

Green AI e Data Augmentation para Triagem de Úlceras do Pé Diabético no Brasil: um Pipeline Eficiente, Sustentável e Socialmente Orientado

Danielle S. Monteiro¹, Marcio Teixeira², Pedro Pizzigatti Corrêa¹

¹Escola Politécnica – Universidade de São Paulo (USP) – São Paulo, SP – Brasil

²Faculdade de Tecnologia – Universidade de Campinas (UNICAMP) – Campinas, SP – Brasil

dsmonteiro@usp.br, mjt@unicamp.br, pedro.correa@usp.br

Abstract. *Diabetic foot ulcers are a severe complication of diabetes mellitus that heavily burdens Brazil's public health system [Armstrong et al. 2017, DATASUS 2025]. We propose a public-health-oriented computer vision pipeline for classifying and segmenting these lesions, guided by computational efficiency (Green AI) and robustness under limited data. It combines EfficientNet-B0 and MobileNetV3-Large with Batch Normalization-only adaptation, MixUp/CutMix, FixRes, and entropy-guided adaptive inference, auditing energy and CO₂ with CodeCarbon [Schmidt et al. 2021] on DFUC2020/DFUC2021. BN-only models reach competitive AUC, F1-macro, and AUPRC at far lower energy cost than full fine-tune.*

Resumo. *As úlceras do pé diabético são uma complicação grave do diabetes mellitus que sobrecarrega o Sistema Único de Saúde [Armstrong et al. 2017, DATASUS 2025]. Propomos um pipeline de visão computacional orientado à saúde pública para classificar e segmentar essas lesões, guiado por eficiência computacional (Green AI) e robustez com dados limitados. Ele combina EfficientNet-B0 e MobileNetV3-Large com adaptação Batch Normalization-only, MixUp/CutMix, FixRes e inferência adaptativa por entropia, auditando energia e CO₂ com o CodeCarbon [Schmidt et al. 2021] nos desafios DFUC2020/DFUC2021. Os modelos BN-only atingem AUC, F1-macro e AUPRC competitivos com custo energético muito inferior ao full fine-tune.*

1. Introdução

O *diabetes mellitus* permanece entre os principais desafios contemporâneos de saúde pública. No Brasil, o inquérito Vigitel 2019 estimou prevalência de 7,8% nas capitais, correspondendo a aproximadamente 16,8 milhões de adultos em todo o país [IDF 2019, Ministério da Saúde 2020]. Entre suas complicações graves, as úlceras do pé diabético destacam-se por sua associação com infecções, amputações e prejuízos significativos à qualidade de vida [Armstrong et al. 2017]. A mortalidade em cinco anos após amputação maior supera 50% [Armstrong et al. 2017]. Além disso, o ônus é desigual: municípios com menor Índice de Desenvolvimento Humano (IDH) apresentam taxas de amputação significativamente maiores [IBGE 2020].

O Sistema Único de Saúde (SUS) enfrenta desafios para o acompanhamento preventivo: escassez de especialistas, sobrecarga da Atenção Primária à Saúde (APS)

e ausência de ferramentas de triagem acessíveis. Nesse cenário, soluções baseadas em visão computacional podem apoiar a triagem automatizada a partir de fotografias clínicas capturadas por agentes comunitários de saúde (ACS) com smartphones [Alkhalefah et al. 2025, Oliveira et al. 2025].

Contudo, a adoção dessas soluções exige modelos leves e viáveis em ambientes com restrições computacionais. Nesse contexto, *Green AI* deixa de ser apenas preocupação ambiental e integra a discussão sobre inclusão tecnológica e justiça em saúde [Schwartz et al. 2020]. Modelos dependentes de GPUs de alto desempenho são estruturalmente excludentes. Quantificar e reportar o custo energético de sistemas de IA para saúde pública torna visível essa barreira [Strubell et al. 2019].

Apesar dos avanços na classificação de úlceras [Cassidy et al. 2022, Alhasson and Alharbi 2025], persistem lacunas referentes à escassez de dados, heterogeneidade clínica e alto custo computacional de arquiteturas modernas. Este artigo busca enfrentar essas limitações de forma integrada, combinando estratégias de eficiência e robustez. Nossas contribuições incluem: (i) um pipeline *end-to-end* leve (adaptação BN-only, FixRes, MixUp/CutMix, inferência por entropia); (ii) metodologia de auditoria de energia e emissão de CO₂ (CodeCarbon); (iii) comparação rigorosa entre desempenho clínico e eficiência energética; e (iv) protocolo aberto e reproduzível.

