário, adotou-se a seleção aleatória de clientes, de modo a isolar os efeitos da heterogeneidade dos dados e da arquitetura empregada, sem influências adicionais de estratégias avançadas de seleção. Para representar diferentes modalidades, foram utilizadas duas arquiteturas:

- Kwt3, aplicada ao dataset *Speech Commands*, para a tarefa de classificação de áudio;
- Lstm, aplicada ao dataset *Shakespeare*, para a tarefa de predição de caracteres em textos;

Ao contrário dos experimentos anteriores, que exploraram dois níveis de heterogeneidade estatística proporcionados pelo α de Dirichlet, este não utilizou um particionamento com base na distribuição de Dirichlet, utilizou um particionamento natural do dataset, agrupando todos os dados pertencentes a uma mesma entidade em um único cliente. Assim como nos outros experimentos, para cada combinação entre arquitetura e algoritmo de ajuste de *fit fraction*, foram realizadas cinco execuções independentes, garantindo estabilidade estatística. Tais configurações do experimento podem ser vistas na Figura 3 abaixo.

Estratégias de Ajuste de Fit Fraction	Estratégias de Seleção de Clientes	Arquitetura da Rede Neural	Valores de α de Dirichlet	Valores de Seeds
FedAvg Constant				
Critical-FI				
Aff	Random	Kwt3 - Speech Commands	Nenhum	
Hetaaff		Lstm - Shakespeare	(Apenas utilizado o particionamento natural do dataset)	1, 2, 3, 4, 5

Figura 3: Variáveis consideradas no terceiro cenário experimental. Foram realizados experimentos para todas as combinações possíveis entre os valores apresentados. A principal diferença dentre este cenário e os anteriores está nas arquiteturas de rede e nas tarefas avaliadas: enquanto os dois primeiros utilizam somente modelos de visão computacional, este incorpora arquiteturas específicas para áudio e texto, mantendo a seleção aleatória de clientes

5 Discussão dos Resultados

5.1 Experimento 1

A partir do gráfico da soma cumulativa de clientes participantes ao longo das rodadas mostrou que o Critical-FL foi a estratégia que mais selecionou clientes durante o treinamento, tanto no cenário não-IID quanto no cenário próximo de IID, enquanto que as outras estratégias possuíram uma quantidade mais próxima de clientes selecionados. Esse comportamento pode ser explicado pelo próprio mecanismo do Critical-FL, que ao detectar um período crítico, caracterizado por uma instabilidade na perda global, o algoritmo duplica a quantidade de clientes selecionados na rodada seguinte. Esse comportamento tem como objetivo estabilizar a direção da otimização, incorporando mais atualizações locais para reduzir divergências estatísticas entre os clientes e corrigir trajetórias de treinamento potencialmente instáveis, tendo como consequência uma grande quantidade de clientes participantes. Esse resultado pode ser visto na Figura 4 abaixo, que contém os gráficos gerados com os resultados obtidos ao longo do treinamento.

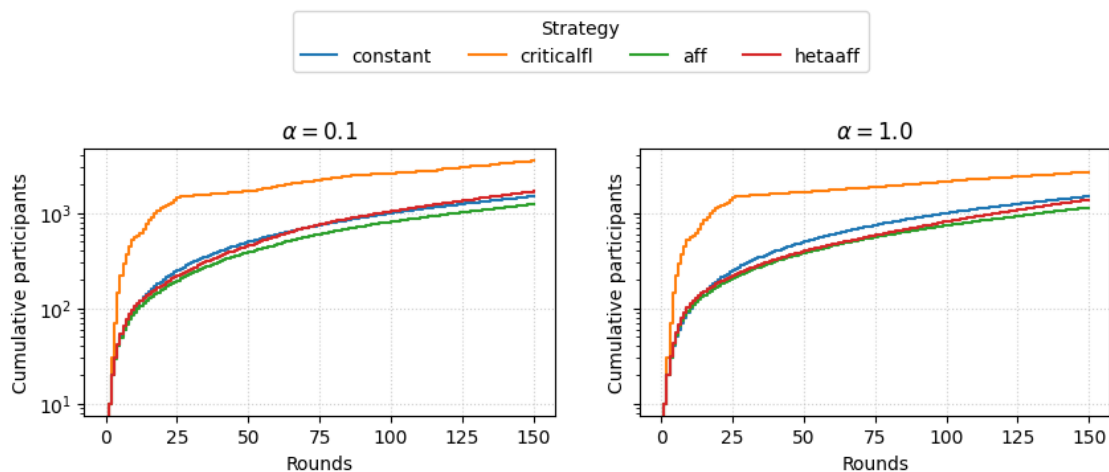


Figura 4: Gráfico da participação cumulativa de clientes durante o treinamento para o experimento 1

Como consequência dessa alta quantidade de clientes selecionados, o Critical-FL também apresentou o maior gasto energético, uma vez que cada cliente adicional implica em maior tempo de processamento agregado. Esse comportamento também pode ser observado para ambos níveis de heterogeneidade. Figura 5 explicita tais resultados.

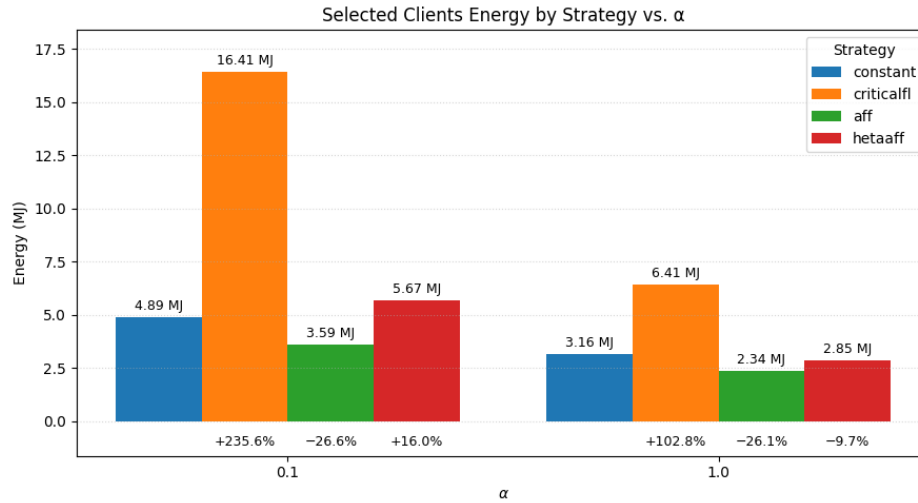


Figura 5: Gráfico do gasto energético durante o treinamento para os clientes selecionados do experimento 1

Além do alto gasto de energia, proporcional à elevada quantidade de clientes selecionados, também foi observado um alto custo em comunicação, uma vez que mais clientes implica em novas transmissões entre clientes e servidor. Figura 6 explicita tal resultado.

