

Contagem Diferencial Automática de Leucócitos em Imagens Microscópicas

Ricardo M. Santos, Murillo Bouzon, Nayara Gerolomo, Daniel L. Costa, Paulo Sérgio Rodrigues

Departamento de Ciência da Computação, Centro Universitário FEI

São Bernardo do Campo, SP, Brasil

{unifrsantos, unifmbouzon, psergio}@fei.edu.br

{nayara.gerolomo,dlcosta}@gmail

Resumo—A contagem diferencial de leucócitos é importante para o diagnóstico de várias doenças. Quando feito manualmente, esse processo é lento e apresenta variância intra e inter-pessoal de resultados, além de requerer um especialista. Este trabalho trata de um método automático de contagem diferencial de neutrófilos e linfócitos a partir de imagens microscópicas. O método proposto é dividido em três etapas principais: pré-processamento, segmentação e reconhecimento dos leucócitos. Enquanto a segmentação baseia-se na limiarização, a classificação dos tipos de leucócitos, contida na etapa de reconhecimento, é feita por máquinas de vetor de suporte (SVMs, *support vector machines*). O método foi construído, parametrizado e validado com base nos resultados de experimentos conduzidos sobre um banco de imagens proveniente de lavagens bronco-alveolares feitas em ratos de laboratório. O método proposto atingiu $92\% \pm 0.04$ de área sob a curva ROC para classificação de neutrófilos e $92\% \pm 0.06$ para linfócitos.

Abstract—The differential leukocyte count is important for the diagnosis of various diseases. When done manually, this process is slow and presents intra- and inter-personal variance of results, in addition to requiring a specialist. This work deals with an automatic method of differential counting of neutrophils and lymphocytes from microscopic images. The proposed method is divided into three main stages: pre-processing, segmentation and recognition of leukocytes. While segmentation is based on thresholding, the classification of the leukocyte types contained in the recognition step is done by support vector machines (SVMs). The method was built, parameterized and validated based on the results of experiments conducted on an image dataset containing bronchoalveolar lavages done in laboratory mice. The proposed method reached $92\% \pm 0.04$ area under ROC for neutrophil classification and $92\% \pm 0.06$ for lymphocytes classification.

I. INTRODUÇÃO

Células sanguíneas humanas são principalmente divididas em três categorias: células vermelhas (RBCs, *red blood cells*), células brancas (WBCs, *white blood cells*) e plaquetas. As WBCs, também chamadas de leucócitos, por sua vez, fazem parte do sistema imunológico do organismo.

A contagem de WBCs no sangue de um indivíduo é um dos fatores mais importantes para o diagnóstico de uma série de doenças. Um paciente soropositivo, por exemplo, tende a apresentar uma quantidade menor de WBCs no sangue [9]. Em contrapartida, anemias e estresse, por exemplo, são condições que aumentam a contagem de WBCs [11].

O termo contagem diferencial refere-se ao processo de contar WBCs, fazendo distinção do tipo de WBC.

A contagem diferencial manual (i.e. contar as células uma a uma em um microscópio) requer o trabalho exaustivo de um especialista. Além de demorado, esse processo apresenta alta intra e intervariância de resultados [14], fato pelo qual não é muito confiável para pesquisas, por exemplo. Assim, faz-se necessária a concepção de métodos de contagem diferencial automáticos ou semi-automáticos rápidos, robustos e de baixo custo do ponto de vista computacional.

Atualmente, uma série de pesquisas tem sido feita a fim de automatizar o processo de contagem diferencial. O trabalho de [7] apresenta um sistema para contagem total de células. Em [6], propõe-se desenvolver um sistema automático de contagem diferencial de WBCs de uma dada amostra de sangue utilizando lógica Fuzzy e *Random Forest*. No entanto, o método apresenta uma taxa de erro elevada, de 7,5%. O trabalho de [12] descreve o uso de algoritmos genéticos para contagem de células coradas por imunocitoquímica em imagens microscópicas. Outros trabalhos similares na área podem ser encontrados em [3], [8], [15], [17]. Um *review* da área é apresentado em [13].

