

---

# **Método Computacional para Segmentação não Supervisionada de Imagens Histológicas de Linfoma**

---

**Thaína Aparecida Azevedo Tosta**



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

Uberlândia  
2016



**Thaína Aparecida Azevedo Tosta**

**Método Computacional para Segmentação não  
Supervisionada de Imagens Histológicas de  
Linfoma**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Marcelo Zanchetta do Nascimento

Uberlândia  
2016

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Sistema de Bibliotecas da UFU, MG, Brasil.**

---

T716m Tosta, Thaína Aparecida Azevedo, 1992-  
2016      Método computacional para segmentação não supervisionada de  
                  imagens histológicas de linfoma / Thaína Aparecida Azevedo Tosta. -  
                  2016.  
                  120 f. : il.

Orientador: Marcelo Zanchetta do Nascimento.  
Dissertação (mestrado) - Universidade Federal de Uberlândia,  
Programa de Pós-Graduação em Ciência da Computação.  
Inclui bibliografia.

1. Computação - Teses. 2. Processamento de imagens - Teses. 3.  
Câncer - Diagnóstico - Teses. 4. Linfoma - Diagnóstico - Teses. I.  
Nascimento, Marcelo Zanchetta do. II. Universidade Federal de  
Uberlândia. Programa de Pós-Graduação em Ciência da Computação.  
III. Título.

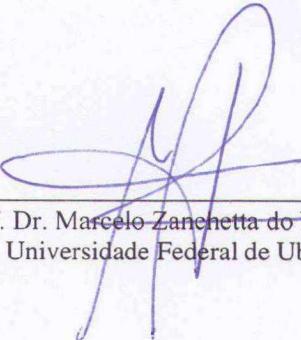
---

CDU: 681.3

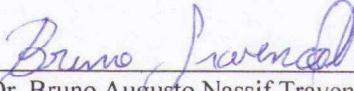
UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada "**Método Computacional para Segmentação não Supervisionada de Imagens Histológicas de Linfoma**" por **Thaína Aparecida Azevedo Tosta** como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação

Uberlândia, 29 de Janeiro de 2016

Orientador:   
Prof. Dr. Marcelo Zanchetta do Nascimento  
Universidade Federal de Uberlândia

Banca Examinadora:

  
Prof. Dr. Bruno Augusto Nassif Travençolo  
Universidade Federal de Uberlândia

  
Prof. Dr. Marcelo Costa Oliveira  
Universidade Federal de Alagoas



*Para minha família, meus amigos e o pequeno Joaquim.*



---

## **Agradecimentos**

Agradeço primeiramente ao meu orientador, Marcelo Zanchetta do Nascimento, pela paciência e confiança no meu trabalho ao longo do desenvolvimento deste projeto. Também agradeço ao professor Paulo Rogério de Faria pelos ensinamentos de histologia e pela imensa ajuda neste trabalho.

Aos professores da Faculdade de Computação da Universidade Federal de Uberlândia, com os quais tive a oportunidade de trabalhar tanto na graduação quanto no mestrado, que contribuíram de forma imensurável para a minha formação pessoal e profissional. Ao secretário da pós-graduação, Erisvaldo Araújo Fialho, pela ajuda e paciência para solucionar milhares de dúvidas. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG - projeto TEC- APQ-02885-15) pelo apoio financeiro.

À minha família por todo o apoio. À família que formei em Uberlândia com Thiago Tosta, Thamiris Rodrigues, Diogo Vilela e Joaquim Tosta, que foram essenciais em tantos momentos felizes e difíceis. Aos meus amigos de Uberaba, da graduação, do mestrado e do laboratório pelo companheirismo incondicional.

E à Andrêssa Finzi, minha companheira nos últimos dois anos.



*“Dream until your dream comes true.”*  
*(Steven Tyler)*



---

## Resumo

A análise de imagens histológicas representa uma das maiores evoluções da medicina moderna. Aliados a essa evolução, métodos computacionais vêm sendo amplamente desenvolvidos para auxiliar especialistas na análise dessas imagens para determinar diagnósticos, prognósticos e tratamentos adequados à condição do paciente. Porém, ao ser realizada por especialistas, essa tarefa torna-se dispendiosa e suscetível a variabilidades inter e intrapatologistas. Para aperfeiçoar tal prática tradicional para diagnósticos de Linfoma de Células do Manto, Linfoma Folicular e Leucemia Linfóide Crônica, este trabalho propõe um método para a segmentação não supervisionada dos componentes nucleares de células indicativas de tais neoplasias utilizando imagens histológicas coradas com Hematoxilina-Eosina. O método proposto foi dividido nas etapas de pré-processamento, segmentação e pós-processamento. Na etapa de pré-processamento, as técnicas de equalização do histograma e filtro gaussiano foram aplicadas sobre os canais componentes do modelo de cores RGB. Na segmentação, foi aplicada uma técnica de limiarização resultante da combinação entre os métodos *fuzzy 3-partition entropy* e algoritmo genético. Por fim, para aperfeiçoamento dos resultados da segmentação, foram utilizadas operações morfológicas e a técnica *valley-emphasis*. Para avaliar o método desenvolvido, imagens histológicas de linfoma com magnificação 20 $\times$  foram selecionadas e segmentadas manualmente por um especialista. Essas imagens de referência (padrão-ouro) permitiram a extração de medidas quantitativas para a comparação entre este método e diferentes técnicas propostas na literatura. Além disso, uma avaliação qualitativa foi realizada levando a resultados relevantes e superiores aos trabalhos comparados. Também foi analisada a sua aplicação sobre as etapas de extração de características e classificação das diferentes lesões consideradas, obtendo resultados de acurácia próximos a 100%.

**Palavras-chave:** Segmentação Nuclear. Imagens Histológicas de Linfoma. Linfoma de Células do Manto. Linfoma Folicular. Leucemia Linfóide Crônica. Algoritmo Genético.



---

## Abstract

Histological image analysis represents a major evolutionary step in modern medicine. Associated with this step, computational methods are being widely developed to help specialists during the analysis of these images to determine diagnostics, prognostics and appropriate treatments in accordance with the condition of the patient. However, when it is performed by specialists, this task becomes time-consuming and susceptible to inter-and intra-pathologist variability. To improve this traditional practice for diagnostics of Mantle Cell Lymphoma, Follicular Lymphoma and Chronic Lymphocytic Leukemia, this study proposes a method for the unsupervised segmentation of nuclear components in indicative cells of such neoplasias using histological images stained with Hematoxylin-Eosin. The proposed method was divided into preprocessing, segmentation and post processing. In the preprocessing step, the techniques used in histogram equalization and Gaussian filter were applied to the channels from RGB color model. In the segmentation, a thresholding technique was applied combining the methods of fuzzy 3-partition entropy and genetic algorithm. Finally, for the improvement of the segmentation results, morphological operations and the valley-emphasis technique were used. For evaluating the developed method, histological images of lymphoma with magnification 20 $\times$  were selected and manually segmented by a specialist. Those reference images (gold standard) allowed the extraction of quantitative measures in order to compare this method with different techniques proposed in the literature. Furthermore, a qualitative evaluation was conducted leading to relevant and improved results over those from compared studies. Its application was also analysed considering the steps of feature extraction and classification of the lesions, obtaining results of accuracy close to 100%.

**Keywords:** Nuclear Segmentation. Histological Images of Lymphoma. Mantle Cell Lymphoma. Follicular Lymphoma. Chronic Lymphocytic Leukemia. Genetic Algorithm.



---

## Listas de ilustrações

Figura 1 – Imagens histológicas de LF coradas com H&E com dimensões de 512 × 512 pixels e magnificação 40×, contendo identificações de centroblastos realizadas por cinco especialistas indicadas por círculos identificados por diferentes cores (SERTEL et al., 2010b) . . . . .	28
Figura 2 – Relação entre as etapas de desenvolvimento, ativação e diferenciação de células B e os linfomas associados a cada uma delas (GERARD; BISHOP, 2012). . . . .	34
Figura 3 – Lâmina de LCM com células “regularmente irregulares” (PATHPEDIA, 2014). . . . .	35
Figura 4 – Lâmina contendo centroblastos, com um deles indicado por uma seta branca, e centrócitos indicados por setas amarelas (UNICAMP, 2013). . . . .	36
Figura 5 – Lâmina de LLC caracterizada por linfócitos tumorais morfológicamente semelhantes aos normais (UNICAMP, 2013) . . . . .	37
Figura 6 – Exemplos de digitalizações de lâminas de LCM (a), LF (b) e LLC (c) . . . . .	39
Figura 7 – Fluxograma do algoritmo genético (KAUSHIK et al., 2013) . . . . .	42
Figura 8 – Exemplo de um processo de <i>crossover</i> executado sobre dois indivíduos para a geração de seus descendentes. . . . .	43
Figura 9 – Exemplo da operação morfológica de dilatação: (a) imagem original, (b) elemento estruturante de tamanho 3 × 3 em forma de cruz e (c) resultado da dilatação. . . . .	44
Figura 10 – Exemplo da operação de erosão: imagem original (a) processada por um elemento estruturante de tamanho 3 × 3 em forma de cruz (b) e seu resultado (c) . . . . .	45
Figura 11 – Exemplo da operação de abertura binária: imagem original (a) e o resultado de sua abertura (c) por um elemento estruturante de tamanho 3 × 3 em forma de cruz (b) . . . . .	45

Figura 12 – Exemplo da operação de fechamento binária: (a) imagem original, (b) elemento estruturante de tamanho $3 \times 3$ em forma de cruz utilizado para seu processamento e (c) resultado de seu fechamento. . . . .	45
Figura 13 – Fluxograma do trabalho proposto. . . . .	68
Figura 14 – Separação de canais do modelo de cores RGB de uma subimagem do caso <i>sj-05-5269-R10_002</i> da classe LLC (a), com alguns núcleos neoplásicos indicados por setas vermelhas e outros núcleos saudáveis indicados por setas em azul, que exemplifica a diferença de contraste entre os canais <i>R</i> (b), <i>G</i> (c) e <i>B</i> (d), e seus respectivos histogramas ilustrados por (e), (f) e (g). . . . .	69
Figura 15 – Imagens com uso da técnica de equalização de histograma: (a) subimagem do caso <i>sj-05-5269-R10_002</i> da classe LLC, (b) seu histograma, (c) imagem após a aplicação da técnica de equalização do histograma e (d) histograma da imagem equalizada. . . . .	71
Figura 16 – (a) Representação gráfica de uma função Gaussiana bidimensional com média $(0,0)$ e desvio padrão igual a 1, e (b) a máscara de tamanho $3 \times 3$ resultante dessa representação. . . . .	71
Figura 17 – Exemplo da aplicação das técnicas de equalização do histograma e filtro Gaussiano sobre o canal <i>G</i> das imagens da classe LLC: (a) imagem original, (b) seu canal <i>G</i> extraído e (c) aplicação das técnicas de pré-processamento utilizadas. . . . .	72
Figura 18 – Representação dos conceitos de indivíduo e população, onde cada linha representa um indivíduo, cada coluna corresponde a uma das variáveis a serem definidas e a matriz como um todo representa uma população. . . . .	73
Figura 19 – Representação obtida pela aplicação das funções de associação, $M_n$ , $M_c$ e $M_f$ , sobre cada nível de intensidade da imagem. . . . .	74
Figura 20 – Gráfico representativo das funções de associação utilizando o método <i>fuzzy 3-partition entropy</i> com indicação dos valores de limiar por setas de cor preta (Figura adaptada de (YIN et al., 2014) © 2014 Elsevier). . . . .	75
Figura 21 – Representação matricial do processo de seleção. Em (a), uma população inicial é representada juntamente com os valores obtidos pela aplicação da função de avaliação ( <i>FA</i> ) de cada indivíduo. Em (b), o resultado do processo de seleção é ilustrado contendo apenas 30% dos indivíduos da população anterior. . . . .	76
Figura 22 – Exemplo de um processo de mutação, onde, em (a), a probabilidade de mutação do terceiro parâmetro é menor que o limiar definido (0,01) o que leva à alteração de seu valor de forma aleatória, representado em (b). . . . .	77

Figura 23 – Exemplo da aplicação do Algoritmo 2 para a definição do critério de parada ao longo de duas gerações ((a) e (b)) por meio da representação do histograma de uma imagem. . . . .	79
Figura 24 – Exemplo ilustrativo de regiões intra-nucleares identificadas pela segmentação na imagem original (a) e sua representação ampliada (b) para percepção das características de cor identificadas como parte integrante de dois núcleos segmentados (c). . . . .	80
Figura 25 – Exemplo da operação de dilatação binária: imagem binária (a), suas regiões correspondentes identificadas sobre a imagem original (b) e o resultado de sua dilatação representado por uma imagem binária (c) e pela imagem original (e) utilizando um elemento estruturante com distribuição em disco com raio igual a 2 (c). . . . .	81
Figura 26 – Exemplo da operação de abertura binária: imagem binária (a), imagem original (b) e o resultado de sua abertura representado por uma imagem binária (d) e seu mapeamento sobre a imagem original (e) obtido por um elemento estruturante de tamanho $3 \times 3$ com distribuição formando um quadrado (c). . . . .	81
Figura 27 – Relação entre regiões identificadas manualmente (R) e automaticamente (S) (OGER; BELHOMME; GURCAN, 2012). . . . .	83
Figura 28 – Exemplo da aplicação de pré-processamento sobre imagens da classe LCM: (a) imagem original, (b) seu canal $R$ extraído e (c) resultado da aplicação das técnicas de pré-processamento que alcançaram entropia mínima. . . . .	85
Figura 29 – Exemplo da aplicação das técnicas equalização do histograma e filtro Gaussiano sobre o canal $B$ das imagens da classe LF: (a) imagem original, (b) seu canal $B$ e (c) aplicação das técnicas de pré-processamento supracitadas. . . . .	86
Figura 30 – Exemplo da aplicação dos métodos de pré-processamento descritos sobre imagens da classe LLC: (a) imagem original, (b) seu canal $G$ e (c) aplicação das técnicas de equalização do histograma e filtro Gaussiano. . . . .	86
Figura 31 – Exemplo da aplicação do filtro Gaussiano sobre uma subimagem do caso <i>sj-04-4525-R4_001</i> da classe LCM, considerando diferentes tamanhos de sua máscara e atribuições à sigma: (a), (e), (i), (m) canal $R$ da imagem original, aplicações com máscaras de tamanho $3 \times 3$ ((b), (f), (j), (n)), $5 \times 5$ ((c), (g), (k), (o)) e $7 \times 7$ ((d), (h), (l), (p)) pixels, com atribuições à variável sigma de 0,5 ((b)-(d)), 2 ((f)-(h)), 3,5 ((j)-(l)) e 5 ((n)-(p)). . . . .	88

Figura 32 – Exemplo da aplicação do AG aliado à técnica <i>fuzzy 3-partition entropy</i> : (a) população inicial com destaque para indivíduos selecionados pela função de avaliação e (b) geração subsequente após a aplicação das etapas de seleção, <i>crossover</i> e mutação. . . . .	90
Figura 33 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso <i>sj-05-768_013</i> da classe LCM: (a) imagem original, (b) imagem resultante do pré-processamento, (c) seu histograma com valores de limiar 73 e 194 indicados por linhas vermelhas, (d) imagem binária resultante da segmentação, (e) mapeamento das regiões segmentadas sobre a imagem original e (f) região ampliada para análise do resultado. . . . .	91
Figura 34 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso <i>sj-05-6124-R3_006</i> da classe LF: (a) imagem original, (b) imagem resultante do pré-processamento descrito, (c) histograma da imagem anterior com valores de limiar 67 e 189 indicados por linhas vermelhas, (d) imagem binária provinda da segmentação, (e) identificação das regiões obtidas na segmentação sobre a imagem original e (f) região ampliada para análise do desempenho da segmentação. . . . .	92
Figura 35 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso <i>sj-05-3344_006</i> da classe LLC: (a) imagem original, (b) imagem resultante da aplicação da equalização do histograma e do filtro Gaussiano, (c) histograma da imagem pré-processada com valores de limiar 80 e 197 indicados por linhas vermelhas, (d) imagem binária da segmentação, (e) mapeamento das regiões segmentadas sobre a imagem original e (f) região ampliada para avaliação do resultado. . . . .	93
Figura 36 – Exemplo da aplicação do método de segmentação <i>valley-emphasis</i> para a separação de células unidas: (a), (d) e (g) representam imagens resultantes da segmentação das classes LCM, LF e LLC, respectivamente, (b), (e) e (h) ilustram regiões ampliadas de suas imagens segmentadas correspondentes, (c), (f) e (i) apresentam tais regiões após esse processamento. . . . .	94
Figura 37 – Exemplo da aplicação das operações morfológicas de dilatação e abertura sobre células identificadas na segmentação: (a), (d) e (g) ilustram imagens segmentadas de LCM, LF e LLC, respectivamente, (b), (e) e (h) apresentam suas regiões ampliadas, (c), (f) e (i) apresentam tais regiões obtidas após esse processamento. . . . .	95





---

## **Lista de tabelas**

Tabela 1 – Estado da arte de estudos relacionados ao processamento de imagens histológicas coradas com H&E de LCM, LF e LLC. . . . .	53
Tabela 2 – Valores de entropia obtidos pela aplicação dos métodos de pré-processamento utilizados sobre cada canal dos modelos de cores <i>Red, Green and Blue</i> (RGB), HSV, LAB, LUV, YCbCr, YIQ e Hematoxilina. . . . .	87
Tabela 3 – Resultados quantitativos obtidos pela aplicação dos métodos de segmentação sobre a classe LCM. . . . .	101
Tabela 4 – Resultados quantitativos obtidos pelo emprego dos métodos de segmentação sobre a classe LF. . . . .	101
Tabela 5 – Resultados quantitativos obtidos pela aplicação dos métodos de segmentação sobre a classe LLC. . . . .	101
Tabela 6 – Resultados da classificação das imagens analisadas com e sem a aplicação do método proposto, e a partir da marcação do especialista, utilizando a métrica de acurácia. . . . .	103
Tabela 7 – Sensibilidades da classificação das imagens analisadas com e sem a aplicação do método proposto, além das características extraídas das regiões segmentadas manualmente. . . . .	103
Tabela 8 – Resultados da classificação das imagens analisadas com e sem a aplicação do método proposto, e aplicação da segmentação manual, por meio da métrica de especificidade. . . . .	104
Tabela 9 – Matrizes de confusão das classificações propostas utilizando o algoritmo de segmentação apresentado. . . . .	105
Tabela 10 – Matrizes de confusão das classificações propostas sem a aplicação do algoritmo de segmentação apresentado. . . . .	105



---

## **Lista de siglas**

**AG** Algoritmo Genético

**CAD** *Computer-Aided Diagnosis*

**CP** *Crossover Point*

**H&E** Hematoxilina-Eosina

**IHC** Imuno-histoquímica

**LH** Linfomas de Hodgkin

**LNH** Linfomas não-Hodgkin

**LCM** Linfoma de Células do Manto

**LF** Linfoma Folicular

**LLC** Leucemia Linfóide Crônica

**NE** Número de Estruturas

**OMS** Organização Mundial da Saúde

**RGB** *Red, Green and Blue*

**ROI** Regiões de Interesse



---

# **Sumário**

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>27</b>
1.1	Motivação . . . . .	29
1.2	Objetivos e Desafios da Pesquisa . . . . .	29
1.3	Hipótese . . . . .	30
1.4	Contribuições . . . . .	30
1.5	Organização da Dissertação . . . . .	31
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>33</b>
2.1	<b>Linfoma . . . . .</b>	<b>33</b>
2.1.1	Linfoma de Células do Manto . . . . .	34
2.1.2	Linfoma Folicular . . . . .	35
2.1.3	Leucemia Linfóide Crônica . . . . .	36
2.2	<b>Histologia . . . . .</b>	<b>37</b>
2.2.1	Preparação do Tecido . . . . .	38
2.2.2	Digitalização de Lâminas Histológicas . . . . .	39
2.3	<b>Processamento Digital de Imagens: Conceitos . . . . .</b>	<b>40</b>
2.3.1	Modelos de Cores . . . . .	40
2.3.2	Algoritmo Genético . . . . .	41
2.3.3	Operações Morfológicas . . . . .	43
<b>3</b>	<b>PROCESSAMENTO DE IMAGENS MÉDICAS: ESTADO DA ARTE . . . . .</b>	<b>47</b>
3.1	<b>Processamento de Imagens Histológicas . . . . .</b>	<b>47</b>
3.1.1	Pré-processamento . . . . .	47
3.1.2	Segmentação . . . . .	48
3.1.3	Pós-processamento . . . . .	50
3.1.4	Extração de Características . . . . .	51
3.1.5	Classificação . . . . .	52

3.2	Segmentação de Imagens Histológicas de Linfoma . . . . .	52
3.3	Considerações Finais . . . . .	63
4	<b>METODOLOGIA . . . . .</b>	67
4.1	Visão Geral do Método . . . . .	67
4.2	Banco de Imagens . . . . .	67
4.3	Pré-processamento . . . . .	68
4.4	Segmentação . . . . .	72
4.5	Pós-processamento . . . . .	79
4.6	Métricas de Avaliação Quantitativa . . . . .	82
5	<b>EXPERIMENTOS E ANÁLISE DOS RESULTADOS . . . . .</b>	85
5.1	Pré-processamento . . . . .	85
5.2	Segmentação . . . . .	89
5.3	Pós-processamento . . . . .	89
5.4	Análise Comparativa de Resultados . . . . .	92
5.5	Avaliação da Classificação de Linfomas . . . . .	103
6	<b>CONCLUSÃO . . . . .</b>	107
6.1	Principais Contribuições . . . . .	108
6.2	Trabalhos Futuros . . . . .	109
6.3	Contribuições em Produção Bibliográfica . . . . .	109
6.4	Agradecimentos . . . . .	110
	<b>REFERÊNCIAS . . . . .</b>	111

# CAPÍTULO 1

## Introdução

Segundo estimativas mundiais realizadas pelo projeto Globocan 2012, pela Agência Internacional para Pesquisa em Câncer e pela Organização Mundial da Saúde (OMS), 14,1 milhões de novos casos e 8,2 milhões de mortes foram registradas em 2012 decorrentes da incidência de câncer. Em 2030, estimam-se 21,4 milhões de novos casos e 13,2 milhões de mortes. A incidência de câncer no Brasil vem ganhando crescente relevância pelo seu perfil epidemiológico. No país, a previsão para o ano de 2014, também válida para 2015, foi de 576 mil novos casos. Esse cenário poderia ser revertido por meio de diagnósticos, tratamentos e acompanhamentos efetivos para evitar mortes prematuras (FACINA, 2014).

Os linfomas são neoplasias malignas que afetam o sistema linfático humano classificadas em pelo menos 38 subtipos, de acordo com a OMS (ORLOV et al., 2010). Esses subtipos são divididos em Linfomas de Hodgkin (LH) e Linfomas não-Hodgkin (LNH), diferenciando-se pelas suas combinações de fatores morfológicos, imunofenotípicos, genéticos e clínicos (MAURIÑO; SIQUEIRA, 2011). Em 2014, estimaram-se no Brasil, 2.180 e 9.790 novos casos de LH e LNH, respectivamente. Os registros mundiais de 2012 apontam que 199 mil óbitos decorrentes do LNH e 254 óbitos do LH ocorreram nesse ano. Em ambos os tipos, as mais altas taxas de incidência podem ser observadas na América do Norte, Austrália/ Nova Zelândia e em algumas partes da Europa. Uma das diferenças entre esses linfomas é a variação etária de suas incidências, onde o risco para desenvolvimento do LNH aumenta com a idade e a ocorrência do LH apresenta uma distribuição bimodal: com o primeiro pico etário entre 15 e 35 anos e o segundo acima dos 60 anos (FACINA, 2014).

Neste trabalho, serão abordados casos de Linfoma de Células do Manto (LCM), Linfoma Folicular (LF) e Leucemia Linfóide Crônica (LLC) que pertencem à classe LNH. A mesma corresponde a 70% dos casos de linfoma e devido a grande variedade de apresentações clínicas e características histopatológicas, sua classificação e segmentação ainda representam um grande desafio a patologistas (SANTOS; FERNANDES, 2008). O LCM é caracterizado por uma evolução clínica agressiva e um mau prognóstico com, praticamente, nenhum sobrevivente a longo prazo. Já o LF é o segundo mais comum entre

os casos de linfoma de células B na classificação definida pela OMS. A LLC representa uma manifestação diferente da mesma doença diagnosticada pelo Linfoma Linfocítico de Células Pequenas, por isso ambas são conduzidas da mesma maneira. Esses casos são os mais frequentes entre as leucemias nos países ocidentais, representando uma incidência de 30% (CANELLOS; LISTER; YOUNG, 2006).

Para realizar diagnósticos de linfoma, amostras de tecidos coradas com Hematoxilina-Eosina (H&E) são submetidas à análise de um especialista pelo microscópio para identificação de células cancerosas que indiquem a sua incidência. Essas distinções são essenciais ao permitirem o acompanhamento da doença, a identificação de seu estágio e orientação a tratamentos eficazes para o paciente (ORLOV et al., 2010). No entanto, a avaliação quantitativa visual é uma tarefa complexa devido ao alto consumo de tempo, a análise subjetiva do especialista, seu nível de experiência e a variabilidades inter e intrapatologistas. A classificação de graus histológicos de LF realizada por diferentes especialistas é um exemplo dessa variação, com taxas de divergência entre 61% e 73% (SERTEL et al., 2010b; OGER; BELHOMME; GURCAN, 2012). A Figura 1 ilustra essa inconsistência por meio da identificação de centroblastos realizada por cinco especialistas em imagens histológicas de LF coradas com H&E.

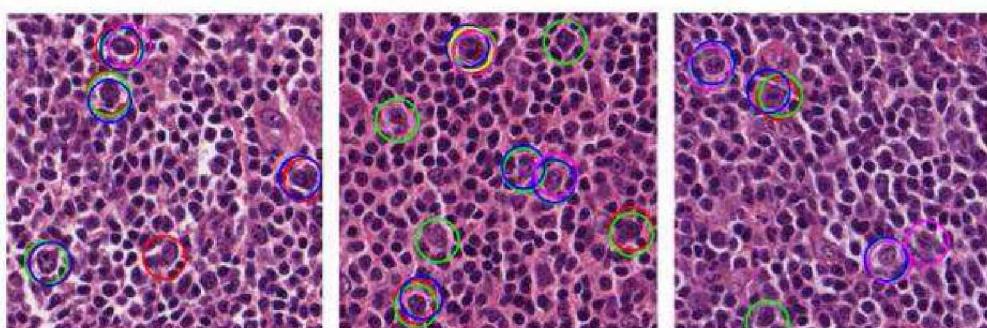


Figura 1 – Imagens histológicas de LF coradas com H&E com dimensões de  $512 \times 512$  pixels e magnificação  $40\times$ , contendo identificações de centroblastos realizadas por cinco especialistas indicadas por círculos identificados por diferentes cores (SERTEL et al., 2010b).

O recente desenvolvimento de tecnologias relacionadas à digitalização de imagens histológicas permite a aceleração de pesquisas em diagnósticos e prognósticos com o auxílio de técnicas computacionais, os chamados *Computer-Aided Diagnosis* (CAD). Esses sistemas têm o potencial para melhoramento da precisão de avaliações visuais quantitativas para auxílio a patologistas em suas decisões e, consequentemente, no aprimoramento de resultados clínicos (SERTEL et al., 2010a; BELKACEM-BOUSSAID et al., 2011). Métodos de visão computacional, como aqueles associados a processamento de imagens, possibilitam a utilização de técnicas capazes de relacionar as informações contidas em imagens a anormalidades características de cada neoplasia (ORLOV et al., 2010).

Em histopatologia, uma das tarefas essenciais é a identificação de estruturas. Esse processo é realizado por meio do emprego de técnicas de segmentação que são utilizadas para reconhecer a presença, distribuição, tamanho e características morfológicas de recursos suficientes para diagnósticos, como tumores, células específicas, núcleos e glândulas (HAGGERTY et al., 2014). Tal tarefa torna-se complexa devido à irregularidade visual dessas estruturas presentes em tecidos, principalmente as nucleares (IRSHAD et al., 2014). A identificação precisa de regiões de interesse é essencial para correlacioná-las a patologias utilizando análise de imagens (ONG et al., 1996). Para lidar com esse problema, um crescente número de trabalhos vêm sendo desenvolvidos possibilitando a integração entre técnicas computacionais e imagens histológicas.

## 1.1 Motivação

Um dos maiores desafios presentes em sistemas de análise de imagens histopatológicas é a segmentação de componentes celulares que possibilitem a separação de estruturas específicas de cada neoplasia (SERTEL et al., 2010b). Além disso, a grande variabilidade de características histológicas existentes em diferentes tipos de linfoma levam à necessidade do desenvolvimento de metodologias e algoritmos que possam contribuir no aprimoramento de etapas e/ou sistemas CAD capazes de identificar as neoplasias LCM, LF e LLC.

Diversos trabalhos são propostos na literatura para a identificação automática de estruturas contidas em imagens histológicas de linfoma. Os estudos de (OGER; BELHOMME; GURCAN, 2012), (SERTEL et al., 2008b) e (MOHAMMED et al., 2013a) exemplificam trabalhos que alcançaram alta acurácia na identificação de regiões foliculares, centroblastos e linfócitos, respectivamente, em imagens histopatológicas.

Diferentes regiões de interesse que contribuem para o diagnóstico das neoplasias tratadas neste trabalho devem ser identificadas. Porém, é necessário verificar o emprego de técnicas de processamento digital de imagens sobre imagens histológicas coradas com H&E obtidas sob diferentes condições de contraste, iluminação e magnificação. Assim, é necessário o desenvolvimento de um método de segmentação robusto a esses tipos de variações que seja capaz de identificar células neoplásicas dos linfomas abordados por este trabalho.

## 1.2 Objetivos e Desafios da Pesquisa

O objetivo geral deste trabalho é o desenvolvimento e análise de um método de segmentação não supervisionado para a identificação de estruturas relacionadas ao diagnóstico das neoplasias: LCM, LF e LLC. Para que esse objetivo pudesse ser alcançado, os procedimentos utilizados no trabalho proposto foram:

- Desenvolvimento de um método não supervisionado para a segmentação nuclear de células indicativas de LCM, LF e LLC subdividido nas etapas de pré-processamento, segmentação e pós-processamento;
- Avaliação do método de segmentação proposto utilizando imagens histológicas des-sas neoplasias, com magnificação 20×;
- Validação do método desenvolvido por meio de análises qualitativas e quantitativas, utilizando métricas obtidas pela comparação entre as regiões segmentadas automa-ticamente e aquelas identificadas pelo especialista;
- Comparação e análise dos resultados de forma qualitativa e quantitativa com méto-dos de segmentação propostos na literatura;
- Análise dos resultados obtidos na classificação entre as diferentes lesões considerando a aplicação do método proposto.

### **1.3 Hipótese**

Diferentes características de imagens histológicas, como suas condições de contraste, iluminação e magnificação, representam um empecilho ao desenvolvimento de técnicas ro-bustas e eficazes para a identificação de estruturas celulares representadas. Para que estru-turas nucleares de imagens histológicas de linfoma pudessem ser identificadas por meio da segmentação automática, uma técnica baseada em limiarização poderia ser desenvolvida e explorada combinando diferentes conceitos de processamento digital de imagens.

Ao propor tal técnica, características relacionadas a distribuição espacial e forma se-riam desconsideradas, e apenas características intrínsecas de núcleos celulares seriam ana-lisadas. Assim, viabiliza-se aperfeiçoar práticas tradicionais no processo de diagnóstico, tornando-o mais objetivo.

### **1.4 Contribuições**

Validada sobre imagens com grandes diferenças de contraste e iluminação, a técnica proposta mostrou-se robusta ao obter resultados relevantes para a segmentação de imagens histológicas. Ao combinar a aplicação de um algoritmo evolutivo e uma técnica amplamente utilizada para a segmentação de diferentes tipos de imagens, foi possível contribuir para o estado da arte do domínio de processamento de imagens médicas. Além disso, a escassa quantidade de trabalhos relacionados a segmentação de imagens histológicas de LCM e LLC torna a contribuição deste trabalho ainda maior.

## 1.5 Organização da Dissertação

Neste capítulo, foram apresentadas as considerações iniciais, motivação e objetivos para o desenvolvimento do método proposto. O restante deste trabalho possui a seguinte organização:

- **Capítulo 2 - Fundamentação Teórica:** apresenta tópicos relacionados a linfoma e histologia, e alguns conceitos de processamento digital de imagens necessários para compreensão deste trabalho;
- **Capítulo 3 - Processamento de Imagens Médicas: Estado da Arte:** apresenta a teoria e estado da arte do processamento digital de imagens médicas e métodos de segmentação aplicados sobre imagens histológicas de linfoma;
- **Capítulo 4 - Metodologia:** apresenta os materiais e métodos utilizados para o desenvolvimento deste projeto, sendo eles: descrição do banco de imagens histológicas utilizado, descrição das técnicas utilizadas nas etapas de pré-processamento, segmentação e pós-processamento, bem como as métricas empregadas para avaliação quantitativa do método proposto;
- **Capítulo 5 - Experimentos e Análise dos Resultados:** apresenta os resultados obtidos em cada etapa empregada, bem como os resultados quantitativos dos experimentos, comparação qualitativa e quantitativa de resultados obtidos pela aplicação de outros métodos de segmentação propostos na literatura, discussão desses resultados, e aplicação e discussão do emprego da segmentação para fins de classificação;
- **Capítulo 6 - Conclusão:** apresenta as conclusões do projeto, suas limitações e trabalhos futuros.



# CAPÍTULO 2

## Fundamentação Teórica

Neste capítulo, serão introduzidos conceitos teóricos necessários para a descrição desta proposta.

### 2.1 Linfoma

Linfoma é uma neoplasia que se desenvolve nos componentes celulares chamados linfócitos (ORLOV et al., 2010). Essas células representam a segunda maior população de glóbulos brancos responsáveis pela defesa imunológica do organismo (GARTNER; HIATT, 2003). Originados na medula óssea, os linfócitos subdividem-se em três categorias funcionais: linfócitos B, linfócitos T e células nulas. Os linfócitos T completam sua maturação no timo, enquanto os linfócitos B saem da medula já como células maduras. Levadas pelo sangue e pela linfa, essas células migram desses órgãos para os órgãos linfáticos periféricos (baço, linfonodos, nódulos linfáticos isolados, tonsilas, apêndice e placas de Payer do íleo), onde se proliferam e completam a sua diferenciação (JUNQUEIRA; CARNEIRO, 2004).

Os casos de linfoma são originários dos gânglios, ou linfonodos, do sistema linfático, localizados em várias regiões do corpo, porém são mais frequentes no pescoço, axila, virilha e ao longo dos grandes vasos. Esses órgãos atuam como filtros para a remoção de bactérias e outras substâncias estranhas que chegam até eles por meio da linfa, um fluido linfático composto por glóbulos brancos, oxigênio, proteínas e glicose (GARTNER; HIATT, 2003; JUNQUEIRA; CARNEIRO, 2004).

Os diferentes tipos de linfoma são geralmente diferenciados por meio de seus padrões de crescimento e suas características citológicas. Essas neoplasias podem se desenvolver tanto em células B quanto em células T, porém 85% de seus casos são derivados de linfócitos B (ORLOV et al., 2010). Os processos de desenvolvimento, ativação e diferenciação de células B normais podem ser divididas em 3 etapas: centro pré-germinativo, centro germinativo e centro pós-germinativo. Durante a etapa pré-germinativa, o linfócito B encontra-se maduro contendo proteínas componentes de anticorpos, as imunoglobulinas. No centro germinativo, os linfócitos B começam a produzir anticorpos específicos

para antígenos para os quais foram expostos. E na etapa pós-germinativa, as células B tornam-se células de memória. A Figura 2 representa a relação entre essas etapas características do desenvolvimento dos linfócitos B e alguns linfomas relacionados a cada uma delas (GERARD; BISHOP, 2012).