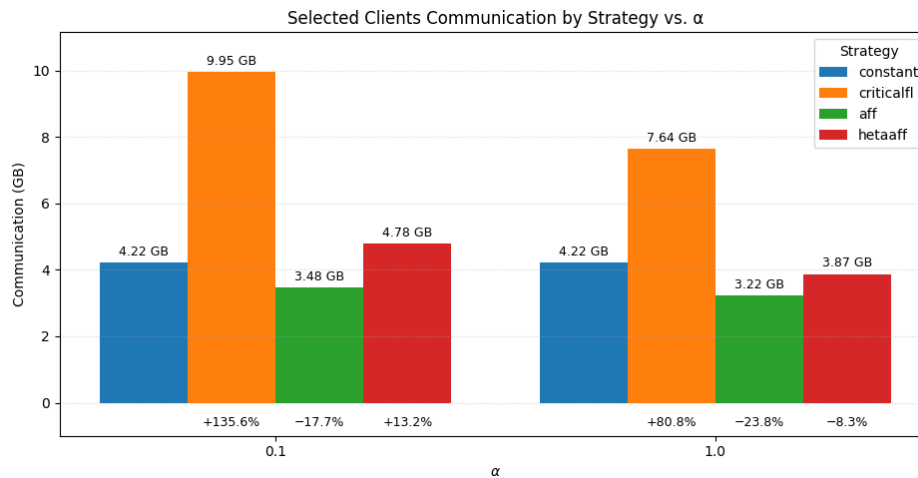


Figura 6: Gráfico do custo em comunicação durante o treinamento para os clientes selecionados do experimento 1

Contudo, esse aumento no número de clientes selecionados não implicou em uma maior performance, uma vez que o Critical-FL empatou em segundo lugar de performance com o HETAFF para o caso de maior heterogeneidade e com o AFF para o caso de menor heterogeneidade estatística. Tal resultado pode ser observado na Figura 7.

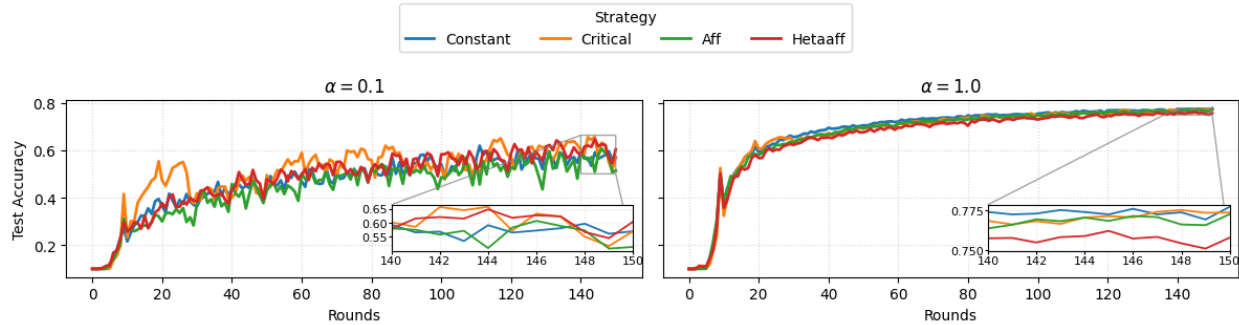


Figura 7: Gráfico da acurácia no conjunto de teste ao longo do treinamento do experimento 1. Nesse gráfico é possível observar que para $\alpha = 0,1$, inicialmente o Critical-FL consegue melhor desempenho, mas ao longo das épocas os outros modelos o alcançam e por fim o HETA AFF termina com a maior acurácia

A respeito do HETA AFF, é possível observar que, no cenário não-IID, ele obteve a segunda maior quantidade de participantes acumulados, embora ainda significativamente mais próximo das demais estratégias do que do Critical-FL. Esse comportamento decorre do mecanismo adaptativo do HETA AFF, o algoritmo ajusta gradualmente o *fit fraction* com base na variação local da perda, ampliando a participação apenas quando há indícios de instabilidade estatística, mas de forma controlada e não tão brusca quanto o Critical-FL, de modo a evitar a explosão de participação observada no Critical-FL, mantendo uma adaptação mais suave e eficiente. Conseqüentemente, ocupou também a segunda colocação em gasto energético, porém ficando com apenas 16% a mais em relação ao FedAvg, enquanto que o Critical-FL ficou com 235,6% a mais. Vale notar ainda que obteve a maior acurácia ao final das 150 rodadas de treinamento. Por outro lado, para o cenário próximo ao IID, o HETA AFF obteve a terceira posição tanto em número de participantes selecionados, quanto em consumo energético. Diferentemente do caso não-IID, nesta configuração ele apresentou o pior resultado de acurácia, indicando menor vantagem quando a heterogeneidade é reduzida.

Além disso, é possível verificar que o AFF foi o que obteve a menor quantidade acumulada de participantes para ambos casos de heterogeneidade. Esse resultado ocorre porque o AFF ajusta o *fit fraction* apenas de forma reativa e pouco agressiva, reduzindo a participação sempre que identifica estabilização na curva de perda local e realizando apenas aumentos modestos quando a perda volta a oscilar. Isso faz com que o AFF priorize economia de recursos e mantenha uma seleção mais conservadora de clientes, resultando consistentemente na menor quantidade de participantes entre todas as estratégias. Contudo, obteve o pior desempenho preditivo no caso não-IID e ficou empatado em segundo lugar no caso IID.

5.2 Experimento 2

Com o experimento 2 finalizado, algumas observações interessantes puderam ser extraídas a partir dos resultados. No que diz respeito à quantidade acumulada de participantes, tanto para a estratégia Div-FL, quanto na Oort, a ordem relativa entre os algoritmos de variação de *fit fraction* permaneceu praticamente a mesma que a do experimento 1. O Critical-FL continuou apresentando uma curva mais alta e com um grande salto no início, devido a sua característica de duplicar a quantidade de clientes selecionados, enquanto que as demais estratégias mantiveram comportamentos mais estáveis e consistentes, semelhantes ao outro experimento. Tais gráficos podem ser vistos na Figura 8 e 9 abaixo.

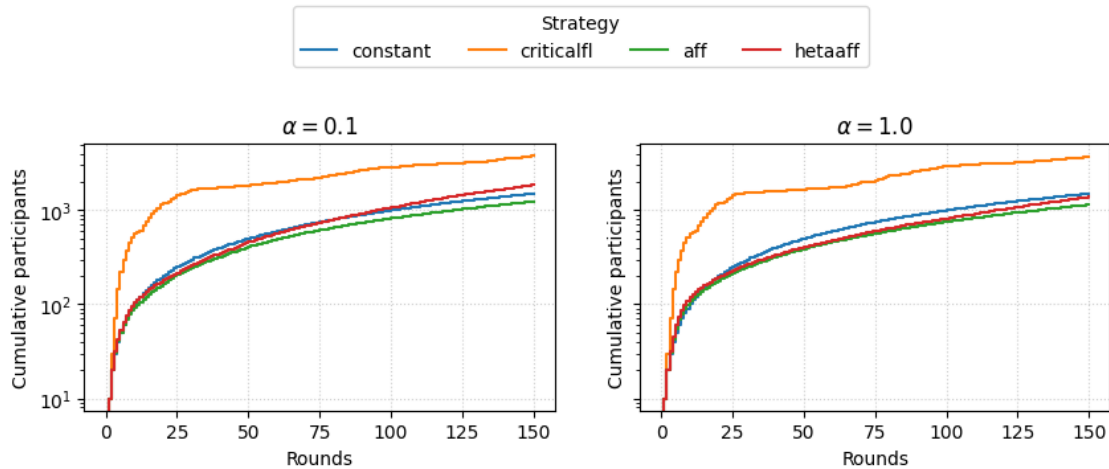


Figura 8: Gráfico da participação cumulativa de clientes durante o treinamento para o experimento 2, com a estratégia Div-Fl de seleção de clientes