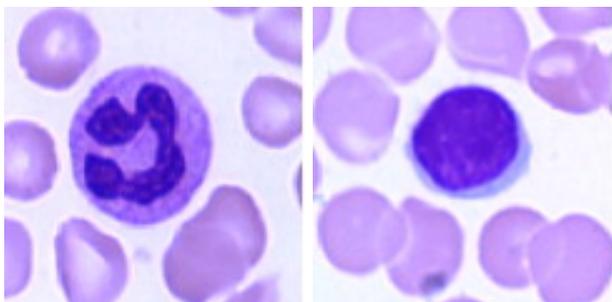
Aglomerados de células podem interferir no resultado final da contagem automática. Por isso, o trabalho de [10] apresenta uma metodologia que lida explicitamente com aglomerados de células. Em [1], é proposta uma forma de segmentar e contar células sanguíneas isoladas e sobrepostas utilizando redes neurais artificiais. O método trabalha apenas com imagens de microscópios de RBCs com escala na ordem de $100x$ e que contenham as mesmas características, como formato e tamanho. Por isso, quanto mais sobrepostas as células, pior é a performance do algoritmo. O artigo de [19] investiga o problema de individualização de células cerebrais, que costumam se aglomerar e ter bordas mal definidas.

As imagens microscópicas podem conter diversas características que dificultam a contagem automática, como baixo contraste e ruídos. Como exemplo, a Figura 1a contém uma imagem de neutrófilos (circunferência central da imagem). Outro exemplo pode ser visto na Figura 1b, que contém uma imagem de linfócitos (também na circunferência central da imagem). Assim, o trabalho de [18] propõe a utilização da transformada *bottom-hat* para melhorar a qualidade da imagem como etapa de pré-processamento para segmentação de imagens histopatológicas de câncer. Outra abordagem para extração de características para contagem automática de

células de pus pode ser vista em [8]. O método, baseado em *gray-thresholding*, pode contabilizar células sobrepostas e individuais. Apesar disso, ainda apresenta erros relacionados à detecção de células. Essa característica reflete diretamente na performance final do algoritmo, que apresenta 87.32% de acurácia e 85.37% de precisão.

Uma das etapas do processo de contagem de células é a segmentação. Essa etapa é primordial, uma vez que impacta diretamente no produto final da contagem. O trabalho de [4] apresenta uma técnica para segmentação de células humanas de câncer de cólon utilizando *fuzzy clustering*. Em [16], é apresentado um método para segmentar células cervicais. Esse processo de segmentação é particularmente desafiador, uma vez que as células têm formatos variados e são comumente sobrepostas. Outro método para segmentação de células de núcleo deformado é proposto em [2]. Neste processo, as células saudáveis são ignoradas na segmentação. O trabalho de [14] categoriza, avalia e discute métodos recentes na área de contagem e detecção automática de leucócitos. Assim, nesse trabalho, a contagem diferencial automática será feita sobre dois tipos de células, os neutrófilos e linfócitos. A contagem é dividida em três etapas. A etapa de pré-processamento, segmentação baseada em limiarização e a etapa de reconhecimento. Nesta etapa final, a classificação será feita com base no *Support Vector Machine (SVM)*.

Este trabalho está organizado da seguinte maneira: na Seção II é descrito o *pipeline* de processamento e classificação das imagens. A Seção III apresenta os resultados e uma breve discussão. A conclusão é apresentada na Seção IV.



(a) Exemplo de neutrófilo na (b) Exemplo de linfócito na circunferência central.

Figura 1: WBCs contadas neste trabalho.

