



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS E TECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA**  
**DOUTORADO EM ENGENHARIA DE TELEINFORMÁTICA**

**DANIEL SILVA FERREIRA**

**TRIAGEM AUTOMÁTICA DO EXAME DE PAPANICOLAOU COM BASE EM**  
**ATENÇÃO VISUAL**

**FORTALEZA**

**2020**

DANIEL SILVA FERREIRA

TRIAGEM AUTOMÁTICA DO EXAME DE PAPANICOLAOU COM BASE EM ATENÇÃO  
VISUAL

Tese apresentada ao Curso de Doutorado em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Ciências e Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Fátima Nelsi-zeuma Sombra de Medeiros

Coorientador: Prof. Dr. Geraldo Luis Bezerra Ramalho

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- F44t Ferreira, Daniel Silva.  
Triagem Automática do Exame de Papanicolaou com base em Atenção Visual / Daniel Silva Ferreira. –  
2020.  
164 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação  
em Engenharia de Teleinformática, Fortaleza, 2020.  
Orientação: Profa. Dra. Fátima Nelsizeuma Sombra de Medeiros.  
Coorientação: Prof. Dr. Geraldo Luis Bezerra Ramalho.
1. Papanicolaou. 2. Atenção Visual. 3. Saliência. 4. Redes Neurais Convolucionais. I. Título.  
CDD 621.38
-

DANIEL SILVA FERREIRA

TRIAGEM AUTOMÁTICA DO EXAME DE PAPANICOLAOU COM BASE EM ATENÇÃO  
VISUAL

Tese apresentada ao Curso de Doutorado em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Ciências e Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 25 de Novembro de 2020

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Fátima Nelsizeuma Sombra de Medeiros (Orientadora)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Geraldo Luis Bezerra Ramalho (Coorientador)  
Instituto Federal de Educação, Ciência e Tecnologia (IFCE)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Cláudia Martins Carneiro  
Universidade Federal de Ouro Preto (UFOP)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Mylène Christine Queiroz de Farias  
Universidade de Brasília (UnB)

---

Prof. Dr. Anselmo Cardoso de Paiva  
Universidade Federal do Maranhão (UFMA)

---

Prof. Dr. José Marques Soares  
Universidade Federal do Ceará (UFC)

Aos meus pais, Francisco José e Maria Aldecir.  
Ao meu filho-anjo, Bernardo.

## AGRADECIMENTOS

A Deus, por estar sempre ao meu lado, dando saúde e motivação para enfrentar as dificuldades.

Aos meus pais Francisco José e Maria Aldecir, pela dedicação empregada em minha formação.

Às minhas irmãs Raquel e Sheila, por todos os momentos de carinho e apoio.

À minha esposa, Ariadine Pimenta, pela paciência, apoio e compreensão.

À minha orientadora, Profa. Dra. Fátima N. Sombra de Medeiros, pela oportunidade, orientação e pelos valiosos conselhos e ensinamentos nesse período.

Ao Dr. Geraldo Bezerra, pela dedicação, disponibilidade e coorientação neste projeto.

À Dra. Daniela Ushizima por me acolher durante o doutorado sanduíche, pelos ensinamentos e esforços dedicados no desenvolvimento deste trabalho.

Aos colegas do projeto CRIC, em especial, Alessandra, Mariana, Cláudia e Andrea, pelo auxílio e suporte sempre que necessário.

À Débora Torres por possibilitar a realização dos experimentos de rastreamento ocular.

Aos professores Dr. José Soares de Andrade Júnior e Dr. Humberto de Andrade Carmona pela autorização de uso do equipamento de rastreamento ocular no Laboratório de Sistemas Complexos (Departamento de Física/UFC).

Aos meus amigos do Laboratório de Visão (LABVIS/UFC), pela divertida companhia diária, conversas, debates filosóficos e apoio na realização deste projeto.

Aos meus amigos do IFCE Maracanaú, pelo total apoio e incentivo neste período.

Ao PPGETI pela oportunidade de participar deste doutorado e ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) pelo apoio.

Ao CNPq pelo apoio financeiro através do projeto Pesquisador Visitante Estrangeiro (PVE) do Programa Ciência sem Fronteiras (CsF). Este projeto possibilitou a realização do doutorado sanduíche no exterior, fato que foi determinante para o desenvolvimento desta tese e para o meu crescimento profissional e pessoal.

À Universidade da Califórnia em Berkeley, ao *Berkeley Institute for Data Science* (BIDS) e ao *Lawrence Berkeley National Laboratory* (LBNL) pelo acolhimento e suporte necessários à realização do meu doutorado sanduíche e pelas contribuições dadas a este projeto.

*“The first observation of cancer cells in the smear of the uterine cervix gave me one of the greatest thrills I ever experienced during my scientific career”*

(George N. Papanicolaou)

## RESUMO

O câncer de colo do útero é um grave problema de saúde pública, sendo a quarta neoplasia maligna que mais afeta mulheres. Pesquisadores têm proposto técnicas computacionais para auxiliar o especialista na identificação de lesões pré-cancerígenas em imagens de exames de Papanicolaou. A abordagem clássica consiste na busca exaustiva por alterações celulares na imagem, classificando células sem considerar características da região em análise. Comumente, essa metodologia tem seu desempenho reduzido por falhas de classificação em regiões com sobreposição celular, ruído e artefatos sem interesse clínico. Esta tese apresenta o modelo *Attention Guided Model for inspection of cells* (AGMic) que analisa imagens de células cervicais a partir da triagem de regiões por um mecanismo de atenção. O modelo proposto identifica áreas que são atrativas a especialistas humanos, ou seja, salientes, e, então, foca a computação na classificação de células dessas sub-imagens específicas. A captura da atenção visual de citopatologistas ocorreu por experimentos de rastreamento ocular, e os dados coletados disponibilizados como uma nova base de dados, chamada *CRIC Visual Attention* (CRICVA). Esta tese introduz ainda a *Region-based Structural Co-occurrence Matrix* (rSCM) para modelar a textura do núcleo e do seu entorno, como parte de um algoritmo que associa as regiões salientes a células potencialmente anormais. Testes em imagens do Sistema Único de Saúde (SUS) revelaram que o modelo AGMic reduz falsos positivos na classificação das imagens e alcança Acurácia ( $ACC$ ) de 0,955 e índice de concordância Kappa ( $K$ ) de 0,919 em três classes (negativa, baixo risco e alto risco) usando apenas as três regiões de maior saliência. Esses resultados superaram a abordagem clássica, a qual obteve  $ACC$ : 0,775,  $K$ : 0,611 quando todos os núcleos são corretamente detectados. Uma vez que se processam menos células, o tempo médio de execução do AGMic foi de  $\sim 5,26$  segundos por imagem, sem GPU, representando uma redução de tempo de  $\sim 46,54\%$  para detecção de regiões com base na segmentação de núcleo. Experimentos complementares em 4.867 imagens de citologia de meio líquido da base pública *Deep Cervical Cytological Lesions* (DCCL) confirmaram o potencial do AGMic para a seleção de regiões de interesse clínico. Estes achados destacam o modelo proposto, habilitando-o ao uso em ferramentas para a análise de exames de Papanicolaou em laboratórios citopatológicos.

**Palavras-chave:** Papanicolaou. Atenção Visual. Saliência. Redes Neurais Convolucionais.



## ABSTRACT

Cervical cancer is a severe public health problem, being the fourth malignant neoplasm that most affects women. Researchers have proposed computational techniques to assist the clinician in identifying pre-cancerous lesions in Pap smears. The classical approach consists of an exhaustive search for cellular changes in the image, classifying cells with no considering region features under analysis. Commonly, this methodology has its performance reduced by classification failures into regions with cell overlap, noise, and artifacts with no clinical interest. This thesis presents AGMic, which analyzes images of cervical cells from the screening of regions by an attention mechanism. The proposed model identifies conspicuous areas to human specialists, i.e., salients, and then focuses computation on cell classification of these specific sub-images. The acquisition of cytopathologists' visual attention occurred through eye-tracking experiments, and the collected data made available as a novel database called CRICVA. This thesis also introduces rSCM to model the nucleus texture and its surroundings, as part of an algorithm that associates the salient regions to potentially abnormal cells. Tests on images from the Brazilian Universal Health System (SUS) revealed that AGMic reduces false positives on image classification and achieves Accuracy (*ACC*) of 0.955 and Kappa concordance index (*K*) of 0.919 in three classes (negative, low risk, and high risk) using only the three highest salient regions. These results overperformed the classical approach performance, which obtained *ACC*: 0.775, *K*: 0.611 when all nuclei are correctly detected. Since fewer cells are processed, the average runtime for AGMIC was  $\sim 5.26$  seconds per image, with no GPU, representing a time reduction of  $\sim 46.54\%$  for region detection based on nuclei segmentation. Complementary experiments on 4,867 liquid-based cytology images from the public dataset DCCL confirmed the potential of AGMic to select clinical interest regions. These findings highlight the proposed model, enabling it to be used in Papanicolaou screening tools in cytopathological laboratories.

**Keywords:** Papanicolaou. Visual Attention. Saliency. Convolutional Neural Networks.

## LISTA DE FIGURAS

- Figura 1 – Campo de visualização a 40x de uma lâmina de exame de Papanicolaou convencional. Nesta tese, esse campo é tratado como imagem de células cervicais ou imagem de exame de Papanicolaou. Célula alterada, com membrana cromatínica irregularmente espessa, obtida em (BARROS *et al.*, 2012). Fonte: elaboração própria. . . . . 35
- Figura 2 – Principais desafios para construção de sistemas de análise automática para exames de Papanicolaou. a) Artefatos presentes nas imagens que resultam de falhas na confecção, coloração e montagem das lâminas, respectivamente. b) Falhas na segmentação de núcleo e citoplasma, exemplificado pelos resultados do método proposto em (ZHANG *et al.*, 2014). c) Elementos comuns em esfregaços cérvico-vaginais, que não derivam de falhas na preparação das lâminas, mas que comumente reduzem o desempenho de métodos computacionais por serem confundidos com objetos de interesse clínico para o câncer do colo do útero. Para detalhes sobre essas estruturas recomendamos a leitura de (WHO; IARC, 2004). Fonte: elaboração própria a partir de imagens obtidas em <<http://www.pro-celula.com.br/home/atlascitologico/atlas/texto/cit-cervice-elem.htm>> . . . . . 38
- Figura 3 – Imagens de células cervicais a 40x disponíveis na base Centro de Reconhecimento para Inspeção de Células (CRIC). Os rótulos (Normal, ASC-US, LSIL, ASC-H, HSIL e Carcinoma) sinalizam a classe do principal achado clínico na imagem, enquanto que as marcações em cores identificam esses achados, se existente, nas diferentes células. As linhas tracejadas simbolizam como as classes do Sistema Bethesda podem ser agrupadas dentro da rotina clínica. Os achados considerados são os reportados na Tabela 1. Fonte: elaboração própria. . . . . 41
- Figura 4 – Distribuição de casos na base CRIC. a) Casos de acordo com o rótulo da imagem, ou seja, classe do principal achado detectado no campo. Total de imagens (campos): 400. b) Distribuição de classes considerando todas as células. Total de células anotadas: 11.473. Fonte: elaboração própria. . . . . 42

Figura 5 – Estrutura típica de um modelo computacional para estimar a atenção visual humana. A saída do sistema é um mapa de saliência, onde a intensidade dos elementos revelam as regiões com maior atratividade. Fonte: elaboração própria a partir de (ITTI; KOCH, 2001) e < <a href="http://www.lizhaoping.org/zhaoping/prints/CosyneForPost.pdf">http://www.lizhaoping.org/zhaoping/prints/CosyneForPost.pdf</a> >. . . . .	45
Figura 6 – Exemplos de padrões psicofísicos que podem ser utilizados para testar a reação de atenção de um observador. Em todos os casos, a modelagem computacional de apenas um atributo a) orientação, b) cor, c) forma ou d) textura pode estimar a atenção <i>bottom-up</i> humana. Fonte: elaboração própria a partir de (ZHANG; LIN, 2013). . . . .	47
Figura 7 – Movimentos sacádicos e fixações oculares do experimento conduzido por Yarbus (1967). As gravações duraram 3 minutos para (b)-(d). a) Cena de pessoas de uma família em uma sala. b) Rastreamento ocular dos participantes em observação livre (sem qualquer pergunta ou tarefa). c) Movimentos oculares quando os participantes são questionados sobre a idade dos indivíduos da cena e d) quando são levados a lembrar da posição dos objetos e pessoas. Fonte: Imagens obtidas em (ZHANG; LIN, 2013). . . . .	49
Figura 8 – Funções básicas que compõem a arquitetura de uma <i>Convolutional Neural Network</i> (CNN). a) Função de não linearidade ( <i>Rectified Linear Unit</i> (ReLU)) e b) função <i>max-pooling</i> para redução de dimensionalidade espacial dos mapas de características. Fonte: adaptado de < <a href="https://cs231n.github.io/convolutional-networks/##pool">https://cs231n.github.io/convolutional-networks/##pool</a> > . . . . .	51
Figura 9 – Estrutura típica de uma arquitetura CNN. No bloco em vermelho estão as características extraídas pelas camadas convolucionais. Fonte: adaptado de (LI <i>et al.</i> , 2017). . . . .	51

Figura 10 – Estratégias comumente adotadas no desenvolvimento de sistemas computacionais para auxílio à análise de exames citopatológicos do colo do útero. a) Abordagem típica, utiliza a segmentação do núcleo e/ou citoplasma para extração de características das células. b) Esquema baseado em um detector de núcleos, não sendo necessário estimar bordas com precisão. As marcações em verde e laranja representam o *Ground Truth* (GT) e os núcleos detectados, respectivamente. c) Utiliza seleção de regiões de interesse para categorização, sem envolver segmentações. Empregamos os resultados dos métodos propostos por Ramalho *et al.* (2015), Phoulady e Mouton (2018), Liang *et al.* (2019) para as ilustrações de (a), (b) e (c), respectivamente. A abordagem em destaque foi a empregada nesta tese. Fonte: elaboração própria. . . . . 60

Figura 11 – Coleta dos dados de atenção visual dos citopatologistas. a) Imagens da Base CRIC com os respectivos rótulos. Os rótulos não foram apresentados aos especialistas. b) Metodologia para selecionar as imagens reduzindo o *viés de seleção* (TORRALBA; EFROS, 2011). c) Uso de um equipamento de rastreamento ocular para coletar os movimentos oculares dos especialistas. d) Dados contidos na nova base de dados CRICVA. e) Imagem colorida de células cervicais. f) Fixações oculares dos citopatologistas. g) Mapa de atenção visual resultante da convolução de uma gaussiana em cada ponto de fixação. h) Regiões salientes da imagem reveladas pela sobreposição da imagem apresentada com o respectivo mapa de atenção. As regiões em vermelho escuro correspondem às áreas com maior atratividade e as em azul escuro, as que demandam menor atenção. Fonte: elaboração própria. . . . . 63

Figura 12 – Amostras da base de dados CRICVA. Os pontos do mapa de fixação foram dilatados para melhor visualização. Nos mapas de atenção, as regiões em vermelho escuro correspondem às áreas com maior atratividade e as em azul escuro, caso contrário. Os rótulos das células estão disponíveis na base CRIC (CRIC, 2018). Referência (Cor: Alteração): Verde: Negativa, Amarelo: ASC-US, Rosa: LSIL, Vermelho: ASC-H, Roxo: HSIL e Preto: Carcinoma. Fonte: elaboração própria. . . . . 66

Figura 13 – Distribuição de casos na base CRICVA, conforme anotações da base CRIC. a) Casos de acordo com o rótulo da imagem, ou seja, classe do principal achado detectado no campo. Total de imagens (campos): 232. b) Distribuição de classes considerando todas as células. Total de células anotadas: 5.654. Fonte: elaboração própria. . . . .	67
Figura 14 – Análise do <i>viés de centro</i> da base de dados CRICVA. a) <i>Average Annotation Map</i> (AAM). Linhas b) horizontal e c) vertical no centro do AAM contrastando com as obtidas a partir de uma gaussiana isotrópica no centro da imagem (GC). Fonte: elaboração própria. . . . .	69
Figura 15 – Modelo AGMic. O primeiro módulo seleciona células de interesse baseado na predição de saliência. O segundo módulo realiza a classificação dessas regiões e sugere um rótulo para a imagem de entrada. Os mapas de saliência são apresentados coloridos apenas para ilustração. Fonte: elaboração própria a partir de uma amostra da Base CRIC. . . . .	70
Figura 16 – Arquitetura do modelo <i>Saliency Attentive Model</i> (SAM) utilizado como a componente $f(\cdot)$ do modelo proposto AGMic. Os mapas de saliência correspondem a matrizes bidimensionais, as quais são apresentadas coloridas para melhor visualização. Fonte: adaptado de (CORNIA <i>et al.</i> , 2018). . . . .	73
Figura 17 – Exemplo de ROI posicionada no centro das células (quadrado com bordas azuis) quando utilizado o pico da saliência (marcador branco) para definir localização. Os quadrados com bordas cinzas sinalizam as células rotuladas pelos especialistas. O mapa de atenção mostrado corresponde aos dados coletados dos especialistas pelo equipamento de rastreamento ocular em sobreposição com a imagem de entrada. A ROI estimada (quadrado com bordas azuis) não contempla satisfatoriamente a região do núcleo. Fonte: elaboração própria. . . . .	76
Figura 18 – Ilustração do cálculo da matriz de coocorrência estrutural proposta (rSCM). a) Imagem de entrada $c$ , onde $i = Q(c)_p \in [0, \dots, N_Q - 1]$ . b) Imagem de entrada modificada $\bar{c}$ , onde $j = Q(\bar{c})_p \in [0, \dots, N_Q - 1]$ . c) Máscara binária para a região $U$ , onde 1 simboliza as posições que devem ser consideradas. d) Matriz rSCM de dimensão $N_Q \times N_Q$ , contabilizando as coocorrências do par $(i, j)$ . Fonte: elaboração própria. . . . .	77

Figura 19 – Exemplo da Matriz rSCM para imagens de entrada quantizadas em 8 níveis. As coocorrências são computadas apenas na região $U$ . Fonte: elaboração própria. . . . .	77
Figura 20 – Interpretação da Matriz rSCM. Fonte: elaboração própria. . . . .	78
Figura 21 – Alterações estruturais em célula cervical normal e anormal que são capturadas pela rSCM. Os gráficos à direita destacam o impacto em uma amostra de textura do núcleo (linha tracejada vermelha) resultante da aplicação do filtro gaussiano com $\sigma = 4$ . O Erro Médio Absoluto (EMA) entre as linhas quantifica as diferenças estruturais (área em cinza), apontando maior sensibilidade à filtragem passa-baixa para os casos anormais. Fonte: elaborada pelo autor. Imagens obtidas da base Herlev (JANTZEN <i>et al.</i> , 2005). . . . .	79
Figura 22 – Espalhamento dos coeficientes não nulos da rSCM quando anormalidade dos núcleos de células cervicais. a) Indicativo de alteração na intensidade dos <i>pixels</i> . b) Indicativo de alteração na textura e c) no tamanho do núcleo. A seta vermelha sinaliza amplitude do ponto $(i, j)$ na matriz proposta. Fonte: elaboração própria. . . . .	80
Figura 23 – Projeção das matrizes rSCM no espaço <i>t-distributed Stochastic Neighbor Embedding</i> (t-SNE) com duas componentes principais (PCA). As amostras selecionadas ilustram os extremos da nuvem de pontos. De acordo com os rótulos de classes disponíveis na base Herlev temos: 1) displasia leve, 2) normal intermediária, 3) normal superficial, 4) normal intermediária, 5) normal colunar, 6) displasia moderada, 7) carcinoma <i>in situ</i> e 8) displasia leve. As marcações nas imagens sinalizam as bordas dos núcleos, conforme o GT, considerando verde para casos normais e vermelho para anormais. Fonte: elaboração própria. . . . .	81
Figura 24 – Comportamento da matriz rSCM em relação ao a) tamanho do núcleo (energia) e b) textura (AP). As linhas tracejadas destacam a mediana para os casos anormais. Medições computadas a partir das 917 imagens disponíveis na base Herlev. Fonte: elaboração própria. . . . .	82

Figura 25 – Estimação da saliência local. Aplicação do algoritmo rSCM para identificar estruturas de interesse. a) Recorte do mapa de saliência estimado globalmente para a imagem de entrada $I$ que equivale ao recorte $I_{crop}$ . b) Mapa obtido a partir de $g(\cdot)$ (Figura 15), enfatizando áreas conforme o índice de interesse $\rho$ . Modulando (a) com (b), alcançamos o mapa de saliência do recorte. Fonte: elaboração própria. A imagem que representa $I_{crop}$ corresponde a uma amostra da base de dados CRIC. . . . .	83
Figura 26 – Arquitetura VGG-16 modificada proposta como Módulo 2. Fonte: elaboração própria. . . . .	85
Figura 27 – Seleção de dados para treinamento do Módulo 2. Fonte: elaboração própria.	86
Figura 28 – Estratégia de <i>data augmentation</i> adotada para treinamento do Modulo 2 do AGMic. Fonte: elaboração própria a partir de recortes de uma imagem da Base CRIC. . . . .	87
Figura 29 – Acurácia ao longo das épocas de treinamento. a) Treino da Etapa A e b) treino da Etapa B. Fonte: elaboração própria. . . . .	88
Figura 30 – Etapas para investigação das hipóteses H1 e H2. 1) Avaliação dos modelos de predição de saliência utilizando métricas estado-da-arte (testar H1) e 2) análise da seleção de Regiões de Interesse (ROIs) para determinar se a alteração que define o rótulo da imagem está destacada pelos mapas de saliência (testar H2). Fonte: elaboração própria. . . . .	89
Figura 31 – Construção da base de dados necessária aos experimentos de predição de saliência. A cardinalidade # representa o quantitativo de imagens existentes em cada conjunto utilizado. Fonte: elaboração própria. . . . .	94
Figura 32 – Mapas de saliência, sobrepostos às imagens de entrada, computados pelos métodos estudados. a) - f) mostram os resultados para cada caso descrito na base CRICVA. Alguns modelos são capazes de indicar regiões conspícuas da imagem, enquanto outros podem ser ineficazes na previsão do olhar do citopatologista, quando as imagens em análise são ruidosas e/ou possuem elevada sobreposição celular. Fonte: elaboração própria. . . . .	100

Figura 33 – Avaliação quantitativa dos métodos a) <i>bottom-up</i> e b) <i>top-down</i> , incluindo os baseados em CNN. A linha horizontal pontilhada representa a faixa interquartil do melhor modelo. Para cada métrica, a linha colorida horizontal marca o desempenho do melhor algoritmo <i>bottom-up</i> . Fonte: elaboração própria. . . . .	101
Figura 34 – Comparações pareadas para todos os métodos analisados usando o teste estatístico de Kruskal-Wallis com o teste <i>post-hoc</i> de Nemenyi. As caixas pretas representam os pares com diferença significativa em $\alpha = 0,05$ . Fonte: elaboração própria. . . . .	102
Figura 35 – Coeficiente de correlação entre as ativações das redes a) VGG-16 e b) ResNet-50 treinadas com amostras da base SALICON e CRICVA. As ativações apresentam correlação acima de 95% nas primeiras camadas, lado esquerdo da linha vertical, em ambas arquiteturas. Fonte: elaboração própria. . . . .	105
Figura 36 – Resultados do modelo AGMic. A seta indica exemplo de falso negativo local que não interfere na classificação da imagem. PC: Classe predita para a imagem. T: Rótulo da imagem a partir dos citologistas (GT). Fonte: elaboração própria a partir de amostras da base de dados CRIC. . . . .	111
Figura 37 – Resultados do modelo AGMic para classificação das imagens e das ROIs em: negativa, baixo risco e alto risco. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria. . . . .	112
Figura 38 – Resultados do modelo AGMic para classificação dos campos (imagens) nas classes: negativa, baixo risco e alto risco. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria. . . . .	113
Figura 39 – Resultados do modelo AGMic para classificação binária: imagens positivas e negativas. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Ambos os gráficos apresentam o desempenho do modelo proposto na classificação dos campos (imagens) e das ROIs localmente. Fonte: elaboração própria. . . . .	114
Figura 40 – Resultados do modelo AGMic para classificação binária das imagens: positiva e negativa. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria. . . . .	116



Figura 41 – Resultados da medida <i>Hit Degree</i> (HD). O uso do algoritmo de refinamento permitiu que o modelo proposto aumentasse a sobreposição entre as ROIs estimadas e as células rotuladas da base CRIC. Fonte: elaboração própria. . . . .	116
Figura 42 – Quantitativo de ROIs estimadas por GT. a) mostra a distribuição sem uso do algoritmo de refinamento e em b) com uso. O algoritmo de refinamento permitiu uma redução de 1,6% em regiões não rotuladas e o aumento de 0,3% para células com alto risco. Fonte: elaboração própria. . . . .	117
Figura 43 – Imagens de células cervicais em meio líquido da base DCCL. Os rótulos na vertical sinalizam o rótulo da imagem, enquanto que as marcações em cores localizam a alteração celular, se existente. Fonte: elaboração própria. . . . .	120
Figura 44 – Distribuição das amostras da base DCCL. a) Quanto à classificação da imagem. Total de imagens: 4.867. b) Quanto às ROIs. Total de regiões rotuladas: 9.723. Fonte: elaboração própria. . . . .	121
Figura 45 – Resultados do Módulo 1 do AGMic em cada classe da base DCCL. As cores do GT representam: verde: normal, amarelo: ASC-US, rosa: LSIL, vermelho: ASC-H, roxo: HSIL e preto: carcinoma. As zonas de atenção estão ranqueadas pela energia do mapa de saliência. O refinamento (etapa local do AGMic) tende a alinhar melhor as células selecionadas à posição definida pelos citopatologistas. Fonte: elaboração própria. . . . .	123
Figura 46 – Distribuição das regiões rotuladas pelos especialistas no ranque das ROIs preditas, conforme energia no mapa de saliência. Aproximadamente a) 65,33% das imagens negativas, b) 71,94% das imagens de baixo risco e c) 54,05% das imagens de alto risco possuem as células sinalizadas nas regiões mais salientes. As porcentagens em cinza indicam a proporção de imagens, assumindo como referência o total de 1.292, 2.341 e 1.234 imagens negativas, baixo risco e alto risco processadas, respectivamente. Fonte: elaboração própria. . . . .	124
Figura 47 – Resultados da medida HD. O algoritmo de refinamento permitiu aumentar a sobreposição entre as ROIs estimadas e as células rotuladas da base DCCL. Fonte: elaboração própria. . . . .	125

Figura 48 – Amostras de recortes de células cervicais, com suas respectivas segmentações, disponíveis na base de dados Herlev. A segmentação do núcleo está representada em azul claro. A área de citoplasma é marcada por azul escuro e as demais cores mostram o fundo ( <i>background</i> ). O agrupamento das 7 classes em normal e anormal segue a proposta de Jantzen <i>et al.</i> (2005). Fonte: elaboração própria a partir dos dados disponíveis na base Herlev (JANTZEN <i>et al.</i> , 2005). . . . .	148
Figura 49 – Metodologia para análise do descritor rSCM como extrator de características de células cervicais normais e anormais. Fonte: elaboração própria. . . . .	150
Figura 50 – Análise da matriz rSCM como descritor de células cervicais. a) Desempenho com máscara de segmentação do núcleo de diferentes tamanhos ( <i>rm</i> ), destacando a abertura do filtro gaussiano ( $\sigma$ ) de melhor desempenho. b) Resultados para os valores de $\sigma$ analisados, considerando $rm = 2$ e c) histograma das 1000 rodadas de classificação para o par ( $rm = 2, \sigma = 4$ ). A linha azul destaca o melhor resultado conforme valor de mediana. Os elementos em cinza correspondem à mesma população de valores de Acurácia ( <i>ACC</i> ). Fonte: elaboração própria. . . . .	152
Figura 51 – Exemplos de células cervicais destacando a área nuclear periférica (bordas brancas) processada pelo algoritmo rSCM. A área em azul representa o núcleo. Fonte: elaboração própria a partir dos dados da base Herlev (JANTZEN <i>et al.</i> , 2005). . . . .	153
Figura 52 – Resultados de a) Acurácia ( <i>ACC</i> ) e b) Kappa ( <i>K</i> ) para os descritores analisados na caracterização de células cervicais normais e anormais. A linha tracejada destaca o desempenho do algoritmo proposto. C#{1,2,3} identifica a categoria por grupo de descritores. M1: (PLISSITI; NIKOU, 2012), M2: (GENÇTAV <i>et al.</i> , 2012), M3: (SARWAR <i>et al.</i> , 2015), M4: (MARINAKIS <i>et al.</i> , 2009), M5: (CHEN <i>et al.</i> , 2014), M6: (CHANKONG <i>et al.</i> , 2014) e M7: (MARIARPUTHAM; STEPHEN, 2015). Fonte: elaboração própria. . . . .	156

Figura 53 – Análise em pares dos descritores usando o teste estatístico de Kruskal-Wallis com o teste de Nemenyi. a) Acurácia ( $ACC$ ) e b) Kappa ( $K$ ). Os quadrados pretos representam os pares com significativa diferença na caracterização das células cervicais ( $\alpha = 0,05$ ). M1: (PLISSITI; NIKOU, 2012), M2: (GENÇTAV *et al.*, 2012), M3: (SARWAR *et al.*, 2015), M4: (MARINAKIS *et al.*, 2009), M5: (CHEN *et al.*, 2014), M6: (CHANKONG *et al.*, 2014) e M7: (MARIARPUTHAM; STEPHEN, 2015). Fonte: elaboração própria. . 157

## LISTA DE TABELAS

Tabela 1	– Sistema Bethesda e sua, respectiva, macro categorização empregada na prática clínica. Fonte: elaboração própria a partir de (NAYAR; WILBUR, 2015a).	36
Tabela 2	– Modelos de predição de saliência. Fonte: elaboração própria.	91
Tabela 3	– Principais características das métricas de avaliação de predição de saliência empregadas nesta tese. A coluna <i>Categoria</i> informa se o mapa de referência GT equivale ao mapa de fixações oculares (Localização) ou mapa de atenção (Distribuição). Fonte: elaboração própria a partir de (BYLINSKII <i>et al.</i> , 2018).	97
Tabela 4	– Ranque do tempo médio de processamento dos métodos de predição de saliência estudados (segundos por imagem). Os algoritmos baseados em CNN estão em negrito. {Mat: Matlab, Py: Python}. Fonte: elaboração própria.	106
Tabela 5	– Coeficiente de concordância $\phi$ (Equação 7.5) entre as $nr = 3$ regiões mais salientes e o achado clínico de maior relevância da imagem. Os melhores resultados para cada alteração celular estão em negrito. Fonte: elaboração própria.	107
Tabela 6	– Resultados comparativos entre o AGMic e a estratégia clássica de busca por lesões em imagens de células cervicais. Classificação das imagens (campos) e das ROIs localmente em três classes: negativa, baixo risco e alto risco. Os valores para a medida Acurácia ( $ACC$ ) e para o índice de concordância Kappa ( $K$ ) são reportados. Fonte: elaboração própria.	118
Tabela 7	– Interpretação adotada para os valores do coeficiente Kappa. Fonte: elaboração própria a partir de (SILVA <i>et al.</i> , 2019).	149
Tabela 8	– Relação dos descritores avaliados e sua respectiva categorização quanto à entrada. Fonte: elaboração própria.	154
Tabela 9	– Ranque de resultados dos descritores quanto à classificação das células cervicais em normais e anormais. O valor da mediana para Acurácia ( $ACC$ ) e Kappa ( $K$ ) são reportados. O resultado do descritor proposto está destacado em cinza. IC corresponde ao intervalo de confiança (95%) para cada decritor. Os métodos estão listados na sequência do melhor para o pior desempenho. Cat: Categoria do método. Fonte: elaboração própria.	155

## LISTA DE ABREVIATURAS E SIGLAS

<i>Attentive ConvLSTM</i>	<i>Attentive Convolutional Long Short-Term Memory</i>
AAM	<i>Average Annotation Map</i>
ACS	<i>American Cancer Society</i>
AGMic	<i>Attention Guided Model for inspection of cells</i>
AUC	Área sob a curva ROC ( <i>Receiver Operating Characteristic</i> )
CAD	<i>Computer Aided Diagnosis</i>
CC	<i>Linear Correlation Coefficient</i>
CNN	<i>Convolutional Neural Network</i>
CRIC	Centro de Reconhecimento para Inspeção de Células
CRICVA	<i>CRIC Visual Attention</i>
DCCL	<i>Deep Cervical Cytological Lesions</i>
EMA	Erro Médio Absoluto
Faster R-CNN	<i>Region-based Convolutional Neural Network</i>
FPN	<i>Feature Pyramid Network</i>
GT	<i>Ground Truth</i>
HD	<i>Hit Degree</i>
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
INCA	Instituto Nacional do Câncer
IP	Índice de Positividade
KL	<i>Kullback-Leibler divergence</i>
LAPAC	Laboratório Piloto de Análises Clínicas
MEQ	Monitoramento Externo da Qualidade
MIQ	Monitoramento Interno da Qualidade
MIT	<i>Massachusetts Institute of Technology</i>
NSS	<i>Normalized Scanpath Saliency</i>
OMS	Organização Mundial da Saúde
QualiCito	Qualificação Nacional em Citopatologia
R-FCN	<i>Region-based Fully Convolutional Networks</i>
ReLU	<i>Rectified Linear Unit</i>
RNA	Redes Neurais Artificiais
ROIs	Regiões de Interesse

RPNs	<i>Region Proposal Networks</i>
rSCM	<i>Region-based Structural Co-occurrence Matrix</i>
SAM	<i>Saliency Attentive Model</i>
SCM	<i>Structural Co-occurrence Matrix</i>
SD	Desvio Padrão
SIM	<i>Similarity</i>
SRC	<i>SAM ResNet-50 for Cells</i>
SRS	<i>SAM ResNet-50 for natural images</i>
SUN	<i>Saliency Using Natural statistics</i>
SUS	Sistema Único de Saúde
SVC	<i>SAM VGG-16 for Cells</i>
SVS	<i>SAM VGG-16 for natural images</i>
t-SNE	<i>t-distributed Stochastic Neighbor Embedding</i>
TBS	Sistema Bethesda
TV	<i>Total Variation</i>

## LISTA DE SÍMBOLOS

$I$	Imagem de entrada
$I_{crop}$	Recorte da imagem de entrada $I$
$f$	Modelo computacional para computação do mapa de saliência global
$g$	Método computacional para computação do mapa de saliência local
$M$	Mapa de saliência global obtido a partir de $f(\cdot)$
$S^{local}$	Mapa de saliência local obtido a partir de $g(\cdot)$
$Z$	Mapa das zonas de atenção estimado a partir do mapa de saliência global
$z_i$	$i$ -ésima zona de atenção
$V_c$	Conjunto de células de interesse
$c_i$	<i>Bounding box</i> que demarca a $i$ -ésima célula de interesse
$\bar{c}$	Versão de $c$ alterada por uma função de transformação
$W$	Pesos do modelo de predição
$y^{fix}$	Mapa de fixações oculares
$y^{aten}$	Mapa de atenção obtido por meio do mapa de fixações oculares
$U$	Área da imagem onde a computação da matriz rSCM deve ocorrer
$U_{circ_i}$	Circularidade da $i$ -ésima região $U$
$U_{prob_i}$	Probabilidade da $i$ -ésima região $U$ ser anormal
$\rho_i$	Índice de interesse da $i$ -ésima região $U$
$\tau_{circ}$	Limiar de circularidade
$Q$	Quantização ou outra função para agrupamento de <i>pixels</i>
$N_Q$	Quantidade de níveis de quantização ou de grupos de <i>pixels</i>
$P_U$	Propriedade entre os <i>pixels</i> de $c$ e $\bar{c}$
$p$	Posição de um elemento em uma matriz bidimensional
$K$	Coefficiente <i>Kappa</i>
$ACC$	Acurácia
$rm$	Raio de um elemento estruturante circular

$\sigma$	Desvio padrão para o filtro passa-baixa gaussiano
$\mu$	Média amostral
$\alpha$	Nível de significância
$nr$	Quantidade de regiões mais salientes
$\varphi$	Taxa de concordância entre as regiões salientes e a alteração celular
$l_e$	Alteração celular mais significativa em uma imagem em análise
$N_{l_e}$	Quantidade de imagens no conjunto de dados com a alteração $l_e$
$TP$	Verdadeiro Positivo
$FP$	Falso Positivo
$FN$	Falso Negativo
$PREC$	Precisão
$REV$	Revocação
$F$	Medida $F$ ( $F$ -measure)
$A$	Matriz binária sinalizando uma região de interesse predita
$G$	Matriz binária sinalizando a região da alteração $l_e$
$B$	Matriz binária sinalizando todas as células rotuladas pelos especialistas em uma imagem de células cervicais
$pRanq$	Posição de ranqueamento



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>27</b>
<b>1.1</b>	<b>Motivação</b>	<b>27</b>
<b>1.2</b>	<b>Objetivos da Tese</b>	<b>30</b>
<b>1.3</b>	<b>Hipóteses</b>	<b>31</b>
<b>1.4</b>	<b>Contribuições</b>	<b>31</b>
<b>1.5</b>	<b>Produção Científica</b>	<b>31</b>
<b>1.5.1</b>	<i>Produção Científica Complementar</i>	<b>32</b>
<b>1.6</b>	<b>Organização da Tese</b>	<b>32</b>
<b>2</b>	<b>EXAME DE PAPANICOLAOU</b>	<b>34</b>
<b>2.1</b>	<b>Principais desafios para alcançar uma análise automatizada</b>	<b>37</b>
<b>2.2</b>	<i>Center for Recognition and Inspection of Cells (CRIC)</i>	<b>39</b>
<b>2.2.1</b>	<i>Base de Imagens CRIC</i>	<b>40</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>43</b>
<b>3.1</b>	<b>Atenção Visual Seletiva</b>	<b>43</b>
<b>3.1.1</b>	<i>Modelagem Computacional</i>	<b>44</b>
<b>3.1.2</b>	<i>Predição de Saliência</i>	<b>45</b>
<b>3.1.2.1</b>	<i>Abordagens Bottom-up e Top-down</i>	<b>46</b>
<b>3.2</b>	<b>Redes Neurais Convolucionais</b>	<b>49</b>
<b>4</b>	<b>REVISÃO DA LITERATURA</b>	<b>55</b>
<b>4.1</b>	<b>Predição de Saliência em Imagens Médicas</b>	<b>55</b>
<b>4.2</b>	<b>Categorização de Exames de Papanicolaou com Base em Regiões de Interesse</b>	<b>58</b>
<b>5</b>	<b>BASE DE DADOS DE ATENÇÃO VISUAL PROPOSTA</b>	<b>62</b>
<b>5.1</b>	<b>Rastreamento da Atenção Visual dos Citopatologistas</b>	<b>62</b>
<b>5.1.1</b>	<i>Seleção de Imagens</i>	<b>62</b>
<b>5.1.2</b>	<i>Participantes</i>	<b>64</b>
<b>5.1.3</b>	<i>Experimento de Rastreamento Ocular</i>	<b>64</b>
<b>5.2</b>	<b>CRIC Visual Attention (CRICVA)</b>	<b>65</b>
<b>5.2.1</b>	<i>Consistência entre os Participantes</i>	<b>68</b>
<b>5.2.2</b>	<i>Viés de Centro</i>	<b>68</b>

<b>6</b>	<b>MODELO PROPOSTO: ATTENTION GUIDED MODEL FOR INSPECTION OF CELLS (AGMIC)</b> . . . . .	<b>70</b>
<b>6.1</b>	<b>Módulo 1: Detecção de Regiões de Interesse com Base em Atenção Visual</b>	<b>72</b>
<b>6.1.1</b>	<b>Etapa Global</b> . . . . .	<b>72</b>
<b>6.1.1.1</b>	<i>Treinamento do Modelo</i> . . . . .	<b>73</b>
<b>6.1.2</b>	<b>Etapa Local</b> . . . . .	<b>75</b>
<b>6.1.2.1</b>	<i>Region-based Structural Co-occurrence Matrix (rSCM)</i> . . . . .	<b>75</b>
<b>6.1.2.1.1</b>	<i>Matriz rSCM aplicada a Imagens de Células Cervicais</i> . . . . .	<b>78</b>
<b>6.1.2.2</b>	<i>Refinamento dos Mapas de Saliência</i> . . . . .	<b>83</b>
<b>6.2</b>	<b>Módulo 2: Classificação das Regiões de Interesse</b> . . . . .	<b>84</b>
<b>7</b>	<b>RESULTADOS EXPERIMENTAIS</b> . . . . .	<b>89</b>
<b>7.1</b>	<b>Experimentos de Predição de Saliência</b> . . . . .	<b>89</b>
<b>7.1.1</b>	<i>Modelos Avaliados</i> . . . . .	<b>90</b>
<b>7.1.2</b>	<i>Base de Imagens</i> . . . . .	<b>93</b>
<b>7.1.3</b>	<i>Métricas de Avaliação</i> . . . . .	<b>94</b>
<b>7.1.4</b>	<i>Análise da Hipótese H1</i> . . . . .	<b>98</b>
<b>7.1.4.1</b>	<i>Modelos bottom-up</i> . . . . .	<b>99</b>
<b>7.1.4.2</b>	<i>Modelos top-down</i> . . . . .	<b>103</b>
<b>7.1.4.3</b>	<i>Tempo de processamento</i> . . . . .	<b>104</b>
<b>7.1.5</b>	<i>Análise da Hipótese H2</i> . . . . .	<b>106</b>
<b>7.2</b>	<b>Experimentos com AGMic</b> . . . . .	<b>107</b>
<b>7.2.1</b>	<i>Base de Imagens</i> . . . . .	<b>108</b>
<b>7.2.2</b>	<i>Medidas de Avaliação</i> . . . . .	<b>109</b>
<b>7.2.3</b>	<i>Análise da Hipótese H3</i> . . . . .	<b>110</b>
<b>7.3</b>	<b>Experimentos com Citologia em Meio Líquido</b> . . . . .	<b>119</b>
<b>7.3.1</b>	<i>Base de Imagens</i> . . . . .	<b>119</b>
<b>7.3.2</b>	<i>Medidas de Avaliação</i> . . . . .	<b>121</b>
<b>7.3.3</b>	<i>Análise</i> . . . . .	<b>122</b>
<b>8</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS</b> . . . . .	<b>126</b>
<b>8.1</b>	<b>Trabalhos Futuros</b> . . . . .	<b>129</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>132</b>
	<b>APÊNDICES</b> . . . . .	<b>145</b>

	<b>APÊNDICE A</b> – Formulário de Consentimento . . . . .	145
	<b>APÊNDICE B</b> – <i>Region-based Structural Co-occurrence Matrix</i> (rSCM) aplicada à caracterização de células cervicais . . . . .	147
<b>B.1</b>	<b>Base de Imagens</b> . . . . .	147
<b>B.2</b>	<b>Medidas de Avaliação</b> . . . . .	148
<b>B.3</b>	<b>Metodologia</b> . . . . .	149
<b>B.4</b>	<b>Resultados</b> . . . . .	150
<b>B.5</b>	<b>Conclusões</b> . . . . .	157
	<b>ANEXOS</b> . . . . .	158
	<b>ANEXO A</b> – Protocolo de Aprovação no Comitê de Ética da UFC . . . . .	159

# 1 INTRODUÇÃO

Atenção visual seletiva consiste de um conjunto de processos cognitivos que permite focar em regiões ou objetos de interesse enquanto ignoram estímulos irrelevantes no ambiente. Essa característica conduz humanos e outros animais à extração de informações importantes do ambiente, facilitando o entendimento de cenas complexas (CARRASCO, 2011). Ao longo dos anos, modelos computacionais foram desenvolvidos com o objetivo de simular a atenção visual humana em diversos processos (ZHANG; LIN, 2013; NGUYEN *et al.*, 2018), dentre os quais, a análise de imagens médicas. Nessa área, esses modelos, em conjunto com dados de rastreamento ocular coletados durante a inspeção manual, têm favorecido o entendimento da atenção seletiva de especialistas (MATSUMOTO *et al.*, 2011; WEN *et al.*, 2017; HELBREN, 2019) e o desenvolvimento de sistemas de diagnóstico por computador (LI *et al.*, 2016; GUAN *et al.*, 2018; LOUKAS *et al.*, 2018; LI *et al.*, 2019). No entanto, a literatura carece de estudos de como esses modelos podem contribuir para automatizar a análise clínica de exames de Papanicolaou, que é a principal ferramenta de rastreamento para detecção precoce do câncer de colo do útero (COX, 2012; INCA, 2019).

Esta tese apresenta um modelo computacional, baseado na predição da atenção visual do citopatologista, para auxílio na triagem de imagens de exames de Papanicolaou. O termo triagem é utilizado nesta tese para refletir o processo de *seleção* dos campos alterados de exames de Papanicolaou realizados por citopatologistas em escrutínio de rotina. O modelo proposto seleciona Regiões de Interesse (ROIs), em conformidade com dados de rastreamento ocular previamente coletados, e então foca na computação de rotinas de categorização de células em sub-imagens para identificar campos com alterações. A abordagem empregada é inspirada no mecanismo natural empregado por especialistas humanos: observação do campo, foco em regiões com características clínicas relevantes, anotação de lesões se existentes. Além disso, esta tese apresenta novos achados que estão alinhados aos de trabalhos publicados recentemente (LI *et al.*, 2019) no que diz respeito à importância de reduzir o volume de dados, por meio da seleção de regiões de interesse, a ser processados por sistemas de visão computacional.

## 1.1 Motivação

O câncer do colo do útero, também conhecido por câncer cervical, é a quarta neoplasia maligna que mais afeta mulheres em todo mundo com 570.000 casos diagnosticados

e 311.000 mortes a cada ano (BRAY *et al.*, 2018). Esses números são inferiores apenas aos relacionados ao câncer de mama, colorretal e pulmão. No Brasil, o Instituto Nacional do Câncer (INCA) reporta o câncer cervical como sendo o terceiro tumor maligno mais frequente em mulheres, também atrás do câncer de mama e colorretal, e a quarta causa de morte de mulheres brasileiras por câncer. Para cada ano do triênio 2020-2022, o INCA estima a ocorrência de 16.590 novos casos (INCA, 2019).

A Organização Mundial da Saúde (OMS) recomenda programas de rastreio e vacinação para reduzir a mortalidade feminina por neoplasia do colo do útero (WHO, 2019). No Brasil, a principal estratégia de rastreamento<sup>1</sup> é o exame citopatológico do colo do útero, ou teste de Papanicolaou, que consiste basicamente em coletar amostras de células do colo uterino para posterior análise no microscópio. De acordo com a OMS, com uma cobertura da população-alvo de, no mínimo, 80% e a garantia de diagnóstico e tratamento adequados dos casos alterados é possível reduzir com a aplicação deste exame, em média, de 60 a 90% a incidência do câncer cervical. Mesmo nos casos onde a neoplasia ocorre, estudos da *American Cancer Society* (ACS) (ACS, 2019a) reportam que se o diagnóstico acontecer no estágio inicial da doença, a pessoa terá 92% de chance de sobreviver por pelo menos mais 5 anos. Por outro lado, se o diagnóstico for no estágio mais avançado, a probabilidade de sobrevivência cai para 17%.

Segundo a ACS (ACS, 2019b), apenas 45% das mulheres com câncer do colo do útero são diagnosticadas precocemente. Dentre as razões, está o fato da análise em laboratório das lâminas do exame de Papanicolaou ser uma tarefa complexa, que exige muita mão-de-obra e cujo resultado depende inteiramente do julgamento humano (KOSS, 1989). Segundo Greer (1997) um esfregaço anormal pode conter apenas algumas células alteradas espalhadas pelas milhares de células normais. A análise cuidadosa de um exame de Papanicolaou é, portanto, uma tarefa demorada, mesmo para um citopatologista muito bem treinado (KOSS, 1989). Potencialmente, fadiga ou falta de concentração podem levar a uma interpretação falsa, sendo, dentre os erros, o mais significativo a ocorrência de falsos negativos<sup>2</sup> (DEMAY, 1997). De acordo com Barut *et al.* (2015), a sensibilidade<sup>3</sup> e especificidade<sup>4</sup> é de respectivamente ~57% e ~76% para o exame de Papanicolaou quando realizado puramente por inspeção manual.

Diante desse cenário, pesquisadores vêm desenvolvendo ao longo dos anos diversos

---

<sup>1</sup> aplicação de um teste ou exame numa população assintomática, aparentemente saudável, com objetivo de identificar lesões sugestivas de câncer e encaminhá-la para investigação e tratamento (INCA, 2019).

<sup>2</sup> testes que falham em identificar células anormais.

<sup>3</sup> é a probabilidade de um indivíduo avaliado e doente de ter seu teste alterado (positivo).

