INVESTIGAÇÃO EXPERIMENTAL DE ESTRATÉGIAS DE TREINAMENTO DA MOBILENETV2 PARA CLASSIFICAÇÃO DE FERIDAS VISANDO APLICAÇÕES EM DISPOSITIVOS MÓVEIS

Pedro Henrique Araújo Mattos Ribeiro[1] - phamr@aluno.ifnnmg.edu.br
Suzana Viana Mota[2] - suzana.mota@ifnmg.edu.br
Karine Alencar Fróes[1] - karine.froes@ifnmg.edu.br
Felipe Augusto Oliveira Mota[1] - felipe.mota@ifnmg.edu.br
[1]Instituto Federal do Norte de Minas Gerais (IFNMG), Campus Januária, Januária,, MG, Brazil
[2]Instituto Federal do Norte de Minas Gerais (IFNMG), Campus Almenara, Almenara, MG, Brazil

Abstract. *O tratamento de feridas representa um desafio para os sistemas de saúde, exigindo métodos de diagnóstico rápidos e precisos. Este trabalho avaliou o desempenho da arquitetura MobileNetV2 na classificação de imagens de feridas para aplicações em dispositivos móveis. Foram realizados experimentos com diferentes configurações de treinamento, incluindo Fine-Tuning, estratégias de particionamento dos dados e cenários com quatro e seis classes. O melhor desempenho foi obtido no cenário com quatro classes, utilizando particionamento 80/20 e Fine-Tuning, alcançando acurácia de 97%. Embora os resultados indiquem o potencial da MobileNetV2 para aplicações móveis, o tamanho reduzido da base de dados limita a generalização do modelo. Como trabalhos futuros, propõe-se ampliar a base de imagens e validar o modelo em dispositivos móveis.*

Keywords: *Dispositivos Móveis, Feridas, Inteligência Artifical Leve, Visão Computacional.*

1. INTRODUÇÃO

O tratamento de lesões crônicas representa um desafio relevante para a saúde pública, pois o diagnóstico tardio pode levar a complicações graves, amputações e aumento da morbimortalidade conforme Sen (2019), Salomé & Gaudencio (2025) e Guest *et al.* (2020). Nesse contexto, Patel *et al.* (2024) defendem que a Inteligência Artificial e as Redes Neurais Convolucionais (CNNs) têm contribuído para o reconhecimento de padrões em imagens clínicas. Paralelamente, a consolidação da saúde móvel (*mHealth*) ampliou o acesso a avaliações em locais com infraestrutura limitada, como defendido por Anisuzzaman *et al.* (2022). Ainda assim, Al-Shabi *et al.* (2025) destacam que a adaptação desses modelos para dispositivos móveis enfrenta limitações de processamento, memória e energia.

Sandler *et al.* (2018), Souza (2025) e Dias *et al.* (2026) defendem o uso do MobileNetV2 por usar convoluções separáveis em profundidade (*depthwise separable convolutions*). Essa configuração diminui a complexidade computacional, trazendo o benefício de viabilizar a inferência em ambientes restritos. De acordo com Anisuzzaman *et al.* (2022), a escassez e o desbalanceamento das bases públicas de imagens tornam-se os principais gargalos no treinamento e no desenvolvimento de modelos leves voltados á identificação de feridas. Essa limitação aumenta o risco de *overfitting* e exige estratégias adequadas de pré-processamento e regularização.

Assim, este trabalho tem como objetivo comparar o desempenho da MobileNetV2 sob diferentes configurações experimentais e estratégias de particionamento de dados, identificando a configuração que ofereça o melhor equilíbrio entre eficiência computacional e capacidade preditiva. Com isso, busca-se fornecer subsídios técnicos e empíricos que propiciem, futuramente, o desenvolvimento de um modelo otimizado para a detecção e identificação de feridas em dispositivos móveis.