II. METODOLOGIA

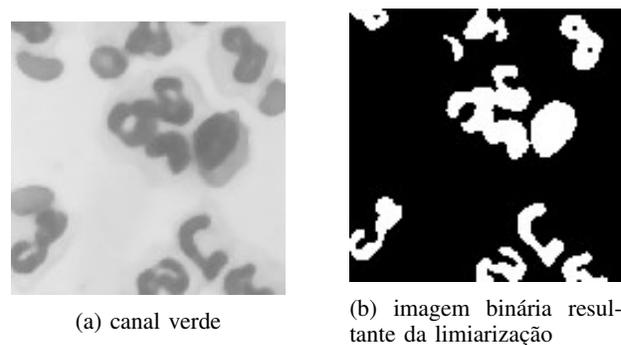
O método proposto neste trabalho objetiva a contagem diferencial automática de linfócitos e neutrófilos. A abordagem é dividida em três etapas, cada qual contendo duas sub-etapas. Ao total, o *pipeline* de processamento compreende seis sub-etapas. As três etapas e suas sub-etapas são exibidas na Figura 3.

Na sub-etapa 1 do pré-processamento, o objetivo é delimitar a região de interesse para resolução do problema. Assim, é empregada uma binarização pelo método de Otsu. O balanceamento de histograma é feito sobre o canal RGB azul,

no qual as intensidades das células são mais próximas das intensidades do fundo da lâmina e mais distantes das intensidades do exterior da lâmina. Essa escolha do canal RGB será validada posteriormente na Seção III. A máscara resultante, como mostrado na Figura 4, pode conter regiões espúrias além da lâmina. A fim de removê-las, são ignorados todos os componentes conexos do *foreground*, exceto o de maior área (lâmina). Por fim, sobre a lâmina, é encaixado um círculo, delimitando a máscara final.

Na sub-etapa 2, é aplicado um filtro de mediana com *kernel* 9×9 , de acordo com os resultados adquiridos e a literatura [14].

Com o objetivo de separar a área da imagem pertencente a leucócitos de seu complemento, na sub-etapa 3 um processo de limiarização é aplicado sobre o histograma do canal RGB verde. Essa escolha foi feita devido ao canal verde apresentar maior diferença de intensidade entre as células e o fundo da lâmina. Neste processo de limiarização, é utilizado o limiar fixo 146. O resultado é exibido na Figura 2.



(a) canal verde

(b) imagem binária resultante da limiarização

Figura 2: Demonstração da limiarização de histograma no limiar de valor 146 sobre o canal RGB verde de uma imagem.

A seguir, é necessário separar as regiões referentes a cada célula ou outro artefato presente na lâmina. Para tal, emprega-se uma extração de BLOBs para encontrar os componentes conexos do *foreground*. O método utilizado é o de seguimento de borda.

A sub-etapa 5 recebe uma região e extrai um conjunto de características referentes ao seu formato e *pixels* internos. As características a serem extraídas, selecionadas da literatura [14], compreendem a área da região e de seu envoltório convexo, solidez (razão entre os dois anteriores), perímetro, compacidade, circularidade, retangularidade e existência de contornos internos. As características de intensidade de *pixels* levadas em conta são média, variância e desvio padrão. Ao todo, nesta sub-etapa, serão levadas em conta essas 11 características.

Na sub-etapa 6, utilizam-se modelos de classificação para distinguir regiões referentes a células de interesse. Dessa forma, diferenciam-se neutrófilos e linfócitos de regiões que representam outras células na lâmina, outros artefatos ou ruído. Para tanto, os vetores de características da sub-etapa 5 são entrada para duas SVMs lineares, cada uma das quais

responsável por reconhecer um dos dois tipos de leucócitos de interesse.

III. EXPERIMENTOS

As imagens utilizadas neste trabalho compõem um banco de dados de 129 imagens microscópicas de resolução 3120×4160 . Com o objetivo de validar e treinar o método proposto foram utilizados alguns tipos de anotações manuais, como posicionamento da lâmina e contornos.