<sup>4</sup> é a probabilidade de um indivíduo avaliado e normal ter seu teste normal (negativo).

métodos computacionais com o objetivo de automatizar a análise de células cervicais (SU *et al.*, 2016; WILLIAM *et al.*, 2019; ALLEHAIBI *et al.*, 2019). A maioria desses modelos estão baseados na extração de características do núcleo e/ou do citoplasma das células, sendo, portanto, dependentes do desempenho de algoritmos de segmentação dessas estruturas. Como reportado em diversos trabalhos (KE *et al.*, 2019; CONCEIÇÃO *et al.*, 2019), a segmentação do núcleo e principalmente do citoplasma em imagens de células cervicais não é uma tarefa trivial, devido à quantidade de artefatos<sup>5</sup> comumente presente neste tipo de exame (MALM *et al.*, 2013). William *et al.* (2018) revisaram 30 artigos, publicados nos últimos 15 anos, com metodologias para automatizar a análise de células cervicais. A conclusão dos autores foi que embora alguns desses métodos alcancem acurácia superior a 93% em bases de dados de *benchmark*, a literatura carece de evidências de como esses algoritmos funcionariam se inseridos na rotina clínica, ou seja, processando imagens de células sob condições reais de ruído, sobreposição e artefatos. Além disso, muito desses trabalhos estão validados apenas em citologia de meio líquido, no qual a qualidade da imagem para fins de segmentação e classificação facilita a abordagem computacional (ZHU *et al.*, 2007). Infelizmente, devido aos custos, a citologia de meio líquido não é a realidade de programas públicos de prevenção desta neoplasia em muitos países em desenvolvimento, como o Brasil. Nessas regiões, onde ocorrem 90% das mortes por câncer do colo do útero (WHO, 2019), a técnica prioritária é o exame de Papanicolaou convencional (esfregaço).

Diante da dificuldade de segmentar adequadamente núcleo e citoplasma, pesquisadores têm proposto técnicas para categorização de lesões celular sem a necessidade de segmentação (ZHANG *et al.*, 2017; GV; REDDY, 2019). Embora a literatura careça de estudos de como esses métodos desempenham em imagens de Papanicolaou convencional, seus resultados são promissores para citologia de meio líquido. É nesta linha de pesquisa que está definida a hipótese principal desta tese: o uso da atenção visual de citopatologistas, reproduzidas por um modelo computacional, pode localizar as células cruciais para a classificação do campo, reduzindo a influência de artefatos sem interesse clínico. Essa ideia está fundamentada na capacidade humana de usar a atenção como seletor de dados de entrada, criando a habilidade necessária para interpretação de cenas complexas. De fato, trabalhos estado-da-arte têm explorado essas premissas para propor sistemas de visão computacional em diversos campos, como robótica,

---

<sup>5</sup> O termo “artefatos” é utilizado nesta tese para se referir aos elementos que podem estar presentes em imagens de células cervicais, mas que não contribuem diretamente para a detecção do câncer do colo do útero. Dentre esses elementos estão os derivados de falhas de confecção da lâmina, como, áreas escuras com alta sobreposição celular e estruturas comuns em esfregaços, como neutrófilos, células inflamatórias, hemácias etc.

desenvolvimento de carros autônomos e análises médicas (SCHLEMPER *et al.*, 2019). Os resultados alcançados por esses estudos, em geral, sugerem melhor desempenho quanto ao tempo de execução da aplicação e melhor acurácia, devido à eliminação de zonas que confundem a interpretação da imagem.

Neste contexto, esta tese introduz uma nova abordagem para a busca de lesões em exames de Papanicolaou, a qual pode ser a chave para instrumentar sistemas automáticos para triagem de exames em tempo real dentro da rotina clínica. No melhor de nosso conhecimento, prever a atenção de citopatologistas para ranquear regiões relevantes em imagens de células cervicais é algo pioneiro na literatura e pode fomentar discussões em áreas diversas, como no treinamento de profissionais e no controle de qualidade na rotina clínica de laboratórios citopatológicos. Além disso, esta tese aborda as imagens citológicas obtidas por exames de Papanicolaou convencional, que além de serem mais desafiadoras que as obtidas em meio líquido, correspondem à realidade do Brasil e da maioria dos países em desenvolvimento.

## 1.2 Objetivos da Tese

Esta tese tem como principal objetivo propor um modelo computacional baseado na predição da atenção visual seletiva de citopatologistas para auxiliar a triagem de imagens de exames de Papanicolaou com alteração celular. Durante o estudo e desenvolvimento deste trabalho, os seguintes objetivos específicos foram alcançados:

- **Objetivo 1:** Avaliar se algoritmos estado-da-arte para a predição de saliência<sup>6</sup> são apropriados para guiar a análise de imagens de células cervicais.
- **Objetivo 2:** Investigar se a predição da atenção visual de citopatologistas permite o ranqueamento e a seleção de partes da imagem que favorecem a identificação precisa de células anormais.
- **Objetivo 3:** Avaliar se células cervicais normais e anormais podem ser descritas a partir da diferença estrutural entre a textura do núcleo e sua versão degradada por uma função de transformação.
- **Objetivo 4:** Disponibilizar para a comunidade científica dados de rastreamento ocular de citopatologistas experientes em cuidados com o colo uterino durante a leitura e análise de exame de Papanicolaou convencional.

---

<sup>6</sup> Mapa de *saliência* é uma forma usual de representar a distribuição de atenção em uma imagem. Sua visualização é similar a um mapa de calor.

### 1.3 Hipóteses

As hipóteses que guiam a pesquisa conduzida nesta tese são:

- **Hipótese 1 (H1):** Métodos *bottom-up* (Seção 3.1.2.1) apresentam mapas de saliência comparáveis aos definidos pela atenção visual de especialistas humanos.
- **Hipótese 2 (H2):** Modelos de predição de saliência (Seção 3.1.2) podem detectar regiões clinicamente relevantes em imagens digitais de exames de Papanicolaou.
- **Hipótese 3 (H3):** O ranqueamento de regiões de interesse com base na predição da atenção visual de citopatologistas é uma estratégia computacional viável para a redução de erros de classificação de imagens de exames de Papanicolaou.

### 1.4 Contribuições

As principais contribuições desta tese são:

- **Contribuição 1:** Criação de uma nova base de *benchmark* com dados de rastreamento ocular de citopatologistas, contendo 5.654 células cervicais de 232 imagens extraídas de exames de Papanicolaou convencional do Sistema Único de Saúde (SUS) brasileiro. Disponível em <<http://dx.doi.org/10.17632/bk45c9yxb9.1>>.
- **Contribuição 2:** Um novo modelo guiado pela predição da atenção visual seletiva de citopatologistas para identificação de lesões celulares em imagens digitais de exames de Papanicolaou convencional. Além disso, o modelo proposto pode fornecer visualização das regiões clinicamente relevantes da imagem usando mapas de saliência.
- **Contribuição 3:** Um novo algoritmo para identificar células cervicais anormais com base na análise estrutural da textura do núcleo, da membrana nuclear, bem como de parte do seu entorno.
- **Contribuição 4:** Avaliação de algoritmos de predição de saliência estado-da-arte quanto à atenção visual seletiva de especialistas em imagens de células cervicais.

### 1.5 Produção Científica

O desenvolvimento desta tese resultou na publicação dos seguintes artigos científicos:

- Ramalho, G. L. B., **Ferreira, D. S.**, Rebouças Filho, P. P., Medeiros, F. N. S. Rotation-invariant feature extraction using a structural co-occurrence matrix. *Measurement*, v. 94, p. 406-415, 2016.



- **Ferreira, D. S.**, Ramalho, G. L. B., Medeiros, F. N. S., Bianchi, A. G., Carneiro, C. M., Ushizima, D. M. Saliency-driven system with deep learning for cell image classification. In: IEEE International Symposium on Biomedical Imaging (ISBI), 2019. p. 1284-1287.
- **Ferreira, D. S.**, Ramalho, G. L. B., Torres, D., Tobias, A. H., Rezende, M. T., Medeiros, F. N. S., Bianchi, A. G., Carneiro, C. M., Ushizima, D. M. Saliency-driven system models for cell analysis with deep learning. *Computer Methods and Programs in Biomedicine*, v. 182, p.105053, 2019.

### 1.5.1 *Produção Científica Complementar*

- Martins, G. L., **Ferreira, D. S.**, Medeiros, F. N. S., Ramalho, G. L. Ensemble of algorithms for multifocal cervical cell image segmentation. In: 9th Brazilian Conference on Intelligent Systems (BRACIS), 2020.

## 1.6 **Organização da Tese**

A estrutura desta tese está organizada como segue nos seguintes capítulos:

- **Capítulo 2:** apresenta uma visão geral do exame de Papanicolaou, e os motivos que o tornam desafiador para uma análise automatizada por técnicas de visão computacional. Discute as políticas de rastreamento do câncer cervical no Brasil e como a metodologia proposta nesta tese pode ser aplicada. Introduce o Centro de Reconhecimento para Inspeção de Células (CRIC), detalhando o projeto e a confecção da base de imagens de exames do SUS utilizada nesta tese.
- **Capítulo 3:** apresenta a teoria e as técnicas necessárias para compreensão desta tese. Trata de como mecanismos da atenção humana podem ser implementados computacionalmente. Apresenta o conceito de saliência e as abordagens usualmente empregadas pelos métodos de predição. Este capítulo também apresenta as Redes Neurais Convolucionais e discute o uso dessas estruturas na predição da atenção visual humana.
- **Capítulo 4:** descreve como os métodos de predição de saliência estão sendo abordados na área médica. Revisa algoritmos estado-da-arte que utilizam técnicas fundamentadas na atenção visual seletiva humana para a construção de sistemas de diagnóstico auxiliado por computador. Este capítulo também apresenta estudos de atenção visual e de seleção de regiões de interesse em imagens de células cervicais.

- **Capítulo 5:** propõe a base de dados CRIC *Visual Attention* (CRICVA). Discute a metodologia para a coleta de dados, detalhando o experimento de rastreamento ocular desenvolvido para gravação da atenção dos citopatologistas.
- **Capítulo 6:** propõe um modelo guiado por predição de saliência para detecção e identificação de células alteradas em imagens de exames de Papanicolaou. Este capítulo discute a metodologia para estimar a atenção visual dos citologistas, selecionar as regiões clinicamente relevantes e categorizar as imagens. Este capítulo também introduz a *Region-based Structural Co-occurrence Matrix* (rSCM), utilizada como parte do algoritmo de definição das ROIs.
- **Capítulo 7:** apresenta os resultados obtidos com os modelos desenvolvidos nesta tese. Cada seção descreve as métricas utilizadas, a seleção dos dados para a validação e a metodologia de avaliação empregada.
- **Capítulo 8:** apresenta as conclusões alcançadas e elenca nossas diretrizes futuras, as quais devem explorar novas investigações e escalar nosso método à prática clínica.

## 2 EXAME DE PAPANICOLAOU

Credita-se ao Dr. George Nicholas Papanicolaou a concepção e desenvolvimento do exame citopatológico do colo do útero, ou exame de Papanicolaou (CIBAS; DUCATMAN, 2003). Em 1928, Papanicolaou relatou que células cancerígenas derivadas do colo uterino podem ser observadas em esfregaços vaginais humanos (KOSS, 1989). No entanto, essa observação foi inicialmente mal recebida (KLINE, 1997), visto que o diagnóstico de câncer era somente feito por meio de biópsia e não havia ligação entre a citologia e a patologia cirúrgica (SILVA, 2004). Vários anos depois, Papanicolaou retomou o método, em colaboração com Herbert Traut, e despertou a atenção da comunidade médica com a publicação dos seguintes trabalhos: (PAPANICOLAOU; TRAUT, 1941) e (PAPANICOLAOU; TRAUT, 1943). Nesta época, a taxa de mortalidade anual por esse câncer nos Estados Unidos era de 26.000 mulheres, e a citologia, ao contrário da biópsia, poderia ser aplicada na escala necessária para revelar lesões precoces (SILVA, 2004). Vários estudos subsequentes na década de 1940, em pequenos grupos de mulheres, documentaram que esfregaços vaginais ou cervicais poderiam levar à descoberta de cânceres ocultos do colo uterino e a alterações pré-cancerosas (KOSS; MELAMED, 2006). Isso abriu o caminho para a introdução de programas de rastreamento de câncer de colo uterino em larga escala no final da década de 1950 (MEHNERT, 2003).

No Brasil, o exame de Papanicolaou é utilizado como principal método para rastreamento do câncer do colo do útero e está disponível na rede pública de saúde por meio do Sistema Único de Saúde (SUS). O procedimento convencional<sup>1</sup> para realização deste exame consiste nas seguintes etapas: 1) colher uma amostra de células da superfície externa e interna do colo do útero usando um cotonete, uma espátula de madeira ou uma escovinha; 2) estender parte dessa amostra em uma lâmina de vidro; 3) aplicar um fixador para preservar a amostra; 4) promover uma coloração da amostra para possibilitar a visualização dos diferentes componentes celulares; 5) inspecionar visualmente a lâmina sob um microscópio rastreando sinais de anormalidade. As etapas de coleta e fixação são geralmente realizadas por um médico que envia a amostra para um laboratório de citologia para coloração e interpretação. A Figura 1 ilustra esse processo em um campo de visualização, com lente objetiva de 40x, de uma lâmina de Papanicolaou convencional. Observa-se a disposição das células e de seus principais componentes, como citoplasma e núcleo.

<sup>1</sup> A nomenclatura “convencional” é utilizada para diferenciar o exame citopatológico de uma variação cuja as amostras coletadas são depositadas em um frasco com um líquido preservativo. Essa técnica conhecida como citologia em meio líquido (ZHU *et al.*, 2007) não é adotada em escala pelo SUS no Brasil e em vários outros países emergentes.

Durante a análise microscópica, o citopatologista busca por indicativos de anormalidade, como aspecto granular na distribuição de cromatina<sup>2</sup> do núcleo e irregularidades na membrana nuclear. Outras informações, como relação de tamanho entre núcleo e citoplasma, coloração, ocorrência de multinucleação também são observadas. Recomendamos a leitura de (BOON; SUURMEIJER, 1996) para melhor compreensão destes atributos, bem como dos procedimentos de coleta do exame de Papanicolaou.

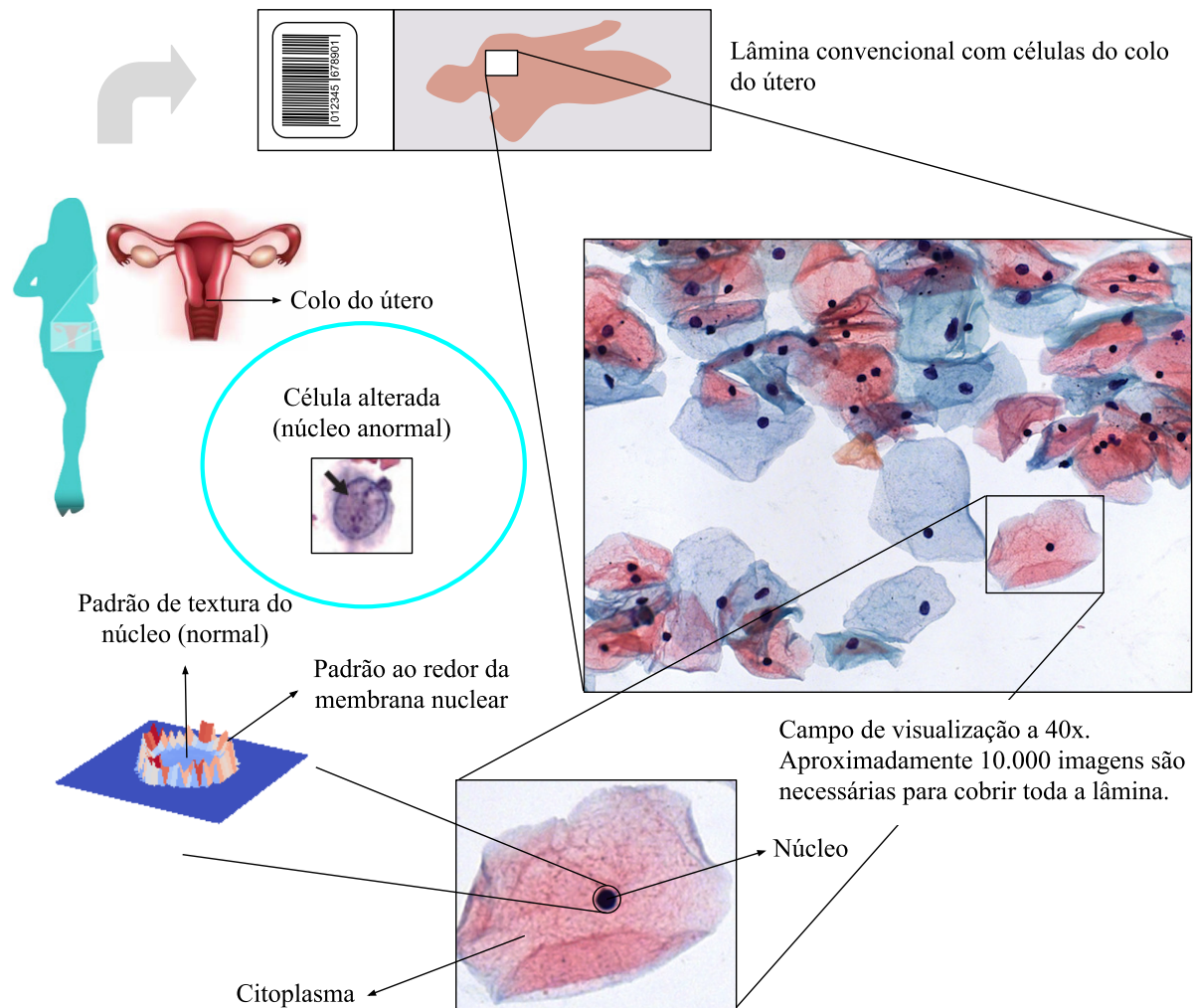


Figura 1 – Campo de visualização a 40x de uma lâmina de exame de Papanicolaou convencional. Nesta tese, esse campo é tratado como imagem de células cervicais ou imagem de exame de Papanicolaou. Célula alterada, com membrana cromatínica irregularmente espessa, obtida em (BARROS *et al.*, 2012). Fonte: elaboração própria.

<sup>2</sup> A cromatina consiste em um material celular composto por ácido desoxirribonucleico (DNA) e histonas de proteínas básicas (STENESH, 1989). Na coloração de Papanicolaou, o DNA fica manchado com hematoxilina, enquanto que algumas das proteínas nucleares que contêm carga negativa também podem ser manchadas (BOON; DRIJVER, 1986). Núcleos normais tendem a apresentar distribuição de cromatina uniforme, homogênea, sem aspecto granular, dando à textura uma aparência com baixa variação de intensidade dos *pixels* (WATANABE *et al.*, 2004). Em determinadas circunstâncias, processos degenerativos e inflamatórios podem tornar a cromatina mais granular, mesmo sem indicar lesão. Nestes casos, a análise de atributos como irregularidade da membrana nuclear, tamanho do núcleo, dentre outros, são cruciais para a decisão dos citopatologistas.

Lesões cancerígenas ou pré-cancerígenas detectadas por meio de exames de Papanicolaou podem ter seu laudo clínico emitido com base em diferentes sistemas de classificação (WHO; IARC, 2005). O padrão adotado no Brasil para laudos citopatológicos é Sistema Bethesda (TBS) (NAYAR; WILBUR, 2015b; NAYAR; WILBUR, 2015a). Como as pesquisas realizadas nesta tese visam fomentar o desenvolvimento de sistemas computacionais para auxílio ao diagnóstico de lesões, também adotamos uma nomenclatura simplificada comumente adotada na rotina clínica. A Tabela 1 apresenta as principais classes do TBS e mostra como elas podem ser agrupadas para a organização de resultados na rotina laboratorial.

Tabela 1 – Sistema Bethesda e sua, respectiva, macro categorização empregada na prática clínica. Fonte: elaboração própria a partir de (NAYAR; WILBUR, 2015a).

Sistema Bethesda		Rotina Clínica
Categoria	Descrição	
Negativa	Células normais, sem a presença de alterações	Negativa (normal)
ASC-US	Células escamosas atípicas de significado indeterminado	Baixo risco
LSIL	Lesão intraepitelial escamosa de baixo grau	
ASC-H	Células escamosas atípicas, não se pode descartar uma lesão de alto grau	Alto risco
HSIL	Lesão intraepitelial escamosa de alto grau	
CA	Carcinoma	

Estudos têm revelado que o exame de Papanicolaou tem sido de baixa sensibilidade (41% a 70%) no Brasil e grande parte dos laboratórios brasileiros apresentam indicadores de qualidade fora dos padrões recomendados, como Índice de Positividade (IP)<sup>3</sup> abaixo de 3,0% (PAULA *et al.*, 2017; TOBIAS *et al.*, 2016; BORTOLON *et al.*, 2012). A principal consequência dessa realidade é a tendência a laudos falso-negativos, induzindo mulheres não diagnosticadas a não procurar acompanhamento e consequentemente desenvolver câncer (TOBIAS *et al.*, 2016). Em países onde o rastreamento do câncer de colo do útero foi bem sucedido, como Estados Unidos, Noruega e Reino Unido, o percentual de positividade observado é de, em média, 6,8%, 4,9% e 6,5%, respectivamente (BORTOLON *et al.*, 2012). Como uma estratégia para minimizar esse problema, a Portaria Federal nº 3.388 de 30 de dezembro de 2013<sup>4</sup> redefiniu as estratégias para Qualificação Nacional em Citopatologia (QualiCito) na prevenção do câncer do colo do útero, promovendo as diretrizes para a revisão dos exames realizados pelos laboratórios públicos

<sup>3</sup> De acordo com Bortolon *et al.* (2012), o índice de positividade expressa a prevalência de alterações celulares nos exames e caracteriza a sensibilidade do processo do rastreamento em detectar lesões na população examinada. É computado pela razão entre o número *total de exames citopatológicos com resultados alterados realizados em determinado local e período* e o *total de exames citopatológicos satisfatórios realizados no mesmo local e período*.

<sup>4</sup> Disponível em: <[http://bvsm.s.saude.gov.br/bvs/saudelegis/gm/2013/prt3388\\_30\\_12\\_2013.html](http://bvsm.s.saude.gov.br/bvs/saudelegis/gm/2013/prt3388_30_12_2013.html)>

e privados credenciados pelo SUS<sup>5</sup>.

A QualiCito acompanha o desempenho dos laboratórios por meio do Programa de Controle de Qualidade, o qual abrange o Monitoramento Interno da Qualidade (MIQ) e o Monitoramento Externo da Qualidade (MEQ). Enquanto o MIQ avalia parâmetros de qualidade para detectar falhas e tomar medidas corretivas apenas no fluxo interno de trabalho do laboratório tipo I, o MEQ consiste na revisão dos esfregaços por laboratório diferente daquele que realizou a primeira leitura. São considerados laboratórios tipo I aqueles que realizam exames citopatológicos do colo do útero e tipo II os responsáveis por realizar estes exames no âmbito do MEQ<sup>6</sup>.

Ferramentas de análise automática que auxiliem os citopatologistas a detectar falso-negativos e eliminar falso-positivos em exames de Papanicolaou podem colaborar significativamente com o MEQ<sup>7</sup>. Laboratórios tipo II comumente possuem volumes elevados de esfregaços para revisar e, embora tenham melhorado os índices de qualidade deste tipo de exame no Brasil, ainda possuem seus resultados comprometidos pelas dificuldades que a inspeção visual clássica das células oferece. Recomendamos a leitura de (INCA, 2012) para compreensão detalhada dos programas de controle de qualidade em exames de Papanicolaou que estão em operação no Brasil.

## 2.1 Principais desafios para alcançar uma análise automatizada

Imagens digitais extraídas de exames de Papanicolaou possuem diversas estruturas que tornam a análise automatizada uma tarefa complexa. As amostras são comumente cobertas por artefatos, muitos deles originados durante a coleta, e por tipos de células sem interesse clínico para o diagnóstico do câncer do colo do útero (BENGTSSON; MALM, 2014). Listamos a seguir alguns dos desafios, comumente reportados na literatura, para a construção de sistemas computacionais baseados em visão computacional para auxílio ao diagnóstico. A Figura 2 ilustra cada tópico.

<sup>5</sup> Fatores socioeconômicos, culturais e/ou educacionais também são relevantes para determinar o sucesso no rastreamento do câncer do colo do útero. Regiões onde as campanhas públicas de prevenção são acessíveis e melhor compreendidas pela população tendem a apresentar melhor controle dessa neoplasia e a reduzir a mortalidade (ALMEIDA, 2018).

<sup>6</sup> Laboratórios tipo II também podem realizar as ações dos laboratórios tipo I.

<sup>7</sup> Conforme o fluxograma interno do laboratório tipo II, uma ferramenta de análise automática pode contribuir como um instrumento de ranqueamento das lâminas, dando preferência aos exames com maior probabilidade de serem positivas. Outra aplicação é funcionar como *segunda opinião* na classificação de lesões ou mesmo selecionando os campos alterados para compor o laudo clínico.

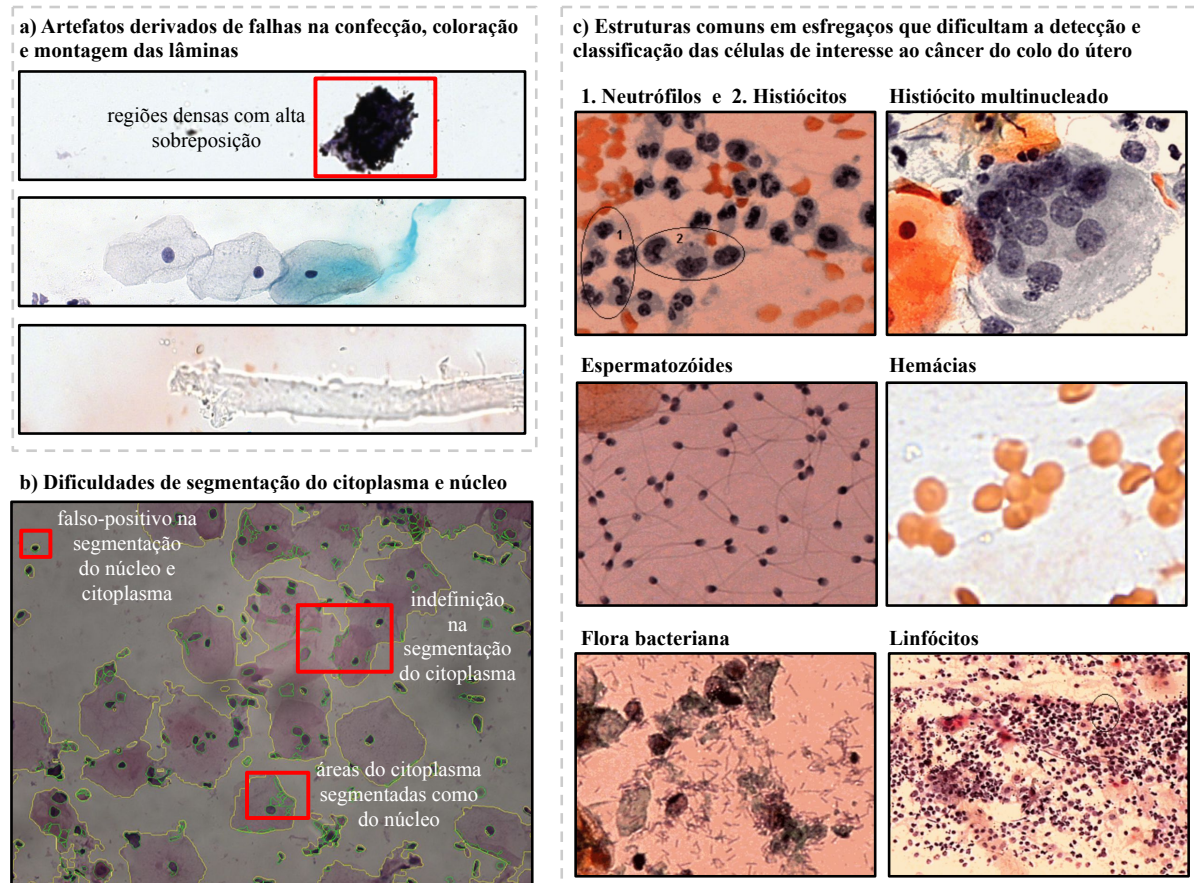


Figura 2 – Principais desafios para construção de sistemas de análise automática para exames de Papanicolaou. a) Artefatos presentes nas imagens que resultam de falhas na confecção, coloração e montagem das lâminas, respectivamente. b) Falhas na segmentação de núcleo e citoplasma, exemplificado pelos resultados do método proposto em (ZHANG *et al.*, 2014). c) Elementos comuns em esfregaços cérvico-vaginais, que não derivam de falhas na preparação das lâminas, mas que comumente reduzem o desempenho de métodos computacionais por serem confundidos com objetos de interesse clínico para o câncer do colo do útero. Para detalhes sobre essas estruturas recomendamos a leitura de (WHO; IARC, 2004). Fonte: elaboração própria a partir de imagens obtidas em <<http://www.pro-celula.com.br/home/atlasCitologico/atlas/texto/cit-cervice-elem.htm>>

- Preparação das amostras: o processo de coloração pode levar a variações na qualidade da amostra. Às vezes, o material celular pode ser distribuído de maneira desigual, levando a aglomerados densos que a luz não consegue penetrar enquanto outras partes da lâmina podem estar vazias. Mesmo quando o esfregaço é bem feito, ainda haverá regiões muito densas e com muitas células sobrepostas para uma interpretação confiável. Um citopatologista experiente consegue contornar essa dificuldade encontrando campos de visualização onde as células estão dispostas apropriadamente para o reconhecimento de lesões.
- Segmentação de células e núcleos: encontrar e delinear cada célula da imagem por

métodos computacionais está sujeito às mesmas complicações que dificultam a análise dos esfregaços por especialistas humanos. A variação na coloração da lâmina, na espessura dos aglomerados celulares e na intensidade de cor e textura dos núcleos e citoplasmas podem comprometer a extração dos atributos necessários ao reconhecimento das lesões. Além disso, é comum nos esfregaços a presença de estruturas que, embora facilmente identificadas por especialistas humanos, podem ser confundidas com os núcleos celulares, como, por exemplo, neutrófilos.

- Rejeição de artefatos: Existem diversas maneiras pelas quais células sanguíneas, células inflamatórias, células dobradas e distorcidas, objetos sobrepostos, muco, erros de coloração etc. influenciam na correta interpretação de uma imagem de uma célula cervical. Os classificadores projetados para diferenciar células normais e pré-cancerígenas produzem resultados imprevisíveis ao classificar esses artefatos. Um esfregaço de Papanicolaou padrão contém normalmente de 100.000 a 200.000 células relevantes, e idealmente uma única célula claramente maligna deve ser suficiente para diagnóstico positivo (BENGTSSON; MALM, 2014). Com a imprevisão da resposta do classificador ao processar esses artefatos, podemos ter um grande número de falsos-positivos, o que torna o sistema inútil. Especialistas humanos conseguem superar essa dificuldade, graças ao conhecimento prévio adquirido sobre essas estruturas e à capacidade do sistema visual humano em ser seletivo para os objetos essenciais de uma cena complexa (CARRASCO, 2011).

## **2.2 *Center for Recognition and Inspection of Cells (CRIC)***

O Centro de Reconhecimento para Inspeção de Células (CRIC) surgiu com o objetivo de pesquisar e desenvolver métodos computacionais para auxiliar na identificação de lesões precursoras do câncer de colo do útero. Proposto inicialmente por pesquisadores da Universidade Federal de Ouro Preto (UFOP), Universidade Federal do Ceará (UFC), Lawrence Berkeley National Laboratory (Berkeley, EUA) e Universidade da Califórnia (Berkeley, EUA), este projeto conta atualmente com a colaboração de cientistas de várias instituições brasileiras e suas contribuições alcançam resultados promissores também em outros tipos de imagens médicas.

Esta tese foi desenvolvida como parte dos trabalhos do CRIC. Com o suporte do SUS e de citopatologistas com experiência na análise de lâminas de Papanicolaou, o CRIC propôs uma base de imagens representativa para o estudo das lesões descritas na Tabela 1. Por ser concebido dentro da rotina clínica do SUS, o CRIC possui relevância no desenvolvimento



de pesquisas tanto para o cenário brasileiro quanto para o de países emergentes, onde a análise laboratorial e as políticas de prevenção do câncer uterino se assemelham às do Brasil. Além do conjunto de dados, detalhado na Seção 2.2.1, o CRIC conta na equipe com citopatologistas experientes na análise de lâminas de Papanicolaou e na saúde da mulher<sup>8</sup>, os quais foram cruciais para alcançar as contribuições desta tese.

### 2.2.1 Base de Imagens CRIC

A base de dados CRIC foi confeccionada a partir de esfregaços cérvico-vaginais disponíveis no setor de citologia do Laboratório Piloto de Análises Clínicas (LAPAC) da Escola de Farmácia da Universidade Federal de Ouro Preto<sup>9</sup>. O laboratório mantém um acervo de aproximadamente 30 mil lâminas oriundas de exames do SUS. A seleção de amostras deu preferência a lâminas mais recentes, seguindo critérios de qualidade, como predominância de células preservadas, bem fixadas e coradas, sem artefatos de montagem, além de ter suficiente representatividade para as lesões descritas na Tabela 1. A Figura 3 apresenta algumas imagens disponíveis na base CRIC e ilustra as anotações (retângulos em distintas cores) disponíveis para estudo das lesões. É importante observar que as características das imagens, como diferentes tonalidades de cores e presença de elementos diversos (Figura 2), dificultam a segmentação das células, mesmo dentro de uma mesma classe.

Os exames de Papanicolaou utilizados no desenvolvimento desta tese são os disponíveis na base do CRIC, versão v.01<sup>10</sup>. Este conjunto é composto por 400 imagens digitalizadas de Papanicolaou, as quais foram coletadas a partir de um microscópio Carl Zeiss equipado com uma câmera Zeiss AxioCam MRc com ampliação óptica de 40x. As imagens adquiridas têm densidade de  $0,255 \mu\text{m}/\text{pixel}$  e uma resolução de  $1392 \times 1040 \text{ pixels}$  (8-bit). As amostras foram preparadas via Papanicolaou convencional e contém células, bem como, artefatos frequentemente coletados como parte dos exames. As células cervicais estão rotuladas em normal (verde) e, quando anormal, usa a classificação descrita na Tabela 1. Também está disponível o rótulo do campo, a qual corresponde à classe da alteração celular mais significativa presente na imagem. Três citopatologistas experientes em análise de exames citopatológicos do SUS realizaram manualmente o procedimento de rotulação. No total, a base de dados reúne anotações para 11.473 células cervicais (#Negativa: 6792, ASC-US: 627, LSIL: 1329, ASC-H: 915, HSIL:

<sup>8</sup> Programa Âmbar. Informações em <<http://www.ambar.net.br>>.

<sup>9</sup> <<https://escoladefarmacia.ufop.br/lapac>>

<sup>10</sup> Nesta tese, simplificamos a nomenclatura para *base de imagens* CRIC, sem mencionar versão.

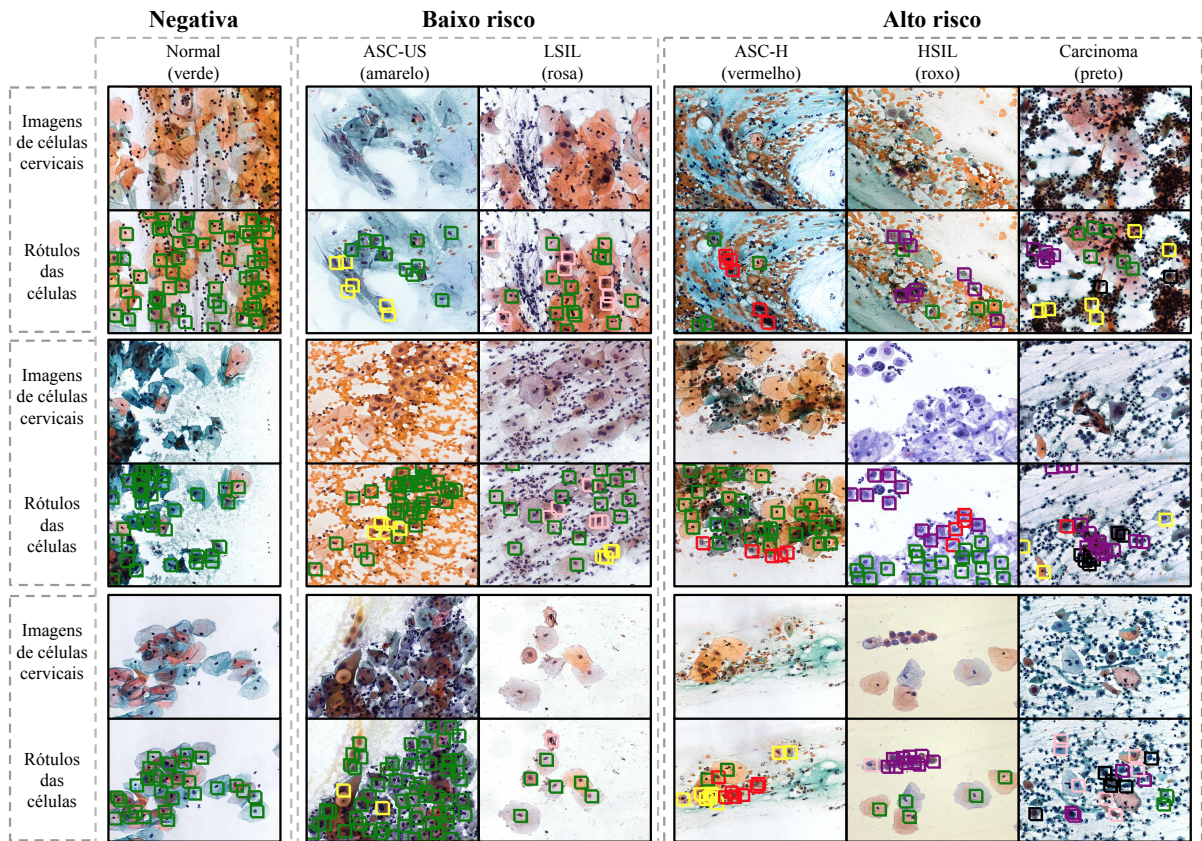
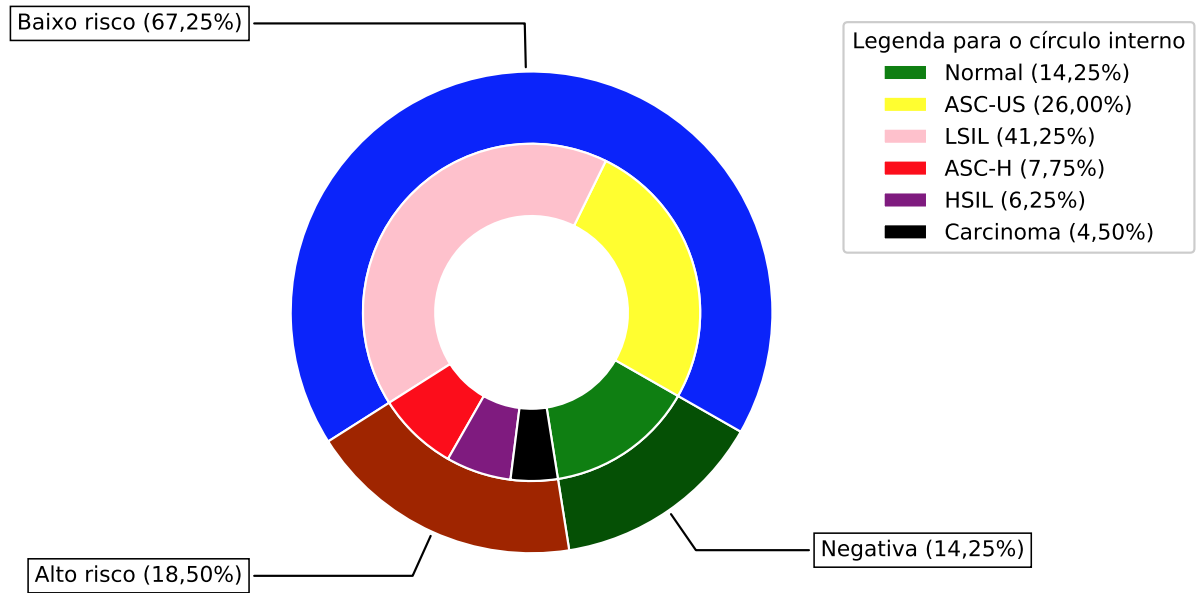
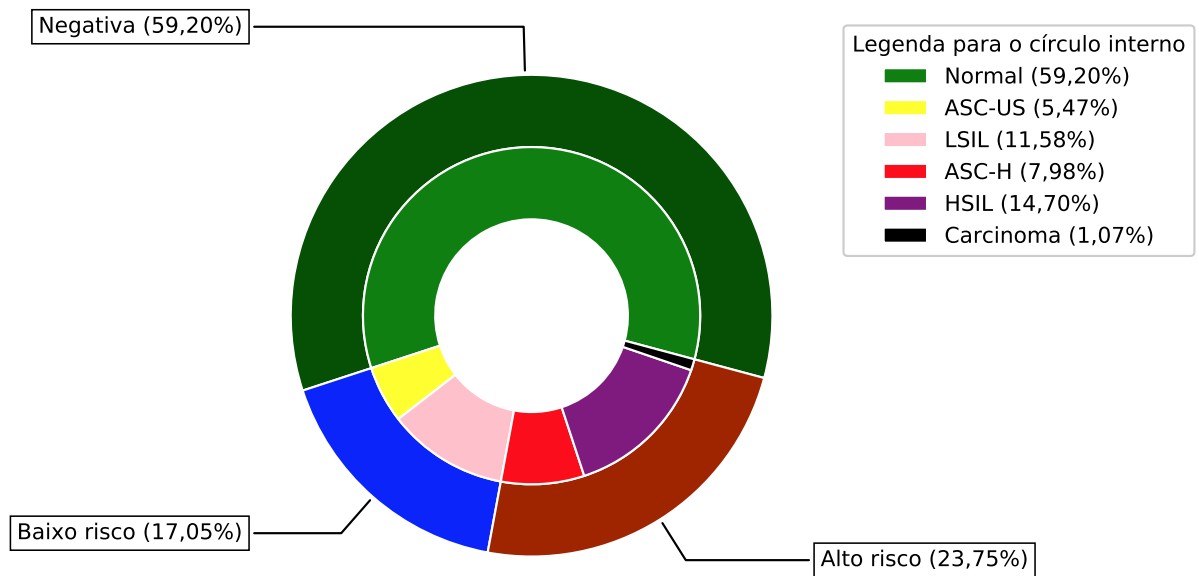


Figura 3 – Imagens de células cervicais a 40x disponíveis na base CRIC. Os rótulos (Normal, ASC-US, LSIL, ASC-H, HSIL e Carcinoma) sinalizam a classe do principal achado clínico na imagem, enquanto que as marcações em cores identificam esses achados, se existente, nas diferentes células. As linhas tracejadas simbolizam como as classes do Sistema Bethesda podem ser agrupadas dentro da rotina clínica. Os achados considerados são os reportados na Tabela 1. Fonte: elaboração própria.

1687, Carcinoma: 123}), as quais estão distribuídas conforme a Figura 4. A média de células por imagem é de 28,68, com desvio padrão de 21,83. As 400 imagens estão distribuídas em #{Negativa: 57, ASC-US: 104, LSIL: 165, ASC-H: 31, HSIL: 25 e Carcinoma: 18}, conforme a Tabela 1.



(a)



(b)

Figura 4 – Distribuição de casos na base CRIC. a) Casos de acordo com o rótulo da imagem, ou seja, classe do principal achado detectado no campo. Total de imagens (campos): 400. b) Distribuição de classes considerando todas as células. Total de células anotadas: 11.473. Fonte: elaboração própria.

### 3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a teoria e as técnicas aplicadas no desenvolvimento desta tese. A primeira seção trata de como mecanismos da atenção humana podem ser implementados computacionalmente. A seção seguinte descreve as Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network* (CNN)s), mostrando como essas redes geralmente utilizadas em problemas de classificação de objetos podem ser adaptadas para a predição da atenção visual seletiva humana.

#### 3.1 Atenção Visual Seletiva

A atenção visual é um dos principais mecanismos de percepção que permite aos seres humanos selecionar com eficiência os dados visuais de maior interesse (ZHANG; LIN, 2013). A atenção é um conceito fundamental para os humanos, mas pesquisadores vêm mostrando que esse conceito também pode beneficiar algoritmos computacionais (ZHANG; LIN, 2013). Neste cenário, vamos abordar dois objetivos da atenção. O primeiro objetivo é lidar com a complexidade da entrada perceptiva. Segundo Tsotsos (2011), a abordagem clássica para detectar objetos nas imagens é o paradigma da janela deslizante, no qual um classificador é treinado para detectar um objeto em uma região da imagem, sendo aplicado repetidamente em janelas de teste de tamanhos diferentes. Algoritmos que seguem esse paradigma de *força-bruta*<sup>1</sup> (busca-exaustiva) para detectar objetos arbitrários (com tamanhos arbitrários), em geral, levam ao consumo excessivo de recursos de processamento (STOCKMEYER; CHANDRA, 1988). Mesmo quando a busca na cena é guiada por uma tarefa, onde o objeto alvo e suas propriedades são conhecidas *a priori*, o desempenho do sistema pode ser comprometido devido ao fato de o objeto aparecer na cena muitas vezes distorcido, sobreposto, contaminado por ruído inerente ao processo de captura, dentre outros fatores dificultantes. Dessa forma, Frintrop (2011) afirma que um mecanismo para priorizar as regiões de uma imagem para processamento adicional é de grande interesse, especialmente se bancos de dados de imagens de alta resolução devem ser investigados ou se o processamento em tempo real for desejado. Isso coloca os modelos guiados por atenção como interessantes para aplicação em análise de exames médicos, onde comumente centenas de imagens precisam ser analisadas e estruturas patológicas identificadas.

O segundo objetivo da atenção é simplificar o cenário para tomada de ações (FRIN-

<sup>1</sup> A técnica de força bruta, conhecida também como busca exaustiva, é um método para resolver um problema através de uma travessia completa (ou parcial) no espaço de busca do problema para se obter uma solução.

TROP, 2011). Essa tarefa é especialmente importante para casos onde a máquina precisa atuar com cenas complexas e possivelmente desconhecidas. Segundo Frintrop (2011), mesmo se estiver disponível um poder computacional ilimitado, ainda assim existirão restrições: em um determinado momento, o sistema poderá ampliar uma ou poucas regiões para análise detalhada ou mesmo apresentar um ou poucos objetos. Na análise de imagens médicas, comumente sistemas computacionais de auxílio ao diagnóstico precisam enfatizar determinadas regiões do exame, desprezar artefatos e ampliar objetos específicos para melhor verificação posterior pelos especialistas. Assim, um mecanismo que seleciona as partes relevantes da cena, permitindo a execução de tarefas focadas a seguir é relevante para esses sistemas.

### 3.1.1 Modelagem Computacional

Embora os princípios biológicos que descrevem a atenção visual humana não sejam ainda completamente esclarecidos, pesquisadores têm, ao longo dos anos, se inspirado em conceitos do sistema visual humano para desenvolver métodos computacionais com o propósito de aplicá-los em tarefas de engenharia (TSOTSOS, 2011; ZHANG; LIN, 2013). A maioria desses métodos possui uma estrutura semelhante, apresentada na Figura 5. Esta estrutura foi idealizada a partir de teorias psicofísicas, como *Feature Integration Theory* (TREISMAN; GELADE, 1980) e o *Guided Search Model* (WOLFE *et al.*, 1989), e aparecem pela primeira vez na forma de algoritmo no trabalho de KOCH (1985). A ideia básica é calcular vários atributos em paralelo, verificar suas ativações e então agrupar essas informações em um mapa, no qual, a atratividade das regiões deve estar codificada. A esse mapa chamamos de *mapa de saliência*.

Dada uma imagem  $I \in D_I \subset \mathbb{Z}^3$  com dimensões  $m \times n$ , podemos computar um mapa de saliência  $M \in D_M \subset \mathbb{R}^2$  por meio de uma função  $f : D_I \rightarrow D_M$ , onde a função  $f$  representa o modelo computacional que está sendo utilizado. Em geral  $M$  possui a mesma dimensão  $m \times n$  de  $I$ , com os pontos de maior amplitude em  $M$  representando maior saliência. Dessa forma, temos que a intensidade do ponto  $M(i, j)$  reflete a conspicuidade de  $I(i, j)$  e, possivelmente, de parte de sua vizinhança. Regiões que atraem maior atenção possuem valores de saliência maiores, e o decréscimo dessa atenção segue proporcionalmente aos valores do mapa.