2. METODOLOGIA

Conforme Prodanov & Freitas (2013), esta pesquisa caracteriza-se como aplicada, de natureza experimental e abordagem incremental. Foram utilizadas as bases públicas *AZH Wound Care Database* e *Medetec Wound Database* (Anisuzzaman *et al.*, 2022). O modelo foi inicializado por Aprendizado por Transferência, utilizando pesos pré-treinados na base *ImageNet*.

Como cenário de referência (*baseline*), apenas a camada classificadora foi treinada, mantendo congeladas as camadas convolucionais da MobileNetV2. Também foram avaliadas quatro estratégias: *Fine-Tuning*, balanceamento de classes, variação do particionamento e redução do número de classes. Em seguida, foram comparados o particionamento tradicional (70/15/15) e divisões compostas apenas por treino e teste (70/30, 75/25 e 80/20).

Os experimentos foram implementados em Python com TensorFlow e treinados com o otimizador Adam. Para o tratamento do desbalanceamento, compararam-se as funções de perda Entropia Cruzada e Perca Focal. O treinamento empregou redução automática da taxa de aprendizado e parada precoce para melhorar a convergência e reduzir o *overfitting*. Os cenários foram avaliados de forma incremental, utilizando o modelo *baseline* como referência e comparando seus efeitos sobre a acurácia global e o *recall* por classe.

3. RESULTADOS E DISCUSSÕES

A primeira avaliação considerou as seis classes (*background*, *normal*, *diabetic*, *pressure*, *venous*, *sirurgical*) do conjunto de dados sob a divisão de 70% para treino, 15% para validação e 15% para teste. A Figura 1 apresenta as curvas de treinamento e a matriz de confusão obtidas nesse cenário, enquanto a Tabela 1 reúne as métricas de classificação alcançadas.

O modelo *baseline* alcançou acurácia global de 73%, com convergência estável, conforme observado na Figura 1(b). Porém, como a acurácia global não reflete o desempenho por classe, também foi analisado o *recall*. Conforme a Tabela 1, as classes *background* e *normal* apresentaram os melhores resultados, ambas com *recall* igual a 1,00, seguidas de *venous*, com 0,84. As classes *pressure* (*recall*=0,52) e *surgical* (*recall*=0,48) tiveram os piores desempenhos. A matriz de confusão da Figura 1(a) mostra confusão frequente entre essas

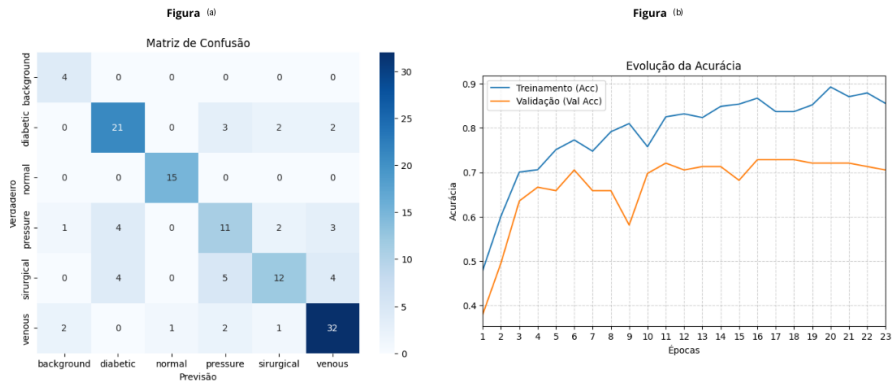

Figure 1: Resultados do Cenário 1: (a) Matriz de confusão; (b) Evolução da acurácia.

Table 1: Relatório de classificação para o Cenário 1.

Classe	**Precisão**	**Recall**	**F1-Score**	**Amostras**
background	0,57	1,00	0,73	4
diabetic	0,72	0,75	0,74	28
normal	0,94	1,00	0,97	15
pressure	0,52	0,52	0,52	21
sirurgical	0,71	0,48	0,57	25
venous	0,78	0,84	0,81	38
Acurácia Global	–	–	**0,73**	**131**
Média Macro	0,71	0,77	0,72	131
Média Ponderada	0,72	0,73	0,72	131

categorias e a classe *diabetic*, indicando sobreposição de características visuais. Assim, embora a MobileNetV2 tenha se mostrado eficiente para classes mais distintas, ainda houve dificuldade na separação de lesões visualmente semelhantes. Esses resultados motivaram a realização de novos experimentos, iniciando-se pela aplicação de *Fine-Tuning*.

