

# Detecção de Micronúcleos em Linfócitos: Um Novo Dataset e Estudo de Caso com YOLOv11

Camile A. Barbosa<sup>1</sup>, Gael F. Lima<sup>2</sup>, Suy F. Hwang<sup>2</sup>,  
Fabiana F. Lima<sup>2</sup>, Filipe R. Cordeiro<sup>1</sup>

<sup>1</sup> Visual Computing Lab, Departamento de Computação,  
Universidade Federal Rural de Pernambuco (UFRPE), Brasil.  
camile.alheiro@ufrpe.br, filipe.rolim@ufrpe.br

<sup>2</sup> Centro Regional de Ciências Nucleares do Nordeste (CRCN-NE / CNEN), Brasil.  
freiresgael@gmail.com, suy.hwangcnen.gov.br,  
fabiana.fariascnen.gov.br

**Abstract.** *The detection of micronuclei (MN) in lymphocytes is essential in biomedicine for assessing genetic damage and chromosomal instability, playing a crucial role in toxicological studies and cancer diagnosis. However, manual identification is time-consuming and prone to errors. In this work, we propose an automated approach using the YOLOv11 neural network for detecting micronuclei in binucleated cells. To achieve this, we built an image dataset collected from the Regional Center for Nuclear Sciences of the Northeast (CRCN), comprising 889 images annotated with the locations of binucleated cells and micronuclei. The results of the detection analysis show that the model achieved a precision of 90.8% and a recall of 92.8%, demonstrating reliability for clinical applications. Additionally, the developed dataset contributes to future research in the field, providing a standardized benchmark for evaluating computer vision models applied to cytogenetics. The dataset is available at <https://doi.org/10.5281/zenodo.14947933>.*

**Resumo.** *A detecção de micronúcleos (MN) em linfócitos é essencial na biomedicina para avaliar danos genéticos e instabilidade cromossômica, sendo importante para estudos toxicológicos e diagnóstico de câncer. No entanto, a identificação manual é demorada e sujeita a erros. Neste trabalho, propomos uma abordagem automatizada utilizando a rede neural YOLOv11 para detecção de micronúcleos em células binucleadas. Para isso, construímos uma base de imagens coletadas do Centro Regional de Ciências Nucleares do Nordeste (CRCN), composta por 889 imagens, com anotações de localização de células binucleadas e micronúcleos. Resultados da análise de detecção mostram que o modelo utilizado alcançou precisão de 90,8% e revocação de 92,8%, demonstrando confiabilidade para aplicações clínicas. Além disso, o conjunto de dados desenvolvido contribui para futuras pesquisas na área, fornecendo uma base padronizada para avaliação de modelos de visão computacional aplicados à citogenética. A base de dados desenvolvida está disponível em <https://doi.org/10.5281/zenodo.14947933>.*

## 1. Introdução

Micronúcleos (MN) são estruturas celulares formadas por fragmentos cromossômicos ou cromossomos inteiros que não se reincorporam ao núcleo principal após a divisão celular,

permanecendo no citoplasma como pequenos núcleos acessórios [Krupina et al. 2021]. A sua detecção é amplamente utilizada em pesquisas de toxicologia genômica, além de contribuir significativamente para estudos sobre câncer e outras doenças relacionadas à instabilidade genética [Krupina et al. 2021].

Tradicionalmente, a detecção e contagem de micronúcleos é realizada manualmente por citogeneticista experientes, o que torna o processo demorado, trabalhoso e propenso a erros de interpretação. A subjetividade na avaliação, aliada à variabilidade na qualidade das imagens microscópicas, pode gerar inconsistências entre observadores, afetando a precisão dos diagnósticos [Rohr et al. 2020]. Nesse contexto, modelos de *Deep Learning*, vêm demonstrando grande potencial para automatizar tarefas de detecção de objetos, incluindo a identificação de micronúcleos em imagens citogenéticas [Yoda et al. 2024, Panchbhai et al. 2023]. Apesar dos trabalhos existentes, a maioria dos trabalhos utiliza bases privadas, o que torna difícil a reproduzibilidade e análise dos modelos. Neste trabalho, desenvolvemos um novo conjunto de dados composto por imagens capturadas no Centro Regional de Ciências Nucleares do Nordeste (CRCN), que foram cuidadosamente anotadas para garantir a precisão na identificação das classes: células binucleadas sem micronúcleo, binucleadas com micronúcleo e micronúcleos. Para análise de detecção, utilizamos o modelo YOLOv11 [Jocher et al. 2023] sob diferentes configurações.