De forma geral, modelos de atenção computacional incluem os seguintes passos: 1) Extrair características de uma imagem de entrada em diferentes escalas. Comumente, os atributos coletados são de baixo nível (ou seja, *bottom-up*), como intensidade, cor e orientação, criando mapas de atributos. 2) Incluir um mecanismo que permita coletar informação de contraste

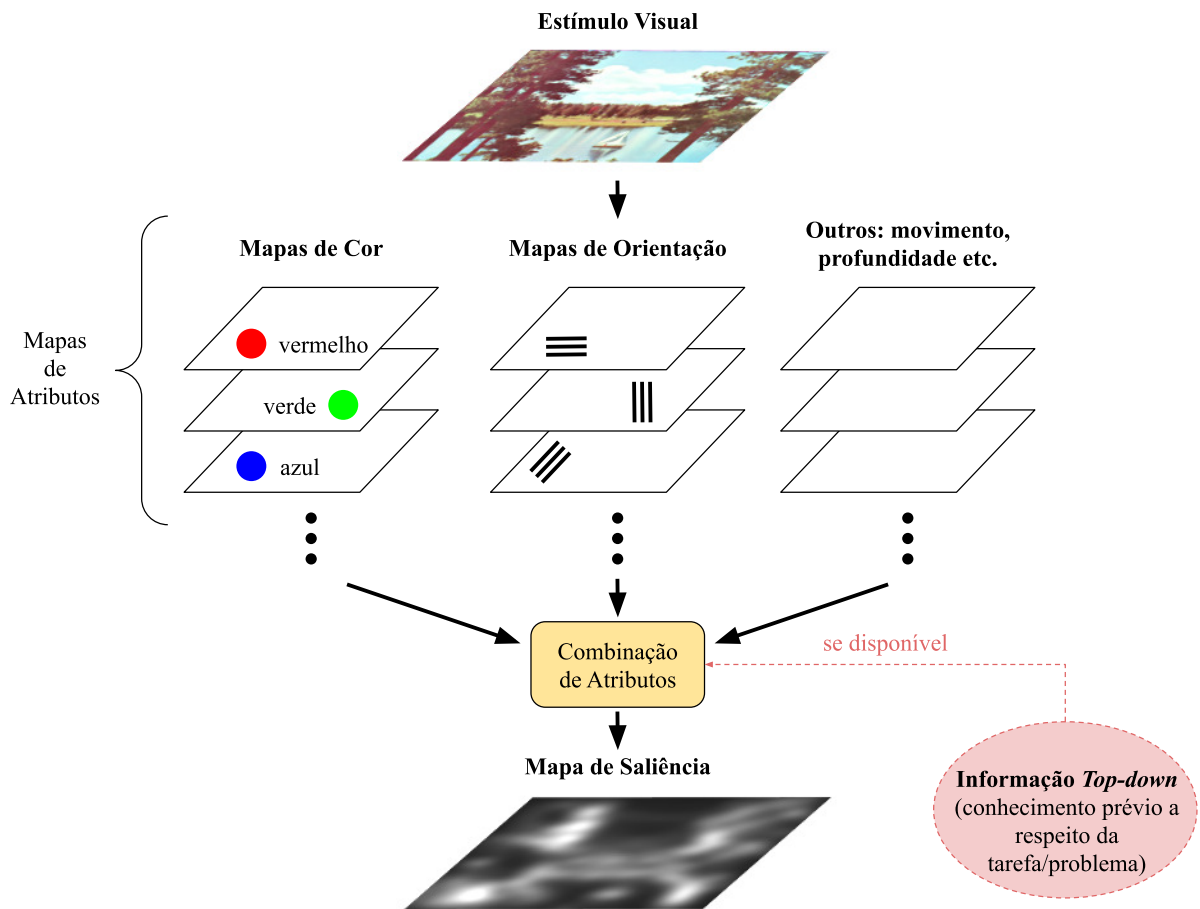


Figura 5 – Estrutura típica de um modelo computacional para estimar a atenção visual humana. A saída do sistema é um mapa de saliência, onde a intensidade dos elementos revelam as regiões com maior atratividade. Fonte: elaboração própria a partir de (ITTI; KOCH, 2001) e <<http://www.lizhaoping.org/zhaoping/prints/CosyneForPost.pdf>>.

dentro desses mapas. Uma abordagem clássica é computar o contraste de uma região (janela) em relação aos seus vizinhos, gerando mapas de características. 3) Combinar todos os mapas de características obtidos, e 4) normalizar, ponderar, combinando esses mapas para formar o mapa de saliência. Quando uma informação prévia sobre o problema está disponível, ou seja, características *top-down*, o mapa de saliência pode ser produzido dando um enfoque maior a uma ou outra região de acordo com o contexto da tarefa. As abordagens *bottom-up* e *top-down* serão detalhadas na Seção 3.1.2.1.

### 3.1.2 Predição de Saliência

A *predição de saliência* difere da *detecção de saliência* pelo fato do primeiro buscar estimar as fixações oculares de um observador, enquanto o segundo visa segmentar um objeto que se diferencia do contexto da cena. Assim, a predição de saliência leva a mapas que correspondem

às probabilidades de diferentes regiões serem conspícuas. Na detecção de saliências, mapas binários são produzidos para sinalizar a localização de um objeto de interesse. Ao longo dos anos, uma variedade de modelos computacionais foram desenvolvidos com ambos propósitos. Esta tese aborda a *predição de saliência*, aplicando o uso de seus conceitos à redução da quantidade de informação a ser processada por rotinas de categorização. A seção seguinte descreve alguns desses modelos. Devido à variedade de trabalhos publicados sobre o assunto na literatura, apenas aqueles com contribuições pioneiras são destacados. Um estudo mais amplo pode ser obtido por meio da leitura de (TSOTSOS, 2011; ZHANG; LIN, 2013; BORJI; ITTI, 2013; BORJI *et al.*, 2014; BORJI *et al.*, 2015; NGUYEN *et al.*, 2018; BORJI, 2019).

### 3.1.2.1 Abordagens Bottom-up e Top-down

Alguns trabalhos experimentais têm sugerido que a atenção visual humana é guiada tanto por estímulos, referido como fatores *bottom-up*, quanto pela tarefa que está sendo executada, fatores *top-down*. A atenção *bottom-up* está baseada em características primitivas da imagem de entrada como orientação, cor, intensidade, contraste, textura, dentre outras (ZHANG; LIN, 2013). Tais fatores podem ser modelados computacionalmente utilizando diversas abordagens, como estruturas probabilísticas, análise espectral ou técnicas clássicas de processamento de imagens visando à representação de processos cognitivos (BORJI; ITTI, 2013). Tipicamente, métodos *bottom-up* são modelados para refletir processos biológicos, revelando regiões da imagem que diferem da sua vizinhança sem lidar com os fenômenos cognitivos que tornam essas regiões relevantes.

De acordo com estudos psicofísicos, a atenção *bottom-up* ocorre no estágio de pré-atenção<sup>2</sup> e é o resultado da extração simples de atributos em todo campo visual. Portanto, uma região com estímulos visuais salientes pode capturar o foco da atenção humana (ZHANG; LIN, 2013). Como ilustrado na Figura 6, uma barra disposta na vertical rodeada por outras barras na horizontal, um círculo com cor diferente dos demais, um objeto diferente dos vizinhos pela forma geométrica e textura podem involuntariamente capturar nossa atenção. Dessa forma, a atenção *bottom-up* vem de um fator exógeno e independentemente de qualquer tarefa ou intenção. Estudos sugerem que esse tipo de atenção ocorre nas primeiras zonas de processamento do cérebro e, por agir em paralelo com as células responsáveis em obter os dados de entrada, o

<sup>2</sup> Segundo Neisser (1967) e Hoffman (1975), a atenção visual é dividida em dois estágios: pré-atenção e atenção. Na pré-atenção, atributos relativos ao fundo da cena e aos objetos da imagem são extraídos, mas apenas os objetos podem ser processados na fase de atenção (NEISSER, 1967; HOFFMAN, 1975; ZHANG; LIN, 2013).

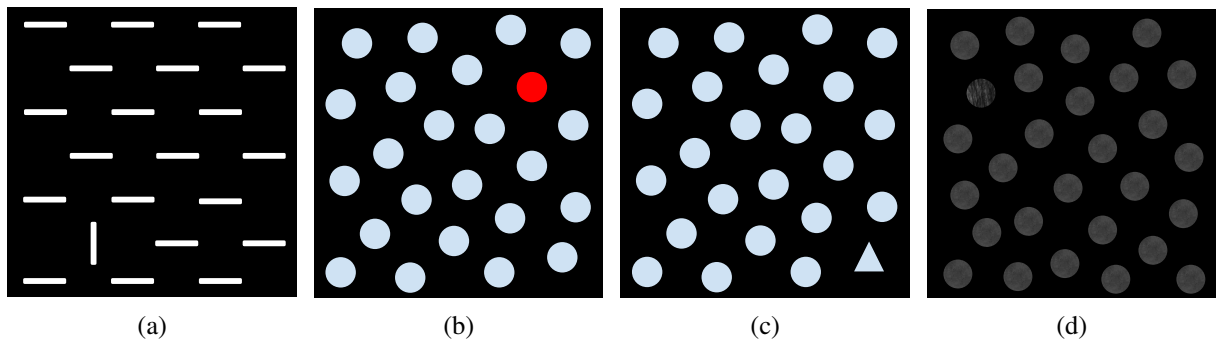


Figura 6 – Exemplos de padrões psicofísicos que podem ser utilizados para testar a reação de atenção de um observador. Em todos os casos, a modelagem computacional de apenas um atributo a) orientação, b) cor, c) forma ou d) textura pode estimar a atenção *bottom-up* humana. Fonte: elaboração própria a partir de (ZHANG; LIN, 2013).

tempo de resposta da atenção *bottom-up* é rápido, da ordem de 25 a 50 ms por item, excluindo o tempo de deslocamento ocular (TAKEDA; YAGI, 2000).

A atenção *top-down* refere-se à percepção visual enviesada pela tarefa, intenção ou emoção do observador. Ao contrário da atenção *bottom-up*, ele é voluntário e lento, demandando um tempo de aproximadamente 200 ms para um jovem adulto sem experiência no problema (TAKEDA; YAGI, 2000). Treinamento e acumulação de conhecimento pelo observador ajudam a reduzir o tempo de reação na atenção *top-down*. Portanto, este mecanismo é dependente de conhecimento prévio que o sujeito tem a respeito do problema. Embora vários modelos computacionais sejam inspirados apenas na atenção *bottom-up*, é bem conhecido que componentes *top-down* têm fonte influência na atenção humana (YARBUS, 1967). Henderson *et al.* (2007) forneceram evidências que a informação *top-down* domina as tarefas de busca no dia-a-dia. Portanto, para imitar corretamente o processo de seleção visual seletiva humana é necessário mesclar com sucesso as influências de fatores *bottom-up* e *top-down*.

Entender o contexto da cena é fundamental para acelerar a busca e o reconhecimento de um objeto (TORRALBA *et al.*, 2006). Por exemplo, se estamos procurando reconhecer uma pessoa, é mais apropriado focar em sua face e não em seus pés. Torralba *et al.* (2006) combinaram mapas de saliência *bottom-up* com informação contextual para orientar a busca de objetos em uma imagem. Áreas salientes que estavam dentro de uma região global selecionada receberam pesos mais altos do que aquelas, embora salientes, estavam fora da seleção. Os resultados alcançados superaram os modelos puramente *bottom-up* na previsão das fixações oculares humanas em uma tarefa de pesquisa. Esse achado confirma que explorar informações no contexto da aplicação pode produzir algoritmos de predição de saliência mais eficazes.



Similar ao modelo proposto por Torralba *et al.* (2006), o método *Saliency Using Natural statistics* (SUN)<sup>3</sup> (ZHANG *et al.*, 2008; KANAN *et al.*, 2009) também combinou informação *top-down* e *bottom-up* para estimar movimentos oculares em experimentos de busca em imagens do dia-a-dia. No entanto, SUN introduz uma modelagem dos alvos por informação *top-down*, superando os algoritmos para predição da atenção humana desenvolvidos até então. Dessa forma, os resultados alcançados por esses trabalhos têm sido uma comprovação computacional das evidências psicofísicas que seres humanos combinam de maneira semelhante informações de baixo e alto nível para orientar sua busca por alvos.

Uma outra forma de adicionar informação *top-down* a um modelo é modular o mapa de saliência obtido por uma matriz de pesos, especialmente projetada de acordo com a tarefa em questão. Essa ideia foi originalmente explorada por Wolfe *et al.* (1989). Por exemplo, se o objeto procurado fosse azul e estivesse na vertical, então, o modelo aumentaria o peso dos mapas de atributos de cor azul e orientação vertical, permitindo que esses componentes tivessem maior ativação. No mapa de saliência resultante, todas as regiões da imagem cujas características fossem semelhantes ao alvo se tornariam mais salientes e com maior chance de chamar atenção. Alguns trabalhos posteriores, como em (ELAZARY; ITTI, 2010) e (GAO *et al.*, 2008), utilizaram essa mesma abordagem, impulsionando essa ideia.

A terceira forma de inserir informação *top-down* nos algoritmos é incorporar o uso de detectores de objetos. Essa ideia foi aplicada nos trabalhos de Cerf *et al.* (2008), Cerf *et al.* (2008), Cerf *et al.* (2009), Cerf *et al.* (2009), confirmando a hipótese de que faces e textos atraem fortemente a atenção humana. Mais tarde, Judd (2011) mostrou que alguns aspectos, como a resolução da imagem, podem influenciar na conspicuidade das regiões, mesmo com faces e textos presentes.

Outro aspecto que devemos considerar é o objetivo do observador ao visualizar a cena. No estudo sobre atenção *top-down* proposto por Yarbus (1967), várias perguntas foram feitas aos participantes quando observavam a imagem de uma família em uma sala (Figura 7a). Os movimentos sacádicos dos observadores variaram para o caso de observação livre (sem perguntas) e quando alguma tarefa os foi dada (com perguntas). O foco da atenção também mudou dependendo de qual tarefa (pergunta) era feita. As Figuras 7b, 7c, 7d mostram os movimentos oculares dos participantes para o caso de nenhuma pergunta ser feita, no caso de serem questionados sobre a idade das pessoas da cena e quando eles são levados a lembrar da

<sup>3</sup> Código fonte disponível em <<https://github.com/MingMingCheng/SalBenchmark/tree/master/Code/matlab/SUN>>.



(a)



(b)



(c)



(d)

Figura 7 – Movimentos sacádicos e fixações oculares do experimento conduzido por Yarbus (1967). As gravações duraram 3 minutos para (b)-(d). a) Cena de pessoas de uma família em uma sala. b) Rastreamento ocular dos participantes em observação livre (sem qualquer pergunta ou tarefa). c) Movimentos oculares quando os participantes são questionados sobre a idade dos indivíduos da cena e d) quando são levados a lembrar da posição dos objetos e pessoas. Fonte: Imagens obtidas em (ZHANG; LIN, 2013).

posição dos objetos, respectivamente. No caso de serem indagados sobre a idade das pessoas, os focos de atenção aconteceram próximo as faces das pessoas. Já para lembrar da posição dos objetos, os movimentos oculares dos participantes ficaram mais distribuídos na imagem, com pontos de foco sobre os locais desses objetos.

### 3.2 Redes Neurais Convolucionais

Aprendizagem profunda (do inglês, *deep learning*) é uma subárea do campo de inteligência artificial que aperfeiçoa a classificação em grandes volumes de dados rotulados. Métodos baseados em aprendizado profundo estão sendo utilizados em uma variedade de

problemas, tais como: carros autônomos, visão robótica, processamento de linguagem natural, saúde preventiva, entre outras áreas (DARGAN *et al.*, 2019). Recentes *surveys* (SHIN *et al.*, 2016; SHEN *et al.*, 2017; LITJENS *et al.*, 2017; ANWAR *et al.*, 2018; HAFIZ; BHAT, 2020) têm demonstrado que técnicas de aprendizado profundo têm alcançado resultados promissores quando aplicadas em problemas de segmentação (DOLZ *et al.*, 2020), classificação (ALYAFEAI; GHOUTI, 2020) e detecção de lesões (RIBLI *et al.*, 2018; WESSELS; HAAR, 2019) em imagens médicas, apesar do custo computacional.

Um dos modelos mais populares de aprendizagem profunda é a CNN (DARGAN *et al.*, 2019). Essa rede corresponde a uma variação do *Perceptron* Multicamadas (MLP) e pode ser descrita, de acordo com Ghariba *et al.* (2019), como uma composição de várias funções

$$f(x) = f_L(\dots f_2(f_1(x; w_1); w_2)\dots); w_L), \quad (3.1)$$

em que cada função  $f_L$  obtém um dado  $x_L$  e um vetor de parâmetros  $w_L$  como entrada e produz outro dado  $x_{L+1}$  como saída. Os parâmetros  $w = (w_1, \dots, w_L)$  correspondem aos pesos da rede, os quais podem ser treinados a partir dos dados de entrada para resolver um problema específico, como, por exemplo, classificação. Para que a rede possa ser generalizada para problemas complexos (e não apenas lineares), existe uma função de ativação não linear que é associada às camadas convolucionais da rede. Uma CNN que não utiliza função de ativação torna-se um modelo linear e nesta categoria de problemas, existem soluções de menor custo computacional. Em arquiteturas atuais, a função *Rectified Linear Unit* (ReLU) (Figura 8a) (HAHNLOSER *et al.*, 2000; NAIR; HINTON, 2010) é amplamente utilizada. Quando a entrada da função ReLU é negativa, seu valor de retorno é zero e o neurônio não é ativado. Isso habilita a ativação apenas de alguns neurônios ao mesmo tempo, tornando a representação da rede esparsa, o que permite acelerar o aprendizado e simplificar o modelo.

Além das camadas convolucionais, as arquiteturas CNN também contêm camadas de agrupamentos, conhecidas como *pooling*. Essas camadas são geralmente usadas imediatamente após camadas convolucionais, simplificando as informações dos mapas de características<sup>4</sup>. O propósito desta operação é reduzir a dimensão do volume de dados de entrada, por meio de um método de sub-amostragem, e preservar a informação discriminante. Existem vários tipos de operador que podem ser utilizados, tais como *max-pooling*, *average-pooling* e *sum-pooling*. A

<sup>4</sup> Capturam o resultado da aplicação dos filtros à entrada, como a imagem de entrada fornecida à rede ou outro mapa de características.

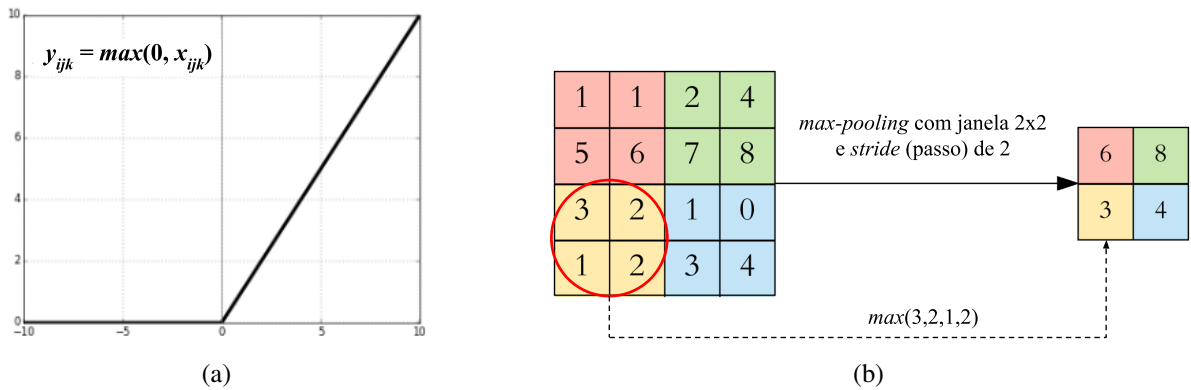


Figura 8 – Funções básicas que compõem a arquitetura de uma CNN. a) Função de não linearidade (ReLU) e b) função *max-pooling* para redução de dimensionalidade espacial dos mapas de características. Fonte: adaptado de <<https://cs231n.github.io/convolutional-networks/#pool>>

Figura 8b ilustra a saída de um operador *max-pooling* com janela de dimensão 2 x 2 e passo (*stride*) de 2.

Várias arquiteturas CNN foram inicialmente projetadas para problemas de reconhecimento e classificação. Dessa forma, elas são compostas basicamente por duas partes: camadas convolucionais, dedicadas à extração de características, e camadas *Perceptron* totalmente conectadas, para a tarefa de classificação. A Figura 9 ilustra a organização típica dessas redes, destacando suas principais partes componentes.

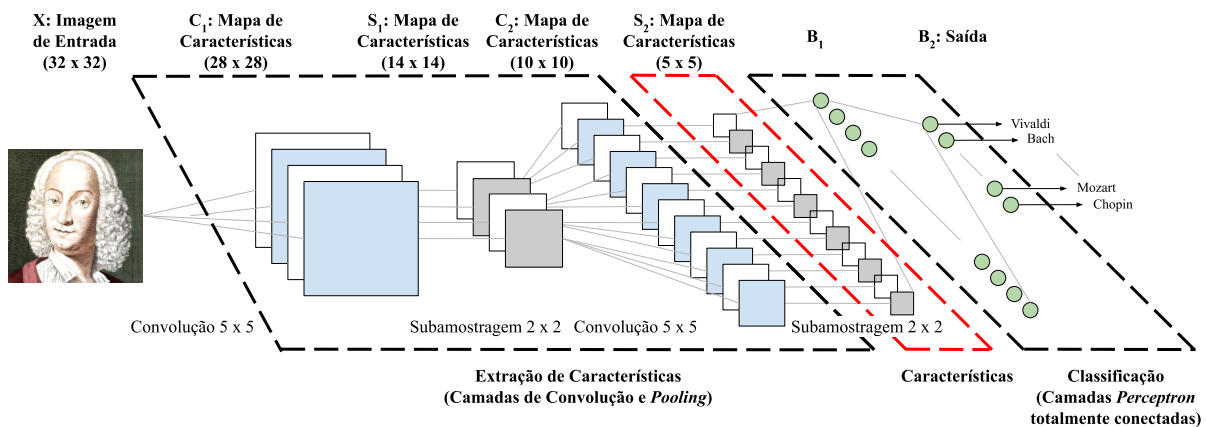


Figura 9 – Estrutura típica de uma arquitetura CNN. No bloco em vermelho estão as características extraídas pelas camadas convolucionais. Fonte: adaptado de (LI *et al.*, 2017).

O treinamento das CNNs segue a mesma metodologia empregada pelas Redes Neurais Artificiais (RNA). Em geral, utiliza-se o algoritmo *backpropagation* (algoritmo de retropropagação do erro) (RUMELHART *et al.*, 1986; GOODFELLOW *et al.*, 2016), em conjunção com uma função custo, para executar essa tarefa. Como essas redes estão sujeitas a

*overfitting* (sobreajuste)<sup>5</sup> durante o treinamento, é necessário utilizar métodos de regularização para amenizar esse efeito. Segundo Goodfellow *et al.* (2016), a regularização corresponde a modificações feitas em um algoritmo de aprendizado com a intenção de reduzir seu erro de generalização<sup>6</sup>, mas não seu erro de treinamento. Existem muitas estratégias de regularização. O *dropout* é uma técnica bastante empregada. Este método consiste em desativar, a cada iteração do treinamento, alguns neurônios da rede seguindo uma determinada probabilidade *pb*. Os neurônios desativados são reamostrados na próxima iteração, de modo que um neurônio desativado em uma iteração pode estar ativo na seguinte. O hiperparâmetro *pb* é conhecido como *taxa de dropout* e geralmente é um número em torno de 0,5, correspondendo a 50% dos neurônios que estão sendo eliminados. Para uma revisão mais detalhada desse mecanismo, consulte (HINTON *et al.*, 2012).

Uma forma de prevenir o *overfitting* é treinar a rede com uma imensa quantidade de amostras (GOODFELLOW *et al.*, 2016). No entanto, essa situação está raramente disponível. Por exemplo, obter imagens médicas para diagnóstico de câncer, rotuladas por vários especialistas, é consideravelmente caro. Uma solução que tem se apresentado eficaz para contornar esse problema, particularmente no reconhecimento de objetos, é a criação de novos dados com base nos existentes. Essa operação é conhecida como *data augmentation* e pode ser alcançada por meio de transformações, como translações, rotações, por contaminação por ruídos ou por estimativas com base na estatística dos dados existentes. Outras técnicas de regularização, que trabalham ajustando condições para o treinamento, também podem ser consideradas. Um exemplo é a parada precoce (*early stopping*), na qual o modelo é treinado enquanto existem ganhos de desempenho no conjunto de validação. Se, durante uma quantidade determinada de épocas<sup>7</sup>, os ganhos não estão mais sendo observados, o modelo termina o treinamento. Assim, não é necessário especificar explicitamente a quantidade de épocas de treinamento, a qual quando mal regulada pode levar a problemas de *overfitting* ou *underfitting*, esse último em situações onde o modelo não encontra relações nos dados de treino.

Predição de saliência utilizando CNN tem sido uma linha de pesquisa bastante investigada nos últimos anos (BYLINSKII *et al.*, 2015). Como a atratividade das regiões de

<sup>5</sup> O *overfitting* é um termo utilizado na estatística para se referir a um erro de modelagem que torna o modelo muito bem adaptado a um conjunto de dados previamente observado, mas ineficaz para novas amostras. O problema surge quando o modelo se ajusta a características relacionadas ao ruído ou variância entre os dados, ao invés da distribuição da qual os dados foram extraídos (WEBB, 2010).

<sup>6</sup> Capacidade do modelo em ser preciso para novas amostras, ou seja, dados que não foram vistos durante o treinamento.

<sup>7</sup> Uma época corresponde a etapa do treinamento em que todas as amostras de treino tiveram a oportunidade de atualizar os parâmetros do modelo. Uma época é composta por um ou mais lotes (*batches*), isto é, grupos de amostras que serão apresentadas à rede.

uma imagem geralmente é dependente de características de baixo e alto nível, o princípio de operação dessas redes revelou-se interessante. Informações de baixo nível, como bordas, cantos e orientações, são capturadas por campos receptivos pequenos, que correspondem às primeiras camadas da rede. Enquanto que características de alto nível, como informações semânticas (por exemplo, partes de objetos ou faces) são extraídos por campos receptivos grandes, que correspondem às últimas camadas da rede. A partir de campos receptivos de diferentes tamanhos, podemos agrupar os resultados dessas múltiplas convoluções, selecionando uma ou várias camadas da rede, e gerar um mapa de saliência combinando informação *bottom-up* e *top-down*.

O modelo proposto por Wang e Shen (2017) ilustra a geração de um mapa de saliência a partir das camadas convolucionais de uma CNN. Os autores propuseram um método baseado na arquitetura codificador-decodificador que permite combinar características *bottom-up* e *top-down* a partir da fusão de informações oriundas de camadas com vários tamanhos de campos receptivos. O codificador corresponde a pilha de camadas convolucionais da CNN, a qual captura informação de alto-nível (contexto) por meio de convoluções e subamostragens dos mapas de características da primeira para última camada convolucional. Portanto, ao longo da rede, esses mapas decrescem de tamanho, devido à operação de *pooling*, conforme ilustrado na Figura 9. O decodificador sobreamostra os mapas de características e constrói uma saída de mesma resolução da imagem de entrada. Assim, dado uma imagem  $I$ , assumindo que seu mapa de características na entrada da  $l$ -ésima camada é  $F^{l-1}$ , então o mapa de característica de saída  $F^l$ , definidos pelos filtros de convolução com pesos  $W^l$ , é obtido por:

$$F^l = f_{con} \left( F^{l-1}; W_{conv}^l \right) = W_{conv}^l * F^{l-1}, l = 1 \dots L, \quad (3.2)$$

onde o símbolo  $*$  indica a operação de convolução e  $L$  corresponde ao número de camadas. É importante notar que  $F^0$  é a imagem  $I$ , a qual em muitas aplicações está disposta como uma pilha de imagens referentes a cada canal de cor RGB. A operação de desconvolução é oposta a operação de convolução e é responsável por realizar o processo de sobreamostragem dos mapas de características. A equação seguinte representa este processo.

$$f_{decon} (F; W_{decon}) = W_{decon} \oslash_s F, \quad (3.3)$$

onde  $\oslash_s$  é o passo (*stride*) de convolução e  $s$  é o fator de sobreamostragem. Para restaurar um mapa de características  $F^l$  subamostrado da  $l$ -ésima camada para o mesmo tamanho da imagem de entrada, um decodificador com múltiplas camadas de desconvolução pode ser implementado. No final, a pilha de mapas de características obtidos, com mesma resolução da imagem, pode ser

fundida em um mapa de saliência. Uma outra solução, é gerar o mapa de saliência sem uso do decodificador, e ao final, levar o mapa para a mesma resolução da imagem por meio de técnicas de interpolação. Essa solução foi implementada por Cornia *et al.* (2018).

Conforme discutimos, quando aplicamos CNNs para a predição de saliência, uma abordagem é desprezar as camadas *Perceptron* totalmente conectadas (Figura 9) e combinar os mapas de características extraídos das camadas convolucionais. A função custo é modelada para computar a distância entre o mapa de saliência predito e um mapa de saliência de referência (*Ground Truth* (GT)) geralmente coletado por um equipamento de rastreamento ocular. Huang *et al.* (2015) foram pioneiros em utilizar métricas de avaliação de mapas de saliência como função custo durante o treinamento. Algumas dessas métricas são revisadas e discutidas por Bylinskii *et al.* (2018) e podem ser utilizadas isoladamente ou combinadas de acordo com a modelagem do problema.

## 4 REVISÃO DA LITERATURA

Este capítulo descreve como os métodos de predição de saliência estão sendo empregados na área médica. A primeira seção revisa os algoritmos estado-da-arte que utilizam técnicas fundamentadas na atenção visual seletiva humana para a construção de sistemas computacionais de auxílio ao diagnóstico. Abordamos também estudos sobre a influência da atenção *top-down* durante a leitura de exames médicos por humanos. A seção seguinte foca na predição da atenção visual em imagens de células cervicais. Daremos enfoque aos métodos computacionais de análise que são guiados pela extração de regiões de interesse e que correspondem ao estado-da-arte.

### 4.1 Predição de Saliência em Imagens Médicas

A maioria dos modelos de predição de saliência foram projetados para estimar as fixações humanas em imagens naturais<sup>1</sup> (BORJI; ITTI, 2013; BORJI *et al.*, 2015). A publicação de várias bases de dados (WINKLER; SUBRAMANIAN, 2013), estudos de *benchmark* (BORJI; ITTI, 2013; BORJI *et al.*, 2015) e desafios públicos (JIANG *et al.*, 2015; BYLINSKII *et al.*, 2015) com os problemas em aberto na literatura impulsionaram a investigação desses modelos para esta categoria de imagens. Diante de promissores resultados, recentemente existe um interesse crescente no estudo dos movimentos oculares de especialistas durante a análise de imagens médicas. Dentre as principais aplicações estão o uso da atenção visual seletiva no projeto de sistemas de diagnóstico auxiliado por computador (do inglês, *Computer Aided Diagnosis* (CAD)) e a melhor compreensão dos padrões de busca dos especialistas. Assim esses estudos podem contribuir também para revelar causas de erros de diagnóstico e melhorar a capacitação da equipe de análise.

O rastreamento ocular, geralmente realizado por um equipamento chamado de *eye-tracker*, é uma importante instrumento para estudos de atenção. A origem desse equipamento remonta a 1879 quando Louis Emile-Javal, oftalmologista francês, notou, com base em observações a olho nu, que os olhos dos leitores fazem movimentos rápidos (como sacadas) misturados com pequenas pausas (como fixações) enquanto lê. O primeiro rastreador ocular era um dispositivo invasivo, construído em 1908 por Edmund Huey. Mas nas décadas de 1970 e 1980, os rastreadores oculares baseados em vídeos foram inventados, permitindo uma leitura não invasiva e mais precisa dos movimentos oculares. Desde então, esse equipamento é usado em uma gama

---

<sup>1</sup> Nesta tese, consideramos imagens naturais como imagens que retratam cenas do dia-a-dia.



de aplicações, incluindo psicologia cognitiva, pesquisa de *marketing*, engenharia de usabilidade, interação homem-computador e qualidade e análise de imagens médicas.

Rastreadores oculares permitem a construção de mapas de saliência de referência mostrando as regiões da imagem que mais demandam atenção do especialista durante a análise. A partir desses mapas é possível desenvolver modelos computacionais e avaliar outros existentes, mesmo que primariamente desenvolvidos para imagens naturais. Matsumoto *et al.* (2011) utilizaram um modelo de predição de saliência *bottom-up*<sup>2</sup> para estimar onde neurologistas olhavam quando visualizavam imagens de tomografia computadorizada do cérebro com acidentes vasculares. Tanto os neurologistas como o grupo de controle, formado por profissionais de saúde com outras especialidades, tendiam a olhar para as áreas de alta saliência estimada. No entanto, os neurologistas focaram frequentemente em regiões discretas, mas clinicamente importantes, que estavam fora dos mapas de saliência preditos. Matsumoto *et al.* (2011) concluíram, portanto, que ambos neurologistas e o grupo de controle usaram a atenção *bottom-up*, embora os neurologistas tenham feito uso mais eficiente da instrução *top-down*.

Jampani *et al.* (2012) avaliaram os algoritmos de predição de saliência *bottom-up* ITTI (ITTI; KOCH, 2001), GBVS (HAREL *et al.*, 2007) e SR (HOU; ZHANG, 2007)<sup>3</sup> em imagens de retina com exsudatos duros, e de raio-X do tórax com lesões de pneumoconiose. Em seu estudo, Jampani *et al.* (2012) observaram que o GBVS apresentou melhor desempenho nas radiografias do tórax, enquanto o SR teve os melhores resultados para as imagens de retina. Os autores propuseram também uma abordagem para inserir informação *top-down* na predição, mas os resultados não alcançaram ganhos significativos ( $\leq 10\%$ ). Dessa forma, os autores concluíram que os fatores *bottom-up* desempenham função importante na predição da atenção nos problemas considerados.

Alzubaidi *et al.* (2010) avaliaram 13 tipos de características, previamente já abordados em métodos de saliência, para responder quais delas poderiam prever com maior precisão as regiões que chamam a atenção de radiologistas experientes na leitura de imagens de radiografia do tórax. Os resultados revelaram que quatro dessas características (*Edge Orientation Histograms*, *wavelet Haar*, filtros Gabor e filtros direcionais) oferecem os melhores resultados. Alzubaidi *et al.* (2010) concluíram que existem diferenças significativas nos mapas de saliência estimados de acordo com os atributos escolhidos, mesmo que eles já tenham sido previamente

<sup>2</sup> O modelo utilizado foi o GBVS, proposto por Harel *et al.* (2007))

<sup>3</sup> Os métodos ITTI, GBVS e SR (*Spectral Residual Approach*) são descritos na Tabela 2. Nesta tese, por simplificação, a nomenclatura ITTI e GBVS foram reduzidas para IT e GB, respectivamente.

validados em outras pesquisas.

Recentemente, Wen *et al.* (2017) compararam o desempenho de 16 algoritmos estado-da-arte em quatro modalidades de imagens médicas: tomografia computadorizada (CT) de tórax, imagens radiográficas do tórax (CXR), tomografia por emissão de pósitrons (PET) de corpo inteiro e mamografias. Seus resultados mostraram que os modelos de saliência avaliados podem prever regiões clinicamente relevantes, mas que o ranque de desempenho desses algoritmos difere do reportado quando as imagens em processamento são naturais. Além disso, o desempenho desses métodos varia de acordo com o tipo de imagem médica que está sendo analisada. Os autores concluíram também que algoritmos projetados para o contexto de imagens naturais podem não funcionar em imagens médicas, necessitando incorporar informação *top-down* do problema. Wen *et al.* (2017) optaram por fornecer informação de localização da região a ser analisada, segmentando uma área específica da imagem. Os autores observaram que métodos que capturavam informação local e global foram mais precisos em prever as fixações oculares dos radiologistas.

Os avanços em técnicas de aprendizado profundo e a crescente disponibilidade de bases de imagens rotuladas, inclusive na área médica, permitiram o desenvolvimento de novos modelos de predição de saliência, mesclando conceitos da atenção *bottom-up* e *top-down*. Esses modelos geralmente estão baseados em Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network* (CNN)s) (Seção 3.2) e podem ser treinados fim-a-fim, melhorando consideravelmente os resultados. Guan *et al.* (2018) apresentou um estudo inovador para categorização de doenças pulmonares e cardíacas utilizando radiografias do tórax. Inspirados no mecanismo de atenção seletiva, os autores exploraram diferentes arquiteturas CNN estado-da-arte e modelaram um sistema que detecta lesões a partir da fusão de informações extraídas de toda imagem (contexto global) e específicas para uma região de interesse (contexto local). Primeiramente, o modelo utiliza uma CNN treinada para localizar a região da imagem clinicamente relevante. Posteriormente, uma segunda CNN captura informações detalhadas dessa área. As características extraídas dentro desses dois contextos são então agrupadas, formando um único vetor para categorização. Os autores concluíram que a inserção de informação local referente a parte clínica de interesse permitiu ao método alcançar melhor descrição das imagens, superando o estado-da-arte. No entanto, na abordagem utilizada por Guan *et al.* (2018), as imagens possuíam apenas uma região relevante e nenhum resultado foi reportado sobre o desempenho do algoritmo caso informações locais de regiões de falso positivo fossem inseridas no sistema.

Em um trabalho similar ao (GUAN *et al.*, 2018), Li *et al.* (2019) propuseram um sistema em dois estágios baseado em atenção para classificação e localização de lesões em exames histológicos da próstata. O modelo seleciona partes clinicamente relevantes de lâminas de biópsia da próstata e dedica rotinas de classificação apenas às regiões selecionadas. Este modelo calcula mapas de saliência em baixa resolução (amplificação de 5x) utilizando uma CNN, realiza um ranqueamento das regiões de acordo com os valores desse mapa e seleciona as regiões salientes que estão nas primeiras posições. Logo após essas regiões são agrupadas, ampliadas para 10x e submetidas a uma outra CNN para categorização da lâmina em três classes: normal, baixo e alto risco de câncer. Os autores reportam ganhos superiores a 5% em relação ao estado-da-arte para classificação de lâmina de biópsia da próstata. No entanto, o trabalho carece de uma avaliação quantitativa do primeiro estágio, não sendo possível portanto definir qual o desempenho do sistema na seleção das regiões de interesse. De qualquer forma, os trabalhos de Guan *et al.* (2018) e Li *et al.* (2019) representam avanços no desenvolvimento de sistemas CAD. Ao superar o estado-da-arte, comumente baseado na seleção supervisionada de Regiões de Interesse (ROIs), esses autores mostram que incorporar mecanismos de seleção visual seletiva a esquemas computacionais de auxílio ao diagnóstico é uma metodologia promissora tanto para aumentar a precisão na detecção de lesões como para aproximar o tempo de processamento do sistema ao tempo-real da rotina clínica.

#### **4.2 Categorização de Exames de Papanicolaou com Base em Regiões de Interesse**

Os estudos que envolvem atenção visual e análise de exames de Papanicolaou são limitados. Coombes e Culverhouse (2003) estudaram a atenção visual seletiva de citopatologistas durante a análise de células cervicais. Os autores utilizaram um equipamento de rastreamento ocular para coletar os dados de atenção e, posteriormente, identificaram atributos que podem contribuir para o desenvolvimento de sistemas de controle de qualidade na leitura de lâminas cérvico-vaginais. A principal conclusão dos autores é que mapas de saliência estimados podem fornecer *feedback* ao citopatologista sobre suas buscas e reduzir a divergência de laudos dentro da equipe em rotinas de laboratório. Outro importante achado foi a correlação existente entre as fixações oculares dos citopatologistas e a coloração das células na imagem. Zhang *et al.* (2013) propuseram um método para detecção de células anormais baseado em mecanismos da atenção *bottom-up* e *top-down*. Os autores exploraram atributos de orientação e intensidade, inspirados pelo trabalho de Itti *et al.* (1998), para criar um mapa de saliência, o qual era modulado pela

resposta de um filtro. Esse filtro era projetado para capturar o tamanho e a intensidade dos *pixels* dos núcleos anormais, características relacionadas à atenção *top-down* na análise citopatológica. Zhang *et al.* (2013) concluíram que o uso de mecanismos de atenção visual seletiva contribuem para a detecção de lesões em exames sem a necessidade de processamento massivo de todos os campos da imagem. No entanto, seus resultados foram validados apenas em imagens de células cervicais obtidas em meio líquido (ZHU *et al.*, 2007).

Os trabalhos desenvolvidos por Coombes e Culverhouse (2003) e Zhang *et al.* (2013) citados anteriormente não utilizam algoritmos de predição de saliência treinados com informação de especialistas humanos. Além disso, esses trabalhos não investigaram o uso de técnicas de aprendizado profundo para estimar estruturas salientes. Muitos algoritmos disponíveis seguem a abordagem clássica, baseada em segmentação de núcleos e/ou citoplasma (CONCEIÇÃO *et al.*, 2019). Destes, alguns utilizam a segmentação como etapa para a extração de características, como textura do núcleo, relação núcleo/citoplasma, cor, etc (Figura 10a); outros dependem da segmentação para localizar núcleos e então classificar recortes da imagem utilizando uma CNN (JITH *et al.*, 2018) (Figura 10b). No entanto, devido às dificuldades discutidas na Seção 2.1, estratégias que não dependem do uso segmentadores para a seleção automática de regiões de interesse têm sido recentemente investigadas para análise de imagens de células cervicais. Devido ao escopo desta tese, continuaremos nossa revisão focando nesta categoria de modelos, a qual está ilustrada na Figura 10c.

Zhang *et al.* (2018) propuseram um método que explora o algoritmo *Region-based Fully Convolutional Networks* (R-FCN) (DAI *et al.*, 2016) para a localizar regiões anormais em imagens de Papanicolaou sem a necessidade de segmentar núcleos e/ou citoplasma. O algoritmo R-FCN corresponde a uma rede neural convolucional desenvolvida para detecção de objetos em imagens naturais. Zhang *et al.* (2018) realizaram treinamento do modelo utilizando uma técnica de transferência de aprendizado (do inglês, *transfer learning*)<sup>4</sup> a partir da base de dados *ImageNet* (KRIZHEVSKY *et al.*, 2012). Então os pesos foram ajustados (*fine-tuning*)<sup>5</sup>, com as imagens de uma base de dados privada fornecida pelo *Guangzhou LBP Medicine Science & Technology Co., Ltd.*. Os resultados obtidos por Zhang *et al.* (2018) são evidências que a transferência de aprendizado a partir da base de dados *ImageNet* é viável para categorização de

<sup>4</sup> Técnica empregada em CNNs que permite reutilizar os pesos previamente treinados em grandes bases de dados em outros problemas. Antes de treinar ou utilizar a rede, os pesos são inicializados com aqueles resultantes de um treinamento anterior com rede de mesma arquitetura. O domínio do problema não precisa ser o mesmo. Para saber mais sobre essa técnica, consulte (YOSINSKI *et al.*, 2014).

<sup>5</sup> Após o procedimento de transferência de aprendizado (YOSINSKI *et al.*, 2014), os pesos da CNN podem ser submetidos a um ajuste fino (do inglês, *fine-tuning*) para melhor convergência com o novo problema.

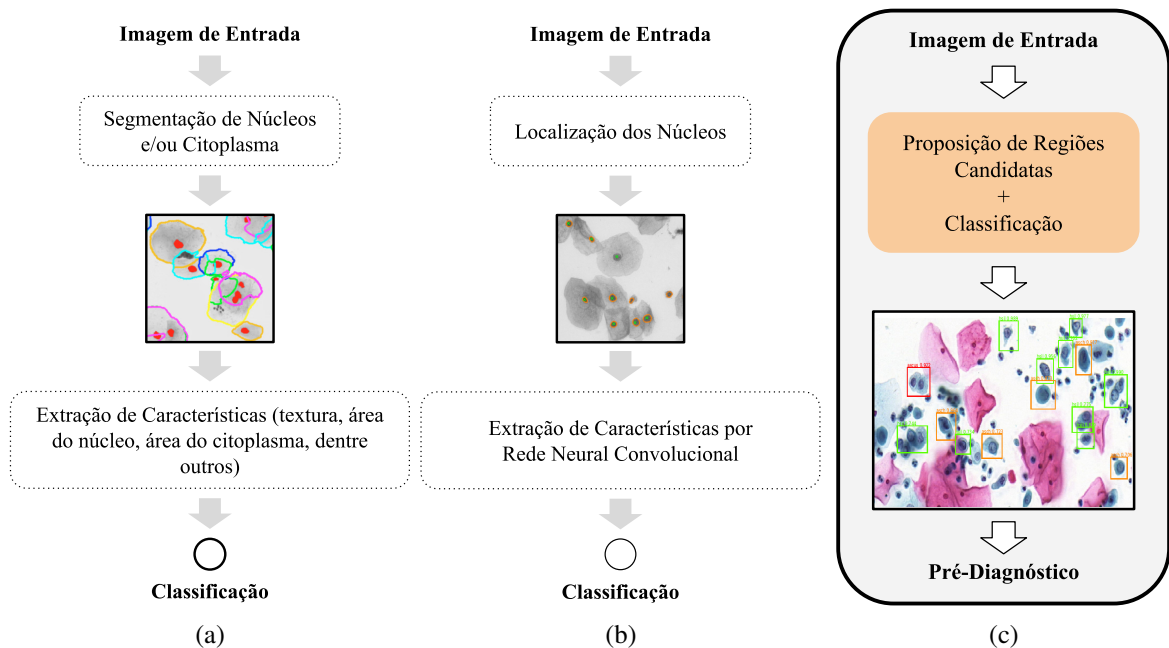


Figura 10 – Estratégias comumente adotadas no desenvolvimento de sistemas computacionais para auxílio à análise de exames citopatológicos do colo do útero. a) Abordagem típica, utiliza a segmentação do núcleo e/ou citoplasma para extração de características das células. b) Esquema baseado em um detector de núcleos, não sendo necessário estimar bordas com precisão. As marcações em verde e laranja representam o *Ground Truth* (GT) e os núcleos detectados, respectivamente. c) Utiliza seleção de regiões de interesse para categorização, sem envolver segmentações. Empregamos os resultados dos métodos propostos por Ramalho *et al.* (2015), Phoulady e Mouton (2018), Liang *et al.* (2019) para as ilustrações de (a), (b) e (c), respectivamente. A abordagem em destaque foi a empregada nesta tese. Fonte: elaboração própria.

imagens de células cervicais. Os autores revelaram também que a classificação a nível de regiões é uma estratégia promissora para os casos onde existam aglomerados de células com elevada sobreposição. Além disso, Zhang *et al.* (2018) mostraram que, em sistemas CAD, sugerir regiões com anormalidades tem maior importância para a rotina clínica do que detectar com exatidão as células doentes, pois a análise posterior do citopatologista humano é indispensável. No entanto, esse trabalho utilizou apenas 62 imagens de células cervicais, todas obtidas em meio líquido. Os resultados também revelaram a tendência a falsos-positivos, prejudicando a precisão do método. Além disso, os autores não apresentaram resultados comparativos com a metodologia clássica baseada em segmentação e se limitaram a estudar apenas duas classes: normal e anormal.

Liang *et al.* (2019) também utilizaram modelos de aprendizado profundo guiados por seleção de regiões, mais precisamente *Region Proposal Networks* (RPNs), para localizar elementos de interesse clínico em exames citopatológicos do colo do útero. Os autores combinaram as técnicas estado-da-arte *Region-based Convolutional Neural Network* (Faster R-CNN) (REN *et*

*al.*, 2017) e *Feature Pyramid Network* (FPN) (LIN *et al.*, 2017) para extrair diversas regiões de uma base de imagens, criar um dicionário com amostras representativas para algumas estruturas e realizar a classificação de imagens de teste com base nas referências obtidas. Diferentemente de Zhang *et al.* (2018), Liang *et al.* (2019) abordaram as diferentes classes do Sistema Bethesda (Tabela 1), incluindo outros elementos, como células glandulares atípicas, *trichomonas*, cândida, flora, herpes e *actinomyces*. Além disso, Liang *et al.* (2019) realizaram experimentos com dois conjuntos de treinamento de tamanhos distintos, confirmando que o desempenho dos modelos RPNs são dependentes da quantidade de amostras disponíveis para treinamento. Os autores observaram também que a diversidade de elementos existentes nas imagens de células cervicais faz as RPNs produzirem um significativo volume de falsos positivos, ou seja, regiões de fundo que não correspondem a estruturas relevantes. Dessa forma, amostras com áreas de fundo das imagens precisaram ser incorporadas na confecção do dicionário deste modelo, o que aumentou a necessidade de mais dados rotulados. Os resultados apresentados mostram que classificar regiões com base na distância para amostras de referência é promissor em imagens de células cervicais, superando a estratégia típica das RPNs, a qual depende de maior disponibilidade de dados. No entanto, Liang *et al.* (2019) validaram seu trabalho apenas em imagens de células cervicais extraídas de exames em meio líquido, os quais possuem melhor qualidade para identificação dos elementos clínicos avaliados.

## 5 BASE DE DADOS DE ATENÇÃO VISUAL PROPOSTA

Este capítulo apresenta a base de dados *CRIC Visual Attention* (CRICVA) com dados de fixações oculares de citopatologistas. Essa base é pioneira em publicar dados de atenção visual em imagens de células cervicais, sendo, portanto, uma das contribuições desta tese (Contribuição 1). Primeiramente, descrevemos a metodologia adotada para o rastreamento da atenção visual dos especialistas, destacando os critérios para a seleção das imagens e os parâmetros necessários para a execução de nossos experimentos. Logo após, detalhamos a base CRICVA, discutindo a validação dos dados e as características da base proposta.