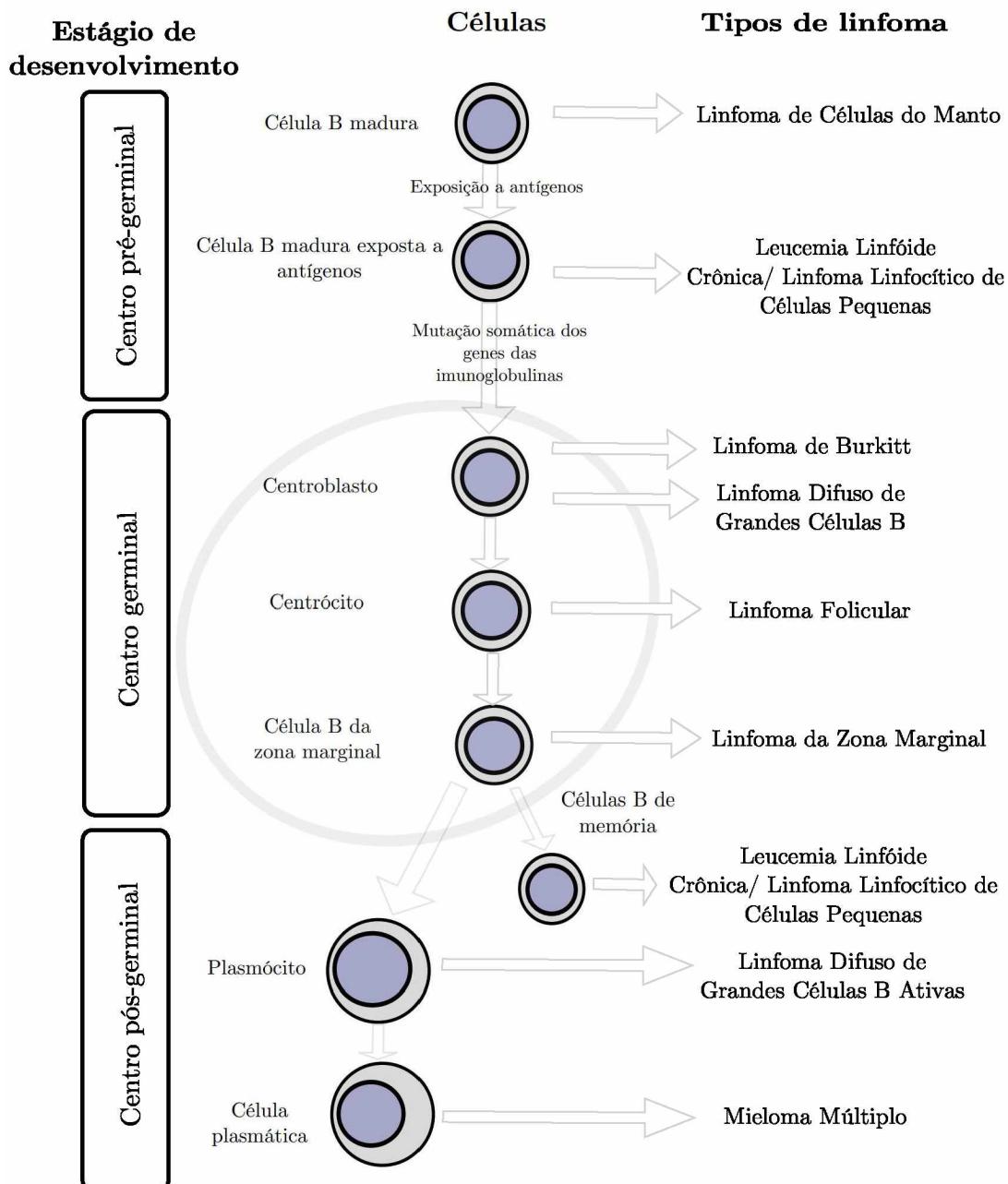


Figura 2 – Relação entre as etapas de desenvolvimento, ativação e diferenciação de células B e os linfomas associados a cada uma delas (GERARD; BISHOP, 2012).

### 2.1.1 Linfoma de Células do Manto

O LCM corresponde a 3 a 10% dos casos de LNH, acometendo indivíduos do sexo masculino de meia-idade ou idosos (MAURIÑO; SIQUEIRA, 2011). Sua incidência é de

aproximadamente 2 a 3 casos a cada 100.000 por ano entre as ocorrências de linfoma na América do Norte e na Europa. Assim como em outros tipos de linfoma, a causa e os fatores condicionais e determinantes desse câncer são pouco conhecidos (CANELLOS; LISTER; YOUNG, 2006).

Essa neoplasia é caracterizada por células clivadas atípicas de tamanhos pequeno e médio que podem ser descritas por uma proliferação vagamente nodular, difusa ou com o alargamento da zona do manto, sendo esta uma região de concentração de linfócitos B que circunda os centros germinativos (GARTNER; HIATT, 2003). Além disso, as células são geralmente de tamanhos comparáveis e compartilham características citológicas similares: tamanho ligeiramente maior que linfócitos normais com cromatinas aglutinadas, pouco citoplasma, nucléolos imperceptíveis e contorno nuclear geralmente irregular e clivado. Células peculiares dessa doença que se assemelham a centroblastos estão essencialmente ausentes, proporcionando uma distinção importante do LF (CANELLOS; LISTER; YOUNG, 2006; MAURIÑO; SIQUEIRA, 2011). A Figura 3 representa uma lâmina de LCM corada com H&E com células ligeiramente irregulares e monomórficas.

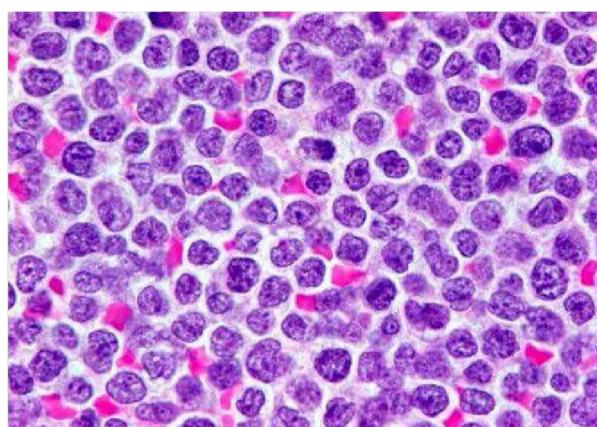


Figura 3 – Lâmina de LCM com células “regularmente irregulares” (PATHPEDIA, 2014).

## 2.1.2 Linfoma Folicular

O LF corresponde a 22% dos casos de LNH (MAURIÑO; SIQUEIRA, 2011). O tipo indolente dessa neoplasia prevalece ocorrendo com frequência de aproximadamente 2/100.000 casos nos países ocidentais (CANELLOS; LISTER; YOUNG, 2006). Sua incidência varia de acordo com as áreas geográficas, onde é menos frequente na Ásia e em países em desenvolvimento que na Europa Ocidental e nos Estados Unidos (CANELLOS; LISTER; YOUNG, 2006). É predominante ainda a ocorrência em indivíduos adultos do sexo masculino com idade mediana de 59 anos e com raros diagnósticos em jovens com menos de 20 anos (MAURIÑO; SIQUEIRA, 2011).

Essa neoplasia é caracterizada por um conjunto de células que revestem as cavidades dos linfonodos chamadas centros foliculares, com um padrão de crescimento folicular par-

cial (CANELLOS; LISTER; YOUNG, 2006). Alguns tipos celulares principais, também presentes em centros foliculares normais, podem ser notados nessas regiões patológicas: os centrócitos e centroblastos. Os centrócitos são caracterizados pelo seu pequeno tamanho, com núcleo clivado, nucléolo pouco visível e citoplasma pálido, e os centroblastos correspondem a células grandes, com formato oval ou arredondado e núcleo caracterizado por 1 a 3 nucléolos. Os folículos neoplásicos são mal definidos e apresentam baixa concentração de linfócitos B (MAURIÑO; SIQUEIRA, 2011). A Figura 4 ilustra a presença de centrócitos e centroblastos em uma região folicular de uma amostra de tecido corada com H&E.

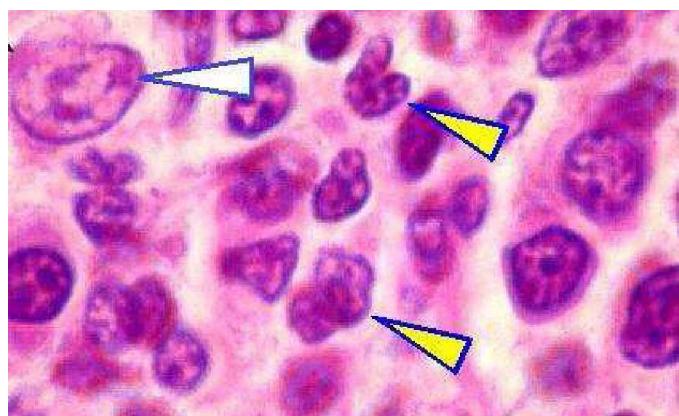


Figura 4 – Lâmina contendo centroblastos, com um deles indicado por uma seta branca, e centrócitos indicados por setas amarelas (UNICAMP, 2013).

Com base no número absoluto de centroblastos no folículo neoplásico, 3 tipos histológicos são obtidos (MAURIÑO; SIQUEIRA, 2011):

- Grau I: de 0 a 5 centroblastos;
- Grau II: de 6 a 15 centroblastos;
- Grau III: mais de 15 centroblastos.

Na nova classificação, o grau III ainda pode ser subdividido em III-A quando centrócitos estão presentes e III-B quando centroblastos formam agrupamentos (MAURIÑO; SIQUEIRA, 2011).

### 2.1.3 Leucemia Linfóide Crônica

A LLC possui incidência mundial de aproximadamente 5 casos a cada 100.000 por ano, aumentando de acordo com a idade (CANELLOS; LISTER; YOUNG, 2006). A média de idade de pacientes diagnosticados é de cerca de 70 anos, com raras ocorrências em pessoas com menos de 40 anos. Essa leucemia é mais frequente no sexo masculino com uma razão de 2:1 e um dos mais importantes fatores de risco associados a ela é o

histórico familiar. Aproximadamente 20% dos pacientes com a doença têm familiares em primeiro grau diagnosticados com LLC e apresentam um risco três vezes maior de desenvolvê-la (BUCCHERI; MAURIÑO, 2011).

Essa é uma doença linfoproliferativa resultante da propagação de células linfóides presas na fase madura do seu processo de diferenciação mas imunologicamente incompetentes (CANELLAS; LISTER; YOUNG, 2006; BUCCHERI; MAURIÑO, 2011). Na maioria dos casos, é caracterizada por linfócitos pequenos com núcleo regular e cromatina condensada, sem nucléolo evidente, citoplasma escasso e ligeiramente basófilo e sem granulações específicas. Além disso, o número de prolinfócitos caracterizados por seu tamanho médio com citoplasma mais abundante e nucléolo evidente é menor do que 10% (BUCCHERI; MAURIÑO, 2011). Histologicamente, centros germinativos podem estar presentes em alguns casos, o que os assemelham ao LCM. Porém, esses casos são distinguidos em níveis celulares. Enquanto as células predominantes são pequenos linfócitos com cromatinas nucleares aglutinadas, é possível observar variações em sua morfologia nuclear (CANELLAS; LISTER; YOUNG, 2006). A Figura 5 ilustra uma amostra de tecido de LLC corada com H&E, onde a infiltração neoplásica apresenta aspecto difuso e homogêneo ao corte.

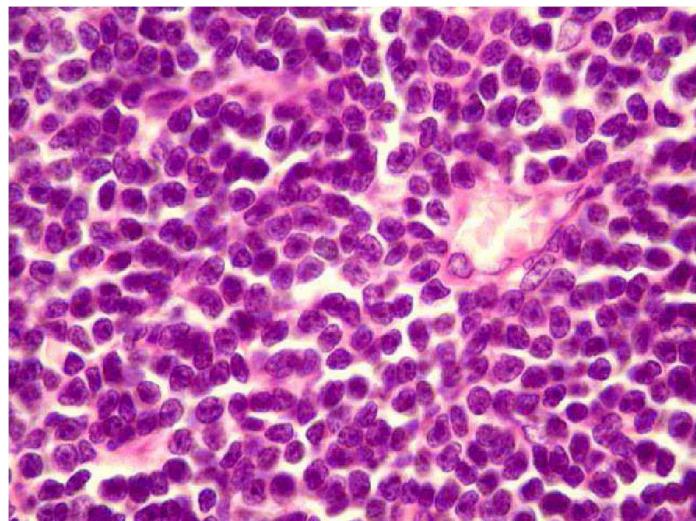


Figura 5 – Lâmina de LLC caracterizada por linfócitos tumorais morfologicamente semelhantes aos normais (UNICAMP, 2013).

## 2.2 Histologia

Histologia é o ramo da anatomia responsável pelo estudo dos tecidos e de como estes se organizam para a formação dos órgãos constituintes do corpo (GARTNER; HIATT, 2003; JUNQUEIRA; CARNEIRO, 2004). Porém, a histologia não mais se limita ao estudo dessas estruturas, abrangendo também sua relação direta com outras áreas, como a patológica (GARTNER; HIATT, 2003).

### 2.2.1 Preparação do Tecido

Diversas técnicas foram desenvolvidas para a preparação de tecidos permitindo a análise dessas amostras com aspecto muito próximo ao seu estado natural vivo (GARTNER; HIATT, 2003). O procedimento mais utilizado em histologia é dependente do uso de microscópios, devido ao pequeno tamanho das células e de seus componentes (JUNQUEIRA; CARNEIRO, 2004). Além disso, para preservar a relação estrutural entre as células, torna-se necessária a utilização de técnicas de cortes de tecidos em fatias muito finas para que os mesmos estejam adequados para análise (CORMACK; HAM, 1983).

As etapas presentes na preparação de tecidos envolvem a manipulação de uma pequena amostra obtida por meio de biópsia, ou de outras técnicas. Esse tecido deve ser retirado de forma minuciosa para que não haja a distorção de sua aparência microscópica, e então ser submetido aos procedimentos de (1) fixação, (2) desidratação e diafanização, (3) inclusão, (4) microtomia e (5) montagem e coloração. Essas etapas são descritas a seguir (CORMACK; HAM, 1983; GARTNER; HIATT, 2003):

1. **Fixação:** essa etapa refere-se à desaceleração de alterações *post mortem* do tecido (ou após a retirada do corpo) e à manutenção de sua arquitetura normal por meio de agentes químicos. A fixação imediata da amostra é essencial para retardar sua degeneração e, possivelmente, permitir a intensificação da coloração subsequente do tecido;
2. **Desidratação e Diafanização:** para que o bloco de tecido possa ser seccionado com facilidade, a parafina passa a substituir a água presente na amostra. Como grande parte do tecido é constituído por água, banhos de álcool são utilizados para a sua desidratação. Entretanto, o álcool não é capaz de reagir como um solvente da parafina, o que torna necessária a diafanização. Essa etapa permite a substituição do álcool pelo xilol, um dos solventes da parafina, por meio de sucessivas trocas desse material até que o tecido esteja preparado para a inclusão;
3. **Inclusão:** nessa etapa, o tecido passa por várias trocas de parafina quente que se dissolve no xilol resultante da etapa anterior. Uma vez impregnada com parafina, a amostra é deixada para endurecer, formando um bloco contendo o tecido;
4. **Microtomia:** após a remoção do excesso de parafina, os cortes no bloco podem ser realizados com o auxílio de um micrótomo. Esse equipamento possui uma navalha e um braço que permitem a secção do bloco de forma progressiva com espessuras definidas e idênticas;
5. **Montagem e Coloração:** os cortes obtidos na microtomia são colocados em lâminas de vidro cobertas por adesivos, compondo a etapa de montagem do tecido. A coloração é realizada sobre soluções aquosas, o que torna necessária a substituição

da parafina por água. Após a reidratação da amostra, o tingimento é realizado para permitir a distinção da densidade óptica dos componentes do tecido. Depois de corado, uma cobertura de proteção chamada lamínula é colocada sobre o corte para sua visualização no microscópio.

Os corantes mais comumente utilizados em histologia são o H&E (GARTNER; HIATT, 2003; JUNQUEIRA; CARNEIRO, 2004). A hematoxilina cora em tons de azul ou violeta estruturas ácidas das células, como o núcleo e porções do citoplasma ricas em ácido ribonucléico (RNA) (JUNQUEIRA; CARNEIRO, 2004). Já a eosina cora os componentes básicos celulares em tons róseos. Como muitos componentes citoplasmáticos possuem pH básico, várias regiões do citoplasma apresentam-se sob a coloração rosa da eosina (GARTNER; HIATT, 2003).

## 2.2.2 Digitalização de Lâminas Histológicas

A digitalização de lâminas histológicas é possível por meio de *scanners* digitais ou microscópios equipados com câmeras. A aquisição de imagens patológicas permite suas análises utilizando algoritmos para seu processamento (VETA et al., 2014; BELSARE; MUSHRIF, 2012; KARSNAS, 2014).

Essa técnica substitui vantajosamente as lâminas físicas para fins educacionais, consultoria a distância e, principalmente, para o desenvolvimento de sistemas computacionais baseados nessas imagens (ROCHA et al., 2009). As Figuras 6(a), 6(b) e 6(c) apresentam imagens histológicas de LCM, LF e LLC, respectivamente, com coloração H&E.

O desenvolvimento de algoritmos capazes de processar imagens digitais histopatológicas tornou-se importante ao permitir a obtenção de diagnósticos e prognósticos para auxílio a especialistas (BELSARE; MUSHRIF, 2012). Essa abordagem computacional possibilita o alcance de rápidos resultados automáticos por meio do emprego de técnicas relacionadas ao processamento digital de imagens aliadas ao conhecimento histológico (KARSNAS, 2014).

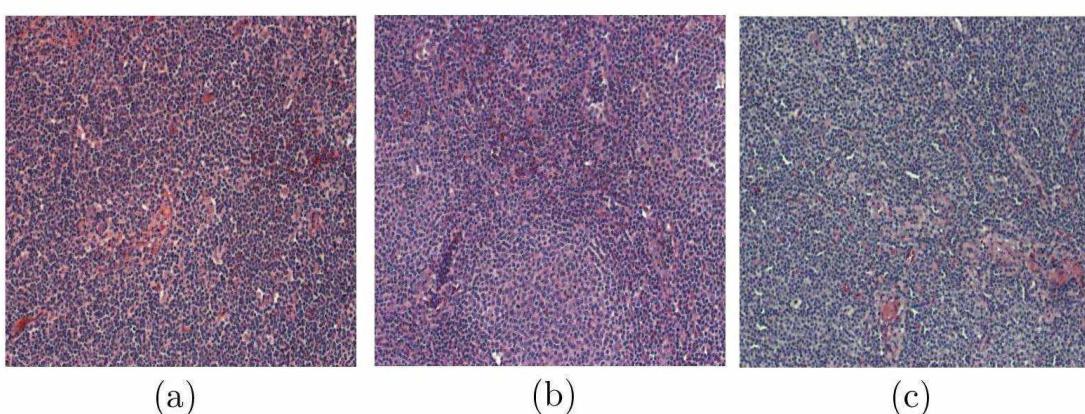


Figura 6 – Exemplos de digitalizações de lâminas de LCM (a), LF (b) e LLC (c).

## 2.3 Processamento Digital de Imagens: Conceitos

Para melhor compreensão deste trabalho, nesta seção, serão apresentados detalhes das técnicas e conceitos utilizados pelo método proposto.

### 2.3.1 Modelos de Cores

Diferentes modelos de cores podem ser empregados em processamento de imagens. Neste trabalho, foram utilizados os modelos de cores e a técnica de deconvolução descritos a seguir:

- **RGB:** o espaço de cores RGB é representado por um sistema de coordenadas cartesianas tridimensional, onde cada uma das coordenadas é composta pela intensidade de vermelho  $R$ , verde  $G$  e azul  $B$  de cada pixel. O processo de formação de cores desse modelo é aditivo, em que a adição de cada um de seus componentes formam novas cores (GANESAN; RAJINI, 2014; PAKHIRA, 2011);
- **HSV:** esse modelo de cores é composto por três componentes: matiz  $H$ , saturação  $S$  e valor  $V$ . A separação da saturação e da luminância desse modelo é muito útil para processamento de imagens (DUAN; YU, 2011; PAKHIRA, 2011);
- **CIE Lab Color (LAB):** é um modelo tridimensional que possui um componente de luminosidade  $L$  e dois componentes de crominância,  $a$  e  $b$ . O componente  $a$  é representado por um eixo que se estende de verde a vermelho enquanto o  $b$  vai de azul a amarelo. Uma característica desse modelo é que ele representa o espectro visível pelo olho humano de maneira uniforme (PAKHIRA, 2011);
- **CIE Luv Color (LUV):** esse modelo é formado por três componentes:  $L$  correspondente à luminância, e  $u$  e  $v$ , que juntos correspondem à crominância. Esse modelo é caracterizado por uma propriedade útil na percepção de cores: imagens com mesmo canal  $L$  são isoluminantes (HEALEY, 1996);
- **YCbCr:** o valor de  $Y$  representa a luminância, utilizado para representar cores RGB em monitores monocromáticos. Os valores  $Cb$  e  $Cr$  são chamados valores de crominância e representam duas coordenadas num sistema que mede a nuance e a saturação das cores. Esses valores quantificam as porções de azul e vermelho em uma cor (PAKHIRA, 2011);
- **YIQ:** o modelo YIQ é resultado de uma combinação linear dos componentes do RGB. Entretanto seus componentes são constituídos por luminância  $Y$ , matiz  $I$  e saturação  $Q$ . O canal  $I$  contém informação da matiz de laranja e ciano e  $Q$  de verde e magenta. Diferente do RGB, seus componentes de crominância e luminância são

separados, justificando seu uso em processamento de imagens (GANESAN; RAJINI, 2014; PAKHIRA, 2011);

- **Deconvolução:** essa técnica permite a separação dos componentes do corante utilizado para coloração das amostras de tecido, de acordo com a densidade óptica proporcional à concentração de cada um destes a uma determinada estrutura celular (JUNQUEIRA; CARNEIRO, 2004). Essas densidades expressam o fator de absorção específico de cada corante, assim, diferentes corantes possuem diferentes densidades ópticas. O método de deconvolução possibilita quantificar a contribuição individual de cada componente do corante utilizado sobre os canais *R*, *G* e *B*. Por meio de transformações ortonormais, essa técnica permite determinar densidades de diferentes componentes mesmo em áreas onde múltiplos constituintes do corante fixaram-se (RUIFROK; JOHNSTON, 2001; TOSTA et al., 2015).

### 2.3.2 Algoritmo Genético

Algoritmo Genético (AG) é uma das ferramentas mais promissoras na busca por soluções otimizadas pela exploração de informações num espaço vasto e desconhecido de forma eficiente (ABBASGHOLIPOUR et al., 2011). Assim, sua aplicação no processamento de imagens tornou-se mais frequente. Sua metodologia baseia-se no princípio da seleção natural, introduzido por Charles Darwin, selecionando os “indivíduos mais aptos” para a alcançar a solução ótima do problema a ser tratado (PAULINAS; UŠINSKAS, 2007).

Considerados computacionalmente simples e de rápido processamento (HARVEY; MARSHALL, 1995; RIBEIRO et al., 2011), esses algoritmos são compostos por uma população formada por indivíduos que correspondem a possíveis soluções. Tais indivíduos são formados por atributos que representam parâmetros de suas soluções. A geração de novas populações é realizada por meio dos processos de seleção, *crossover* e mutação, e pelo emprego de uma função de avaliação. A Figura 7 ilustra a sequência de etapas empregadas por um AG (PAULINAS; UŠINSKAS, 2007).

Após a geração da população inicial, uma função de avaliação é utilizada e calculada a partir dos atributos de cada indivíduo para definir o quão boa é a solução proposta por cada um deles (SHETA; BRAIK; ALJAHDALI, 2012). Assim como o processo de evolução, o AG seleciona as adaptações benéficas da população e as incorpora às gerações posteriores.

Durante o processo de formação de novas gerações, operações genéticas são realizadas, sendo elas: seleção, *crossover* e mutação. A etapa de seleção é empregada para selecionar indivíduos considerados como soluções mais adequadas para o problema analisado, conforme a função de avaliação. Esses indivíduos são mantidos na próxima geração e a partir

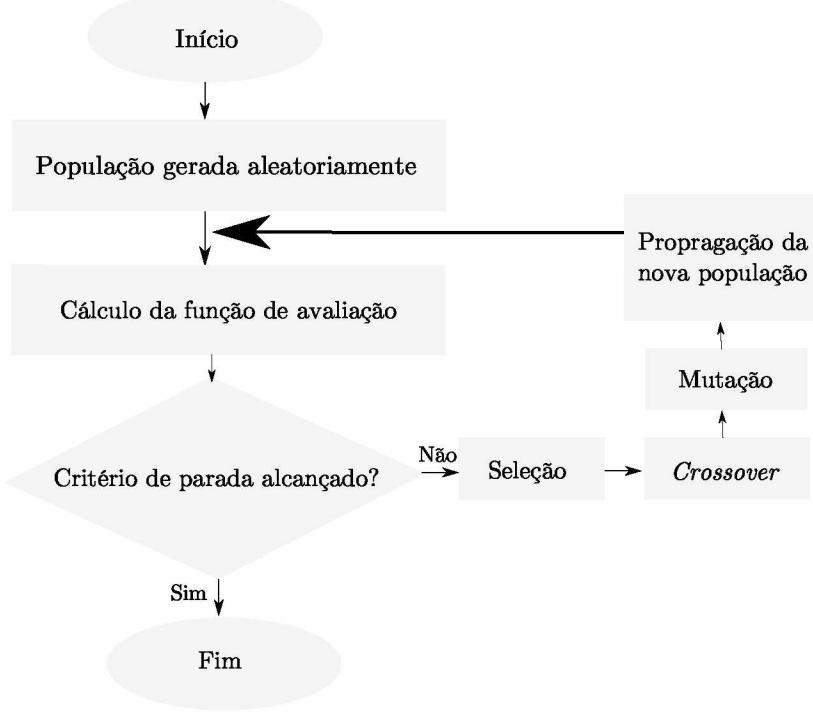


Figura 7 – Fluxograma do algoritmo genético (KAUSHIK et al., 2013).

deles novos indivíduos serão gerados pelas etapas subsequentes (PAULINAS; UŠINSKAS, 2007).

Com a obtenção dos melhores indivíduos na seleção, a etapa de *crossover* é aplicada para a geração da diversidade populacional. Nessa etapa, indivíduos escolhidos serão combinados para gerar novos indivíduos até que toda a população seja novamente preenchida. Para que essa combinação de indivíduos seja realizada, deve-se definir a probabilidade de *crossover* e o *Crossover Point* (CP) (JIANLI; BAOQI, 2009). A probabilidade de *crossover*, comumente contida no intervalo de 0,5 à 1,0, definirá um limiar para que um indivíduo seja escolhido para reprodução. Em seguida, um CP deve ser escolhido para determinar a combinação desses indivíduos, definindo quais atributos serão “herdados” por cada descendente a ser gerado (JIANLI; BAOQI, 2009; LIN; LEE; HONG, 2003). A Figura 8 ilustra um exemplo do processo de *crossover*, em que dois indivíduos foram utilizados para a geração de dois descendentes baseado na atribuição de CP ao valor 2.

A mutação também é responsável pelo aumento da diversidade populacional durante a execução do AG, o que leva à alteração dos valores dos atributos dos indivíduos. Tal diversidade ocorre de forma efetiva a partir da definição de uma probabilidade de mutação (ZHANG et al., 2014). Essa probabilidade define um limiar para que a mutação ocorra em determinado parâmetro de um indivíduo (MCCALL, 2005).

Para que o algoritmo seja finalizado, é definido o critério de parada que objetiva alcançar uma solução ótima, ou próxima disso. Para tanto, diferentes abordagens podem ser empregadas, como a definição do limite de gerações a serem obtidas na execução do

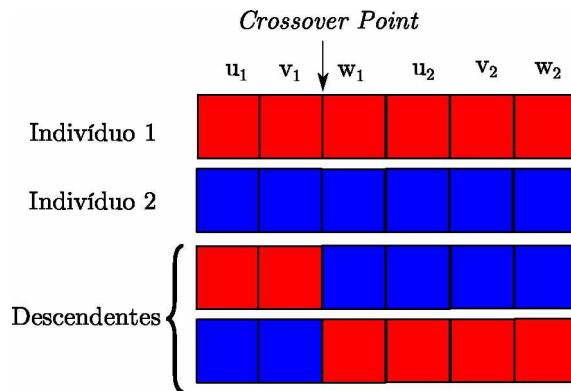


Figura 8 – Exemplo de um processo de *crossover* executado sobre dois indivíduos para a geração de seus descendentes.

AG (RIBEIRO et al., 2011; GUDMUNDSSON; KABUKA, 1998).

### 2.3.3 Operações Morfológicas

Operações morfológicas apresentam-se como uma poderosa abordagem para diferentes aplicações em processamento de imagens, permitindo a construção de operadores úteis para a descrição de objetos. No contexto de tais operações, a teoria dos conjuntos é utilizada para representar objetos em uma imagem. Assim, essas operações são aplicáveis em imagens binárias, compostas apenas por pixels pretos e brancos. Um operador morfológico é um mapeamento entre o conjunto A que define a imagem e um conjunto B, chamado elemento estruturante. O elemento estruturante corresponde a um conjunto binário com definição prévia de seu tamanho, distribuição de seus elementos e sua origem. A aplicação do elemento estruturante sobre uma imagem binária baseia-se nos deslocamentos da origem desse elemento sobreposto aos elementos não nulos da imagem (GONZALEZ; WOODS, 2000; PEDRINI; SCHWARTZ, 2007).

As operações de dilatação e erosão são as operações fundamentais em morfologia matemática. A operação de dilatação entre o conjunto A e o elemento estruturante B é definida como a adição de Minkowski, matematicamente expressa por (GONZALEZ; WOODS, 2000; PEDRINI; SCHWARTZ, 2007):

$$A \oplus B = x | (\hat{B}_x) \cap A \neq 0, \quad (1)$$

onde,  $\hat{B}$  representa a reflexão de B. A dilatação consiste na união do elemento estruturante à imagem. Ao se deslocar pela imagem, verifica-se a sobreposição do pixel de origem do elemento estruturante sobre um pixel não nulo. Caso isso ocorra, a imagem é preenchida pontualmente pela distribuição do elemento estruturante. Caso contrário, a imagem resultante não é alterada (GONZALEZ; WOODS, 2000). A Figura 9 ilustra a execução dessa operação com um elemento estruturante de tamanho  $3 \times 3$  pixels com

distribuição em forma de cruz, onde as regiões superior, inferior, esquerda e direita do pixel central são preenchidas conforme ilustrado pela Figura 9(c).

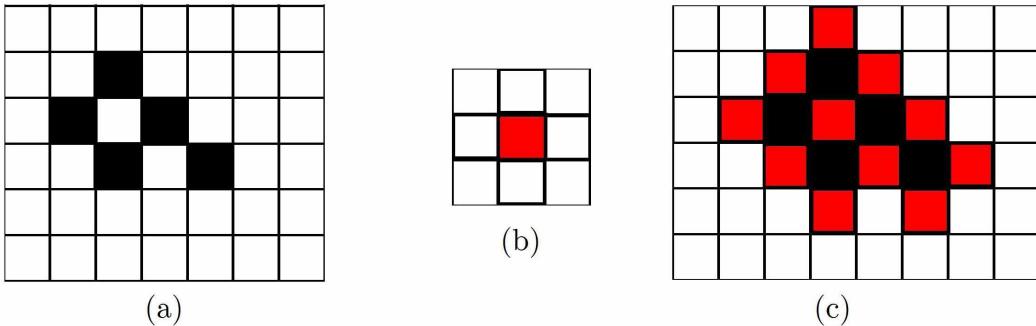


Figura 9 – Exemplo da operação morfológica de dilatação: (a) imagem original, (b) elemento estruturante de tamanho  $3 \times 3$  em forma de cruz e (c) resultado da dilatação.

A operação de erosão entre o conjunto A e o elemento estruturante B é definida como a subtração de Minkowski, matematicamente expressa por (GONZALEZ; WOODS, 2000; PEDRINI; SCHWARTZ, 2007):

$$A \ominus B = x | (B_x) \cap A_c = \emptyset, \quad (2)$$

onde,  $A_c$  representa o conjunto complemento de A. A operação de erosão é realizada considerando a translação do elemento estruturante sobre a imagem. Assim, ao se deslocar ao longo de elementos não nulos, deve-se verificar se todo o elemento estruturante está contido na região do objeto. Se sim, elementos do objeto que coincidam com a distribuição do elemento estruturante são apagados e apenas a origem desse elemento é mantida no resultado. Caso contrário, todos os elementos da imagem são apagados. A Figura 10 ilustra a execução desse processo sobre o objeto representado pela Figura 10(a) com um elemento estruturante de tamanho  $3 \times 3$  pixels (Figura 10(b)) com distribuição em forma de cruz. O resultado da erosão é ilustrado pelo objeto em vermelho contido na Figura 10(c).

Outras operações importantes em análise de imagens são a abertura e o fechamento, definidos, respectivamente, por (GONZALEZ; WOODS, 2000; PEDRINI; SCHWARTZ, 2007):

$$A \circ B = (A \ominus B) \oplus B, \quad (3)$$

$$A \bullet B = (A \oplus B) \ominus B, \quad (4)$$

onde,  $\ominus$  e  $\oplus$  correspondem, respectivamente, às operações de erosão e dilatação. Assim, a abertura é definida como a união de todas as translações de B contidas na imagem

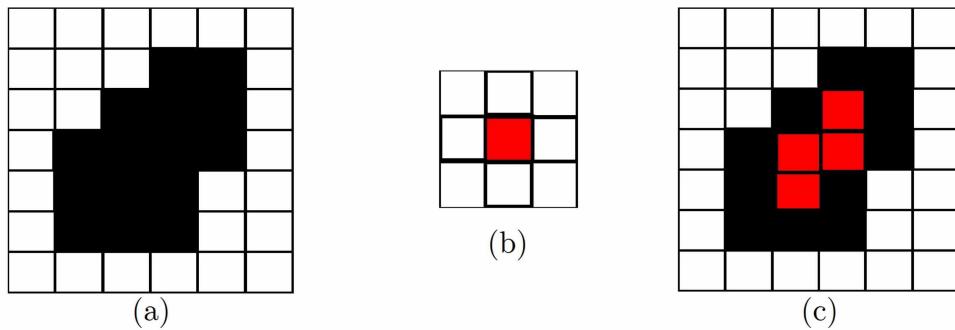


Figura 10 – Exemplo da operação de erosão: imagem original (a) processada por um elemento estruturante de tamanho  $3 \times 3$  em forma de cruz (b) e seu resultado (c).

A, e o fechamento pode ser interpretado geometricamente como a união de todas as translações de B que não estão contidas em A (PEDRINI; SCHWARTZ, 2007). As Figuras 11 e 12 ilustram as operações de abertura e fechamento, respectivamente, por meio de um elemento estruturante de tamanho  $3 \times 3$  com distribuição em forma de cruz.