Durante o decorrer deste trabalho, foram conduzidos 6 experimentos. Cada experimento compreende uma sub-etapa da metodologia. Neste artigo, apenas o experimento final é descrito devido à limitação de espaço e por ser o mais importante na avaliação do método. Entretanto, nos demais experimentos, foram obtidos os seguintes resultados:

- Sub-etapa de delimitação da lâmina: a melhor performance é obtida sob o canal RGB azul e com os passos de remoção de regiões espúrias e encaixe de círculo presentes;
- Sub-etapa de remoção de ruídos: o filtro de mediana obteve os melhores resultados no melhoramento de imagem;
- Sub-etapa de limiarização: para esta sub-etapa, foram comparados os métodos de segmentação com um limiar fixo e utilizando divergência Fuzzy. Para ambos os métodos, notou-se que o canal RGB verde é mais apropriado para a segmentação de leucócitos por limiarização (de acordo com [5]). Entretanto, a segmentação utilizando divergência Fuzzy apresentou qualidade inferior quando comparada à segmentação com um limiar fixo. Dessa forma, neste trabalho, considerou-se a segmentação com um limiar fixo.
- Sub-etapa do operador morfológico de abertura: a utilização do operador morfológico de abertura é prejudicial ao método.

A. Reconhecimento

Este experimento (sub-etapa 6) constitui o mais importante deste trabalho, por isso é descrito em mais detalhes. Neste experimento, objetivou-se avaliar a capacidade da etapa de reconhecimento na delimitação de objetos de interesse. Nesta etapa, foram providas as regiões que representem os neutrófilos e linfócitos. Foi utilizado o esquema de validação K-Fold para particionar, em 4 iterações, as regiões de rótulo anotado em um conjunto de treinamento e um de teste. Os resultados do classificador foram comparados com as anotações e analisados por curvas de ROC, de acordo com a Figura 5. A Figura 5a mostra o desempenho do método proposto de acordo com a curva ROC. Nessa figura, a metodologia proposta atingiu $92\% \pm 0.04$ de área sob essa curva para a classificação de neutrófilos. O mesmo pode ser observado na Figura 5b, onde obteve-se $92\% \pm 0.06$ de área sob a curva ROC para a classificação de linfócitos.

IV. CONCLUSÃO

Neste trabalho, foi construída e validada uma metodologia de contagem diferencial automática de neutrófilos e

linfócitos. A metodologia proposta compreende três etapas: pré-processamento, segmentação e reconhecimento. Cada etapa é sub-dividida em outras duas sub-etapas.

Foram realizados seis experimentos no total. Cada experimento foi aplicado com o objetivo de avaliar e justificar uma sub-etapa da metodologia. Devido à limitação do espaço, apenas o experimento referente à sub-etapa 6 (experimento mais relevante) foi descrito neste artigo. Entretanto, os demais resultados referentes às outras sub-etapas foram citados. Foram feitos testes K-Fold para validar o método. Tanto para a classe de neutrófilos quanto para os linfócitos, a área sob a curva ROC foi de 92%. Assim, o método proposto demonstrou ser eficiente na contagem diferencial automática.

Como trabalhos futuros, pretende-se aplicar o método proposto em outras classes de células e comparar com as propostas na literatura.

V. AGRADECIMENTOS

Agradecimentos à FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo - Processo nº 2018/02612-7), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e ao Centro Universitário da FEI.