2. Trabalhos Relacionados

2.1. Arquiteturas Eficientes e Adaptação Leve

Arquiteturas modernas que exploram escalonamento sistemático, como EfficientNet [Tan and Le 2019], e otimizações específicas para dispositivos móveis, como MobileNetV3 [Howard et al. 2019], oferecem ganhos expressivos de eficiência. Embora abordagens baseadas em Vision Transformers e arquiteturas híbridas (ex: MobileViT) apresentem avanços teóricos [Mehta and Rastegari 2022], sua latência e consumo de memória em hardware legado no SUS ainda configuram um desafio logístico em aberto.

Do ponto de vista de adaptação eficiente, [Frankle et al. 2021] demonstraram que treinar exclusivamente os parâmetros de *Batch Normalization* (BN-only) de redes pré-treinadas reduz o risco de sobreajuste (*overfitting*), viabilizando transferências robustas em domínios de dados limitados. Estratégias estruturais complementares incluem o FixRes [Touvron et al. 2019], que mitiga discrepâncias de resolução entre treinamento e teste de forma econômica, e mecanismos de inferência adaptativa por entropia [Yang et al. 2020], que otimizam o processamento baseando-se na incerteza da predição.

2.2. Segmentação de Úlceras e *Green AI*

Desafios públicos como o DFUC2021 [Cassidy et al. 2022] impulsionaram o desenvolvimento de modelos multiclasse para detecção de isquemia e infecção [Galdran et al. 2022, Niri et al. 2025]. No entanto, a viabilidade de implantação e a eficiência energética são frequentemente negligenciadas nos *benchmarks*. O paradigma de *Green AI* [Schwartz et al. 2020] propõe que o custo computacional seja tratado como métrica científica central, utilizando ferramentas como o CodeCarbon [Schmidt et al. 2021] para auditar o impacto do treinamento de aprendizado profundo [Strubell et al. 2019]. Em paralelo, para lidar com o forte desbalanceamento clínico, técnicas de regularização

geométrica e de textura (MixUp [Zhang et al. 2018] e CutMix [Yun et al. 2019]) são amplamente empregadas. Em domínios médicos, a literatura adverte que tais técnicas exigem hiperparâmetros conservadores para manter a plausibilidade biológica das lesões geradas.

3. Materiais e Métodos

3.1. Dados e Protocolo Experimental

Utilizamos as bases públicas DFUC2020 e DFUC2021 [Yap et al. 2021, Cassidy et al. 2022], com aprovação ética internacional prévia. O DFUC2021 contém 15.683 amostras divididas em classes severamente desbalanceadas (ex: 2.555 infecção vs. 227 isquemia). Do conjunto de treino, 20% foram reservados para validação estratificada por paciente, evitando vazamento de dados. As imagens foram pré-processadas com normalização RGB e padronização de resolução. Para conformidade futura com a Lei Geral de Proteção de Dados (LGPD) no SUS [Brasil 2018], a minimização e anonimização dos dados clínicos são imperativas.

A Tabela 1 sumariza os hiperparâmetros e configurações detalhadas (incluindo parâmetros do MixUp/CutMix, resoluções das fases de adaptação e regularizadores), garantindo a eliminação de ambiguidades metodológicas e a reprodução integral dos experimentos.

Tabela 1. Hiperparâmetros do pipeline para reprodutibilidade integral.

Configuração	Valor adotado
Backbones	EfficientNet-B0 e MobileNetV3-Large (pré-treino ImageNet)
Estratégia BN-only	Camadas conv. congeladas; treina apenas BN (γ , β) + cabeça (GAP $\rightarrow$ Dropout 0,3 $\rightarrow$ Linear)
Resolução (FixRes)	Fase 1: 224 $\times$ 224 px (30 épocas) / Fase 2: 320 $\times$ 320 px (10 épocas)
Otimizador e <i>Scheduler</i>	AdamW (lr=1 $\times$ 10 ⁻³ , wd=1 $\times$ 10 ⁻⁴) / CosineAnnealingLR
MixUp / CutMix	α = 0,4; probabilidade de 0,5 por técnica por amostra
Aumento de dados base	<i>Random resized crop</i> , espelhamento horizontal, <i>color jitter</i> leve
Limiar de entropia (τ)	0,60 (B0) e 0,55 (MobileNetV3-L), definidos via validação
Hardware / Ambiente	NVIDIA RTX 3090 (24 GB) / PyTorch 2.0.1, CUDA 11.8