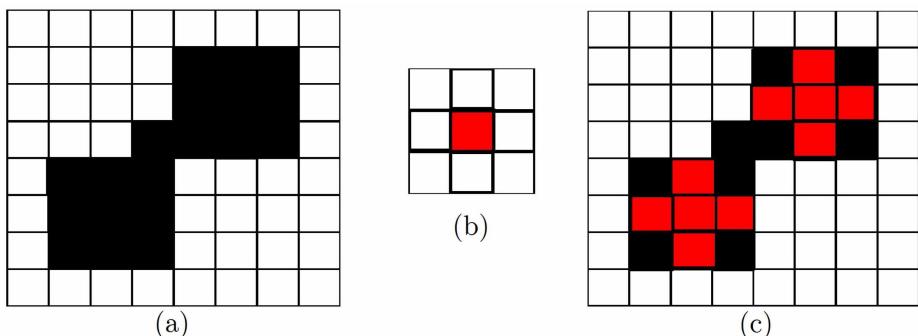


Figura 11 – Exemplo da operação de abertura binária: imagem original (a) e o resultado de sua abertura (c) por um elemento estruturante de tamanho  $3 \times 3$  em forma de cruz (b).

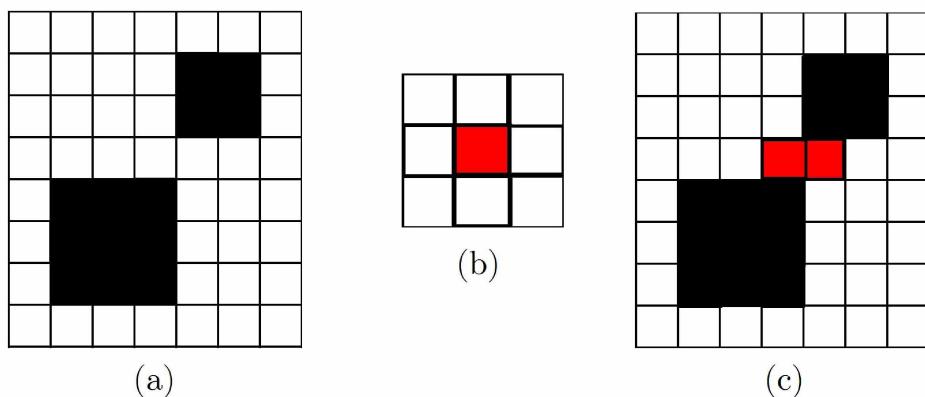


Figura 12 – Exemplo da operação de fechamento binária: (a) imagem original, (b) elemento estruturante de tamanho  $3 \times 3$  em forma de cruz utilizado para seu processamento e (c) resultado de seu fechamento.



# CAPÍTULO 3

## Processamento de Imagens Médicas: Estado da Arte

Neste capítulo, serão descritos estudos da literatura relacionados, direta e indiretamente, ao trabalho proposto.

### 3.1 Processamento de Imagens Histológicas

A análise de imagens histológicas representa uma das maiores evoluções da medicina moderna. Essa técnica possibilitou aos especialistas um avanço na análise e processamento de informações contidas em imagens. Aliados a essa evolução, métodos computacionais vêm sendo largamente desenvolvidos para melhoramento dessas análises, desde a identificação de Regiões de Interesse (ROI) por meio da segmentação até a classificação de neoplasias (IRSHAD et al., 2014).

O estudo de processamento de imagens aplicado à histologia permite aperfeiçoar certas práticas tradicionais, levando a maior objetividade e menor variabilidade no processo de diagnóstico (HAGGERTY et al., 2014). Para que essas técnicas sejam empregadas, torna-se necessária a definição de etapas para sua aplicação (OLIVEIRA et al., 2013). Nesse contexto, as principais etapas para análise e quantificação de informações são: pré-processamento, segmentação, pós-processamento, extração de características e classificação.

#### 3.1.1 Pré-processamento

A etapa de pré-processamento é necessária na análise de imagens médicas para remoção de representações irrelevantes, aprimoramento de contraste e eliminação de ruídos, entre outros aperfeiçoamentos. Basicamente, seu objetivo é melhorar a qualidade da imagem para o processo de segmentação, a ser realizado posteriormente.

No contexto de imagens histológicas, diversas técnicas são utilizadas nessa fase, como o filtro Gaussiano para a suavização da imagem (IRSHAD et al., 2014; HE et al., 2010; CARRERAS; GONZALEZ; SOLORZANO, 2004). Além deste, também são utilizadas: normalização de cores (IRSHAD et al., 2014; HAGGERTY et al., 2014; TADROUS, 2010) para promover a classificação por meio dessa característica (KOTHARI et al., 2011), filtro de média para remoção de pequenos ruídos (KRISHNAN et al., 2010), filtro de suavização aliado ao filtro de Sobel para realce de contornos (CARRERAS; GONZALEZ; SOLORZANO, 2004) e correções de iluminação (IRSHAD et al., 2014).

Entre as técnicas de pré-processamento aplicadas a imagens histológicas de linfoma estão: a seleção de canais de modelos de cores, para detecção de estruturas nucleares, e conversão desses modelos, para melhor representação do conteúdo da imagem. Diversos estudos exploram diferentes modelos de cores para a etapa de pré-processamento em imagens de linfoma. Dentre eles, destacam-se os trabalhos de (ZORMAN et al., 2007) e (NEUMAN et al., 2010) que utilizam o modelo de cores RGB para a seleção de um canal que seja capaz de representar as informações mais relevantes para etapas posteriores. Com a mesma finalidade, ainda existem estudos que buscam a conversão entre modelos. Nesse caso, o trabalho de (SERTEL et al., 2010a) converte o modelo de cores RGB para LUV de forma a permitir maior uniformidade entre as cores presentes na imagem. O estudo de (HAGGERTY et al., 2014) também utiliza esse mecanismo de conversão. Nesse caso, o modelo RGB é convertido para LAB, e seleciona-se qual deles apresenta as informações com maior contraste entre os objetos da imagem para aplicação da etapa de segmentação.

### **3.1.2 Segmentação**

O processo de segmentação é utilizado para permitir o isolamento de objetos de interesse que levam a uma análise precisa de estruturas específicas (GONZALEZ; WOODS, 2000). Esses objetos podem ser utilizados para detecção e classificação de cânceres (IRSHAD et al., 2014). A segmentação é uma das tarefas mais complexas e essenciais em processamento digital de imagens pois, a partir dela, pode-se garantir o sucesso em análise de imagens (GONZALEZ; WOODS, 2000).

A segmentação é um processo capaz de subdividir a imagem em suas partes constituintes, de acordo com o contexto do problema a ser resolvido. O resultado desse processo é o isolamento total dos objetos de interesse do restante da imagem por meio do emprego de técnicas particulares para cada situação específica (GONZALEZ; WOODS, 2000). Imagens são segmentadas com base em regiões semelhantes pela avaliação de critérios de homogeneidade, como cor e textura (KANDWAL; KUMAR; BHARGAVA, 2014).

Essa técnica é dividida em duas categorias: baseadas em descontinuidade e em similaridades. Na primeira, a segmentação pode ser executada por meio de mudanças abruptas nos níveis de cinza da imagem. A detecção de descontinuidades fragmenta-se em três tipos básicos: de pontos, de linhas e de bordas. A detecção de pontos baseia-se na aplicação de

uma máscara capaz de medir grandes diferenças entre um determinado ponto central e seus vizinhos. A detecção de linhas também ocorre pela aplicação de uma máscara eficaz na identificação de linhas na imagem de acordo com a disposição de seus elementos. Já a detecção de borda, abordagem mais comum em processamento de imagens, é representada pela computação de um operador local diferencial no qual a derivada primeira da imagem pode detectar a presença de bordas e o sinal da derivada de segunda ordem pode determinar se um pixel de borda localiza-se no lado escuro ou claro da imagem. A segmentação baseada em similaridades é composta por técnicas que consideram semelhanças de intensidade ou de níveis de cinza (KANDWAL; KUMAR; BHARGAVA, 2014). Entre as técnicas que compõem essa abordagem estão: limiarização e agrupamento de regiões. A limiarização é uma das mais importantes abordagens para a segmentação de imagens. Em tal método, os pixels do objeto e do fundo passam a ter seus níveis de cinza agrupados em duas classes dominantes. Já a segmentação orientada a regiões representa a divisão completa da imagem em  $n$  regiões distintas (GONZALEZ; WOODS, 2000).

Entre as práticas de segmentação de imagens histológicas, podem ser citadas: limiarização global, limiarização local e *multithresholding*. A limiarização global é utilizada quando a distribuição de intensidade entre os objetos e o fundo da imagem é muito distinta fazendo com que um único valor de limiar seja capaz de separar essas regiões (KANDWAL; KUMAR; BHARGAVA, 2014). Uma das técnicas pertencentes a esse tipo de limiarização, e utilizada nos trabalhos de (HAGGERTY et al., 2014) e (KRISHNAN et al., 2010), é a segmentação por Otsu, que permite a separação entre os objetos de interesse e o fundo da imagem por meio da maximização da variância entre essas diferentes regiões e a minimização da variância interna das mesmas (MOHAMMED et al., 2013a). A limiarização local divide a imagem em várias sub-regiões e então seleciona vários valores de limiar para cada uma delas (KANDWAL; KUMAR; BHARGAVA, 2014). Já a técnica de *multithresholding*, utilizada em (JANSSENS et al., 2013), objetiva encontrar diferentes limiarizações para separar diferentes objetos, como núcleo, citoplasma e estroma (BANIMELHEM; YAHYA, 2011; HE et al., 2010). Tais técnicas não são bem sucedidas em imagens complexas. Para tanto, outros métodos são utilizados, como o crescimento de regiões (ONG et al., 1996), onde o agrupamento de sub-regiões ocorre de acordo com pontos “semente” que agregam pixels com propriedades similares a suas sub-regiões (GONZALEZ; WOODS, 2000).

Além dessas técnicas, outros algoritmos também podem ser utilizados no contexto de imagens histológicas. Tem-se observado o aumento do número de trabalhos que propõem o uso de técnicas de classificação em conjunto com informações locais dos pixels. Entre os mesmos, destacam-se: o *K-Nearest Neighbors* ( $k$ -*nn*), empregado no trabalho de (VINITSKI et al., 1996), onde pixels são agrupados de acordo com as propriedades presentes em seus  $k$  vizinhos mais próximos, e o *K-Means Clustering* ( $k$ -*means*), utilizado no trabalho de (SIEREN et al., 2010), que associa pixels componentes da imagem em diferentes agrupamentos definidos de acordo com o valor  $k$ . Para a detecção de regiões celulares e

nucleares sobrepostas, a transformada de *watershed* é usualmente utilizada por meio de sua inicialização sobre pixels específicos e preenchimento gradual de regiões circundantes dos mesmos, relacionando seus valores a topografias locais (IRSHAD et al., 2014), como utilizado por (VAHADANE; SETHI, 2013).

Em imagens histológicas de linfoma, destacam-se as técnicas: *watershed* e o modelo de contorno ativo, além de métodos já citados, como o algoritmo de Otsu. Em (NEUMAN et al., 2010), o algoritmo *watershed* foi utilizado para a separação das regiões nucleares detectadas na imagem. O modelo de contorno ativo é utilizado para descrever o contorno de objetos por meio de informações de gradiente, procurando minimizar uma função de energia (IRSHAD et al., 2014). Esse método foi empregado por (BELKACEM-BOUSSAID et al., 2010) devido ao pequeno contraste característico de suas imagens, assim como bordas turvas e efeitos de sombra. A partir de sua utilização, esse modelo pode fornecer contornos suaves, o que evita o pós-processamento após a detecção de bordas (HE et al., 2010). Além disso, classificadores também foram utilizados, como nos estudos de (OZTAN et al., 2012) e (SERTEL et al., 2008a) que exploram o classificador *k-means* para a separação de regiões foliculares e de centroblastos, respectivamente.

### 3.1.3 Pós-processamento

As técnicas de pós-processamento são utilizadas para corrigir os limites resultantes da segmentação para alcançar resultados próximos àqueles obtidos por especialistas, denominado padrão-ouro (JANSSENS et al., 2013; OLIVEIRA et al., 2013). Entre as técnicas utilizadas sobre imagens histológicas, nessa etapa, estão: morfologia matemática para remoção de falsas regiões detectadas (VAHADANE; SETHI, 2013) e de objetos pequenos ou próximos às bordas (MEESTER et al., 1991), normalização da imagem (CARRERAS; GONZALEZ; SOLORZANO, 2004), algoritmos para refinamento de contorno nuclear (WANG; OZOLEK; ROHDE, 2010) e mapeamento probabilístico (VEILLARD; BRESSAN; RACOCEANU, 2012).

Essa etapa também pode ser empregada em imagens histológicas de linfoma. Em (SERTEL et al., 2008a), a operação morfológica de fechamento foi utilizada sobre amostras de tecidos coradas com Imuno-histoquímica (IHC) para a modelagem de regiões foliculares no formato de elipses, contribuindo também para remoção de ruídos, refinamento de contornos e o preenchimento de espaços nessas estruturas. No estudo de (BELKACEM-BOUSSAID et al., 2011), o pós-processamento também foi realizado por meio da operação de fechamento para a remoção de ruídos presentes na imagem, além da utilização de uma limiarização empiricamente determinada para a eliminação de regiões menores que as nucleares e o emprego de descritores de forma de Fourier para redução de irregularidades características dos contornos obtidos na segmentação. Além disso, filtros de suavização também podem ser utilizados, como no trabalho de (SERTEL et al., 2010a), onde o filtro

Gaussiano foi empregado para a melhor identificação de células individuais acompanhado por operações morfológicas e pela transformada *watershed*.

### 3.1.4 Extração de Características

A extração de características compõe uma das etapas do processamento de imagens histológicas que permite a comparação entre diferentes estruturas de forma quantitativa (WANG; OZOLEK; ROHDE, 2010). Após a segmentação, tais características são extraídas das regiões de interesse para detectar e classificar potenciais neoplasias que podem ser distinguidas por meio desses recursos (HE et al., 2012). O bom desempenho dessa etapa implica diretamente na otimização dessa classificação, a ser executada posteriormente (ONG et al., 1996). Diversas características podem ser extraídas de imagens, entre elas: recursos morfométricos, como forma e tamanho, topológicos, de intensidade, cor e textura (HE et al., 2012). Essa etapa pode ser utilizada para obter diversas características de imagens histológicas. No trabalho de (KRISHNAN et al., 2010), essa etapa baseou-se na análise do formato das células segmentadas por meio de suas características como área, perímetro, compacidade e excentricidade. Em (ATUPELAGE et al., 2013), medidas de textura foram obtidas por meio de medidas fractais que permitiram a observação de disparidades contidas na imagem.

Com linfoma, (SERTEL et al., 2009) utilizou medidas estatísticas capazes de representar características morfológicas e topológicas para a classificação de estruturas de LF. Essas medidas foram extraídas por meio da representação das imagens utilizando *Model-Based Intermediate Representation* (MBIR), onde características morfológicas de comprimento do maior e do menor eixos e a área das regiões nucleares e celulares foram obtidas. As medidas topológicas foram dadas por meio da distância do núcleo mais próximo, a média da distância entre núcleo e citoplasma mais próximos, a identificação de componentes vizinhos, entre diversas outras. Além disso, também foi desenvolvido um mecanismo para análise de cor e textura das imagens utilizando *Self-Organizing Maps* (SOM), para promover a melhor representação da imagem por meio de sua quantização não linear, e construção da matriz de co-ocorrência para extração de informações de textura.

O trabalho de (BELKACEM-BOUSSAID et al., 2009) também passa pela etapa de extração de características ao aplicar técnicas de extração de textura considerando a identificação de objetos por meio do cálculo das áreas dos mesmos. Ao combinar essas informações, torna-se possível classificar as imagens utilizadas. A análise de textura baseou-se na avaliação estatística da mesma no domínio de Fourier. A variação de frequência dessa característica representou um bom descritor da mesma. Assim, a técnica *Principal Components Analysis* (PCA) foi utilizada para quantificar essas variações.

### 3.1.5 Classificação

Essa etapa representa a última operação a ser aplicada no processamento de imagens histológicas. A classificação pode ser executada por uma grande quantidade de métodos capazes de determinar se uma imagem, em níveis celular ou histológico, corresponde a uma patologia (WANG; OZOLEK; ROHDE, 2010). Diversas técnicas podem ser utilizadas para essa classificação, como redes neurais artificiais, *Support Vector Machine* (SVM), *k-nn*, classificador bayesiano e árvores de decisão. Além destas, algumas técnicas baseiam-se na definição de critérios relacionados às características específicas das estruturas de interesse, como em (HAGGERTY et al., 2014), onde os objetos presentes em imagens histológicas epiteliais segmentadas foram classificados de acordo com o cálculo de sua área e da área de sua região delimitante, nas quais partes constituintes da epiderme possuíam área menor que 2.000 pixels e a razão entre essas áreas maior que 0,44. Caso contrário, essas regiões não compunham a epiderme. No estudo de (WANG; OZOLEK; ROHDE, 2010), o classificador bayesiano foi incorporado ao sistema para classificar vários tipos de subepitélios. E no trabalho de (ATUPELAGE et al., 2013), o modelo de classificação *bag-of-feature* (BOF) foi utilizado para classificação de núcleos de células pulmonares.

A aplicação da etapa de classificação também pode ocorrer em imagens histológicas de linfoma, como o trabalho de (ARORA; BANERJEE, 2013) que desenvolveu uma abordagem para classificação de imagens de LF entre os 3 graus dessa neoplasia. A classificação foi realizada por meio de redes neurais artificiais com dois parâmetros: curvatura do núcleo e centro dessa mesma região. Essa etapa também foi empregada no trabalho de (OZTAN et al., 2012) que utilizou os classificadores SVM, *k-nn* e o classificador bayesiano para também classificar imagens de LF entre seus possíveis graus histológicos. Com a mesma finalidade, o trabalho de (SERTEL et al., 2009) também utilizou o classificador bayesiano aliado aos algoritmos *Linear Discriminant Analysis* (LDA) e PCA.

## 3.2 Segmentação de Imagens Histológicas de Linfoma

Linfoma é um amplo campo de pesquisa que inclui grande variedade de neoplasias do sistema linfático. Alguns trabalhos concentram-se nesse tema aliados a aplicação de técnicas de processamento de imagens. Dentre estes, grande parte dedica-se ao estudo de LF. Porém, pesquisas foram conduzidas sobre outros tipos dessa patologia, como LCM e LLC (ORLOV et al., 2010). Trabalhos referentes a LCM, LF e LLC foram descritos conforme representado pela Tabela 1, contendo suas etapas de pré-processamento, segmentação e pós-processamento, além de seus resultados e os bancos de imagens (BI) utilizados para sua validação.

Tabela 1 – Estado da arte de estudos relacionados ao processamento de imagens histológicas coradas com H&E de LCM, LF e LLC.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(ARORA; BANERJEE, 2013) 110 imagens de LF.	Regiões foliculares	-	Contorno ativo utilizando <i>locally clustering energy</i> .	-	Não houve avaliação da etapa de segmentação.
(BELKACEM-BOUSSAID et al., 2010) 40 imagens de LF.	Regiões foliculares	Mapeamento sobre o canal <i>L</i> do modelo LAB e aplicação de <i>matching filter</i> .	Modelo de contorno ativo.	Aplicação de fechamento, remoção de pequenas regiões e aplicação de descriptores de forma de Fourier.	0,71 com desvio padrão de 0,12 utilizando a métrica de Zijbendos.
(BELKACEM-BOUSSAID et al., 2011) 15 imagens de LF.	Regiões foliculares	Seleção do canal <i>B</i> do modelo de cores RGB e aplicação das técnicas <i>multi-channel decorrelation stretching</i> , filtro de média e equalização adaptativa do histograma.	Segmentação baseada em regiões utilizando <i>curve evolution</i> .	Aplicação de fechamento, remoção de objetos, <i>novel concavity index</i> , <i>recursive watershed transform</i> , análise de saturação de cores e descriptores de forma de Fourier.	78,33% pelo índice de similaridade de Zijdenbos com desvio padrão de 2,83.
(DIMITROPOULOS et al., 2014) 300 células contidas em imagens de LF.	Centroblastos	Conversão para níveis de cinza, filtro Gausiano e equalização do histograma.	Limiarização e Otsu.	Aplicação de dilatação, <i>hole-filling</i> , <i>Gaussian mixture modeling</i> utilizando <i>expectation maximization</i> , extração de características morfológicas e texturais e classificador <i>neuro-fuzzy</i> .	Acurácia de detecção de 90,35%.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(KONG; BELKACEM-BOUSSAID; GURCAN, 2011a) 10 imagens de LF.	Núcleos celulares	Conversão para <i>most discriminant color space</i> (MDC) baseada em análises discriminativas Fisher-Rao.	Limiarização, <i>Efficient Local Fourier Transform</i> , <i>k-means</i> e <i>k-nn</i> .	-	Acurácia média de 0,769.
(KONG; GURCAN; BELKACEM-BOUSSAID, 2011b) 10 imagens de LF.	Núcleos celulares	Conversão para MDC baseada em análises discriminativas Fisher-Rao.	Limiarização, <i>Efficient Local Fourier Transform</i> , <i>k-means</i> e <i>k-nn</i> .	<i>Radial symmetry point detector</i> , detecção de concavidades e aplicação de descritores de forma de Fourier.	Acurácia de células corretamente separadas de 94,74%.
(LUO; CELENK; BEJAI, 2006) Conjunto de imagens celulares de sangue de LCM, Leucemia de Células Pilosas e Leucemia de Células Plasmáticas.	Células neoplásicas	Conversão para níveis de cinza e inversão desses níveis de intensidade.	Otsu e <i>watershed</i> .	-	Não houve avaliação da etapa de segmentação.
(MICHAEL et al., 2014) 12 imagens de LF.	Centroblastos	Conversão para níveis de cinza, filtro Gaussiano e equalização do histograma.	Limiarização com valor 0,37 para eliminação de glóbulos vermelhos e Otsu para a segmentação nuclear.	Remoção de pequenos objetos, aplicação de dilatação, <i>hole-filling</i> , erosão, <i>Gaussian mixture modeling</i> , limiares baseados no formato nuclear, <i>singular value decomposition</i> e análises lineares discriminantes.	Acurácia de 82,58%.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(MOHAMMED et al., 2013a) 140 imagens de sangue.	Linfócitos	-	Otsu, <i>k-means</i> , SVM e subtração entre segmentações celular e nuclear.	-	Acurárias máximas de 98,43%, 98,69% e 99,85% nas segmentações nuclear, celular e citoplasmática, respectivamente.
(MOHAMMED et al., 2013b) 132 imagens de sangue.	Linfócitos	Conversão para níveis de cinza.	Otsu, <i>Canny edge detector</i> e subtração entre segmentações celular e nuclear.	Aplicação de dilatação, <i>hole-filling</i> , erosão e <i>watershed</i> com remoção de 1% de mínimos locais.	Acurárias de 99,92%, 99,85% e 99,63% nas segmentações nuclear, celular e citoplasmática, respectivamente.
(OGER; BELHOMME; GURCAN, 2012) 12 imagens de LF coradas com H&E e IHC.	Regiões foliculares	Aplicação de <i>decorrelation stretching</i> , filtro de média, equalização adaptativa do histograma e separação dos canais <i>R</i> e <i>B</i> do modelo RGB.	Otsu e intersecção entre máscaras binárias obtidas sobre os canais <i>R</i> e <i>B</i> .	Aplicação de dilatação, <i>watershed</i> e remoção de pequenos objetos.	Índices de sensibilidade, especificidade, conformidade e Jaccard com médias máximas de 0,935, 0,843, 0,198 e 0,57, respectivamente.
(OZTAN et al., 2012) 17 imagens de LF coradas com H&E.	Centroblastos	Conversão para o modelo de cores LAB.	<i>k-means</i> .	Aplicação de MBIR.	Não houve avaliação da etapa de segmentação.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(SERTEL et al., 2008a) 30 imagens de LF coradas com IHC e 11 imagens LF coradas com H&E.	Centroblastos e regiões foliculares.	-	Para a detecção de regiões foliculares, aplicação de matriz de co-ocorrência e canal <i>S</i> do modelo de cores <i>Hue Saturation Value</i> (HSV). Para detecção de centroblastos, características extraídas foram reduzidas pela aplicação de PCA e <i>k-means</i> .	Remoção de pequenas regiões, aplicação de fechamento e transformada de <i>watershed</i> .	Detecção de folículos com sensibilidade de $85,5\% \pm 9,8\%$ e especificidade de $92,5\% \pm 4,0\%$ . Detecção de centroblastos com sensibilidade de 92,68% e especificidade de 90,57%.
(SERTEL et al., 2008b) 17 imagens de LF.	Centróцитos e centroblastos.	Conversão para o modelo LAB.	Limiarização para eliminação de glóbulos vermelhos e fundo da imagem sobre o modelo RGB e aplicação de <i>k-means</i> sobre o modelo LAB.	-	Não houve avaliação da etapa de segmentação.
(SERTEL et al., 2009) 17 de imagens de LF.	Centroblastos	Conversão para o modelo LAB.	Limiarização e <i>k-means</i> .	Remoção de áreas irrelevantes, aplicação da transformada <i>watershed</i> e MBIR.	Não houve avaliação da etapa de segmentação.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(SERTEL et al., 2010a) 100 ROI de imagens de LF.	Centroblastos	Conversão para o modelo LUV.	<i>Mean-shift</i> com eliminação de agrupamentos com quantidade de dados menor que 2% do total.	Atribuição de pesos para componentes nucleares e citoplasmáticos, aplicação de filtro Gaussiano, filtro <i>top-hat</i> , transformada <i>watershed</i> , construção de medidas de vizinhança de centroblastos e aplicação de limiarização adaptativa.	Acurácia de 89%.
(SERTEL et al., 2010b) 100 ROI de imagens de LF.	Centroblastos	Conversão para imagens 1-D <i>unitone</i> utilizando PCA e normalização.	<i>Gaussian mixture modeling</i> com parâmetros estimados utilizando <i>expectation maximization</i> .	Limiarização adaptativa, transformada <i>fast radial symmetry</i> , remoção de regiões considerando tamanho, excentricidade e características de textura extraídas por meio da técnica <i>gray-level run length matrix</i> (GLRLM).	Acurácia de 80,7%.

Ref. and BI	Objetos de interesse	Pré-processamento	Segmentação	Pós-processamento	Métricas e Resultados
(YANG et al., 2008) 18 casos de LCM, 9 de LF, 20 de LLC, e 54 casos de outros tipos de linfoma, resultando em 3.898 imagens celulares.	Células de hematopatologias sobrepostas.	Conversão para o modelo LUV.	Algoritmos L2E e GVF para estimativa de contornos, detecção de pontos de alta curvatura, aplicação de <i>canny edge detector</i> para modelagem de bordas internas e <i>concave vertex graph</i> .	<i>Quadratic splines</i>	Acurácia de 90,1%.
(ZORMAN et al., 2007) O sistema foi treinado sobre 3.627 objetos e testado sobre 1.813 objetos de LF.	Regiões foliculares	Separação do canal <i>R</i> do modelo RGB.	Limiarização utilizando médias dos valores de brilho.	Transformada de matriz e árvores de decisão.	Máxima acurácia de 86,65%.

Entre os métodos de segmentação utilizados sobre imagens histológicas de linfoma coradas com H&E, presentes na literatura, destacam-se os estudos de (MICHAIL et al., 2014), (MOHAMMED et al., 2013b), (SERTEL et al., 2008b) e (YANG et al., 2008) por serem trabalhos realizados recentemente com resultados relevantes por meio da utilização de técnicas que contribuem para este trabalho.

Entre os trabalhos relacionados ao processamento de imagens histológicas de LF, está o estudo de (SERTEL et al., 2008b) que propõe um método de classificação dessa patologia entre os graus indolente (I e II) e alto (III) por meio de análises de textura e cor das imagens coradas com H&E. Para discriminar os graus indolentes, subdivididos entre baixo (I) e médio (II), um novo método é apresentado para a extração de características. O algoritmo foi dividido nas etapas de segmentação, extração de características e classificação. Na segmentação, as imagens foram convertidas do modelo de cores RGB para o modelo LAB, onde a diferença entre cores é uniforme e permite a utilização da distância euclidiana como uma medida a ser aplicada para quantização não uniforme de cores. Entretanto, ambos os modelos de cores foram utilizados. Sobre o modelo RGB, uma limiarização foi aplicada considerando valores de intensidade para remoção de glóbulos vermelhos. Outras estruturas foram segmentadas sobre o modelo LAB por meio do algoritmo *k-means*, contendo 3 agrupamentos: núcleo, citoplasma e material extracelular. A distinção entre os graus indolente e alto foi possível por meio do cálculo do raio entre quantidades de citoplasma e núcleo, considerando que células de alto grau são maiores. O objetivo da etapa de extração de características foi diferenciar os graus baixo e médio. Essa abordagem foi executada pela modificação da matriz de co-ocorrência utilizando quantização não uniforme de cores por meio da técnica *Self Organizing Feature Maps* (SOFM) que corresponde a uma rede neural artificial com treinamento não supervisionado. Esse tipo de quantização pode aperfeiçoar o espectro limitado de cores apresentado por imagens H&E. Além disso, a quantização uniforme não é capaz de representar cores de forma natural, o que é essencial para a matriz de co-ocorrência (SERTEL et al., 2008b). Posteriormente, tal matriz foi computada para obter informações de textura, permitindo a extração de cinco características: homogeneidade, energia, contraste, correlação e entropia. Essas informações constituem um conjunto suficiente para a classificação subsequente, utilizando um classificador bayesiano associado à técnica de redução de dimensionalidade PCA. O sistema foi validado sobre um banco de imagens constituído por 17 imagens com magnificação 40× classificadas por três patologistas, resultando em seis imagens de grau I, oito de grau II e três de grau III. A extração de ROI obtida por especialistas resultou em 510 imagens. A avaliação da etapa de classificação foi executada pela abordagem *k-fold cross-validation* com cinco conjuntos utilizados para treinamento e teste com a mesma quantidade de imagens. A classificação entre os graus indolente e alto foi efetiva enquanto a diferenciação entre os graus baixo e médio mostrou-se complexa devido a similaridade entre seus componentes citológicos. A acurácia do sistema atingiu

88,9%. O algoritmo proposto não incluiu a avaliação da etapa de segmentação porém exemplifica a importância do emprego dessa etapa para tal finalidade.

Em (MICHAIL et al., 2014), uma técnica foi proposta para detecção automática de centroblastos em imagens de LF coradas com H&E. O método foi dividido em pré-processamento, segmentação, rotulação de células, seleção de regiões de interesse e classificação. Na etapa de pré-processamento, imagens RGB foram submetidas à conversão para níveis de cinza, e um filtro Gaussiano de tamanho  $3 \times 3$  pixels foi aplicado para remoção de ruídos e detalhes irrelevantes. A equalização do histograma também foi utilizada para aperfeiçoamento de contraste e detecção de estruturas nucleares de forma simples e eficiente. Após o pré-processamento, glóbulos vermelhos foram eliminados por meio de uma máscara aplicada sobre o modelo RGB com limiarização global empiricamente determinada com valor 0,37. O método de Otsu foi executado para separação de regiões nucleares e material extracelular. Na etapa de pós-processamento, rotulação de componentes conectados foi utilizada para identificar regiões individuais e uma operação morfológica foi aplicada para remoção de pequenos objetos. Após a segmentação, regiões internas de grandes células poderiam apresentar uma aparência oca devido às suas transparências. Para o aperfeiçoamento dessas regiões, operações de dilatação, *hole-filling* e erosão foram aplicadas. Como esse processo pode apresentar certas limitações quanto à sobreposição de células, um método baseado em *Gaussian mixture models* foi implementado por meio do algoritmo *expectation maximization*. Essa abordagem permite a definição de critérios para a identificação de coordenadas de possíveis núcleos sobrepostos. Para selecionar as regiões de interesse, critérios de intensidade de brilho, proporções de células e distâncias médias considerando perímetros dessas estruturas foram utilizados. Após a etapa de seleção, possíveis centroblastos foram classificados por meio da extração de características morfológicas, topológicas e texturais utilizando LDA. Esse algoritmo foi executado sobre imagens com magnificação  $400\times$  caracterizadas por pequena espessura durante o processo de aquisição das amostras histológicas. Tal característica permite uma descrição detalhada de regiões nucleares, apesar de torná-las homogêneas em relação às suas características de cor e textura. Nesse banco, nove imagens foram utilizadas e divididas em 70 centroblastos e 100 não centroblastos. O sistema foi testado sobre três imagens de LF de grau II. Para avaliar o algoritmo, o conjunto foi dividido aleatoriamente  $k$  vezes com distribuição de 80% dos dados para treinamento e 20% para validação por meio da técnica *k-fold cross-validation*. A quantidade de objetos de interesse identificados alcançou acurácia de 82,58% considerando a análise realizada por patologistas, por meio do cálculo da média de centroblastos detectados em três imagens. A média de regiões falso positivas identificadas nas imagens de teste foi de 50. De acordo com os autores, essa alta taxa pode ser explicada pela semelhança entre certos tipos celulares, como células endoteliais e centroblastos. Em trabalhos futuros, o uso de informações de textura será investigado para detectar diferenças entre grandes células.

Entre os estudos relacionados à LLC, o trabalho de (MOHAMMED et al., 2013b) propôs um método de segmentação de linfócitos. Seu objetivo é a detecção precoce de LLC por meio da identificação das estruturas de núcleo, citoplasma e célula, essenciais para esse propósito. Além disso, outra abordagem foi utilizada para redução dos efeitos de super e subsegmentação resultantes do algoritmo *watershed*. O método foi dividido em segmentação nuclear, segmentação celular e segmentação citoplasmática. A segmentação nuclear foi inicializada por meio da conversão para níveis de cinza com posterior aplicação da técnica de Otsu, que foi capaz de isolar regiões nucleares. Isso foi possível já que essas estruturas eram caracterizadas por níveis de intensidade escuros, quando comparados às intensidades de outras regiões. Entretanto, ainda nessa etapa, o algoritmo *Canny edge detector* foi utilizado seguido pelas operações morfológicas de dilatação, *hole-filling* e erosão para a remoção de objetos irrelevantes e regiões com áreas menores que o tamanho nuclear. Os resultados da segmentação foram representados por uma máscara contendo apenas as regiões nucleares. A segmentação celular foi também executada sobre imagens em níveis de cinza utilizando as técnicas de Otsu, *Canny edge detector* e as operações morfológicas de dilatação, *hole-filling* e erosão. Durante essa etapa, o algoritmo *watershed* foi utilizado e dividido em uma etapa de pré-processamento composta pela eliminação de mínimos locais para prevenção dos efeitos de super e subsegmentação. O uso do método *watershed* pode levar à união de linfócitos e glóbulos vermelhos. Assim, apenas 1% dos mínimos locais foram removidos. Depois desse processo, o algoritmo *watershed* foi aplicado resultando em uma matriz de componentes conectados de cada imagem, o que resultou em uma máscara contendo representações celulares. A segmentação citoplasmática foi obtida pela subtração pixel a pixel das máscaras geradas nas etapas anteriores, o que também resultou em uma máscara contendo apenas as regiões de citoplasma. O sistema foi testado sobre 140 imagens com 93 linfócitos neoplásicos de LLC e 47 linfócitos normais com resolução de  $363 \times 360$  pixels e magnificação  $100\times$ . Entretanto, de acordo com os autores, apenas 132 imagens foram processadas já que oito delas possuíam linfócitos muito próximos a glóbulos vermelhos, o que afetaria negativamente a execução do sistema em relação à supersegmentação na etapa de segmentação celular. Estimativas de acurácia foram executadas considerando comparações entre imagens processadas pelo algoritmo proposto e aquelas segmentadas manualmente por um especialista. Os resultados finais demonstraram 99,92% de acurácia na segmentação nuclear, 99,85% na segmentação celular e 99,63% na segmentação citoplasmática. Em trabalhos futuros, imagens com diferentes condições de iluminação e coloração em suas aquisições serão utilizadas para validação do sistema.