3.1 Cenário 2 – 6 Classes com Fine-Tunning

Neste novo cenário, realizou-se o descongelamento das camadas superiores da MobileNetV2 para treinamento por *Fine-Tuning*. A Figura 2 apresenta a matriz de confusão e a evolução da acurácia, enquanto a Tabela 2 reúne as métricas obtidas. A aplicação do *Fine-Tuning* não produziu ganhos consistentes em relação ao cenário *baseline*. A acurácia global caiu para 70%, e a curva de validação apresentou oscilações ao longo do treinamento, o que sugere que o ajuste dos pesos não foi suficiente para melhorar a generalização do modelo.

A análise da Figura 2(a) e da Tabela 2 mostra que as maiores dificuldades permaneceram nas classes *pressure* e *surgical*, com *recall* de 0,43 e 0,40, respectivamente. Essas categorias continuaram sendo confundidas com as classes *diabetic* e *venous*, indicando que a arquitetura enfrenta dificuldades para extrair atributos discriminativos robustos capazes de diferenciar tais morfologias nessa configuração de classes.

Dessa forma, verificou-se que o *Fine-Tuning*, isoladamente, não foi suficiente para superar as limitações observadas no cenário com seis classes. Esses resultados indicaram a necessidade de novos experimentos, com foco na reorganização do conjunto de dados e na redução da granularidade do problema.

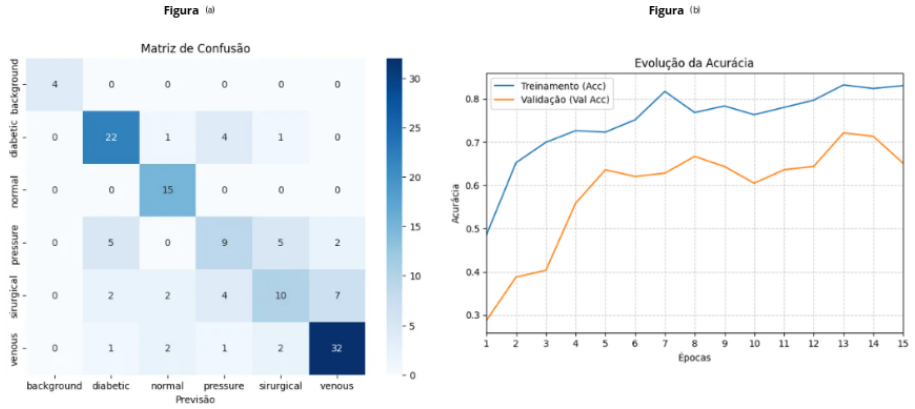

Figure 2: Resultados para o Cenário 2: (a) Matriz de Confusão; (b) Evolução da Acurácia.

Table 2: Relatório de classificação para o Cenário 2.

Classe	**Precisão**	**Recall**	**F1-Score**	**Amostras**
background	1,00	1,00	1,00	4
diabetic	0,73	0,79	0,76	28
normal	0,75	1,00	0,86	15
pressure	0,50	0,43	0,46	21
sirurgical	0,56	0,40	0,47	25
venous	0,78	0,84	0,81	38
Acurácia Global	–	–	**0,70**	**131**
Média Macro	0,72	0,74	0,73	131
Média Ponderada	0,69	0,70	0,69	131

3.2 Cenário Otimizado – 4 Classes

Foi desenvolvido um novo conjunto de experimentos com quatro classes (*background*, *diabetic*, *normal* e *venous*), excluindo temporariamente as categorias *pressure* e *surgical*. Essa decisão se justifica pela elevada variabilidade morfológica interna dessas duas classes, que dificulta a extração de padrões consistentes por arquiteturas leves. Além da redefinição das classes, foram avaliadas diferentes estratégias de particionamento dos dados. Para isso, confrontou-se o arranjo composto por treino, validação e teste na proporção (70/15/15), com arranjos compostos por treino e teste nas proporções (70/30, 75/25 e 80/20).