## 2. Trabalhos Relacionados

A detecção automatizada de micronúcleos (MN) em imagens celulares tem sido foco de diversas pesquisas recentes. Dentre essas abordagens, os modelos da família YOLO (You Only Look Once) [Jiang et al. 2022] têm recebido atenção especial pela alta precisão e velocidade na detecção de objetos em tempo real. Wei et al. [Wei et al. 2023] propuseram uma versão aprimorada do YOLOv5, denominada YOLOv5-CEB, que incorpora mecanismos de atenção para melhorar a precisão na detecção de micronúcleos. Os resultados experimentais desse trabalho mostram um aumento de 2% na precisão média, 2% na acurácia e 1,3% na taxa de revocação em comparação com o algoritmo YOLOv5 original. Ibarra-Arellano et al. [Ibarra-Arellano et al. 2024] apresentaram a ferramenta micronuclAI, uma aplicação com interface gráfica voltada para a quantificação automatizada de micronúcleos de diferentes tamanhos, morfologias e localizações. Utilizando uma rede neural convolucional treinada para analisar recortes de células individuais em conjuntos de dados provenientes de diversas linhagens celulares humanas e murinas, o classificador alcançou desempenho próximo ao nível humano. Outra contribuição significativa foi proposta por Wei et al. [Wei et al. 2024], que desenvolveram o STD-YOLOv7, um detector de pequenos alvos para micronúcleos baseado na arquitetura YOLOv7. Esse modelo incorpora o mecanismo de atenção coordenada (CA) e o Res-ACmix, melhorando a precisão na detecção de micronúcleos ao permitir que a rede se concentre em regiões críticas da imagem. Apesar dos avanços na área, os trabalhos descritos se baseiam em bases de dados privadas, limitando a reproduzibilidade dos experimentos e a comparação direta entre abordagens.

Nosso trabalho difere dessas abordagens ao empregar a versão mais recente do YOLO (YOLOv11). Além disso, disponibilizamos um novo conjunto de dados, desenvolvido a partir de imagens coletadas no CRCN, contendo anotações de localização de células binucleadas e micronúcleos. Ao tornar o dataset público, buscamos padronizar a

avaliação de modelos e fomentar a reprodutibilidade de experimentos na área.

### 3. Metodologia

#### 3.1. Aquisição e Anotação da Base de Dados

A construção do conjunto de dados utilizado neste estudo foi realizada com imagens de linfócitos binucleados obtidas diretamente das amostras biológicas processadas no CRCN. A captura das imagens foi realizada manualmente, com acompanhamento de especialista, digitalizadas por meio de um microscópio óptico Leica DM 500, com resolução de  $2048 \times 1536$ .

Para assegurar a qualidade e validade científica das amostras, as coletas foram periodicamente revisadas e validadas por citogeneticistas experientes do CRCN. A validação foi realizada com base nas diretrizes do projeto HUMN (Human Micronucleus Project) [Fenech et al. 2002], um protocolo amplamente reconhecido na comunidade científica para padronização da análise de micronúcleos. Seguindo esses critérios, durante a fase de validação, algumas imagens foram descartadas por apresentarem irregularidades, como sobreposição de núcleos, danos ao citoplasma ou baixa qualidade de contraste e foco. A Figura 1 no detalhe (a) apresenta um exemplo de uma célula válida e no detalhe (b) outra inválida, evidenciando os critérios utilizados na seleção das amostras.



**Figura 1. (a) Imagem válida usada na base de dados construída (a) e (b) imagem descartada como célula inválida.**

Ao todo, foram feitas 889 capturas, entre elas, 543 imagens de células binucleadas sem micronúcleo (BN) e 346 imagens de células binucleadas com micronúcleo (BNMN). As imagens foram originalmente capturadas em formato TIFF e convertidas para PNG. A Tabela 1 apresenta os dados da base.

**Tabela 1. Distribuição da base de dados desenvolvida. A segunda coluna mostra o número de imagens contendo cada classe e a terceira coluna mostra o número de objetos totais da classe**

Classe	Qtd. de imagens	Qtd. de objetos
Binucleada sem micronúcleo (BN)	543	600
Binucleada com micronúcleo (BNMN)	346	347
Micronúcleo (MN)	346	472
Total	889	1419

### 3.2. Treinamento do Modelo de Detecção

O modelo YOLOv11 foi treinado por 200 épocas, utilizando os parâmetros padrões da arquitetura, que incluem taxa de aprendizagem 0.01 e otimização com gradiente estocástico (SGD). Foi utilizado também o modelo YOLOv11 pré-treinado na base COCO [Jain et al. 2022] e refinado para a nossa base. Para os experimentos, usamos tamanho de *batch* 8 e redimensionamos as imagens para tamanho  $1080 \times 1080$  pixels. Os experimentos foram conduzidos utilizando o Kaggle Notebook[Kaggle 2025], um ambiente de desenvolvimento baseado em nuvem que oferece recursos computacionais gratuitos, incluindo GPUs e TPUs.