Considerando casos de LCM, (YANG et al., 2008) apresentou um método de segmentação de células sobrepostas contidas em imagens histológicas de linfoma. Estruturas de LF, LLC e de outras classes de linfoma também foram segmentadas para validação do algoritmo. O estudo foi dividido nas etapas de extração do contorno de bordas, detecção de pontos côncavos, detecção de bordas internas, segmentação de células próximas, cál-

culo de função de custo e pós-processamento. O algoritmo de contorno de bordas objetiva extraír bordas de células sobrepostas. Esse processo foi inicializado pela aplicação de L2E para obter estimativas de bordas exteriores das regiões de interesse. Para separar objetos do fundo da imagem, o método GVF *snake* foi aplicado, utilizando os gradientes de cor do modelo de cores LUV. Utilizando o resultado da etapa anterior, pontos côncavos foram detectados considerando pontos de alta curvatura nas bordas identificadas. Em cada ponto de borda, um conjunto de triângulos foi construído e analisado para identificar pontos côncavos e conexos. Entretanto, apenas pontos côncavos foram considerados para cálculos realizados sobre as bordas dos referidos triângulos. Para a detecção de bordas internas, os algoritmos *Canny edge detector* e *straight line fitting* foram utilizados. Com pontos côncavos e bordas internas detectadas, foi possível determinar curvas que separavam os objetos. Posteriormente, a segmentação de células próximas foi aplicada baseado na construção de um grafo côncavo. Essa estrutura considera todas as bordas e pontos identificados anteriormente, onde seus vértices representam o conjunto de pontos de bordas internas e o conjunto de pontos côncavos. Suas arestas correspondem ao conjunto de bordas internas e o conjunto formado por *filling edges* que associam vértices não conectados a pontos internos. Para identificar possíveis bordas para a separação de células, pesos foram associados a cada aresta considerando bordas internas e *filling edges*. Então, o algoritmo de Dijkstra foi aplicado para que o menor caminho entre vértices fosse encontrado, o que leva à segmentação utilizando bordas internas e não apenas *filling edges* no grafo, que levariam a caminhos maiores. Uma etapa de minimização de funções de custo foi empregada para a procura pelo melhor caminho. Essa função representa três critérios utilizados por patologistas: células deveriam ser salientes, com formato próximo a elipses, e com áreas biologicamente razoáveis. Todos esses critérios foram representados por funções matemáticas que foram combinadas em um custo final que deveria ser minimizado. Na etapa de pós-processamento, *quadratic splines* foram utilizadas para suavização das bordas dos objetos. Considerando que núcleo e citoplasma são caracterizados por cores distintas, esse algoritmo foi capaz de separar apenas o citoplasma de células próximas. O sistema foi testado sobre 86 casos representados por imagens coradas com H&E e com magnificação 60 $\times$ , compostos por 18 casos de LCM, 9 de LF, 20 de LLC e 54 casos de outros tipos de linfoma, e 19 casos benignos. Os resultados obtidos foram comparados à segmentação manual executada por especialistas em 207 casos de células próximas, alcançando uma taxa de acurácia de 88,9%, e em um banco composto por 3.898 casos com uma acurácia de 90,1%. Além disso, esse método obteve desempenho superior ao algoritmo *watershed*, comumente utilizado para a segmentação de células próximas.

### 3.3 Considerações Finais

No contexto de imagens histológicas de linfoma, diversas contribuições em segmentação e detecção de estruturas citológicas vêm sendo realizadas. Entretanto, algumas limitações ainda podem ser encontradas nesses estudos da literatura. Uma dessas limitações pode ser atribuída ao uso de apenas um modelo de cores para a execução de algoritmos propostos. Muitos estudos indicam os benefícios do uso do modelo LAB devido a sua representação de cores homogênea (BELKACEM-BOUSSAID et al., 2010; OZTAN et al., 2012; SERTEL et al., 2008b; SERTEL et al., 2009), o modelo HSV e sua capacidade de obter informações com mais detalhes (SERTEL et al., 2008a), além do modelo LUV (SERTEL et al., 2010a) e a técnica de deconvolução (KORZYNSKA et al., 2013; ORLOV et al., 2010; ZORMAN et al., 2007). Em (OZTAN et al., 2012), o modelo LAB foi utilizado devido à sua aplicação eficiente para o cálculo da distância euclidiana. Entretanto, de acordo com (SERTEL et al., 2010a), o modelo LUV também alcança alta eficiência quando utilizado para o cálculo desse tipo de distância.

A técnica de deconvolução pode ser aplicada para a separação de cores em diferentes canais da imagem. Para correção de iluminação, essa técnica é útil para a extração do canal *L* do modelo LAB já que esse canal possui maiores similaridades com a percepção humana de luz (KURU, 2014). Além disso, a deconvolução pode ser utilizada sobre imagens RGB para separar o canal *R* que possui maior contraste comparado ao fundo da imagem (ZORMAN et al., 2007), e o canal *B*, que alcança maior representação descritiva de regiões foliculares (BELKACEM-BOUSSAID et al., 2010). Imagens H&E também podem ser submetidas à deconvolução de cor onde o canal *H* é capaz de representar, principalmente, regiões nucleares e o canal *E* representa áreas de citoplasma (ORLOV et al., 2010). Assim, a comparação de diferentes modelos de cores e diferentes técnicas para tratamento de cores pode contribuir para melhores desempenhos, o que é refletido por resultados quantitativos e qualitativos dos sistemas.

Entre os estudos analisados, alguns utilizam pequena quantidade de casos em seus bancos de imagens para avaliação de seus métodos. De acordo com (BELKACEM-BOUSSAID et al., 2011), a aplicação de métodos de segmentação em quantidade limitada de casos pode não ser capaz de corroborar com a eficiência dos sistemas propostos. Além da pequena quantidade de imagens utilizadas, os estudos de (OGER; BELHOMME; GURCAN, 2012) e (SERTEL et al., 2008a) utilizaram imagens de referência obtidas de amostras coradas com outro corante (IHC) para auxílio à segmentação de imagens de LF coradas com H&E. Bancos utilizados nesses trabalhos foram compostos por imagens muito semelhantes em seu conteúdo, porém coradas com esses diferentes tipos de corantes, fazendo necessária a presença de pares de imagens para seu processamento. O uso desses dois tipos de imagens pode ser considerado um obstáculo para a aplicação prática desses métodos.

Devido a variações no espectro de cores de imagens, métodos de segmentação não

supervisionados podem tornar sistemas de processamento de imagens histológicas mais robustos a esse tipo de variação se comparado a aplicações de técnicas supervisionadas (SERTEL et al., 2009). Diversos trabalhos utilizaram a técnica *k-means* na etapa de segmentação, como em (OZTAN et al., 2012), (SERTEL et al., 2008a), (SERTEL et al., 2008b) e (SERTEL et al., 2009). Entretanto, outros métodos não supervisionados podem ser utilizados, como o algoritmo *mean-shift*, o qual não é caracterizado pelo conhecimento prévio da quantidade de grupos nos quais as informações da imagem serão divididas. Essa abordagem pode ser benéfica ao não considerar essa quantidade determinada, que pode não ser adequada às variações de cores (SERTEL et al., 2010a).

Entre os trabalhos descritos pela Tabela 1, é possível notar que alguns métodos não utilizaram as etapas de pré-processamento e/ou pós-processamento em seus algoritmos. Essas etapas podem ser consideradas importantes já que a partir delas é possível a remoção de áreas irrelevantes antes e depois da segmentação (BELKACEM-BOUSSAID et al., 2010; BELKACEM-BOUSSAID et al., 2011), aperfeiçoamento da visibilidade de folículos (BELKACEM-BOUSSAID et al., 2011), identificação de células individuais (MICHAEL et al., 2014; SERTEL et al., 2010a) e outros aprimoramentos capazes de contribuir para o desempenho do sistema.

Além disso, a implementação de sistemas propostos por alguns trabalhos foram conduzidos sobre células localizadas na região central das imagens contidas nos bancos utilizados. Tal abordagem permite a aplicação de tais trabalhos apenas sobre células individuais. Os estudos de (MOHAMMED et al., 2013b) e (MOHAMMED et al., 2013a) foram caracterizados por essa análise, com alta acurácia obtida em seus resultados, porém, sem o tratamento de imagens compostas por diferentes distribuições espaciais de suas células. Assim, testes realizados em diferentes tipos de imagens podem corroborar o bom desempenho e alta robustez de algoritmos propostos.

Outro importante aspecto no desenvolvimento de CAD são as métricas utilizadas para validação de resultados. (OGER; BELHOMME; GURCAN, 2012) recomenda a utilização das métricas de sensibilidade, especificidade e conformidade ao invés do índice de Jaccard, já que este pode não ser capaz de expressar apropriadamente a qualidade da segmentação comparada à segmentação manual. Nos estudos de (MOHAMMED et al., 2013b) e (MOHAMMED et al., 2013a), as métricas utilizadas foram caracterizadas pelo uso de apenas pixels não zero e o número de pixels identificados nas segmentações automática e manual, desconsiderando a sobreposição espacial entre essas áreas. Para que comparações possam ser realizadas com algoritmos da literatura, é necessário estabelecer um padrão para métricas de avaliação (GURCAN et al., 2009). Considerando os estudos relacionados à segmentação de imagens de linfoma descritos neste trabalho, a métrica mais utilizada é a acurácia. Entretanto, essa métrica pode ser calculada de diversas maneiras. Em (MOHAMMED et al., 2013b) e (MOHAMMED et al., 2013a), a acurácia foi calculada como descrito anteriormente, enquanto em (SERTEL et al., 2010a), essa

métrica foi calculada utilizando a localização dos centróides de centroblastos detectados e centroblastos identificados por especialistas. Porém, ambas são chamadas de acurácia mesmo com diferenças observadas em seus cálculos.

Diferentes processos de preparação de tecidos podem ser uma limitação de sistemas propostos. Essa é uma consideração importante, já que a coloração e a espessura das amostras de tecidos podem impactar nos resultados obtidos na segmentação (BELKACEM-BOUSSAID et al., 2011). Assim, tais sistemas devem considerar condições de contraste e iluminação para se tornarem robustos a diferentes tipos de amostras. (BELKACEM-BOUSSAID et al., 2011) sugere que um algoritmo proposto possa considerar a qualidade das imagens como um parâmetro para sua execução. (OGER; BELHOMME; GURCAN, 2012) também destaca a aplicação de técnicas para estimativa de parâmetros internos de métodos para alcançar os melhores resultados possíveis. Esses métodos podem ser aplicados e, iterativamente, os resultados obtidos podem ser avaliados. Diversas técnicas podem ser utilizadas para esse propósito e uma delas são os algoritmos genéticos, por exemplo.

De acordo com (MCCANN et al., 2015), o desenvolvimento de técnicas robustas que possam ser aplicadas sobre diferentes imagens histológicas ainda é um desafio. Os estudos descritos neste trabalho não mencionam o uso de imagens obtidas de um banco público. Todos os trabalhos foram analisados e seus bancos de imagens foram pesquisados, mas ainda assim não foram encontrados. O desenvolvimento de algoritmos aplicáveis em tais bancos é importante ao demonstrar robustez em relação a diferentes métodos de preparação de tecidos. Um banco de imagens público desenvolvido por (NIA, 2008) é descrito e utilizado neste trabalho.

Considerando abordagens de detecção, alguns trabalhos possuem limitações relacionadas à alta quantidade de regiões falso positivas. De acordo com os autores, análises de diferentes características são necessárias para contornar essa dificuldade (MICHAEL et al., 2014; SERTEL et al., 2010a). Além disso, características, como as topológicas e morfológicas, podem ser obtidas por refinamentos em MBIR, como sugerido por (SERTEL et al., 2009).

Uma frequente dificuldade apresentada por trabalhos relacionados à segmentação de imagens de linfoma é a separação de células sobrepostas, como presente em (BELKACEM-BOUSSAID et al., 2010), (MOHAMMED et al., 2013b), (MOHAMMED et al., 2013a) e (OZTAN et al., 2012). Essa limitação leva a resultados da segmentação que poderiam ser ainda melhores, caso contassem com seu tratamento. (BELKACEM-BOUSSAID et al., 2010) sugere considerar conectividade espacial, variações de coloração, ajustamento de curvas e características geométricas de ROI para o tratamento dessa limitação.



# CAPÍTULO 4

## Metodologia

### 4.1 Visão Geral do Método

O desenvolvimento do sistema proposto baseou-se em três etapas: pré-processamento, segmentação e pós-processamento. Cada uma delas engloba técnicas de processamento digital de imagens aplicadas sobre as imagens de entrada do sistema. Não há interação com o usuário para definição de parâmetros de configuração. A entrada necessária para execução do algoritmo é uma imagem histológica de linfoma do modelo de cores RGB. Cada imagem é processada, individualmente, e o resultado é uma imagem binária, contendo regiões segmentadas de componentes nucleares que indicam a incidência de LCM, LF ou LLC. A Figura 13 apresenta um esquema geral da sequência metodológica desenvolvida na análise dos componentes celulares de linfoma.

Todo o sistema foi desenvolvido utilizando linguagem MATLAB® e os experimentos foram realizados utilizando um *ultrabook* (Acer M5-481T-6417) com processador de 1,7 GHz (Intel Core i5-3317U) e 6 GB de memória RAM. Também foi utilizado o *plugin deconvolution* do *software* ImageJ (IMAGEJ, 2015) para análise de imagens resultantes do processo de deconvolução, a ser descrito no Capítulo 5.

### 4.2 Banco de Imagens

Os diferentes casos de linfoma tratados neste trabalho foram digitalizados por meio de um microscópio de luz branca Zeiss Axioscope com lente objetiva de  $20\times$  e câmera colorida CCD AxioCam MR5. Todas as imagens foram obtidas sob as mesmas configurações dos instrumentos utilizados e mesmas lentes objetivas, câmera e fonte luminosa. Tais amostras histológicas foram preparadas com aplicação do corante H&E e digitalizadas no padrão de cores RGB com 24 bits de quantização, disponíveis para *download* em (NIA, 2008).

O conjunto de imagens descrito foi analisado pelo Professor Doutor Paulo Rogério Faria, do Departamento de Morfologia do Instituto de Ciências Biomédicas da Universidade Federal de Uberlândia, para segmentação manual de subconjuntos de cada classe.

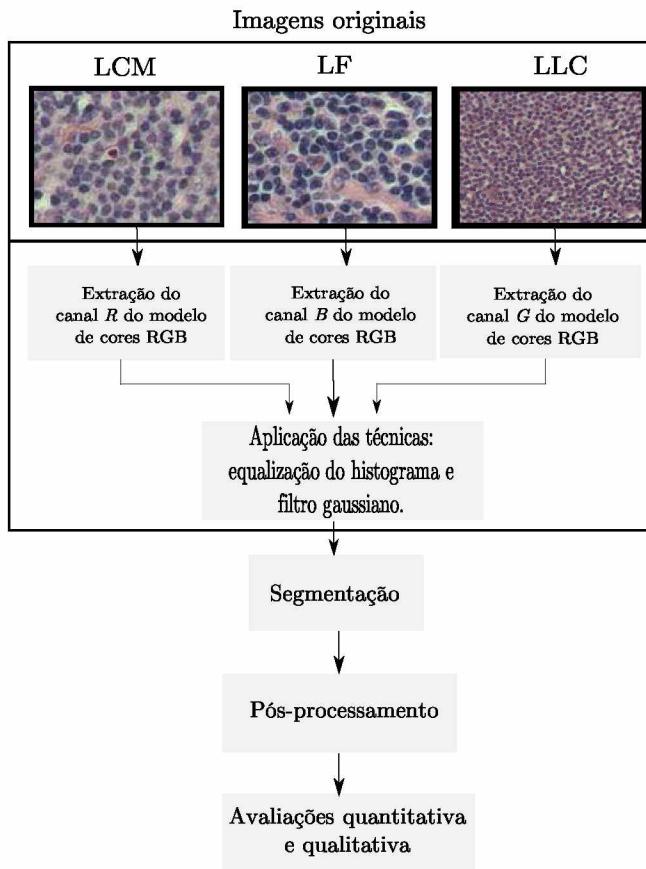


Figura 13 – Fluxograma do trabalho proposto.

Originalmente, a base de imagens é composta por 122 imagens de LCM, 139 de LF e 113 de LLC. Devido a ruídos e desfoque gerados no processo de aquisição das imagens, algumas delas foram eliminadas pelo especialista já que a identificação de anormalidades tornou-se complexa com esses empecilhos. As imagens resultantes dessa seleção foram escolhidas de forma aleatória e sem consideração da distribuição das células das diferentes classes de lesões. Com isso, o conjunto de imagens utilizado neste estudo foi de 99, 62 e 12 imagens de casos de LCM, LF e LLC, respectivamente, com dimensões inferiores ou iguais a 1040 × 1388 pixels.

### 4.3 Pré-processamento

Para que os processos posteriores possam ser bem sucedidos, diferentes modelos de cores, canais desses modelos e diversos filtros podem ser utilizados na definição de uma combinação mais adequada para tratamento e realce de contraste das imagens antes da etapa de segmentação.

Vários modelos de cores foram analisados para a definição do canal mais adequado para a etapa de segmentação. Baseado na avaliação visual e na métrica de entropia (Seção 4.6), os canais componentes do modelo de cores RGB foram escolhidos para aplicação das

técnicas de pré-processamento sobre cada tipo de lesão. As Figuras 14(b), 14(c) e 14(d) demonstram, respectivamente, diferenças de contraste entre os canais *R*, *G* e *B* do modelo RGB de uma imagem do grupo LLC, representada pela Figura 14(a). Os histogramas, contendo a frequência de cada nível de intensidade dos canais *R*, *G* e *B* são ilustrados pelas Figuras 14(e), 14(f) e 14(g), respectivamente. É possível notar que a distribuição dos níveis de intensidade do canal *G* (Figura 14(f)) é mais uniforme, indicando maior contraste e contrapondo-se às distribuições dos canais *R* e *B* (GONZALEZ; WOODS, 2000). Esse procedimento foi também empregado para os demais grupos, onde os canais *R* e *B* apresentaram melhor contraste visual de suas estruturas para as classes LCM e LF, respectivamente.

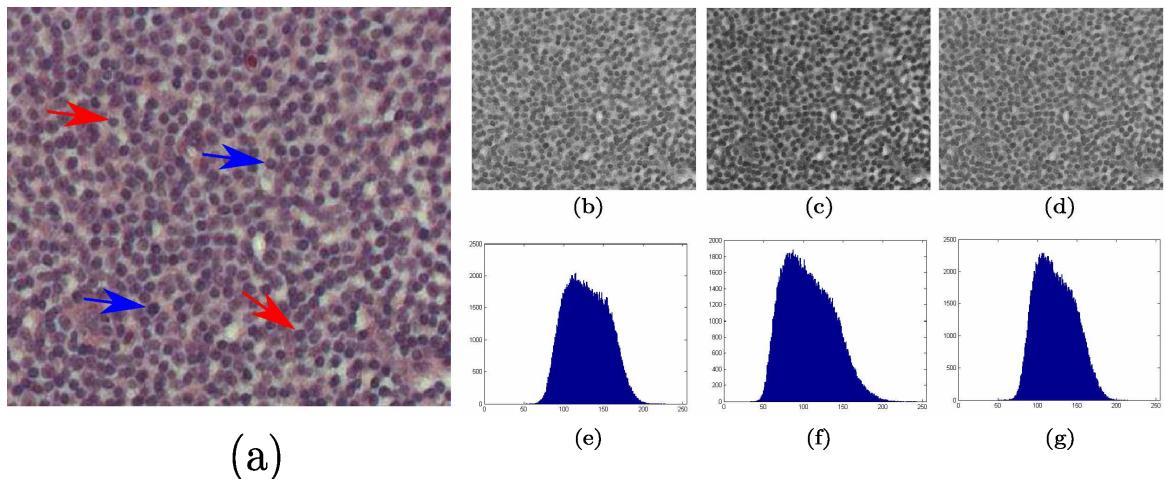


Figura 14 – Separação de canais do modelo de cores RGB de uma subimagem do caso *sj-05-5269-R10\_002* da classe LLC (a), com alguns núcleos neoplásicos indicados por setas vermelhas e outros núcleos saudáveis indicados por setas em azul, que exemplifica a diferença de contraste entre os canais *R* (b), *G* (c) e *B* (d), e seus respectivos histogramas ilustrados por (e), (f) e (g).

Após a definição do canal de cores mais adequado de cada grupo de lesão, a técnica de equalização do histograma foi empregada. A equalização é um método que realça o contraste utilizando informações do histograma. O método busca redistribuir esses níveis de intensidade para que o histograma da imagem processada tenha uma distribuição uniforme. Assim, imagens escuras e claras que antes possuíam altas frequências de seus pixels localizados à esquerda ou à direita, respectivamente, passam a ter uma representação gráfica do histograma com uma distribuição “espalhada” ao longo do intervalo dos níveis de cinza da imagem (GONZALEZ; WOODS, 2000).

Para que esse processo seja executado, deve-se considerar  $r$  a variável que representa os níveis de cinza da imagem a ser realçada, compreendida no intervalo  $[0, L-1]$ , onde  $L$  representa o total de níveis de intensidade da imagem. Para qualquer  $r$ , adota-se uma

função de transformação:

$$s = T(r), \quad (5)$$

que resulta em um novo valor  $s$  a ser atribuído a cada pixel de intensidade  $r$  da imagem. Essa função de transformação deve satisfazer as condições:

1.  $T(r)$  é univariada e monotonicamente crescente no intervalo  $0 \leq r \leq 1$ , considerando níveis de intensidade normalizados;
2.  $0 \leq T(r) \leq 1$  para  $0 \leq r \leq 1$ .

As condições descritas asseguram que o mapeamento esteja compreendido no intervalo delimitado pelos níveis de brilho da imagem preservando a ordem crescente dos mesmos (GONZALEZ; WOODS, 2000).

Assim, dada uma imagem contendo  $n = M \times N$  pixels, assumindo valores discretos para os níveis de cinza  $r$ , a equalização utiliza a função de distribuição acumulada de probabilidade como sua função de transformação, expressa por:

$$g_r = T(f_r) = \sum_{i=0}^r p_f(f_i) = \sum_{i=0}^r \frac{n_i}{n}, \quad (6)$$

onde,  $n_i$  representa o número de ocorrências do nível de cinza  $i$ , e  $p_f(f_i)$  é a probabilidade do  $i$ -ésimo nível de cinza. Para que essa função seja aplicada na equalização, sua normalização é necessária. Nas Figuras 15(a) e 15(b), é representada uma imagem da classe LLC e seu respectivo histograma. Nas Figuras 15(c) e 15(d), são apresentadas a aplicação desse método sobre a Figura 15(a) e a modificação na distribuição do histograma da mesma, tornando-se mais uniforme ao longo do intervalo de intensidades, respectivamente.

Posteriormente, o filtro Gaussiano foi empregado para remoção de ruídos e suavização dos níveis de brilho presentes na imagem. Essa técnica consiste no deslocamento de uma máscara, uma matriz caracterizada por tamanho e distribuição de seus elementos previamente definidos, que resulta na soma dos produtos entre seus elementos e os valores de intensidade da imagem. A distribuição dos coeficientes da máscara do filtro é definida a partir de uma função Gaussiana bidimensional, como representada a seguir:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (7)$$

onde,  $x$  e  $y$  representam as coordenadas dos pixels da imagem e o desvio padrão é representado pela variável  $\sigma$ . Considerando uma função Gaussiana com média  $(0,0)$  e  $\sigma = 1$ , sua representação gráfica é ilustrada pela Figura 16(a). A Figura 16(b) apresenta a

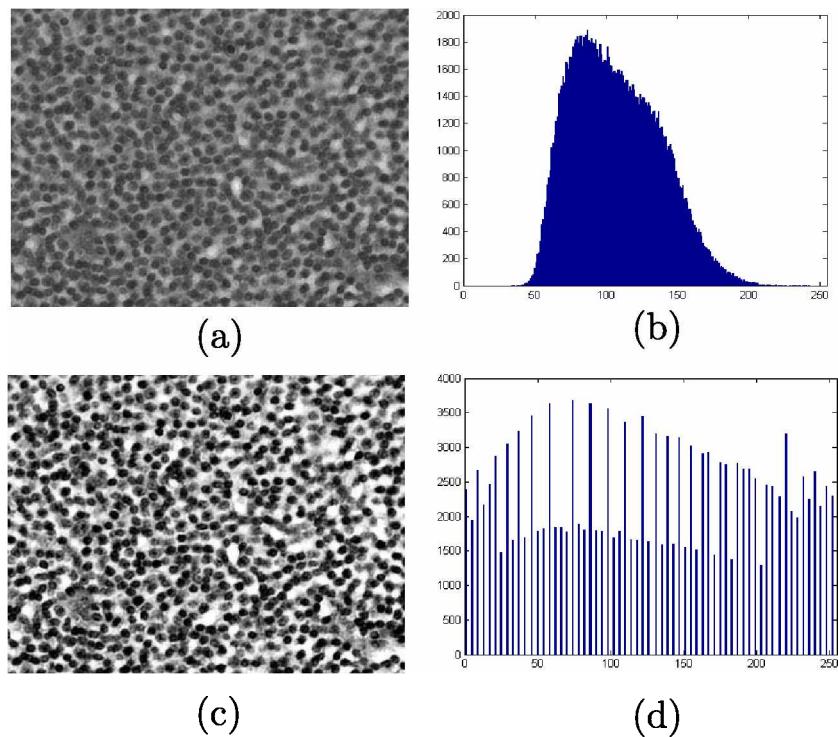


Figura 15 – Imagens com uso da técnica de equalização de histograma: (a) subimagem do caso *sj-05-5269-R10\_002* da classe LLC, (b) seu histograma, (c) imagem após a aplicação da técnica de equalização do histograma e (d) histograma da imagem equalizada.

máscara resultante dessa função de tamanho  $3 \times 3$  pixels a ser aplicada sobre a imagem. A suavização é realizada por meio da substituição de cada pixel por uma média ponderada dos pixels vizinhos, tal que o peso dos vizinhos decresce conforme maior é a distância do pixel central (PEDRINI; SCHWARTZ, 2007). Neste trabalho, diferentes valores atribuídos ao tamanho da máscara e ao parâmetro  $\sigma$  da função Gaussiana foram testados empiricamente, porém o tamanho da máscara  $3 \times 3$  e  $\sigma = 2$  foram mais adequados para os diferentes grupos de lesões.

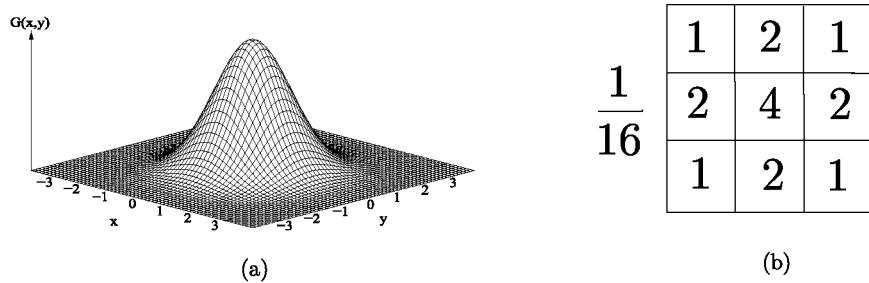


Figura 16 – (a) Representação gráfica de uma função Gaussiana bidimensional com média  $(0,0)$  e desvio padrão igual a 1, e (b) a máscara de tamanho  $3 \times 3$  resultante dessa representação.

As Figuras 17(a), 17(b) e 17(c) ilustram uma imagem original da classe LLC, a ex-

tração do seu canal de cor  $G$  e a aplicação das técnicas descritas, respectivamente. O Algoritmo 1 apresenta a descrição sequencial dos métodos utilizados nessa etapa.

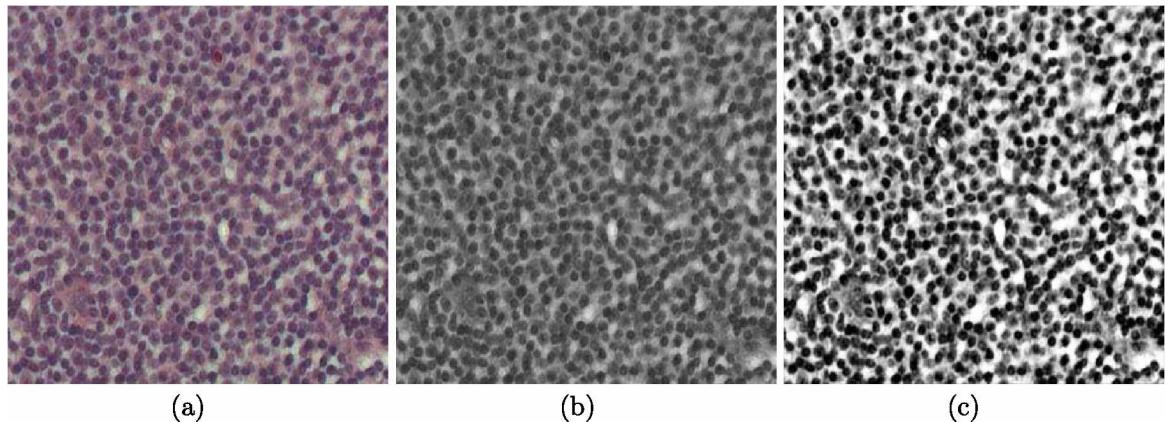


Figura 17 – Exemplo da aplicação das técnicas de equalização do histograma e filtro Gaussiano sobre o canal  $G$  das imagens da classe LLC: (a) imagem original, (b) seu canal  $G$  extraído e (c) aplicação das técnicas de pré-processamento utilizadas.

---

#### **Algoritmo 1** Pré-processamento

**Entrada:** Imagem RGB  $I$  de uma das lesões: LCM, LF ou LLC;

**Saída:** Imagem pré-processada  $I_{PP}$ ;

1. **Se** Lesão da Imagem de Entrada é do tipo LCM:
  2.        Extrair o canal de cor  $R$  da imagem  $I$  para a imagem  $I_{LCM}$ ;
  3. **Senão se** Lesão da Imagem de Entrada é do tipo LF:
  4.        Extrair o canal de cor  $B$  da imagem  $I$  para a imagem  $I_{LF}$ ;
  5. **Senão se** Lesão da Imagem de Entrada é do tipo LLC:
  6.        Extrair o canal de cor  $G$  da imagem  $I$  para a imagem  $I_{LLC}$ ;
  7. **Fim**
  8. Aplicar equalização do histograma nas imagens  $I_{LCM}$ ,  $I_{LF}$  ou  $I_{LLC}$ ;
  9. Aplicar filtro Gaussiano nas imagens  $I_{LCM}$ ,  $I_{LF}$  ou  $I_{LLC}$  para criação da imagem  $I_{PP}$ .
- 

## 4.4 Segmentação

Para o desenvolvimento de uma técnica de segmentação a ser aplicada sobre imagens de linfoma, foi empregado um método baseado em limiarização. As técnicas *fuzzy 3-partition entropy* e AG foram utilizadas para a definição de valores de limiar adequados para a separação de núcleos de células neoplásicas (YIN et al., 2014).

As imagens da base utilizada contêm regiões de núcleo e citoplasma com características muito próximas, onde, segundo o especialista, a diferenciação entre essas regiões é muito delicada e pouco distingível. Porém, mesmo com tal dificuldade, áreas de citoplasma foram consideradas no processo de separação a fim de se obter melhores resultados dos contornos nucleares. Então, inicialmente, o parâmetro denominado Número de Estruturas (NE) foi definido. Neste estudo, NE foi atribuída ao valor 3, o qual representa as regiões de núcleo, citoplasma e fundo da imagem, correspondente à quantidade de classes a serem identificadas.

Um AG foi elaborado para a seleção dos valores de limiar da imagem de entrada. Na primeira etapa do AG, o histograma da imagem pré-processada foi calculado e sua normalização foi realizada para a definição da população inicial. Neste trabalho, cada população foi definida por 60 indivíduos, sendo cada um deles composto por seis variáveis relacionadas a níveis de intensidade do histograma normalizado. A Figura 18 ilustra a representação matricial de uma população, seus indivíduos e suas variáveis.

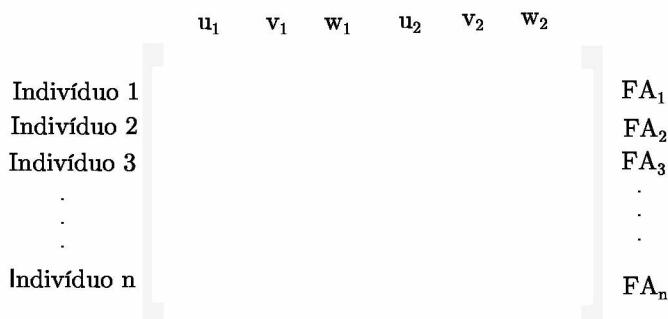


Figura 18 – Representação dos conceitos de indivíduo e população, onde cada linha representa um indivíduo, cada coluna corresponde a uma das variáveis a serem definidas e a matriz como um todo representa uma população.

Para a definição das variáveis características de cada indivíduo, foi empregada a técnica *fuzzy 3-partition entropy*. As funções  $S$  e  $Z$ , representadas pelas Equações 8 e 9, foram utilizadas para quantificar o grau de associação de cada nível de cinza ( $k$ ) a cada uma das três classes investigadas. Como o uso dessas funções permite a definição de apenas um valor de limiar, então, neste estudo, foram necessários dois pares dessas funções para a segmentação, considerando as três classes analisadas (núcleo, citoplasma e fundo da imagem). Esse procedimento permitiu determinar as variáveis  $u_1$ ,  $v_1$ ,  $w_1$ ,  $u_2$ ,  $v_2$  e  $w_2$  (YIN et al., 2014).

$$S(k, u, v, w) = \begin{cases} 1, & k \leq u \\ 1 - \frac{(k-u)^2}{(w-u) \cdot (v-u)}, & u < k \leq v \\ \frac{(k-w)^2}{(w-u) \cdot (w-v)}, & v < k \leq w \\ 0, & k > w \end{cases} \quad (8)$$

$$Z(k, u, v, w) = 1 - S(k, u, v, w). \quad (9)$$

A partir das variáveis  $u_1$ ,  $v_1$ ,  $w_1$ ,  $u_2$ ,  $v_2$  e  $w_2$ , graus de associação ( $M$ ) devem ser definidos para quantificar o quanto próximo cada nível de intensidade está de cada um dos grupos a serem identificados. O cálculo do grau de associação é representado pelas Equações 10, 11 e 12, onde cada nível de intensidade possui três graus de associação, representados por vetores:  $M_n$  relacionado ao grau de associação às regiões nucleares,  $M_c$  às regiões de citoplasma e  $M_f$  às regiões de fundo da imagem. A Figura 19 ilustra os vetores supracitados, com 256 níveis de intensidade e graus de associação relacionados a cada grupo de estruturas investigadas.