REFERÊNCIAS

- [1] S. S. Adagale and S. S. Pawar. Image segmentation using pcnn and template matching for blood cell counting. In *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–5, Dec 2013.
- [2] P. Faridi, H. Danyali, M. S. Helfroush, and M. A. Jahromi. An automatic system for cell nuclei pleomorphism segmentation in histopathological images of breast cancer. In *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5, Dec 2016.
- [3] J. Ge, Z. Gong, J. Chen, J. Liu, J. Nguyen, Z. Y. Yang, C. Wang, and Y. Sun. A system for automated counting of fetal and maternal red blood cells in clinical kb test. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1706–1711, May 2014.
- [4] A. Gharipour and A. W. C. Liew. Fuzzy clustering using local and global region information for cell image segmentation. In *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 216–222, July 2014.
- [5] Madhumala Ghosh, Devkumar Das, Chandan Chakraborty, and Ajoy K. Ray. Automated leukocyte recognition using fuzzy divergence. *Micron*, 41(7):840 – 846, 2010.
- [6] Pramit Ghosh, Debotosh Bhattacharjee, and Mita Nasipuri. Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique. *Applied Soft Computing*, 46:629 – 638, 2016.
- [7] X. Guo and F. Yu. A method of automatic cell counting based on microscopic image. In *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 1, pages 293–296, Aug 2013.
- [8] R. Hamid, F. Naim, N. W. Arshad, and Z. Mohamed. Automated detection and counting of pus cells on sputum images. In *2013 International Conference on Electronics, Computer and Computation (ICECCO)*, pages 5–8, Nov 2013.
- [9] Thomas H Brannagan III and Yili Zhou. Hiv-associated guillain-barré syndrome. *Journal of the Neurological Sciences*, 208(1–2):39 – 42, 2003.
- [10] S. Kothari, Q. Chaudry, and M. D. Wang. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 795–798, June 2009.
- [11] Lorraine O. De Labry, Edward W. Campion, Robert J. Glynn, and Pantel S. Vokonas. White blood cell count as a predictor of mortality: Results over 18 years from the normative aging study. *Journal of Clinical Epidemiology*, 43(2):153 – 157, 1990.

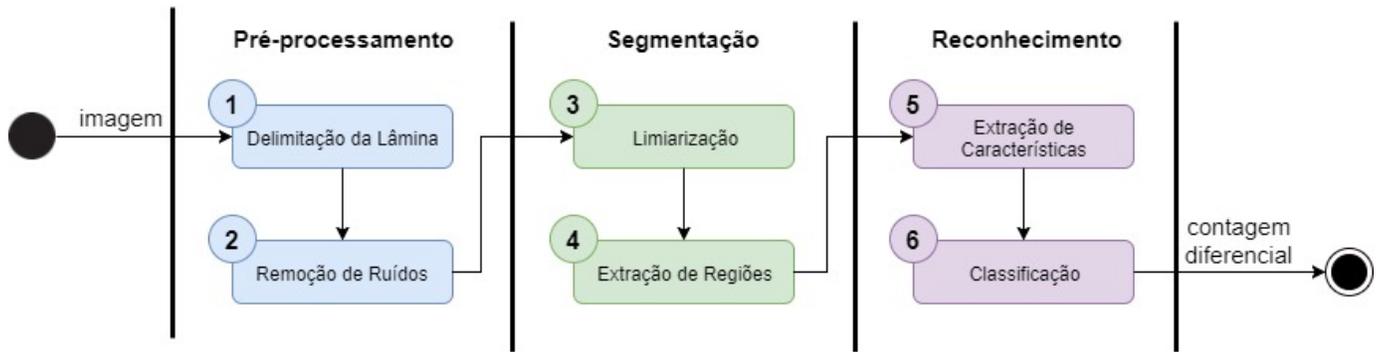
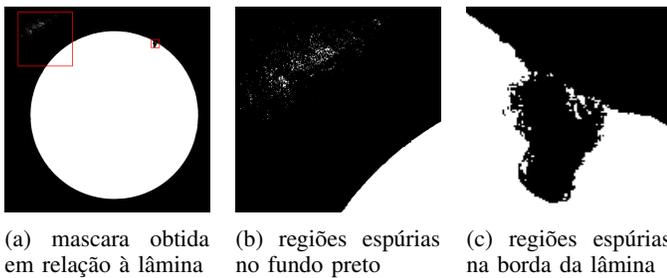