A separação da base em conjunto de treino e teste foi feita separando 2/3 das amostras para treinamento e 1/3 para teste, resultando em uma base de dados final de 621 imagens para treino (379 BNs e 242 BNMNs) e 268 imagens de teste (164 BNs e 104 BNMNs).

## 4. Resultados

A análise de detecção realizada utilizando diferentes configurações do YOLOv11: YOLOv11\_s (*small* - 9.4M parâmetros), YOLOv11\_m (*medium* - 20.1 M parâmetros) e YOLOv11\_l (*large* - 25.3 M par), aplicadas ao conjunto de dados desenvolvido. Essas versões diferem em quantidade de parâmetros do modelo e tamanho do modelo. Os experimentos utilizaram as métricas de precisão, revocação e  $mAP_{50}$ , que são métricas bastante utilizadas para detecção de objetos [Padilla et al. 2020]. A Tabela 2 mostra os resultados obtidos para cada modelo e cada classe. Os resultados apresentados mostram que o modelo YOLOv11\_m apresentou os melhores resultados para a métrica de  $mAP_{50}$ , que indica a qualidade da localização da detecção comparada com a localização esperada. Considerando as métricas de precisão e revocação, o YOLOv11\_m também obteve um melhor balanceamento entre as duas métricas, comparado com as outras abordagens. É possível observar também que a classe MN é a classe mais difícil de ser detectada entre os modelos. Isso se deve ao tamanho dos micronúcleos nas imagens, e mostra que ainda é um problema que pode ser investigado em trabalhos futuros.

Também foi feita uma análise qualitativa sobre a detecção, considerando o melhor modelo obtido, o YOLOv11\_m. A Figura 2 mostra os resultados de detecção, onde o retângulo em verde corresponde à detecção do modelo feita corretamente (verdadeiro positivo), com sua respectiva classe e o valor ao lado corresponde à métrica de Interseção Sobre União (IoU) [Padilla et al. 2020], que diz o quanto próxima a detecção está da marcação esperada. Na Figura 2, temos a detecção na imagem original, na coluna esquerda, e na coluna direita temos um redimensionamento na região da detecção para melhor visualização. Na Figura 2 (b) o retângulo vermelho indica um falso negativo, onde há um micronúcleo, mas não foi identificado. Podemos observar a partir dessas imagens que, apesar dos ruídos das imagens e de outras células na imagem, isso não atrapalhou o modelo de identificar as células corretamente. As outras células são automaticamente descartadas pelo modelo por não serem binucleadas ou por possuírem estruturas atípicas, tais como núcleos de células com tamanhos diferentes, as quais também foram descartadas pela anotação do citogeneticista.

**Tabela 2. Resultados de detecção dos modelos YOLOv11\_s, YOLOv11\_m e YOLOv11\_l, para as métricas de precisão, revocação e  $mAP_{50}$  (%). Os melhores valores de cada métrica estão representados em negrito.**

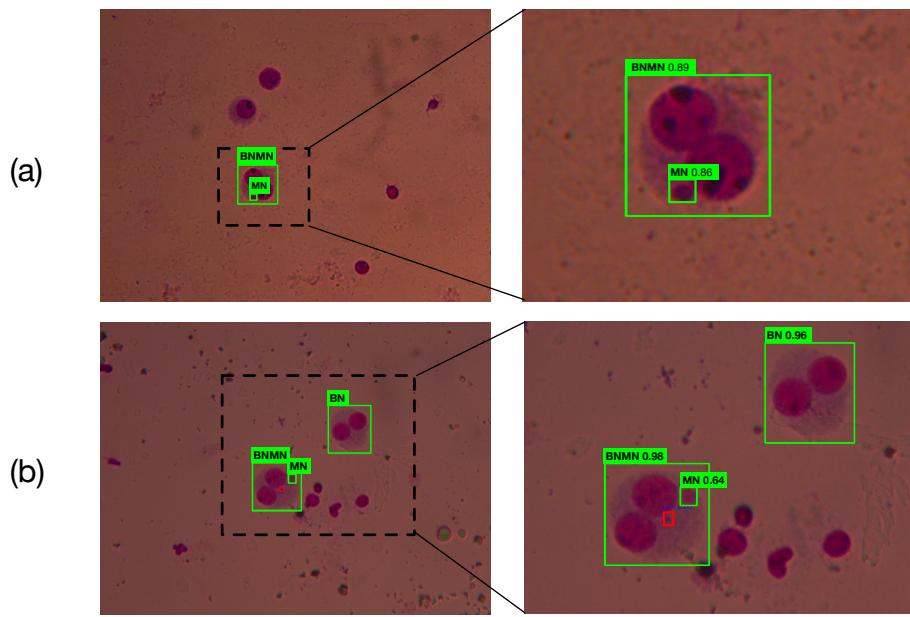
Modelo	Classe	Precisão	Revocação	$mAP_{50}$ (%)
YOLOv11_s	BN	87,8	<b>99,4</b>	96,7
	BNMN	93,0	93,3	98,1
	MN	<b>95,3</b>	79,6	92,7
	Todas	<b>92,0</b>	90,8	95,8
YOLOv11_m	BN	87,9	98,9	<b>98,3</b>
	BNMN	95,2	94,1	<b>98,4</b>
	MN	89,3	<b>84,9</b>	<b>94,2</b>
	Todas	90,8	<b>92,7</b>	<b>96,9</b>
YOLOv11_l	BN	<b>91,1</b>	98,7	98,2
	BNMN	<b>96,1</b>	<b>94,8</b>	<b>98,4</b>
	MN	88,1	84,4	91,4
	Todas	91,8	92,6	96,0