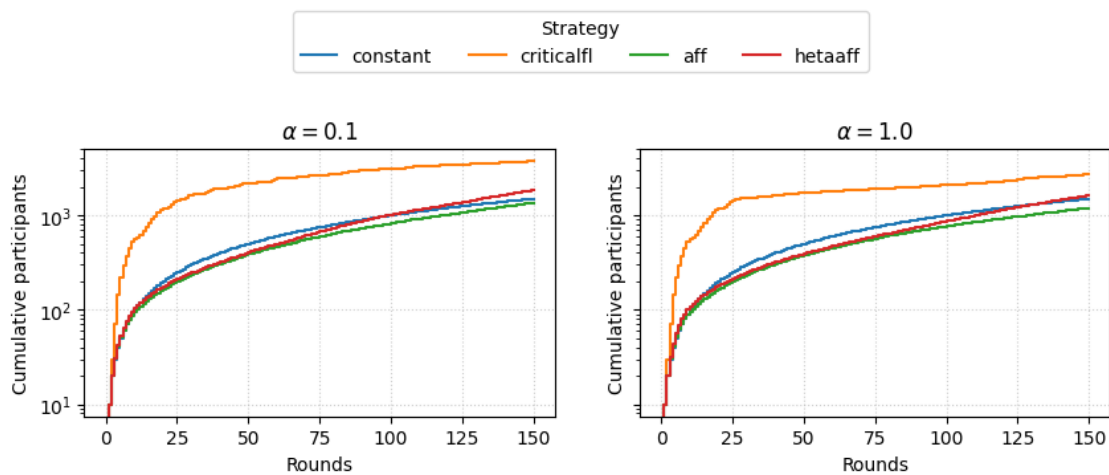


Figura 9: Gráfico da participação cumulativa de clientes durante o treinamento para o experimento 2, com a estratégia Oort de seleção de clientes

Quanto ao custo energético, a manutenção de uma ordem semelhante à da quantidade acumulada de clientes fez com que os padrões observados fossem, novamente, próximos aos do experimento anterior. Sob a política Div-Fl, o HETA AFF ficou em segundo lugar em consumo energético para $\alpha = 0,1$ e em terceiro para $\alpha = 1$, apresentando, nessa estratégia de seleção, um gasto de energia maior do que no caso da seleção aleatória. Esse aumento ocorre porque o Div-FL tende a selecionar clientes mais diversos entre si, o que frequentemente inclui dispositivos com maior custo de processamento local ou com tempos de execução mais longos devido à variabilidade de suas cargas de trabalho, refletindo diretamente no consumo energético total. Já sob o Oort, a posição relativa do HETA AFF só mudou para o $\alpha = 1$, ficando em segundo lugar em transmissão de dados. Esse algoritmo de seleção tende a priorizar clientes com maior utilidade estatística, geralmente clientes com mais dados ou com contribuições mais relevantes para a convergência, resultando em treinamentos locais mais intensivos. Por conta disso, o gasto energético do HETA AFF sob o Oort aumentou em

relação à seleção aleatória, mas ainda assim permaneceu menor do que o observado com o Div-FL, já que o Div-FL amplia a diversidade de forma mais agressiva e frequentemente inclui clientes mais custosos energeticamente. Tais resultados podem ser vistos nas Figuras 10 e 11 abaixo.

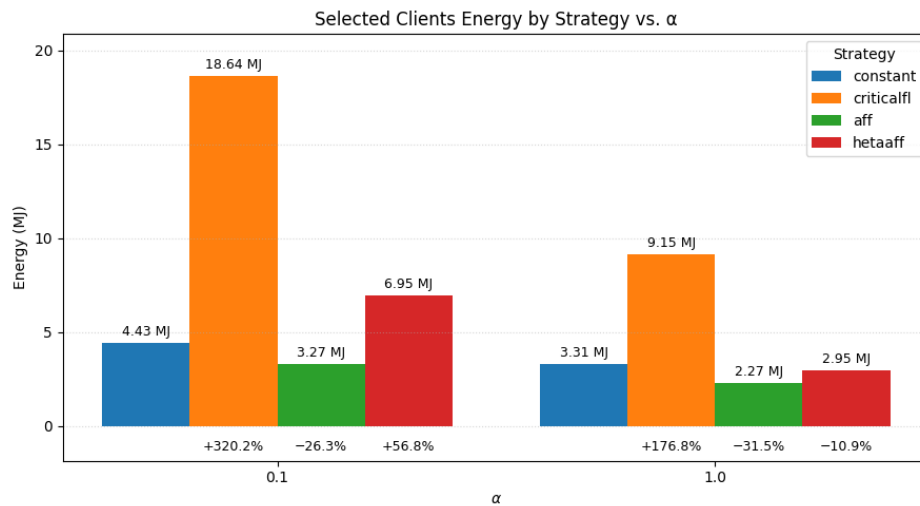


Figura 10: Gráfico do gasto energético dos clientes selecionados durante o treinamento para o experimento 2, com a estratégia Div-FL de seleção de clientes

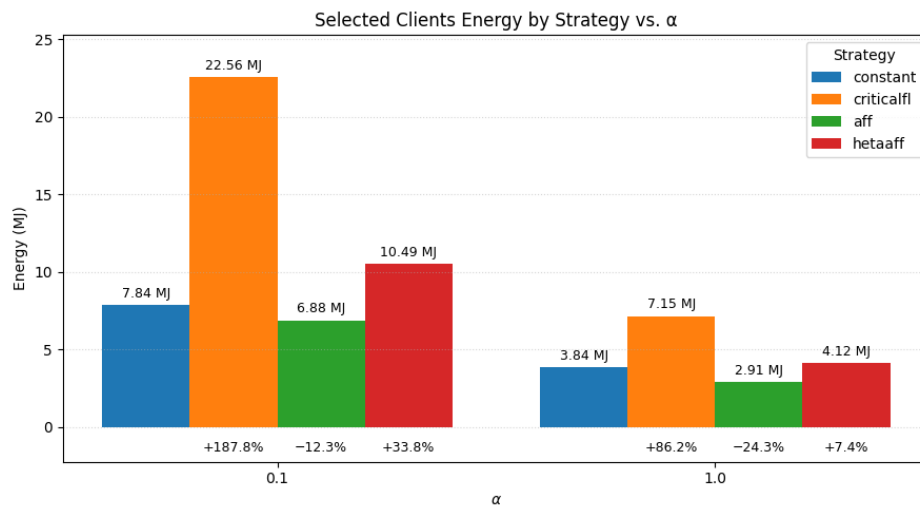


Figura 11: Gráfico do gasto energético dos clientes selecionados durante o treinamento para o experimento 2, com a estratégia Oort de seleção de clientes

Em relação aos custos de comunicação, observou-se um comportamento análogo ao verificado no consumo energético, uma vez que o volume total de dados transmitidos acompanha diretamente a quantidade acumulada de clientes selecionados. Com isso, sob a política Div-FL, o HETA AFF ficou em segundo lugar em quantidade de dados transmitidos para $\alpha = 0,1$ e em terceiro para $\alpha = 1$, apresentando, nessa estratégia, um volume de comunicação maior do que no caso da seleção aleatória, efeito esperado, dado que o Div-FL seleciona clientes mais diversos e, portanto, aumenta a frequência e o custo das transmissões. Já sob o Oort, a posição relativa do HETA AFF só mudou

para $\alpha = 1$, passando a ocupar o segundo lugar; e, embora o Oort tenha levado a um volume maior de comunicação do que a seleção aleatória, ainda assim manteve custos menores do que os observados com o Div-FL. De modo geral, o Oort foi o que mais aumentou a quantidade de dados transmitidos no caso $\alpha = 0,1$, enquanto o Div-FL foi o que apresentou maior volume quando $\alpha = 1$. Esses padrões reforçam que a escolha da estratégia de seleção de clientes influencia diretamente os custos sistêmicos de comunicação, conforme ilustrado nas Figuras 12 e 13.

