PICKCELLS

Uso em pesquisa/triagem (RUO) — apoio à decisão, não diagnóstico

HematoCase: detecção e classificação assistidas por IA em mielograma e sangue periférico — nota técnica de validação

PickCells · Plataforma PickCells-H · Fortaleza/Recife, Brasil · Setembro de 2026 · Dados de validação gerados em 17/09/2026.

Resumo. O HematoCase é uma bancada digital que oferece **apoio à decisão** em hematologia: modelos de IA localizam e classificam células em campos de **mielograma** e **sangue periférico**, e o profissional habilitado revisa, valida e conclui. Esta nota descreve a metodologia e os resultados de validação frente à curadoria humana. Os detectores de mielograma foram avaliados em 50 caixas anotadas (6 campos, IoU ≥ 0,5) e os classificadores em 180 imagens de 9 classes (20 por classe). O melhor classificador (**YOLOv8-cls**) obteve acurácia de **98,3%** (IC Wilson 95%: 95,2–99,4); o melhor detector (variante **letterbox**) alcançou revocação de **0,86** (73,8–93,0), precisão **1,00** e acerto de classe de **88,4%**. O tamanho reduzido do conjunto anotado (n = 50) produz intervalos amplos; a diferença entre *letterbox* e *squash* não foi estatisticamente significativa (McNemar, p ≈ 1), embora teoricamente motivada. Todos os resultados são apoio à decisão (RUO), nunca laudo.

1. Introdução

A avaliação morfológica do esfregaço de sangue periférico e do aspirado de medula óssea (mielograma) é etapa central do diagnóstico hematológico, apoiada na contagem diferencial e na identificação de linhagens celulares[2]. Limiares morfológicos têm peso diagnóstico direto — por exemplo, o percentual de blastos que define leucemia aguda na classificação da OMS[1] — e a própria contagem diferencial carrega imprecisão amostral bem descrita, que justifica reportar intervalos de confiança em vez de pontos isolados[3].

Métodos de aprendizado profundo tornaram-se referência em visão computacional [8] e, em hematologia, redes convolucionais já demonstraram reconhecimento de blastos e ti-

pagem de leucócitos em nível comparável ao humano[14,15]. O HematoCase aplica **detecção de objetos**[9] para localizar células e **classificação** para tipá-las, mantendo o profissional como autoridade final. Esta nota reporta a validação desses modelos com quantificação explícita de incerteza.

2. Materiais e métodos

2.1 Modelos

Foram avaliados três **detectores de mielograma** — dois de 2024 (arquitetura YOLOv8n), um deles com pré-processamento *letterbox* (preservando a proporção do campo, igual ao treino) e outro com *squash*, e um de 2023 (YOLOv5s)[10] — e cinco **classificadores** de 9 classes celulares: VGG16 e VGG19[11], ResNet50[12], DenseNet121[13] e YOLOv8-cls[10].

2.2 Conjuntos de avaliação e verdade de referência

A **curadoria humana** constitui a verdade de referência. A classificação foi avaliada em 180 imagens (9 classes — basófilo, eosinófilo, linfócito, monócito, neutrófilo, granulócito imaturo, plaqueta, eritroblasto e fundo — 20 por classe, amostragem com semente fixa 42, pré-processamento raw255). A detecção foi avaliada em 6 campos de mielograma totalizando **50 caixas anotadas**.

2.3 Métricas

Uma detecção é considerada verdadeiro-positivo quando a sobreposição com a caixa de referência atinge **IoU ≥ 0,5**[16,17]. Reporta-se precisão, revocação e **acerto de classe sobre os verdadeiro-positivos**; para os classificadores, acurácia global e matriz de confusão.

3. Análise estatística

Cada proporção é acompanhada de **intervalo de confiança de Wilson (score) 95%**[4], preferido à aproximação normal e à Clopper-Pearson[5] por bom comportamento em amostras pequenas e proporções próximas de 0 ou 1. Como verificação independente, os mesmos intervalos foram estimados por **bootstrap percentil 95%** (10.000 rea-

mostragens, semente 42)[6]. A comparação entre as variantes *letterbox* e *squash*, avaliadas sobre **as mesmas** 50 caixas de referência, usou o **teste pareado de McNemar**[7]. Reprodutibilidade: sementes fixas, pré-processamento congelado e serviço de inferência determinístico (verificado por teste-ouro de regressão).