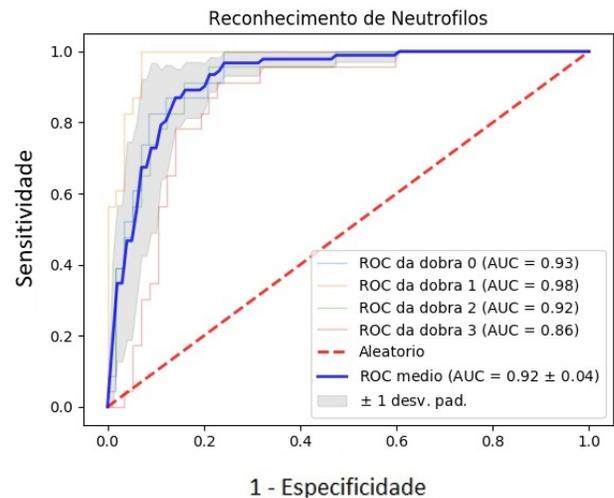
Figura 3: Etapas do método automático de contagem diferencial de leucócitos deste trabalho.



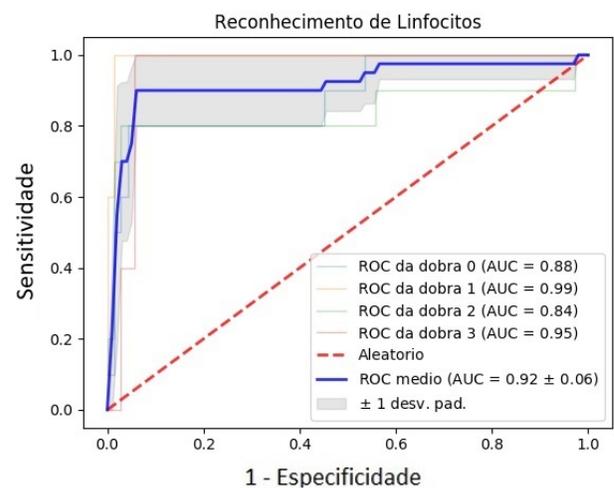
(a) máscara obtida em relação à lâmina (b) regiões espúrias no fundo preto (c) regiões espúrias na borda da lâmina

Figura 4: Regiões espúrias presentes na máscara obtida da binarização pelo método de Otsu para delimitação da lâmina.

- [12] M. Ramin, A. Sepas-Moghaddam, P. Ahmadvand, and M. M. Dehshibi. Counting the number of cells in immunocytochemical images using genetic algorithm. In *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, pages 185–190, Dec 2012.
- [13] L. E. H. Reyes, L. X. B. Rozo, and F. A. R. Morales. Automatic leukocyte image segmentation: A review. In *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*, pages 1–9, Sept 2015.
- [14] Mukesh Saraswat and K.V. Arya. Automated microscopic image analysis for leukocytes identification: A survey. *Micron*, 65:20 – 33, 2014.
- [15] W. Tangsuksant, C. Pintavirooj, S. Taertulakarn, and S. Daochai. Development algorithm to count blood cells in urine sediment using ann and hough transform. In *The 6th 2013 Biomedical Engineering International Conference*, pages 1–4, Oct 2013.
- [16] Afaf Tareef, Yang Song, Weidong Cai, Heng Huang, Hang Chang, Yue Wang, Michael Fulham, Dagan Feng, and Mei Chen. Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation. *Neurocomputing*, 221:94 – 107, 2017.
- [17] Razali Tomari, Wan Nurshazwani Wan Zakaria, Muhammad Mahadi Abdul Jamil, Faridah Mohd Nor, and Nik Farhan Nik Fuad. Computer aided system for red blood cell classification in blood smear image. *Procedia Computer Science*, 42:206 – 213, 2014.
- [18] Pin Wang, Xianling Hu, Yongming Li, Qianqian Liu, and Xinjian Zhu. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, 122:1 – 13, 2016.
- [19] Z. You, M. E. Vandenberghe, Y. Balbastre, N. Souedet, P. Hantraye, C. Jan, A. S. Herard, and T. Delzescaux. Automated cell individualization and counting in cerebral microscopic images. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3389–3393, Sept 2016.



(a) acurácia no reconhecimento de neutrófilos



(b) acurácia no reconhecimento de linfócitos

Figura 5: Curvas ROC do reconhecimento de neutrófilos e linfócitos.