3.2. Estratégia de Treinamento e Eficiência Adaptativa

Como sumarizado na Tabela 1, adotamos EfficientNet-B0 e MobileNetV3-Large como pilares devido à proporção favorável entre desempenho e pegada de memória. Aplicamos pesos de classe inversamente proporcionais às frequências para combater o desbalanceamento intrínseco à isquemia, além de *temperature scaling* [Guo et al. 2017] para calibração probabilística. Como referência de alto custo (*baseline* interno), treinamos arquiteturas pesadas (ResNet-50 e EfficientNet-B7) em regime de *full fine-tune* mantendo as demais transformações geométricas.

Na inferência adaptativa [Yang et al. 2020], as imagens são processadas inicialmente na resolução base (224 $\times$ 224). O reprocessamento custoso em alta resolução (320 $\times$ 320) ocorre exclusivamente se a incerteza probabilística (Entropia de Shannon) da

primeira etapa exceder o limiar τ estabelecido. Auditorias de consumo elétrico foram realizadas pelo CodeCarbon [Schmidt et al. 2021], reportando energia por lote de predições e gramas de CO₂ equivalente por acerto diagnóstico, adotando o fator de emissão padrão brasileiro (74 gCO₂/kWh).

4. Resultados

4.1. Desempenho Clínico e Pegada de Carbono

A Tabela 2 posiciona o método proposto frente à literatura e aos nossos *baselines* custos no DFUC2021. Ressalta-se que a literatura primária frequentemente omite métricas estritas de desbalanceamento (AUPRC) e indicadores de energia, o que exige cautela em comparações absolutas.

Tabela 2. Comparação de desempenho e eficiência multiclasse (DFUC2021). *
Nosso método; ‡ Baseline de alto custo; N/R: Não Reportado.

Modelo	AUC	F1	AUPRC	AUPRC (Isq)	Wh/1000	CO ₂ e/TP (g)	Ref.
DeiT-Base	0,72	N/R	N/R	N/R	N/R	N/R	[Cassidy et al. 2022]
Inception-V3	0,85	N/R	N/R	N/R	N/R	N/R	[Cassidy et al. 2022]
CNN+Transformer	0,83	0,86	N/R	N/R	N/R	N/R	[Galdran et al. 2022]
ResNet-50 (full FT) ‡	0,85	0,87	~0,71	~0,69	~4,2	~8,1	Atual
EfficientNet-B7 (full FT) ‡	0,88	0,89	~0,74	~0,72	~12,7	~24,3	Atual
EfficientNet-B0 (BN-only) *	0,83	0,91	~0,74	~0,71	~1,8	~3,5	Atual
MobileNetV3-L (BN-only) *	0,81	0,90	~0,72	~0,69	~1,1	~2,1	Atual

Os modelos leves (BN-only + FixRes) entregam AUPRC competitivo e penalidade mínima em AUC frente aos modelos pesados (queda de apenas 2 a 4 pontos percentuais). Em contrapartida, a eficiência energética é notável: o MobileNetV3-L apresenta um consumo energético $\sim 11,5\times$ inferior ao EfficientNet-B7, mitigando drasticamente a emissão de carbono por verdadeiro positivo (2,1g de CO₂e contra 24,3g).

4.2. Ablação e Análise Qualitativa de Segmentação

A Tabela 3 isola a contribuição de cada módulo do *pipeline*. A restrição *BN-only* age como forte regularizador, evitando que a rede decore a base minúscula, enquanto o FixRes e as estratégias de mistura de *patches* (CutMix/MixUp) garantem estabilidade nas métricas F1 para classes minoritárias.

Tabela 3. Impacto da remoção de componentes (medianas sobre 5 sementes).