---

4. Resultados

4.1 Classificação (9 classes)

As acurácias variaram de 36,1% (VGG16) a 98,3% (YOLOv8-cls). A Figura 1 mostra os pontos e os IC de Wilson 95%; as matrizes de confusão dos dois melhores modelos aparecem na Figura 2.

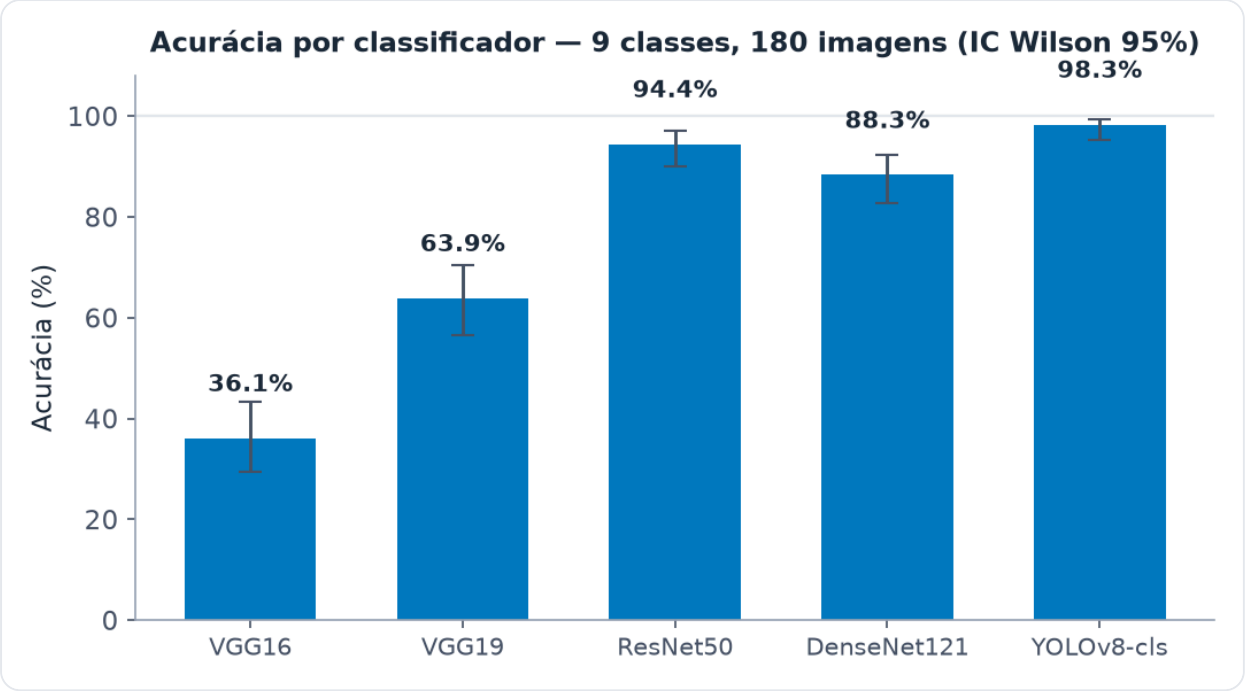


Figura 1. Acurácia por classificador (9 classes, 180 imagens) com intervalos de confiança de Wilson 95%.

Classificador	Acurácia	Acertos	IC Wilson 95%	IC bootstrap 95%
VGG16	36,1%	65/180	29,5–43,4	28,9–43,3
VGG19	63,9%	115/180	56,6–70,5	56,7–70,6
DenseNet121	88,3%	159/180	82,8–92,2	83,3–92,8
ResNet50	94,4%	170/180	90,1–97,0	91,1–97,8
YOLOv8-cls	**98,3%**	177/180	95,2–99,4	96,1–100

Tabela 1. Acurácia por classificador com IC de Wilson e bootstrap 95%.