### 5.1 Rastreamento da Atenção Visual dos Citopatologistas

Registramos os dados de atenção visual dos citopatologistas utilizando um equipamento de rastreamento ocular (do inglês, *eye tracker*)<sup>1</sup>. O estudo foi aprovado pelo comitê de ética da Universidade Federal do Ceará sob o protocolo 2.439.252 (Anexo A). Todas as atividades foram conduzidas de acordo com preceitos éticos de pesquisa pautados na Declaração de Helsink<sup>2</sup> e nas leis brasileiras. A Figura 11 ilustra o procedimento adotado para a coleta de dados e as subseções seguintes detalham cada uma das etapas.

#### 5.1.1 Seleção de Imagens

Selecionamos um conjunto de 232 imagens de células cervicais da base de dados CRIC para serem exibidas aos especialistas. A escolha dessas imagens obedeceu aos seguintes requisitos:

1. imagens com células distribuídas por toda área de visualização da imagem;
2. imagens contendo diferentes artefatos, como células de sangue, inflamatórias, distorcidas, com sobreposição, além de outros elementos discutidos na Seção 2.1;
3. imagens escolhidas em comum acordo entre dois pesquisadores<sup>3</sup> do projeto;
4. 80 imagens selecionadas aleatoriamente das remanescentes da base CRIC (Figura 11b), totalizando o conjunto de 232 imagens da base proposta.

<sup>1</sup> Experimentos detalhados na Seção 5.1.3.

<sup>2</sup> A Declaração de Helsink<sup>2</sup> corresponde a um documento de princípios éticos para orientar médicos e outros participantes de pesquisas clínicas envolvendo seres humanos. Acesso: <<https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/>>

<sup>3</sup> Os pesquisadores que participaram desta etapa do projeto tinham experiência em visão computacional e conhecimento para identificar os principais elementos presentes nas imagens.

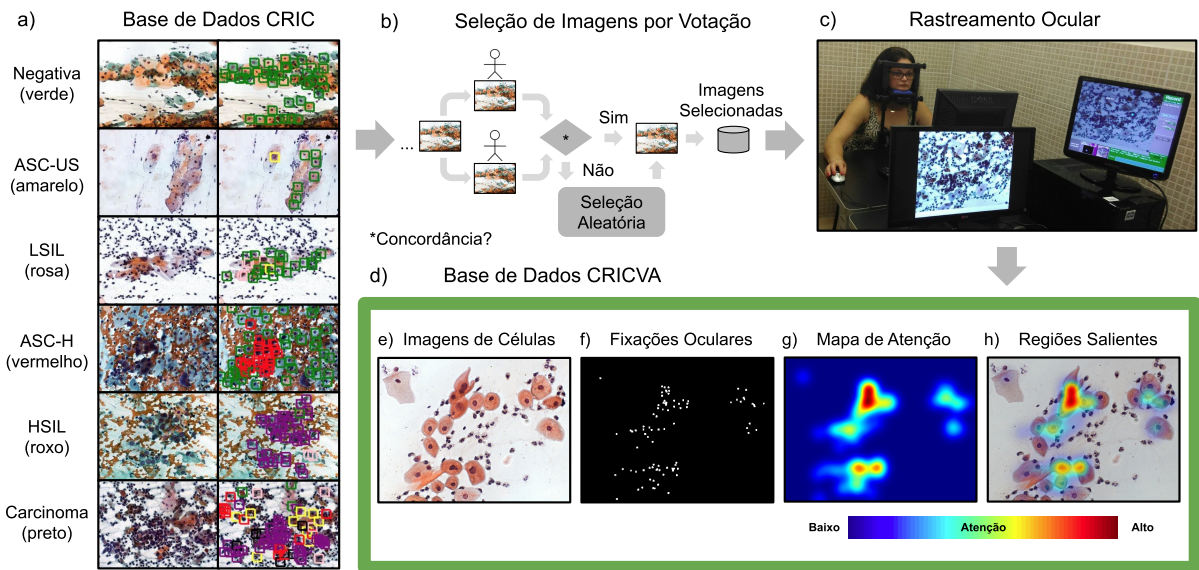


Figura 11 – Coleta dos dados de atenção visual dos citopatologistas. a) Imagens da Base Centro de Reconhecimento para Inspeção de Células (CRIC) com os respectivos rótulos. Os rótulos não foram apresentados aos especialistas. b) Metodologia para selecionar as imagens reduzindo o *viés de seleção* (TORRALBA; EFROS, 2011). c) Uso de um equipamento de rastreamento ocular para coletar os movimentos oculares dos especialistas. d) Dados contidos na nova base de dados CRICVA. e) Imagem colorida de células cervicais. f) Fixações oculares dos citopatologistas. g) Mapa de atenção visual resultante da convolução de uma gaussiana em cada ponto de fixação. h) Regiões salientes da imagem reveladas pela sobreposição da imagem apresentada com o respectivo mapa de atenção. As regiões em vermelho escuro correspondem às áreas com maior atratividade e as em azul escuro, as que demandam menor atenção. Fonte: elaboração própria.

Para atender ao Requisito 3, os critérios 1 e 2 listados foram executados por dois pesquisadores do projeto. Cada um deles fez a escolha das imagens, conforme seu próprio julgamento e sem ter conhecimento da decisão do outro avaliador. Dessa forma, o pesquisador 1 escolheu o conjunto de imagens  $J_1$  a partir da população de imagens  $J_{cric}$  existentes na base CRIC. O pesquisador 2, por sua vez, selecionou o conjunto de imagens  $J_2$ . O conjunto final de amostras  $J_{cricva}$  pertencentes à base CRICVA pode ser visto da seguinte forma:

$$J_{cricva} = \{(J_1 \cap J_2) + rand(J_{cric} - (J_1 \cap J_2); q) \mid J_1, J_2 \in J_{cric} \text{ e } q = 80\},$$

onde  $rand(\cdot)$  é uma função para a seleção de amostras aleatórias. Essa componente de aleatoriedade foi inserida para reduzir o *viés de seleção*<sup>4</sup> na escolha dos conjuntos.

<sup>4</sup> O viés de seleção está relacionado à preferência por um tipo particular de imagem durante a composição da base de dados (TORRALBA; EFROS, 2011)



### 5.1.2 *Participantes*

Os dados de atenção visual necessários ao desenvolvimento desta tese foram coletados de três citopatologistas, com visão normal ou corrigidas por lentes de grau. Todos os participantes são especialistas na leitura de esfregaços cérvico-vaginais convencionais e em cuidados com a saúde do colo uterino. A média de idade dos citopatologistas é de 34,3 anos (46, 30 e 27) e suas carreiras em laboratório especializado em citologia é 20, 8 e 3 anos, respectivamente. Os dois participantes mais experientes possuem título de doutorado, enquanto o outro de mestrado, em cursos de citopatologia. Os participantes assinaram um termo de consentimento (Apêndice A) para sua participação neste projeto.

### 5.1.3 *Experimento de Rastreamento Ocular*

Utilizamos um rastreador ocular para coletar os dados de atenção visual dos citopatologistas durante a análise de imagens de exames de Papanicolaou. Nesta tese, enfatizamos a *fixação ocular*, que é uma das informações obtidas pelo equipamento. A fixação é o período em que o olho permanece imóvel, refletindo a atratividade de uma área específica da imagem (CARRASCO, 2011). Para cada imagem, representamos a localização dessas fixações como uma matriz binária, conforme mostrado na Figura 11f. A partir desses pontos, construímos um mapa de atenção, convoluindo com uma gaussiana bidimensional isotrópica com 1 (1 grau) do ângulo visual centrado em cada fixação. Essa abertura é escolhida para corresponder ao tamanho estimado da fóvea (TORRALBA *et al.*, 2006; MEUR; BACCINO, 2013). As regiões salientes foram identificadas sobrepondo a imagem apresentada com os mapas de atenção obtidos. Vermelho escuro sinaliza as regiões com maior atratividade e azul escuro o caso contrário.

Os experimentos foram realizados utilizando o equipamento de rastreamento ocular *EyeLink 1000*, projetado pela *SR Research Ltd., Mississauga, Canadá*. Configuramos o equipamento para obter as fixações oculares com uma taxa de amostragem de 1000 Hz, fazendo gravações do olho direito<sup>5</sup> do especialista (Figura 11c). As tarefas foram criadas utilizando o *software* proprietário do equipamento (*SR Research Experiment Builder V4*) e uma sequência de 8 conjuntos de imagens (*trials*) foram apresentadas a cada participante em um monitor Dell E178FPC a 60 Hz. Cada *trial* foi composta por imagens de células cervicais de acordo com a seguinte distribuição: [*Trial*: Quantidade de Imagens]: 01: 26; 02: 26; 03–06: 25, e 07–08:

<sup>5</sup> Em equipamento similar, gravações do olho direito também foram obtidas em (MATSUMOTO *et al.*, 2011).

40. Os citopatologistas foram expostos às mesmas imagens em cada *trial*. A motivação para ter diferentes quantidade de imagens em cada conjunto foi permitir pesquisas futuras a respeito do desempenho dos citopatologistas ao longo do tempo. Antes de cada *trial*, os citopatologistas realizaram uma calibração de nove pontos (DUCHOWSKI, 2017) para mapear as fixações oculares nas coordenadas da tela. Além disso, os participantes tiveram a oportunidade de descansar e reportar qualquer desconforto durante os intervalos de apresentação dos conjuntos de imagens.

Conduzimos um experimento guiado por tarefa<sup>6</sup>, no qual os citopatologistas participantes deveriam interpretar as imagens e sinalizar as células cervicais alteradas utilizando cliques do *mouse*. O tempo de análise era livre para cada participante e os cliques registrados do *mouse* não eram visíveis no monitor<sup>7</sup>. As imagens exibidas foram redimensionadas para 1280 x 1024 *pixels*, mantendo a razão de tela (*aspect ratio*) original por meio da adição de *pixels* branco na parte inferior da imagem. Os participantes utilizaram a tecla de espaço do teclado para passar as imagens, as quais eram exibidas em sequência aleatória, em cada *trial*, para reduzir a influência do padrão sequencial. Os participantes não podiam retornar as imagens passadas. No final de cada *trial*, os dados de atenção visual eram salvos e uma nova rodada podia ser então iniciada.

## 5.2 CRIC Visual Attention (CRICVA)

A base de dados proposta, chamada CRICVA<sup>8</sup>, é composta por 232 ({Negativa: 36, ASC-US: 75, LSIL: 94, ASC-H: 7, HSIL: 6 e Carcinoma: 14}) imagens contendo 5.654 células cervicais ({Negativa: 4285, ASC-US: 408, LSIL: 488, ASC-H: 97, HSIL: 293 e Carcinoma: 83}), sendo em média 24,37 células por imagem e Desvio Padrão (SD) de 16,36. Essa base disponibiliza dados de fixação ocular de três citopatologistas com experiência na identificação de lesões pré-cancerígenas em exames de Papanicolaou. A Figura 12 apresenta algumas dessas imagens, com seus respectivos mapas de atenção. A Figura 13 detalha a distribuição dos casos conforme anotações da base CRIC.

<sup>6</sup> Neste tipo de experimento, os observadores olham a imagem e procuram realizar alguma tarefa, como, por exemplo, procurar um alvo ou interpretar a cena para posterior questionário. Essa modalidade contrasta com a *observação livre*, onde nenhum objetivo é dado aos participantes e eles apenas olham a imagem por um tempo determinado ou não. Uma discussão a respeito do tema foi introduzido na Seção 3.1.2.1

<sup>7</sup> A não apresentação dos pontos já clicados evita a geração de elementos de atração artificiais na cena em análise.

<sup>8</sup> Todos os dados estão disponíveis para *download* e citáveis em <<http://dx.doi.org/10.17632/bk45c9yxb9.1>>.

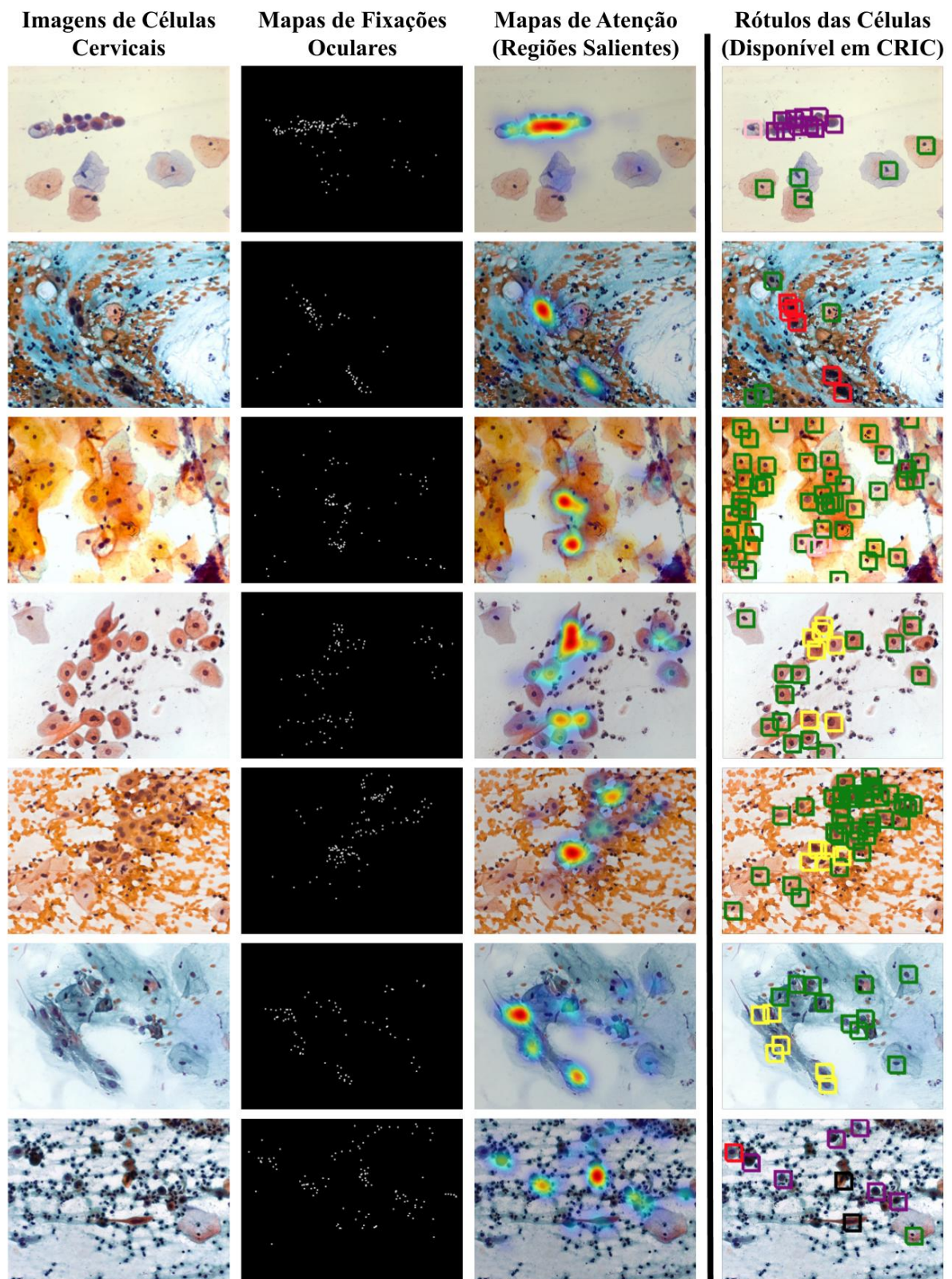
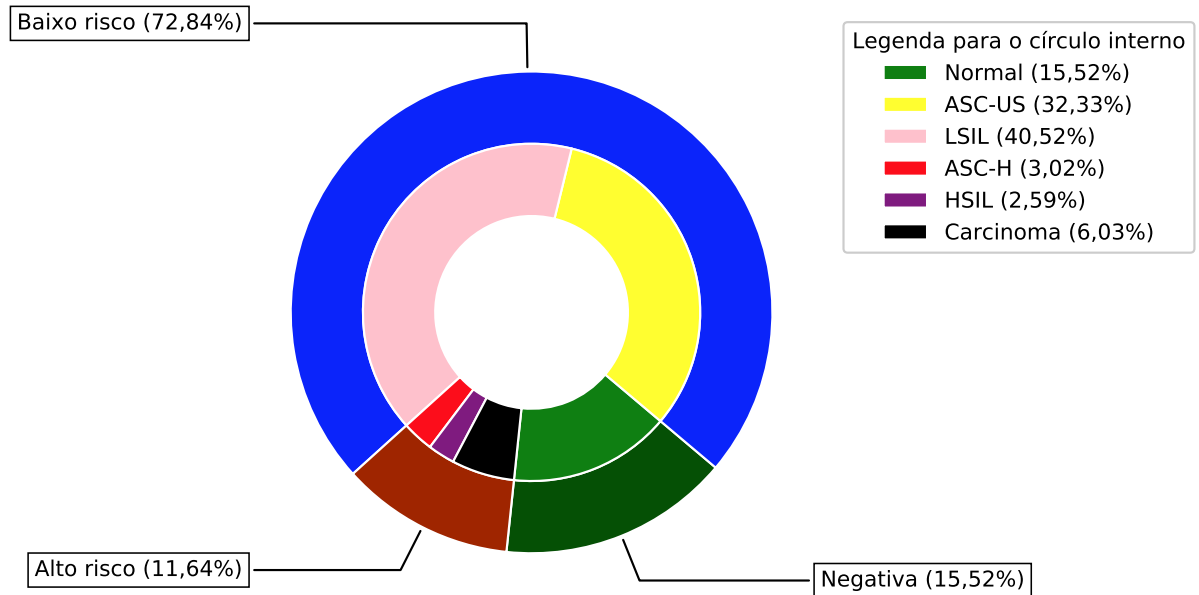
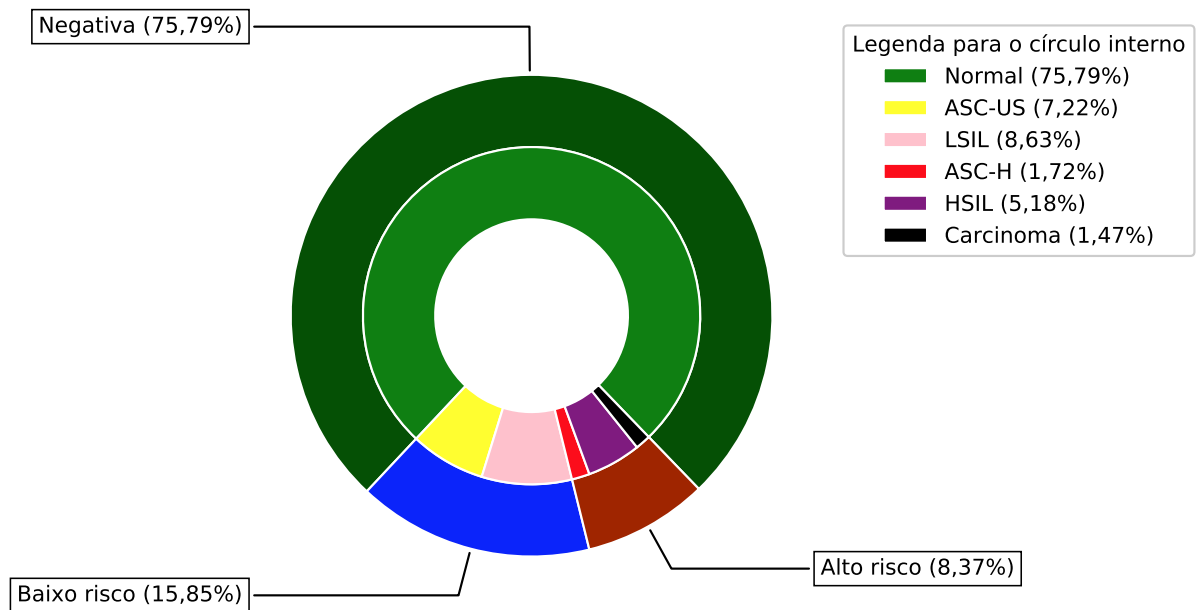


Figura 12 – Amostras da base de dados CRICVA. Os pontos do mapa de fixação foram dilatados para melhor visualização. Nos mapas de atenção, as regiões em vermelho escuro correspondem às áreas com maior atratividade e as em azul escuro, caso contrário. Os rótulos das células estão disponíveis na base CRIC (CRIC, 2018). Referência (Cor: Alteração): Verde: Negativa, Amarelo: ASC-US, Rosa: LSIL, Vermelho: ASC-H, Roxo: HSIL e Preto: Carcinoma. Fonte: elaboração própria.



(a)



(b)

Figura 13 – Distribuição de casos na base CRICVA, conforme anotações da base CRIC. a) Casos de acordo com o rótulo da imagem, ou seja, classe do principal achado detectado no campo. Total de imagens (campos): 232. b) Distribuição de classes considerando todas as células. Total de células anotadas: 5.654. Fonte: elaboração própria.

### 5.2.1 *Consistência entre os Participantes*

Inspirados por Volokitin *et al.* (2016) e Bylinskii *et al.* (2018), medimos a concordância entre os mapas de atenção visual de cada citopatologista para cada imagem. O objetivo foi identificar se a distribuição da atenção dedicada a cada imagem ocorreu de forma similar para todos os citopatologistas. Para esta avaliação, julgamos não ser importante regiões de falso positivo, ou seja, regiões observadas por um determinado especialista que não foram vistas pelos demais. Nossa metodologia consistiu então em criar um mapa de fixação ocular para cada especialista e então utilizá-lo para prever as fixações dos demais. Utilizamos a Área sob a curva ROC (*Receiver Operating Characteristic*) (AUC), na variação proposta por Judd *et al.* (2009), para quantificar essa concordância, pois regiões de falso positivo possuem baixo impacto nos resultados desta métrica (Seção 7.1.3).

O valor médio da métrica AUC entre todos os participantes foi de 0,823 (SD=0,007) e o *p-value* de 0,117 (Kruskal-Wallis ANOVA *test* com  $\alpha = 0,05$ ) (KRUSKAL; WALLIS, 1952), considerando todas as imagens da base CRICVA. Estes resultados sugerem uma alta consistência entre as fixações oculares dos citopatologistas participantes, evidenciando a existência de áreas de análise clínica bem definidas. Isso valida a base CRICVA para o desenvolvimento de estudos em imagens de células cervicais que buscam explorar a atratividade de determinadas regiões pela atenção humana.

### 5.2.2 *Viés de Centro*

A tendência humana de olhar para objetos próximos a parte central da imagem corresponde a um viés comum em base de dados de visão computacional (TATLER, 2007; BORJI *et al.*, 2014). Além disso, essas bases de dados estão sujeitas ao *viés de captura*, ocasionado pela preferência dos fotógrafos em posicionar o objeto de interesse próximo ao centro da imagem (BORJI *et al.*, 2015). Para entender como o *viés de centro* ocorre na base de dados proposta, avaliamos o *Average Annotation Map* (AAM) da base CRICVA, como mostrado na Figura 14. O AAM corresponde à média dos mapas de atenção coletados para todas as imagens da base (BORJI *et al.*, 2015). Para analisar a dispersão do AAM, contrastamos linhas de centro na vertical e horizontal com aquelas obtidas a partir de uma gaussiana no centro da imagem, como ilustrado nas Figuras 14b e 14c, respectivamente. Embora o AAM tenha forte ativação próximo ao centro da imagem, existe uma variação significativa na distribuição espacial

das ativações, indicando que os citopatologistas dedicaram tempo para analisar objetos distantes da região central da imagem.

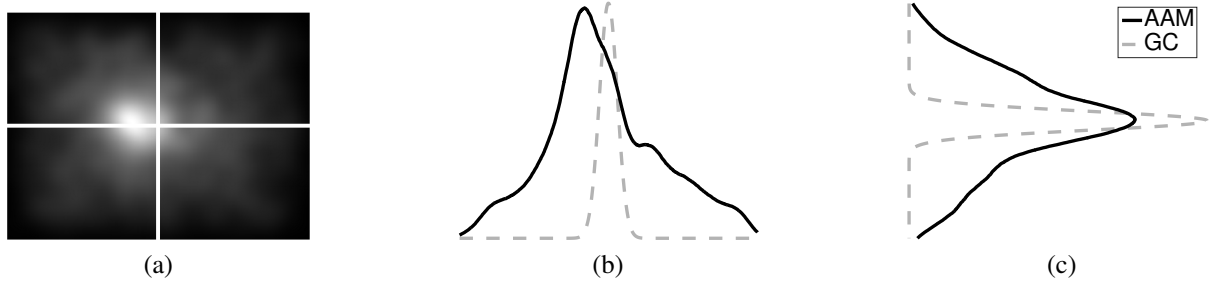


Figura 14 – Análise do *viés de centro* da base de dados CRICVA. a) *Average Annotation Map* (AAM). Linhas b) horizontal e c) vertical no centro do AAM contrastando com as obtidas a partir de uma gaussiana isotrópica no centro da imagem (GC). Fonte: elaboração própria.

## 6 MODELO PROPOSTO: *ATTENTION GUIDED MODEL FOR INSPECTION OF CELLS* (AGMIC)

Esta tese propõe um modelo em dois estágios, chamado *Attention Guided Model for inspection of cells* (AGMic), para análise automática de imagens de células cervicais (Figura 15). O primeiro módulo (Módulo 1) corresponde à atenção computacional, estimando Regiões de Interesse (ROIs) por predição de saliência e selecionando células de potencial interesse clínico. No segundo módulo (Módulo 2), o modelo classifica as células selecionadas, bem como a imagem, em três classes: normal, baixo risco e alto risco (Tabela 1).

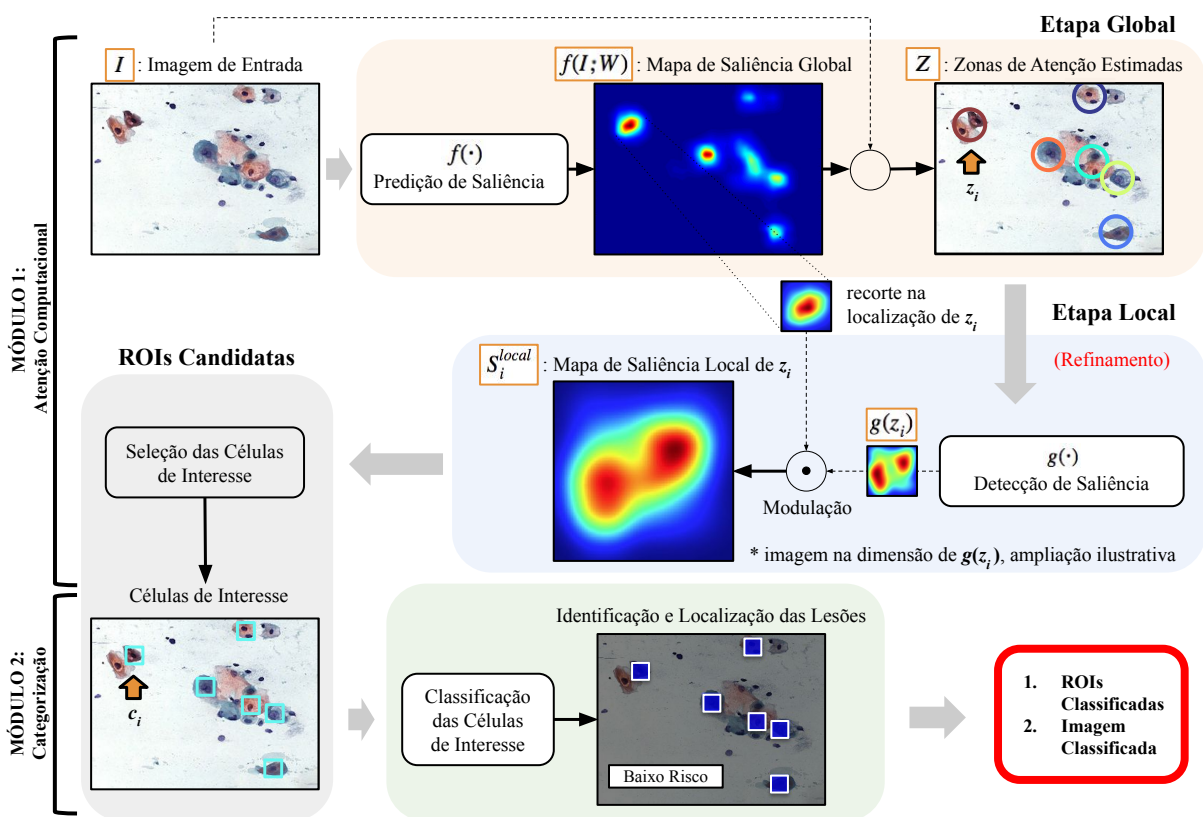


Figura 15 – Modelo AGMic. O primeiro módulo seleciona células de interesse baseado na predição de saliência. O segundo módulo realiza a classificação dessas regiões e sugere um rótulo para a imagem de entrada. Os mapas de saliência são apresentados coloridos apenas para ilustração. Fonte: elaboração própria a partir de uma amostra da Base Centro de Reconhecimento para Inspeção de Células (CRIC).

Considerando  $I$  uma imagem de células cervicais definida em  $D_I \subset \mathbb{Z}^3$  com dimensões  $m \times n$  e valores de *pixels* em  $\{0, \dots, 255\}^3$ , o Módulo 1 computa mapas de saliência definidos em  $D_M \subset \mathbb{R}^2$  por meio de  $f : D_I \rightarrow D_M$  e  $g : D_I \rightarrow D_M$ . Conforme ilustrado na Figura 15, a etapa global do Módulo 1 mapeia as zonas conspícuas da imagem em uma matriz  $Z$  a partir do

mapa de saliência  $f(I; W)$ , onde  $W$  reflete o mecanismo de atenção (*bottom-up* e/ou *top-down*) adotado.<sup>1</sup>

O Módulo 1 usa  $g(\cdot)$  para capturar características indicativas de alteração celular e então refinar o mapa de saliência obtido globalmente. Para cada zona de atenção  $z_i \in Z$ , o modelo proposto gera um mapa de saliência  $S_i^{local} \in D_M$  unindo a distribuição da atenção quanto à imagem completa, predita por  $f(\cdot)$ , a probabilidade das estruturas locais de  $z_i$  indicarem alteração celular. A equação que descreve esse processo é:

$$S_i^{local} = g(z_i) \odot crop(f(I; W); (p, q)), \quad \forall z_i \in Z \quad (6.1)$$

onde  $(p, q)$  são as coordenadas em  $Z$  do centroide de  $z_i$ ,  $crop(\cdot)$  é uma função de recorte que retorna uma matriz bidimensional com mesmo tamanho de  $g(z_i)$  e  $\odot$  simboliza a operação de modulação por amplitude. Cada  $S_i^{local}$  gerado é posicionado em relação a  $I$  por meio de  $(p, q)$ . A *i-ésima* célula de interesse é demarcada pelo *bounding box*  $c_i$  de dimensões  $(a, b)$  e centroide no ponto de máximo de  $S_i^{local}$ . Assim, para  $qz$  zonas de atenção, temos  $V_c$  como o conjunto de todos  $c_i$  (ROIs):

$$V_c = \left\{ boundingBox(I; max(S_{i=0}^{local}); (a, b)), \dots, boundingBox(I; max(S_{i=qz-1}^{local}); (a, b)) \right\}, \quad (6.2)$$

onde  $max(\cdot)$  captura o ponto máximo de  $S_i^{local}$ . A função  $boundingBox(\cdot)$  define uma área retangular de  $I$ , com dimensões  $(a, b)$ , cujo centroide equivale à posição do *pixel* de maior saliência.

Tendo  $V_c$  como entrada do Módulo 2 do AGMic, a classe da imagem é inferida a partir da classificação isolada de cada  $c$ . Portanto, a saída do modelo proposto é dupla: localização e rotulação local de partes da imagem definidas de acordo com a atenção visual estimada de citopatologistas e a classe do campo. Nossa abordagem para determinar a classe da imagem é assumi-la com o mesmo rótulo da alteração de maior impacto dentre as ROIs. Fundamentamos essa estratégia na necessidade de reportar lesões críticas, mesmo que suas ocorrências sejam pontuais na imagem em análise (BENGTSSON; MALM, 2014). Assim, o método proposto permite a triagem<sup>2</sup> de imagens de células cervicais, a qual pode ser executada a partir da existência de alteração no campo ou pela quantidade de áreas conspícuas aos citopatologistas com a presença de células anormais. O AGMic corresponde à Contribuição 2 desta tese.

<sup>1</sup> As zonas de atenção definidas em  $Z$  são obtidas pela aplicação de um filtro de máximo local em  $f(I; W)$ . As coordenadas dos pontos de máximo obtidos correspondem aos centroides das diferentes zonas de atenção. Todas as zonas possuem o mesmo tamanho, com o valor de raio obtido experimentalmente.

<sup>2</sup> O termo triagem é empregado para representar o processo de *seleção* de campos feito por especialistas para apontar quais exames possuem maior potencial de anormalidade.



## 6.1 Módulo 1: Detecção de Regiões de Interesse com Base em Atenção Visual

Exploramos os dados de atenção visual dos citopatologistas, coletados nos experimentos de rastreamento ocular descritos na Seção 5.2, para modelar a informação *top-down* na predição de saliência. Dessa forma, alcançamos um modelo *Convolutional Neural Network* (CNN) que simula a atenção visual de especialistas na busca por partes clinicamente relevantes da imagem. Esse modelo corresponde à componente  $f(\cdot)$  do AGMic e produz um mapa de saliência de aspecto global, onde a imagem  $I$  é vista por completo. Em seguida, o mapa de saliência estimado é refinado em etapa local, onde propomos um algoritmo que explora a intensidade, textura e o tamanho do núcleo celular para posicionar as ROIs<sup>3</sup> em áreas com potencial anormalidade.

### 6.1.1 Etapa Global

O modelo *Saliency Attentive Model* (SAM) é um algoritmo desenvolvido por Cornia *et al.* (2018) para predição de saliência em imagens naturais<sup>4</sup>. Este modelo encontra-se validado no *benchmark* do MIT (BYLINSKII *et al.*, 2015) para uso em imagens naturais. Empregamos o modelo SAM como a componente  $f(\cdot)$  do AGMic, propondo, portanto, um modelo de aprendizagem profunda otimizado para a predição de saliência em imagens de células cervicais. Teoricamente,  $f(\cdot)$  pode assumir qualquer algoritmo de predição de saliência, seja com características *bottom-up*, *top-down* ou combinação de ambas. Essa condição levou à investigação de duas hipóteses, cujos resultados são discutidos na Seção 7.1. Como o algoritmo SAM está validado apenas para imagens naturais, esta tese apresenta um modelo SAM capaz de reproduzir a atenção visual dos citopatologistas em imagens de células cervicais, a partir da aprendizagem dos padrões codificados na base de dados CRIC *Visual Attention* (CRICVA) (Seção 5.1) proposta.

A Figura 16 sintetiza as etapas do algoritmo SAM, a partir do exposto na Figura 9, e o empregamos nesta tese conforme a arquitetura concebida por Cornia *et al.* (2018). De forma geral, trata-se de um algoritmo que utiliza CNN como extrator de características da imagem e como mecanismo para modelar atenção *top-down*. O vetor de características (representado pelo conjunto de mapas de ativação em vermelho) é processado por uma rede *Attentive Convolutional Long Short-Term Memory* (*Attentive ConvLSTM*) para realce das regiões de saliência. Um

<sup>3</sup> As ROIs correspondem aos *bounding boxes* que identificam as células de interesse a ser submetidas à categorização (Módulo 2).

<sup>4</sup> Nesta tese, imagens naturais correspondem a imagens do dia-a-dia, ou seja, fotografias que revelam cenas comuns do cotidiano.

provável viés de centro do conjunto de dados de rastreamento ocular é modelado por gaussianas isotrópicas. Então, o vetor de mapas é fundido em um único mapa por meio de uma camada convolucional (1x1), gerando o mapa de saliência final. Para informações complementares sobre o algoritmo SAM, bem como o funcionamento da rede *Attentive ConvLSTM* e sua aplicação em imagens naturais, vide (CORNIA *et al.*, 2018).

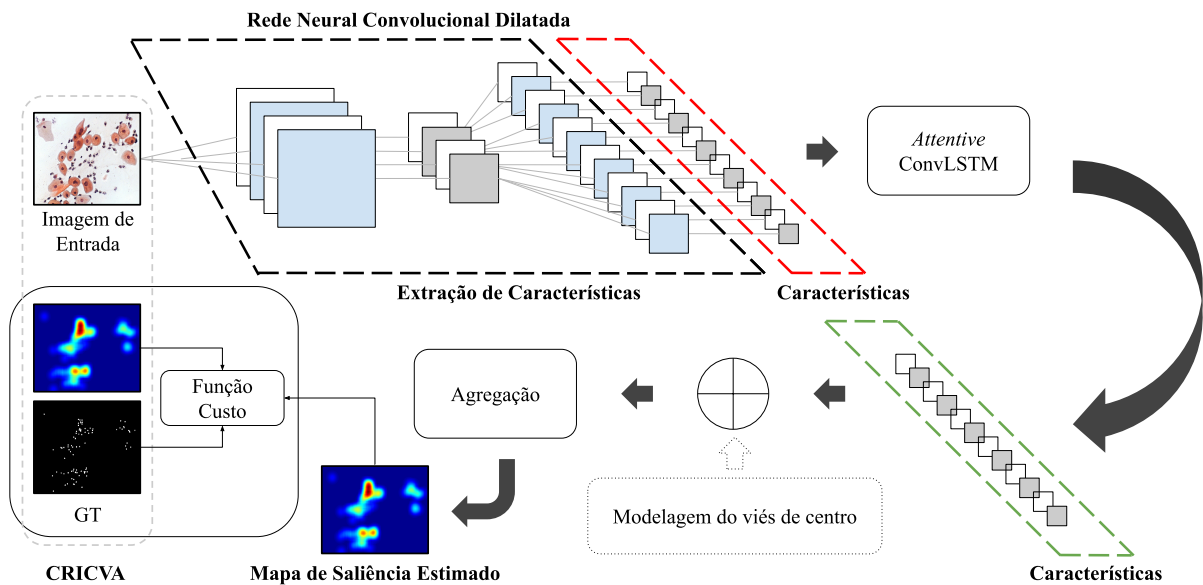


Figura 16 – Arquitetura do modelo SAM utilizado como a componente  $f(\cdot)$  do modelo proposto AGMic. Os mapas de saliência correspondem a matrizes bidimensionais, as quais são apresentadas coloridas para melhor visualização. Fonte: adaptado de (CORNIA *et al.*, 2018).

#### 6.1.1.1 Treinamento do Modelo

O protocolo para treinamento do algoritmo SAM inclui a definição da função custo conforme os achados de Huang *et al.* (2015). Em seu trabalho, Huang *et al.* (2015) reportaram a utilização de métricas projetadas para avaliar modelos de previsão de saliência como parte da função de treinamento. Dessa forma, seguimos a proposta de Cornia *et al.* (2018) para definir a combinação linear das seguintes métricas como função custo: *Normalized Scanpath Saliency* (NSS) (PETERS *et al.*, 2005), *Linear Correlation Coefficient* (CC) e *Kullback-Leibler divergence* (KL). NSS utiliza o mapa de fixações oculares (Figura 11f) como referência, enquanto CC e KL, os mapas de atenção (Figura 11g). Por meio do NSS, a função custo capta informação precisa quanto à estrutura da imagem fixada pelo olhar do observador, enquanto torna o treinamento

robusto a imprecisões no *Ground Truth* (GT)<sup>5</sup> pelas métricas CC e KL. Além disso, NSS e CC penalizam regiões de falsos negativos e falsos positivos igualmente, e KL é sensível a falsos negativos. Assim, a combinação dessas métricas prioriza a redução de falsos negativos, que no contexto de imagens médicas é especialmente relevante. As equações para computação de NSS, CC e KL, bem como suas principais propriedades, estão detalhadas na Seção 7.1.3.

A Equação 6.3 foi introduzida por Cornia *et al.* (2018) e validada como função custo para treinamento do método SAM na predição de saliência em imagens naturais. Por combinar métricas sensíveis a falsos negativos, estendemos sua aplicação a imagens médicas, particularmente a imagens de células cervicais. Assim como em (CORNIA *et al.*, 2018), assumimos  $\zeta = -1$ ,  $\beta = -1$ ,  $\gamma = 10$  como fatores de ponderação para as métricas NSS ( $L_{nss}$ ), CC ( $L_{cc}$ ) e KL ( $L_{kl}$ ), respectivamente, dando destaque à métrica com maior sensibilidade aos falsos negativos (Tabela 3).

$$L_{custo}(\tilde{y}, y^{aten}, y^{fix}) = \zeta L_{nss}(\tilde{y}, y^{fix}) + \beta L_{cc}(\tilde{y}, y^{aten}) + \gamma L_{kl}(\tilde{y}, y^{aten}), \quad (6.3)$$

onde  $\tilde{y}$  é o mapa estimado em cada iteração de treinamento,  $y^{fix}$  corresponde ao mapa de fixações oculares e  $y^{aten}$  representa o mapa de atenção.

NSS e CC são métricas que precisam ser maximizadas, daí seus pesos serem negativos. A medida KL é uma dissimilaridade, portanto, precisa ser minimizada. Cornia *et al.* (2018) conduziram um estudo exaustivo para a formulação da Equação 6.3, revelando que essa combinação permite que o modelo alcance resultados satisfatórios e balanceados quanto às principais métricas de avaliação estado-da-arte. O peso de valor 10, dado à métrica KL, prioriza a redução de falsos negativos, característica crucial quando se processa imagens médicas.

Treinamos o método SAM para propor os modelos SAM VGG-16 *for Cells* (SVC) e SAM ResNet-50 *for Cells* (SRC), com base nas arquiteturas CNN VGG-16 (SIMONYAN; ZISSERMAN, 2014) e ResNet-50 (HE *et al.*, 2016), respectivamente. Empregamos os dados de atenção visual referente a 159 imagens de células cervicais, sendo 130 para treinamento e 29 para validação. Todas as imagens são da base CRICVA. A metodologia consistiu em: 1) *transferência de aprendizado*: inicialização dos pesos da CNN com aqueles obtidos a partir de treinamento na base de dados ImageNet (KRIZHEVSKY *et al.*, 2012)<sup>6</sup>; 2) *ajuste fino*: refinamos os pesos a partir

<sup>5</sup> Durante o rastreamento ocular, podem haver fixações oculares isoladas em partes não observadas da imagem. Essas fixações geralmente são causadas por desvio rápido do olhar do especialista ou mesmo por ruído na leitura do equipamento.

<sup>6</sup> O projeto ImageNet consiste de um grande banco de dados para uso em pesquisas e desenvolvimento de modelos de visão computacional. Trata-se de uma base de propósito geral, cujas imagens representam cenas do

dos dados disponíveis em nosso conjunto de treinamento. Não realizamos operações de *data augmentation*<sup>7</sup> nos dados extraídos pelo rastreador ocular, evitando interferências artificiais nos padrões de visualização implementados pelos especialistas. Os pesos referentes à rede *Attentive ConvLSTM* foram inicializados como matrizes ortogonais aleatórias, conforme sugerido por Cornia *et al.* (2018). Redimensionamos as imagens de entrada para 320 x 240 *pixels* (preservando o *aspect ratio*). Configurações adicionais para treinamento da CNN podem ser encontradas em (CORNIA *et al.*, 2018).

### 6.1.2 Etapa Local

Os mapas estimados por  $f(\cdot)$  se apresentam esparsos em uma região saliente, com perda de precisão espacial quanto a estruturas celulares relevantes. Isso acontece porque as fixações oculares dos citopatologistas estão naturalmente distribuídas entre o objeto foco e sua periferia, resultado da busca por atributos contextuais<sup>8</sup> e da susceptibilidade da atenção humana a fatores *bottom-up*<sup>9</sup>. Assim, não é possível assumir que o pico da atenção estimada estará sobre uma determinada estrutura celular. A Figura 17 exemplifica essa situação, onde o ponto máximo da saliência leva a uma área entre as células rotuladas pelos especialistas.

Trabalhos recentes (SILVA *et al.*, 2019) têm reforçado a importância do núcleo na detecção de lesões em células cervicais. Esta tese introduz um algoritmo para refinar os mapas preditos por  $f(\cdot)$ , centralizando as ROIs no núcleo com maior potencial de lesão. O algoritmo proposto explora os padrões de textura do núcleo, definidos por variações na distribuição de cromatina, bem como da membrana nuclear, para inferir presença de anormalidade.

#### 6.1.2.1 Region-based Structural Co-occurrence Matrix (rSCM)

Esta tese propõe a matriz de coocorrência estrutural baseada em região, chamada *Region-based Structural Co-occurrence Matrix* (rSCM). Esse método relaciona estruturas de

---

cotidiano. Mais de 14 milhões de imagens foram rotuladas à mão pelo projeto e em pelo menos 1 milhão delas as localizações das ROIs também são fornecidas. Desde 2010, o projeto realiza um concurso anual, *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), para incentivar a criação de novos modelos. Neste concurso, pesquisadores competem para elaborar algoritmos capazes de detectar e classificar corretamente objetos e cenas em uma lista de mil classes não sobrepostas. Muitas das arquiteturas CNN estado-da-arte foram propostas a partir da ILSVRC. Detalhes em: <<http://www.image-net.org/>>

<sup>7</sup> Técnica que permite a geração de novos exemplares de dados a partir dos existentes. É bastante usada no treinamento de modelos CNN para melhor generalização.

<sup>8</sup> Citopatologistas tendem a buscar informação de contexto, como relação de tamanho núcleo/citoplasma, células vizinhas, dentre outras, para emitir laudo (WHO; IARC, 2004).

<sup>9</sup> Falhas na coleta e confecção das lâminas podem gerar artefatos (Figura 2) que distraem a atenção do especialista de forma rápida e automática durante a análise.

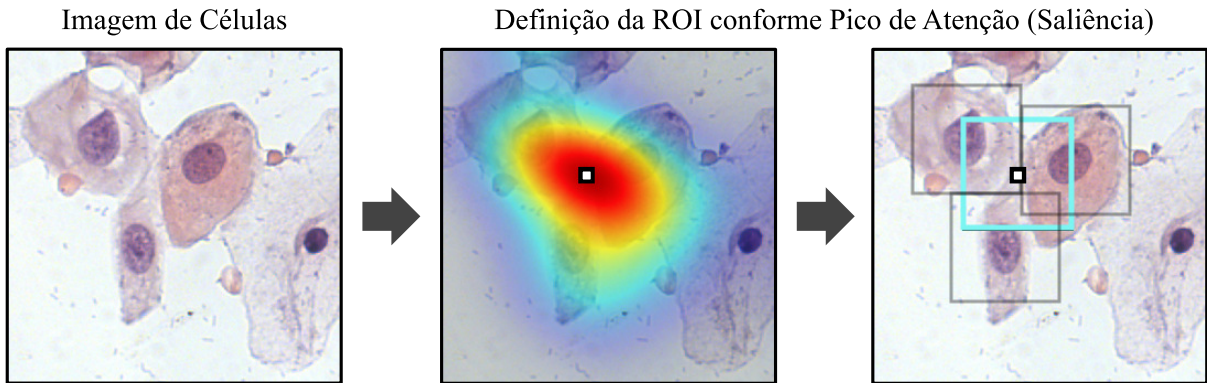


Figura 17 – Exemplo de ROI posicionada no centro das células (quadrado com bordas azuis) quando utilizado o pico da saliência (marcador branco) para definir localização. Os quadrados com bordas cinzas sinalizam as células rotuladas pelos especialistas. O mapa de atenção mostrado corresponde aos dados coletados dos especialistas pelo equipamento de rastreamento ocular em sobreposição com a imagem de entrada. A ROI estimada (quadrado com bordas azuis) não contempla satisfatoriamente a região do núcleo. Fonte: elaboração própria.

baixo nível de duas imagens de entrada, com mesma resolução espacial e de contraste, e mapeia as coocorrências entre suas estruturas em um histograma bidimensional, obedecendo áreas definidas em um mapa. Esse mapa demarca uma região  $U$  da imagem onde as coocorrências devem ser contabilizadas. Particularmente, definir áreas de processamento na computação da rSCM tem duas vantagens: a) permite avaliar apenas regiões específicas da imagem, descartando outras com baixo poder discriminante e b) permite que a matriz capture informação quanto ao tamanho da região em processamento. Tendo  $c$  como a imagem de entrada (*bounding box* contendo uma célula de interesse),  $\bar{c}$  a sua versão alterada por uma função de transformação<sup>10</sup>,  $Q$  como uma função de agrupamento de *pixels* (como, por exemplo, quantização), podemos computar cada elemento  $m_{i,j}$  da matriz rSCM para uma região específica  $U$ :

$$m_{i,j} = \#\{(i,j) | P_U(i,j), i = Q(c)_p, j = Q(\bar{c})_p, p \in U\}, \quad (6.4)$$

onde  $\#$  corresponde a cardinalidade do subconjunto do par  $(i,j)$  na posição  $p$  que satisfaz a propriedade  $P_U$ .  $Q$  define a resolução de contraste da imagem e limita o tamanho da matriz rSCM, minimizando o custo computacional. Assim, se  $Q$  corresponde a uma função de quantização e retorna  $N_Q$  valores de intensidade definidos em  $[0, \dots, N_Q - 1]$ , a matriz rSCM será quadrada de dimensão  $N_Q \times N_Q$ . A propriedade  $P_U$  expressa o quão semelhantes as intensidades de um *pixel*

<sup>10</sup> A função de transformação reflete a processos que modificam a imagem de entrada, como, por exemplo, realce, alterações de contraste, dentre outros. Se existe informação *a priori* do problema, essa transformação pode ser ajustada para maximizar o desempenho da matriz em capturar informação relevante no escopo do problema.