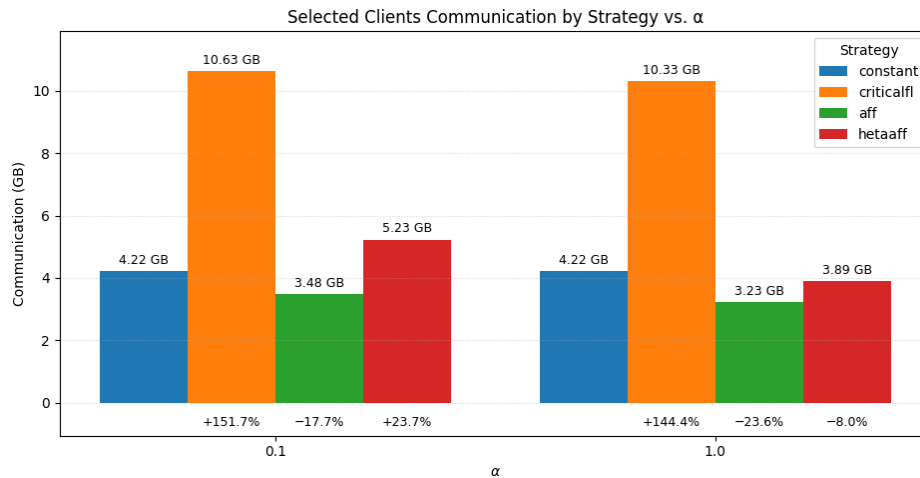


Figura 12: Gráfico do volume de comunicação de clientes selecionados durante o treinamento para o experimento 2, com a estratégia Div-FL de seleção de clientes

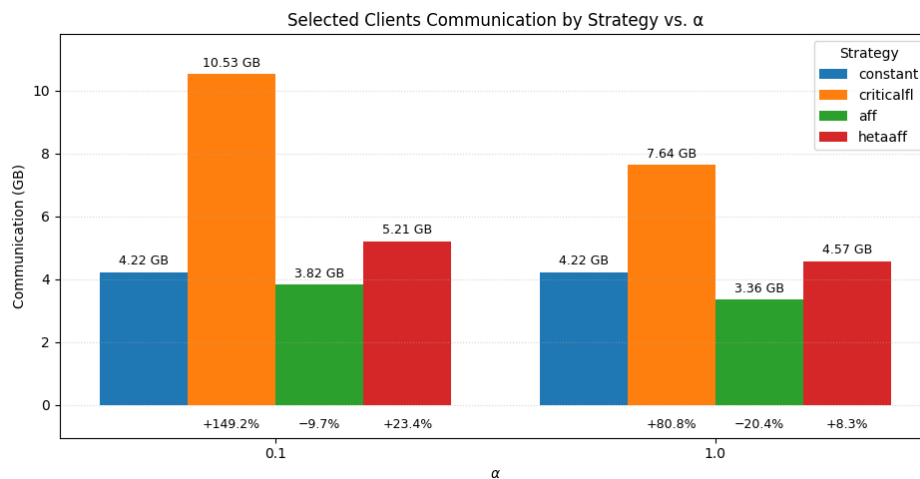


Figura 13: Gráfico do volume de comunicação de clientes selecionados durante o treinamento para o experimento 2, com a estratégia Oort de seleção de clientes

Apesar desse acréscimo nos custos, a melhora no desempenho preditivo do HETAFF ao utilizar o Oort foi pouco significativa, indicando que, para esse algoritmo, a política orientada por utilidade não trouxe benefícios proporcionais ao aumento de consumo energético. Além disso, observa-se que o Critical-FL atinge valores de acurácia mais elevados nas primeiras rodadas, destacando-se em relação às demais estratégias, tanto no Div-FL quanto no Oort para $\alpha = 0,1$. Esse comportamento

inicial ocorre porque a política do Critical-FL duplica a participação de clientes nas primeiras rodadas, aumentando rapidamente a diversidade e o volume de atualizações enviadas ao servidor, o que acelera a queda inicial da perda. Contudo, essa vantagem não se sustenta ao longo do treinamento, pois a elevada heterogeneidade dos gradientes resultante desta seleção ampliada torna a convergência menos estável, fazendo com que as demais estratégias eventualmente alcancem o seu desempenho de forma mais controlada e segura. Tais resultados podem ser visualizados nas Figuras 14 e 15 abaixo.

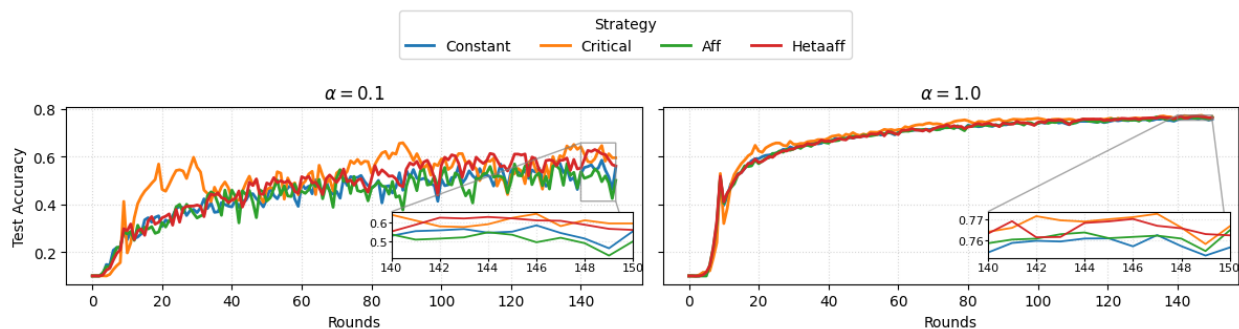


Figura 14: Gráfico da acurácia no conjunto de teste ao longo do treinamento do experimento 2, com a estratégia Div-Fl de seleção de clientes

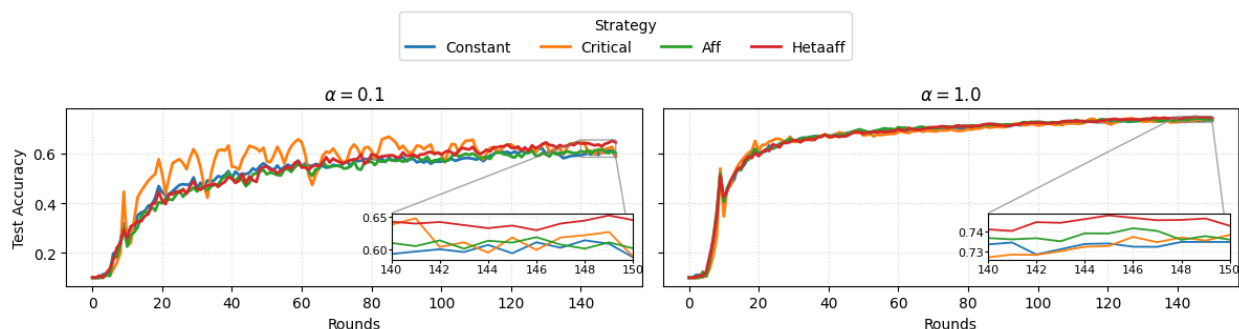


Figura 15: Gráfico da acurácia no conjunto de teste ao longo do treinamento do experimento 2, com a estratégia Oort de seleção de clientes

5.3 Experimento 3

Os resultados do experimento 3 forneceram observações relevantes sobre o comportamento das estratégias em tarefas que utilizam arquiteturas e tipos de dados distintos dos cenários anteriores. No caso da arquitetura Kwt3, aplicada ao reconhecimento de comandos de voz, observou-se que a quantidade acumulada de participantes manteve o mesmo padrão já identificado nos experimentos anteriores, tendência que também se repetiu para a arquitetura Lstm, usada na tarefa de predição de texto. Padrão este que é explicado pela forma como os algoritmos foram estabelecidos. Resultados de quantidade acumulada para cada tarefa podem ser vistos nas Figuras 16 e 17.