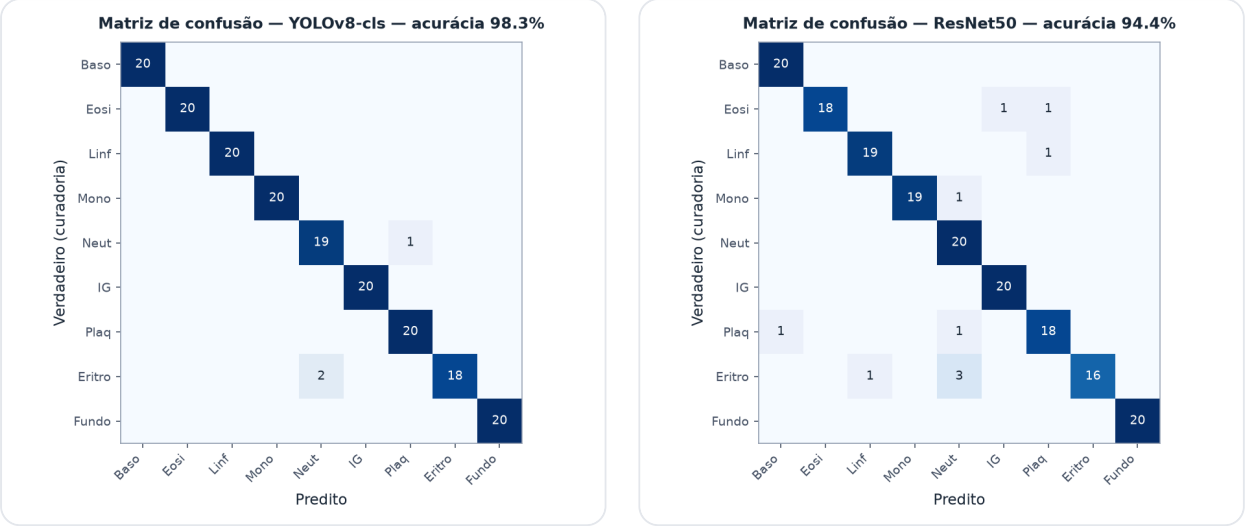


Figura 2. Matrizes de confusão dos dois melhores classificadores (YOLOv8-cls, à esquerda; ResNet50, à direita). Linhas = classe verdadeira (curadoria), colunas = predição.

4.2 Detecção (mielograma)

A variante *letterbox* apresentou o melhor desempenho global (revocação 0,86; acerto de classe 88,4%), com precisão de 1,00 e nenhum falso-positivo em 43 detecções. A Figura 3 resume as métricas com IC 95%.