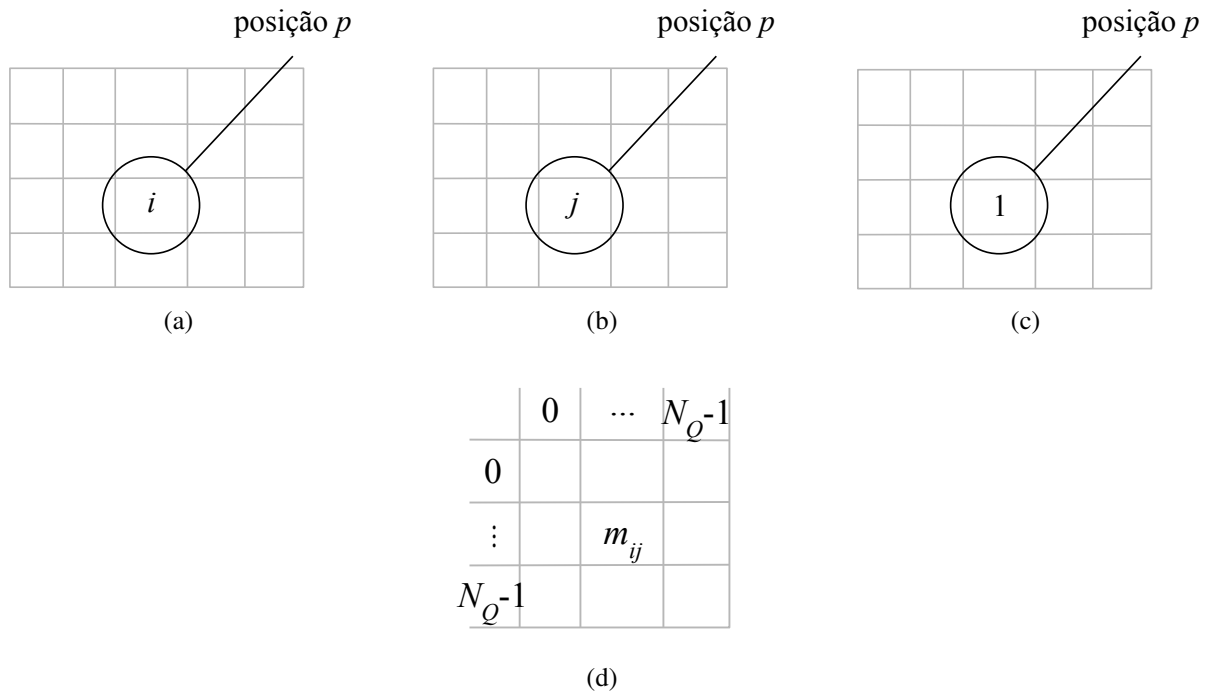


Figura 18 – Ilustração do cálculo da matriz de coocorrência estrutural proposta (rSCM). a) Imagem de entrada  $c$ , onde  $i = Q(c)_p \in [0, \dots, N_Q - 1]$ . b) Imagem de entrada modificada  $\bar{c}$ , onde  $j = Q(\bar{c})_p \in [0, \dots, N_Q - 1]$ . c) Máscara binária para a região  $U$ , onde 1 simboliza as posições que devem ser consideradas. d) Matriz rSCM de dimensão  $N_Q \times N_Q$ , contabilizando as coocorrências do par  $(i, j)$ . Fonte: elaboração própria.

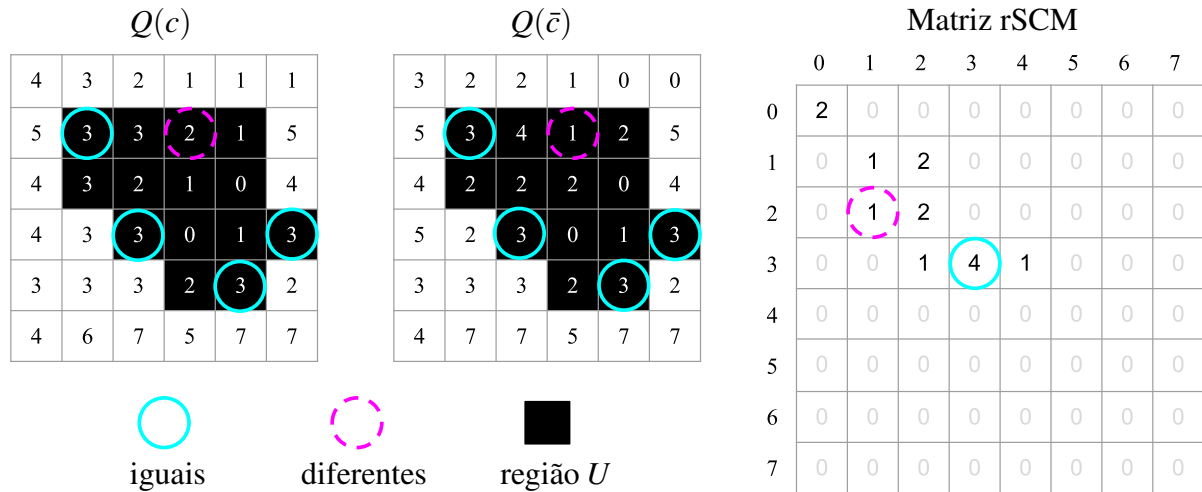


Figura 19 – Exemplo da Matriz rSCM para imagens de entrada quantizadas em 8 níveis. As coocorrências são computadas apenas na região  $U$ . Fonte: elaboração própria.

de  $c$  e seu correspondente em  $\bar{c}$  quantizados devem ser para serem contadas. A Figura 18 ilustra a computação da rSCM, a qual é exemplificada para imagens de entrada quantizadas em 8 níveis na Figura 19. Observe que a rSCM contém as contagens do par  $(i, j)$  pertencentes à região  $U$  sinalizada em negrito.

A matriz rSCM captura as alterações estruturais oriundas de uma função de trans-

formação<sup>11</sup> aplicada a uma imagem  $c$ . É importante ressaltar que a imagem transformada  $\bar{c}$  deve preservar o domínio de  $c$ . As diferenças estruturais entre  $c$  e  $\bar{c}$  são compactadas na matriz e podem ser analisadas com base em seus quadrantes (Figura 20). Quando a transformação é insuficiente para causar alteração na imagem, toda informação fica retida na diagonal principal da matriz, sinalizando que  $c = \bar{c}$ . A rSCM apresenta ainda invariância quanto a transformações que resultem em rotações de  $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$  e seus elementos são inteiros positivos.

Segundo Quadrante (2°Q) Região Escura	Primeiro Quadrante (1°Q) Transições Escuro - Claro
Terceiro Quadrante (3°Q) Transições Claro - Escuro	Quarto Quadrante (4°Q) Região Clara

Figura 20 – Interpretação da Matriz rSCM. Fonte: elaboração própria.

#### 6.1.2.1.1 Matriz rSCM aplicada a Imagens de Células Cervicais

Esta tese propõe a matriz rSCM para avaliar padrões de textura na região do núcleo de imagens de células cervicais. Uma vez que a distribuição irregular de cromatina é um potencial indicador de anormalidade (WATANABE *et al.*, 2004), estamos interessados em capturar diferenças estruturais no núcleo entre  $c$  e  $\bar{c}$ , utilizando como função de transformação o filtro passa-baixa gaussiano ( $\sigma = 4$ )<sup>12</sup>. Distribuição irregular de cromatina gera imagens do núcleo celular com aspecto granular, ou seja, com padrões de textura heterogênea. Tal situação contrasta com casos normais, onde o núcleo tem aparência mais homogênea. Como filtros passa-baixa tem efeito relevante em sinais de textura heterogênea, a matriz rSCM captura informações que podem discriminar células anormais de normais. A Figura 21 ilustra o papel do filtro passa-baixa adotado, destacando as diferenças estruturais (área cinza), a ser capturada pela matriz proposta. Para melhor visualização, essa diferença está quantificada pela métrica Erro Médio Absoluto (EMA)<sup>13</sup> (BOTCHKAREV, 2018), mostrando que as estruturas tendem a sofrer maior degradação em casos anormais.

<sup>11</sup> Esta tese utiliza um filtro passa-baixa gaussiano para gerar degradação na imagem  $c$ . Detalhes na Seção 6.1.2.1.1.

<sup>12</sup> O Apêndice B descreve os experimentos realizados para alcançar esse parâmetro.

<sup>13</sup> Erro Médio Absoluto. Dado por  $EMA = \frac{1}{N_p} \sum_{i=1}^{N_p} |p_{c_i} - p_{\bar{c}_i}|$ , onde  $p_{c_i}$  e  $p_{\bar{c}_i}$  correspondem à intensidade do  $i$ -ésimo pixel da imagem  $c$  e  $\bar{c}$ , respectivamente.  $N_p$  é o número de pixels.

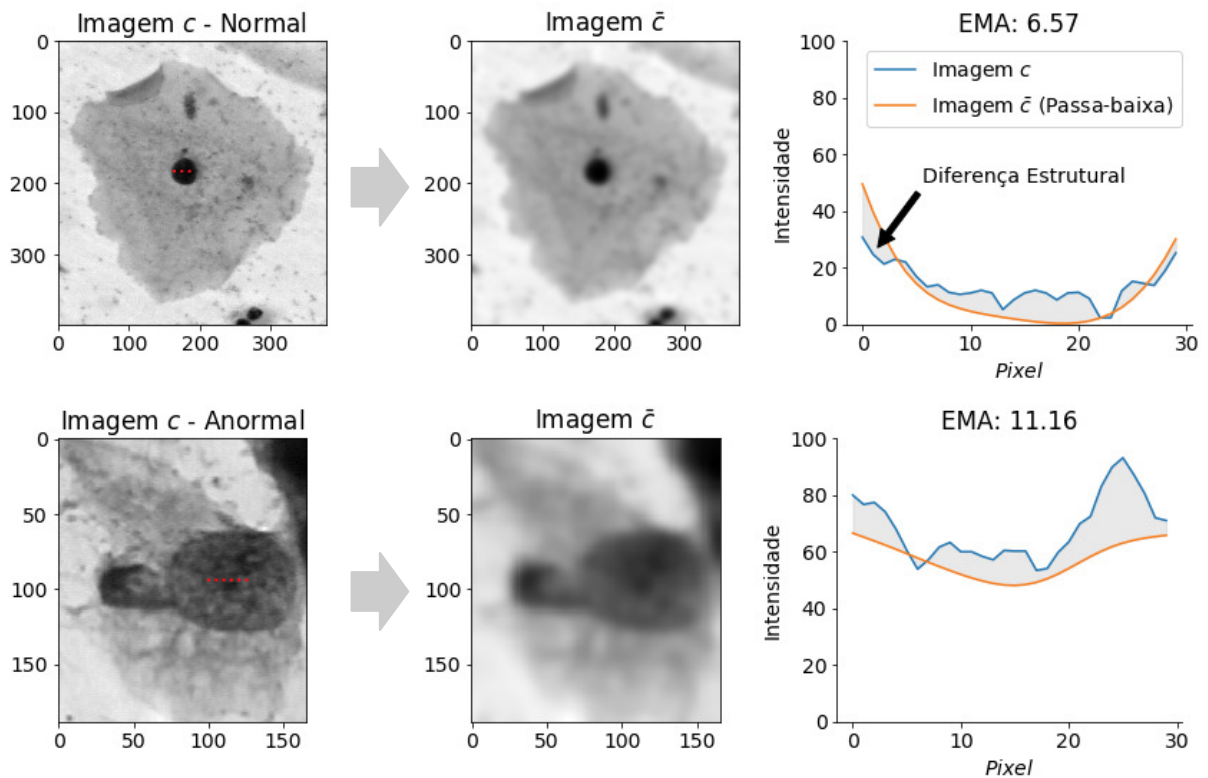


Figura 21 – Alterações estruturais em célula cervical normal e anormal que são capturadas pela rSCM. Os gráficos à direita destacam o impacto em uma amostra de textura do núcleo (linha tracejada vermelha) resultante da aplicação do filtro gaussiano com  $\sigma = 4$ . O EMA entre as linhas quantifica as diferenças estruturais (área em cinza), apontando maior sensibilidade à filtragem passa-baixa para os casos anormais. Fonte: elaborada pelo autor. Imagens obtidas da base Herlev (JANTZEN *et al.*, 2005).

Se a região  $U$  a ser processada corresponde à área do núcleo celular, a matriz rSCM captura três informações principais para caracterização de células cervicais: a) intensidade: dada pelo valor dos *pixels*; b) textura: dada pela degradação de estruturas pela filtragem passa-baixa e c) tamanho do núcleo: dado pela energia<sup>14</sup> da matriz. Essas características habilitam a matriz proposta a convergir para padrões que indicam alteração celular. A Figura 22 ilustra como essas informações podem ser analisadas, seguindo de detalhes do comportamento esperado da rSCM quando presença de anormalidade no núcleo das células.

- a) **Intensidade:** núcleos são geralmente estruturas escuras, cujas intensidades dos *pixels* se aproximam de zero. Como os núcleos normais se apresentam comumente como estruturas homogêneas, matrizes rSCM que possuem coeficientes não nulos concentrados no segundo quadrante, mas espalhados em direção aos demais indicam potencial anormalidade.
- b) **Textura:** se  $c = \bar{c}$  dentro da região  $U$  em análise, então temos uma matriz rSCM diagonal. Uma vez que núcleos normais possuem geralmente padrões de textura homogêneos, a

<sup>14</sup> Assumimos **energia** como o valor resultante do somatório de todos os coeficientes da matriz.



filtragem passa-baixa tende a degradar menos. Logo, matrizes rSCM cujos coeficientes não nulos se dispersam da diagonal principal são indicativas de anormalidade.

- c) **Tamanho do núcleo:** é bem conhecido que o tamanho do núcleo é uma característica relevante para identificar células cervicais doentes (WHO; IARC, 2004; BENGTTSSON; MALM, 2014; WESOŁA *et al.*, 2014). Núcleos anormais são geralmente maiores que os normais, salvo algumas circunstâncias<sup>15</sup>. Dessa forma, matrizes rSCM com maior energia tendem a indicar lesão celular, uma vez que a quantidade de *pixels* avaliados é maior.

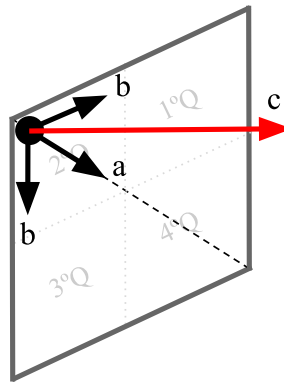


Figura 22 – Espalhamento dos coeficientes não nulos da rSCM quando anormalidade dos núcleos de células cervicais. a) Indicativo de alteração na intensidade dos *pixels*. b) Indicativo de alteração na textura e c) no tamanho do núcleo. A seta vermelha sinaliza amplitude do ponto  $(i, j)$  na matriz proposta. Fonte: elaboração própria.

A modelagem da rSCM está exemplificada para as 917 imagens da base Herlev (JANTZEN *et al.*, 2005) (Seção B.1). A projeção *t-distributed Stochastic Neighbor Embedding* (t-SNE)<sup>16</sup> (MAATEN; HINTON, 2008) com duas componentes principais (PCA) (Figura 23) demonstra o comportamento da matriz em casos normais e anormais, considerando a região *U* analisada como a área do núcleo. Podemos analisar as matrizes obtidas: *i*) quanto à intensidade: observação da distribuição dos coeficientes não nulos entre os quadrantes da matriz; *ii*) quanto à textura: por meio da medida AP, a qual é definida pela quantidade de coeficientes não nulos ponderados por sua distância à diagonal principal. Essa medida corresponde à dispersão dos coeficientes não nulos em relação à diagonal principal e *iii*) quanto ao tamanho do núcleo: computação da energia. Os gráficos da Figura 24 quantificam a análise de *ii* e *iii* para as amostras da base Herlev, evidenciando a existência de padrões distintos na matriz para os casos normais e anormais. É importante ressaltar que a energia e a medida AP são apenas para ilustrar o

<sup>15</sup> O tamanho do núcleo varia de acordo com o tipo de célula (por exemplo, escamosas, glandulares) e sua maturidade (WHO; IARC, 2004; WESOŁA *et al.*, 2014). Dessa forma, a análise apenas do tamanho do núcleo tende a ser insuficiente para confirmar anormalidade.

<sup>16</sup> Computado com parâmetros em valores padrão, exceção para *perplexity*=40. Erro após 1000 iterações: 0,74.

comportamento da matriz proposta. Nesta tese, a vetorização da matriz rSCM é usada como vetor de características.

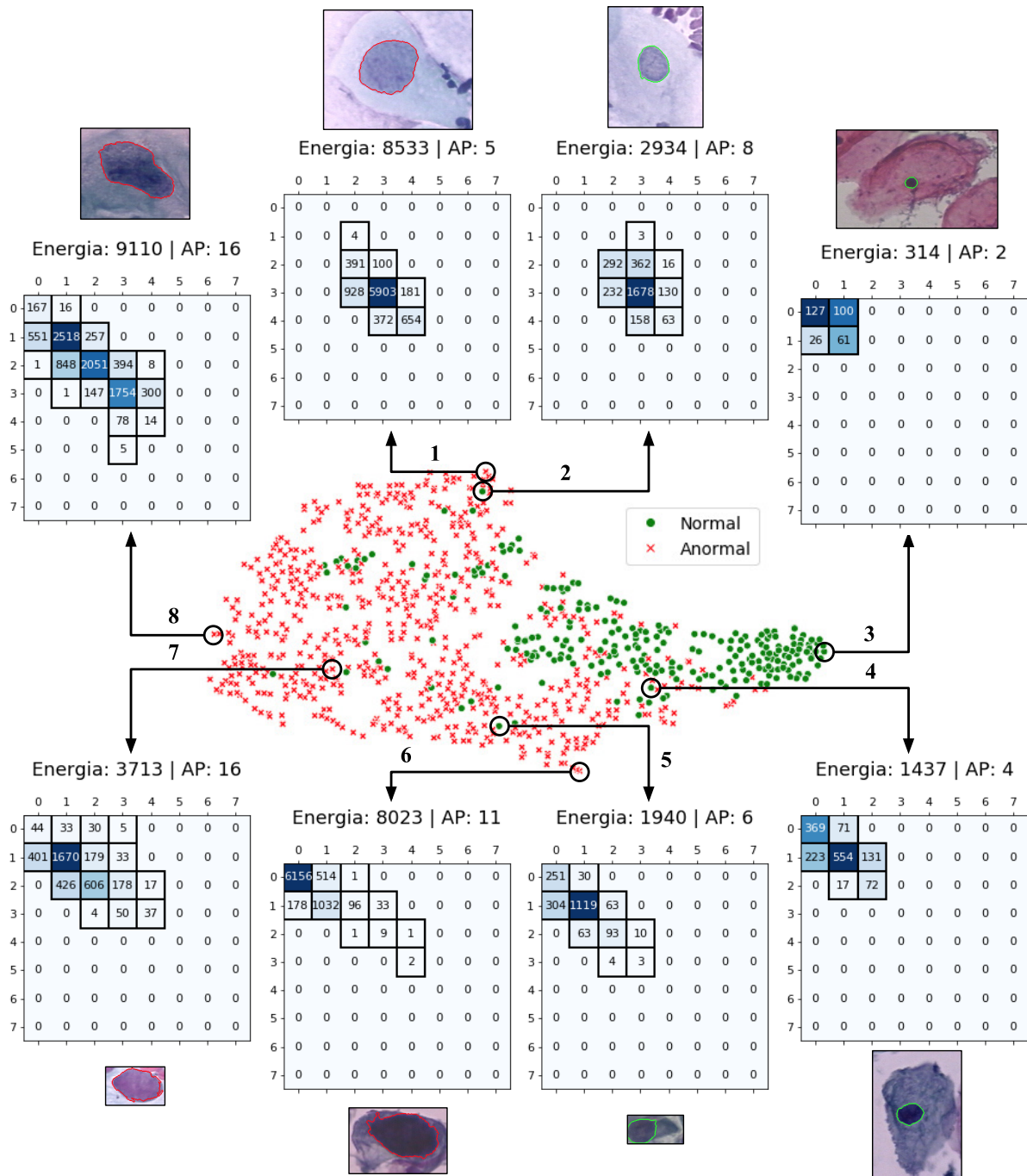


Figura 23 – Projeção das matrizes rSCM no espaço t-SNE com duas componentes principais (PCA). As amostras selecionadas ilustram os extremos da nuvem de pontos. De acordo com os rótulos de classes disponíveis na base Herlev temos: 1) displasia leve, 2) normal intermediária, 3) normal superficial, 4) normal intermediária, 5) normal colunar, 6) displasia moderada, 7) carcinoma *in situ* e 8) displasia leve. As marcações nas imagens sinalizam as bordas dos núcleos, conforme o GT, considerando verde para casos normais e vermelho para anormais. Fonte: elaboração própria.

Experimentalmente, adotamos  $Q$  como função de quantização, com 3 *bits* de resolução (8 níveis de quantização), resultando em um vetor de características com 64 elementos. Quantidades superiores a 8 níveis de quantização tornam a matriz mais sensível a variações de intensidade na textura do núcleo normal, levando-a a capturar mais diferenças estruturais e, conseqüentemente, aproximar a matriz desses casos às observadas para núcleos alterados. Em outras palavras, a dispersão dos elementos não nulos fora da diagonal principal, nos casos normais, tende a se aproximar do esperado para os casos com alteração, resultando em perda de discriminância. Adotamos a Propriedade  $P_U$  como uma função que minimiza a diferença absoluta entre o par  $(i, j)$  na posição  $p$ , ou seja,  $P_U$  equivalente a  $|Q(c)_p - Q(\bar{c})_p| < N_Q$ , onde  $N_Q$  é a quantidade de partições retornada pela função de quantização  $Q$ . Escolhemos  $N_Q$  como critério na Propriedade  $P_U$  para tornar a matriz sensível às transições de todos os níveis quantizados. A matriz rSCM é calculada para imagens em níveis de cinza (256 níveis) convertidas pela equação  $Y = 0,2125R + 0,7154G + 0,0721B$ , onde  $R$ ,  $G$  e  $B$  representam os canais de cor da imagem colorida.

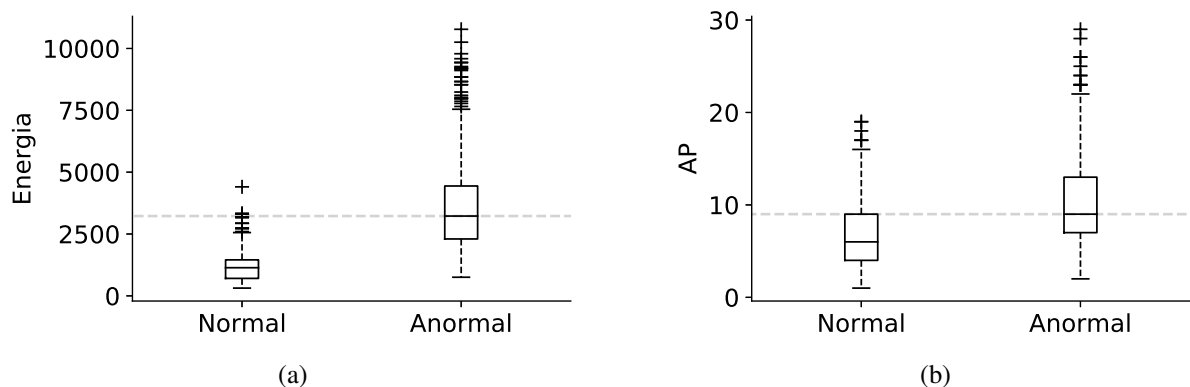


Figura 24 – Comportamento da matriz rSCM em relação ao a) tamanho do núcleo (energia) e b) textura (AP). As linhas tracejadas destacam a mediana para os casos anormais. Medições computadas a partir das 917 imagens disponíveis na base Herlev. Fonte: elaboração própria.

A matriz rSCM corresponde à Contribuição 3 desta tese. Como parte dessa contribuição, o Apêndice B apresenta experimentos para avaliar a matriz rSCM<sup>17</sup> como descritor de células cervicais em problemas de classificação binária: normal e anormal. O Apêndice B discute esses resultados, bem como a metodologia adotada para estimação de parâmetros.

<sup>17</sup> A matriz rSCM pode ser estendida para um descritor de textura de propósito geral para sinais  $n$ -dimensionais, chamada genericamente de *Structural Co-occurrence Matrix* (SCM). Para outros detalhes sobre nosso descritor, bem como suas aplicações, recomendamos a leitura do artigo “*Rotation-invariant Feature Extraction using a Structural Co-occurrence Matrix*”, disponível em <<https://doi.org/10.1016/j.measurement.2016.08.012>>.

### 6.1.2.2 Refinamento dos Mapas de Saliência

Propomos nesta tese um algoritmo para identificar potenciais estruturas de interesse em cada zona de atenção  $z$  (Figura 15) e codificá-las em um mapa de saliência (local). Esse mapa local traz um refinamento do mapa de saliência obtido globalmente, sendo, portanto, endereçado para posicionar as ROIs. A Figura 25 apresenta o algoritmo de refinamento.

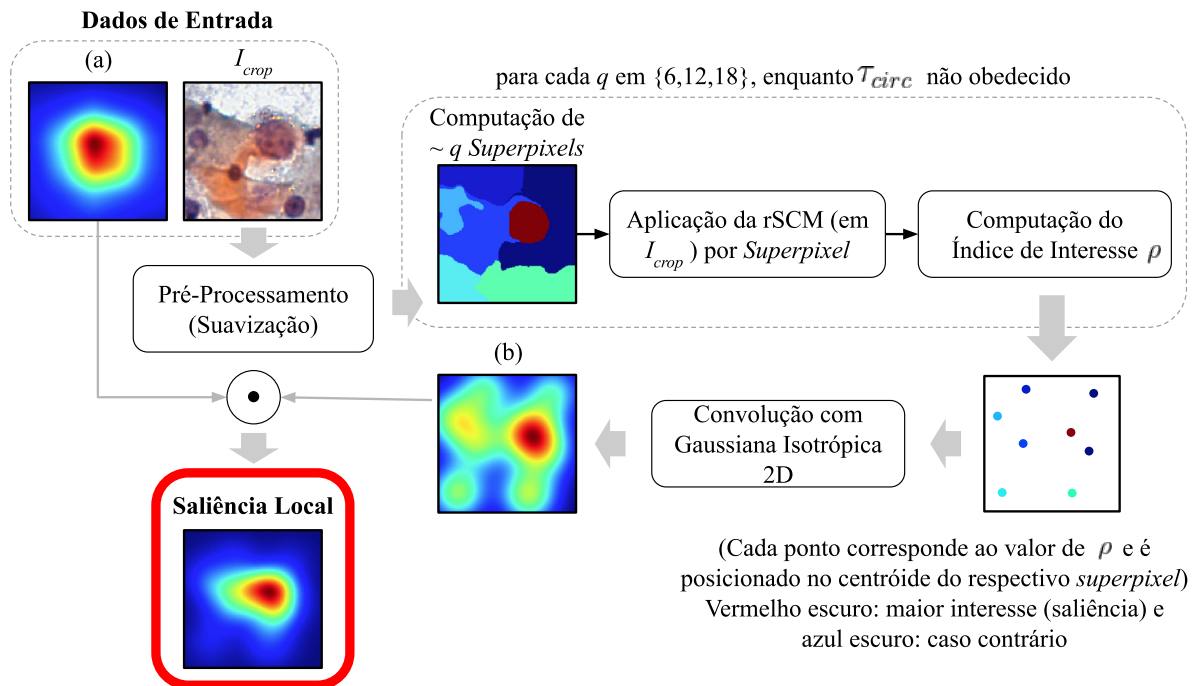


Figura 25 – Estimação da saliência local. Aplicação do algoritmo rSCM para identificar estruturas de interesse. a) Recorte do mapa de saliência estimado globalmente para a imagem de entrada  $I$  que equivale ao recorte  $I_{crop}$ . b) Mapa obtido a partir de  $g(\cdot)$  (Figura 15), enfatizando áreas conforme o índice de interesse  $\rho$ . Modulando (a) com (b), alcançamos o mapa de saliência do recorte. Fonte: elaboração própria. A imagem que representa  $I_{crop}$  corresponde a uma amostra da base de dados CRIC.

Utilizamos um algoritmo baseado em *superpixels* para definir diferentes regiões na imagem de entrada ( $I_{crop}$ ). A computação de *Superpixels*<sup>18</sup> aglutina *pixels* da imagem segundo um critério de similaridade na intensidade dos *pixels*, formando estruturas que, por vezes, definem objetos (ACHANTA *et al.*, 2012). *Superpixels* têm sido empregados na segmentação de células cervicais em (RAMALHO *et al.*, 2015; LEE; KIM, 2016), conseguindo resultados promissores em bases de dados de *benchmark*. O refinamento proposto nesta tese utiliza *superpixels* para alcançar estruturas de baixo nível na imagem, para as quais é inferida algum nível de saliência. Para evitar a formação de aglutinados referentes a ruídos, realizamos um pré-processamento

<sup>18</sup> Implementamos *superpixels* a partir do algoritmo SLIC proposto por Achanta *et al.* (2012).

com o algoritmo *Total Variation* (TV) (RUDIN *et al.*, 1992). O algoritmo TV suaviza a imagem, minimizando o impacto sobre as bordas de objetos. Essa preservação de bordas permite uma melhor definição das estruturas da imagem, incluindo o núcleo.

Assumimos cada *superpixel* como uma máscara para uma região  $U_i$  na imagem fornecida. Para  $U_i$ , é computado um índice de interesse  $\rho_i$ , o qual definimos pela combinação linear entre a circularidade do  $i$ -ésimo *superpixel* ( $U_{circ_i} \in [0, 1]$ ) e a probabilidade da região correspondente ser anormal ( $U_{prob_i} \in [0, 1]$ ). Nesta tese,  $U_{prob}$  é obtido por meio de um classificador previamente treinado, o qual processa as características extraídas pela rSCM para região em análise. Formalmente, temos:  $\rho_i = 0,5 \times U_{circ_i} + 0,5 \times U_{prob_i}$ , onde  $\rho_i \in [0, 1]$ . Esse índice tem sua sustentação na morfologia ovóide do núcleo e a na relevância da análise de textura nuclear como atributo clínico (Apêndice B).

A quantidade de *superpixels* é aproximadamente  $q$ , com  $q \in \{6, 12, 18\}$ . A busca pela quantidade de regiões a analisar ocorre iterativamente, da menor para maior quantidade, até que a região de maior interesse  $\rho$  tenha  $U_{circ} > \tau_{circ} = 0,75$ . Se nenhuma região atender à condição,  $q = 18$  é assumido. Uma matriz de pontos, posicionados no centroide dos *superpixels* e com amplitude correspondente ao seu valor de  $\rho$ , é criada e posteriormente convoluída por uma gaussiana isotrópica bidimensional. O mapa resultante modula, então, o recorte correspondente em  $f(I, W)$  (Figura 15) e o mapa de saliência local é gerado.

No algoritmo de refinamento proposto, a matriz rSCM extraída na região de cada *superpixel* tem tamanho  $8 \times 8$ , resultando em um vetor de característica de 64 elementos (vetorização da matriz). A Propriedade  $P_U$  foi igual para todas as regiões, seguindo formulação descrita na Seção 6.1.2.1.1. Não houve melhorias no desempenho do método proposto com vetor de características de maior tamanho, obtidos a partir de matrizes com maior dimensão espacial.

## 6.2 Módulo 2: Classificação das Regiões de Interesse

Diferentemente do Módulo 1 que processa a imagem por completo, o Módulo 2 recebe as ROIs selecionadas como entrada e as processa isoladamente. Para isso, esta tese utiliza uma CNN com base na arquitetura VGG-16. Substituímos as camadas do último bloco, ou seja, as camadas totalmente conectadas, por uma com 1024 neurônios e uma outra com 3 neurônios para computar a probabilidade de classificação para cada classe utilizando regressão *softmax*. Aplicamos *dropout* (HINTON *et al.*, 2012) antes da primeira camada ( $rate=0,25$ ) e entre as duas ( $rate=0,50$ ) para reduzir o *overfitting*. A Figura 26 apresenta a arquitetura VGG-16 modificada

adotada nesta tese.



Figura 26 – Arquitetura VGG-16 modificada proposta como Módulo 2. Fonte: elaboração própria.

O Módulo 2 recebe ROIs com tamanho de  $100 \times 100$  *pixels*. Essa dimensão de entrada permite a representação de imagens com núcleos normais e anormais em todas as patologias estudadas, na resolução de  $0,255 \mu\text{m}/\text{pixel}$ . O núcleo é utilizado como referência para determinar o tamanho dos recortes por essa ser a principal estrutura na categorização de células cervicais (WESOŁA *et al.*, 2014; SILVA *et al.*, 2019). Recortes com tamanhos maiores adicionam artefatos de fundo nos recortes de células normais superficiais, aumentando sua dispersão intra-classe. Tamanhos menores comprometem a representação do núcleo em toda sua extensão, principalmente para os casos HSIL, para o qual Wesola *et al.* (2014) mostraram ter o maior tamanho de núcleo. As dimensões dos filtros do estágio convolucional foram mantidas conforme definido na arquitetura padrão da VGG-16. Realizamos transferência de aprendizado (*transfer-learning*) a partir da base de imagens *ImageNet* e executamos treinamento de acordo com o seguinte procedimento: 1) inicialização dos pesos da CNN com aqueles treinados na *ImageNet*; 2) congelamento (*freezing*)<sup>19</sup> das camadas convolucionais e treinamento das camadas totalmente conectadas (FC) por 30 épocas (Etapa A); 3) ajuste fino (*fine-tuning*) das últimas 4 camadas convolucionais e das FC utilizando estratégia de parada precoce (*early stopping*) (Etapa B). Empregamos o *Stochastic Gradient Descent* (SGD) (ZHANG *et al.*, 2017) em *mini-batch* de 128 para computar os gradientes da função custo<sup>20</sup> e interativamente ajustar os pesos durante o treinamento. Utilizamos taxa de aprendizado<sup>21</sup> (*learning rate*) de 0,0001 e aplicamos o momento de *Nesterov*<sup>22</sup> (SUTSKEVER *et al.*, 2013) em ambos treinamentos (etapas A e B).

<sup>19</sup> Processo que mantém constante os pesos de camadas CNN durante o treinamento.

<sup>20</sup> Empregamos *categorical crossentropy* como função custo (KETKAR, 2017).

<sup>21</sup> Hiperparâmetro que controla o quanto os pesos estão sendo ajustados na rede em relação ao gradiente de perda. Quanto menor o valor, maior é a chance da rede encontrar um mínimo local adequado. No entanto, pode significar que levará muito tempo para convergir. É um importante hiperparâmetro a ser ajustado, fundamental para alcançar um modelo CNN com desempenho satisfatório para o problema.

<sup>22</sup> Em linhas gerais, o momento e o momento de Nesterov são pequenas variações no gradiente que podem acelerar o treinamento e melhorar significativamente a convergência. Detalhes em (SUTSKEVER *et al.*, 2013).

O treinamento é encerrado por parada precoce<sup>23</sup> (*early stopping*) com 5 épocas de observação. Então, selecionamos o modelo que contém o espaço de hiper-parâmetros com maior Acurácia de validação ao longo das épocas. O Módulo 2 foi treinado para reconhecer amostras em três classes distintas: normal, baixo risco e alto risco (Tabela 1). O procedimento descrito foi alcançado experimentalmente.

A Figura 27 ilustra a criação das bases de imagens para treinamento e validação do Módulo 2. Essas bases contêm recortes de células, conforme GT do CRIC, oriundos das imagens empregadas para treinar o Módulo 1. Recortes de imagens da base CRIC que não foram empregadas em CRICVA foram selecionadas aleatoriamente e complementam esses conjuntos. Foram separadas 75% das imagens por classe rotulada. As imagens remanescentes foram utilizadas posteriormente para validar o modelo AGMic (Seção 7.2).

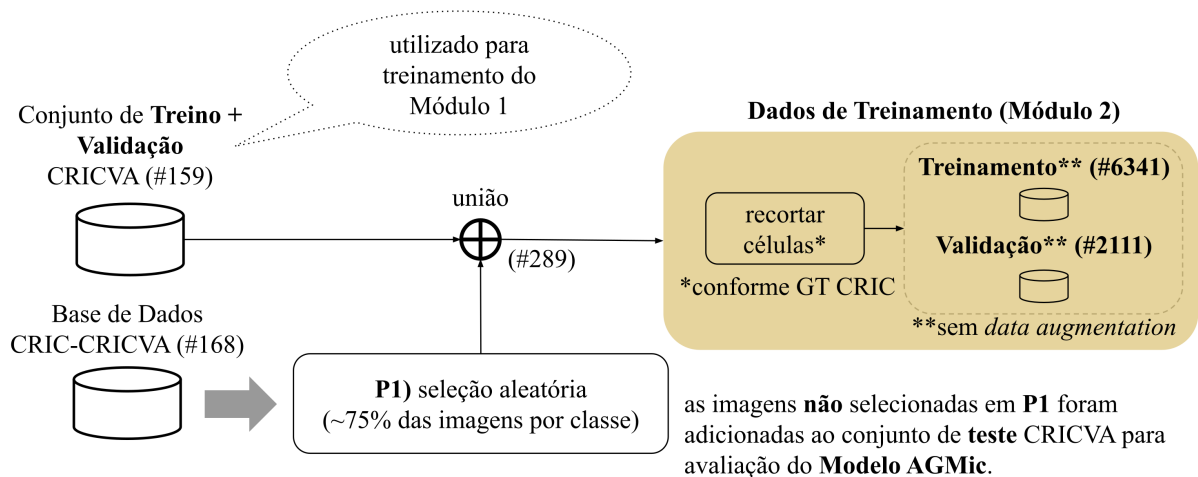


Figura 27 – Seleção de dados para treinamento do Módulo 2. Fonte: elaboração própria.

<sup>23</sup> Técnica que permite encerrar o treinamento quando a precisão parar de melhorar no conjunto de validação. Uma vez que quantidade excessiva de épocas pode levar o modelo ao sobreajuste do conjunto de dados (*overfitting*) de treinamento e quantidade insuficiente pode resultar em um modelo subajustado (*underfitting*), a parada precoce é útil por evitar a determinação explícita da quantidade de épocas de treinamento. Assim, o modelo encontra uma condição, dentro de algumas épocas de observação, que pode representar a melhor generalização para o problema.

Em problemas de classificação, *data augmentation* é uma técnica conhecida por melhorar o desempenho da CNN, reduzindo o *overfitting*. Essa técnica consiste em ampliar o conjunto de dados disponível a partir da geração de novos exemplares com base nos existentes. No caso de imagens de células cervicais, várias operações podem ser realizadas nos dados reais, incluindo rotações, translações, contaminação por ruído, dentre outras. A partir de um conjunto mais diversificado de amostras, é possível alcançar um modelo com melhor capacidade de generalização.

A Figura 28 ilustra a estratégia adotada nesta tese para *data augmentation*. Inspirados por Zhang *et al.* (2017), realizamos 10 rotações ( $\theta = 36$ ) e 9 translações (10 *pixels*), resultando em 90 versões<sup>24</sup> para cada recorte de célula de nosso conjunto de treinamento. Dessa forma, alcançamos 721.710 recortes para treino e validação.

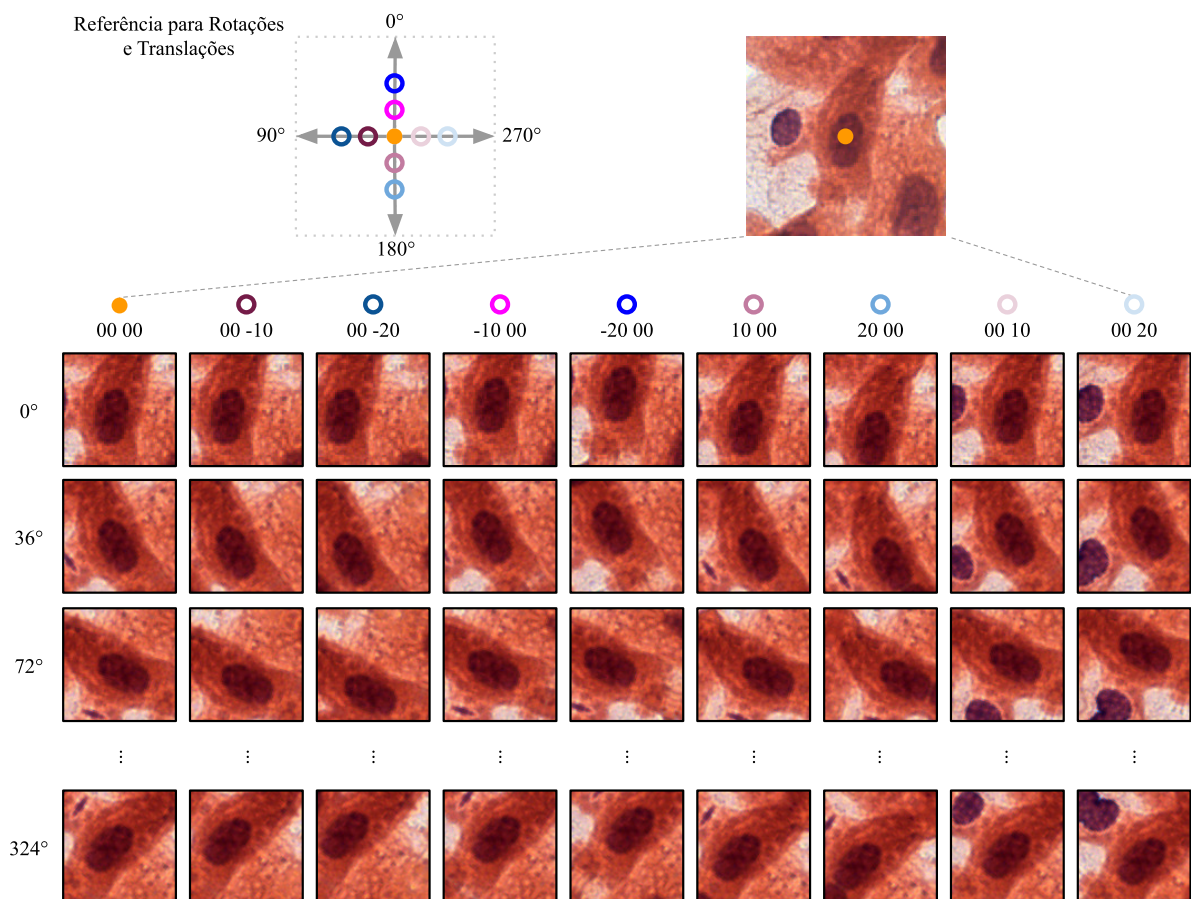


Figura 28 – Estratégia de *data augmentation* adotada para treinamento do Módulo 2 do AGMic. Fonte: elaboração própria a partir de recortes de uma imagem da Base CRIC.

<sup>24</sup> Nas células próximas às bordas da imagem, as translações eventualmente podem levar a regiões fora dos limites da imagem. Nesses casos, as versões derivadas foram descartadas.



A Figura 29 apresenta as curvas de Acurácia observadas ao longo das épocas de treinamento em ambas etapas A e B. A etapa B convergiu com 20 épocas. Os algoritmos foram desenvolvidos em *Python* utilizando o *Keras* (versão: 2.1.5)<sup>25</sup>.

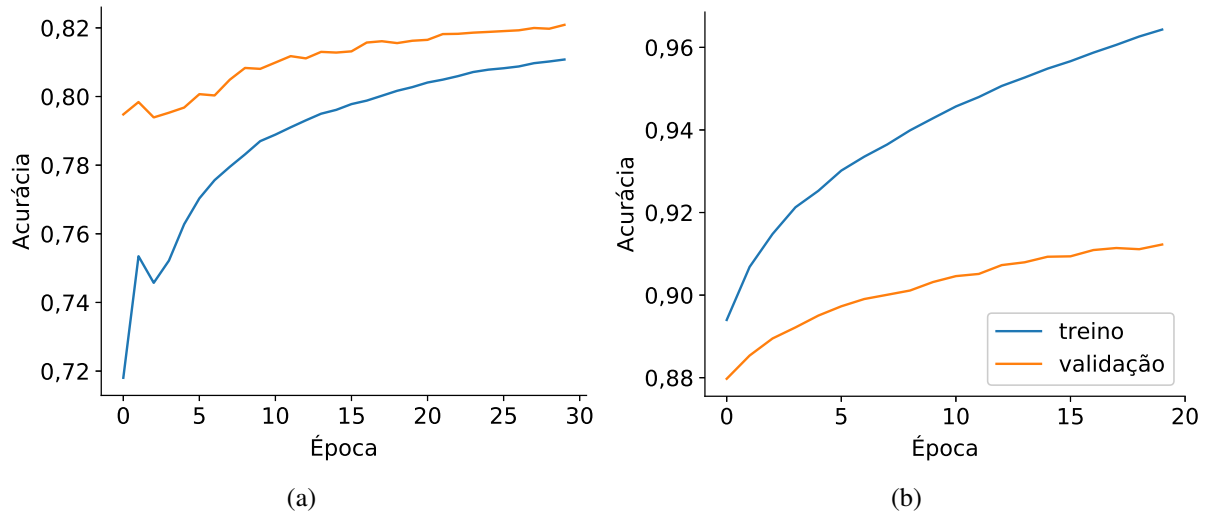


Figura 29 – Acurácia ao longo das épocas de treinamento. a) Treino da Etapa A e b) treino da Etapa B. Fonte: elaboração própria.

<sup>25</sup> Keras é uma biblioteca de código aberto escrita em linguagem *Python* para desenvolvimento de algoritmos com redes neurais. Detalhes em <<https://keras.io/>>

## 7 RESULTADOS EXPERIMENTAIS

Este capítulo apresenta os resultados obtidos com os modelos desenvolvidos nesta tese. Em cada seção, descrevemos as métricas utilizadas, a seleção dos dados para a validação e a metodologia de avaliação empregada.

### 7.1 Experimentos de Predição de Saliência

A Figura 30 ilustra a metodologia proposta para avaliar a predição de saliência nas imagens de células cervicais. Particularmente, estamos interessados em confirmar duas hipóteses:

- **Hipótese 1 (H1):** Métodos *bottom-up* apresentam mapas de saliência comparáveis aos preditos pela atenção visual de especialistas humanos.
- **Hipótese 2 (H2):** Modelos de predição de saliência podem detectar regiões clinicamente relevantes em imagens digitais de exames de Papanicolaou.

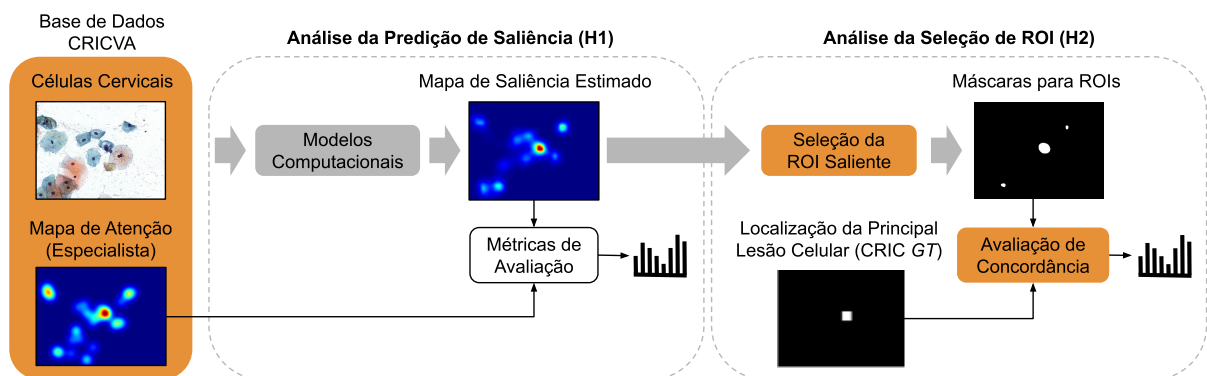


Figura 30 – Etapas para investigação das hipóteses H1 e H2. 1) Avaliação dos modelos de predição de saliência utilizando métricas estado-da-arte (testar H1) e 2) análise da seleção de Regiões de Interesse (ROIs) para determinar se a alteração que define o rótulo da imagem está destacada pelos mapas de saliência (testar H2). Fonte: elaboração própria.