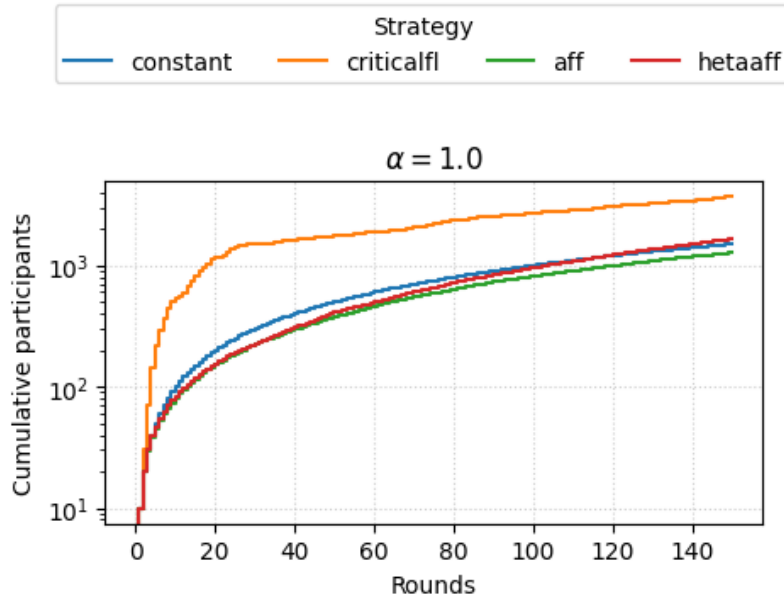


Figura 16: Gráfico da participação cumulativa de clientes durante o treinamento para o experimento 3, com o modelo Kwt3

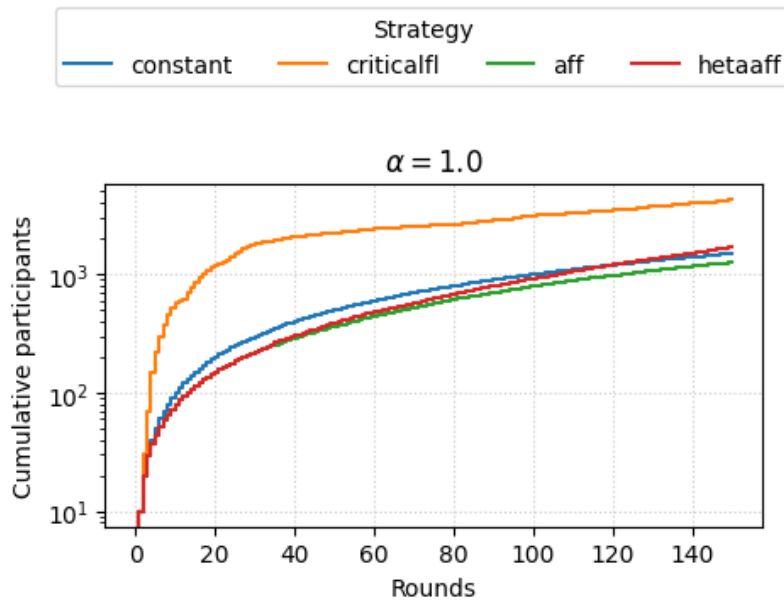


Figura 17: Gráfico da participação cumulativa de clientes durante o treinamento para o experimento 3, com o modelo Lstm

No que se refere ao gasto energético, verificou-se novamente a manutenção da ordem relativa entre as estratégias, mudando apenas a magnitude do consumo total, explicado pela característica e estrutura de cada arquitetura, sendo a Lstm a que se destacou com um gasto mais elevado em relação às outras arquiteturas, por conta da sua complexidade interna de funcionamento. Tais resultados podem ser observados nas Figuras 18 e 19 abaixo.

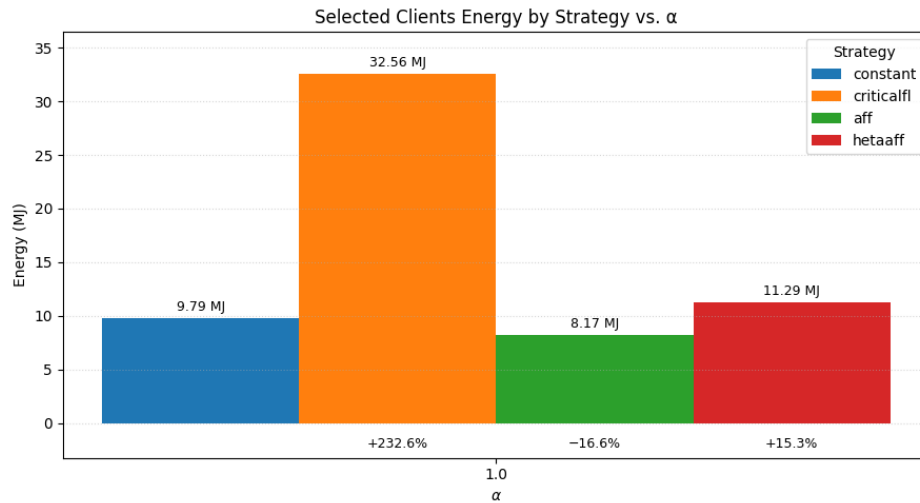


Figura 18: Gráfico do gasto energético dos clientes selecionados durante o treinamento para o experimento 3, com o modelo Kwt3

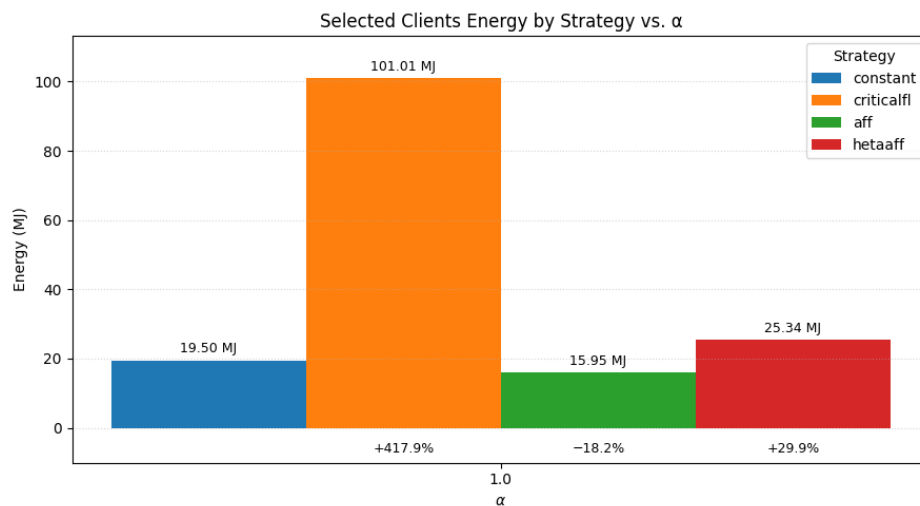


Figura 19: Gráfico do gasto energético dos clientes selecionados durante o treinamento para o experimento 3, com o modelo Lstm

No critério de custo de transmissão, a ordem relativa permaneceu praticamente inalterada, entretanto, a quantidade absoluta de dados enviados aumentou significativamente, devido ao maior tamanho dos modelos utilizados nesse cenário. Por isso, a Kwt3 apresentou o maior volume de transmissão entre as arquiteturas avaliadas, uma vez que seus parâmetros e tensores de atualização são relativamente maiores do que os da Lstm. Esse aumento impacta diretamente o custo de comunicação, pois no aprendizado federado o tamanho do modelo determina o volume de dados transmitidos a cada rodada, independentemente da estratégia de seleção ou do nível de heterogeneidade estatística. Tais conclusões podem ser obtidas conferindo as Figuras 20 e 21 abaixo.