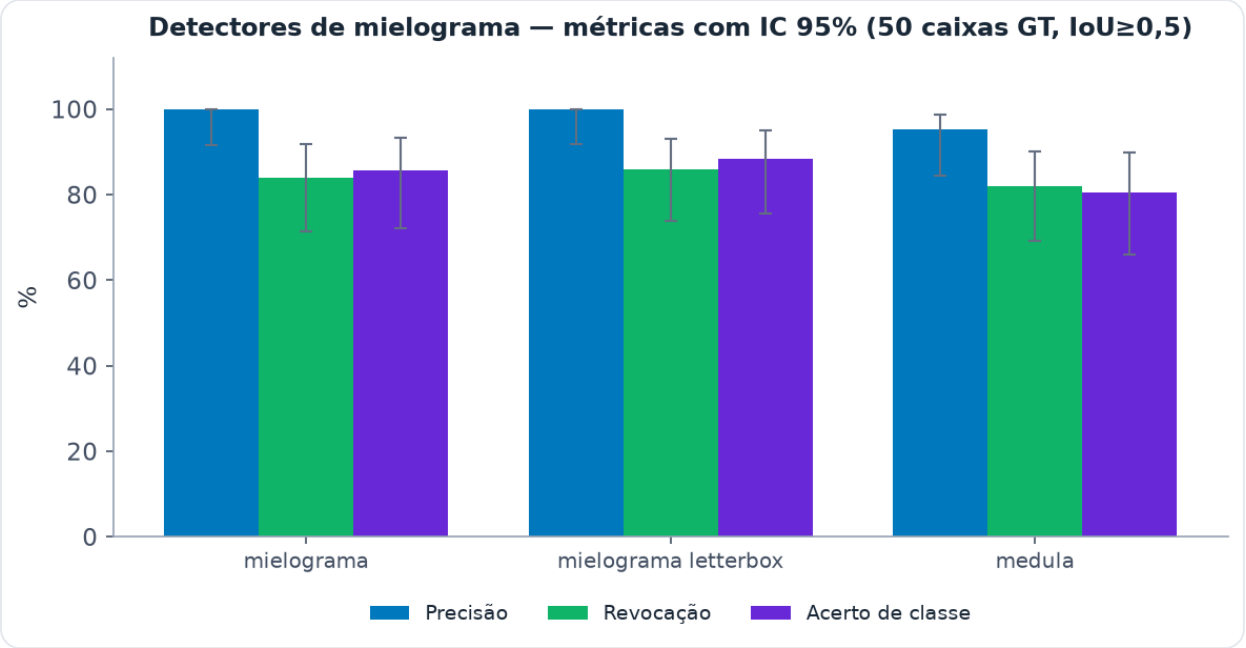


Figura 3. Detectores de mielograma — precisão, revocação e acerto de classe com IC 95% (50 caixas de referência, IoU ≥ 0,5).

Detector	Arquit.	Precisão (IC95)	Revocação (IC95)	Acerto classe (IC95)	TP/FP/FN
Mielograma 2024 · letterbox	YOLOv8n	1,00 (91,8–100)	0,86 (73,8–93,0)	88,4% (75,5–94,9)	43/0/7
Mielograma 2024 · squash	YOLOv8n	1,00 (91,6–100)	0,84 (71,5–91,7)	85,7% (72,2–93,3)	42/0/8
Medula 2023	YOLOv5s	0,95 (84,5–98,7)	0,82 (69,2–90,2)	80,5% (66,0–89,8)	41/2/9

Tabela 2. Detectores de mielograma — métricas com IC 95% (50 caixas GT, IoU ≥ 0,5).

4.3 Letterbox vs. squash (teste pareado)

Sobre as mesmas 50 caixas, o *letterbox* recuperou 43 e o *squash* 42 (Δrevocação = +2,0 pp; Δacerto de classe = +2,7 pp). O teste de McNemar não rejeitou a hipótese nula ($p \approx 1$), ou seja, a diferença observada é compatível com ruído amostral neste tamanho de amostra.

Interpretação honesta. O ganho do *letterbox* é **teoricamente motivado** — a geometria de inferência passa a coincidir com a do treino, evitando distorção de aspecto — e, nos dados, **nunca foi pior**; porém, com n = 50 caixas, **não é estatisticamente significativo**. Adotou-se o *letterbox* por consistência metodológica e ausência de prejuízo, não por superioridade comprovada.