A investigação das hipóteses H1 e H2 alcança a Contribuição 4 desta tese, por meio da: 1) avaliação quantitativa de modelos de predição de saliência estado-da-arte na estimação da atenção visual seletiva de citopatologistas em imagens extraídas de exames de Papanicolaou convencional. Além disso, reportamos métodos que podem detectar áreas de interesse clínico nesse tipo de exame usando apenas características *bottom-up*; 2) verificação sob uma nova abordagem dos achados de Coombes e Culverhouse (2003) sobre a alta correlação entre características de baixo nível das células cervicais, como cor e brilho, e o foco de atenção dos citopatologistas durante a busca por lesões. Adicionalmente também mostramos que fatores

*top-down* têm influência na forma que citopatologistas observam as imagens durante a inspeção e 3) análise quantitativa de como as lesões em células cervicais são detectadas por algoritmos de predição de saliência quando aplicados como extrator de ROIs.

### 7.1.1 Modelos Avaliados

Investigamos 10 métodos *bottom-up*<sup>1</sup> e duas arquiteturas CNN em um algoritmo de predição de aprendizagem profunda, abrangendo, portanto, diferentes categorias de métodos (BORJI; ITTI, 2013). Também consideramos o modelo proposto por Zhang *et al.* (2013), pois este foi projetado para explorar características *top-down* em imagens de células cervicais. Resumimos cada um desses algoritmos na Tabela 2 e introduzimos as respectivas abreviaturas, conforme abordamos no restante da tese. Todos os algoritmos, com exceção do proposto por Zhang *et al.* (2013), foram validados em bases públicas de dados e escolhidos para nossos experimentos de acordo com os seguintes critérios: 1) receber uma única imagem como entrada; 2) código fonte está publicamente disponível; 3) apresentar reconhecido desempenho em *rankings* públicos de algoritmos de predição de saliência; 4) tempo de processamento reportado ser inferior a 3 segundos por imagem e 5) corresponder ao estado-da-arte ou ser um modelo de referência na literatura. Com base nesses requisitos, visitamos o *benchmark* de predição de saliência do *Massachusetts Institute of Technology* (MIT) (BYLINSKII *et al.*, 2015). Nele, ranqueamos os modelos pela métrica *Normalized Scanpath Saliency* (NSS)<sup>2</sup> (PETERS *et al.*, 2005) e selecionamos o algoritmo CNN de melhor desempenho em imagens naturais: *Saliency Attentive Model* (SAM)<sub>v</sub> e SAM<sub>r</sub> (CORNIA *et al.*, 2018), nos quais os mapas de características são extraídos a partir das arquiteturas VGG-16 e ResNet-50, respectivamente. Coletamos também 5 algoritmos não baseados em CNN: BMS (ZHANG; SCLAROFF, 2013), LDS (FANG *et al.*, 2017), FES (TAVAKOLI *et al.*, 2011), SWD (DUAN *et al.*, 2011) e UHM (TAVAKOLI; LAAKSONEN, 2016). Complementamos nosso estudo a partir do extensivo *benchmark* introduzido por Borji *et al.* (2015), selecionando os modelos *bottom-up* IT (ITTI *et al.*, 1998), GB (HAREL *et al.*, 2007), SR (HOU; ZHANG, 2007), SS (HOU *et al.*, 2012) e SIM (MURRAY *et al.*, 2011). Os parâmetros utilizados em cada algoritmo são aqueles reportados nos artigos originais.

<sup>1</sup> Não contém informação sobre o contexto de análise de células cervicais.

<sup>2</sup> Conforme discutiremos na Seção 7.1.3, a métrica NSS considera falsos positivos e negativos igualmente, além de não depender de transformação parametrizada dos dados. Dessa forma, é recomendada pelo *benchmark* do MIT (BYLINSKII *et al.*, 2015) como referência para ranqueamento dos modelos de predição de saliência.

Tabela 2 – Modelos de predição de saliência. Fonte: elaboração própria.

#	Abreviação		Descrição	Ano	Benchmark
	Literatura	Tese			
1	SAMv ( <i>Saliency Attentive Model - VGG-16</i> ) (CORNIA <i>et al.</i> , 2018)	SVS, SVC	Usa a VGG-16 e incorpora a <i>Attentive Convolutional Long Short-Term Memory network (Attentive ConvLSTM)</i> para predição de saliência em imagens, sem sequência temporal.	2018	MIT (BY-LINSKII <i>et al.</i> , 2015)
2	SAMr (SAM - ResNet-50) (CORNIA <i>et al.</i> , 2018)	SRS, SRC	A mesma abordagem do modelo SAMv, mas usando a ResNet-50.	2018	
3	BMS ( <i>Boolean Map based Saliency</i> ) (ZHANG; SCLAROFF, 2013)	BMS	Caracteriza uma imagem de entrada por um conjunto de imagens binárias, as quais foram criadas a partir de limiarizações aleatórias no espaço de características. Esse modelo realiza uma análise dos mapas binários construídos para encontrar regiões vizinhas e estimar o mapa de saliência.	2013	
4	LDS ( <i>Learning Discriminative Subspaces</i> ) (FANG <i>et al.</i> , 2017)	LDS	Este algoritmo baseia-se no aprendizado de um conjunto de subespaços discriminativos. Estes elementos são otimizados para destacar possíveis objetos na cena e suprimir artefatos. O LDS emprega análise de componentes principais para criar os subespaços candidatos.	2017	
5	FES ( <i>Fast and Efficient Saliency</i> ) (TAVAKOLI <i>et al.</i> , 2011)	FES	Usa a estratégia <i>center-surround</i> <sup>3</sup> para estimar saliência a partir de características locais em uma estrutura bayesiana. As distribuições de probabilidades são estimadas por meio de <i>sparse sampling</i> e o <i>kernel density estimation</i> .	2011	
6	SWD ( <i>Spatially Weighted Dissimilarity Saliency</i> ) (DUAN <i>et al.</i> , 2011)	SWD	Este algoritmo baseia-se na combinação de dissimilaridades e na distância espacial entre <i>patches</i> da imagem e a informação de centro. A distância espacial pondera as dissimilaridades correspondentes. A análise de componentes principais é adotada para redução de dimensionalidade.	2011	
7	UHM ( <i>Unsupervised Hierarchical Models</i> ) (TAVAKOLI; LAAKSONEN, 2016)	UHM	É um modelo hierárquico, multiescala e não supervisionado, que explora conceitos de saliência local e global. Utiliza <i>Independent Subspace Analysis (ISA)</i> , que equivale a uma rede neural de duas camadas. Este algoritmo obtém uma representação hierárquica da entrada, empilhando as redes ISA juntas, como é feito em modelos <i>deep learning</i> .	2016	

Continua na próxima página

<sup>3</sup> Paradigma em que a hipótese é que existe uma janela local dividida em um centro e um entorno, e o centro contém um objeto.

Tabela 2 – Continuação da página anterior

#	Abreviação		Descrição	Ano	Benchmark
	Literatura	Tese			
8	ITTI ( <i>Itti's Saliency Model</i> ) (ITTI <i>et al.</i> , 1998) - implementação por Harel <i>et al.</i> (2007)	IT	É o pioneiro em predição de saliência. Esse algoritmo extrai características de baixo nível usando diferenças locais, em uma abordagem <i>center-surround</i> , de intensidade, cor e orientação em várias escalas espaciais. Em seguida, a fusão entre as diferentes escalas e a normalização destes mapas produz três mapas de saliência que são combinados para gerar o mapa final.	1998	(BORJI <i>et al.</i> , 2015)
9	GBVS ( <i>Graph-Based Visual Saliency</i> ) (HAREL <i>et al.</i> , 2007)	GB	Extraí características de baixo nível de forma similar ao algoritmo IT. GB usa uma cadeia de Markov para construir um grafo totalmente conectado. Neste grafo, os nós estão relacionados com os mapa de características. O peso entre dois nós é definido pela dissimilaridade entre os mapas de características e sua distância espacial. O mapa de saliência final é estimado baseado na distribuição de equilíbrio <sup>4</sup> .	2016	
10	SR ( <i>Spectral Residual Approach</i> ) (HOU; ZHANG, 2007)	SR	Não possui extração de características e não depende de conhecimento prévio sobre os objetos da cena. SR avalia o <i>log-spectrum</i> <sup>5</sup> , de uma imagem e extrai o resíduo espectral <sup>6</sup> . Esse resíduo é transformado no domínio espacial, pela transformada de Fourier Inversa, para o obter o mapa de saliência.	2007	
11	SS ( <i>Sparse Salient Regions</i> ) (HOU <i>et al.</i> , 2012)	SS	Os autores usaram a função de sinal da Transformada Discreta de Cosseno (DCT) de uma imagem para gerar uma assinatura, contendo principalmente informações sobre o primeiro plano da imagem. O algoritmo explora essas informações para detectar regiões e gerar mapas de saliência.	2012	
12	SIM ( <i>Saliency by Induction Mechanisms</i> ) (MURRAY <i>et al.</i> , 2011)	SIC	Essa metodologia consiste em processar os estímulos visuais de acordo o mecanismo primário do sistema visual humano (por exemplo, oponente de cor, canais de luminância e decomposição em várias escalas). Após, o algoritmo simula os mecanismos de inibição das células do córtex visual e integra informações em escalas múltiplas por uma transformada wavelet inversa. É baseado no modelo de indução de cores unificado desenvolvido por Otazu <i>et al.</i> (2010).	2012	

Continua na próxima página

<sup>4</sup> Exprime a tendência à estabilização da distribuição de porcentagens da cadeia ao longo do tempo. Em outras palavras, significa que, a longo prazo, a probabilidade do sistema está em cada um dos seus estados pouco ou nada varia (equilíbrio). No método GB, a distribuição de equilíbrio da cadeia deve convergir para maiores probabilidades nos nós que tem os maiores valores de dissimilaridades com seus vizinhos.

<sup>5</sup> O *log-spectrum*  $L(f)$  é dado por  $L(f) = \log(AMP(f))$ , onde  $AMP(f)$  é a amplitude do espectro médio de Fourier.

<sup>6</sup> O resíduo espectral é obtido por:  $R(f) = L(f) - AMP(f)$ .

Tabela 2 – Continuação da página anterior

#	Abreviação		Descrição	Ano	Benchmark
	Literatura	Tese			
13	ZH ( <i>Detection of Abnormal Nuclei in Cervical Smear Images</i> ) (ZHANG <i>et al.</i> , 2013)	ZH	Combina um mecanismo de atenção <i>bottom-up</i> com uma estratégia de busca por células anormais para a predição de saliência em imagens de células cervicais em meio líquido. Este algoritmo captura saliência de acordo com as características de brilho e orientação da imagem de entrada e depois modula o mapa produzido por um modelo <i>annular template matching</i> . Os autores projetaram esse <i>template</i> a partir de características de tamanho e forma comuns aos núcleos anormais. Nesta tese, consideramos o trabalho de Wesola <i>et al.</i> (2014) para configurar esses parâmetros.	2013	*

\* modelo desenvolvido para predição de saliência em imagens de células cervicais.

Os modelos SAM VGG-16 *for natural images* (SVS) e SAM ResNet-50 *for natural images* (SRS) foram alcançados a partir da base de imagens naturais SALICON (JIANG *et al.*, 2015). Dessa forma, esses algoritmos foram otimizados para a predição de saliência em imagens naturais, seguindo proposta de Cornia *et al.* (2018). A base de dados SALICON possui 10.000 imagens de treinamento, 5.000 imagens de validação e 5.000 imagens de teste divididas em 80 categorias, sendo o maior conjunto de dados disponível para previsão de saliência em imagens naturais. Para imagens de células cervicais, os modelos SAM VGG-16 *for Cells* (SVC) e SAM ResNet-50 *for Cells* (SRC) foram otimizados com dados da base CRIC *Visual Attention* (CRICVA), considerando 130 imagens para treinamento e 29 para validação, conforme metodologia descrita na Seção 6.1.1.1. Abordamos SVS e SRS como referência para investigar se de fato SVC e SRC modelaram a atenção *top-down* quanto à análise clínica de células cervicais.

### 7.1.2 Base de Imagens

Investigamos a Hipótese H1 a partir de 73 imagens da base de dados CRICVA. Essas imagens foram selecionadas aleatoriamente para a fase de teste e não estão contidas nos conjuntos de treinamento e validação de nosso modelo. Como dados de fixação ocular não são necessários para o estudo de H2, podemos ampliar nossa análise para as imagens anormais remanescentes da base Centro de Reconhecimento para Inspeção de Células (CRIC), totalizando 207 imagens para avaliação. A distribuição de rótulos por imagem, conforme o principal achado clínico presente no campo, é #{ASC-US: 52, LSIL: 100, ASC-H: 26, HSIL: 21 and Carcinoma:

8}. Todas as características das imagens, descritas nas Seções 2.2.1 e 5.2, foram mantidas. A Figura 31 ilustra a construção das bases de teste.

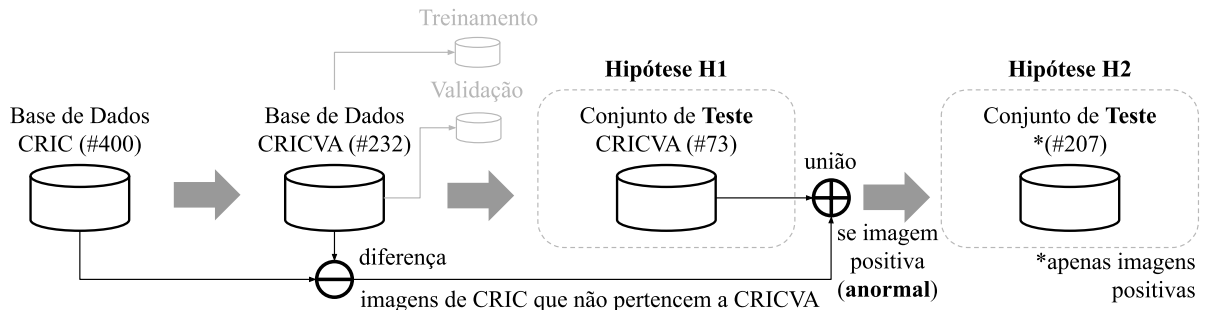


Figura 31 – Construção da base de dados necessária aos experimentos de predição de saliência. A cardinalidade # representa o quantitativo de imagens existentes em cada conjunto utilizado. Fonte: elaboração própria.

O estudo de H1 e H2 exige bases de dados com imagens de exames de Papanicolaou rotuladas tanto quanto à classe das células quanto à atenção visual seletiva dos especialistas. Além disso, as imagens precisam representar campos de visualização, contendo, portanto, um conjunto de células e não apenas recortes de células únicas. Esses requisitos inviabilizaram a inclusão das bases de dados de *benchmark* descritas a seguir em nosso estudo de H1 e H2.

- Herlev (JANTZEN *et al.*, 2005): composta por 917 recortes de células de exames de Papanicolaou com amostras anormais e normais. Utilizamos esta base para validar a matriz *Region-based Structural Co-occurrence Matrix* (rSCM) como descritor para células cervicais (Apêndice B). Não existem dados de atenção visual para esta base.
- ISBI2014 (LU *et al.*, 2016): composta por 16 imagens reais de citologia em meio líquido e 917 sintéticas. Não possui células rotuladas e não existem dados de atenção visual para esta base.
- ISBI2015 (RAMALHO *et al.*, 2015): composta por 17 imagens reais de citologia em meio líquido. Não possui células rotuladas e não existem dados de atenção visual para esta base.
- Cervix93 (PHOULADY; MOUTON, 2018): composta por 93 imagens reais de citologia em meio líquido. Apresenta imagens com células anormais e normais, no entanto, não possui rótulos por células e não dispõe de dados de atenção visual.

### 7.1.3 Métricas de Avaliação

Métricas de avaliação para modelos de predição de saliência podem ser categorizadas em duas classes, conforme a disposição dos dados de *Ground Truth* (GT) (BYLINSKII *et al.*,

2018). As *métricas baseadas em localização* têm como referência mapas binários, no qual os pontos sinalizam as localizações das fixações oculares (Figura 11f). Assim essas métricas permitem examinar mapas de saliência quanto à posição exata capturada do olhar do observador na imagem. Bylinskii *et al.* (2018) classificam as métricas que têm como referência mapas de atenção (Figura 11g) como *métricas baseadas em distribuição*<sup>7</sup>. De acordo com Bylinskii *et al.* (2018), as *métricas baseadas em distribuição* permitem incorporar incerteza nas medições, tais como erros de calibração no equipamento de rastreamento ocular e eventuais imprecisões quanto à posição do olho humano na tela durante a observação das imagens. Assim, ambas as categorias de medidas têm relevância na avaliação dos modelos, sendo, portanto, reportadas em conjunto nesta tese.

Empregamos cinco métricas<sup>8</sup> de avaliação estado-da-arte contemplando as duas categorias. Focamos o uso dessas métricas para avaliação quanto a falsos positivos e falsos negativos, ou seja, se as regiões salientes preditas nos mapas correspondem (e com qual intensidade) as regiões de atenção capturadas pelo rastreador ocular. As *métricas baseadas em localização* adotadas são AUC-Judd (JUDD *et al.*, 2009) e NSS. AUC-Judd tende a 1 quando as fixações oculares estão localizadas nas regiões de maior saliência do mapa predito. No entanto, essa métrica reduz sua sensibilidade à medida que os valores das regiões salientes tendem a zero, tornando-a tolerante a falsos positivos. A métrica NSS mede a correspondência entre o mapa de saliência predito e as fixações oculares, sendo, por sua vez, igualmente afetado por falsos positivos e negativos. Dado  $M$  como sendo o mapa de saliência estimado,  $y^{fix}$  como mapa binário de fixações oculares GT,  $\mu$  como a média e  $\sigma$  como desvio padrão, podemos computar a métrica NSS como:

$$NSS(M, y^{fix}) = \frac{1}{N} \sum_i \bar{M}_i \times y_i^{fix}, \quad (7.1)$$

onde  $N = \sum_i y_i^{fix}$ ,  $\bar{M} = \frac{M - \mu(M)}{\sigma(M)}$ ,  $i$  indexa o  $i$ -ésimo *pixel* e  $N$  é o número total de *pixels* fixados. Valores positivos de NSS sinalizam correspondência entre  $M$  e  $y^{fix}$ . NSS retorna zero quando não há correspondência e negativo nos casos onde existe uma anti-correspondência, ou seja, as fixações oculares estão localizadas nas regiões cujos *pixels* são predominantemente negativos no mapa de saliência normalizado  $\bar{M}$  (média zero e desvio padrão unitário). NSS produz valores definidos em  $[-\infty, +\infty]$ .

<sup>7</sup> Essas métricas assumem os mapas como distribuição de probabilidade.

<sup>8</sup> Todas as métricas possuem código fonte disponível em <[https://github.com/cvzoya/saliency/tree/master/code\\_forMetrics](https://github.com/cvzoya/saliency/tree/master/code_forMetrics)> e compõem o *benchmark* de modelos de predição de saliência do MIT (BYLINSKII *et al.*, 2015), disponível em <<http://saliency.mit.edu/>>.



Adotamos as seguintes métricas baseadas em distribuição: *Linear Correlation Coefficient* (CC), *Similarity* (SIM) e *Kullback-Leibler divergence* (KL). A métrica CC avalia o nível de correlação linear entre dois conjuntos de dados. Dessa forma, CC tende a +1 nas regiões onde  $M$  e  $y^{aten}$  possuem magnitudes similares, assumindo  $y^{aten}$  como o mapa de atenção GT. Quando CC está próximo de -1 ou +1, há quase uma relação perfeitamente linear entre as duas variáveis. Essa métrica penaliza falsos positivos e falsos negativos igualmente. CC resulta em valores em  $[-1, +1]$  e pode ser computada por:

$$CC(M, y^{aten}) = \frac{cov(M, y^{aten})}{\sigma(M) \times \sigma(y^{aten})}, \quad (7.2)$$

onde  $cov(M, y^{aten})$  é a covariância de  $M$  e  $y^{aten}$ .

A métrica SIM calcula a interseção entre duas distribuições. Trata-se de uma medida originalmente concebida para avaliar casamento de padrões entre duas imagens (RUBNER *et al.*, 2000), (SWAIN; BALLARD, 1991). No contexto de saliência, SIM verifica a interseção entre os mapas de referência e o predito, visto como histogramas, sendo mais sensível a falsos negativos que a falsos positivos. Seus valores resultantes estão definidos em  $[0, 1]$ , onde  $SIM = 1$  indica que as distribuições são a mesma e  $SIM = 0$  que não existe sobreposição, portanto, diferentes. Podemos computar SIM como a soma dos valores mínimos em cada *pixel*, após normalizar os mapas de entrada:

$$SIM(M, y^{aten}) = \sum_i \min(M_i, y_i^{aten}), \quad (7.3)$$

onde  $\sum_i M_i = \sum_i y_i^{aten} = 1$ , com  $i$  representando a localização de cada *pixel*.

KL é uma medida de dissimilaridade assimétrica altamente sensível a falsos negativos. Essa métrica faz uma interpretação probabilística entre o mapa estimado  $M$  e a referência  $y^{aten}$ , avaliando a perda de informação quando a distribuição  $M$  é utilizada para aproximar a distribuição  $y^{aten}$ . KL resulta em  $[0, \infty]$ , com baixos valores indicando os melhores resultados. A métrica KL é tão sensível a valores nulos que um conjunto esparsa de previsões é bastante penalizado. Formalmente, temos:

$$KL(M, y^{aten}) = \sum_i y_i^{aten} \log \left( \varepsilon + \frac{y_i^{aten}}{\varepsilon + M_i} \right), \quad (7.4)$$

onde  $\varepsilon$  é uma constante de regularização<sup>9</sup> e  $i$  representa o  $i$ -ésimo *pixel* do mapa.

A Tabela 3 resume as principais características das métricas de predição adotadas nesta tese, apontando aplicações alinhadas às propriedades de cada uma delas. Embora limitemos

<sup>9</sup> Neste trabalho, utilizamos  $\varepsilon = 2.2204e - 16$ , conforme empregado por Bylinskii *et al.* (2015).

Tabela 3 – Principais características das métricas de avaliação de predição de saliência empregadas nesta tese. A coluna *Categoria* informa se o mapa de referência GT equivale ao mapa de fixações oculares (Localização) ou mapa de atenção (Distribuição). Fonte: elaboração própria a partir de (BYLINSKII *et al.*, 2018).

Métricas*	Intervalo Teórico	Categoria	Características
AUC-Judd ↑	[0, 1]	Localização	1. Variação da Área sob a curva ROC ( <i>Receiver Operating Characteristic</i> ) (AUC) que é historicamente a métrica mais usada para avaliar algoritmos de predição de saliência; 2. Tende a 1 quando as fixações oculares estão localizadas nas regiões de maior saliência do mapa predito. Reduz a sensibilidade à medida que os valores das regiões salientes tendem a zero, tornando-a tolerante a falsos positivos; 3. Ótimo indicador para aplicações de detecção.
NSS ↑	$[-\infty, +\infty]$	Localização	1. Apresenta igual sensibilidade a falsos positivos e a falsos negativos; 2. Por ser computado a partir dos pontos de fixação ocular (mapa binário obtido diretamente pelo rastreador ocular), fornece melhor avaliação quanto maior for o número de observadores; 3. Ótimo indicador para aplicações de casamento de padrões e recuperação de conteúdo.
CC ↑	[-1, 1]	Distribuição	1. Penaliza falsos positivos e falsos negativos igualmente; 2. Pode ser utilizado para procurar mapas de saliência correlacionados ao longo de uma base de imagens e/ou observadores, sendo útil, portanto, em aplicações de casamento de padrões e recuperação de conteúdo.
SIM ↑	[0, 1]	Distribuição	1. Apresenta maior sensibilidade a falsos negativos que a falsos positivos; 2. Importante métrica para avaliar algoritmos de predição aplicados a problemas de busca parcial de padrões, ou seja, em partes da imagem, e recuperação de conteúdo.
KL ↓	[0, $\infty$ ]	Distribuição	1. Extremamente sensível a falsos negativos. 2. Ótimo indicador para aplicações de detecção.

\* ↑ : métrica de similaridade. Os melhores resultados tendem ao limite superior do intervalo teórico.

↓ : métrica de dissimilaridade. Os melhores resultados tendem ao limite inferior do intervalo teórico.

a aplicação dos mapas de saliência à identificação de ROIs, a análise por variadas métricas contribui com evidências de como os algoritmos estudados podem desempenhar em outros contextos, como quando aplicado para casamento de padrões e recuperação de conteúdo. Recomendamos a leitura de (BYLINSKII *et al.*, 2018) para um estudo mais amplo dessas medidas.

Analisamos a Hipótese H2 medindo a probabilidade das células de maior relevância clínica estarem contidas em regiões salientes preditas. Inspirados por Bylinskii *et al.* (2016), definimos como células salientes aquelas cujas posições pertencem às zonas de *pixels* com maior amplitude nos mapas estimados. Assim, o Algoritmo 1 foi utilizado para delinear até  $nr$  regiões com os maiores valores de saliência para cada mapa estimado. Experimentalmente, observamos que  $nr = 3$  potencializa nossa análise de H2, revelando as maiores diferenças entre os métodos estudados. Quando o mapa tem menos que três ROIs salientes, o mapa de saliência é limiarizado em 0.95.

Definimos a taxa de concordância  $\varphi \in [0, 1]$  entre as ROIs e a região da alteração

**Algoritmo 1:** Algoritmo para seleção de ROIs de acordo com os níveis de densidade das diferentes regiões contidas nos mapas de saliência estimados. Empregado na análise da Hipótese H2.

```

1 function ROIselect ( $M, nr$ );
   Input : ( $float\ matrix \in [0,1]$ )  $M$ : Mapa de saliência estimado; ( $int$ )  $nr$ : Número de
           regiões requisitadas
   Output : ( $logical\ matrix$ )  $Z$ : Máscara binária sinalizando as locações das ROIs
2  $th = 0.95$ ;
3  $step = 0.05$ ;
4  $count = 0$ ;
5 while ( $count < nr$ ) and ( $th > 0$ ) do
6   |  $Z = logical(M * (M \geq th))$ ;
7   |  $count = getNumberOfSeparatedRegions(Z)$ ;
8   |  $th = th - step$ ;
9 end
10 if ( $th \leq 0$ ) then
11 |  $Z = logical(M \geq 0.95)$ ;
12 end
13 Ranquear as regiões em ordem decrescente com base na energia dos respectivos  $pixels$  e
    retornar até  $nr$  regiões mais salientes.
14  $Z = getNRMostSalientRegions(Z, M, nr)$ ;
15 return  $Z$ ;

```

celular mais significativa ( $le$ ) como:

$$\varphi = \frac{1}{N_{le}} \sum_{i=1}^{N_{le}} \#\{\forall r \in Z_i \mid (r \cap G_i) \neq \emptyset\}, \quad (7.5)$$

onde  $\#\{\cdot\}$  representa a cardinalidade,  $N_{le}$  corresponde ao número de imagens de teste com a alteração  $le$ ,  $Z$  é a máscara binária resultante do Algoritmo 1,  $r$  é a ROI e  $G$  representa a máscara binária de referência (GT) sinalizando a área de  $100 \times 100$   $pixels$  contendo a alteração  $le$ . Consideramos *identificado* se qualquer parte da área de alteração casar com a ROI. Caso contrário, assumimos que o modelo não pode identificar a localização da alteração celular de forma apropriada.

#### 7.1.4 Análise da Hipótese H1

Investigamos H1 em 73 imagens, as quais compõem o conjunto de teste da base CRICVA (Figura 31). Inicialmente os resultados dos algoritmos que utilizam apenas características *bottom-up* foram analisados. Esses algoritmos não foram otimizados para reconhecer informações específicas no contexto de análise de células cervicais. Logo após, discutiremos o desempenho dos modelos *top-down*, incluindo os modelos baseados em *Convolutional Neural*

*Network* (CNN). Incluímos também mapas de saliência que simulam o viés de centro. Esses mapas correspondem ao *baseline* de centro, os quais consistem em uma bolha gaussiana no centro da imagem. Empregamos este modelo para verificar a existência de viés de captura<sup>10</sup> na base CRIC e investigar a influência do viés de centro no desempenho dos métodos estudados.

#### 7.1.4.1 Modelos bottom-up

A Figura 32 mostra mapas de saliência preditos pelos algoritmos abordados, incluindo os modelos SVC e SRC. As regiões menos salientes estão representadas na cor azul escuro enquanto a cor vermelho escuro destaca o contrário. Um estudo quantitativo dos resultados desses algoritmos é apresentado na Figura 33.

Os algoritmos *bottom-up* tendem a destacar regiões de falsos positivos, ou seja, dar destaque a partes da imagem que não foram observadas pelos citopatologistas. Como essa categoria de modelos é comumente projetada para simular a atenção humana seguindo estímulos de baixo nível, fatores como contraste entre as células e o fundo da imagem e coloração das estruturas da lâmina tem impacto significativo na predição da saliência. O método SIC, por exemplo, é uma evidência desse nosso achado. Por ser um algoritmo fundamentado na percepção de cor, utilizando por base a metodologia desenvolvida por Otazu *et al.* (2010), SIC foi sensível a regiões com maior presença de contraste, como transição células-fundo da imagem, artefatos e ruído. Portanto, em aplicações onde falsos positivos são prejudiciais, o método SIC é raramente adequado<sup>11</sup>. Por outro lado, observamos que métodos sensíveis a contraste podem contribuir para encontrar regiões com núcleos normais, neutrófilos e agrupamentos de células, devido essas estruturas se apresentarem geralmente em cor escura e homogênea quanto ao citoplasma celular e ao fundo da imagem.

O algoritmo FES produziu mapas de saliência visualmente próximos aos mapas de atenção dos citopatologistas. Esses resultados sugerem que o paradigma *center-surround*<sup>12</sup>, com uso de características modeladas por uma estrutura Bayesiana, contribui para a geração de mapas de saliência precisos em regiões de células anormais. Como o modelo Bayesiano

<sup>10</sup> O viés de captura transmite alguma tendência dos fotógrafos de posicionar os alvos na imagem durante a imagem da cena (TORRALBA; EFROS, 2011).

<sup>11</sup> Como exemplo de aplicações que são sensíveis a falsos positivos, destacamos: o ranqueamento de imagens por demanda de atenção visual e recuperação de imagens com base em saliência. Em outras aplicações, como para detecção de alvos, mapas com falsos positivos podem levar ao processamento de regiões desnecessárias, aumentando o custo computacional e a chance de perda do tempo real do sistema.

<sup>12</sup> Paradigma em que a hipótese é que existe uma janela local dividida em um centro e um entorno, e o centro contém o objeto.

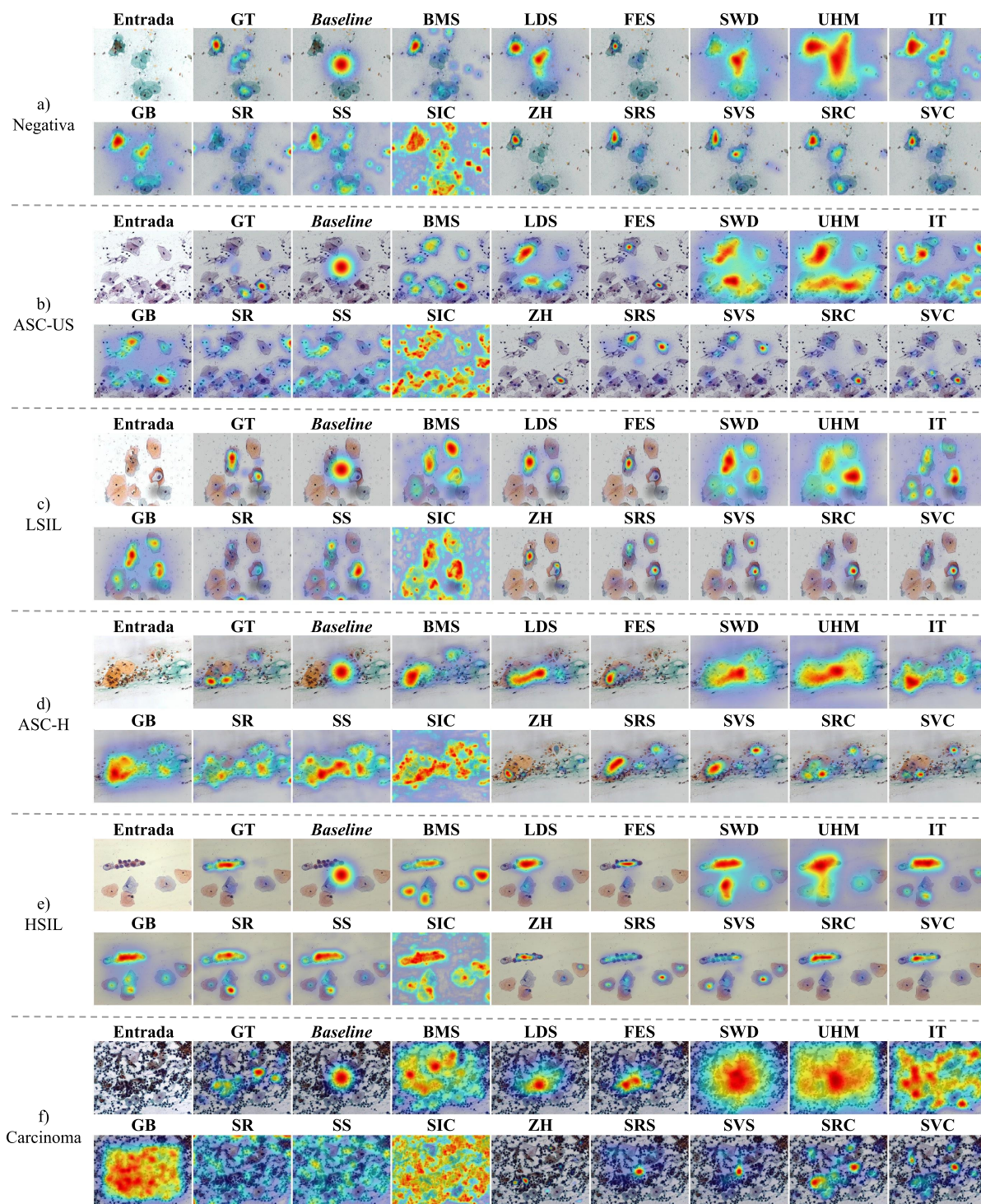


Figura 32 – Mapas de saliência, sobrepostos às imagens de entrada, computados pelos métodos estudados. a) - f) mostram os resultados para cada caso descrito na base CRICVA. Alguns modelos são capazes de indicar regiões conspícuas da imagem, enquanto outros podem ser ineficazes na previsão do olhar do citopatologista, quando as imagens em análise são ruidosas e/ou possuem elevada sobreposição celular. Fonte: elaboração própria.

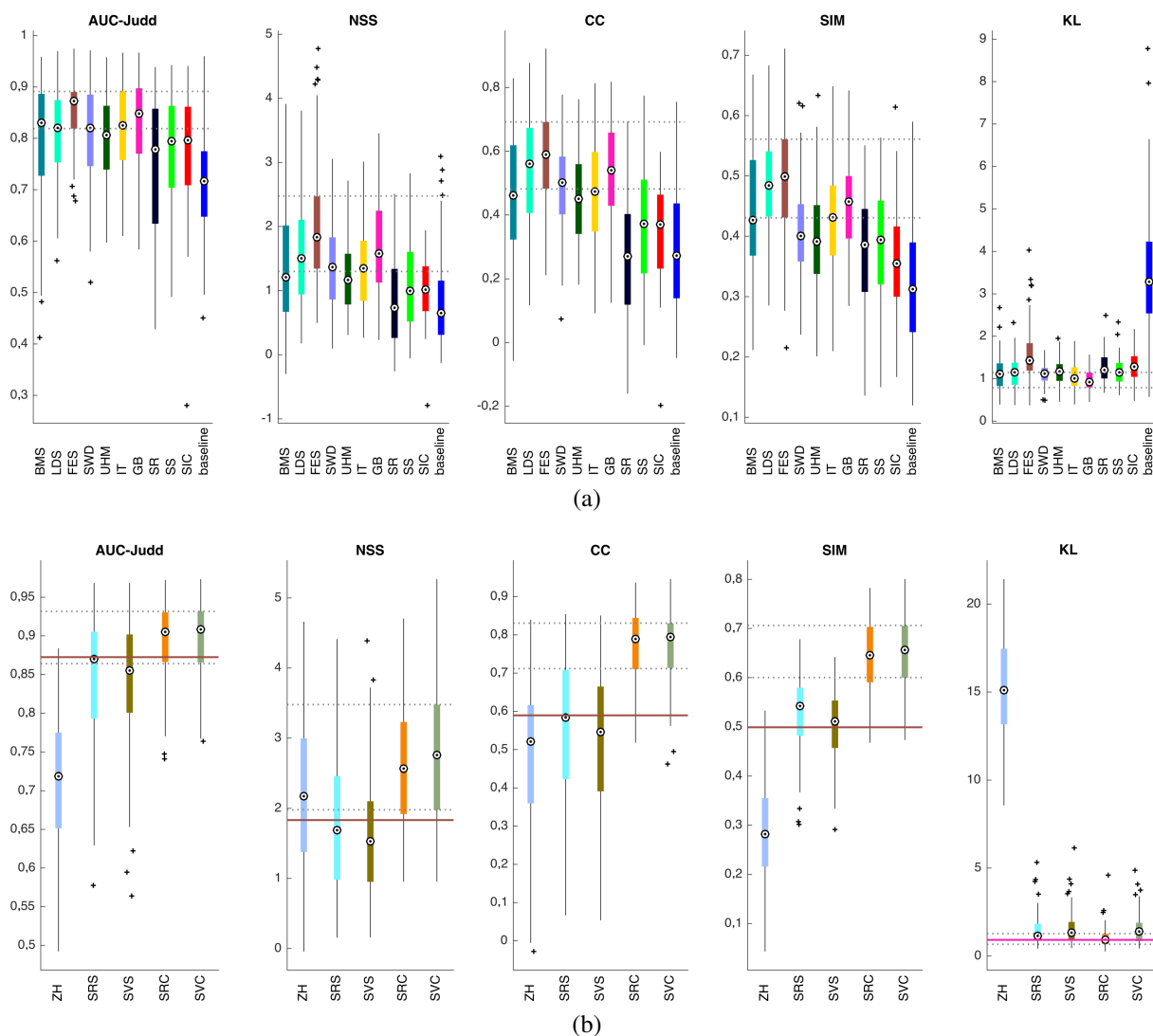


Figura 33 – Avaliação quantitativa dos métodos a) *bottom-up* e b) *top-down*, incluindo os baseados em CNN. A linha horizontal pontilhada representa a faixa interquartil do melhor modelo. Para cada métrica, a linha colorida horizontal marca o desempenho do melhor algoritmo *bottom-up*. Fonte: elaboração própria.

necessita de estimação das distribuições *a priori*, processamos o *Average Annotation Map* (AAM) de CRICVA (Figura 14a), excluindo os dados de teste, para este fim, conforme proposto por Tavakoli *et al.* (2011). Embora o algoritmo FES tenha enfatizado regiões próximas ao centro da imagem, devido à informação AAM, esse algoritmo encontrou regiões com células anormais na periferia da imagem. Não encontramos diferenças significativas no desempenho do modelo a partir de uma distribuição *a priori* uniforme, ou seja, sem priorizar regiões da imagem. Os modelos BMS, LDS e GB também apresentaram altos valores de saliência nas localizações das fixações oculares dos citopatologistas, mas mostraram-se mais sensíveis a falsos positivos.

Computamos o teste estatístico de Kruskal-Wallis (KRUSKAL; WALLIS, 1952) com o teste *post-hoc* Nemenyi ( $\alpha = 0,05$ ) (HOLLANDER *et al.*, 2013) para encontrar os

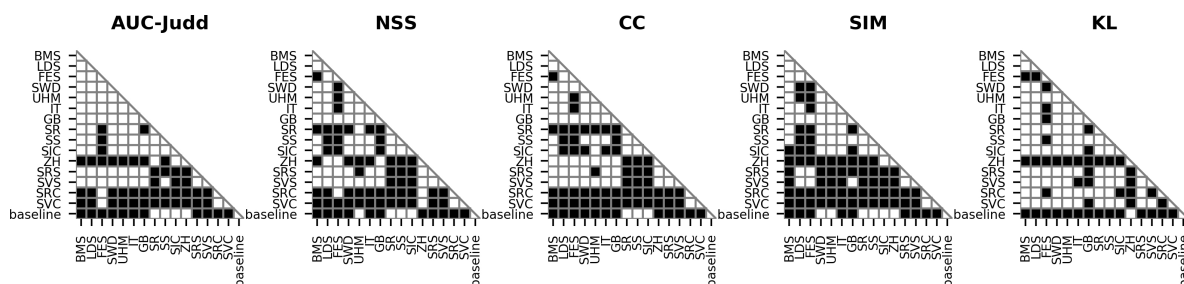


Figura 34 – Comparações pareadas para todos os métodos analisados usando o teste estatístico de Kruskal-Wallis com o teste *post-hoc* de Nemenyi. As caixas pretas representam os pares com diferença significativa em  $\alpha = 0,05$ . Fonte: elaboração própria.

resultados dos modelos que diferem estatisticamente um dos outros. A Figura 34 relata todas as comparações aos pares para cada métrica de avaliação estudada.

Em termos da métrica AUC-Judd, o algoritmo FES superou todos os modelos baseados apenas em características *bottom-up*. No entanto, não observamos diferenças estatísticas entre seus resultados e os obtidos por BMS, LDS, SWD, UHM, IT e GB. Baseado nas propriedades das métricas adotadas (Tabela 3) (BYLINSKII *et al.*, 2018), apontamos que esses modelos são adequados a aplicações de predição de saliência em que falsos positivos não influenciam consideravelmente os resultados. Se os efeitos de falsos positivos e falsos negativos são igualmente considerados, a métrica NSS apontou LDS, FES e GB como os métodos de melhor desempenho, e as métricas CC e SIM confirmaram esse achado. Dessa forma, argumentamos que os algoritmos LDS, FES e GB são viáveis para previsão de saliência em imagens de células cervicais que requerem precisão na identificação de regiões conspícuas. Encontramos algumas exceções, que correspondem aos casos em que há um grande número de artefatos de imagem, como o caso exemplificado na Figura 32f. Em casos similares ao da Figura 32f, observamos que algoritmos de predição de saliência baseados apenas em estímulos de baixo nível podem ser inefetivos, devido à predominância de regiões escuras e ausência de informação de contraste que possam definir células e, conseqüentemente, ROIs. Embora o algoritmo FES tenha pontuado bem a tarefa de localização das regiões conspícuas na maioria dos casos, seus *pixels* salientes apresentaram-se mais concentrados em torno do pico quando comparado ao GT. Esse fato prejudicou a avaliação do método FES pela métrica KL, que é altamente sensível a falsos negativos em *pixels*. Essas são algumas indicações de que as configurações adicionais do FES podem aumentar seu desempenho, mas outras investigações podem ser necessárias para confirmar esses vieses.

O *baseline* de centro confirmou que parte da atenção visual do especialista foi dedicada aos alvos no centro da imagem. No entanto, os melhores métodos *bottom-up* atingiram

regiões salientes distantes do centro da imagem. Isso demonstra a capacidade destes algoritmos em capturar as características de baixo nível dos alvos, mesmo sem considerar qualquer informação de contexto (*top-down*). Com base nessa análise, validamos a Hipótese H1 argumentando que a atenção visual do citopatologista está correlacionada com as características de baixo nível das células anormais. Além disso, a modelagem apropriada dos atributos das células pode permitir o uso de algoritmos rápidos *bottom-up*, como parte de uma estrutura de previsão de saliência para análise de imagens de células.

#### 7.1.4.2 Modelos *top-down*

A Figura 32 mostra que a abordagem *top-down* implementada pelo modelo ZH é uma estratégia promissora para localizar áreas de atenção em imagens de células cervicais. No entanto, observamos que os mapas de saliência estimados por esse modelo são esparsos em relação às fixações oculares dos especialistas. Como o modelo ZH é guiado pela correspondência de um *template*, modelado a partir do tamanho característico de células anormais, algumas regiões salientes tendem a ser negligenciadas pelo casamento não satisfatório entre o *template* e o formato de apresentação das células. Em exames de Papanicolaou convencional, os núcleos comumente aparecem distorcidos e/ou sobrepostos com artefatos (Figura 2) com variações na intensidade de cor resultantes do processo de coloração da lâmina. Estes fatores modificam o formato circular esperado para o núcleo, reduzindo o desempenho do *template* anular adotado por ZH. Confirmamos essas observações por meio da análise quantitativa apresentada na Figura 33. A métrica AUC-Judd revelou que o modelo ZH subponderou as áreas de atenção relevantes (falso negativo), embora o NSS tenha confirmado a assertividade de regiões. Além disso, as *métricas baseadas em distribuição* (CC, SIM, KL) apontaram uma distância significativa entre as áreas de atenção exigidas pelos citopatologistas e aquelas estimadas pelo modelo ZH.

Os experimentos mostraram que os modelos SVC e SRC codificam satisfatoriamente a atenção *top-down* relacionada à análise dos citopatologistas, levando-os a estimar mapas de saliência mais precisos que os outros métodos estudados. Como SVS e SRS possuem desempenho similar aos algoritmos *bottom-up* de melhor rendimento (Figura 34), argumentamos que SVC e SRC aprenderam atributos relevantes quanto à atenção visual dos especialistas, mesmo com treinamento em um conjunto reduzido de imagens de células cervicais. Particularmente, demonstramos que a estratégia de *transferência de aprendizado* a partir de outro domínio (imagens naturais) é viável para a predição de saliência em imagens de células cervicais, mesmo



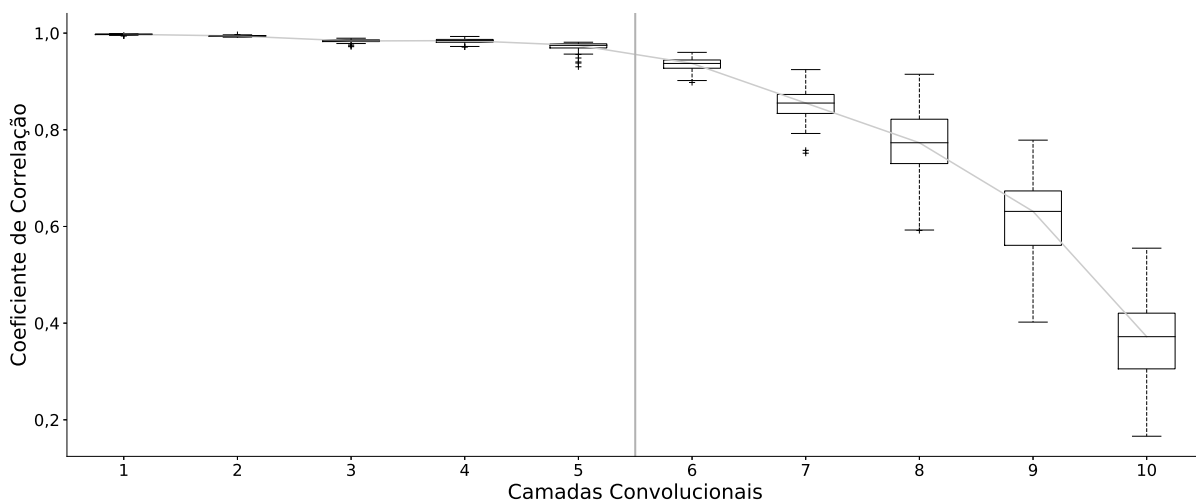
quando poucos dados específicos estão disponíveis para ajuste. Embora os resultados obtidos confirmem diferenças entre fatores *top-down* na análise de imagens naturais e de células cervicais, encontramos evidências que os *frameworks* baseados em CNN podem prever mapas de saliência valiosos em imagens de células quando incorporam conhecimento de outros domínios. Uma interpretação para isso é que existem características comuns que são relevantes em ambos casos, possivelmente quanto às primitivas de visão de baixo nível. Além disso, os resultados apontaram que os fatores *top-down* orientam a atenção do citopatologista durante a análise dos campos, reduzindo principalmente a sensibilidade ao brilho e áreas de alto contraste.

Explorando as arquiteturas CNN estudadas, analisamos a correlação entre as ativações produzidas pelas redes quando treinadas quanto a fatores *top-down* em imagens naturais (SALICON) e quanto a imagens de células cervicais (CRICVA) (Figura 35). Observamos que aproximadamente a primeira metade das camadas tem coeficiente de correlação acima de 95% em suas ativações, reforçando que as características de baixo nível são comuns em ambos domínios (imagens naturais e de células cervicais) e há codificação satisfatória da atenção *top-down* dos especialistas pelos modelos SVC e SRC nas camadas superiores da rede. Para o treinamento da CNN em imagens de células, esse achado pode contribuir para estabelecer parâmetros de ajuste fino (*fine-tuning*), como o número de camadas inferiores a serem congeladas e a taxa de aprendizado. Além disso, nossa análise mostrou que as ativações das camadas inferiores da CNN tendem a ser similares para imagens naturais e de células cervicais quando aplicado à predição da atenção visual humana. Assim, confirmamos essa propriedade das CNNs também para o contexto de predição de saliência.