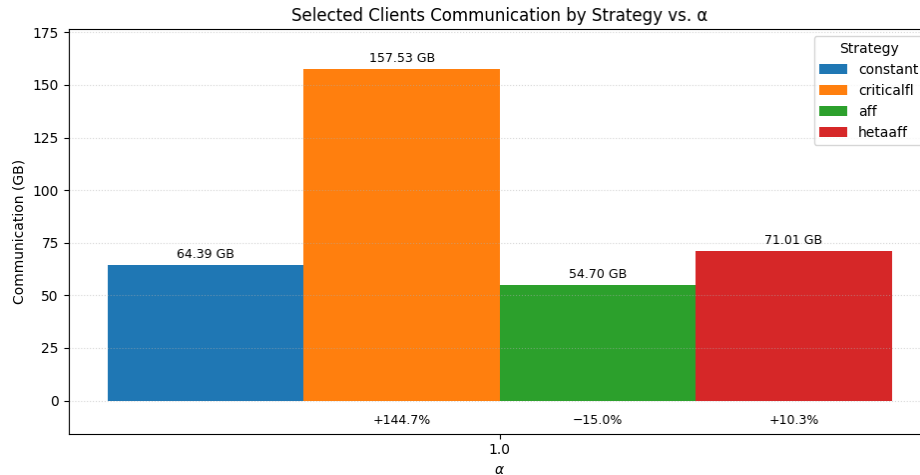


Figura 20: Gráfico do volume de comunicação de clientes selecionados durante o treinamento para o experimento 3, com o modelo Kwt3

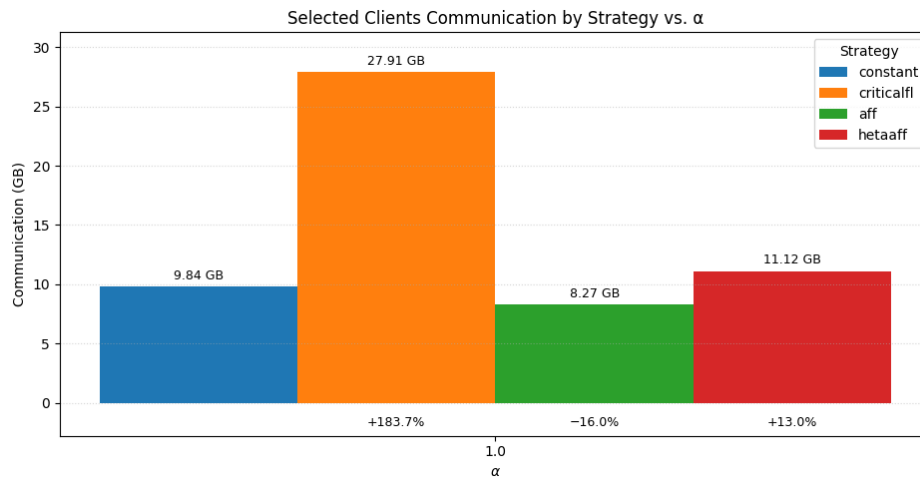


Figura 21: Gráfico do volume de comunicação de clientes selecionados durante o treinamento para o experimento 3, com o modelo Lstm

Quanto à acurácia, com a utilização de outros contextos de tarefas, como reconhecimento de áudio e predição de texto, a diferença entre as estratégias se tornou mais evidente. Nos experimentos com o Kwt3, os resultados ficaram mais distantes entre si, enquanto que para a Lstm, as diferenças entre as estratégias foram menores. Além disso, é possível verificar que o uso da estratégia HETAFF é a que obteve maior desempenho preditivo, ao final das 150 rodadas de treinamento, em ambos cenários. Tais resultados podem ser vistos nas Figuras 22 e 23.

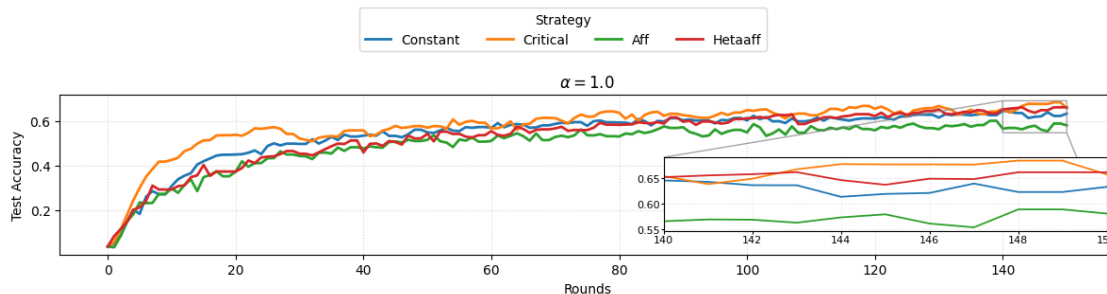


Figura 22: Gráfico da acurácia no conjunto de teste ao longo do treinamento do experimento 3, com o modelo Kwt3

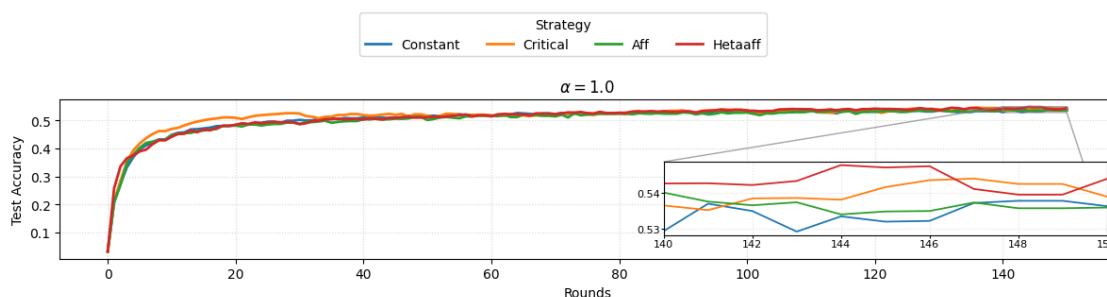


Figura 23: Gráfico da acurácia no conjunto de teste ao longo do treinamento do experimento 3, com o modelo Lstm

5.4 Impacto da Heterogeneidade

A partir dos resultados obtidos nos experimentos, é possível analisar de forma mais clara o impacto da heterogeneidade estatística sobre o desempenho das estratégias de variação do *fit fraction*. Nos cenários não-IID, tanto no experimento base quanto no experimento 2, o HETA AFF apresentou, de forma consistente, um desempenho preditivo superior ao das demais estratégias, evidenciando sua maior robustez em contextos com forte divergência entre os clientes. Já no cenário mais próximo de IID, observou-se uma redução nessa vantagem: no experimento base, o HETA AFF passou a apresentar desempenho inferior às demais estratégias, embora a diferença observada tenha sido menor do que nos casos não-IID.