Níveis de Intensidades (k)	0	1	2	...	254	255
$M_n$	$M_{n0}$	$M_{n1}$	$M_{n2}$	...	$M_{n254}$	$M_{n255}$
$M_c$	$M_{c0}$	$M_{c1}$	$M_{c2}$	...	$M_{c254}$	$M_{c255}$
$M_f$	$M_{f0}$	$M_{f1}$	$M_{f2}$	...	$M_{f254}$	$M_{f255}$

Figura 19 – Representação obtida pela aplicação das funções de associação,  $M_n$ ,  $M_c$  e  $M_f$ , sobre cada nível de intensidade da imagem.

$$M_n(k) = S(k, u_1, v_1, w_1), \quad (10)$$

$$M_c(k) = \begin{cases} Z(k, u_1, v_1, w_1), & k \leq w_1, \\ S(k, u_2, v_2, w_2), & k > w_1, \end{cases} \quad (11)$$

$$M_f(k) = Z(k, u_2, v_2, w_2), \quad (12)$$

onde,  $k$  representa intensidades de brilho da imagem, neste trabalho,  $0 \leq k \leq 255$ ; e  $u_1$ ,  $v_1$ ,  $w_1$ ,  $u_2$ ,  $v_2$  e  $w_2$ , onde  $0 \leq u_1 < v_1 < w_1 < u_2 < v_2 < w_2 \leq 255$ , são parâmetros que determinam a distribuição dos graus de associação de cada nível de intensidade. A Figura 20 ilustra a relação entre o grau de associação de cada função e cada um dos níveis de brilho. Nesse exemplo, é possível perceber que níveis de intensidade compreendidos entre  $u_1$  e o ponto de intersecção das curvas  $M_n$  e  $M_c$  possuem maior grau de associação com o conjunto escuro, que é característico da representação das estruturas nucleares. Os pixels com níveis de intensidade contidos no intervalo definido pelos pontos de intersecção das curvas são associados à região de citoplasma já que possuem maior grau de associação a

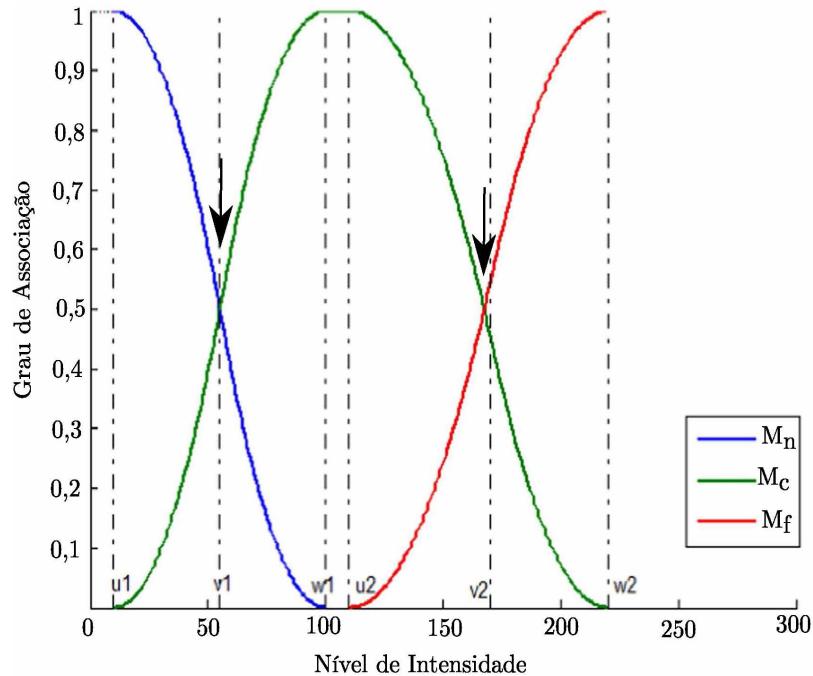


Figura 20 – Gráfico representativo das funções de associação utilizando o método *fuzzy 3-partition entropy* com indicação dos valores de limiar por setas de cor preta (Figura adaptada de (YIN et al., 2014) © 2014 Elsevier).

esse grupo. Por fim, níveis de intensidade contidos no intervalo do ponto de intersecção das curvas  $M_c$  e  $w_2$  são associados às regiões de fundo e outras informações irrelevantes.

Em seguida, os valores determinados por cada indivíduo foram analisados por meio da função de avaliação baseada no cálculo da entropia, definida pela Equação 13:

$$H(u_1, v_1, w_1, u_2, v_2, w_2) = -P_n \cdot \log(P_n) - P_c \cdot \log(P_c) - P_f \cdot \log(P_f), \quad (13)$$

onde, os valores de probabilidade ( $P$ ) de cada uma das estruturas são obtidas por:

$$P_n = \sum_{k=0}^{255} h(k) \cdot M_n(k), \quad (14)$$

$$P_c = \sum_{k=0}^{255} h(k) \cdot M_c(k), \quad (15)$$

$$P_f = \sum_{k=0}^{255} h(k) \cdot M_f(k), \quad (16)$$

onde,  $h(\cdot)$  representa o histograma normalizado da imagem. Nessa etapa, busca-se maximizar o valor de  $H$ , indicando maior quantidade de informação extraída e permitindo a definição da melhor combinação de parâmetros  $(u_1, v_1, w_1, u_2, v_2, w_2)$  que determinam a melhor separação entre as estruturas analisadas por meio dos valores de limiar definidos. Esses valores são representados pela Figura 20 com setas de cor preta nos pontos

de intersecção das funções  $M_n$ ,  $M_c$  e  $M_f$ . Tais valores de limiar também podem ser obtidos considerando os graus de associação, onde ambos os valores correspondem àqueles com grau de associação igual a 0,5 (YIN et al., 2014). Esse processo foi empregado em toda a população, e o valor de  $H$  foi armazenado com os parâmetros de cada indivíduo. A Figura 18 ilustra uma população com seus valores obtidos pela função de avaliação representados pelas variáveis  $FA_i$ .

Na etapa de seleção, adotou-se o processo de elitismo, onde, considerando os valores obtidos pela função de avaliação, apenas determinada quantidade dos melhores indivíduos da população é mantida na próxima geração (PAULINAS; UŠINSKAS, 2007). Nessa etapa, a seleção foi realizada a partir de 30% da população, ou seja, dos 60 indivíduos que compõem cada população, apenas 18 foram considerados em gerações subsequentes. A Figura 21 ilustra um exemplo da execução da seleção, com a representação da população inicial e os valores obtidos em suas funções de avaliação, bem como o resultado da seleção sobre essa população. Nesse exemplo, os indivíduos 2 e 6 possuem os maiores valores de entropia, e assim foram mantidos na nova geração, representando 30% da população total.

	$u_1$	$v_1$	$w_1$	$u_2$	$v_2$	$w_2$	$FA$		$u_1$	$v_1$	$w_1$	$u_2$	$v_2$	$w_2$	$FA$
Indivíduo 1	1	50	75	150	210	250	20	Indivíduo 6	45	93	125	179	224	254	70
<b>Indivíduo 2</b>	25	80	115	170	230	246	<b>50</b>	Indivíduo 2	25	80	115	170	230	246	50
Indivíduo 3	13	62	98	164	218	253	10	...	...	...	...	...	...	...	...
Indivíduo 4	28	79	102	158	245	249	15	...	...	...	...	...	...	...	...
Indivíduo 5	32	73	108	167	231	238	35	...	...	...	...	...	...	...	...
<b>Indivíduo 6</b>	45	93	125	179	224	254	<b>70</b>	...	...	...	...	...	...	...	...

(a)

(b)

Figura 21 – Representação matricial do processo de seleção. Em (a), uma população inicial é representada juntamente com os valores obtidos pela aplicação da função de avaliação ( $FA$ ) de cada indivíduo. Em (b), o resultado do processo de seleção é ilustrado contendo apenas 30% dos indivíduos da população anterior.

Após a seleção dos melhores indivíduos, a etapa de *crossover* foi aplicada para a combinação de tais indivíduos utilizando as definições de probabilidade de *crossover* e CP. Para a definição da probabilidade de *crossover*, uma probabilidade foi gerada aleatoriamente para cada indivíduo selecionado na etapa anterior. Caso esse valor fosse menor que a probabilidade de *crossover*, o indivíduo seria selecionado como um dos pais dos descendentes a serem gerados. Esse processo gerou continuamente probabilidades para os melhores indivíduos até que dois deles fossem selecionados para combinação (LIN; LEE; HONG, 2003). Neste trabalho, tal probabilidade foi atribuída ao valor 0,65, definido empiricamente com testes realizados sobre a população considerando valores contidos no intervalo de 0,5 à 1,0.

Neste trabalho, o valor do CP foi selecionado entre os parâmetros  $u_1, v_1, w_1, u_2, v_2$  e  $w_2$ , variando de 1 à 6, para a geração de dois descendentes. Então, o primeiro descendente foi determinado pelos parâmetros do indivíduo 1 compreendidos entre as posições 1 e CP e pelas informações do indivíduo 2 compreendidas entre  $(CP + 1)$  e 6. Já o segundo descendente foi formado pelos parâmetros do indivíduo 2 contidos entre as posições 1 e CP e pelas informações do indivíduo 1 contidas no intervalo entre  $(CP + 1)$  e 6.

Na etapa de mutação, probabilidades foram geradas aleatoriamente para cada parâmetro de cada indivíduo até satisfazer a condição de se obter um valor inferior ao da probabilidade de mutação (MCCALL, 2005). Neste trabalho, essa probabilidade assumiu valor igual a 0,01 para que os novos valores de atributos gerassem soluções semelhantes entre si a cada execução (LAD; AGRAWAL; PANDYA, 2014). A Figura 22 ilustra um exemplo desse processo executado sobre um indivíduo da população.

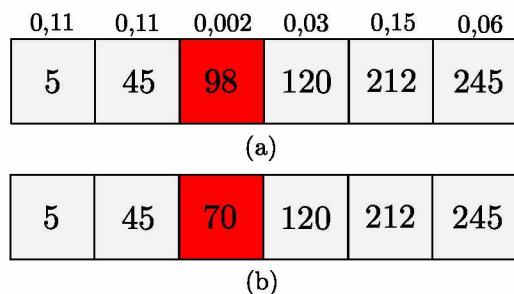


Figura 22 – Exemplo de um processo de mutação, onde, em (a), a probabilidade de mutação do terceiro parâmetro é menor que o limiar definido (0,01) o que leva à alteração de seu valor de forma aleatória, representado em (b).

Finalmente, o critério de parada é definido para se obter o melhor resultado. Neste estudo, o algoritmo de Hammouche *et al.* (HAMMOUCHE; DIAF; SIARRY, 2008) foi utilizado já que este é capaz de obter a convergência do AG em poucas iterações (Algoritmo 2). Nesse caso, a variável  $t$  foi atribuída a dois valores considerando as três estruturas analisadas nas imagens de linfoma. O processo é interrompido quando um dos valores armazenados por  $t$  permanece o mesmo ao longo de duas gerações consecutivas.

A Figura 23 representa um exemplo de uma condição suficiente para a finalização do algoritmo. Na Figura 23(a), é apresentado o resultado obtido em uma geração e a Figura 23(b) ilustra o resultado obtido na geração subsequente. A média dos valores entre as classes 1 e 2 ( $t_1$ ), e 2 e 3 ( $t_2$ ) são também apresentadas. Como as médias entre as classes 2 e 3 foram as mesmas em ambas as gerações, resultando no valor 250, o algoritmo é finalizado.

O Algoritmo 3 apresenta a sequência metodológica dessa etapa.

---

**Algoritmo 2** Critério de parada
 

---

1. Calcular a média dos níveis de intensidade  $m_i(s)$  da estrutura  $C_i$ , onde  $i$  representa o número de estruturas e  $s$  corresponde à geração executada;
  2. Atualizar o valor de  $t_i(s)$  de acordo com a seguinte equação:
- $$t_i(s+1) = \frac{m_i(s) + m_{i+1}(s)}{2}; \quad (17)$$
3. Repita as etapas 1 e 2 até que a iteração estabilize-se, onde  $t_i(s+1) = t_i(s)$ ,  $i = 1, 2, \dots, k-1$ .
- 

---

**Algoritmo 3** Segmentação
 

---

**Entrada:** Imagem pré-processada  $I_{PP}$ ;

**Saída:** Imagem binária  $I_{bin}$ ;

1. Criação aleatória de 60 indivíduos caracterizados por seis atributos para formação da população inicial;
  2. Cálculo dos graus de associação considerando os atributos de cada indivíduo;
  3. Cálculo dos valores de entropia (função de avaliação) de cada indivíduo;
  4. **Enquanto** Critério de parada proposto por (HAMMOUCHE; DIAF; SIARRY, 2008) não é alcançado **faça**
  5. Seleção de 30% dos indivíduos que obtiveram os maiores valores de entropia;
  6. Execução da etapa de *crossover* sobre esses indivíduos;
  7. Execução da etapa de mutação sobre a população;
  8. Cálculo do grau de associação de cada indivíduo;
  9. Cálculo dos valores de entropia (função de avaliação) de cada indivíduo;
  10. Seleção do melhor indivíduo da geração obtida;
  11. Cálculo dos valores de limiar desse indivíduo;
  12. **Fim**
  13. Binarização da imagem  $I_{PP}$  considerando a separação de regiões de núcleo das regiões de citoplasma e fundo da imagem para formação da imagem  $I_{bin}$ .
-

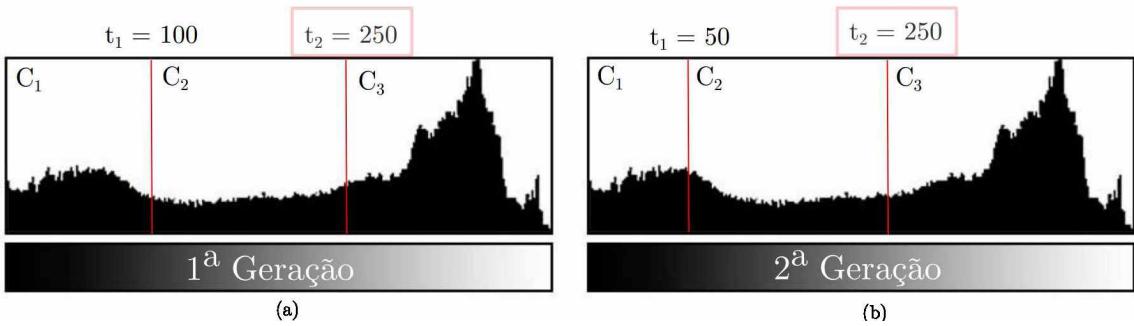


Figura 23 – Exemplo da aplicação do Algoritmo 2 para a definição do critério de parada ao longo de duas gerações ((a) e (b)) por meio da representação do histograma de uma imagem.

## 4.5 Pós-processamento

Para refinar as estruturas segmentadas e eliminar regiões falso positivas, inicialmente, os valores de área das regiões segmentadas foram calculados e regiões com áreas inferiores à determinado valor  $n$  foram eliminadas. Por meio de uma avaliação empírica,  $n$  foi atribuído ao valor 10 pixels, o qual representa o tamanho de pequenos ruídos presentes nas imagens de linfoma segmentadas.

Também foi possível notar que o processo de segmentação não foi capaz de separar regiões nucleares entre si. As regiões intra-nucleares eram caracterizadas por níveis de intensidade mais claros que os núcleos que as formavam, resultando em um único objeto. Um exemplo de região intra-nuclear é ilustrado pela Figura 24. Na Figura 24(a), é apresentado o resultado da segmentação sobre uma imagem original com a eliminação de regiões menores que 10 pixels. O objeto indicado pelo círculo vermelho destaca a identificação de dois núcleos unidos em uma única região. Na Figura 24(b), é apresentada a imagem correspondente à anterior com a mesma região em destaque porém com seus níveis de intensidade expostos, onde é possível notar que a região intra-nuclear possui níveis de intensidade mais claros em relação aos dois núcleos que formam o objeto segmentado. Neste estudo, regiões com tamanho maior que 100 pixels foram consideradas para refinamento nessa etapa. Esse valor foi definido analisando o tamanho médio das regiões marcadas pelo especialista para estruturas verdadeiro positivas.

Então, após a identificação dos objetos segmentados contendo mais de um núcleo, as regiões intra-nucleares deveriam ser eliminadas. Para tanto, o algoritmo *valley-emphasis* (NG, 2006) foi selecionado para aplicação sobre as regiões da imagem pré-processada com as técnicas definidas no item 4.3.

O método *valley-emphasis* objetiva determinar o valor de limiar por meio da região de vale presente no histograma. Para encontrar as regiões de vale, esse método considera a menor probabilidade de ocorrência ( $p_t$ ) de determinado nível de intensidade. Além disso, a variância entre as classes consideradas é maximizada, de acordo com o segundo termo

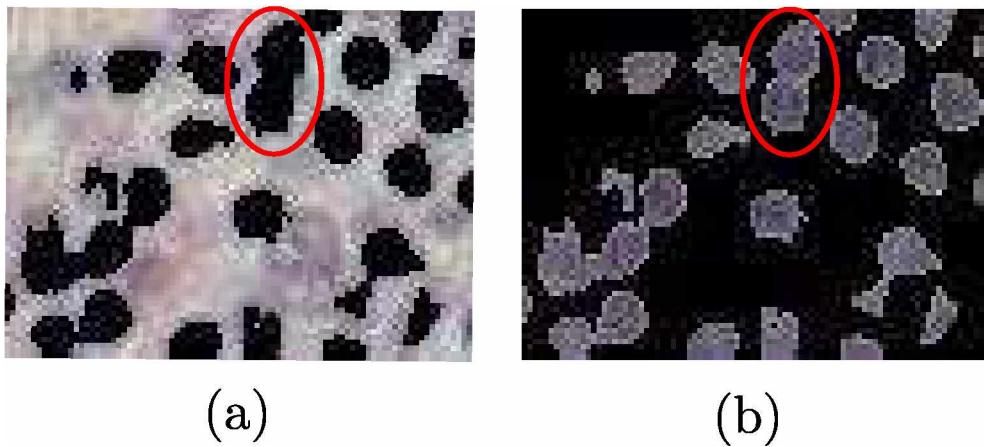


Figura 24 – Exemplo ilustrativo de regiões intra-nucleares identificadas pela segmentação na imagem original (a) e sua representação ampliada (b) para percepção das características de cor identificadas como parte integrante de dois núcleos segmentados (c).

da Equação 18, conforme definido pelo método de Otsu (OTSU, 1975):

$$T = \text{Max}\{(1 - p_t) \cdot (\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t))\}, \quad (18)$$

onde,  $\omega$  representa a probabilidade de ocorrência das classes consideradas e  $\mu$  corresponde à média dos níveis de intensidade de cada classe. Ao maximizar o primeiro termo, a menor probabilidade entre os níveis de intensidade será encontrada. Esse termo assegura que o valor de limiar esteja localizado na região de vale da distribuição do histograma.

Por fim, as operações de dilatação e abertura foram aplicadas sobre a imagem segmentada pelo método *valley-emphasis* com o objetivo de se obter o resultado final do processo de segmentação não supervisionado. A dilatação permite o aumento da área dos objetos representados e o preenchimento de seus espaços nulos. A Figura 25 ilustra a execução dessa operação com um elemento estruturante com distribuição em disco com raio igual a 2. A Figura 25(a) apresenta a imagem binária, onde regiões segmentadas são representadas por pixels com intensidade em preto e na Figura 25(b), é apresentado seu mapeamento sobre a imagem original. Também é apresentado seu elemento estruturante (Figura 25(c)), a imagem binária resultante de sua aplicação (Figura 25(d)) e sua identificação correspondente sobre a imagem original (Figura 25(e)).

A abertura objetiva aumentar espaços entre objetos, suavizar contornos e eliminar ruídos. A Figura 26 ilustra a execução desse processamento utilizando os objetos representados em uma imagem binária (Figura 26(a)) e seus correspondentes identificados na imagem original (Figura 26(b)), onde é possível notar objetos unidos por finas ligações e pequenos ruídos, utilizando um elemento estruturante de tamanho  $3 \times 3$  pixels com distribuição quadrada (Figura 26(c)). Os objetos resultantes da abertura são apresentados pelas Figuras 26(d) e 26(e).

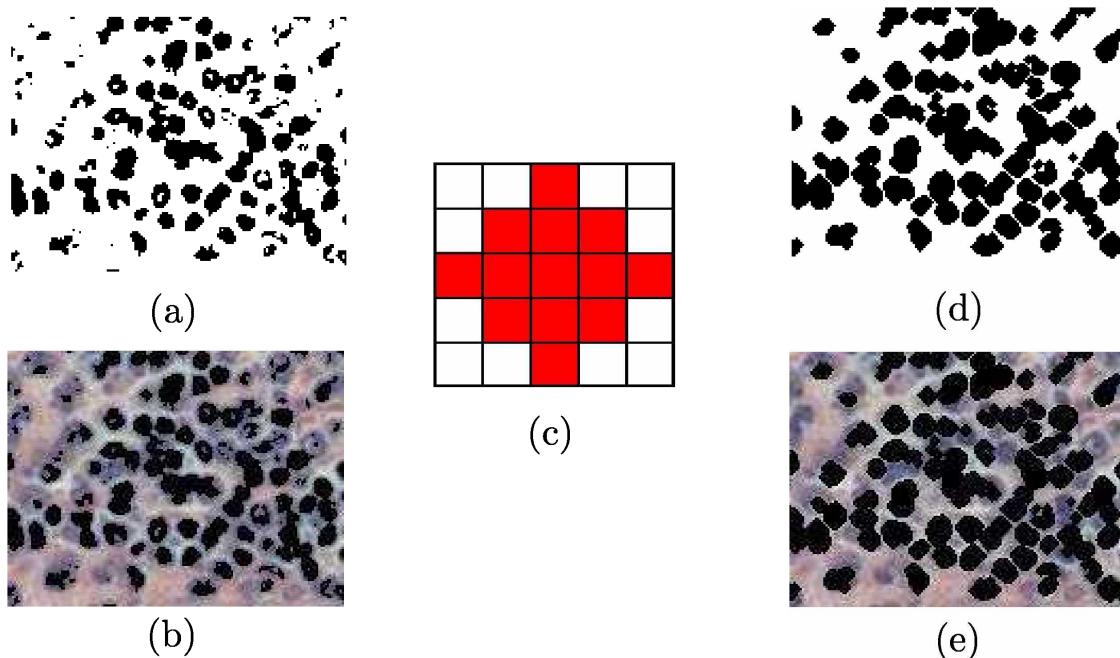


Figura 25 – Exemplo da operação de dilatação binária: imagem binária (a), suas regiões correspondentes identificadas sobre a imagem original (b) e o resultado de sua dilatação representado por uma imagem binária (c) e pela imagem original (e) utilizando um elemento estruturante com distribuição em disco com raio igual a 2 (c).

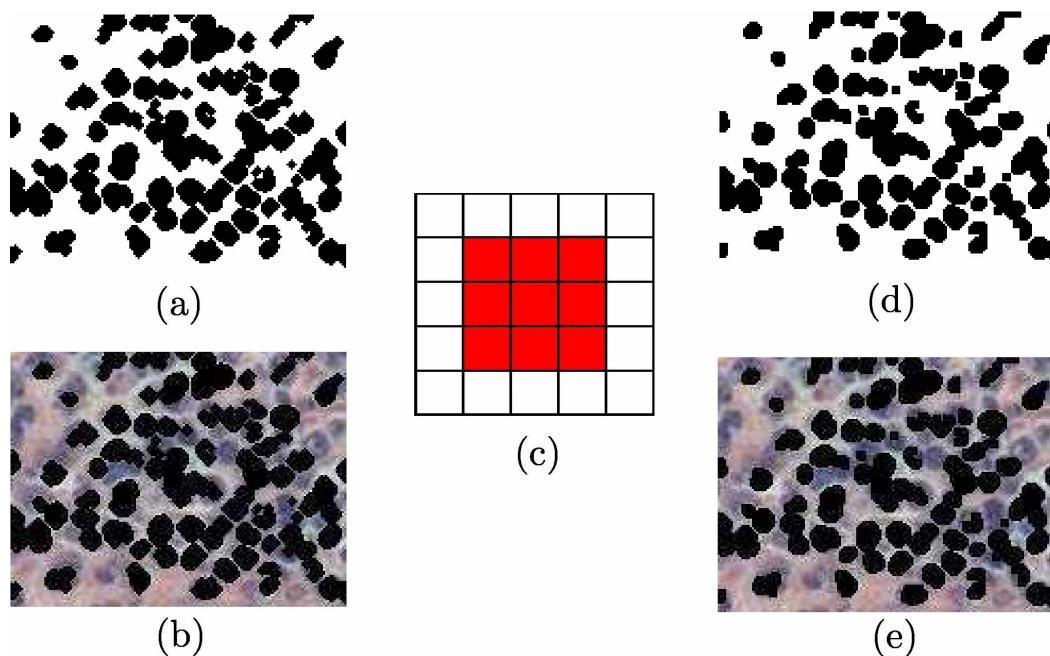


Figura 26 – Exemplo da operação de abertura binária: imagem binária (a), imagem original (b) e o resultado de sua abertura representado por uma imagem binária (d) e seu mapeamento sobre a imagem original (e) obtido por um elemento estruturante de tamanho  $3 \times 3$  com distribuição formando um quadrado (c).

O Algoritmo 4 apresenta a compilação dos métodos descritos e utilizados nessa etapa.

---

#### **Algoritmo 4** Pós-processamento

---

**Entrada:** Imagem binária  $I_{bin}$ ;

**Saída:** Imagem pós-processada correspondente ao resultado do método proposto  $I_{PosP}$ ;

1. Eliminação de regiões com áreas menores que 10 pixels;
  2. Mapeamento das regiões identificadas na segmentação sobre as imagens pré-processadas ( $I_{PP}$ );
  3. Aplicação do método *valley-emphasis* sobre essas regiões com áreas maiores que 100 pixels;
  4. Aplicação da operação morfológica de dilatação com elemento estruturante distri-buído no formato disco de tamanho 2;
  5. Aplicação da operação morfológica de abertura com elemento estruturante distri-buído no formato quadrado de tamanho 2 para obtenção da imagem  $I_{PosP}$ .
- 

## 4.6 Métricas de Avaliação Quantitativa

A fim de definir o modelo de cores e filtros aplicados sobre as imagens, na etapa de pré-processamento, além da análise visual também foi empregada a métrica baseada no cálculo da entropia. Essa medida foi adotada considerando que o baixo valor da entropia indica alta homogeneidade e, consequentemente, a presença de poucos ruídos nas imagens, os quais poderiam representar problemas na etapa de segmentação (TSAI; LEE; MAT-SUYAMA, 2008). Além disso, a mínima entropia também indica o desempenho eficaz de filtros capazes de lidar com desigualdade de iluminação que também podem levar a complicações na segmentação (LEONG; BRADY; MCGEE, 2003). Assim, o baixo valor da entropia corresponde à situação desejada para imagens resultantes do pré-processamento. Para tanto, a entropia foi empregada para quantificar o grau de homogeneidade das imagens da base por meio das Equações 19 e 20 (BELKACEM-BOUSSAID et al., 2010):

$$E = \sum_{i=1}^N p_i \cdot \log_2(p_i), \quad (19)$$

$$p = \frac{H_t(I(p, q))}{\sum_{t=1}^N H_t(I(p, q))}, \quad (20)$$

onde,  $H$  representa o histograma da imagem,  $N$  é o número total de níveis de cinza,  $t$  é o índice que define o nível de intensidade considerado, e  $i$  é o índice que demonstra a variação do histograma.

Diferentes métricas podem ser utilizadas para avaliar a segmentação de imagens médicas, as quais são capazes de quantificar o quanto próximos são os resultados obtidos por

meio das segmentações não supervisionada e manual (GHOSE et al., 2012; ESTRADA; JEPSON, 2009). Neste trabalho, foram utilizadas cinco métricas para essas avaliações: acurácia, sensibilidade, especificidade (INSANA; MEYERS; GROSSMAN, 2000), *dice index* e *variation of information* (WU et al., 2015).

A partir da comparação entre regiões obtidas na segmentação não supervisionada e regiões demarcadas por um especialista, é possível obter quatro classificações possíveis para os pixels dessas regiões: verdadeiro positivos ( $V_P$ ), correspondentes ao número de pixels detectados corretamente, os verdadeiro negativos ( $V_N$ ) que representam a quantidade de pixels não detectados corretamente, os falso positivos ( $F_P$ ) que apresentam o número de pixels incorretamente detectados, e os falso negativos ( $F_N$ ) que denotam pixels incorretamente não detectados (CHANG et al., 2014). A Figura 27 ilustra a relação entre esses conceitos e regiões detectadas de formas manual e não supervisionada.

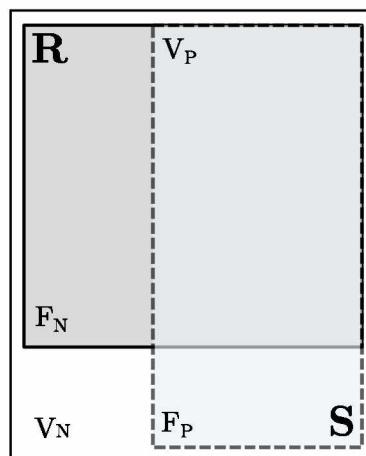


Figura 27 – Relação entre regiões identificadas manualmente (R) e automaticamente (S) (OGER; BELHOMME; GURCAN, 2012).

A partir desses conceitos, as métricas de sensibilidade ( $Se$ ), especificidade ( $Es$ ) e acurácia ( $Ac$ ) podem ser definidas, respectivamente, pelas Equações (21), (22) e (23). A sensibilidade reflete a quantidade de pixels detectados manualmente que também foram corretamente segmentados pelo sistema proposto. A especificidade permite calcular a quantidade de pixels que não foram detectados manualmente assim como foram corretamente excluídos na segmentação automática (OGER; BELHOMME; GURCAN, 2012). Já a acurácia corresponde a taxa de pixels corretamente identificados em toda a área da ROI (BYRD; ZENG; CHOUIKHA, 2007).

$$Se = \frac{V_P}{V_P + F_N} \quad (21)$$

$$Es = \frac{V_N}{V_N + F_P} \quad (22)$$

$$Ac = \frac{V_P + V_N}{V_P + V_N + F_P + F_N} \quad (23)$$

A métrica *dice index* (*DI*) é definida como a razão entre o tamanho da intersecção e o tamanho das regiões segmentadas automaticamente (*S*) e marcadas manualmente (*R*). Tal operação é expressa pelo cálculo a seguir:

$$DI(S, R) = \frac{2|S \cap R|}{|S| + |R|} \quad (24)$$

A métrica *variation of information* (*VoI*) é utilizada para quantificar a distância entre *S* e *R*, como demonstrado pela Equação 25, onde *H(X)* representa a entropia de *X* e *I(X, Y)* corresponde a informações mútuas entre *X* e *Y*.

$$VoI(S, R) = H(R) + H(S) - 2I(R, S) \quad (25)$$

Para expressar maior similaridade entre as segmentações automática e manual, altos valores de *DI* e baixos valores de *VoI* são requeridos (WU et al., 2015). Pelas equações (21), (22) e (23), é possível notar que, assim como a métrica *DI*, altos valores obtidos por meio desses cálculos indicam melhores resultados da segmentação automática.

Além disso, o método proposto foi avaliado considerando etapas de extração de características e classificação empregadas em sistemas CAD. Para tanto, neste estudo, foi selecionado o trabalho de (DONG et al., 2014) que propõe o uso de descritores de intensidade extraídos de diferentes canais de modelos de cores para a classificação de imagens histológicas. Assim, os descritores de características foram analisados considerando a presença e ausência da etapa de segmentação utilizando o classificador SVM, como descrito por (OZTAN et al., 2012). Considerando a execução desse classificador para a separação de amostras entre duas classes, as combinações foram: LCM × LF, LCM × LLC e LF × LLC.

As regiões identificadas pela aplicação das técnicas descritas foram mapeadas sobre os canais *R*, *G* e *B*, e imagens em níveis de cinza obtidas das imagens originais. A extração de nove características foi realizada sobre cada uma dessas imagens, totalizando 36 características a serem avaliadas pelo classificador, sendo elas: média, mediana, desvio padrão, curtose, assimetria, variância, 1-norma euclidiana, 2-norma euclidiana e entropia.

# CAPÍTULO 5

## Experimentos e Análise dos Resultados

### 5.1 Pré-processamento

Por meio da utilização das técnicas descritas no item 4.3, os resultados obtidos são ilustrados pelas Figuras 28, 29 e 30. As Figuras 28(a), 29(a) e 30(a) ilustram as imagens originais das classes LCM, LF e LLC, respectivamente. Nas Figuras 28(b), 29(b) e 30(b), são apresentados os canais  $R$ ,  $B$  e  $G$  extraídos de suas imagens correspondentes das lesões LCM, LF e LLC, respectivamente. Por fim, as Figuras 28(c), 29(c) e 30(c) representam os resultados finais da etapa de pré-processamento de cada lesão processadas pela equalização do histograma e o filtro Gaussiano.

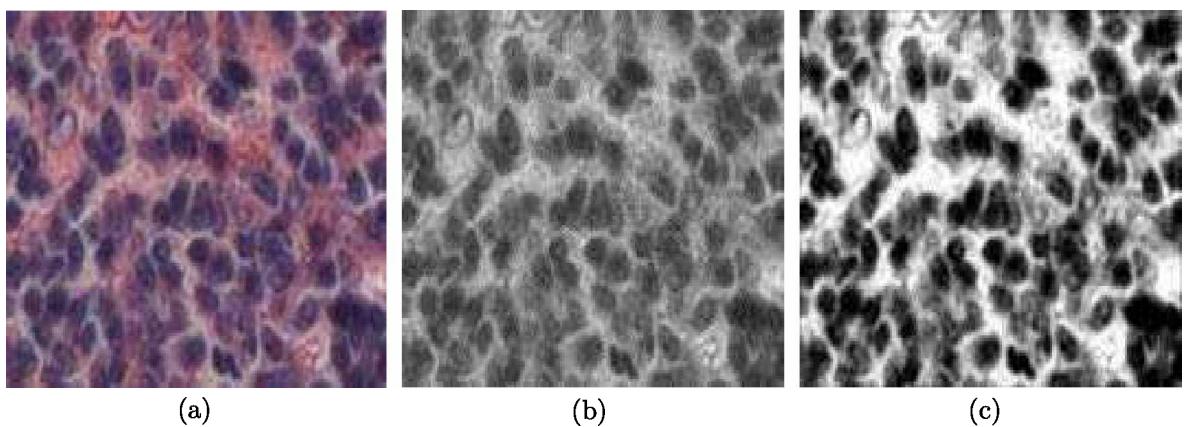


Figura 28 – Exemplo da aplicação de pré-processamento sobre imagens da classe LCM: (a) imagem original, (b) seu canal  $R$  extraído e (c) resultado da aplicação das técnicas de pré-processamento que alcançaram entropia mínima.

