

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Computação Inteligente

Título: Avaliação Sustentável de Auto e Semissupervisão em Aprendizado Federado

Orientador – Leandro Honorato de Souza Silva (leandro.ssilva@upe.br)

Coorientador – Bruno José Torres Fernandes (bjtf@ecomp.poli.br)

O aprendizado federado (FL) emergiu como uma abordagem promissora para o treinamento colaborativo de modelos de aprendizagem de máquina a partir de dados distribuídos, preservando a privacidade ao evitar o compartilhamento direto das informações brutas entre os participantes [1]. Essa característica torna o FL especialmente relevante em cenários sensíveis, como saúde, indústria e aplicações IoT, nos quais a descentralização e a confidencialidade são fundamentais [2]. Entretanto, ambientes federados reais frequentemente apresentam forte heterogeneidade de dados (non-IID) e escassez de rótulos, o que compromete a convergência, a estabilidade e o desempenho final dos modelos [1], [3], [4].

Nos últimos anos, técnicas de aprendizado semissupervisionado e autossupervisionado (SSL – *Self-supervised Learning*) mostraram-se eficazes para mitigar esses desafios, ao permitir que modelos aprendam representações robustas a partir de dados não rotulados, tornando o treinamento mais estável sob heterogeneidade severa e reduzem a dependência de rótulos de alta qualidade [2]. Abordagens como SSFL [5], SelfFed [6] e frameworks contrastivos personalizáveis [7], reforçam que SSL é particularmente eficaz em condições de escassez de rótulos, desbalanceamento e variação significativa entre clientes. Estudos adicionais mostram que o SSL federado não apenas reduz o impacto da heterogeneidade, mas também pode superar o aprendizado supervisionado descentralizado em tarefas de classificação, detecção, segmentação e séries temporais industriais [8].

Apesar desses avanços, observa-se que grande parte da literatura se concentra predominantemente em métricas de acurácia, negligenciando um aspecto relevante: o custo computacional e consequentemente energético associado ao treinamento de modelos federados auto ou semissupervisionados. O aumento da complexidade dos modelos, do número de rodadas de comunicação e da quantidade de operações necessárias para técnicas como a geração de pseudo-rotulos, aprendizagem por contraste e geração de múltiplas vistas intensifica o consumo energético — um ponto crítico à luz das ODS da ONU, que destacam a necessidade de tecnologias mais eficientes, sustentáveis e de baixo impacto ambiental (ODS 7, 9, 12 e 13).

Diante desse cenário, esta proposta de mestrado visa investigar estratégias auto e semi supervisionadas aplicadas ao aprendizado federado que incorporem explicitamente o custo computacional como métrica de otimização, buscando conciliar desempenho, privacidade e sustentabilidade. O objetivo é avaliar os métodos da literatura quanto à eficiência computacional, propondo uma métrica integrada que considere simultaneamente acurácia, eficiência de comunicação e consumo energético. Espera-se demonstrar que é possível obter modelos federados robustos e precisos, reduzindo significativamente o custo computacional, promovendo práticas de IA alinhadas aos princípios de sustentabilidade e uso responsável de recursos.

Referências Bibliográficas

- [1] Z. Song, X. Yang, Y. Zhang, X. Fu, Z. Xu, e I. King, “A systematic survey on federated semi-supervised learning”, em *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, Jeju, Korea: International Joint Conferences on Artificial Intelligence Organization, ago. 2024, p. 8195–8203. doi: 10.24963/ijcai.2024/911.
- [2] L. Wang, K. Zhang, Y. Li, Y. Tian, e R. Tedrake, “Does Learning from Decentralized Non-IID Unlabeled Data Benefit from Self Supervision?”, em *The Eleventh International Conference on Learning Representations*, fev. 2023. doi: 10.48550/arXiv.2210.10947.
- [3] W. Kim, K. Park, K. Sohn, R. Shu, e H.-S. Kim, “Federated Semi-Supervised Learning with Prototypical Networks”, maio 2022, [Online]. Disponível em: <http://arxiv.org/abs/2205.13921>
- [4] A. Psaltis, A. Kastellos, C. Z. Patrikakis, e P. Daras, “FedLID: Self-Supervised Federated Learning for Leveraging Limited Image Data”, em *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, out. 2023, p. 1039–1048.
- [5] C. He, Z. Yang, E. Mushtaq, S. Lee, M. Soltanolkotabi, e S. Avestimehr, “SSFL: Tackling Label Deficiency in Federated Learning via Personalized Self-Supervision”, em *The Tenth International Conference on Learning Representations*, 2022. [Online]. Disponível em: www.aaai.org
- [6] S. Ali Khowaja, K. Dev, S. Muhammad Anwar, e M. George Linguraru, “SelfFed: Self-supervised federated learning for data heterogeneity and label scarcity in medical images”, *Expert Syst Appl*, vol. 261, fev. 2025, doi: 10.1016/j.eswa.2024.125493.
- [7] C. Huang, X. Chen, Y. Zhang, e H. Wang, “FedCoSR: Personalized Federated Learning With Contrastive Shareable Representations for Label Heterogeneity in Non-IID Data”, *IEEE Trans Cybern*, vol. 55, no 12, p. 5841–5854, dez. 2025, doi: 10.1109/TCYB.2025.3599370.
- [8] M. Russell e P. Wang, “Maximizing model generalization for machine condition monitoring with Self-Supervised Learning and Federated Learning”, *J Manuf Syst*, vol. 71, p. 274–285, dez. 2023, doi: 10.1016/j.jmsy.2023.09.008.