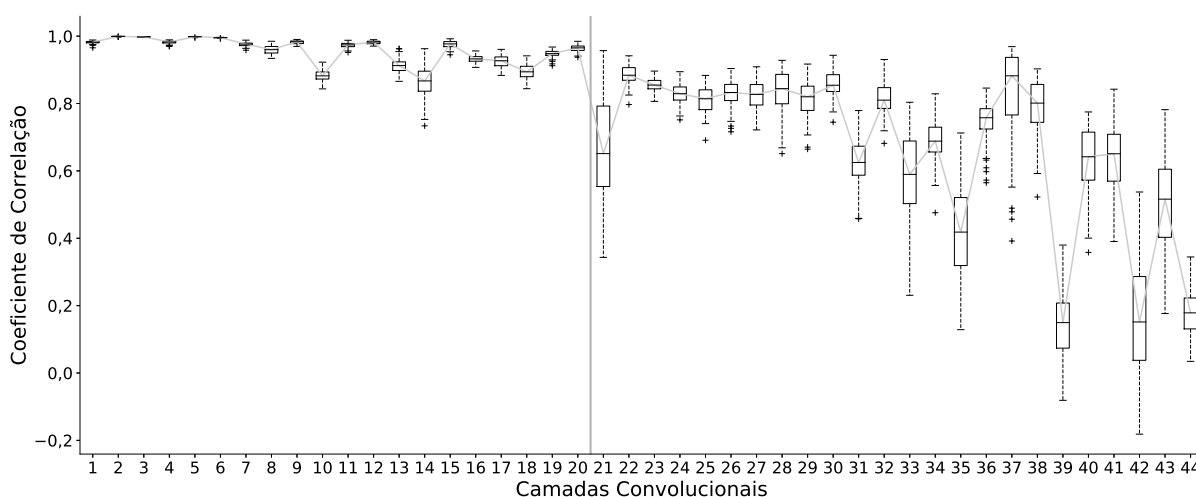
Os testes estatísticos de Kruskal-Wallis e *post-hoc* Nemenyi não identificaram diferenças estatísticas ( $\alpha = 0,05$ ) entre SVC e SRC para todas as métricas de avaliação, exceto KL (Figura 34). Portanto, argumentamos que ambas as variantes podem ser igualmente escolhidas para aplicações em células cervicais, de acordo com a disponibilidade dos recursos computacionais.

#### 7.1.4.3 Tempo de processamento

A Tabela 4 apresenta o tempo de processamento médio dos métodos de predição de saliência estudados, com base nos experimentos no banco de dados CRICVA. As abordagens *bottom-up*, com exceção do método SIC, foram mais rápidas que as de aprendizado profundo. Analisando em conjunto a Tabela 4 e a Figura 33, encontramos que alguns métodos *bottom-up*



(a)



(b)

Figura 35 – Coeficiente de correlação entre as ativações das redes a) VGG-16 e b) ResNet-50 treinadas com amostras da base SALICON e CRICVA. As ativações apresentam correlação acima de 95% nas primeiras camadas, lado esquerdo da linha vertical, em ambas arquiteturas. Fonte: elaboração própria.

podem ser uma solução viável para a análise de imagens de células cervicais em aplicações nas quais os recursos computacionais são limitados. Isso nos motivou a relatar resultados usando um computador Intel (I7-4770HQ) CPU (2,2 GHz) e 16 GB de RAM.

Resumimos as contribuições da Hipótese H1 relatando técnicas rápidas *bottom-up* que podem identificar as regiões mais conspícuas em uma imagem de células cervicais durante a análise por citopatologistas. Além disso, confirmamos um achado descrito por Coombes e Culverhouse (2003), o qual afirma existir alta correlação entre algumas características de baixo nível de células anormais e o padrão de fixação ocular dos especialistas. Também apresentamos um modelo CNN, em duas diferentes arquiteturas (VGG-16 e ResNet-50), para predição de

Tabela 4 – Ranque do tempo médio de processamento dos métodos de predição de saliência estudados (segundos por imagem). Os algoritmos baseados em CNN estão em negrito. {Mat: Matlab, Py: Python}. Fonte: elaboração própria.

<b>Método</b>	SR	SS	FES	IT	ZH	GB	BMS	LDS
<b>Código</b>	Mat	Mat	Mat	Mat	Mat	Mat	Py	Mat
<b>Tempo</b>	0,01	0,02	0,22	0,23	0,50	0,51	0,63	0,96
<b>Método</b>	UHM	SWD	<b>SVC</b>	<b>SVS</b>	<b>SRC</b>	<b>SRS</b>	SIC	
<b>Código</b>	Mat	Mat	<b>Py</b>	<b>Py</b>	<b>Py</b>	<b>Py</b>	Mat	
<b>Tempo</b>	1,36	2,38	<b>3,23</b>	<b>3,27</b>	<b>3,85</b>	<b>3,86</b>	10,03	

saliência em imagens de células cervicais convencionais. Esses modelos foram alcançados a partir dos mapas de atenção de três especialistas e obtiveram predições que superam o estado-da-arte.

### 7.1.5 Análise da Hipótese H2

A investigação de H2 foi realizada em 207 imagens de células cervicais (Figura 31). A Tabela 5 apresenta os resultados da métrica  $\phi$  para todos os algoritmos de saliência estudados. Considerando apenas três ROIs, o modelo SRC identificou a localização da alteração celular de maior relevância na imagem para todas as patologias consideradas, com exceção de 1 caso LSIL onde quatro regiões foram necessárias. Esse resultado demonstrou a potencialidade desse modelo como extrator de regiões de interesse para análise de células com propósito clínico. De fato, observamos que lesões relevantes foram realçadas nos mapas estimados, podendo ser isoladas para análise. O SVC também mostrou resultados valiosos para sistemas baseados em ROIs, no entanto, foi menos robusto na detecção de regiões com lesões ASC-US e carcinoma do que o método SRC. O modelo ZH, que implementa um mecanismo *top-down*, alcançou seu melhor resultado para extração de ROIs com LSIL. Como os autores do ZH relataram apenas resultados de experimentos em citologia de meio líquido (ZHANG *et al.*, 2013), os trabalhos desenvolvidos nesta tese estendem a aplicação desse método aos exames de Papanicolaou convencionais. Algumas estratégias *bottom-up*, como as adotadas pelo FES e GB, podem ser adequadas para a seleção de ROIs em imagens de células cervicais, especialmente em amostras com aglomerados celular menos contaminados por artefatos escuros. Embora SWD e UHM tenham tido um bom desempenho nos experimentos de H2, as Figuras 32 e 33 mostraram a existência considerável de falsos positivos na predição da saliência, o que pode restringir o uso desses métodos para algumas aplicações. Esses resultados, além dos descritos na Figura 33, apontam que o método FES é a melhor estratégia analisada para seleção de ROIs em imagens de células cervicais sem uso de características *top-down*.

Tabela 5 – Coeficiente de concordância  $\phi$  (Equação 7.5) entre as  $nr = 3$  regiões mais salientes e o achado clínico de maior relevância da imagem. Os melhores resultados para cada alteração celular estão em negrito. Fonte: elaboração própria.

	ASC-US	LSIL	ASC-H	HSIL	Carcinoma
BMS	0,731	0,820	0,808	0,857	0,500
LDS	0,692	0,860	0,885	0,952	0,625
FES	0,865	0,860	0,962	<b>1,000</b>	0,750
SWD	0,808	0,900	0,923	<b>1,000</b>	0,625
UHM	0,712	0,810	0,846	<b>1,000</b>	0,625
IT	0,462	0,680	0,923	<b>1,000</b>	0,375
GB	0,730	0,870	0,923	0,904	0,500
SR	0,173	0,280	0,384	0,571	0,000
SS	0,385	0,450	0,731	0,762	0,250
SIC	0,153	0,390	0,730	0,571	0,375
ZH	0,846	0,930	0,884	0,761	0,375
SRS	0,865	0,860	0,846	0,952	0,625
SVS	0,731	0,840	0,885	0,952	0,500
SRC	<b>1,000</b>	0,990	<b>1,000</b>	<b>1,000</b>	<b>1,000</b>
SVC	0,981	<b>1,000</b>	<b>1,000</b>	<b>1,000</b>	0,875
Especialistas*	1,000	0,989	1,000	1,000	0,786
<i>Baseline</i> de centro	0,365	0,570	0,653	0,904	0,250

calculado a partir de todas as imagens de CRICVA. Os especialistas alcançaram  $\phi = 1$  para LSIL e Carcinoma com  $nr = 4$  e  $nr = 5$  ROIs salientes, respectivamente.

O *baseline* de centro não revelou viés de captura significativo para as lesões ASC-US, LSIL e ASC-H, confirmando o desempenho dos métodos pesquisados. Uma vez que mostramos a viabilidade da predição de atenção em imagens de células cervicais, novos trabalhos podem aperfeiçoar os estudos para os casos HSIL e carcinoma. Novas diretrizes podem explorar bases de dados com maior diversificação celular e campos HSIL onde as células de interesse estejam mais afastadas do centro. Portanto, argumentamos como principal contribuição de H2, o levantamento quantitativo da confiabilidade de algoritmos estado-da-arte, bem como de nossa proposta, para predição de saliência em imagens de células cervicais.

## 7.2 Experimentos com *Attention Guided Model for inspection of cells* (AGMic)

Esta seção analisa os resultados obtidos a partir do modelo proposto AGMic. São apresentados os resultados de experimentos com e sem a etapa local de refinamento. Particularmente, estamos interessados em investigar a hipótese:

- **Hipótese 3 (H3):** O ranqueamento de regiões de interesse com base na predição da atenção visual de citopatologistas é uma estratégia computacional viável para a redução de erros de classificação de imagens de exames de Papanicolaou.

Em todos os experimentos com o AGMic, o modelo SVC compõe o Módulo 1. Embora o modelo SRC apresente melhores resultados de predição de saliência para a maioria das classes de alterações (Tabela 5), diferenças estatísticas para o SVC dentro das métricas avaliadas (Figuras 33 e 34) não foram observadas. Além disso, o modelo SVC é composto por CNN com menor custo em tempo de execução (Tabela 4). O algoritmo de refinamento foi ajustado da seguinte forma: 1) rSCM com filtro gaussiano de desvio padrão  $\sigma = 4$  (detalhes na Seção B.4); 2) algoritmo SLIC<sup>13</sup> (ACHANTA *et al.*, 2012) para computação dos *superpixels*; 3) classificador *Random Forest* (BREIMAN, 2001) treinado na base Herlev, *cross-validation* com 10 *folds*<sup>14</sup>. Optamos pelo uso de um classificador treinado na base Herlev devido o conjunto de dados CRIC não apresentar segmentação de núcleos, ou padrão ouro, por especialistas. Embora os experimentos reportados na Seção B.4 apontem que dilatar a máscara de segmentação do núcleo com elemento estruturante de raio igual 2 aumenta a discriminância entre células normais e anormais, as máscaras estimadas por *superpixels* comumente contemplam parte do citoplasma por imprecisão em torno do núcleo. Dessa forma, optamos pelo não emprego da dilatação de máscara<sup>15</sup> na etapa local do Módulo 1 do AGMic. O Módulo 2, por sua vez, corresponde a um modelo CNN VGG-16 modificado (Seção 6.2).

### 7.2.1 Base de Imagens

A Hipótese H3 é analisada a partir de 111 imagens da base de dados CRIC. Este conjunto é composto pelas imagens que não foram utilizadas para treinamento dos Módulos 1 e 2 (Figura 27). As amostras empregadas contemplam 3.183 células cervicais, com densidade média de 28,68 células/imagem (SD = 21,85). Na prática clínica, as classes de células podem ser agrupadas por compartilharem de protocolos similares de acompanhamento e tratamento,

<sup>13</sup> Implementação do pacote *python scikit-image* (versão: 0.14.0) com os seguintes parâmetros: *sigma = 2*, *compactness = 25* e *enforce\_connectivity = True*.

<sup>14</sup> Parâmetros reportados conforme experimento *grid search* (pacote *python scikit-learn* - versão: 0.19.1): *RandomForestClassifier(bootstrap=True, class\_weight=None, criterion='gini', max\_depth=None, max\_features='auto', max\_leaf\_nodes=None, min\_impurity\_decrease=0.0, min\_impurity\_split=None, min\_samples\_leaf=1, min\_samples\_split=2, min\_weight\_fraction\_leaf=0.0, n\_estimators=128, n\_jobs=1, oob\_score=False, random\_state=None, verbose=0, warm\_start=False)*.

<sup>15</sup> A Figura 50a mostra que a inclusão excessiva de áreas do citoplasma que estão ao redor do núcleo leva à perda de desempenho da matriz rSCM como descritor na identificação de células anormais.

logo os dados estão organizados como descrito na Tabela 1: alteração de alto risco (ASC-H, HSIL e carcinoma), baixo risco (ASC-US e LSIL) e negativas (normal). Assim, temos a seguinte distribuição: #{Alto risco: 21, Baixo risco: 73 e Negativa: 17}.

### 7.2.2 Medidas de Avaliação

O desempenho do modelo AGMic é avaliado a partir de cinco medidas. A Acurácia (*ACC*) e o índice de concordância Kappa (*K*), descritos na Seção B.2, quantificam o desempenho do sistema em termos de classificação local das ROIs e das imagens. Embora a Acurácia seja sensível ao desbalanceamento de classes, seu valor é comumente reportado na literatura como referência. O índice Kappa, por sua vez, é mais robusto ao desbalanceamento e pode ser visto em conjunto para melhor análise. Particularmente, na avaliação da classificação das imagens, medimos também a Precisão, Revocação e *F*, as quais são descritas a seguir.

A Precisão (do inglês, *Precision*) (POWERS, 2011) corresponde à fração de instâncias efetivamente relevantes entre as que o modelo classificou como relevantes. Essa medida reflete a exatidão do modelo, ou seja, sua qualidade em identificar como positiva apenas os casos que são de fato positivos. Em sistemas médicos aponta a probabilidade de um paciente com teste positivo ter a doença, revelando também as chances de um paciente saudável ser submetido a tratamentos desnecessários. Formalmente, o valor da Precisão (*PREC*) é computado por:

$$PREC = \frac{TP}{TP + FP}, \quad (7.6)$$

em que *TP* e *FP* correspondem às taxas de verdadeiros positivos e falsos positivos, respectivamente. Os valores de *PREC* estão definidos no intervalo  $[0, 1]$ , com 1 representando o melhor caso.

A Revocação (do inglês, *Recall*) (POWERS, 2011), também conhecida por *sensibilidade*, é uma medida que revela o potencial do modelo em identificar os casos relevantes. Em sistemas médicos, a Revocação aponta a quantidade de casos positivos dentre todos que são de fato positivos. Assim, essa medida é relevante por mostrar a confiabilidade do algoritmo em não perder casos de interesse. A equação seguinte descreve a Revocação (*REV*):

$$REV = \frac{TP}{TP + FN}, \quad (7.7)$$

em que *TP* e *FN* são as taxas de verdadeiros positivos e falsos negativos, respectivamente. Os valores de *REV* estão definidos no intervalo  $[0, 1]$ , com 1 representando o melhor caso.

A medida  $F$  (do inglês, *F-measure*) (BLAIR, 1979; POWERS, 2011) permite uma análise em conjunto das medidas Precisão e Revocação, dada por meio de uma média harmônica entre ambas. É possível ajustar essa medida para dar mais importância à Precisão ou à Revocação. Nesta tese, é computado o  $F_1$ , no qual  $PREC$  e  $REV$  são igualmente considerados. Essa medida é particularmente interessante para avaliar algoritmos em base de dados com desbalanceamento de classes. Quando Precisão e Revocação apresentam valores próximos,  $F_1$  se aproxima da média aritmética das duas. Formalmente:

$$F_1 = 2 \cdot \frac{PREC \cdot REV}{PREC + REV}, \quad (7.8)$$

com  $F_1 = 1$ , representando um modelo ideal e  $F_1 = 0$ , caso contrário.

Assumimos como GT da imagem (campo) a classe da célula de maior relevância clínica. Sendo assim, se uma determinada imagem contiver, por exemplo, uma única célula de alto risco e todas as outras negativas, o GT dessa imagem é assumido como sendo de alto risco. O GT local é obtido a partir da classe de maior importância dentro da região  $100 \times 100$  pixels anotada na base CRIC (Figura 3). Quando a ROI predita não casa com uma região rotulada pelo especialista, ela é desconsiderada da avaliação local. No entanto, essas ROIs participam da predição da classe da imagem.

A medida *Hit Degree* (HD) (ZHANG *et al.*, 2018) quantifica o nível de sobreposição entre uma ROI predita e a rotulada, avaliando a precisão do algoritmo em localizar e posicionar ROIs. Para uma imagem  $X$  com dimensões  $(p, q)$ , temos  $A$  como sendo uma matriz binária de dimensões  $(p, q)$  sinalizando a área  $100 \times 100$  da ROI e  $B$  como matriz binária representando regiões  $100 \times 100$  referentes a todas as células rotuladas em  $X$ . Formalmente, temos:

$$HD = \frac{A \cap B}{A}, \quad (7.9)$$

onde  $HD \in [0, 1]$ .

### 7.2.3 Análise da Hipótese H3

Em conformidade com os achados descritos na Seção 7.1.5 e na Tabela 5, a Figura 36 apresenta alguns dos resultados obtidos pelo modelo AGMic, considerando apenas as quatro regiões mais salientes. O ranqueamento por saliência priorizou áreas de células com lesão, levando a uma redução de amostras que precisam ser classificadas. Dessa forma, uma parte

significativa da imagem é descartada. De fato, clinicamente a existência de uma única célula com lesão é suficiente para reportar anormalidade. Observamos também que eventuais erros de falso negativo no Módulo 2 podem ser compensados em outras ROIs, devido o ranqueamento por saliência priorizar regiões potencialmente anormais em regiões da imagem que são conspícuas aos citopatologistas.

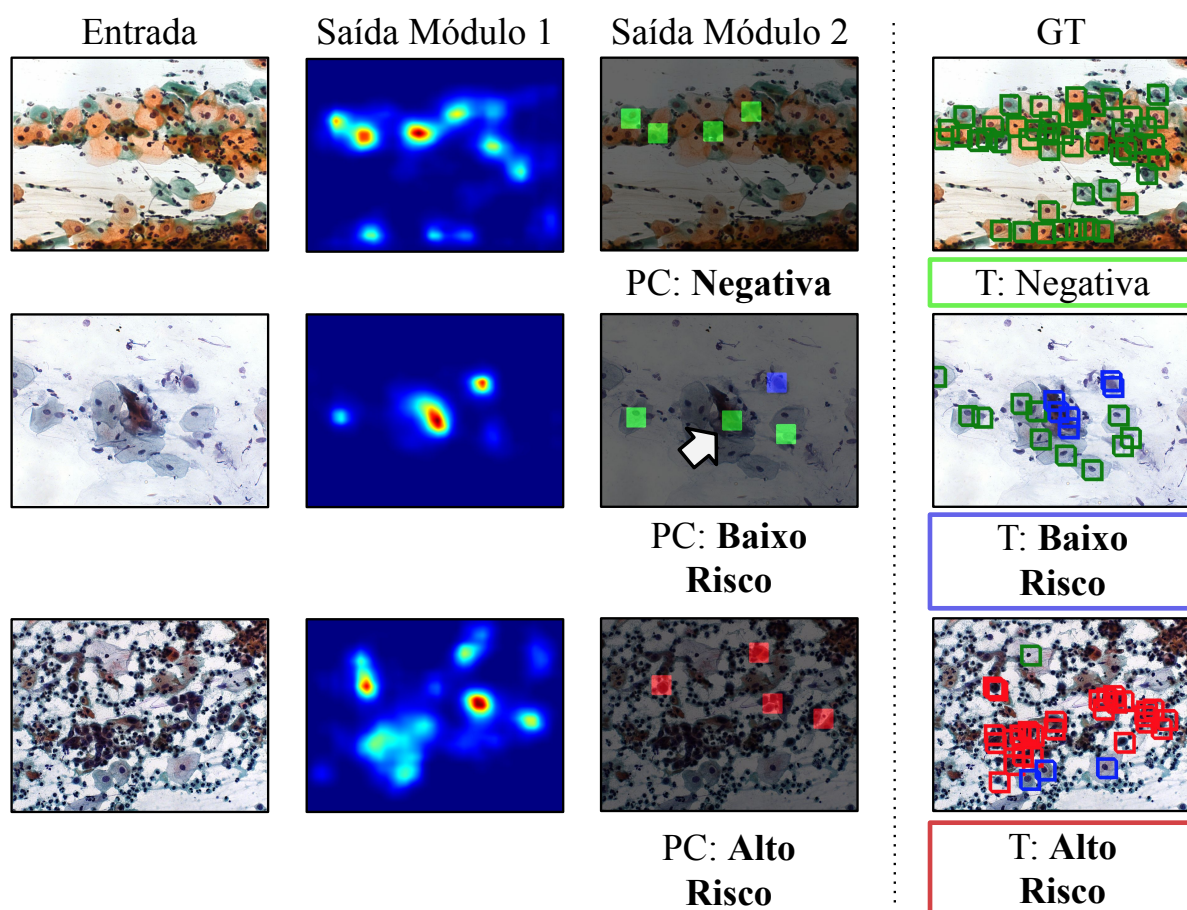


Figura 36 – Resultados do modelo AGMic. A seta indica exemplo de falso negativo local que não interfere na classificação da imagem. PC: Classe predita para a imagem. T: Rótulo da imagem a partir dos citologistas (GT). Fonte: elaboração própria a partir de amostras da base de dados CRIC.

Avaliamos quantitativamente o modelo AGMic variando a quantidade de ROIs salientes consideradas em cada imagem. Assim, cada posição de ranqueamento ( $pRanq$ ) contém o desempenho do algoritmo com diferentes quantidades de ROIs. Uma vez que as ROIs estão ranqueadas por saliência, foi possível analisar o comportamento do modelo proposto quando apenas as regiões mais conspícuas são processadas. Por exemplo, para  $pRanq = 1$ , o desempenho do AGMic foi avaliado apenas com o resultado de classificação da região mais saliente em cada imagem. Para  $pRanq = 3$ , as três regiões mais salientes foram consideradas, e assim por diante.



Vale salientar que a quantidade de ROIs detectadas pelo Módulo 1 pode ser diferente em cada imagem. Os experimentos com as imagens da base CRIC revelaram a existência de até 12 regiões salientes em algumas imagens.

A Figura 37 apresenta os resultados de Acurácia ( $ACC$ ) e Kappa ( $K$ ) obtidos pelo AGMic na base de dados CRIC. Três classes foram consideradas, negativa, baixo risco e alto risco, conforme agrupamento definido na Tabela 1. A classificação da imagem corresponde à alteração predita de maior relevância. Os resultados apontaram que o melhor desempenho de classificação é obtido nas quatro regiões mais salientes, destacada pela área cinza. Nesta faixa cinza, as medidas Acurácia e Kappa alcançam os maiores valores e decrescem nas regiões seguintes. Esse achado confirma que uma classificação mais precisa dos campos pode ser alcançada a partir do ranqueamento de ROIs por saliência. Outra observação é que quanto maior  $pRanq$ , maior o número de ROIs processadas por imagem e conseqüentemente maiores as chances de erro do Módulo 2. Para  $pRanq = 1$ , apenas a região mais saliente de cada imagem é processada. Para  $pRanq = 2$ , o número de ROIs se aproxima do dobro de  $pRanq = 1$ , e assim por diante. Esse acréscimo de erro na classificação das ROIs ao longo de  $pRanq \leq 4$  sustenta o achado da Figura 36 de que eventuais falsos negativos de ROIs tendem a não prejudicar a classificação da imagem, pois podem ser compensados pelo sucesso de classificação em outras regiões salientes.

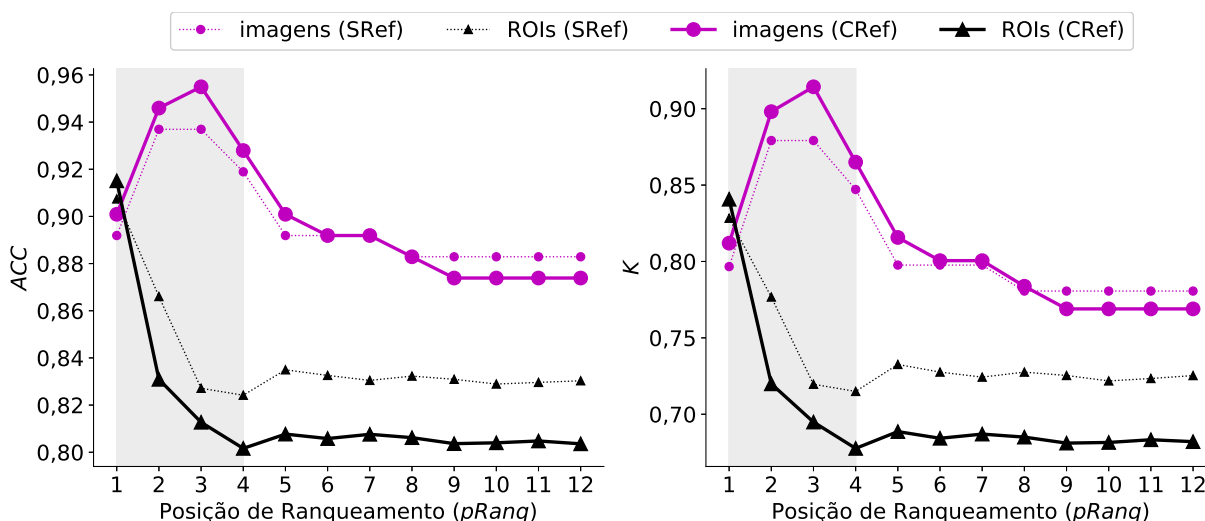


Figura 37 – Resultados do modelo AGMic para classificação das imagens e das ROIs em: negativa, baixo risco e alto risco. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria.

Assumir a classificação da imagem como a classe de maior relevância dentre as

ROIs pode levar a erros de falsos positivos. Embora do ponto de vista clínico falsos positivos não sejam tão prejudiciais quanto falsos negativos, sua existência geralmente resulta em custos, seja por reanálise por especialistas ou por procedimentos desnecessários ao paciente (JANTZEN *et al.*, 2005). Para entender como isso acontece no modelo AGMic, a Figura 38 apresenta os gráficos de Precisão (*PREC*), Revocação (*REV*) e  $F_1$ , considerando os resultados de classificação para as imagens da base CRIC.

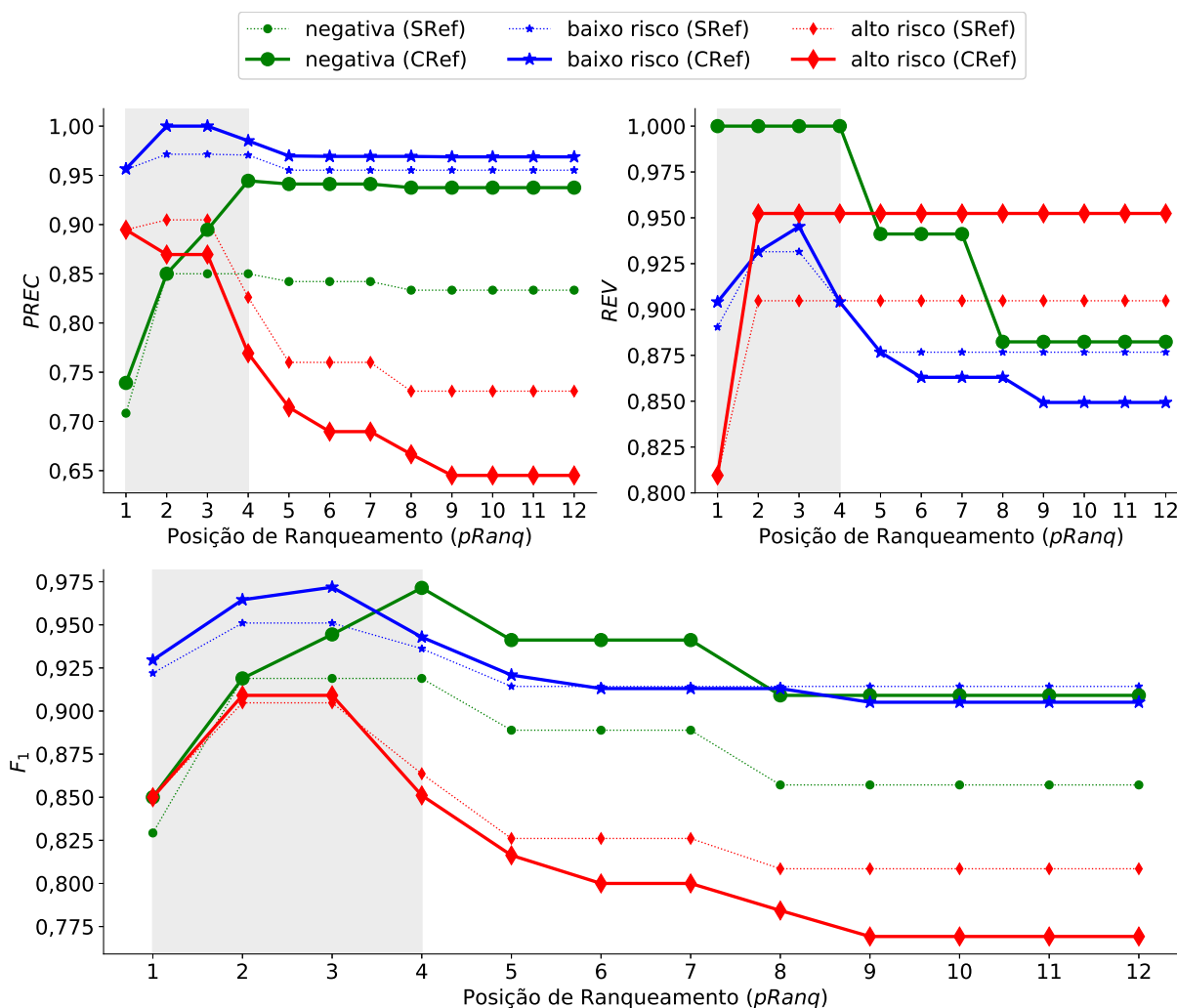


Figura 38 – Resultados do modelo AGMic para classificação dos campos (imagens) nas classes: negativa, baixo risco e alto risco. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria.

A medida  $F_1$  indica que a melhor combinação entre *PREC* e *REV* ocorreu com  $pRanq \leq 4$ , área dos gráficos destacada em cinza. Este resultado, em conjunção com as evidências da Figura 37, sustenta que o ranqueamento por saliência melhora o desempenho na classificação de imagens de células cervicais. Quanto à Revocação (*REV*), o modelo proposto alcançou seu

melhor desempenho em  $pRanq = 3$ , o que condiz com os achados na análise de H2 (Seção 7.1.5). Assim, com apenas 3 regiões salientes, o AGMic tem 95,23% de chances de identificar imagens com lesões de alto risco, com precisão de 86,95%, e de 94,52% para lesões de baixo risco, com precisão de 100%. Em termos de *REV*, o uso do algoritmo de refinamento alcançou ganhos de  $\sim 5,26\%$  no desempenho do AGMic quanto à classe de alto risco,  $\sim 1,47\%$  quanto a baixo risco. Também houve melhoria nos valores de *PREC* na classe de baixo risco em  $\sim 2,94\%$  e negativa em  $\sim 5,26\%$ , mesmo com perdas de  $\sim 3,89\%$  na precisão da classe de alto risco. Isso sugere que a etapa local (refinamento) do AGMic torna o método mais sensível a casos anormais, em especial ao alto risco. Como *REV* reflete a sensibilidade do método em identificar casos alterados, os resultados mostraram que o algoritmo de refinamento reduz as chances de perder casos anormais, mesmo com a observada ampliação de falsos positivos em casos de alto risco.

Sistemas de triagem automática para exames de Papanicolaou podem requerer classificação binária, marcando a imagem como positiva ou negativa. De fato, encontrar atributos discriminantes entre as diferentes classes de células anormais é uma tarefa desafiadora pelo fato de imagens representantes dessas classes compartilharem de atributos similares quanto à anormalidade, como textura do núcleo de aspecto granular, tamanho do núcleo aumentado em relação ao citoplasma, dentre outros. Assim, a Figura 39 mostra os resultados do AGMic com as classes de baixo risco e alto risco agrupadas em uma única classe, a qual denominamos *positiva*.

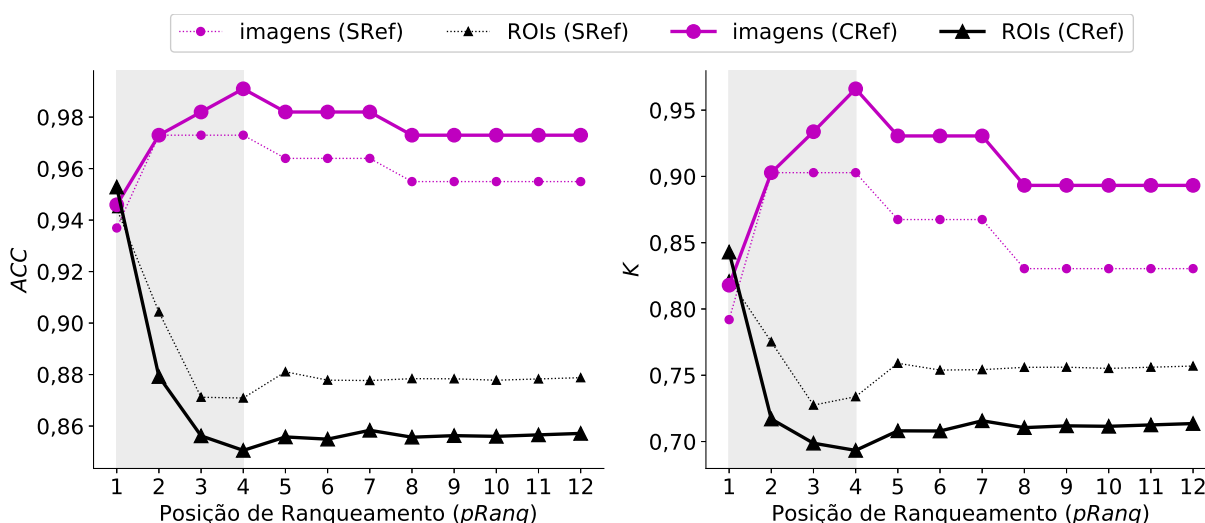


Figura 39 – Resultados do modelo AGMic para classificação binária: imagens positivas e negativas. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Ambos os gráficos apresentam o desempenho do modelo proposto na classificação dos campos (imagens) e das ROIs localmente. Fonte: elaboração própria.

Assim como no problema de três classes (alto risco, baixo risco e negativa - Figuras 37 e 38), o melhor desempenho do modelo proposto ocorreu para  $pRanq \leq 4$ . Processando as quatro regiões mais salientes, o AGMic alcançou  $ACC = \sim 0,99$  na classificação das imagens e  $ACC = \sim 0,85$  na classificação das ROIs. Para o índice Kappa, os valores foram  $K = \sim 0,96$  e  $K = \sim 0,70$  para imagens e ROIs, respectivamente. Esses resultados sugerem robustez do AGMic quanto a eventuais erros de classificação das ROIs pelo Módulo 2, uma vez que, as taxas de acerto para as imagens foi superior a das ROIs. No cenário de classificação binária, os ganhos do algoritmo de refinamento foram mais evidentes. Houve acréscimos de  $\sim 7\%$  em  $K$  na classificação binária das imagens contra  $\sim 4\%$  ( $pRanq = 3$ ) quando três classes são consideradas (Figura 37).

A Figura 40 apresenta  $PREC$ ,  $REV$  e  $F_1$  para o cenário de classificação binária. Como na classificação de três classes (Figura 38), o modelo proposto obteve seu melhor desempenho processando apenas as quatro regiões mais salientes (área destacada em cinza nos gráficos). Com  $pRanq = 4$ , o AGMic alcançou  $F_1 = 0,995$  para a classe positiva, detectando  $\sim 98\%$  das imagens alteradas com precisão máxima, ou seja, sem gerar falsos positivos. Este resultado, em conjunto com a Figura 38, mostra que Módulo 2 tende a diferenciar melhor células anormais de normais, e confunde com alguma frequência as amostras de baixo risco e alto risco.

A Figura 40 mostra também que o algoritmo de refinamento no Módulo 1 promoveu ganhos de  $2,19\%$  nos valores de Revocação da classe positiva. É interessante notar que, fora da área cinza, os valores de  $REV$  nesta classe permanecem constantes, uma vez que a imagem é positiva se foi estimada alteração em alguma das ROIs. No entanto, a partir de  $pRanq = 4$ , a curva de  $PREC$  é decrescente, revelando que áreas ruidosas da imagem podem ter sido inseridas no processamento, levando à queda de desempenho do Módulo 2. Esses achados, em conjunção com os discutidos anteriormente, reforçam que a proposta de ranquear regiões por saliência, processando apenas as ROIs mais salientes, reduz a ocorrência de falsos positivos e negativos na identificação de alterações celulares em exames de Papanicolaou.

O algoritmo de refinamento<sup>16</sup> permitiu alcançar melhores resultados na classificação das imagens (campos) para todas as medidas analisadas (Figuras 37, 38, 39 e 40). Isso ocorreu pelo melhor alinhamento da ROI à célula de interesse<sup>17</sup>, aproximando as coordenadas de recortes às mesmas utilizadas na geração das amostras para treinamento do Módulo 2 (Figura 28). Esse melhor ajuste é também verificado nos resultados da medida HD (Equação 7.9), os quais estão

<sup>16</sup> Etapa local do Módulo 1 do AGMic.

<sup>17</sup> Problema descrito na Figura 17.

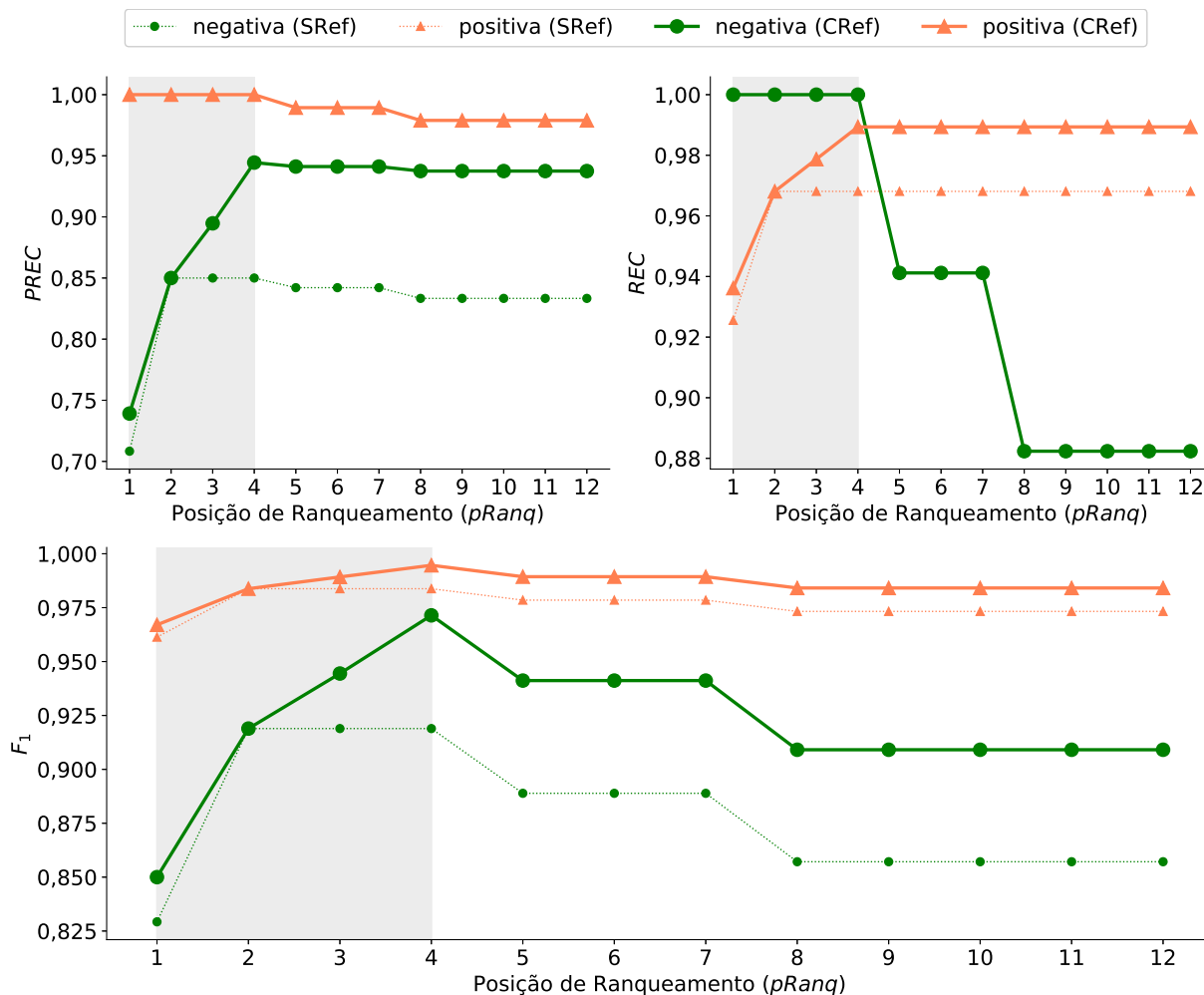


Figura 40 – Resultados do modelo AGMic para classificação binária das imagens: positiva e negativa. SRef e CRef correspondem aos resultados sem e com o algoritmo de refinamento local, respectivamente. Fonte: elaboração própria.

exibidos na Figura 41 para todas as imagens de teste. Para esta medida, o método de refinamento permitiu alcance de  $\sim 10,77\%$  de ganho na mediana dos valores observados, além de deslocar o primeiro quartil de 0,555 para 0,616.

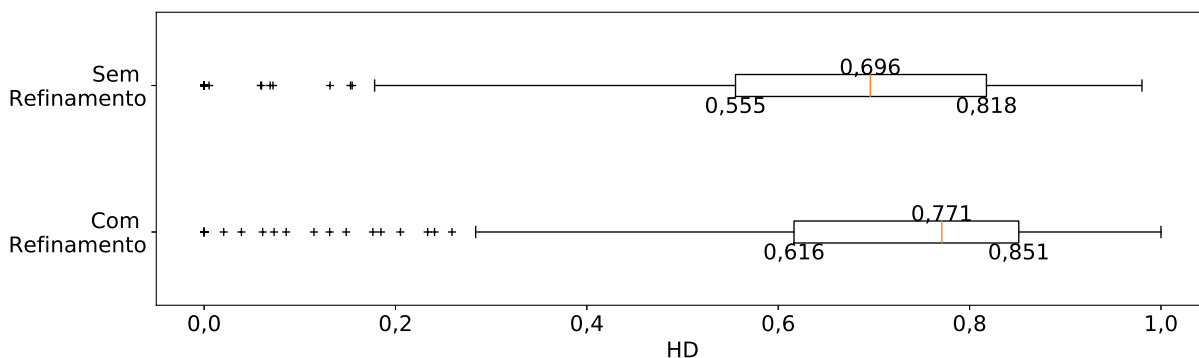


Figura 41 – Resultados da medida HD. O uso do algoritmo de refinamento permitiu que o modelo proposto aumentasse a sobreposição entre as ROIs estimadas e as células rotuladas da base CRIC. Fonte: elaboração própria.

A Figura 42 apresenta outra contribuição do algoritmo de refinamento: redução da quantidade de ROIs detectadas em regiões que não foram rotuladas pelos especialistas. O uso do algoritmo de refinamento melhorou o desempenho de localização em 1,6%, concentrando maior parte na detecção de novas células normais e 0,3% em células com lesões de alto risco. Estes resultados mostram que a análise de textura dos núcleos é valiosa para identificar células de interesse, aproximando o posicionamento de recortes aos critérios visuais adotados pelos citopatologistas. Especificamente, os resultados sugerem que a matriz rSCM proposta nesta tese (Seção 6.1.2.1) pode ser aplicada satisfatoriamente entre diferentes bases de dados de células cervicais, desde que o algoritmo de refinamento foi concebido e treinado a partir de amostras da base Herlev e aplicado nas imagens do CRIC.

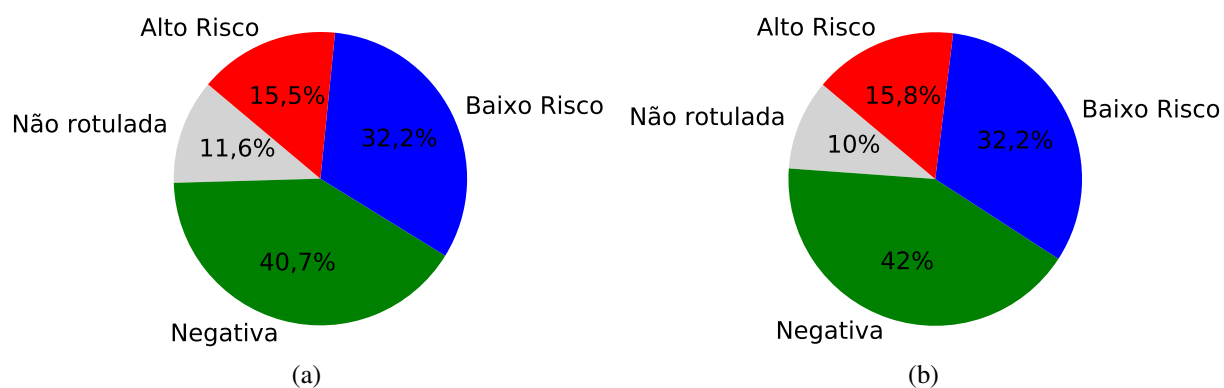


Figura 42 – Quantitativo de ROIs estimadas por GT. a) mostra a distribuição sem uso do algoritmo de refinamento e em b) com uso. O algoritmo de refinamento permitiu uma redução de 1,6% em regiões não rotuladas e o aumento de 0,3% para células com alto risco. Fonte: elaboração própria.

Posicionamos o modelo AGMic na literatura comparando os melhores resultados do mesmo com a estratégia clássica de busca automatizada por lesões em células cervicais, a qual baseia-se em segmentação/detecção de núcleos. Uma vez que este trabalho é pioneiro em utilizar predição de saliência para seleção de ROIs em exames de Papanicolaou para efeito de classificação de células, não foi possível compará-lo a algoritmos similares. Arquiteturas CNN baseadas na proposição de regiões *Region Proposal Networks* (RPNs)<sup>18</sup> requerem grandes volumes de imagens rotuladas para serem treinadas, o que impossibilitou sua inserção nos experimentos. Sendo assim, estudamos o modelo proposto da seguinte forma: 1) verificando o desempenho do Módulo 2 considerando a situação ideal onde todas as células são detectadas e 2) avaliando o desempenho do Módulo 2 quando um algoritmo de segmentação é utilizado.

<sup>18</sup> Estão nessa categoria os seguintes modelos: R-FCN (DAI *et al.*, 2016), Faster-RCNN (REN *et al.*, 2017), Yolo (REDMON; FARHADI, 2017).

Em ambos os casos, o Módulo 2 é considerado para reduzir o viés do modelo de classificação, permitindo-nos concluir sobre o ranqueamento por saliência. A Tabela 6 exhibe os resultados alcançados para as medidas Acurácia ( $ACC$ ) e Kappa ( $K$ ) considerando três classes: negativa, baixo risco e alto risco. Para melhor analisar o AGMic quanto à aplicabilidade em laboratórios de análise clínica, o tempo de processamento<sup>19</sup> (em segundos) em uma máquina Intel (I7-4770HQ) CPU (2.2 GHz) 16 GB RAM, sem uso de GPU, é reportado. Quando utilizada GPU, o tempo de processamento do Módulo 1 (etapa global) reduz de  $\sim 3,23$  para  $\sim 0,06$  segundos por imagem e o do Módulo 2 de  $\sim 0,96$  para  $\sim 0,02$ , com  $pRanq = 3$ . Dado que a etapa de refinamento adiciona  $\sim 1,07$  segundos ao modelo, o AGMic (processando apenas as três regiões mais salientes<sup>20</sup>) classifica uma imagem em tempo médio inferior a 1,2 segundos, incluindo o refinamento, com GPU.

Tabela 6 – Resultados comparativos entre o AGMic e a estratégia clássica de busca por lesões em imagens de células cervicais. Classificação das imagens (campos) e das ROIs localmente em três classes: negativa, baixo risco e alto risco. Os valores para a medida Acurácia ( $ACC$ ) e para o índice de concordância Kappa ( $K$ ) são reportados. Fonte: elaboração própria.