Os canais analisados para a aplicação desses métodos foram os canais dos modelos de cores RGB, HSV, LAB, LUV, YCbCr e YIQ. Também foram analisadas imagens resultantes do processo de deconvolução (RUIFROK; JOHNSTON, 2001). Considerando que a melhor representação das regiões nucleares ocorre na imagem resultante da deconvolução da hematoxilina, esta foi utilizada nessa comparação. A Tabela 2 ilustra os valores de

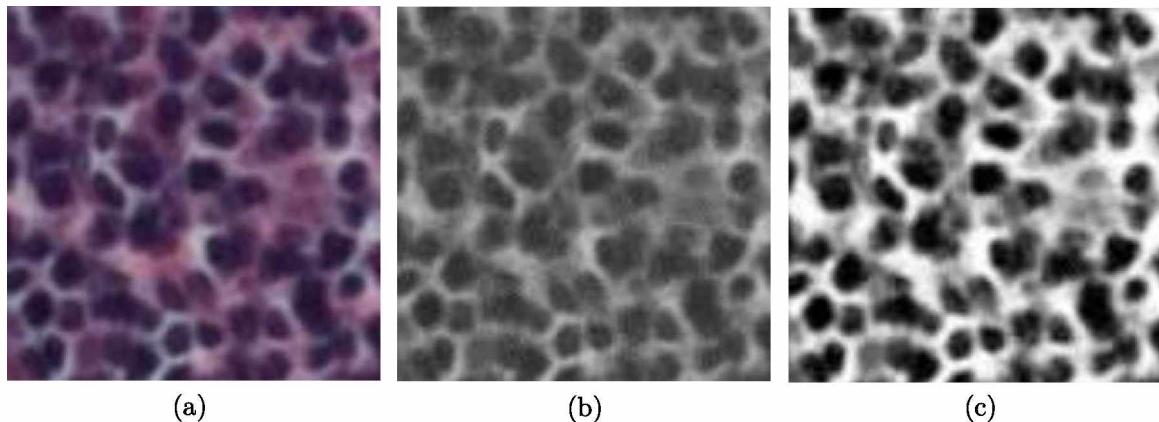


Figura 29 – Exemplo da aplicação das técnicas equalização do histograma e filtro Gaussiano sobre o canal  $B$  das imagens da classe LF: (a) imagem original, (b) seu canal  $B$  e (c) aplicação das técnicas de pré-processamento supracitadas.

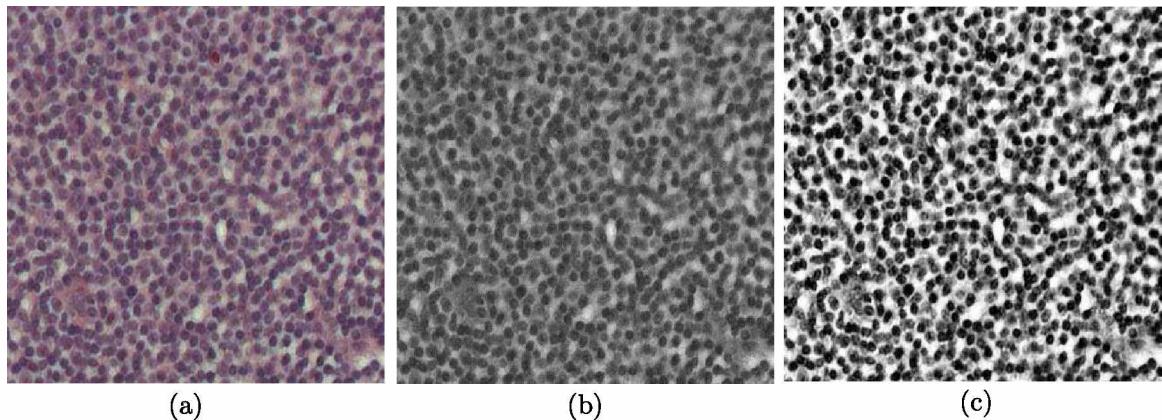


Figura 30 – Exemplo da aplicação dos métodos de pré-processamento descritos sobre imagens da classe LLC: (a) imagem original, (b) seu canal  $G$  e (c) aplicação das técnicas de equalização do histograma e filtro Gaussiano.

entropia obtidos pela aplicação da equalização do histograma e do filtro Gaussiano sobre cada canal dos modelos de cores citados.

Pela análise da Tabela 2, é possível notar que os mínimos valores de entropia entre todos os canais de cores, representados em negrito, foram obtidos pelos canais  $R$ ,  $B$  e  $G$  para as lesões LCM, LF e LLC, respectivamente. A entropia é utilizada para quantificar o grau de homogeneidade das imagens (BELKACEM-BOUSSAID et al., 2010). Seu valor mínimo na etapa de pré-processamento é um indicativo de que filtros aplicados permitiram a obtenção de imagens com poucos ruídos (TSAI; LEE; MATSUYAMA, 2008) e sem desigualdades de iluminação (LEONG; BRADY; MCGEE, 2003). Tais características levaram à utilização dessa métrica para avaliação dessa etapa, assim como foi realizado por (BELKACEM-BOUSSAID et al., 2010). Além disso, a avaliação visual também foi empregada ao considerar pequenas diferenças entre os valores de entropia dos canais analisados.

Tabela 2 – Valores de entropia obtidos pela aplicação dos métodos de pré-processamento utilizados sobre cada canal dos modelos de cores RGB, HSV, LAB, LUV, YCbCr, YIQ e Hematoxilina.

Canais e seus Modelos de Cores	LCM	LF	LLC
<i>R</i> (RGB)	<b>-7,9830</b>	-7,7889	-7,7922
<i>G</i> (RGB)	-7,4814	-7,7889	<b>-7,9900</b>
<i>B</i> (RGB)	-7,4822	<b>-8,0021</b>	-7,7914
<i>H</i> (HSV)	-7,7969	-7,7505	-7,7155
<i>S</i> (HSV)	-7,7753	-7,7981	-7,7847
<i>V</i> (HSV)	-7,7816	-7,7889	-7,7910
<i>L</i> (LAB)	-7,7823	-7,7954	-7,4909
<i>a</i> (LAB)	-7,4377	-7,8108	-7,5604
<i>b</i> (LAB)	-7,4539	-7,8081	-7,4346
<i>L</i> (LUV)	-7,0504	-7,6281	-7,0308
<i>U</i> (LUV)	-6,8782	-7,5759	-6,8558
<i>V</i> (LUV)	-7,6759	-7,6906	-7,4842
<i>Y</i> (YCbCr)	-7,7813	-7,7949	-7,7702
<i>Cb</i> (YCbCr)	-7,3814	-7,5534	-7,3676
<i>Cr</i> (YCbCr)	-7,2217	-7,5603	-7,2570
<i>Y</i> (YIQ)	-7,6817	-7,7949	-7,6907
<i>I</i> (YIQ)	-5,9798	-6,5765	-5,1639
<i>Q</i> (YIQ)	-7,2159	-7,6344	-7,4339
Hematoxilina	-7,6084	-7,7884	-7,4469

Após a escolha das técnicas de pré-processamento para aplicação sobre os diferentes tipos de lesões, foi necessário definir os parâmetros do filtro Gaussiano, que são o valor de  $\sigma$  e o tamanho de sua máscara. A variável sigma foi analisada com atribuição a valores contidos no intervalo [0,5, 5,0], enquanto o tamanho da máscara desse filtro recebeu as atribuições  $3 \times 3$ ,  $5 \times 5$  e  $7 \times 7$  pixels. A Figura 31 apresenta uma série de experimentos realizados, onde as Figuras 31(a), 31(e), 31(i) e 31(m) apresentam o canal *R* de uma mesma imagem para melhor visualização e comparação dos resultados apresentados ao longo das linhas dessa representação. As Figuras 31(b), 31(c) e 31(d) apresentam os resultados obtidos pela aplicação das máscaras de tamanhos  $3 \times 3$ ,  $5 \times 5$  e  $7 \times 7$  pixels, respectivamente, com a variável sigma atribuída ao valor 0,5. As Figuras 31(f), 31(g) e 31(h) são resultados da atribuição do valor 2 à sigma utilizando máscaras  $3 \times 3$ ,  $5 \times 5$  e  $7 \times 7$ , respectivamente. As Figuras 31(j)-(l) e 31(n)-(p) representam as imagens resultantes da atribuição dos valores 3,5 e 5 à variável sigma da função Gaussiana bidimensional, respectivamente.

Pela análise das imagens apresentadas pela Figura 31, é possível observar que quanto maior o tamanho da máscara, menor é o nível de nitidez da imagem resultante, como pode ser comprovado pela comparação entre as Figuras 31(f) e 31(h). O menor nível de nitidez, apresentado pelas Figuras 31(h), 31(l) e 31(p), leva à menor distinção entre as regiões de citoplasma e núcleo das células, e, consequentemente, a resultados insatisfatórios na

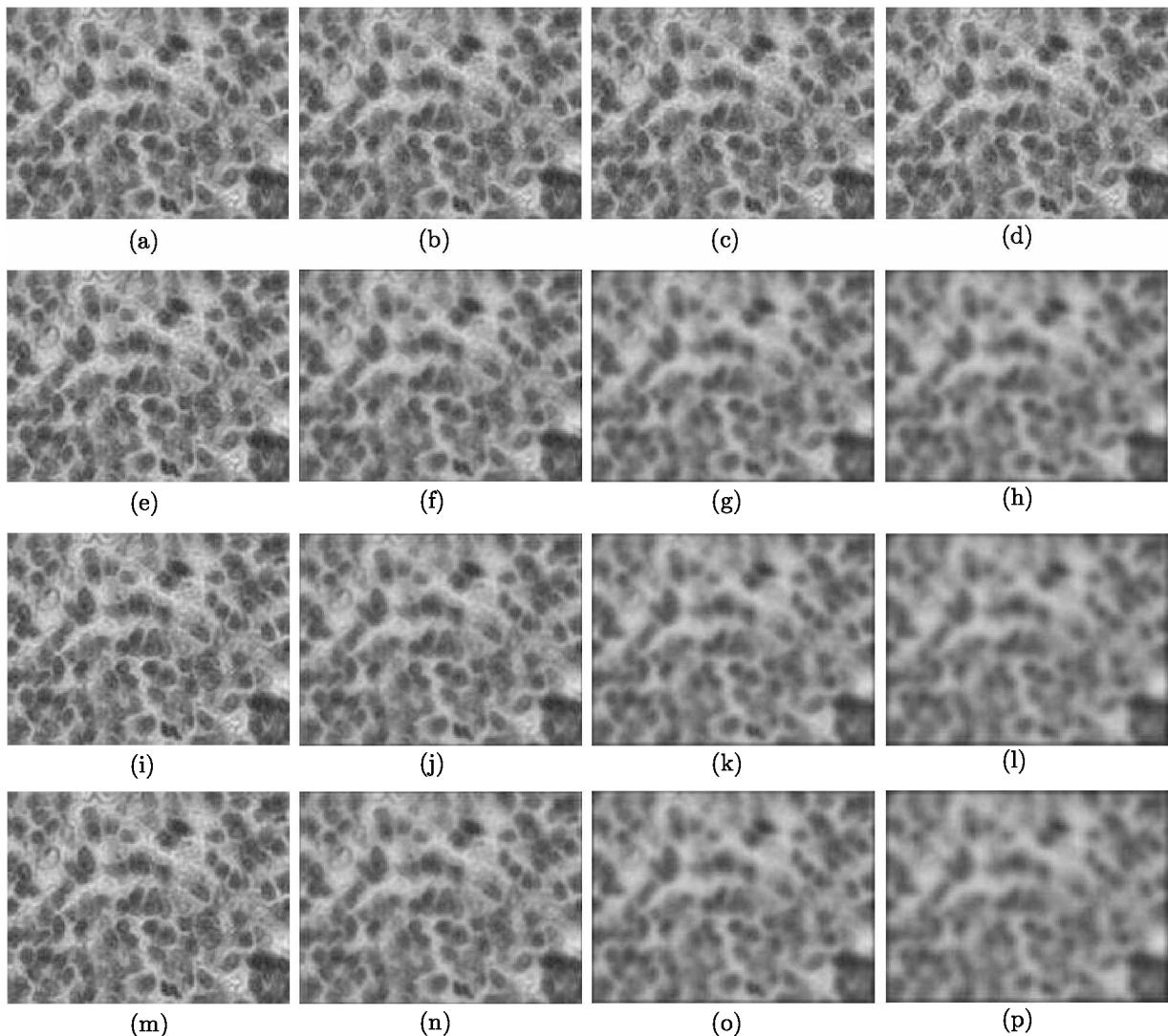


Figura 31 – Exemplo da aplicação do filtro Gaussiano sobre uma subimagem do caso *sj-04-4525-R4\_001* da classe LCM, considerando diferentes tamanhos de sua máscara e atribuições à sigma: (a), (e), (i), (m) canal *R* da imagem original, aplicações com máscaras de tamanho  $3 \times 3$  ((b), (f), (j), (n)),  $5 \times 5$  ((c), (g), (k), (o)) e  $7 \times 7$  ((d), (h), (l), (p)) pixels, com atribuições à variável sigma de 0,5 ((b)-(d)), 2 ((f)-(h)), 3,5 ((j)-(l)) e 5 ((n)-(p)).

segmentação. Essa característica também é expressa pelas Figuras 31(g), 31(k) e 31(o), o que levou à escolha do tamanho  $3 \times 3$  pixels para a máscara do filtro Gaussiano. Tal definição permitiu a eliminação de pequenos ruídos, porém não reforçou a diferenciação complexa entre as regiões de núcleo e citoplasma. A atribuição do valor sigma mais relevante para a aplicação proposta foi o valor 2, representada pela Figura 31(f). A atribuição de valor 0,5 não levou a diferenças significativas entre os diferentes tamanhos da máscara e a imagem original, como observado pelas Figuras 31(a)-(d). Em comparação com as atribuições de 3,5 e 5, o valor 2 promove melhores distinções entre as regiões nucleares e informações circundantes a elas, justificando sua escolha.

## 5.2 Segmentação

Para ilustrar a execução do AG, a Figura 32 é apresentada. Na Figura 32(a), é apresentada uma população inicial, criada aleatoriamente, composta por seus seis parâmetros e o valor resultante da aplicação da função de avaliação sobre cada indivíduo. Os indivíduos em destaque representam as melhores soluções obtidas nessa geração de acordo com o cálculo da maximização da entropia. A Figura 32(b) apresenta a geração posterior após a aplicação das etapas de seleção, *crossover* e mutação. Nessa figura, ainda estão destacados os indivíduos selecionados no processo de seleção.

A aplicação do método proposto na etapa de segmentação permitiu a identificação de núcleos celulares em imagens histológicas de linfoma. As Figuras 33(a), 34(a) e 35(a) apresentam imagens originais das classes LCM, LF e LLC, respectivamente, para exemplificar essa aplicação. As Figuras 33(b), 34(b) e 35(b) apresentam as imagens resultantes da etapa de pré-processamento de cada classe, e as Figuras 33(c), 34(c) e 35(c) ilustram os histogramas das imagens pré-processadas, bem como os valores de limiar obtidos pela etapa de segmentação indicados por linhas vermelhas. As Figuras 33(d), 34(d) e 35(d) ilustram os resultados da segmentação por meio das imagens binárias resultantes. As Figuras 33(e), 34(e) e 35(e) apresentam o preenchimento das regiões identificadas na segmentação sobre as imagens originais. Por fim, as Figuras 33(f), 34(f) e 35(f) apresentam regiões, em destaque nas suas correspondentes anteriores, ampliadas para análise dessa etapa.

Por meio das Figuras 33(f), 34(f) e 35(f), é possível notar contornos irregulares das regiões identificadas em todas as lesões, o que se contrapõe aos resultados esperados para essa aplicação. Assim, a etapa de pós-processamento faz-se necessária para correção dos contornos obtidos. Além disso, também são apresentados pequenos ruídos que devem ser eliminados por corresponderem a regiões falso positivas. Para tanto, uma análise empírica foi necessária para determinar o tamanho mínimo das regiões verdadeiro positivas, o que levou à eliminação dessas pequenas regiões. Também foram obtidas segmentações nucleares contendo dois núcleos identificados em um único objeto, o que demandou um método para as suas separações.

## 5.3 Pós-processamento

Na etapa de pós-processamento, inicialmente foi aplicado o método de segmentação *valley-emphasis* para a separação de células unidas resultantes da segmentação. A Figura 36 ilustra a aplicação desse método sobre um caso de cada lesão. As Figuras 36(a), 36(b) e 36(c) representam uma imagem segmentada de LCM, uma região da imagem resultante da segmentação, e a região correspondente obtida pela aplicação desse método, respectivamente. As Figuras 36(d), 36(e) e 36(f) ilustram um caso da lesão LF

$u_1$	$v_1$	$w_1$	$u_2$	$v_2$	$w_2$	FA	$u_1$	$v_1$	$w_1$	$u_2$	$v_2$	$w_2$	FA
40	71	139	244	246	248	0.7311	66	67	125	152	154	181	0.7148
36	108	124	204	234	244	0.9139	37	89	131	140	159	218	1.0959
9	167	202	217	238	245	0.8959	15	102	106	111	160	167	0.7800
44	100	167	173	189	193	1.0799	41	47	61	173	181	192	0.9874
8	12	25	71	180	210	0.7858	61	120	125	139	148	246	0.7399
9	81	97	112	177	242	1.0439	59	101	125	133	170	195	0.8022
48	114	125	165	195	203	1.0799	75	120	125	128	182	250	0.9705
41	70	167	173	181	192	1.0881	48	114	125	165	195	203	1.0799
30	57	87	127	149	245	1.0319	44	100	167	188	189	193	0.9413
65	129	178	192	227	245	0.9637	42	67	79	135	154	167	1.0695
35	38	65	66	140	214	1.0103	28	86	94	97	205	230	0.8786
50	62	64	89	208	237	1.0245	77	107	142	170	198	251	0.9081
90	121	140	149	157	212	1.0508	53	121	140	149	157	212	0.7842
73	97	145	192	193	234	1.0337	38	132	168	200	208	209	0.9778
14	19	33	135	199	238	0.7419	9	89	97	126	177	242	0.9090
3	41	86	120	145	203	0.9935	27	126	167	182	199	230	1.0417
42	67	79	135	154	167	1.0695	44	58	58	79	111	235	1.0390
21	58	115	176	191	233	0.9818	50	53	59	95	120	215	0.9810
20	39	113	137	211	254	0.9471	66	67	61	173	181	192	0.7947
1	27	198	208	222	245	0.8898	41	47	125	146	154	181	0.6823
22	66	102	110	204	232	1.0352	15	102	106	149	157	212	0.9324
35	37	46	67	148	222	0.9758	53	121	140	134	160	167	0.6253
37	89	131	140	159	218	1.0959	59	101	125	133	170	215	0.6013
19	31	47	61	61	102	0.7722	50	53	59	77	120	195	1.0078
13	106	125	125	230	241	1.0381	75	114	132	165	195	203	0.9462
28	86	94	99	199	230	1.0667	48	120	125	125	182	250	0.9328
25	34	62	103	240	244	0.8654	28	86	94	97	177	242	0.9315
4	15	60	90	147	209	0.8938	9	89	97	126	205	230	0.5609
11	43	115	165	166	187	1.0294	27	126	94	97	205	230	0.8748
47	48	76	139	175	190	1.0366	28	86	167	182	199	230	0.6968
21	94	160	198	199	237	1.0058	44	58	58	79	111	218	1.0309
78	111	114	124	130	130	0.9689	37	89	131	136	159	235	0.7147
97	136	164	203	207	208	0.9969	48	114	125	165	198	251	0.9830
89	140	150	159	223	239	1.0085	77	107	142	170	195	203	0.8099
50	53	59	77	120	215	1.0390	77	100	167	173	189	193	0.8166
44	58	58	79	111	235	1.0390	44	107	142	170	198	251	0.7366
28	47	110	112	231	250	0.9789	75	120	125	128	170	195	0.7065
66	67	104	152	154	181	1.0961	59	101	125	133	182	250	0.5236
30	57	76	81	108	130	0.9748	42	67	97	126	177	242	0.8408
7	22	67	186	204	237	0.7386	9	89	79	135	154	167	0.9994
61	117	125	139	148	246	1.0874	44	100	167	173	120	215	0.9475
59	101	125	133	159	173	1.0834	50	53	59	77	189	193	1.0264
10	94	203	226	233	252	0.8620	75	120	140	149	157	212	0.7743
25	35	67	86	173	184	0.9919	53	121	125	128	182	250	0.9216
27	126	167	182	199	230	1.0417	48	114	125	132	154	181	0.9034
8	50	85	178	190	227	0.9032	66	67	125	165	195	203	0.6145
122	128	156	158	219	231	0.9937	44	67	79	135	154	167	0.7674
7	47	61	147	205	226	0.8651	42	58	58	79	111	235	0.8317
43	120	125	128	182	250	1.0818	9	89	97	134	160	167	0.7134
11	15	18	25	133	174	0.7942	15	102	106	126	177	242	1.0333
38	132	168	184	208	209	1.0461	28	86	94	97	160	167	0.7798
110	116	165	204	210	248	0.9569	15	102	106	134	205	230	1.0202
21	34	44	100	205	212	0.8942	75	120	168	200	208	209	0.7462
15	102	106	134	160	167	1.0913	38	132	125	128	182	250	0.9169
4	27	43	74	110	251	0.8838	27	126	167	182	199	246	0.9235
51	87	95	125	235	243	1.0214	61	120	125	139	148	230	0.9006
13	69	108	140	188	240	1.0153	48	114	125	182	199	230	0.7131
77	107	137	170	179	251	1.0511	27	126	167	165	195	203	0.9278
33	44	45	170	178	255	0.8408	50	53	59	77	120	193	1.0071
8	49	94	143	171	225	0.9846	44	100	167	174	189	215	0.4228

(a)

(b)

Figura 32 – Exemplo da aplicação do AG aliado à técnica *fuzzy 3-partition entropy*: (a) população inicial com destaque para indivíduos selecionados pela função de avaliação e (b) geração subsequente após a aplicação das etapas de seleção, crossover e mutação.

contendo a imagem segmentada, uma região segmentada ampliada e essa mesma região após a aplicação da técnica supracitada. Por fim, as Figuras 36(g), 36(h) e 36(i) cor-

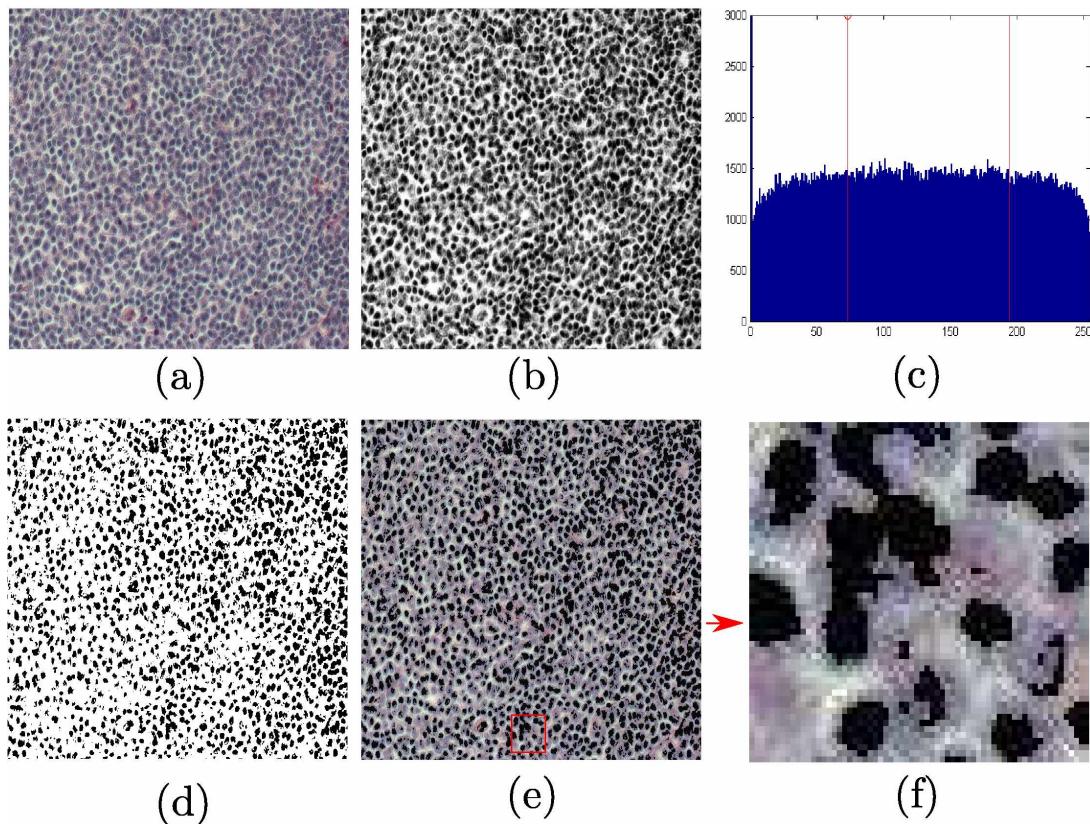


Figura 33 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso *sj-05-768\_013* da classe LCM: (a) imagem original, (b) imagem resultante do pré-processamento, (c) seu histograma com valores de limiar 73 e 194 indicados por linhas vermelhas, (d) imagem binária resultante da segmentação, (e) mapeamento das regiões segmentadas sobre a imagem original e (f) região ampliada para análise do resultado.

respondem à imagem segmentada de um caso de LLC, uma de suas regiões obtidas pela segmentação e o resultado da aplicação dessa etapa sobre essa região, respectivamente.

Pela análise das Figuras 36(c), 36(f) e 36(i), é possível notar que algumas células resultantes são caracterizadas por espaços vazios em seus interiores e com contornos irregulares. Na tentativa de corrigir tais falhas, as operações morfológicas de dilatação e abertura foram empregadas. A Figura 37 ilustra um exemplo de cada lesão após a aplicação dessa abordagem. As Figuras 37(a), 37(d) e 37(g) apresentam as imagens segmentadas com aplicação do método *valley-emphasis* das lesões LCM, LF e LLC, respectivamente. As Figuras 37(b), 37(e) e 37(h) ilustram regiões ampliadas dos casos LCM, LF e LLC, respectivamente. Por fim, as Figuras 37(c), 37(f) e 37(i) correspondem aos resultados da aplicação das operações morfológicas para a obtenção do resultado final.

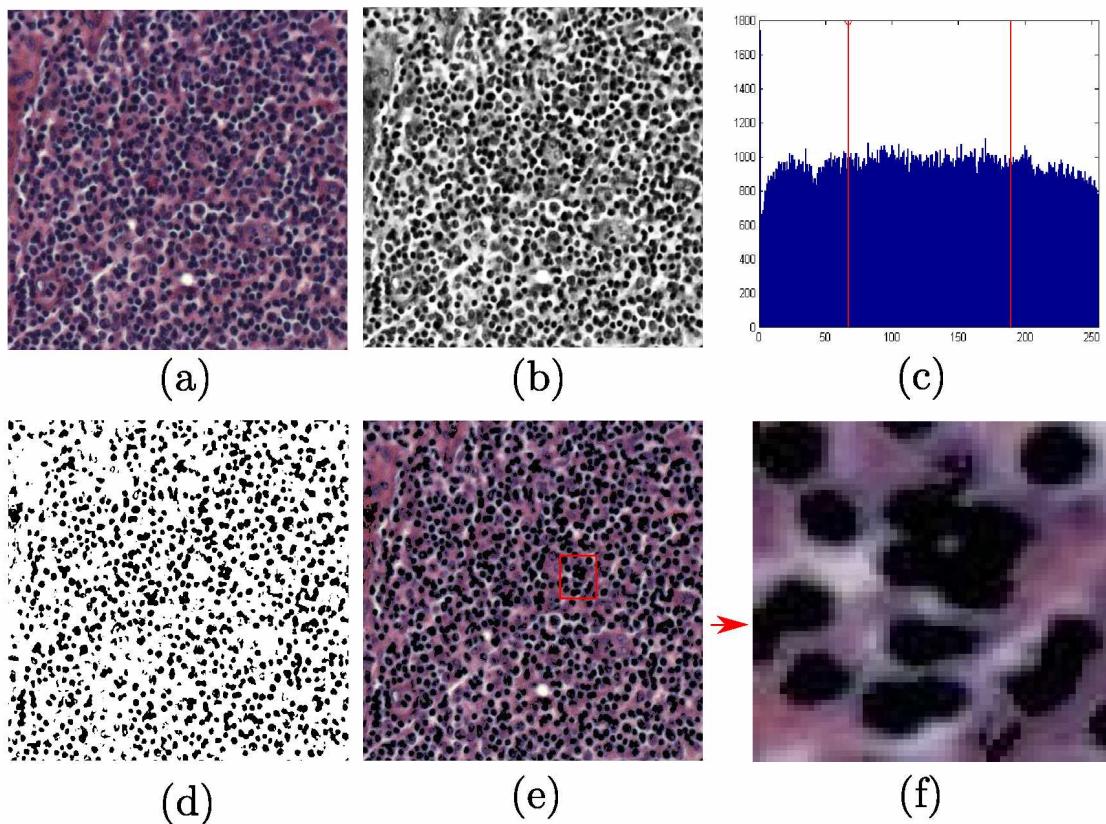


Figura 34 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso *sj-05-6124-R3\_006* da classe LF: (a) imagem original, (b) imagem resultante do pré-processamento descrito, (c) histograma da imagem anterior com valores de limiar 67 e 189 indicados por linhas vermelhas, (d) imagem binária provinda da segmentação, (e) identificação das regiões obtidas na segmentação sobre a imagem original e (f) região ampliada para análise do desempenho da segmentação.

## 5.4 Análise Comparativa de Resultados

Conforme descrito no Capítulo 3 deste estudo, não são apresentadas na literatura técnicas computacionais desenvolvidas para a análise de lesões LCM e LLC utilizando imagens histológicas obtidas por biópsia do linfonodo, apenas com imagens de sangue de alta magnificação. Considerando métodos de segmentação de imagens de LF, também é possível notar a ausência de trabalhos que se dediquem exclusivamente à segmentação de imagens histológicas dessa lesão por meio da identificação de ambas as suas estruturas características, centrócitos e centroblastos, como proposto por este trabalho. Com isso, alguns estudos propostos para a segmentação de estruturas nucleares em imagens histológicas coradas com H&E foram escolhidos para comparação de resultados. O método *mean-shift* (COMANICIU; MEER, 2002) foi utilizado devido à sua ampla aplicação por diferentes trabalhos propostos para a segmentação de imagens médicas, como (LIU; XING; YANG, 2014), (XING; YANG, 2013), (XING; YANG, 2015), (ZHANG et al., 2014) e (ZHOU et al., 2013). As técnicas propostas por (VAHADANE; SETHI, 2013), (WIE-

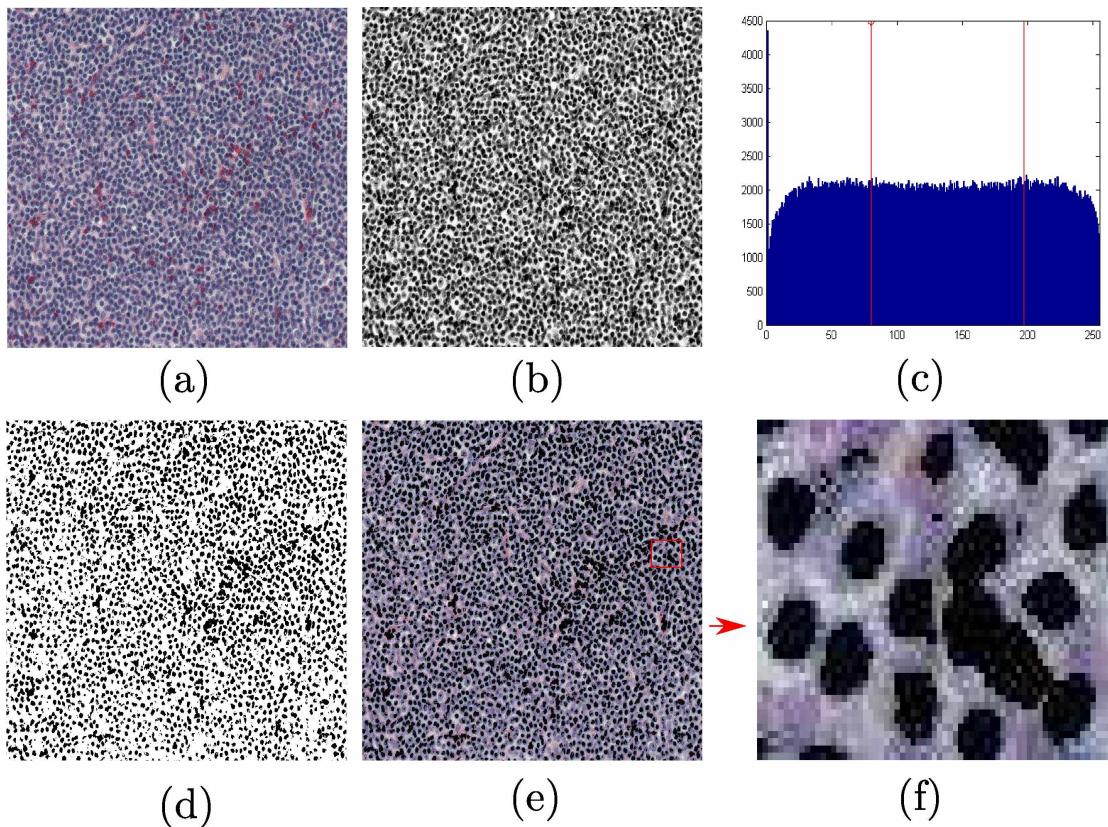


Figura 35 – Exemplo da aplicação da etapa de segmentação sobre uma subimagem do caso *sj-05-3344\_006* da classe LLC: (a) imagem original, (b) imagem resultante da aplicação da equalização do histograma e do filtro Gaussiano, (c) histograma da imagem pré-processada com valores de limiar 80 e 197 indicados por linhas vermelhas, (d) imagem binária da segmentação, (e) mapeamento das regiões segmentadas sobre a imagem original e (f) região ampliada para avaliação do resultado.

NERT et al., 2012) e (OLIVEIRA et al., 2013) também foram empregadas nessa comparação por apresentarem métodos de segmentação nuclear de imagens histológicas, como proposto por este trabalho. Os códigos dos trabalhos utilizados foram fornecidos pelos próprios autores dos artigos para que não houvesse viés entre implementações.

O algoritmo *mean-shift* é um método de agrupamento para reconhecimento de padrões que possui diferentes aplicações, como a segmentação. Essa técnica utiliza estimativas de densidade para a atribuição de cada pixel a uma classe baseada em informações de cores. Para tanto, os parâmetros largura de banda do estimador de densidades, neste trabalho atribuída ao valor 7, e tamanho mínimo da área das regiões identificadas, atribuído ao valor 10, foram definidos. A aplicação desse método descrito por (COMANICIU; MEER, 2002) não contém a aplicação de etapas de pré- e pós-processamento. Por isso foram utilizadas as etapas descritas neste estudo.