## 5. Conclusão

Este trabalho apresentou uma abordagem eficiente para a detecção de micronúcleos em linfócitos utilizando o modelo YOLOv11, alcançando alta precisão e revocação em todas as classes analisadas. Este trabalho também contribui com uma nova base de dados com anotações de células binucleadas e micronúcleos, padronizando a avaliação de modelos e promovendo a reprodutibilidade de experimentos futuros. Além disso, os resultados obtidos demonstram que o sistema proposto é uma solução confiável para aplicações clínicas e laboratoriais, reduzindo a subjetividade e o tempo necessário para a análise das amostras.

## Referências

- Fenech, M., Chang, W. P., Kirsch-Volders, M., et al. (2002). Humn project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 534(1-2):65–75. Accessed: 2025-02-09.
- Ibarra-Arellano, M. A., Caprio, L. A., Hada, A., Stotzem, N., Cai, L., Shah, S., Melms, J. C., Wünneman, F., Izar, B., and Schapiro, D. (2024). micronuclai: Automated quantification of micronuclei for assessment of chromosomal instability. *bioRxiv [Preprint]*, 2024(05.24.595722). Accessed: 2025-02-09.
- Jain, S., Dash, S., Deorari, R., et al. (2022). Object detection using coco dataset. In *2022 International Conference on Cyber Resilience (ICCR)*, pages 1–4. IEEE.
- Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia computer science*, 199:1066–1073.
- Jocher, G., Qiu, J., and Chaurasia, A. (2023). Ultralytics YOLO.
- Kaggle (2025). Kaggle notebooks. url:<https://www.kaggle.com/code>. Acessado em: 20/02/2025.



**Figura 2. Detecção de células binucleadas e micronúcleos em imagens, utilizando o YOLOv11\_m. Retângulos em verde indicam a detecção do modelo feita corretamente, indicando a classe e métrica IoU associada. Retângulos em vermelho indicam falso negativo.**

- Krupina, K., Goginashvili, A., and Cleveland, D. W. (2021). Causes and consequences of micronuclei. *Current Opinion in Cell Biology*, 70:91–99. Accessed: 2025-02-09.
- Padilla, R., Netto, S. L., and Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*, pages 237–242. IEEE.
- Panchbhai, A., Ishanzadeh, M. C. S., Sidali, A., Solaiman, N., Pankanti, S., Kanagaraj, R., Murphy, J. J., and Surendranath, K. (2023). A deep learning workflow for quantification of micronuclei in dna damage studies in cultured cancer cell lines: a proof of principle investigation. *Computer Methods and Programs in Biomedicine*, 232:107447.
- Rohr, P., da Silva, G. F., Vicentini, V. E. P., de Almeida, I. V., Dos Santos, R. A., Takahashi, C. S., Goulart, M. O., da Silva, G. N., de Oliveira, L. B., Grisolia, C. K., et al. (2020). Buccal micronucleus cytome assay: Inter-laboratory scoring exercise and micronucleus and nuclear abnormalities frequencies in different populations from brazil. *Toxicology Letters*, 333:242–250.
- Wei, W., Leng, Y., Cao, L., et al. (2024). Std-yolov7: A small target detector for micro-nucleus based on yolov7. Accessed: 2025-02-09.
- Wei, W., Li, J., Wu, X., and Zhang, H. (2023). High-through cell micronucleus image detection method combining multi-attention mechanism and yolov5. *Biomedical Signal Processing and Control*, 87:105496. Accessed: 2025-02-09.
- Yoda, H., Abe, K., Takeo, H., Takamura-Enya, T., and Koike-Takeshita, A. (2024). Application of image-recognition techniques to automated micronucleus detection in the in vitro micronucleus assay. *Genes and Environment*, 46(1):11.