No experimento 2, o comportamento foi influenciado pelas políticas de seleção de clientes. Sob o Oort, o HETA AFF obteve o melhor desempenho em ambos os níveis de heterogeneidade, mantendo-se competitivo mesmo quando a distribuição dos dados era mais homogênea. Já sob o Div-Fl, a estratégia apresentou o segundo melhor desempenho para $\alpha = 0,1$ e a terceira colocação para $\alpha = 1$. Esse comportamento reforça que o HETA AFF tende a se beneficiar mais dos cenários não-IID, onde sua adaptação baseada em heterogeneidade tem maior impacto, evidenciando e reforçando o seu princípio de funcionamento.

Por fim, observa-se que, em termos absolutos, todas as estratégias alcançaram valores de acurácia mais elevados nos cenários com $\alpha = 1$, o que é esperado, já que distribuições mais homogêneas entre clientes facilitam a convergência do modelo global.

5.5 Efeito das Estratégias de Seleção

A partir dos resultados obtidos no experimento 2, é possível avaliar o impacto das estratégias de seleção de clientes combinadas com as estratégias de variação de *fit fraction*. No contexto do Oort, o HETA AFF apresentou o melhor desempenho preditivo nos dois níveis de heterogeneidade avaliados, superando consistentemente as demais estratégias. Isso sugere que a priorização baseada em utilidade adotada pelo Oort tende a favorecer a adaptação sensível à heterogeneidade realizada pelo HETA AFF.

Por outro lado, sob a política Div-Fl, os resultados mostram um comportamento distinto: o HETA AFF alcançou a segunda melhor posição para $\alpha = 0,1$ e a terceira posição para $\alpha = 1$, indicando que essa política de seleção não maximiza seu potencial de desempenho. Nesse caso, o Critical-Fl apresentou o melhor resultado em ambos os cenários de heterogeneidade, sugerindo uma combinação mais favorável entre esse método de ajuste e a política de seleção baseada em diversidade.

Além disso, quando comparados aos resultados do Experimento 1 (com seleção aleatória), observou-se que o desempenho numérico do HETA AFF sob o Oort foi superior ao obtido no cenário anterior, reforçando a importância de ambas estratégias de seleção na qualidade do aprendizado global.

5.6 Influência das Arquiteturas de Redes

A partir dos resultados obtidos no Experimento 3, é possível avaliar o impacto das arquiteturas de rede sobre o comportamento das estratégias de variação do *fit fraction* em diferentes modalidades de tarefa no contexto federado. Para essa análise, consideram-se principalmente os Experimentos 1 e 3, uma vez que ambos utilizam seleção aleatória de clientes, permitindo isolar com maior precisão o efeito da arquitetura, ainda que os particionamentos adotados nos dois experimentos não sejam idênticos.

Nos cenários de classificação de áudio (Kwt3) e predição de texto (Lstm), observou-se que o HETA AFF apresentou o melhor desempenho preditivo entre todas as estratégias avaliadas. Como esses cenários utilizam particionamentos baseados na própria estrutura dos dados, refletindo padrões reais de uso dos usuários, o grau de heterogeneidade é naturalmente elevado, aproximando-se do comportamento observado no Experimento 1 para $\alpha = 0,1$, caracterizado como fortemente não-IID. Com isso, tem-se que esses resultados indicam que o HETA AFF mantém um comportamento estável e competitivo em cenários altamente heterogêneos, independentemente da tarefa executada ou da arquitetura utilizada. Assim, sua vantagem não se restringe ao domínio de visão computacional, demonstrando potencial de generalização para outras modalidades, como áudio e texto.

5.7 Análise de Custos: Comunicação e Energia

Para finalizar as análises dos resultados, é necessário observar os custos envolvidos no processo de treinamento, tanto em termos de transmissão de dados quanto de energia computacional consumida pelos clientes. Embora detalhes específicos já tenham sido discutidos nos experimentos individuais, esta seção consolida as informações por meio da Figura 24, que resume o comportamento de cada estratégia em todos os cenários avaliados.

Cenários	Experimento 1		Experimento 2				Experimento 3	
Seletor Clientes	Random		Div-FI		Oort		Random	
Arquitetura Rede	ShuffleNet		ShuffleNet		ShuffleNet		Kwt3	Lstm
Particionamentos	$\alpha = 0,1$	$\alpha = 1$	$\alpha = 0,1$	$\alpha = 1$	$\alpha = 0,1$	$\alpha = 1$	Particionamento Natural	
Gasto Energético (MJ)	Critical-FI (16,41) Hetaaff (5,67) Constant (4,89) Aff (3,59)	Critical-FI (6,41) Constant (3,16) Hetaaff (2,85) Aff (2,34)	Critical-FI (18,64) Hetaaff (6,95) Constant (4,43) Aff (3,27)	Critical-FI (9,15) Constant (3,31) Hetaaff (2,95) Aff (2,27)	Critical-FI (22,56) Hetaaff (10,49) Constant (7,84) Aff (6,88)	Critical-FI (7,15) Hetaaff (4,12) Constant (3,84) Aff (2,91)	Critical-FI (32,56) Hetaaff (11,29) Constant (9,79) Aff (8,17)	Critical-FI (101,01) Hetaaff (25,34) Constant (19,50) Aff (15,95)
Gasto em Comunicação (GB)	Critical-FI (9,95) Hetaaff (4,78) Constant (4,22) Aff (3,48)	Critical-FI (7,64) Constant (4,22) Hetaaff (3,87) Aff (3,22)	Critical-FI (10,63) Hetaaff (5,23) Constant (4,22) Aff (3,48)	Critical-FI (10,33) Constant (4,22) Hetaaff (3,89) Aff (3,23)	Critical-FI (10,53) Hetaaff (5,21) Constant (4,22) Aff (3,82)	Critical-FI (7,64) Hetaaff (4,57) Constant (4,22) Aff (3,36)	Critical-FI (157,53) Hetaaff (71,01) Constant (64,39) Aff (54,70)	Critical-FI (27,91) Hetaaff (11,12) Constant (9,84) Aff (8,27)

Figura 24: Resultados obtidos em termos de custos energéticos e de transmissão de dados para todos os cenários e configurações disponíveis. Em destaque em cada célula, encontra-se a estratégia que apresentou o melhor desempenho preditivo

A partir da análise da tabela, observa-se que a estratégia Critical-FI é consistentemente a que apresenta maior consumo energético e maior volume de transmissão de dados. Contudo, esse custo elevado não se traduz em melhor desempenho preditivo na maioria dos casos. A única exceção ocorre quando combinada à política de seleção Div-FI utilizando a arquitetura ShuffleNet, no qual o Critical-FI atinge a melhor performance nos dois níveis de heterogeneidade avaliados.

Em contraste, a estratégia AFF demonstrou ser a opção mais econômica, apresentando o menor custo energético e de transmissão em todos os cenários avaliados, reforçando o seu princípio de atuação, diminuir os custos. No entanto, essa redução de custos vem acompanhada de uma limitação importante: o AFF não alcançou o melhor desempenho preditivo em nenhum dos experimentos, indicando que sua economia é obtida em detrimento de sua performance. Ao analisar o HETA AFF, nota-se que seu custo energético permanece relativamente próximo ao AFF, particularmente quando comparado ao aumento substancial observado no Critical-FI. Essa característica reforça o equilíbrio alcançado pelo HETA AFF entre custo e desempenho.