O trabalho de (VAHADANE; SETHI, 2013) descreve um método para a segmentação de núcleos celulares em imagens histológicas utilizando o algoritmo *watershed*. Inicial-

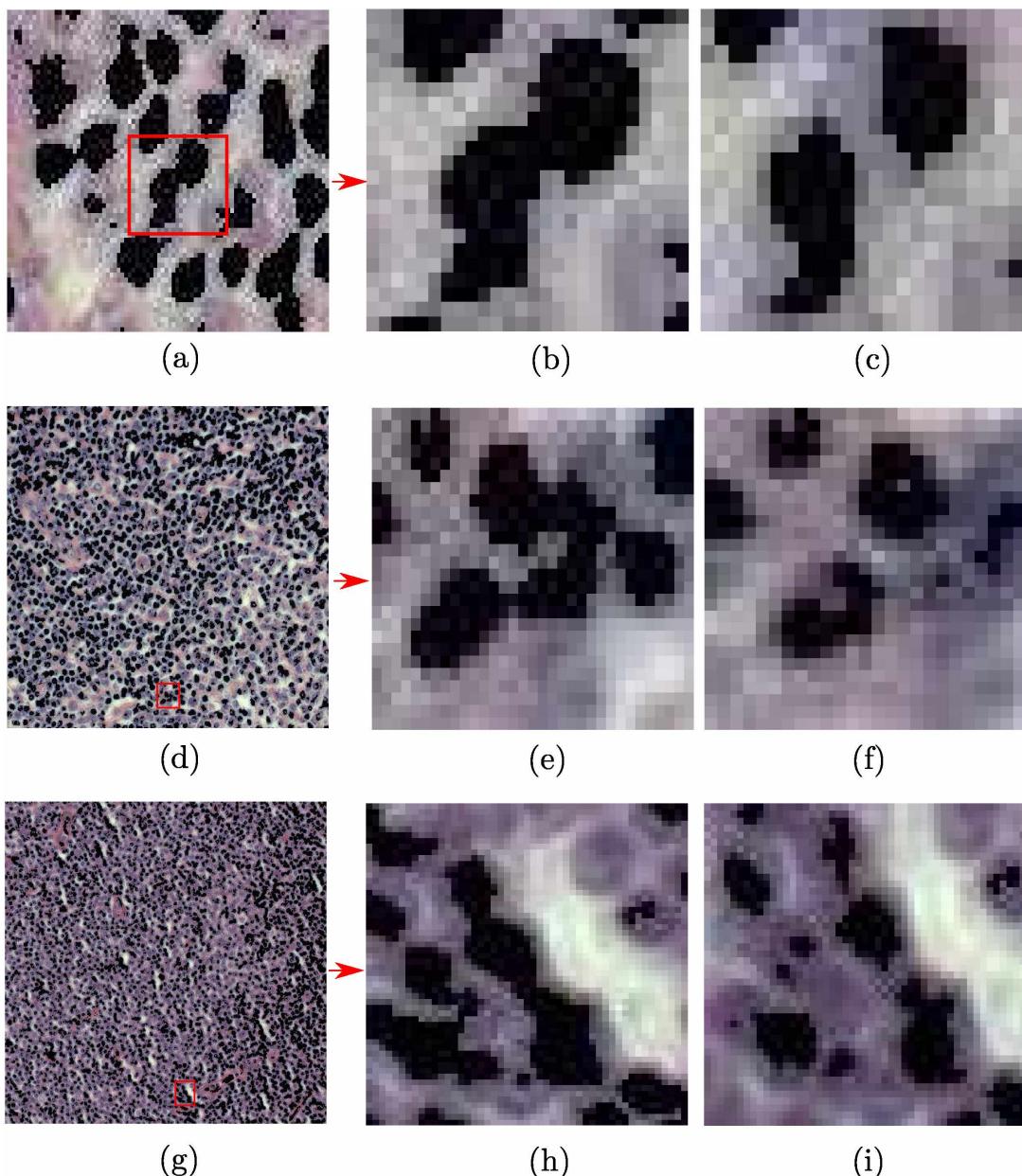


Figura 36 – Exemplo da aplicação do método de segmentação *valley-emphasis* para a separação de células unidas: (a), (d) e (g) representam imagens resultantes da segmentação das classes LCM, LF e LLC, respectivamente, (b), (e) e (h) ilustram regiões ampliadas de suas imagens segmentadas correspondentes, (c), (f) e (i) apresentam tais regiões após esse processamento.

mente, seu trabalho propõe a extração do canal *R* das imagens de entrada e as aplicações do filtro Gaussiano e das operações morfológicas de erosão e dilatação. Na etapa de segmentação, esse trabalho utiliza as técnicas *radial symmetry transform* e a segmentação de Otsu para aperfeiçoar o algoritmo *watershed* e evitar a supersegmentação, resultante desse método. Por fim, as operações morfológicas são novamente aplicadas para eliminação de regiões falso positivas.

O estudo de (WIENERT et al., 2012) também propõe a segmentação de núcleos celula-

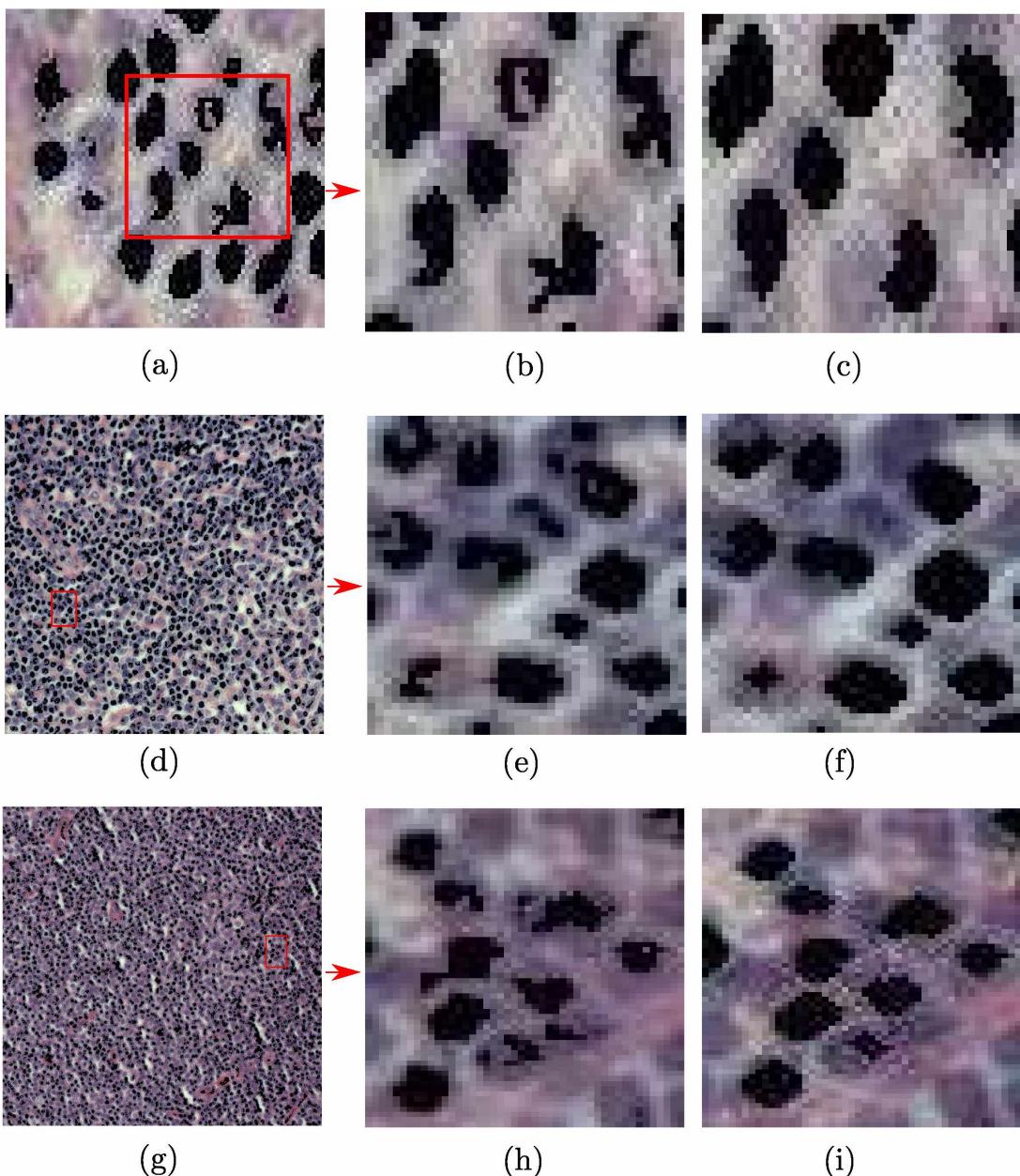


Figura 37 – Exemplo da aplicação das operações morfológicas de dilatação e abertura sobre células identificadas na segmentação: (a), (d) e (g) ilustram imagens segmentadas de LCM, LF e LLC, respectivamente, (b), (e) e (h) apresentam suas regiões ampliadas, (c), (f) e (i) apresentam tais regiões obtidas após esse processamento.

res em imagens histológicas, porém baseada em contornos utilizando mínimas informações *a priori* dos formatos nucleares a serem identificados. Inicialmente, a detecção de possíveis contornos foi realizada com posterior avaliação dos mesmos considerando seus gradientes de máximos locais calculados pelo método de Sobel. Transformadas de distância foram utilizadas para remover pequenos objetos não pertencentes às bordas e a separação de objetos identificados foi realizada pela avaliação da concavidade dos mesmos. Por fim, para a classificação dos objetos em núcleos celulares, foram utilizadas informações de cores

extraídas por meio da técnica de deconvolução, e remoção de regiões com áreas menores que 50 pixels.

O trabalho de (OLIVEIRA et al., 2013) apresenta um método para a segmentação de núcleos de células cubóides em imagens histológicas de próstata por meio do emprego de etapas de pré-processamento, segmentação e pós-processamento. Na primeira etapa, o canal  $R$  foi extraído com posterior aplicação do alargamento de contraste da imagem. A segmentação foi baseada no emprego da técnica de mínima entropia cruzada e o pós-processamento utilizou a rotulação de componentes conexos, cálculo das quantidades de pixels de cada agrupamento e eliminação de regiões com áreas menores que a média dessas quantidades.

As Figuras 38(a), 39(a) e 40(a) apresentam, respectivamente, as imagens originais com magnificação  $20\times$  das classes LCM, LF e LLC. As imagens de referência marcadas manualmente pelo especialista são representadas pelas Figuras 38(b), 39(b) e 40(b) para as classes supracitadas, respectivamente. As Figuras 38(c), 39(c) e 40(c) ilustram os resultados obtidos pela aplicação do método de segmentação descrito por este trabalho sobre as classes LCM, LF e LLC, respectivamente.

As Figuras 38(d), 39(d) e 40(d) exemplificam os resultados obtidos pela técnica *mean-shift*. São apresentados nas Figuras 38(e), 39(e) e 40(e) os resultados do método de (VAHADANE; SETHI, 2013). As Figuras 38(f), 39(f) e 40(f) ilustram os resultados da técnica de (WIENERT et al., 2012). Por fim, nas Figuras 38(g), 39(g) e 40(g), são apresentados os resultados de (OLIVEIRA et al., 2013).

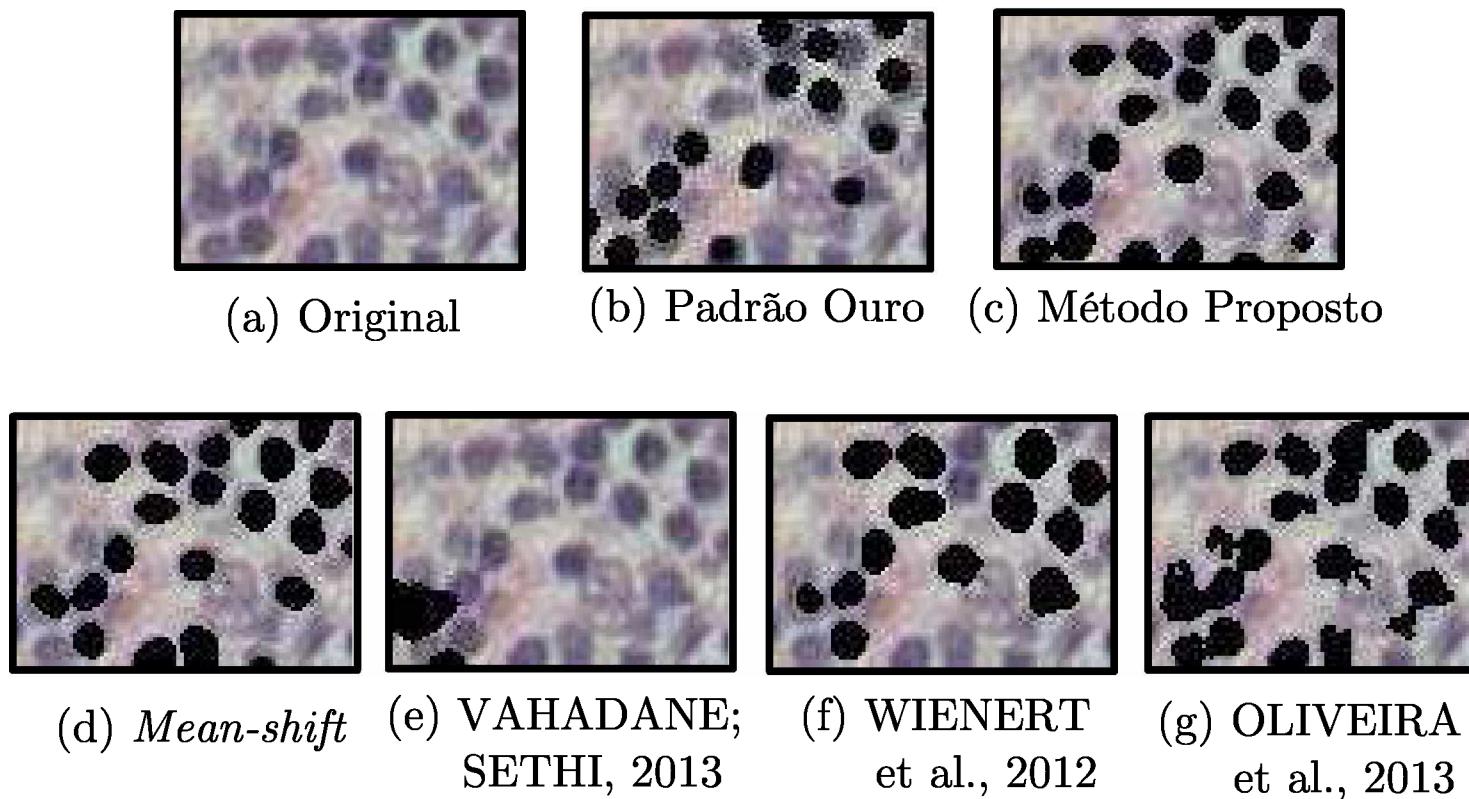


Figura 38 – Comparaçāo de resultados com aplicacāo de tēcnicas de segmentaçāo sobre uma amostra de tecido diagnosticado com LCM: (a) imagem original, (b) imagem segmentada pelo especialista, (c) mētodo proposto por este trabalho, (d) *mean-shift*, (e) tēcnica de (VAHADANE; SETHI, 2013), (f) proposta de (WIENERT et al., 2012) e (g) mētodo de (OLIVEIRA et al., 2013).

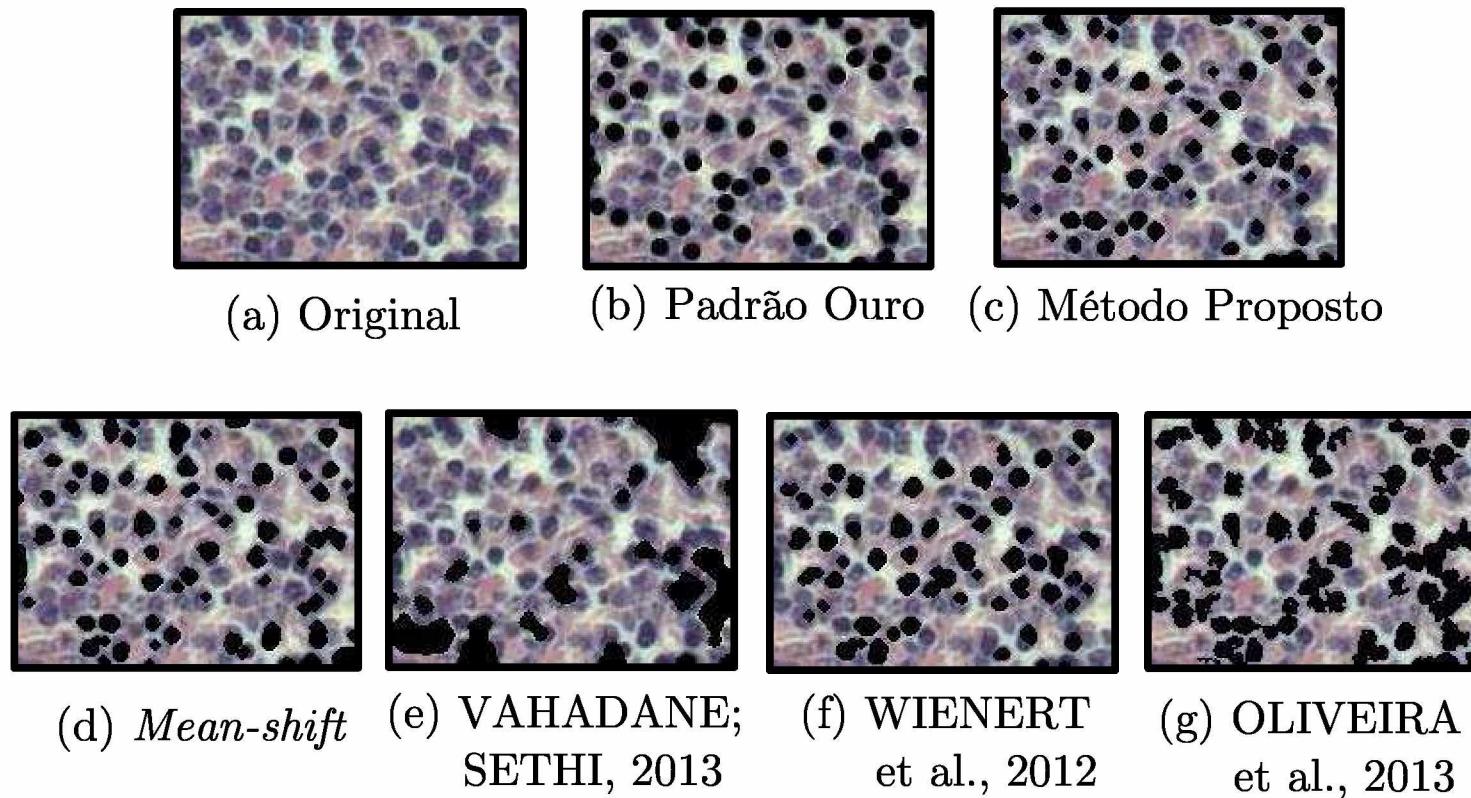


Figura 39 – Comparaçāo de resultados com aplicāo de tēcnicas de segmentaçāo sobre uma amostra de tecido diagnosticado com LF: (a) imagem original, (b) imagem segmentada pelo especialista, (c) mētodo proposto por este trabalho, (d) *mean-shift*, (e) mētodo de (VAHADANE; SETHI, 2013), (f) tēcnica de (WIENERT et al., 2012) e (g) proposta de (OLIVEIRA et al., 2013).

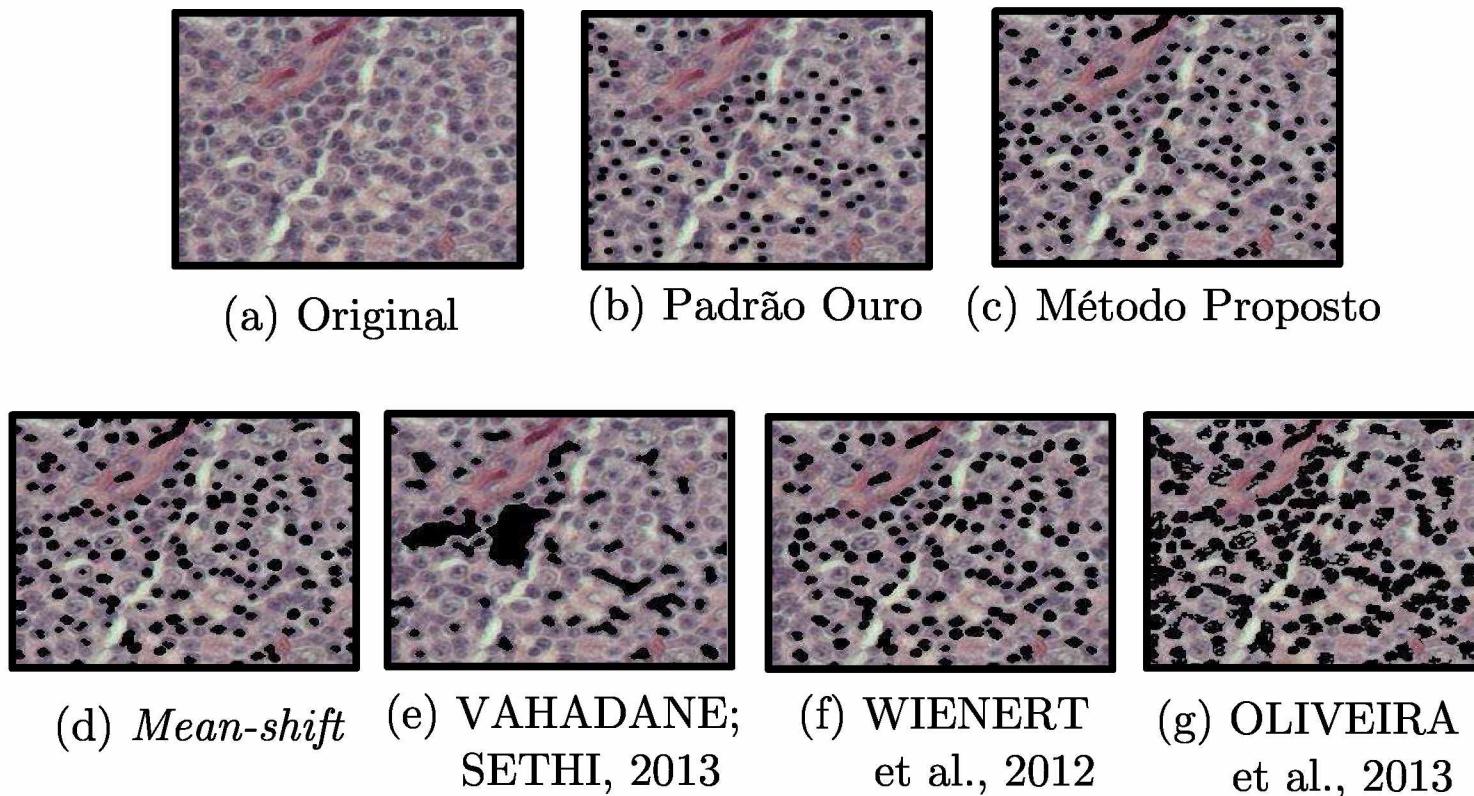


Figura 40 – Comparaçāo de resultados com aplicāo de tēcnicas de segmentaçāo sobre uma amostra de tecido diagnosticado com LLC: (a) imagem original, (b) imagem segmentada pelo especialista, (c) mētodo proposto por este trabalho, (d) *mean-shift*, (e) proposta de (VAHADANE; SETHI, 2013), (f) mētodo de (WIENERT et al., 2012) e (g) tēcnicade (OLIVEIRA et al., 2013).

De acordo com os resultados da segmentação proposta (Figuras 38(c), 39(c) e 40(c)), é possível observar resultados relevantes na separação das regiões nucleares nas imagens das diferentes classes consideradas. Em comparação com a segmentação realizada pelo especialista (Figuras 38(b), 39(b) e 40(b)), os resultados obtidos apresentam maior quantidade de células identificadas, compondo um pequeno conjunto de regiões falso positivas. A área das regiões segmentadas automaticamente apresentaram-se, em pequena proporção, maiores que as identificadas manualmente. Porém, ainda assim, sua distribuição espacial foi semelhante à marcação do especialista.

Em contrapartida, a aplicação do método de (VAHADANE; SETHI, 2013) (Figuras 38(e), 39(e) e 40(e)) sobre as diferentes classes não apresentou resultados coerentes em nenhuma delas. É possível observar que muitas regiões foram identificadas unindo diferentes núcleos. Além disso, a quantidade de estruturas falso negativas apresentou-se de forma expressiva. Essas características podem ser observadas nas três classes de lesões avaliadas neste estudo.

O método de (OLIVEIRA et al., 2013) (Figuras 38(g), 39(g) e 40(g)) apresentou sobreposição em todas as classes, indicando excesso na demarcação das regiões identificadas (MOHAMMED et al., 2013b; OLIVEIRA, 2013). Utilizando suas próprias etapas de pré- e pós-processamento, essa proposta não foi capaz de separar regiões identificadas contendo mais de um núcleo celular. Essa limitação levaria a análises errôneas em possíveis etapas de extração de características e classificação. Além disso, grande quantidade de regiões falso positivas apresentaram-se nos resultados finais.

Em comparação com os resultados de (VAHADANE; SETHI, 2013) e (OLIVEIRA et al., 2013), a aplicação dos métodos *mean-shift* e (WIENERT et al., 2012) apresentaram segmentações mais próximas às realizadas pelo especialista. Os resultados do método *mean-shift* (Figuras 38(d), 39(d) e 40(d)) foram próximos aos obtidos pelo método proposto neste trabalho, porém regiões falso positivas ainda se fizeram presentes numa análise qualitativa das imagens. Também é possível notar que regiões identificadas por essa técnica apresentaram áreas maiores que as segmentadas manualmente, caracterizando a sobreposição. A quantidade de regiões falso positivas é ainda mais expressiva nos resultados de (WIENERT et al., 2012) (Figuras 38(f), 39(f) e 40(f)). É também notório que regiões falso negativas apresentaram-se em maior quantidade nesse método.

A avaliação quantitativa dos resultados obtidos pelo método proposto e por aqueles citados anteriormente foi realizada a partir da comparação com a segmentação manual. Os resultados quantitativos contendo medidas de acurácia, sensibilidade, especificidade, *dice index* e *variation of information*, para o método proposto, *mean-shift*, e as técnicas de (VAHADANE; SETHI, 2013), (WIENERT et al., 2012) e (OLIVEIRA et al., 2013) são apresentadas pelas Tabelas 3, 4 e 5.

Tabela 3 – Resultados quantitativos obtidos pela aplicação dos métodos de segmentação sobre a classe LCM.

Técnica	Acurácia	Sensibilidade	Especificidade	Dice Index	Variation of Information
Método Proposto	80,85%±4,2866	0,4164±0,1604	0,8633±0,0612	0,8884±0,0316	1,0778±0,1848
<i>Mean-shift</i>	79,50%±2,1604	0,48±0,1129	0,8395±0,0191	0,8779±0,0144	1,1514±0,1007
(VAHADANE; SETHI, 2013)	80,67%±7,3243	0,1735±0,1225	0,8949±0,0895	0,8879±0,0524	0,9674±0,2330
(WIENERT et al., 2012)	79,95%±2,5423	0,4128±0,1037	0,8545±0,0331	0,8819±0,0171	1,1239±0,1106
(OLIVEIRA et al., 2013)	67,10%±6,3462	0,6802±0,0780	0,6685±0,0743	0,7788±0,0527	1,3739±0,1370

Tabela 4 – Resultados quantitativos obtidos pelo emprego dos métodos de segmentação sobre a classe LF.

Técnica	Acurácia	Sensibilidade	Especificidade	Dice Index	Variation of Information
Método Proposto	82,77%±1,9957	0,5105±0,1580	0,8731±0,0366	0,8999±0,0135	1,0421±0,1491
<i>Mean-shift</i>	81,62%±1,1042	0,5869±0,0844	0,8476±0,0189	0,8905±0,0074	1,1053±0,0704
(VAHADANE; SETHI, 2013)	79,38%±5,2147	0,2567±0,0814	0,8678±0,0596	0,8797±0,0359	1,0845±0,1858
(WIENERT et al., 2012)	81,05%±1,8416	0,5332±0,0914	0,8507±0,0247	0,8874±0,0126	1,1169±0,0913
(OLIVEIRA et al., 2013)	70,47%±4,0786	0,6997±0,0720	0,7043±0,0524	0,8062±0,0362	1,3357±0,1001

Tabela 5 – Resultados quantitativos obtidos pela aplicação dos métodos de segmentação sobre a classe LLC.

Técnica	Acurácia	Sensibilidade	Especificidade	Dice Index	Variation of Information
Método Proposto	80,82%±2,6783	0,4131±0,1784	0,8864±0,0470	0,8840±0,0202	1,1335±0,2199
<i>Mean-shift</i>	79,74%±2,0699	0,4836±0,0666	0,86±0,0161	0,8760±0,0162	1,2175±0,1256
(VAHADANE; SETHI, 2013)	77,07%±5,4066	0,2101±0,0668	0,8810±0,0454	0,8639±0,0379	1,1676±0,2022
(WIENERT et al., 2012)	78,53%±2,0171	0,4742±0,0542	0,8489±0,0164	0,8682±0,0161	1,2518±0,1248
(OLIVEIRA et al., 2013)	70,60%±2,6045	0,67±0,0767	0,7156±0,0445	0,8018±0,0256	1,4346±0,1241

Por meio da avaliação da métrica de acurácia, que quantifica o quão próxima está a segmentação automática em relação à manual, é possível observar que o método desenvolvido alcançou resultados médios de 80,85%, 82,77% e 80,82% para as classes LCM, LF e LLC, respectivamente. Esses resultados são superiores aos obtidos pelos métodos utilizados para comparação, porém, pela avaliação visual é notório que essa métrica não é suficiente para a análise dos métodos de segmentação. Considerando os resultados da técnica de (VAHADANE; SETHI, 2013), é possível observar identificações errôneas na aplicação dessa técnica mas que apresenta resultados quantitativos próximos aos do método proposto, que possui claro desempenho superior, fazendo com que essa métrica não seja suficiente para avaliar a segmentação.

A medida de sensibilidade é capaz de expressar a quantidade de pixels corretamente identificados na segmentação automática. Em todas as classes de neoplasias analisadas, os métodos *mean-shift*, (WIENERT et al., 2012) e (OLIVEIRA et al., 2013) apresentaram resultados superiores ou próximos aos obtidos pelo método proposto. Porém, quando a análise dessa métrica é realizada considerando a medida de especificidade, é possível inferir sobreposições de regiões. Nessa situação, a métrica de sensibilidade é caracterizada por um alto valor e a especificidade é definida por um baixo valor. Em comparação com os resultados do método proposto, é possível notar essa condição, sobre todas as classes, pelas técnicas supracitadas. Assim, tais técnicas levam a resultados com maior sobreposição que o algoritmo desenvolvido. Além disso, a medida de sensibilidade também foi capaz de expressar o baixo desempenho do método de (VAHADANE; SETHI, 2013).

A métrica *dice index* quantifica a sobreposição espacial entre as segmentações automática e manual. O método proposto obteve os maiores valores dessa métrica em todas as classes de imagens. Porém, a técnica de (VAHADANE; SETHI, 2013) obteve valores semelhantes e superiores a métodos que obtiveram resultados visuais mais coerentes com a avaliação do especialista. Assim, é possível inferir que a análise dessa única medida não é suficiente para a avaliação dos resultados dessa aplicação já que não corrobora os resultados visuais ilustrados pelas Figuras 38, 39 e 40.

A medida *variation of information* teve comportamento semelhante à métrica *dice index*. A métrica *variation of information* é capaz de quantificar as distâncias entre as segmentações automática e manual. Pela avaliação dos resultados quantitativos, o método proposto e a técnica de (VAHADANE; SETHI, 2013) apresentaram os menores valores dessa medida em todas as classes, condição correspondente a segmentações semelhantes à avaliação manual. Porém, os resultados visuais obtidos por (VAHADANE; SETHI, 2013) não representam essa condição, o que também leva a inadequação da análise única do emprego dessa medida sobre a presente aplicação.

## 5.5 Avaliação da Classificação de Linfomas

Baseado no estudo de (DONG et al., 2014), os descritores de características foram obtidos por meio das métricas de acurácia, sensibilidade e especificidade para as imagens segmentadas pelo método proposto e para imagens sem a aplicação de qualquer tipo de processamento. A fim de avaliar a influência da segmentação proposta sobre a etapa de classificação das imagens analisadas, o método SVM foi utilizado por meio da função de base radial (RBF) com parâmetro  $\sigma = 1$ . A divisão das amostras de treinamento e teste foi realizada pela técnica *cross-validation* com atribuição do valor 10 à variável  $k$ . As classificações das lesões LCM × LF, LCM × LLC e LF × LLC utilizaram 161, 111 e 74 imagens, respectivamente, divididas em 10 conjuntos de mesmo tamanho. Assim, a cada iteração, um desses conjuntos foi utilizado para treinamento e o restante foi aplicado para testes. A Tabela 6 apresenta os resultados obtidos com e sem a aplicação do método proposto neste trabalho, e os resultados da classificação utilizando características extraídas das regiões segmentadas manualmente, por meio da métrica de acurácia. A Tabela 7 apresenta os valores obtidos pela aplicação da métrica de sensibilidade considerando as características extraídas da imagem, das regiões individuais segmentadas pelo algoritmo proposto e pelo especialista. Por fim, a Tabela 8 exibe os resultados da medida de especificidade das classificações com e sem a aplicação do método apresentado por este estudo, e considerando as regiões identificadas na segmentação manual.

Tabela 6 – Resultados da classificação das imagens analisadas com e sem a aplicação do método proposto, e a partir da marcação do especialista, utilizando a métrica de acurácia.

Classes	Acurárias do Método Proposto	Acurárias da Segmentação Manual	Acurárias sem Aplicação da Segmentação
LCM × LF	99,38%	99,32%	91,93%
LCM × LLC	96,40%	99,32%	89,19%
LF × LLC	95,95%	88,33%	86,49%

Tabela 7 – Sensibilidades da classificação das imagens analisadas com e sem a aplicação do método proposto, além das características extraídas das regiões segmentadas manualmente.

Classes	Sensibilidades do Método Proposto	Sensibilidades da Segmentação Manual	Sensibilidades sem Aplicação da Segmentação
LCM × LF	1	0,7903	0,8871
LCM × LLC	0,6667	0,7742	0
LF × LLC	0,75	0,9167	0,1667

Pela análise dos resultados apresentados, é possível observar que a aplicação da segmentação tem impacto direto sobre a classificação das imagens analisadas. Considerando a aplicação do método proposto, as classificações entre as classes LCM × LF e LCM

Tabela 8 – Resultados da classificação das imagens analisadas com e sem a aplicação do método proposto, e aplicação da segmentação manual, por meio da métrica de especificidade.

Classes	Especificidades do Método Proposto	Especificidades da Segmentação Manual	Especificidades sem Aplicação da Segmentação
LCM × LF	0,9899	0,9899	0,9394
LCM × LLC	1	0,9899	1
LF × LLC	1	0,6774	1

× LLC apresentaram resultados de acurácia próximos a 100%, permitindo melhor distinção que o emprego da classificação sem a segmentação, com resultados de 91,93% e 89,19%, respectivamente. Ao serem comparados com os resultados obtidos pela aplicação da segmentação manual, a classificação realizada sobre as imagens sem a segmentação apresenta resultados de acurácia inferiores em todas as classificações, enquanto a aplicação do método proposto apresenta resultados dessa métrica próximos e até superiores, como a classificação das classes LF × LLC. A separação dessas classes demonstrou o melhor desempenho utilizando o algoritmo proposto, com resultado de 95,95%.

Considerando a medida de sensibilidade, a aplicação da segmentação manual alcança média de 0,8270 entre as diferentes classificações propostas, valor pouco superior à média obtida pela aplicação do método proposto, 0,8055, e muito superior à média da classificação sem a segmentação, 0,3512. Assim, infere-se que maior quantidade de amostras foram classificadas corretamente gerando altas taxas de amostras verdadeiro positivas. Já a medida de especificidade alcança maiores médias com a aplicação do método proposto, 0,9966. As classificações utilizando a segmentação manual e sem a aplicação da segmentação resultam em médias de 0,8857 e 0,9798. Portanto, a aplicação do método proposto é capaz de alcançar maiores taxas de amostras verdadeiro negativas.