4.4 Exemplos de detecção

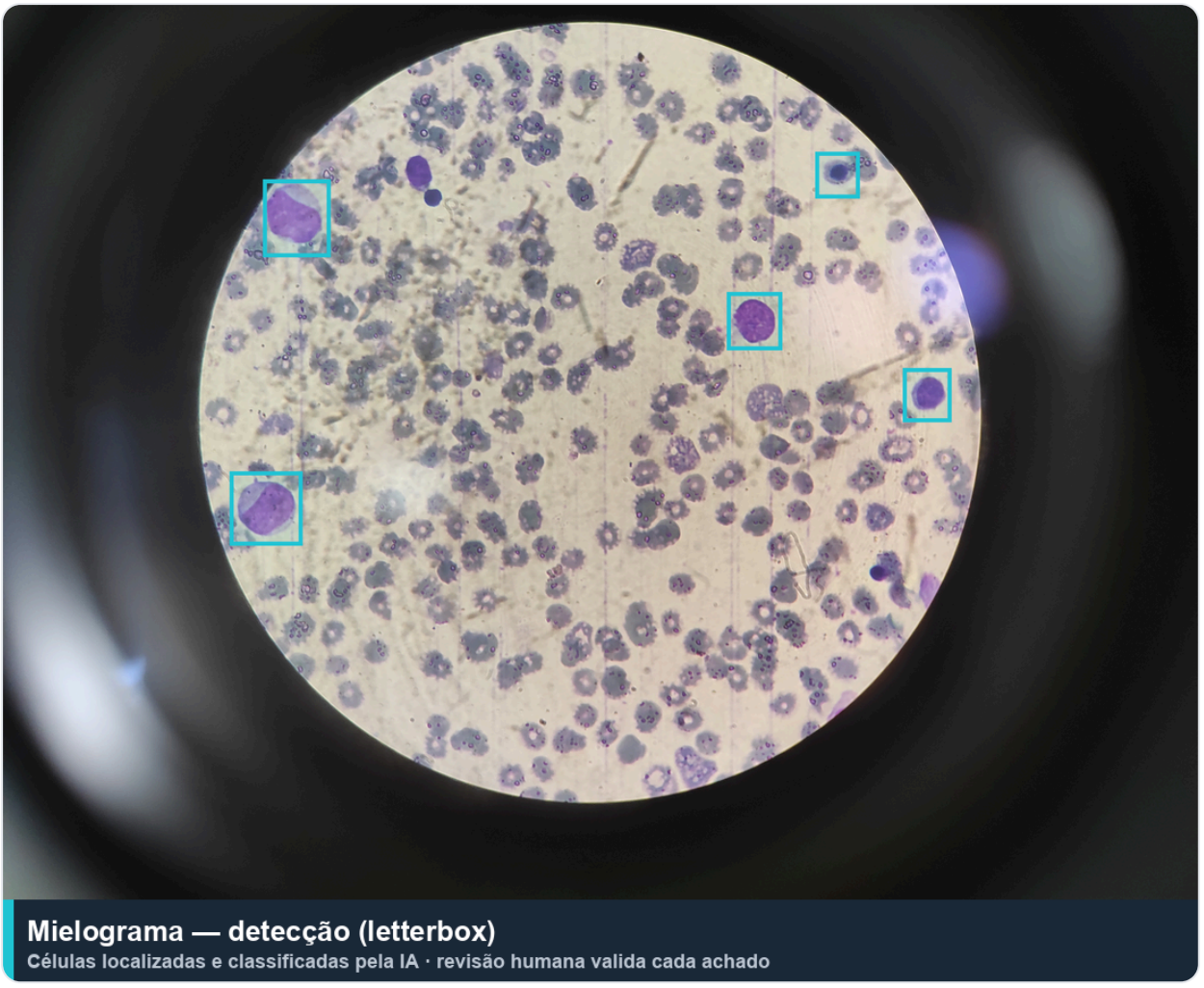


Figura 4. Detecção real em campo de mielograma (modelo letterbox). Cada caixa é revisada pelo profissional.