Considerando os 8 cenários totais avaliados, obtêm-se as seguintes frequências de melhor desempenho:

- HETA AFF: 5 vezes
- Critical-FI: 2 vezes
- FedAvg: 1 vez
- AFF: 0 vezes

Assim, o HETA AFF obteve o melhor resultado em aproximadamente 63% dos cenários, consolidando-se como a estratégia de variação de *fit fraction* mais eficaz no conjunto geral de experimentos. Essa frequência reforça sua robustez frente a diferentes combinações de arquitetura, heterogeneidade estatística e políticas de seleção de clientes, mantendo boa performance mesmo com custos de comunicação e energia apenas moderadamente superiores ao AFF, e substancialmente menores que os do Critical-FI.

6 Conclusões

Este trabalho teve como objetivo ampliar a avaliação do algoritmo HETA AFF, originalmente proposto como uma estratégia adaptativa de ajuste do *fit fraction* em ambientes federados heterogêneos. A fim de expandir a avaliação proposta pelo artigo original, esta avaliação buscou

ser mais abrangente, incorporando além da métrica de acurácia, os custos de comunicação, consumo energético e participação de clientes, além de explorar novos contextos experimentais, como diferentes arquiteturas de rede neural, diferentes tipos de dados e diferentes políticas de seleção de clientes.

Os experimentos conduzidos revelaram que o HETA AFF apresenta um comportamento consistente e robusto em condições de alta heterogeneidade estatística, superando, na maioria dos casos, as demais estratégias avaliadas (FedAvg, AFF e Critical-Fl). No cenário não-IID, sua vantagem preditiva foi bastante evidente, reforçando que sua adaptação dinâmica da fração de clientes é particularmente eficaz quando a diversidade estatística entre os participantes representa um desafio significativo para a convergência global.

Em termos de custo, observou-se que o HETA AFF mantém um equilíbrio, embora não seja a estratégia mais econômica, seu consumo de energia e volume de transmissão permanecem significativamente inferiores ao Critical-Fl, ao mesmo tempo em que entrega desempenho superior na maior parte dos cenários. Comparativamente ao AFF, que é a opção mais econômica, o HETA AFF obtém um aumento modesto de custo, mas com ganhos expressivos de desempenho, tornando-se uma alternativa competitiva para aplicações reais onde o equilíbrio entre custo e eficácia é essencial.

Por fim, este trabalho ampliou de forma significativa o entendimento sobre o comportamento do HETA AFF em diferentes cenários, fornecendo uma avaliação mais completa sobre seu equilíbrio entre eficiência, desempenho e custos.

Referências

- [1] F. Souza, P. Alves, M. Rodrigues, F. Maciel, A. Souza, L. Villas and L. F. Bittencourt, “Tackling the effect of local model updates heterogeneity on adaptive fit fraction”, *FLTA*. 2025.
- [2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. *Proceedings of Machine Learning Research*, A. Singh and J. Zhu, Eds., vol. 54. PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: <https://proceedings.mlr.press/v54/mcmahan17a.html>
- [3] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council. [Online]. Available: <https://data.europa.eu/eli/reg/2016/679/oj>
- [4] S. C. of the Thirteenth National People’s Congress. Personal information protection law of the people’s republic of china. [Online]. Available: <http://en.npc.gov.cn.cdurl.cn/2021-12/29/c694559.htm>
- [5] DM. Schreyer, T. Sattarov, and D. Borth, “Federated and privacy-preserving learning of accounting data in financial statement audits,” in *Proceedings of the Third ACM International Conference on AI in Finance*, ser. *ICAIF ’22*. New York, NY, USA: Association for Computing Machinery, 2022, p. 105–113. [Online]. Available: <https://doi.org/10.1145/3533271.3561674>
- [6] Z. L. Teo, L. Jin, N. Liu, S. Li, D. Miao, X. Zhang, W. Y. Ng, T. F. Tan, D. M. Lee, K. J. Chua, J. Heng, Y. Liu, R. S. M. Goh, and D. S. W. Ting, “Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture,” *Cell Reports Medicine*, vol. 5, no. 2, p. 101419, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666379124000429>.

- [7] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, “Federated learning for the internet of things: Applications, challenges, and opportunities,” *IEEE Internet of Things Magazine*, vol. 5, no. 1, pp. 24–29, 2022.
- [8] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2, 429-450.
- [9] G. Yan, H. Wang, X. Yuan, and J. Li, “Criticalfl: A critical learning periods augmented client selection framework for efficient federated learning,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD ’23. New York, NY, USA: Association for Computing Machinery, 2023, p. 2898–2907. [Online]. Available: <https://doi.org/10.1145/3580305.3599293>
- [10] F. Maciel, J. B. D. da Costa, L. F. G. Gonzalez, A. M. de Souza, L. A. Villas, and L. F. Bittencourt, “Adaptive fit fraction based on model performance evolution in federated learning,” in *2024 11th International Conference on Future Internet of Things and Cloud (FiCloud)*, 2024, pp. 77–84.
- [11] D. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. Li, T. Parcollet, P. de Gusmão, Ni. Lane, “Flower: A Friendly Federated Learning Research Framework”, 2022, arXiv:2007.14390 [cs.LG]
- [12] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: an imperative style, high-performance deep learning library*. Red Hook, NY, USA: Curran Associates Inc., 2019.
- [13] F. Labs, “Flower Datasets 0.5.0 — flower.ai,” <https://flower.ai/docs/datasets/>, 2025, [Accessed 27-06-2025].
- [14] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, arXiv:1704.04861 [cs.CV].
- [15] X. Zhang, X. Zhou, M. Lin, J. Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices”, arXiv:1707.01083 [cs.CV].
- [16] L. Fu, H. Zhang, G. Gao, M. Zhang and X. Liu, ”Client Selection in Federated Learning: Principles, Challenges, and Opportunities,” in *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21811-21819, 15 Dec.15, 2023, doi: 10.1109/JIOT.2023.3299573.
- [17] F. Lai, X. Zhu, H. V. Madhyastha, M. Chowdhury, “Oort: Efficient Federated Learning via Guided Participant Selection”, University of Michigan, 2021, <https://www.usenix.org/system/files/osdi21-lai.pdf> [Accessed 21-11-2025]
- [18] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, D. Bacon, “Federated Learning: Strategies for Improving Communication Efficiency“, arXiv:1610.05492v2 [cs.LG]
- [19] F. Sattler; S. Wiedemann; K. Müller; W. Samek, “Robust and Communication-Efficient Federated Learning From Non-i.i.d. Data”, *IEEE Transactions on Neural Networks and Learning Systems* (Volume: 31, Issue: 9, September 2020), <https://ieeexplore.ieee.org/abstract/document/8889996>

- [20] S. Wang et al., "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," in *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205-1221, June 2019, doi: 10.1109/JSAC.2019.2904348, <https://ieeexplore.ieee.org/abstract/document/8664630>
- [21] T. Nishio and R. Yonetani, "Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge," *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, 2019, pp. 1-7, doi: 10.1109/ICC.2019.8761315, <https://ieeexplore.ieee.org/abstract/document/8761315>
- [22] A. Berg, M. O'Connor, M. T. Cruz, "Keyword Transformer: A Self-Attention Model for Keyword Spotting", *Interspeech 2021* 30 August, 3 September, 2021, Brno, Czechia
- [23] Hochreiter, Sepp Schmidhuber, Jürgen. (1997). "Long Short-Term Memory", *Neural Computation*. 9. 1735-1780. 10.1162/neco.1997.9.8.1735
- [24] Speech Commands Dataset, available on https://huggingface.co/datasets/google/speech_commands
- [25] Shakespeare Dataset, available on <https://huggingface.co/datasets/flwrlabs/shakespeare>