Os resultados dos experimentos utilizando o método *cross-validation 10-fold* são apresentados pelas Tabelas 9 e 10 para as classificações binárias entre as classes analisadas neste trabalho com e sem a aplicação do método proposto, respectivamente. Nessas tabelas, são apresentadas as matrizes de confusão com quantificações das taxas de  $V_P$ ,  $F_N$ ,  $F_P$  e  $V_N$ . As amostras foram classificadas entre as classes LCM × LF, LCM × LLC e LF × LLC.

Pela análise dessas tabelas, é possível notar a alta taxa de amostras  $V_P$  identificadas na classificação (Tabela 9). Além disso, o classificador aliado ao emprego do método proposto alcançou baixos valores nos índices  $F_N$  e  $F_P$ , o que corrobora seu desempenho superior quando comparado à sua aplicação sem a utilização da segmentação. Tal observação é comprovada pela avaliação da Tabela 10, onde o índice  $F_N$  apresenta valores superiores aos apresentados pela Tabela 9 e valores inferiores de amostras  $V_P$ , principalmente nas classificações entre as lesões LCM × LLC e LF × LLC. Considerando as taxas de  $V_N$ , é possível observar pequenas diferenças entre os valores apresentados pelas tabelas supra-

Tabela 9 – Matrizes de confusão das classificações propostas utilizando o algoritmo de segmentação apresentado.

Folds	LCM×LF				LCM×LLC				LF×LLC			
	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>
<i>Fold 1</i>	5	1	0	9	2	0	0	10	1	0	0	6
<i>Fold 2</i>	11	1	0	19	3	0	0	20	1	2	0	12
<i>Fold 3</i>	18	1	0	29	3	1	0	30	1	3	0	19
<i>Fold 4</i>	24	1	0	39	3	2	0	40	2	3	0	25
<i>Fold 5</i>	31	1	0	49	4	2	0	49	3	3	0	31
<i>Fold 6</i>	37	1	0	59	4	3	0	59	4	3	0	38
<i>Fold 7</i>	43	1	0	69	5	3	0	69	6	3	0	44
<i>Fold 8</i>	49	1	0	79	6	3	0	79	6	4	0	50
<i>Fold 9</i>	55	1	0	89	7	3	0	89	7	4	0	56
<i>Fold 10</i>	60	2	0	99	8	4	0	99	7	5	0	62

Tabela 10 – Matrizes de confusão das classificações propostas sem a aplicação do algoritmo de segmentação apresentado.

Folds	LCM×LF				LCM×LLC				LF×LLC			
	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>	V <sub>P</sub>	F <sub>N</sub>	F <sub>P</sub>	V <sub>N</sub>
<i>Fold 1</i>	6	0	1	9	0	1	0	9	0	1	0	7
<i>Fold 2</i>	12	1	1	19	0	2	0	19	0	3	0	13
<i>Fold 3</i>	18	1	2	28	0	3	0	29	0	5	0	19
<i>Fold 4</i>	24	1	2	38	0	5	0	39	0	6	0	25
<i>Fold 5</i>	29	2	2	48	0	6	0	49	0	7	0	31
<i>Fold 6</i>	35	2	3	56	0	7	0	59	1	7	0	37
<i>Fold 7</i>	39	4	3	66	0	9	0	69	2	7	0	43
<i>Fold 8</i>	44	5	3	76	0	10	0	79	3	7	0	50
<i>Fold 9</i>	50	6	3	86	0	11	0	89	3	8	0	56
<i>Fold 10</i>	55	7	3	96	0	12	0	99	3	9	0	62

citadas. Porém, os resultados da Tabela 9 ainda são superiores, mesmo que em menores proporções.

A Figura 41 apresenta graficamente os valores obtidos por cada atributo considerado para a classificação proposta. O índice  $c$  equivale às imagens em níveis de cinza e os índices  $r$ ,  $g$  e  $b$  correspondem aos canais do modelo de cores RGB. As abreviações  $M$ ,  $Me$ ,  $DP$ ,  $C$ ,  $As$ ,  $V$ ,  $NE1$ ,  $NE2$  e  $E$  equivalem aos atributos média, mediana, desvio padrão, curtose, assimetria, variância, 1-norma euclidiana, 2-norma euclidiana e entropia, respectivamente. Nessa figura, são apresentados valores referentes a um caso de cada lesão deste estudo, os quais permitem a apresentação das diferenças entre esses grupos analisados. Na Figura 41(a), são apresentados os valores obtidos com o emprego da segmentação em alguns casos dos grupos analisados. Já a Figura 41(b) apresenta os valores obtidos pelos atributos sem o emprego da segmentação. Pela análise de tais figuras, é possível notar a alta ordem dos valores obtidos sem o emprego da segmentação, alcançando valores como

$10^7$ . Por meio desses gráficos, observa-se que a técnica de segmentação permite reduzir os valores de alguns atributos, o que contribui para níveis de classificação mais relevantes entre os grupos analisados neste estudo.

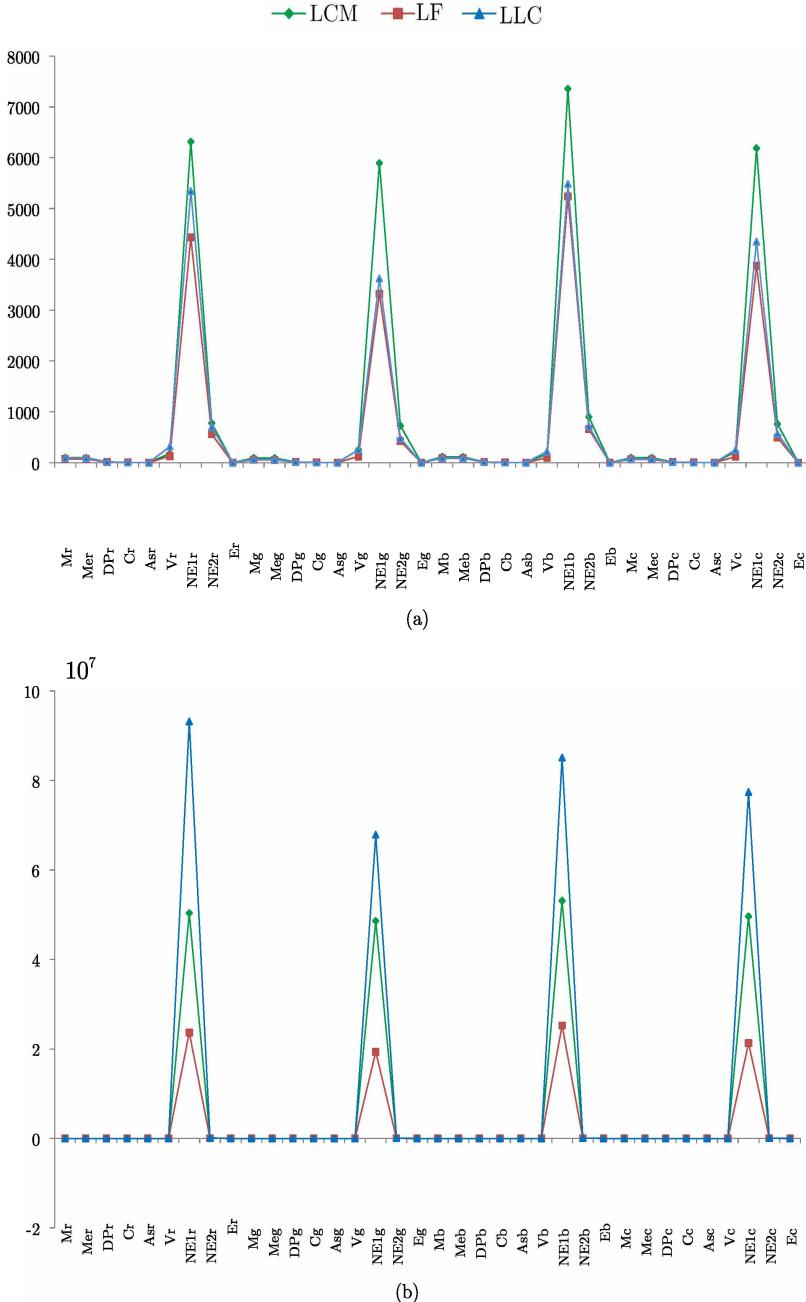


Figura 41 – Valores obtidos pela extração de características de subimagens dos casos *sj-05-768\_013*, *sj-05-894-R3\_004* e *sj-03-852-R2\_012*, das lesões LCM, LF e LLC, respectivamente: (a) utilizando segmentação nuclear proposta e (b) sem a segmentação nuclear proposta.

# CAPÍTULO 6

## Conclusão

Neste trabalho, foi apresentado um método não supervisionado para a segmentação nuclear de células neoplásicas em imagens histológicas de LCM, LF e LLC, coradas com H&E. O método proposto foi desenvolvido em três etapas de processamento. Na etapa de pré-processamento, foram empregados a equalização do histograma e o filtro Gaussiano. A combinação dessas técnicas permitiu eliminar ruídos e realizar correções de iluminação não uniforme decorrentes do processo de aquisição das imagens. Os experimentos realizados demonstraram que de acordo com cada classe de lesão um canal do modelo de cores RGB foi mais relevante para a representação das regiões de interesse. Na etapa de segmentação, a combinação das técnicas *fuzzy 3-partition entropy* e AG mostrou-se eficiente ao permitir a identificação dos núcleos celulares neoplásicos. Ao utilizar a entropia como função de avaliação, foi possível avaliar os parâmetros definidos por cada indivíduo da população e selecionar a combinação de valores que possibilitou a melhor separação das regiões de interesse. Na etapa de pós-processamento, o método de segmentação *valley-emphasis* foi eficaz em seu propósito de separação de diferentes núcleos identificados como um único objeto na etapa anterior. Além disso, o emprego das operações morfológicas de dilatação e abertura foi capaz de corrigir a forma e preencher espaços não nulos das regiões identificadas. A avaliação qualitativa deste método obteve resultados relevantes para imagens com magnificação 20 $\times$ . Além disso, o método também foi avaliado quantitativamente utilizando imagens de referência segmentadas manualmente por um especialista, e comparado com técnicas presentes na literatura, sendo elas: *mean-shift* e de (VAHADANE; SETHI, 2013), (WIENERT et al., 2012) e (OLIVEIRA et al., 2013). Por meio de tal abordagem, o método proposto alcançou médias de acurácia de 80,85%, 82,77% e 80,82% para as lesões LCM, LF e LLC, respectivamente. Também foi analisada a aplicação da técnica proposta para a classificação de imagens de linfoma. As etapas de extração de características e classificação utilizadas para investigação do uso deste método de segmentação demonstraram, de forma quantitativa, a importância dessa etapa na análise dos diferentes tipos de linfoma. Em análises quantitativas da métrica de acurácia, a classificação realizada, considerando a aplicação prévia do método proposto, alcança

diferenças de até 9% em relação à classificação por meio de características extraídas das imagens sem qualquer tipo de processamento. Ao utilizar características extraídas das regiões identificadas pelo especialista, a técnica descrita neste trabalho obtém resultados de acurácia similares e até superiores, alcançando diferenças de até 7%. Para os grupos investigados, há uma diferença média da medida de acurácia de 8,04% entre a classificação sem o emprego do método proposto e com esse método, e de 4,76% em relação à classificação utilizando a segmentação manual do especialista.

As técnicas utilizadas para comparação foram escolhidas devido ao seu desenvolvimento dedicado à segmentação nuclear em imagens histológicas com o mesmo tipo de coloração (H&E). As análises visuais mostraram que tais técnicas não foram capazes de proporcionar uma boa representação das estruturas segmentadas pelo especialista. Os resultados quantitativos obtidos pelas diferentes métricas empregadas neste estudo mostraram também tal diferença para os grupos de lesões considerados. Os resultados da medida de acurácia obtidos pelos diferentes métodos de comparação alcançaram médias de 76,80%, 78,13% e 76,48% para as classes LCM, LF e LLC, respectivamente.

## 6.1 Principais Contribuições

O método desenvolvido foi eficiente em todas as classes de lesões analisadas, considerando sua aplicação sobre imagens com grandes diferenças de contraste e iluminação. Para que essa eficiência fosse observada nesta técnica, foram utilizadas características intrínsecas dos núcleos celulares, desconsiderando suas distribuições espaciais e formas. Essa característica permite sua aplicação em ambientes reais ou para fins de extração de características e classificação, gerando análises mais precisas das classes LCM, LF e LLC.

Outro aspecto importante é que o método proposto foi capaz de contribuir para o estado da arte de processamento de imagens de linfoma. Pela escassa quantidade de trabalhos relacionados à segmentação de imagens de LCM e LLC, a contribuição desta técnica torna-se ainda maior. Considerando imagens histológicas de LF, apesar de maior quantidade de trabalhos dedicados ao seu processamento, este estudo ainda é relevante ao considerar a identificação dos dois tipos celulares característicos dessa neoplasia: centrócitos e centroblastos. Além disso, este trabalho propõe uma combinação de técnicas ainda não utilizadas por outros estudos, como o AG e *fuzzy 3-partition entropy*.

O método proposto apresenta resultados robustos ao alcançar maior representatividade dos contornos nucleares. Comparada aos outros métodos da literatura considerados, a técnica descrita apresenta contornos mais consistentes e coerentes com a marcação do especialista. Tal efeito é decorrente do emprego da segmentação aliada à etapa de pós-processamento que foi capaz de identificar com maior exatidão tais limites. O uso do AG aliado à técnica *fuzzy 3-partition entropy* também se mostrou eficiente ao permitir a avaliação de diferentes soluções propostas por indivíduos da população levando à combi-

nação de valores que possibilitasse a remoção de algumas regiões não neoplásicas ainda nessa etapa, contrapondo-se aos resultados finais dos métodos comparados. Por meio das avaliações qualitativa e quantitativa, a combinação dos métodos utilizados nas diferentes etapas deste trabalho foi capaz de alcançar resultados mais eficazes que os métodos empregados para comparação.

Uma limitação da técnica apresentada é a permanência de regiões falso positivas nos resultados finais, semelhante ao que foi apresentado pelos métodos comparados, porém nesses em maiores proporções. Por meio dessa análise, é possível concluir que apenas a característica de intensidade não é suficiente para a diferenciação entre núcleos neoplásicos e não neoplásicos dos diferentes tipos de lesões abordadas por este trabalho.

## 6.2 Trabalhos Futuros

Os trabalhos futuros deste projeto podem se dedicar a duas vertentes: otimização do método proposto e desenvolvimento de um sistema CAD utilizando as etapas descritas neste estudo.

Para otimização do método desenvolvido, os trabalhos futuros podem incluir:

- Validação do método proposto utilizando imagens de referência segmentadas por diferentes especialistas;
- Validação por meio de imagens caracterizadas por magnificações maiores que 20×.

Para o desenvolvimento de um sistema CAD, os seguintes temas podem ser investigados:

- Exploração de características das regiões identificadas para eliminação de objetos falso positivos, compondo uma etapa de detecção das regiões de interesse;
- Investigação de classificadores capazes de separar, de forma ainda mais eficiente, as classes de imagens analisadas.

## 6.3 Contribuições em Produção Bibliográfica

- Tosta, T. A. A., de Abreu, A. F., Travencolo, B. A. N., do Nascimento, M. Z., e Neves, L. A. (2015, Junho). *Unsupervised Segmentation of Leukocytes Images Using Thresholding Neighborhood Valley-Emphasis*. 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS), (pp. 93-94). IEEE.
- Tosta, T. A. A., de Abreu, A. F., Vilela, D. S., Travencolo, B. A. N., do Nascimento, M. Z., e Neves, L. A. (2015, Outubro). *Unsupervised Segmentation of Leukocytes*

*Images Using Thresholding Neighborhood Valley-Emphasis.* XI Workshop de Visão Computacional (WVC), (pp. 112-117).

- Neves, L. A., Pavarino, E., Souza, M. P., Valencio, C. R., Zafalon, G. F. D., Zanchetta do Nascimento, M., e Tosta, T. A. A. (2015, Junho). *Multiscale Tetrahedral Meshes for FEM Simulations of Esophageal Injury.* 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS), (pp. 103-108). IEEE.

## 6.4 Agradecimentos

T.A.A.T. e M.Z.N. agradecem à CAPES e FAPEMIG (projeto TEC- APQ-02885-15) pelo apoio financeiro.

---

## Referências

- ABBASGHOLIPOUR, M. et al. Color image segmentation with genetic algorithm in a raisin sorting system based on machine vision in variable conditions. **Expert Systems with Applications**, v. 38, n. 4, p. 3671–3678, 2011.
- ARORA, B.; BANERJEE, S. Computer assisted grading schema for follicular lymphoma based on level set formulation. In: IEEE. **2013 Students Conference Engineering and Systems (SCES)**. [S.l.], 2013. p. 1–6.
- ATUPELAGE, C. et al. Computational cell classification methodology for hepatocellular carcinoma. In: IEEE. **2013 International Conference Advances in ICT for Emerging Regions (ICTer)**. [S.l.], 2013. p. 21–27.
- BANIMELHEM, O.; YAHYA, Y. A. Multi-thresholding image segmentation using genetic algorithm. In: **World Congress in Computer Science, Computer Engineering and Applied Computing**. [S.l.: s.n.], 2011.
- BELKACEM-BOUSSAID, K. et al. Segmentation of follicular regions on h&e slides using a matching filter and active contour model. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **SPIE Medical Imaging**. [S.l.], 2010. p. 762436–762436.
- BELKACEM-BOUSSAID, K. et al. Automatic detection of follicular regions in h&e images using iterative shape index. **Computerized Medical Imaging and Graphics**, Elsevier, v. 35, n. 7, p. 592–602, 2011.
- BELKACEM-BOUSSAID, K. et al. Extraction of color features in the spectral domain to recognize centroblasts in histopathology. In: IEEE. **Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE**. [S.l.], 2009. p. 3685–3688.
- BELSARE, A. D.; MUSHRIF, M. M. Histopathological image analysis using image processing techniques: An overview. **Signal & Image Processing : An International Journal (SIPIJ)**, v. 3, n. 4, p. 23–36, 2012.
- BUCCHERI, V.; MAURIÑO, B. B. **Clínica Médica**. [S.l.], 2011. v. 3, 165-173 p.
- BYRD, K. A.; ZENG, J.; CHOUIKHA, M. A validation model for segmentation algorithms of digital mammography images. **Journal of Applied Science & Engineering Technology**, v. 1, 2007.

- CANELLOS, G. P.; LISTER, T. A.; YOUNG, B. **The Lymphomas**. 2<sup>nd</sup> edition. [S.l.], 2006.
- CARRERAS, A.; GONZALEZ, R. F.; SOLORZANO, C. O. Automatic registration of serial mammary gland sections. In: IEEE. **26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004. IEMBS'04**. [S.l.], 2004. v. 1, p. 1691–1694.
- CHANG, V. et al. Gold-standard and improved framework for sperm head segmentation. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 117, n. 2, p. 225–237, 2014.
- COMANICIU, D.; MEER, P. Mean shift: A robust approach toward feature space analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 24, n. 5, p. 603–619, 2002.
- CORMACK, D. H.; HAM, A. W. **Histologia**. 8. ed. [S.l.], 1983.
- DIMITROPOULOS, K. et al. Using adaptive neuro-fuzzy inference systems for the detection of centroblasts in microscopic images of follicular lymphoma. **Signal, Image and Video Processing**, Springer, v. 8, n. 1, p. 33–40, 2014.
- DONG, F. et al. Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. **PloS one**, Public Library of Science, v. 9, n. 12, p. e114885, 2014.
- DUAN, J.; YU, L. A wbc segmentation method based on hsi color space. In: IEEE. **4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)**. [S.l.], 2011. p. 629–632.
- ESTRADA, F. J.; JEPSON, A. D. Benchmarking image segmentation algorithms. **International Journal of Computer Vision**, Springer, v. 85, n. 2, p. 167–181, 2009.
- FACINA, T. **Estimativa 2014 – Incidência de Câncer no Brasil**. 2014.
- GANESAN, P.; RAJINI, V. Yiq color space based satellite image segmentation using modified fcm clustering and histogram equalization. In: IEEE. **International Conference on Advances in Electrical Engineering (ICAEE)**. [S.l.], 2014. p. 1–5.
- GARTNER, L. P.; HIATT, J. L. **Tratado de Histologia em Cores**. 2. ed. [S.l.], 2003. 223-244 p.
- GERARD, E. J.; BISHOP, M. R. Overview of non-hodgkin's lymphoma. **Dis Mon**, v. 58, p. 208–218, 2012.
- GHOSE, S. et al. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 108, n. 1, p. 262–287, 2012.
- GONZALEZ, R. C.; WOODS, R. E. **Processamento de Imagens Digitais**. [S.l.]: Edgard Blucher, 2000.
- GUDMUNDSSON, M.; KABUKA, M. R. Edge detection in medical images using a genetic algorithm. **IEEE Transactions on Medical Imaging**, v. 17, n. 3, p. 469–474, 1998.

- GURCAN, M. N. et al. Histopathological image analysis: A review. **IEEE Reviews in Biomedical Engineering**, IEEE, v. 2, p. 147–171, 2009.
- HAGGERTY, J. M. et al. Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin. **BMC Medical Imaging**, BioMed Central Ltd, v. 14, n. 1, p. 7, 2014.
- HAMMOUCHE, K.; DIAF, M.; SIARRY, P. A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. **Computer Vision and Image Understanding**, Elsevier, v. 109, n. 2, p. 163–175, 2008.
- HARVEY, N. R.; MARSHALL, S. The design of different classes of morphological filter using genetic algorithms. In: IET. **Fifth International Conference on Image Processing and its Applications**. [S.l.], 1995. p. 227–231.
- HE, L. et al. Histology image analysis for carcinoma detection and grading. **Computer Methods and Programs in Biomedicine**, v. 107, n. 3, p. 538–556, 2012.
- HE, L. et al. Sequence and genome analysis: Methods and applications. **Sequence and Genome Analysis: Methods and Applications**, p. 271–287, 2010.
- HEALEY, C. G. Choosing effective colours for data visualization. In: IEEE. **Visualization'96. Proceedings**. [S.l.], 1996. p. 263–270.
- IMAGEJ. **ImageJ (Image Processing and Analysis in Java)**. 2015. Accessado em: 04/11/2015. Disponível em: <<http://imagej.nih.gov/ij/index.html>>.
- INSANA, M.; MEYERS, K.; GROSSMAN, L. **Handbook of Medical Imaging: Medical Image Processing and Analysis**. [S.l.]: SPIE Press, Chapter: Signal Modelling for Tissue Classification, 2000.
- IRSHAD, H. et al. Methods for nuclei detection, segmentation, and classification in digital histopathology: A review - current status and future potential. **IEEE Reviews in Biomedical Engineering**, IEEE, v. 7, p. 97–114, 2014.
- JANSSENS, T. et al. Charisma: An integrated approach to automatic h&e-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting. **Medical Image Analysis**, v. 17, n. 8, p. 1206–1219, 2013.
- JIANLI, L.; BAOQI, Z. The segmentation of skin cancer image based on genetic neural network. In: IEEE. **2009 WRI World Congress on Computer Science and Information Engineering**. [S.l.], 2009. v. 5, p. 594–599.
- JUNQUEIRA, J. C.; CARNEIRO, J. **Histologia Básica**. 10. ed. [S.l.], 2004. 254-283 p.
- KANDWAL, R.; KUMAR, A.; BHARGAVA, S. Review: Existing image segmentation techniques. **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 4, n. 4, p. 153–156, 2014.
- KARSNAS, A. **Image Analysis Methods and Tools for Digital Histopathology Applications Relevant to Breast Cancer Diagnosis**. Dissertação (Mestrado) — Uppsala Universitet, 2014.

- KAUSHIK, D. et al. Medical image segmentation using genetic algorithm. **Image**, Citeseer, v. 81, n. 18, 2013.
- KONG, H.; BELKACEM-BOUSSAID, K.; GURCAN, M. Cell nuclei segmentation for histopathological image analysis. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **SPIE Medical Imaging**. [S.l.], 2011a. p. 79622R–79622R.
- KONG, H.; GURCAN, M.; BELKACEM-BOUSSAID, K. Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting. **IEEE Transactions on Medical Imaging**, IEEE, v. 30, n. 9, p. 1661–1677, 2011b.
- KORZYNSKA, A. et al. Validation of various adaptive threshold methods of segmentation applied to follicular lymphoma digital images stained with 3, 3-diaminobenzidine&haematoxylin. **Diagn Pathol**, v. 8, n. 48, p. 1596–8, 2013.
- KOTHARI, S. et al. Automatic batch-invariant color segmentation of histological cancer images. In: IEEE. **2011 IEEE International Symposium Biomedical Imaging: from Nano to Macro**. [S.l.], 2011. p. 657–660.
- KRISHNAN, M. R. et al. Automated characterization of sub-epithelial connective tissue cells of normal oral mucosa: Bayesian approach. In: IEEE. **2010 IEEE Students' Technology Symposium (TechSym)**,. [S.l.], 2010. p. 44–48.
- KURU, K. Optimization and enhancement of h&e stained microscopical images by applying bilinear interpolation method on lab color model. **Theor Biol Med Model**, 2014.
- LAD, K.; AGRAWAL, M.; PANDYA, M. M. Survey on genetic algorithms & basic operators. **International Journal of Advanced Information Science and Technology (IJAIST)**, v. 22, n. 22, p. 44–48, 2014.
- LEONG, F. W.; BRADY, M.; MCGEE, J. Correction of uneven illumination (vignetting) in digital microscopy images. **Journal of Clinical Pathology**, BMJ Publishing Group Ltd and Association of Clinical Pathologists, v. 56, n. 8, p. 619–621, 2003.
- LIN, W.; LEE, W.; HONG, T. Adapting crossover and mutation rates in genetic algorithms. **J. Inf. Sci. Eng.**, Citeseer, v. 19, n. 5, p. 889–903, 2003.
- LIU, F.; XING, F.; YANG, L. Robust muscle cell segmentation using region selection with dynamic programming. In: IEEE. **2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)**. [S.l.], 2014. p. 521–524.
- LUO, Y.; CELENK, M.; BEJAI, P. Discrimination of malignant lymphomas and leukemia using radon transform based-higher order spectra. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging**. [S.l.], 2006. p. 61445K–1–61445K–10.
- MAURIÑO, B. B.; SIQUEIRA, S. A. C. **Classificação dos Linfomas**. [S.l.], 2011. 174-188 p.
- MCCALL, J. Genetic algorithms for modelling and optimisation. **Journal of Computational and Applied Mathematics**, Elsevier, v. 184, n. 1, p. 205–222, 2005.

- MCCANN, M. T. et al. Automated histology analysis: Opportunities for signal processing. **IEEE Signal Processing Magazine**, IEEE, v. 32, n. 1, p. 78–87, 2015.
- MEESTER, U. et al. Towards a quantitative grading of bladder tumors. **Cytometry**, v. 12, n. 7, p. 602–613, 1991.
- MICHAIL, E. et al. Detection of centroblasts in h&e stained images of follicular lymphoma. In: **IEEE. 2014 22nd Signal Processing and Communications Applications Conference (SIU)**. [S.l.], 2014. p. 2319–2322.
- MOHAMMED, E. A. et al. Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation. In: **IEEE 15th International Conference on e-Health Networking, Applications and Services**. [S.l.: s.n.], 2013.
- MOHAMMED, E. A. et al. Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In: **IEEE. 26th Annual IEEE Canadian Conference Electrical and Computer Engineering (CCECE), 2013**. [S.l.], 2013. p. 1–5.
- NEUMAN, U. et al. **Information Technologies in Biomedicine**. [S.l.], 2010. v. 69, 101-113 p.
- NG, H. Automatic thresholding for defect detection. **Pattern Recognition Letters**, Elsevier, v. 27, n. 14, p. 1644–1649, 2006.
- NIA. **Lymphoma - NIA Intramural Research Program Laboratory of Genetics**. 2008.
- OGER, M.; BELHOMME, P.; GURCAN, M. N. A general framework for the segmentation of follicular lymphoma virtual slides. **Computerized Medical Imaging and Graphics**, Elsevier, v. 36, n. 6, p. 442–451, 2012.
- OLIVEIRA, D. L. L. de. **Método Computacional para Segmentação Não Supervisionada de Componentes Histológicos da Próstata**. Dissertação (Mestrado) — Universidade Federal do ABC, 2013.
- OLIVEIRA, D. L. L. de et al. Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy. **Expert Systems with Applications**, Elsevier, v. 40, n. 18, p. 7331–7340, 2013.
- ONG, S. et al. Image analysis of tissue sections. **Computers in Biology and Medicine**, Elsevier, v. 26, n. 3, p. 269–279, 1996.
- ORLOV, N. V. et al. Automatic classification of lymphoma images with transform-based global features. **IEEE Transactions on Information Technology in Biomedicine**, IEEE, v. 14, n. 4, p. 1003–1013, 2010.
- OTSU, N. A threshold selection method from gray-level histograms. **Automatica**, v. 11, n. 285-296, p. 23–27, 1975.
- OZTAN, B. et al. Follicular lymphoma grading using cell-graphs and multi-scale feature analysis. In: **SPIE Medical Imaging**. [S.l.: s.n.], 2012. v. 8315.

PAKHIRA, M. K. **Digital Image Processing and Pattern Recognition.** [S.l.]: PHI Learning Private Limited, 2011.

PATHPEDIA. Mantle cell lymphoma. Global Online Pathology Resource. 2014. Disponível em: <[http://www.pathpedia.com/education/eatlas/histopathology/lymph\\_node/mantle\\_cell\\_lymphoma.aspx](http://www.pathpedia.com/education/eatlas/histopathology/lymph_node/mantle_cell_lymphoma.aspx)>.

PAULINAS, M.; UŠINSKAS, A. A survey of genetic algorithms applications for image enhancement and segmentation. **Information Technology and Control**, v. 36, n. 3, 2007.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações.** [S.l.]: Editora Thomson Learning, 2007. 528 p. ISBN 978-85-221-0595-3.

RIBEIRO, A. et al. An image segmentation based on a genetic algorithm for determining soil coverage by crop residues. **Sensors**, v. 11, n. 6, p. 6480–6492, 2011.

ROCHA, R. et al. Digital slides: Present status of a tool for consultation, teaching, and quality control in pathology. **Pathology-Research and Practice**, v. 205, n. 11, p. 735–741, 2009.

RUIFRON, A. C.; JOHNSTON, D. A. Quantification of histochemical staining by color deconvolution. **Analytical and Quantitative Cytology and Histology/ The International Academy of Cytology and American Society of Cytology**, v. 23, n. 4, p. 291–299, 2001.

SANTOS, F. P. S.; FERNANDES, G. S. **Linfomas não-Hodgkin.** [S.l.], 2008.

SERTEL, O. et al. An image analysis approach for detecting malignant cells in digitized h&e-stained histology images of follicular lymphoma. In: **IEEE. 20th International Conference on Pattern Recognition (ICPR), 2010.** [S.l.], 2010a. p. 273–276.

SERTEL, O. et al. Computerized microscopic image analysis of follicular lymphoma. In: **INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. Medical Imaging.** [S.l.], 2008. v. 6915.

SERTEL, O. et al. Texture classification using nonlinear color quantization: Application to histopathological image analysis. In: **IEEE. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008.** [S.l.], 2008. p. 597–600.

SERTEL, O. et al. Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. **Journal of Signal Processing Systems**, v. 55, n. 1-3, p. 169–183, 2009.

SERTEL, O. et al. Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 57, n. 10, p. 2613–2616, 2010b.

SHETA, A.; BRAIK, M. S.; ALJAHDALI, S. Genetic algorithms: A tool for image segmentation. In: **IEEE. International Conference on Multimedia Computing and Systems (ICMCS).** [S.l.], 2012. p. 84–90.

- SIEREN, J. C. et al. An automated segmentation approach for highlighting the histological complexity of human lung cancer. **Annals of Biomedical Engineering**, v. 38, n. 12, p. 3581–3591, 2010.
- TADROUS, P. J. Digital stain separation for histological images. **Journal of Microscopy**, v. 240, n. 2, p. 164–172, 2010.
- TOSTA, T. A. A. et al. Unsupervised segmentation of leukocytes images using thresholding neighborhood valley-emphasis. In: IEEE. **IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2015.
- TSAI, D.; LEE, Y.; MATSUYAMA, E. Information entropy measure for evaluation of image quality. **Journal of Digital Imaging**, Springer, v. 21, n. 3, p. 338–347, 2008.
- UNICAMP, F. **Anatomia Patológica Especial, Módulo - Hemopatologia**. [S.l.], 2013. Disponível em: <<http://anatpat.unicamp.br/lamhemo.html>>.
- VAHADANE, A.; SETHI, A. Towards generalized nuclear segmentation in histological images. In: IEEE. **IEEE 13th International Conference Bioinformatics and Bioengineering (BIBE), 2013**. [S.l.], 2013. p. 1–4.
- VEILLARD, A.; BRESSAN, S.; RACOCEANU, D. Svm-based framework for the robust extraction of objects from histopathological images using color, texture, scale and geometry. In: **11th International Conference Machine Learning and Applications (ICMLA), 2012**. [S.l.: s.n.], 2012. v. 1, p. 70–75.
- VETA, M. et al. Breast cancer histopathology image analysis: A review. **IEEE Transactions on Biomedical Engineering**, v. 61, n. 5, p. 1400–1411, 2014.
- VINITSKI, S. et al. Validation of tissue segmentation based on 3d feature map in an animal model of a brain tumor. In: IEEE. **Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE**. [S.l.], 1996. v. 2, p. 740–742.
- WANG, W.; OZOLEK, J. A.; ROHDE, G. K. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. **Cytometry Part A**, v. 77, n. 5, p. 485–494, 2010.
- WIENERT, S. et al. Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach. **Scientific Reports**, Nature Publishing Group, v. 2, 2012.
- WU, G. et al. Histological image segmentation using fast mean shift clustering method. **Biomedical Engineering Online**, BioMed Central Ltd, v. 14, n. 1, p. 24, 2015.
- XING, F.; YANG, L. Robust selection-based sparse shape model for lung cancer image segmentation. In: **Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013**. [S.l.]: Springer, 2013. p. 404–412.
- XING, F.; YANG, L. Unsupervised shape prior modeling for cell segmentation in neuroendocrine tumor. In: IEEE. **12th International Symposium on Biomedical Imaging (ISBI), 2015 IEEE**. [S.l.], 2015. p. 1443–1446.

- YANG, L. et al. Automatic image analysis of histopathology specimens using concave vertex graph. In: **Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2008.** [S.l.]: Springer, 2008. v. 11, n. 1, p. 833–841.
- YIN, S. et al. Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization. **Pattern Recognition**, Elsevier, v. 47, n. 9, p. 2894–2907, 2014.
- ZHANG, J. et al. An improved quantum-inspired genetic algorithm for image multilevel thresholding segmentation. **Mathematical Problems in Engineering**, Hindawi Publishing Corporation, v. 2014, 2014.
- ZHOU, Y. et al. Stain guided mean-shift filtering in automatic detection of human tissue nuclei. **Journal of Pathology Informatics**, Medknow Publications, v. 4, n. Suppl, 2013.
- ZORMAN, M. et al. Symbol-based machine learning approach for supervised segmentation of follicular lymphoma images. In: **IEEE. Twentieth IEEE International Symposium Computer-Based Medical Systems (CBMS), 2007.** [S.l.], 2007. p. 115–120.