Componente Ausente	Δ AUC	Δ Custo treino	Observação Principal
(Subst. por Full fine-tune)	+0,02–0,04	$\sim 4\times$ maior	Risco crítico de <i>overfitting</i> computacional
Sem FixRes	–0,02 –0,04	a —	Perda de robustez espacial
Sem MixUp/CutMix	–0,03 –0,05 (F1)	a —	Queda drástica nas classes raras
Sem inferência adaptativa	Sem Δ AUC	$\sim 2,3\times$ maior	Desperdício de inferência em casos triviais

Na tarefa de segmentação, o EfficientNet-B0 obteve IoU global de 81,4% (Dice = 87,2%). A análise qualitativa dos erros demonstrou que bordas bem definidas apresentam

alta precisão ($\text{IoU} > 85\%$), enquanto infecções agudas sofrem subestimação sutil devido à confusão gerada por exsudato espesso. O erro morfológico mais crítico, contudo, ocorre em casos de isquemia severa, onde o modelo superestima a lesão ao confundir tecido necrótico adjacente e intacto com a área ativamente ulcerada, devido à enorme similaridade espectral das cores escuras. Tais achados reforçam a necessidade premente de bases de dados maiores especificamente mapeadas para necrose e isquemia.

5. Discussão e Limitações

5.1. Viabilidade Operacional e Equidade Algorítmica no SUS

Os testes de inferência confirmam a viabilidade prática da arquitetura na ponta assistencial: a latência do MobileNetV3-L opera entre 1 e 3 segundos em *smartphones* comuns via *quantização offline*, caindo para 0,3 a 0,8 segundos em CPUs de Unidades Básicas de Saúde (UBS), ou menos de 0,05 segundos caso roteada para servidores centrais da Rede Nacional de Dados em Saúde (RNDS). Tal perfil permite triagem de risco instantânea por Agentes Comunitários de Saúde sem dependência de internet estável.

Contudo, o principal risco de governança do modelo reside no viés de fototipo e equidade. A base DFUC2021 reflete pacientes do Reino Unido, majoritariamente de fototipos caucasianos (Fitzpatrick I–III). No Brasil, 55,5% da população autodeclara-se preta ou parda [IBGE 2022]. A ausência de amostras clinicamente anotadas de peles ricas em melanina inviabiliza a calibração justa de limiares diagnósticos — como a identificação visual de eritema, que se manifesta de forma radicalmente diferente em fototipos Fitzpatrick V e VI. Modelos implantados sem essa validação interseccional podem apresentar taxas de falsos negativos mais elevadas em pacientes negros, institucionalizando disparidades assistenciais através da IA algorítmica.

5.2. Lacunas de Dados e Perspectivas

O estudo possui limitações inerentes. As aferições ambientais (CodeCarbon) refletem estritamente o *hardware* laboratorial e o mix elétrico local, e não simulam plenamente a exaustão térmica em dispositivos móveis no campo. Estruturalmente, a ausência de uma base nacional e federada de imagens DFU documentando desfechos clínicos é a principal barreira. Para transcender provas de conceito técnica rumo ao impacto, é mandatória a criação de protocolos abertos regidos pela LGPD [Brasil 2018], estabelecendo consentimento livre, captura com iluminação difusa padronizada e licenciamento não comercial.

6. Conclusão

Demonstramos que a adoção de adaptação restrita (BN-only), estratégias conservadoras de aumento de dados e roteamento por entropia permite classificação multiclasse robusta de lesões de pé diabético a uma fração do custo computacional convencional. O *Green AI* mostra-se não apenas viável, mas mandatório para contextos de infraestrutura restrita como a Atenção Primária do SUS. O gargalo para a translação clínica real, entretanto, deixa de ser puramente algorítmico para tornar-se sociotécnico: a implantação ética exige bases de dados diversificadas, auditoria por fototipos de pele e total alinhamento legal. Tratar a tecnologia com responsabilidade estrutural e sustentabilidade é o único caminho para não amplificarmos as desigualdades que a inteligência artificial tem o potencial de corrigir.

Código fonte reproduzível disponível em: <https://github.com/danimonteirodba/dfu-green-ai>.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Uso de Inteligência Artificial

Ferramentas de IA generativa foram utilizadas no processo de escrita para revisão gramatical, concisão e apoio na estruturação do texto. A responsabilidade integral pelo conteúdo metodológico, análises científicas, afirmações factuais e conclusões permanece exclusivamente com os autores humanos, que conduziram a validação técnica de toda a pesquisa.