Table 3: Comparação dos diferentes particionamentos avaliados para o cenário de quatro classes.

Split	**Fine-Tuning**	**Acurácia**	**Precisão Média**	**Recall Média**	**F1-score Médio**
70/15/15	Não	88%	93%	91%	92%
70/15/15	Sim	85%	84%	90%	86%
70/30	Não	93%	93%	93%	93%
70/30	Sim	92%	92%	92%	92%
75/25	Não	89%	89%	89%	89%
75/25	Sim	93%	93%	93%	93%
80/20	Não	87%	87%	87%	87%
80/20	Sim	97%	98%	97%	97%

Conforme evidenciado na Tabela 3, o arranjo 80/20 com *fine-tuning* apresentou o melhor desempenho. Os splits 75/25 e 70/30 também obtiveram resultados competitivos, indicando resposta positiva da MobileNetV2 ao ajuste dos pesos e à reorganização dos dados.

Diante disso, o split 80/20 foi adotado como configuração padrão. Os detalhes da distribuição de erros e da evolução das métricas por época estão apresentados na Figura 3 e na Tabela 4.

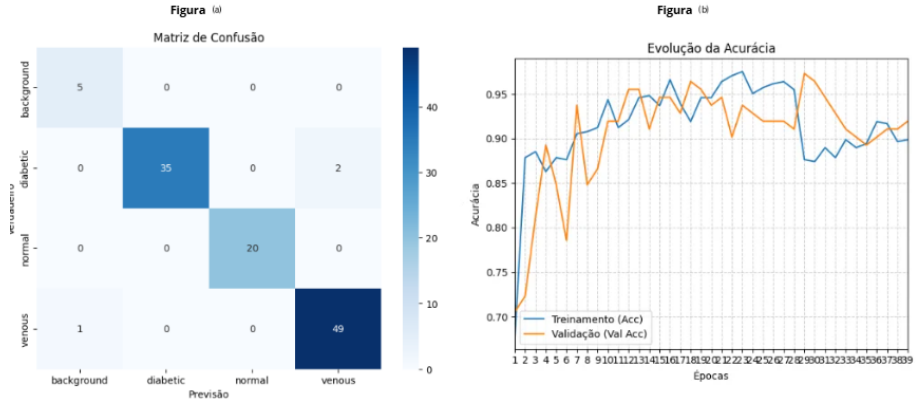

Figure 3: (a) Matriz de Confusão; (b) Evolução da Acurácia de treinamento e validação.

Table 4: Relatório de classificação para o Cenário Otimizado.

Classe	**Precisão**	**Recall**	**F1-Score**	**Amostras**
background	0,83	1,00	0,91	5
diabetic	1,00	0,95	0,97	37
normal	1,00	1,00	1,00	20
venous	0,96	0,98	0,97	50
Acurácia Global	–	–	**0,97**	**112**
Média Macro	0,95	0,98	0,96	112
Média Ponderada	0,98	0,97	0,97	112

Como resultado, o cenário de 4 classes alcançou acurácia global de 97%. As curvas de aprendizado da Figura 3(b) indicam convergência rápida e ausência de *overfitting*. Conforme a Tabela 4, o modelo obteve sensibilidade de 1,00 para *Background* e *Normal*, além de bom desempenho em *Diabetic* (0,95) e *Venous* (0,98). Ainda assim, esses valores devem ser interpretados com cautela, pois o número reduzido de amostras de teste pode elevar pontualmente os índices. Mesmo com essa limitação, a baixa ocorrência de erros indica um foco correto na estruturação metodológica.

4. Conclusões

Este trabalho avaliou o desempenho da arquitetura MobileNetV2 na classificação de imagens de feridas sob diferentes configurações experimentais. Os resultados demonstraram que arquiteturas leves podem alcançar elevado desempenho quando associadas a estratégias adequadas de treinamento, particionamento dos dados e ajuste fino, atingindo acurácia de 97% no cenário de quatro classes.

