

# Classificação de Glóbulos Brancos usando Descritores de Textura e XGBoost

Domingos Alves Dias Júnior<sup>1</sup>, Geraldo Braz Junior<sup>1</sup>

<sup>1</sup>Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)  
CEP 65080-805 – São Luís – MA – Brasil

{domingos.adj, geraldo}@nca.ufma.br

## 1. Introdução

Os glóbulos brancos são um tipo de célula sanguínea que é produzido na medula óssea e encontrado no sangue e no tecido linfático. Os glóbulos brancos fazem parte do sistema imunológico do corpo e ajuda o corpo a combater infecções e outras doenças. Os tipos de glóbulos brancos são neutrófilos, eosinófilos, basófilos, monócitos e linfócitos [NCI 2020, ACS 2018]. O diagnóstico de doenças sanguíneas geralmente envolve a identificação e caracterização de amostras de sangue de pacientes. Entre eles, a contagem e classificação de glóbulos brancos, pois ajudam os patologistas a diagnosticar doenças como leucemia e outras.

O objetivo deste trabalho é apresentar um novo método automático para classificação de glóbulos brancos em imagens microscópicas, utilizando uma abordagem de análise de textura e o classificador XGBoost.

## 2. Materiais e Métodos

O método proposto consiste em quatro etapas principais: detecção de glóbulos brancos; extração de características; classificação e validação, como mostrado na Figura 1. Para a realização dos experimentos, foram obtidas imagens microscópicas do banco de dados BCCD do desafio Kaggle, composta por 410 imagens das quatro classes consideradas: 88 eosinófilos, 33 linfócitos, 21 monócitos e 207 neutrófilos.

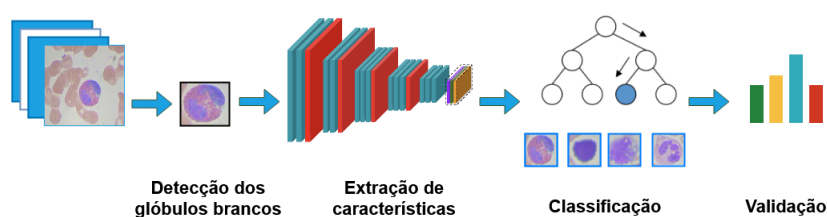


Figura 1. Ilustração do método proposto.

Na etapa de detecção dos glóbulos brancos, a técnica de limiarização foi utilizada para localizar o núcleo celular, uma vez que o núcleo possui uma pigmentação que o distingue de outros elementos da imagem. No entanto, é necessário identificar todo o corpo celular. Para isso, o maior núcleo celular é encontrado e uma caixa delimitadora com um espaço de 10 pixels é criada. Finalmente, as imagens foram centralizadas com um tamanho final de 310x310. Na etapa de extração de características, foram utilizados três descritores de textura e uma de cor: Haralick; Hu Moments; Threshold Adjacency

Statistics (TAS) e o Histograma de Cores. Para a construção do modelo preditivo, foi utilizado o algoritmo baseado em árvores de decisão com aumento de gradiente XGBoost [Chen and Guestrin 2016]. Finalmente, para validar o método proposto, foram utilizadas as métricas acurácia, precisão e F-Measure.

### 3. Resultados e Discussão

Esta seção apresenta e discute os resultados obtidos pelo método proposto. Para fins de validação, a base de dados foi dividida em 80% para treinamento e 20% para teste. Além disso, foi verificada a eficiência do método proposto comparando com outras técnicas de classificação, a extração de características usando a rede neural VGG16 e ResNet50 pré-treinadas com os pesos da base ImageNet. A Tabela 1 apresenta os resultados obtidos por todas as técnicas usando o mesmo conjunto de teste. Pode-se observar que o método proposto obteve o melhor resultado, alcançando 90% de acurácia, 89% de precisão e 85 de F-Measure.

**Tabela 1. Resultados da classificação dos glóbulos brancos.**

Experimentos	Acc (%)	Precisão (%)	F-Measure
VGG16	64,42	65	60
Resnet50	65,38	65	60
<b>Método proposto</b>	<b>90</b>	<b>89</b>	<b>85</b>

Analisando o melhor resultado alcançado e comparando-o com os trabalhos relacionados, verifica-se que [Vatathanavaro et al. 2018] obteve 88,29% de acurácia, usando a base BCCD e LISC. Entretanto, obtivemos resultados superiores. E em [Mohamed et al. 2020], foi obtido 97% de acurácia, usando vários modelos pré-treinados e uma alta taxa de validação, 90% para treino e 10% para teste. Portanto, os resultados alcançados no método proposto mostram que a combinação de técnicas de análise de textura e o classificador XGBoost produz boa discriminação das classes dos glóbulos brancos, o que comprovam a eficiência do método proposto.

### 4. Conclusões

Este artigo apresentou um método automático para classificação dos glóbulos brancos. Uma de suas principais vantagens é que possui um custo computacional consideravelmente baixo, comparado às abordagens de aprendizado profundo. O método proposto mostrou resultados promissores usando as técnicas descritas. Como trabalhos futuros, pretende-se analisar outras técnicas de textura, além de descritores de forma, como HOG. Por fim, explorar alternativas relacionadas à detecção dos glóbulos brancos usando redes neurais. Acredita-se que as modificações mencionadas possam proporcionar melhores resultados de classificação.

### Referências

- ACS (2018). About chronic lymphocytic leukemia. <https://cancer.org/cancer/chronic-lymphocytic-leukemia/about/normal-tissue.html>. Acesso em: 2020-06-07.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM.
- Mohamed, E., El-Behaidy, W., Khoriba, G., and Li, J. (2020). Improved white blood cells classification based on pre-trained deep learning models. *Journal of Communications Software and Systems*, 16.
- NCI (2020). White blood cell. <https://cancer.gov/publications/dictionaries/cancer-terms/def/white-blood-cell>. Acesso em: 2020-06-07.
- Vatathanavaro, S., Tungjitnob, S., and Pasupa, K. (2018). White blood cell classification: A comparison between vgg-16 and resnet-50 models. page 3.