Referências

- Alhasson, H. F. and Alharbi, S. S. (2025). Advances in image-based diagnosis of diabetic foot ulcers using deep learning and machine learning: a systematic review. *Biomedicine*, 13(12):2928.
- Alkhalefah, S., Alturaiki, I., and Altwaijry, N. (2025). Advancing diabetic foot ulcer care: Ai and generative ai approaches for classification, prediction, segmentation, and detection. *Healthcare*, 13(6):648.
- Armstrong, D. G., Boulton, A. J. M., and Bus, S. A. (2017). Diabetic foot ulcers and their recurrence. *New England Journal of Medicine*, 376(24):2367–2375.
- Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. lei geral de proteção de dados pessoais (lgpd). Diário Oficial da União, Brasília, DF, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 10 mai. 2026.
- Cassidy, B., Kendrick, C., Reeves, N. D., et al. (2022). Diabetic foot ulcer grand challenge 2021: evaluation and summary. In *Diabetic Foot Ulcers Grand Challenge (DFUC 2021)*, *Lecture Notes in Computer Science*, volume 13183. Springer, Cham.
- DATASUS (2025). Morbidade hospitalar do sus (sih/sus): procedimentos relacionados a amputações de membros inferiores. Departamento de Informática do SUS, Ministério da Saúde. Disponível em: <http://tabnet.datasus.gov.br>. Acesso em: jan. 2025.
- Frankle, J., Schwab, D. J., and Morcos, A. S. (2021). Training BatchNorm and only BatchNorm. In *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=vYeQQ29TbvX>.
- Galdran, A., Carneiro, G., and Ballester, M. A. G. (2022). Convolutional nets versus vision transformers for diabetic foot ulcer classification. In *Diabetic Foot Ulcers Grand Challenge (DFUC 2021)*, *Lecture Notes in Computer Science*, volume 13183. Springer.
- Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, *PMLR 70*, pages 1321–1330. <https://proceedings.mlr.press/v70/guo17a.html>.

- Howard, A., Sandler, M., Chu, G., et al. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324.
- IBGE (2020). Pesquisa nacional de saúde 2019: informações sobre domicílios, acesso e utilização dos serviços de saúde. Technical report, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro. <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101748>.
- IBGE (2022). Censo demográfico 2022: resultados do universo — características da população. Technical report, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-demografico-2022.html>. Acesso em: 10 mai. 2026.
- IDF (2019). Idf diabetes atlas. Technical report, International Diabetes Federation. <https://www.diabetesatlas.org>.
- Mehta, S. and Rastegari, M. (2022). Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In *International Conference on Learning Representations (ICLR)*.
- Ministério da Saúde (2020). Vigitel brasil 2019: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico. Technical report, Ministério da Saúde, Brasília. https://bvsms.saude.gov.br/bvs/publicacoes/vigitel_brasil_2019_vigilancia_fatores_risco.pdf.
- Niri, R., Zahia, S., Stefanelli, A., et al. (2025). Wound segmentation with u-net using a dual attention mechanism and transfer learning. *Journal of Imaging Informatics in Medicine*.
- Oliveira, T. C. D., Oliveira, A. F., Araújo, L. C., et al. (2025). Digital health technologies for diabetic foot ulcers: a systematic review of clinical evidence, access inequities, and public health integration. *International Journal of Environmental Research and Public Health*, 22(9):1430.
- Schmidt, V., Goyal, K., Joshi, A., et al. (2021). CodeCarbon: estimate and track carbon emissions from machine learning computing.
- Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12):54–63.
- Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3645–3650.
- Tan, M. and Le, Q. V. (2019). EfficientNet: rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, PMLR 97, pages 6105–6114. <https://proceedings.mlr.press/v97/tan19a.html>.
- Touvron, H., Vedaldi, A., Douze, M., et al. (2019). Fixing the train-test resolution discrepancy. In *Advances in Neural Information Processing Systems (NeurIPS)*, vo-

- lume 32, pages 8250–8260. <https://proceedings.neurips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html>.
- Yang, L., Mehta, N., Chen, D., et al. (2020). Resolution adaptive networks for efficient inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2369–2378.
- Yap, M. H., Cassidy, B., Pappachan, J. M., et al. (2021). Analysis towards classification of infection and ischaemia of diabetic foot ulcers. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4.
- Yun, S., Han, D., Oh, S. J., et al. (2019). CutMix: regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6023–6032.
- Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2018). mixup: beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=r1Ddp1-Rb>.