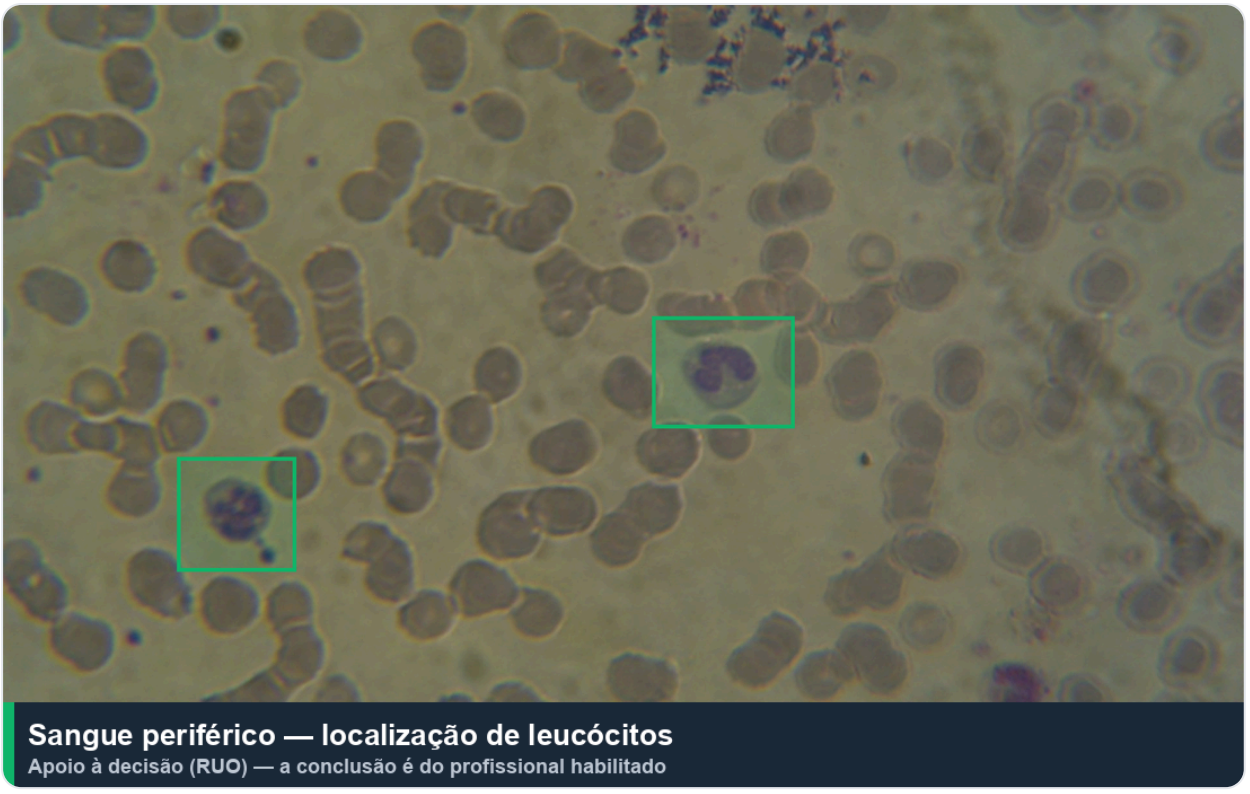


Figura 5. Localização de leucócitos em sangue periférico. Apoio à decisão (RUO).

5. Discussão

No conjunto avaliado, a classificação atinge desempenho alto e estável para os modelos mais recentes (ResNet50 e YOLOv8-cls), enquanto arquiteturas VGG servem como linha de base inferior — padrão consistente com a literatura de *backbones*[12,13]. Na detecção, a precisão elevada (ausência de falsos-positivos nos modelos de 2024) é desejável em triagem, pois limita alarmes espúrios; a revocação de ~0,85 indica que uma fração das células ainda escapa e reforça a necessidade de revisão humana. A tipagem de leucócitos do sangue periférico é coberta pelo classificador de 9 classes; a verdade de referência de **detecção em campo** concentra-se, por ora, no mielograma.

6. Limitações e reprodutibilidade

O principal limite é o **tamanho do conjunto anotado de detecção** (50 caixas em 6 campos), que amplia os intervalos de confiança e reduz o poder do teste pareado; comparações entre variantes devem ser lidas como exploratórias. As imagens provêm de aquisição limitada e não substituem uma avaliação clínica prospectiva. A plataforma é **RUO**: as saídas são apoio à decisão e não constituem laudo. Para reprodutibilidade, fi-

xaram-se sementes (42), pré-processamento e um serviço de inferência determinístico, verificado por teste-ouro de regressão a cada implantação.

7. Conclusão

O HematoCase entrega detecção e classificação assistidas com **incerteza quantificada** (IC de Wilson e bootstrap) e comparações honestas (McNemar). Os melhores modelos alcançam desempenho alto em classificação e precisão perfeita em detecção no conjunto avaliado, sempre sob **curadoria humana obrigatória**. A ampliação do conjunto anotado é o passo natural para estreitar os intervalos e sustentar afirmações comparativas com significância estatística.

Referências

1. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022;36(7):1703–1719.
2. Bain BJ. *Blood Cells: A Practical Guide*. 5th ed. Wiley-Blackwell; 2015.
3. Rümke CL. The imprecision of the ratio of two percentages observed in differential white blood cell counts: a warning. *Blood Cells*. 1985;11(1):137–140.
4. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*. 1927;22(158):209–212.
5. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26(4):404–413.
6. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC; 1993.
7. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153–157.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
9. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In: *Proc. IEEE CVPR*; 2016. p. 779–788.
10. Jocher G, et al. Ultralytics YOLO (v5/v8) [software]. https://github.com/ultralytics/ultralytics
11. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *Proc. ICLR*; 2015.
12. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proc. IEEE CVPR*; 2016. p. 770–778.
13. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *Proc. IEEE CVPR*; 2017. p. 4700–4708.

14. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat Mach Intell*. 2019;1(11):538–544.
15. Acevedo A, Merino A, Alférez S, Molina Á, Boldú L, Rodellar J. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data Brief*. 2020;30:105474.
16. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge. *Int J Comput Vis*. 2010;88(2):303–338.
17. Padilla R, Netto SL, da Silva EAB. A survey on performance metrics for object-detection algorithms. In: *Proc. IWSSIP*; 2020.

PickCells · HematoCase · Nota técnica de validação (RUO). Projeto apoiado pelo Edital FUNDECI 01/2021 — Banco do Nordeste (BNB) · Termo de Outorga FUNDECI 2023.0003.