Os experimentos também evidenciaram que a disponibilidade e a organização dos dados influenciam diretamente o desempenho do modelo. Embora a abordagem tenha apresentado resultados promissores, a discriminação de lesões com alta complexidade morfológica continua desafiadora para o modelo, devido à escassez de dados, o que impede a extração de atributos altamente discriminativos. Esse cenário reforça a necessidade de bases mais representativas, permitindo que a rede aprenda a diferenciar os padrões visuais específicos de cada patologia de forma mais eficaz.

Como trabalhos futuros, propõe-se ampliar e diversificar a base de imagens, investigar novas estratégias de treinamento capazes de melhorar a distinção entre as lesões, bem como realizar validações em ambientes clínicos reais. Além disso, pretende-se desenvolver um aplicativo para dispositivos móveis integrando o modelo treinado, disponibilizando uma ferramenta de apoio à identificação de feridas para profissionais da saúde.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro concedido por meio do projeto OET-00243-26. Agradecem também ao Instituto Federal do Norte de Minas Gerais (IFNMG) pelo apoio institucional.

REFERENCES

Al-Shabi, M., *et al.* (2025), "Edge AI: A Comprehensive Survey of Technologies, Applications, and Challenges", *Journal of Edge Computing and Artificial Intelligence*, 15(2), 112–128.

Anisuzzaman, D.M., Patel, Y., Rostami, B., Niezgoda, J., Gopalakrishnan, S. and Yu, Z. (2022), "Multi-modal wound classification using wound image and location by deep neural network", *Scientific Reports*, 12, 20057.

Dias, K. M., Rodrigues, J. P. S., Rohmer, E. e Mota, F. A. O. (2026), "Avaliação de Inferência em Edge AI sob Restrições Embarcadas em um Sistema Robótico Simulado Baseado na Internet das Coisas Robóticas", In *Anais de Eventos do IFNMG*, Januária.

Guest, J.F., Fuller, G. and Vowden, P. (2020), "Cohort study evaluating the burden of wounds to the UK's National Health Service in 2017/2018", *BMJ Open*, 10(12), e045253.

Lindholm, C. and Searle, R. (2016), "Wound management for the 21st century: combining effectiveness and efficiency in modern clinical practice", *International Wound Journal*, 13, 5–15.

Patel, Y., Shah, T., Dhar, M.K., Zhang, T., Niezgoda, J., Gopalakrishnan, S. and Yu, Z. (2024), "Integrated image and location analysis for wound classification: a deep learning approach", *Scientific Reports*, 14, 7043.

Prodanov, C.C. e Freitas, E.C. (2013), *Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico*, 2ª ed., Editora Feevale, Novo Hamburgo.

Salomé, G.M. and Gaudencio, L. (2025), "Jogo educativo "Feridas Crônicas": Orientação para profissionais na avaliação, prevenção e tratamento de feridas", *Revista Brasileira de Cirurgia Plástica*, 40. DOI: 10.1055/s-0045-1807722.

Sandler, M., Howard, A., Zhmoginov, A., Chen, L.C. and Zhu, M. (2018), "MobileNetV2: Inverted Residuals and Linear Bottlenecks", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520.

Sen, C.K. (2019), "Human Wounds and Its Burden: An Updated Compendium of Estimates", *Advances in Wound Care*, 8(2), 39–48.

Souza, V. A. (2025), "Classificação automática de imagens utilizando redes neurais convolucionais mobilenet: um estudo de caso com reconhecimento de felinos", In *Anais do IV Congresso Brasileiro de Ciência e Saberes Multidisciplinares*, Volta Redonda. UniFOA.

Whitelaw, S., Mamas, M.A., Topol, E. and Spall, H.G.C. (2020), "Applications of digital health technologies in the COVID-19 pandemic response", *The Lancet Digital Health*, 2(8), e435–e440.