	Imagens		ROIs		tempo médio (s)
	$ACC$	$K$	$ACC$	$K$	
GT: todas as células	0,775	0,611	0,782	0,631	5,96 ( $\pm 4,74$ )
Segmentação: detecção de núcleos por Bora <i>et al.</i> (2017)	0,495	0,302	0,722	0,360	9,84 ( $\pm 9,98$ )
Saliência: AGMic (Proposta)* - Sem Refinamento	0,937	0,879	<b>0,827</b>	<b>0,720</b>	<b>4,16</b> ( $\pm 0,27$ )
Saliência: AGMic (Proposta)* - Com Refinamento	<b>0,955</b>	<b>0,919</b>	0,813	0,695	5,26 ( $\pm 0,29$ )

\* Consideramos AGMic com  $pRanq = 3$  regiões salientes.

A etapa de refinamento corresponde a  $\sim 1,07$  (s) do tempo de processamento.

<sup>19</sup> Os algoritmos desenvolvidos nesta tese não foram otimizados quanto ao tempo de execução.

<sup>20</sup> O modelo AGMic alcança melhores resultados de classificação, nas classes negativa, baixo risco e alto risco, processando apenas as três regiões de maior saliência (Figuras 37 e 38).

### 7.3 Experimentos com Citologia em Meio Líquido

Esta seção analisa o desempenho do Módulo 1 do AGMic, codificado a partir dos dados de atenção visual da base CRICVA, na seleção e ranqueamento de ROIs em imagens de células cervicais em meio líquido. Não houve treinamento do modelo neste conjunto de dados, ocorrendo, portanto, transferência de aprendizado da citologia convencional para a citologia em meio líquido.

#### 7.3.1 Base de Imagens

Os experimentos foram realizados na base de dados *Deep Cervical Cytological Lesions (Deep Cervical Cytological Lesions (DCCL))*, introduzida por Zhang *et al.* (2019). Essa base possui amostras de 1.167 lâminas, sendo 933 de pacientes positivos e 234 de casos normais, todas em meio líquido. Os exames foram coletados por quatro centro médicos chineses entre os anos de 2016 e 2018 e digitalizados por um dos três equipamentos (Nanozoomer2.0HT, KFBIO KF-RPO-400 e AperioAT2) com ampliação de 200x e em cores de 24 *bits*.

A Figura 43 apresenta algumas imagens da base de dados DCCL. Essas imagens foram extraídas a partir de recortes de 2.000 x 1.200 *pixels* das lâminas digitalizadas. Dessas imagens, 5.874 estão rotuladas e disponíveis no conjunto de dados. Os rótulos foram determinados de acordo com a avaliação de seis especialistas, oferecendo para cada imagem a classe do achado clínico (ou normal, para casos negativos) e a localização de células de interesse. Os especialistas participantes tinham oito anos de experiência em média e realizaram a análise dos campos assumindo o sistema Bethesda (Tabela 1).

Os seguintes procedimentos foram realizados para habilitar a base DCCL a este trabalho: 1) Recorte<sup>21</sup> das imagens, produzindo uma amostra de 1.388 x 1.040 *pixels* para cada. Isso é necessário para manter a entrada dos algoritmos com o mesmo tamanho das imagens da base CRIC. Apenas as imagens com células rotuladas dentro da área do recorte foram consideradas. 2) Desconsideramos imagens com ROIs de diferentes classes (negativa, baixo risco e alto risco). Diferentemente da base CRIC, onde grande parte das células estão rotuladas (Figura 3), a base DCCL identifica apenas as células com maior suspeita de alteração. Assim, optamos por imagens onde a classe das regiões selecionadas são consistentes.

No total, 4.867 imagens da base DCCL foram utilizadas nos experimentos. As

---

<sup>21</sup> Procedimento realizado assumindo como ponto inicial o *pixel* de posição (0,0).



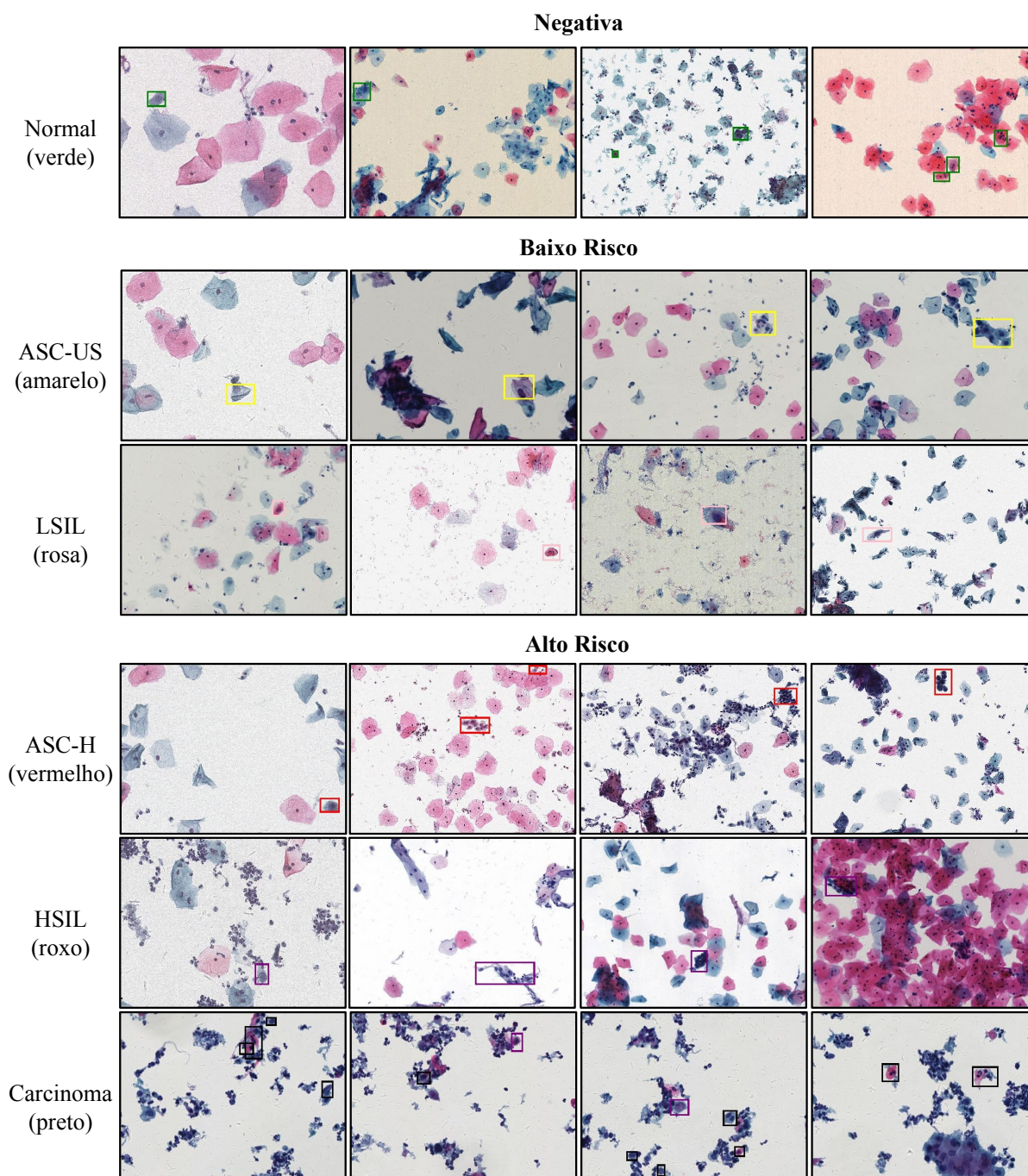


Figura 43 – Imagens de células cervicais em meio líquido da base DCCL. Os rótulos na vertical sinalizam o rótulo da imagem, enquanto que as marcações em cores localizam a alteração celular, se existente. Fonte: elaboração própria.

imagens estão distribuídas em #{negativa: 1.292, baixo-risco: 2.341, alto-risco: 1.234}<sup>22</sup>. Temos 9.723 regiões rotuladas<sup>23</sup>, identificando partes da imagem com interesse clínico. As ROIs negativas foram selecionadas como aquelas que podem ser facilmente confundidas como alteradas. A Figura 44 detalha a distribuição de classe das imagens, bem como de suas ROIs.

<sup>22</sup> Detalhadamente: #{Normal: 1.292, ASC-US: 1.700, LSIL: 641, ASC-H: 728, HSIL: 328 e CA: 178}.

<sup>23</sup> Detalhadamente: #{Normal: 2.713, ASC-US: 2.341, LSIL: 1.108, ASC-H: 1.093, HSIL: 1.642 e CA: 826}.

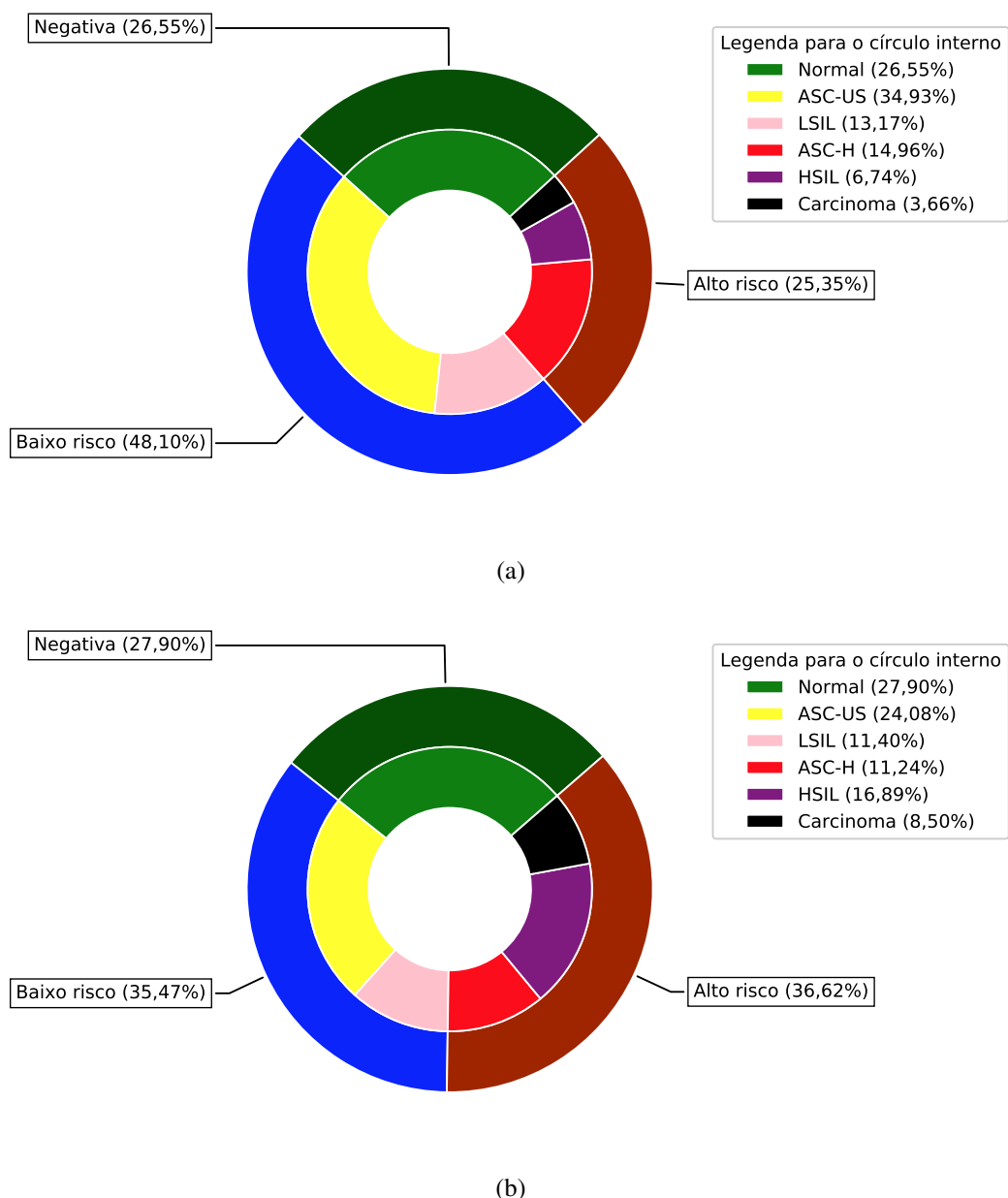


Figura 44 – Distribuição das amostras da base DCCL. a) Quanto à classificação da imagem. Total de imagens: 4.867. b) Quanto às ROIs. Total de regiões rotuladas: 9.723. Fonte: elaboração própria.

### 7.3.2 Medidas de Avaliação

Os algoritmos propostos foram analisados quanto ao ranqueamento e posicionamento das ROIs estimadas. Para quantificar o ranqueamento, computamos a frequência com que as regiões de maior saliência coincidem com as áreas definidas pelos especialistas.

Para avaliar o posicionamento das ROIs, calculamos a interseção entre uma área de 100 x 100 *pixels* estimada e a marcação do GT. Utilizamos a medida HD (Equação 7.9), assumindo a variável *A* como a região de menor área. Essa condição é necessária devido à

necessidade de considerar total sobreposição se a ROI estimada estiver completamente sobreposta à marcação do GT. Na base CRIC, tanto as ROIs estimadas quanto o GT possuem dimensão 100 x 100 *pixels*, enquanto que na base DCCL as regiões possuem diferentes tamanhos.

### 7.3.3 *Análise*

A Figura 45 apresenta alguns resultados do Módulo 1 do AGMic em imagens da base DCCL. Assim como observado para a citologia convencional (base CRIC), o modelo de saliência encontrou ROIs em diferentes partes da imagem, evitando áreas com bastante células aglomeradas nas primeiras posições de ranqueamento. Isso permitiu o modelo sugerir células em regiões de boa visualização. A etapa local alcançou ROIs com melhor posicionamento em relação ao GT. O não uso do refinamento potencializou o problema descrito na Figura 17, estando, portanto, consistente com os resultados obtidos na base CRIC.

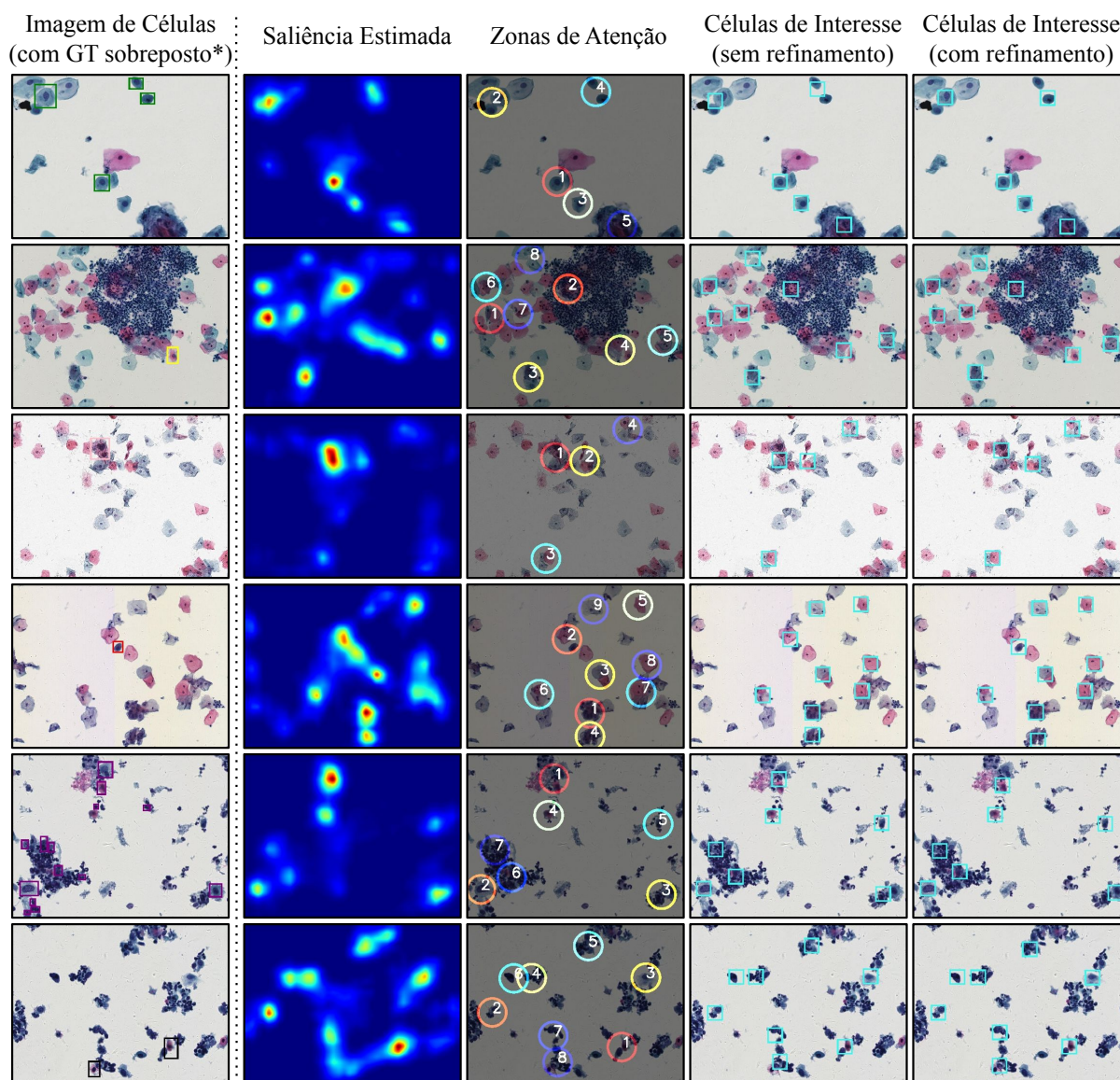
A Figura 46 apresenta os resultados quantitativos quanto ao ranqueamento. Um total de 3.195 imagens das 4.867 analisadas apresentaram ROIs entre as regiões salientes. Mesmo quando a região rotulada corresponde à última posição no ranque<sup>24</sup> ( $pRanq = 17$ ), a área de busca por células alteradas fica restrita a apenas 29,97% do campo visual da imagem. Essa ganho ficou ainda mais significativo para 23,92% dos casos, onde as células anotadas estão contidas na região de maior saliência, a qual representa somente 1,76% da imagem.

Aproximadamente 34,35% das imagens analisadas não possuem ROIs rotuladas dentre as regiões salientes. Da população de negativas, baixo risco e alto risco, temos 34,67%, 28,06% e 45,94%, respectivamente. No entanto, é provável que existam células de mesma classe em outras regiões da imagem que não foram identificadas no GT pelos citopatologistas. Em outras palavras, para ter certeza que a região do achado clínico não foi detectada pelo modelo proposto é necessário ressubmeter as imagens para análise por citopatologistas para saber se a alteração rotulada na imagem não se repete em alguma das regiões destacadas no mapa de saliência, visto que a base DCCL não possui rótulos para todas as células do campo.

A Figura 47 mostra os resultados da medida HD quanto à precisão de posicionamento das ROIs. Observando a mediana, o algoritmo de refinamento<sup>25</sup> permitiu um acréscimo de 9,97% na área de interseção entre a região estimada e o GT, quando comparado a ROIs centradas no ponto de máxima saliência. Houve também um deslocamento do primeiro quartil de 0,352 para

<sup>24</sup> Os experimentos apontaram que os mapas de saliência estimados conseguem distinguir no máximo 17 zonas de atenção na base DCCL.

<sup>25</sup> Etapa local do Módulo 1 do AGMic.



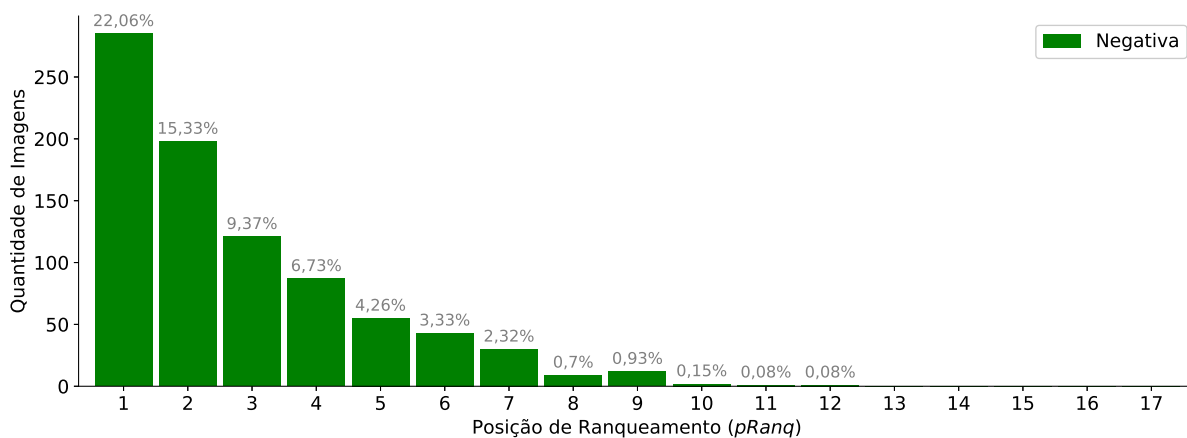
\* De forma ilustrativa, essa figura sobrepõe a imagem processada ao rótulo (GT) anotado na base DCCL.

Figura 45 – Resultados do Módulo 1 do AGMic em cada classe da base DCCL. As cores do GT representam: verde: normal, amarelo: ASC-US, rosa: LSIL, vermelho: ASC-H, roxo: HSIL e preto: carcinoma. As zonas de atenção estão ranqueadas pela energia do mapa de saliência. O refinamento (etapa local do AGMic) tende a alinhar melhor as células selecionadas à posição definida pelos citopatologistas. Fonte: elaboração própria.

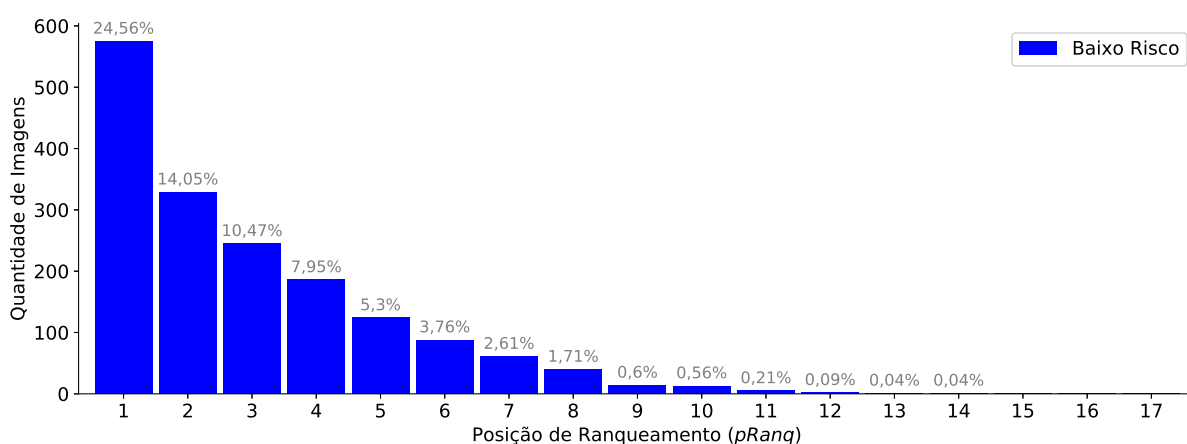
0,387. Esses achados são consistentes com os reportados na Figura 41, mostrando que o uso da matriz rSCM, em combinação com a metodologia descrita na Seção 6.1.2.2, alcança ROIs mais próximas às células de interesse. O tempo médio de processamento<sup>26</sup> desta etapa foi de 2,23 (SD=1,03) segundos por imagem.

Devido à base DCCL identificar suas ROIs em diferentes tamanhos, não foi pos-

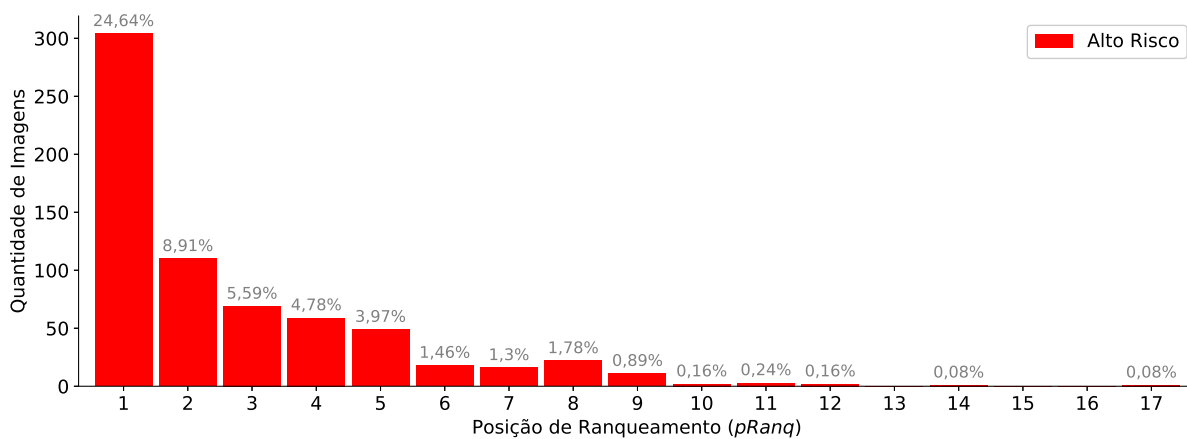
<sup>26</sup> Os algoritmos desenvolvidos neste trabalho não estão otimizados quanto ao custo computacional.



(a)



(b)



(c)

Figura 46 – Distribuição das regiões rotuladas pelos especialistas no ranque das ROIs preditas, conforme energia no mapa de saliência. Aproximadamente a) 65,33% das imagens negativas, b) 71,94% das imagens de baixo risco e c) 54,05% das imagens de alto risco possuem as células sinalizadas nas regiões mais salientes. As porcentagens em cinza indicam a proporção de imagens, assumindo como referência o total de 1.292, 2.341 e 1.234 imagens negativas, baixo risco e alto risco processadas, respectivamente. Fonte: elaboração própria.

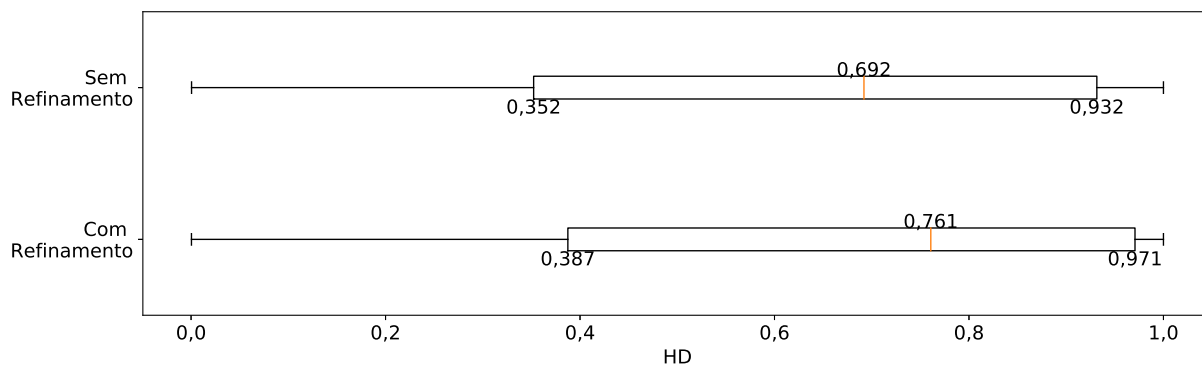


Figura 47 – Resultados da medida HD. O algoritmo de refinamento permitiu aumentar a sobreposição entre as ROIs estimadas e as células rotuladas da base DCCL. Fonte: elaboração própria.

sível treinar e avaliar o Módulo 2 do AGMic. Recortes dentro das ROIs rotuladas necessitam de supervisão humana para que estruturas celulares de interesse, como núcleo, possam estar inteiramente contidas no recorte. Como a literatura carece de trabalhos que utilizam dados de atenção de citopatologistas para seleção de ROIs, não foi possível estabelecer referência para estudos comparativos.

## 8 CONCLUSÕES E TRABALHOS FUTUROS

Esta tese propõe um modelo computacional para análise de imagens de células cervicais guiado por saliência. O objetivo principal foi investigar se a codificação da atenção visual de citopatologistas experientes pode contribuir para a triagem de regiões clinicamente relevantes, destinando rotinas de categorização a partes específicas da imagem. Essa abordagem tem como vantagem a priorização de áreas de busca por lesões nas imagens, evitando erros de classificação devido ao processamento de regiões confusas, além de oferecer ganho computacional no processamento. A partir da confecção de uma base de dados de atenção visual de citopatologistas, investigamos 13 algoritmos de predição de saliência, sendo 1 método *top-down* que explora características do núcleo celular e 2 variações de um modelo CNN, os quais codificamos para reconhecer atenção *top-down* em imagens de células cervicais. A partir da fusão de modelos de aprendizado profundo a mapas de saliência que reproduzem a atenção de especialistas, desenvolvemos e implementamos um modelo que dispensa a segmentação celular, identificando padrões anormais em microscopia de Papanicolaou sob condições ruidosas reais, artefatos e oclusão. Dessa forma, esta tese alcança seu objetivo principal.

Os resultados revelaram que fatores *top-down* guiam a atenção dos citopatologistas e podem ser codificados para apontar regiões de interesse clínico em sistemas computacionais. Fatores *bottom-up* também indicam células anormais, no entanto, com maiores taxas de falsos positivos. A análise do mapa de ativações das redes neurais convolucionais (VGG-16 e Resnet-50) mostrou que as características de baixo nível modeladas para a predição de saliência em imagens citológicas se assemelham às obtidas em imagens naturais. Entre esses domínios (imagens naturais e de citologia), observamos correlação superior a 95% na primeira metade das camadas convolucionais. Assim, contribuímos com uma estratégia que sugere a quantidade de camadas que devem ser ajustadas para a predição de saliência em imagens de células cervicais por meio de *fine-tuning*. Esse achado também revela que a transferência de aprendizagem a partir de imagens naturais permite que metodologias baseadas em CNN obtenham mapas de saliência precisos em imagens de citologia do colo do útero. Trabalhos futuros podem explorar semelhanças entre imagens naturais e de células cervicais como parte de esquemas para abordar a interpretabilidade da CNN.

A codificação da atenção *top-down* coletada pelo rastreamento ocular levou a algoritmos menos susceptíveis a falsos positivos. Por outro lado, os algoritmos que usam apenas características *bottom-up* tenderam a marcar regiões de alto contraste como salientes, especi-

almente regiões de transição entre agrupamentos de células e o fundo da imagem. Elementos escuros, como núcleos, neutrófilos, alguns artefatos de lâmina ou mesmo agrupamentos de células (Figura 2), encontraram-se comumente salientes por técnicas baseadas na análise no domínio da frequência. Assim, concluímos que os algoritmos analisados que realizam análise no domínio da frequência não são adequados para predição de saliência em células cervicais com propósitos de busca por lesões, pois leva ao processamento desnecessário de regiões em áreas de artefatos e sobreposição celular. No entanto, nossos resultados sugerem que alguns algoritmos, em especial o SIC, podem contribuir em aplicações que visem a remoção do fundo da imagem, por exemplo. Dentre os métodos *bottom-up*, os melhores resultados foram alcançados por técnicas que realizam análise de regiões com base em sua vizinhança (paradigma *center-surround*), extraindo características no domínio espacial que possam ser relacionadas com aquelas provenientes de outras partes da imagem por meio de *frameworks* estatísticos. Embora apresente tendência a falsos positivos, os experimentos realizados revelaram o potencial de alguns métodos em predizer regiões de atenção dos citopatologistas, mesmo sem qualquer informação sobre o contexto de células cervicais. Assim, assumimos confirmada a Hipótese H1<sup>1</sup>.

Quanto à extração de ROIs, o modelo proposto confirmou aplicabilidade para todos os padrões anormais estudados, destacando a principal alteração celular do campo nas regiões mais salientes da imagem. Para imagens com ocorrência de alterações LSIL, ASC-H e HSIL, os resultados revelaram também métodos *bottom-up* com resultados promissores, a um custo de até 6,8% do tempo de processamento dos métodos de aprendizado profundo. Estes algoritmos correspondem a métodos rápidos que oferecem predição satisfatória quanto à localização dessas lesões, apesar da maior probabilidade de falsos positivos. Dessa forma, esses métodos são opções para cenários onde CNNs são inviáveis por indisponibilidade de recursos computacionais apropriados para processamento. Outro achado é que a atenção *top-down* dos citopatologistas, codificada computacionalmente, pode ser aplicada com precisão satisfatória para imagens de citologia em meio líquido. Para essas imagens, a correspondência entre as regiões rotuladas manualmente e as ROIs salientes preditas foi promissora. Isso mostra que as estruturas da imagem que são atrativas para especialistas são comuns nos dois meios de coleta, e os atributos dessas estruturas podem ser capturados computacionalmente. Sendo assim, diante dos resultados alcançados tanto para as imagens de citologia convencional quanto para as de meio líquido,

---

<sup>1</sup> Hipótese H1: Métodos *bottom-up* apresentam mapas de saliência comparáveis aos definidos pela atenção visual de especialistas humanos.



confirmamos a Hipótese H2<sup>2</sup>.

Classificar clinicamente um campo de citologia cervical com base no ranqueamento de ROIs por saliência melhora a sensibilidade e a precisão na identificação das imagens anormais. Isso acontece porque o conhecimento dos citopatologistas, codificado computacionalmente, previne agrupamentos de células, onde a análise é bem conhecida pelos citologistas como imprecisa. Assim, rotinas de classificação precisam lidar apenas com partes específicas da imagem, reduzindo as chances de ter seu desempenho comprometido por erros de categorização em áreas irrelevantes. O processamento restrito apenas às regiões mais salientes torna a categorização de imagens de exames de Papanicolaou convencional em três classes (negativa, baixo risco e alto risco) mais precisa, com ganhos em tempo de execução significativos em relação à abordagem clássica, em que todas as células do campo são analisadas. Também observamos que a codificação da atenção *top-down* dos especialistas a partir de imagens de citologia convencional permite priorizar áreas de busca por lesões em imagens também de meio líquido, bem como demonstra evitar aglomerados com forte sobreposição celular. Esses resultados, portanto, estão consistentes nos dois meios de coleta, apontando que a estratégia de ranquear ROIs por saliência é benéfica para a análise de células cervicais tanto na citologia convencional quanto na citologia em meio líquido. Dessa forma, confirmamos a Hipótese H3<sup>3</sup>.

A saliência estimada a partir do rastreamento ocular mostrou-se esparsa, ou seja, distribuída em uma região da imagem, não podendo assumir que o seu pico estará ajustado ao núcleo da célula (Figura 17). A partir da codificação de padrões de textura, por meio da proposta de uma matriz de coocorrência estrutural, desenvolvemos um algoritmo para sugerir ROIs salientes com melhor ajuste ao núcleo celular. Com este método de refinamento, ampliamos a sensibilidade de nosso modelo para os casos de alto risco e baixo risco, sem comprometer a precisão de triagem dos campos anormais. Assim, verificamos que a modelagem da textura do núcleo a partir das diferenças estruturais geradas pela aplicação de um filtro passa-baixa, habilita ajustes nos mapas de saliência preditos a partir de dados de rastreamento ocular. Isso permitiu um melhor casamento das regiões preditas com as anotações dos especialistas e, conseqüentemente, um melhor desempenho do modelo de classificação. Esse melhor ajuste no posicionamento das ROIs foi confirmado também para imagens de citologia em meio líquido, sugerindo que a matriz

<sup>2</sup> Hipótese H2: Modelos de predição de saliência podem detectar regiões clinicamente relevantes em imagens digitais de exames de Papanicolaou.

<sup>3</sup> Hipótese H3: O ranqueamento de regiões de interesse com base na predição da atenção visual de citopatologistas é uma estratégia computacional viável para a redução de erros de classificação de imagens de exames de Papanicolaou.

de coocorrência estrutural proposta pode ser aplicada satisfatoriamente para imagens obtidas em ambos meios de coleta. Adicionalmente, mostramos que a matriz proposta codifica a informação de cromatina do núcleo, da membrana nuclear e do entorno do núcleo, podendo ser aplicada como um descritor para classificação de células cervicais em normais e anormais (Apêndice B). Os experimentos apontaram que o descritor proposto alcança resultados equivalentes ao estado-da-arte em problemas de classificação, com a vantagem de depender apenas da segmentação do núcleo. Assim, alcançamos uma solução com potencial de ser incorporado a sistemas automáticos de auxílio à análise clínica, pois evita a etapa de segmentação do citoplasma que é geralmente desafiadora em imagens de citologia reais.

Esta tese é pioneira em reunir e publicar dados de atenção visual de citopatologistas experientes em imagens de células cervicais oriundas de exames de Papanicolaou convencional. Além da coleta dos dados de fixações oculares explorados nesta tese, os experimentos de rastreamento ocular permitiram a coleta de informações como movimentos sacádicos<sup>4</sup>, *dwell time*<sup>5</sup>, dentre outras, as quais serão publicados à medida que novos trabalhos forem desenvolvidos. Por fim, os experimentos de rastreamento ocular realizados, bem como seus achados, reforçam o potencial de trabalhos que se baseiam na captura da atenção humana para confecção de modelos computacionais.

## 8.1 Trabalhos Futuros

O modelo proposto corresponde a uma nova abordagem para classificação de imagens de células cervicais, utilizando como base zonas de atenção em análise similar à empregada por especialistas humanos. Nesta tese o potencial dessa abordagem é demonstrado, assim, novos estudos devem ser encaminhados com os seguintes propósitos:

- Otimizar parâmetros do modelo AGMic

Pesquisas futuras são recomendadas para investigar a existência de um conjunto de parâmetros que maximize o desempenho dos algoritmos propostos.

- Ampliar o conjunto de dados CRICVA

A base CRICVA corresponde a um avanço em bases de dados de *benchmark* para predição de saliência em imagens de células cervicais, apresentando à literatura dados de atenção

---

<sup>4</sup> Movimentos oculares rápidos que permitem mover a fóvea de um ponto de interesse para outro, ou seja, entre as fixações oculares. Desses movimentos, podem ser extraídas informações como amplitude, velocidade, duração, direção, dentre outros.

<sup>5</sup> Quantidade de tempo total que o participante permanece observando uma região de interesse da cena.

visual em amostras de exames do Sistema Único de Saúde (SUS). Novos trabalhos serão conduzidos para ampliar a quantidade de citopatologistas e aumentar o volume de amostras disponíveis.

- Avaliar o modelo AGMic em células de exames de Papanicolaou em meio líquido

Esta tese apresenta resultados do Módulo 1 do AGMic em imagens de células cervicais de meio líquido. Trabalhos futuros devem explorar esses dados para investigação também do Módulo 2 e proposição de aplicações.

- Avaliar o modelo AGMic em diferentes amostras citológicas

Vários atributos que definem anormalidade em células cervicais, como relação entre os tamanhos do núcleo e citoplasma, variações na textura do núcleo, dentre outros, são comuns a células provenientes de outros tecidos, como, por exemplo, o mamário. Assim, trabalhos futuros podem investigar se a atenção *top-down* dos citopatologistas, codificada pelo modelo proposto, é extensível à identificação de patologias em células de outros tecidos humanos.

- Avaliar novos algoritmos para categorização de células cervicais

Os experimentos realizados nesta tese aplicam uma arquitetura CNN modificada para categorização das regiões de interesse. Observamos que uma solução combinada entre CNN e a matriz de coocorrência estrutural proposta pode oferecer ganhos na precisão de classificação em especiais circunstâncias. Dessa forma, estudos futuros serão realizados com o objetivo de aperfeiçoar o desempenho do Módulo 2 do AGMic.

- Investigar o Módulo 1 do AGMic em outras aplicações, tais como:

- compressão de imagens citopatológicas: definição de partes relevantes da imagem que precisam ser armazenadas e/ou transmitidas com melhor qualidade.
- controle de qualidade em laboratórios citopatológicos: a correlação entre os mapas estimados e dados de atenção visual coletados de citologistas podem orientar programas de reciclagem na prática clínica.
- *Software* educacional: estudantes podem ser avaliados quanto ao seu padrão de busca por lesões nas imagens.

- Operacionalizar sistema para ranqueamento de exames de Papanicolaou em laboratórios tipo II do SUS

Trabalhos futuros devem ser realizados para implementar uma versão do AGMic que habilite o ranqueamento de campos (imagens) escaneados de exames de Papanicolaou.

O objetivo é adaptar nosso modelo para contribuir com novas metodologias que visem a melhor distribuição dos recursos humanos deste tipo de laboratório, dando prioridade para a análise de campos de exames que são clinicamente relevantes.

- Estudar o AGMic como parte de projetos que visem desenvolver microscópios autônomos  
Durante o desenvolvimento desta tese, alguns ensaios foram realizados por pesquisadores parceiros para o desenvolvimento de um microscópio autônomo. Projetos futuros devem investigar como a metodologia proposta pode contribuir com esses projetos.
- Outras propostas:
  - Estudar os demais dados coletados pelo rastreador ocular, como: movimentos sacádicos, *dwell time* e dilatação da pupila.
  - Estudar a atenção visual seletiva de estudantes em citologia a partir de um grupo de controle formado pelos especialistas.
  - Construir uma plataforma *online* que permita:
    - \* classificação *online* de imagem de células cervicais conforme predição pelo modelo computacional proposto.
    - \* submissão de novos modelos com o propósito de criação de um *benchmark*.
    - \* divulgação científica.

## REFERÊNCIAS

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. SLIC superpixels compared to state-of-the-art superpixel methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 34, n. 11, p. 2274–2282, 2012.
- ACS. **Survival Rates for Cervical Cancer**. 2019. Acesso em: 17-12-2019. Disponível em: <<https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/survival.html>>.
- ACS. **Survival Rates for Cervical Cancer**. 2019. Acesso em: 17-12-2019. Disponível em: <<https://www.cancer.net/cancer-types/cervical-cancer/statistics>>.
- ALLEHAIBI, K. H. S.; NUGROHO, L. E.; LAZUARDI, L.; PRABUWONO, A. S.; MANTORO, T. *et al.* Segmentation and classification of cervical cells using deep learning. **IEEE Access**, IEEE, v. 7, p. 116925–116941, 2019.
- ALMEIDA, K. I. V. Desigualdade social e câncer do colo do útero: uma revisão sistemática. **Anais do XVI Encontro Nacional de Pesquisadores em Serviço Social**, v. 16, n. 1, 2018.
- ALYAFEAI, Z.; GHOUTI, L. A fully-automated deep learning pipeline for cervical cancer classification. **Expert Systems with Applications**, Elsevier, v. 141, p. 112951, 2020.
- ALZUBAIDI, M.; BALASUBRAMANIAN, V.; PATEL, A.; PANCHANATHAN, S.; JR., J. A. B. What catches a radiologist’s eye? A comprehensive comparison of feature types for saliency prediction. In: **Medical Imaging 2010: Computer-Aided Diagnosis**. San Diego, CA: SPIE, 2010. v. 7624, p. 262 – 271.
- ANWAR, S. M.; MAJID, M.; QAYYUM, A.; AWAIS, M.; ALNOWAMI, M.; KHAN, M. K. Medical image analysis using convolutional neural networks: A review. **Journal of Medical Systems**, Springer, v. 42, n. 11, p. 226, 2018.
- BARROS, A. L. de S.; LIMA, D. N. de O.; AZEVEDO, M. D.; OLIVEIRA, M. de L. **Caderno de Referência 1: Citopatologia Ginecológica**. Rio de Janeiro: CEPESC, 2012. ISBN 978-85-324-0033-8.
- BARUT, M. U.; KALE, A.; KUYUMCUOĞLU, U.; BOZKURT, M.; AĞAÇAYAK, E.; ÖZEKINCI, S.; GUL, T. Analysis of sensitivity, specificity, and positive and negative predictive values of smear and colposcopy in diagnosis of premalignant and malignant cervical lesions. **Medical Science Monitor**, v. 21, p. 3860, 2015.
- BENGTSSON, E.; MALM, P. Screening for cervical cancer using automated analysis of Pap-smears. **Computational and Mathematical Methods in Medicine**, Hindawi, v. 2014, p. 1–12, 2014.
- BLAIR, D. C. **Information Retrieval**. 2nd. ed. London: Butterworths, 1979.
- BOON, M. E.; DRIJVER, J. S. **Routine Cytological Staining Techniques: Theoretical Background and Practice**. Basingstoke and London: Macmillan International Higher Education, 1986. ISBN 9780333397138.
- BOON, M. E.; SUURMEIJER, A. J. **The Pap Smear**. 3rd. ed. Boca Raton, FL: CRC Press, 1996. ISBN 3718658577.

- BORA, K.; CHOWDHURY, M.; MAHANTA, L. B.; KUNDU, M. K.; DAS, A. K. Automated classification of Pap smear images to detect cervical dysplasia. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 138, p. 31–47, 2017.
- BORJI, A. Saliency prediction in the deep learning era: Successes and limitations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, p. 1–20, 2019.
- BORJI, A.; CHENG, M.-M.; HOU, Q.; JIANG, H.; LI, J. Salient object detection: A survey. **arXiv preprint arXiv:1411.5878**, 2014.
- BORJI, A.; CHENG, M.-M.; JIANG, H.; LI, J. Salient object detection: A benchmark. **IEEE Transactions on Image Processing**, IEEE, v. 24, n. 12, p. 5706–5722, 2015.
- BORJI, A.; ITTI, L. State-of-the-art in visual attention modeling. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 35, n. 1, p. 185–207, 2013.
- BORTOLON, P. C.; SILVA, M. A. F. da; CORRÊA, F. de M.; DIAS, M. B. K.; KNUPP, V. M. de A. O.; ASSIS, M. de; CLARO, I. B. Avaliação da qualidade dos laboratórios de citopatologia do colo do útero no Brasil. **Revista Brasileira de Cancerologia**, v. 58, n. 3, p. 435–444, 2012.
- BOTCHKAREV, A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. **arXiv preprint arXiv:1809.03006**, 2018.
- BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, Wiley Online Library, v. 68, n. 6, p. 394–424, 2018.
- BREIMAN, L. Random forests. **Machine Learning**, Springer, v. 45, n. 1, p. 5–32, 2001.
- BYLINSKII, Z.; JUDD, T.; BORJI, A.; ITTI, L.; DURAND, F.; OLIVA, A.; TORRALBA, A. **MIT Saliency Benchmark**. 2015. <<http://saliency.mit.edu/>>.
- BYLINSKII, Z.; JUDD, T.; OLIVA, A.; TORRALBA, A.; DURAND, F. What do different evaluation metrics tell us about saliency models? **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 41, n. 3, p. 740–757, 2018.
- BYLINSKII, Z.; RECASENS, A.; BORJI, A.; OLIVA, A.; TORRALBA, A.; DURAND, F. Where should saliency models look next? In: **Proceedings of the 14th European Conference on Computer Vision (ECCV)**. Amsterdam, Netherlands: Springer, 2016. p. 809–824.
- CARRASCO, M. Visual attention: The past 25 years. **Vision Research**, Elsevier, v. 51, n. 13, p. 1484–1525, 2011.
- CERF, M.; FRADY, E. P.; KOCH, C. Using semantic content as cues for better scanpath prediction. In: **Proceedings of the 2008 Symposium on Eye Tracking Research & Applications**. New York, NY, USA: Association for Computing Machinery, 2008. p. 143–146.
- CERF, M.; FRADY, E. P.; KOCH, C. Faces and text attract gaze independent of the task: Experimental data and computer model. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 9, n. 12, p. 10.1–15, 2009.

CERF, M.; HAREL, J.; EINHÄUSER, W.; KOCH, C. Predicting human gaze using low-level saliency combined with face detection. In: **Advances in Neural Information Processing Systems 20 (NIPS)**. Red Hook, NY, USA: Curran Associates, Inc., 2008. p. 241–248.

CERF, M.; HAREL, J.; HUTH, A.; EINHÄUSER, W.; KOCH, C. Decoding what people see from where they look: Predicting visual stimuli from scanpaths. In: **Attention in Cognitive Systems**. Berlin: Springer, 2009. p. 15–26.

CHANKONG, T.; THEERA-UMPON, N.; AUEPHANWIRIYAKUL, S. Automatic cervical cell segmentation and classification in Pap smears. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 113, n. 2, p. 539–556, 2014.

CHEN, Y.-F.; HUANG, P.-C.; LIN, K.-C.; LIN, H.-H.; WANG, L.-E.; CHENG, C.-C.; CHEN, T.-P.; CHAN, Y.-K.; CHIANG, J. Y. Semi-automatic segmentation and classification of Pap smear cells. **Journal of Biomedical and Health Informatics**, IEEE, v. 18, n. 1, p. 94–108, 2014.

CIBAS, E. S.; DUCATMAN, B. S. **Cytology: Diagnostic Principles and Clinical Correlates**. 2nd. ed. London: Saunders, 2003. ISBN 145574462X.

CONCEIÇÃO, T.; BRAGA, C.; ROSADO, L.; VASCONCELOS, M. J. M. A review of computational methods for cervical cells segmentation and abnormality classification. **International Journal of Molecular Sciences**, Multidisciplinary Digital Publishing Institute, v. 20, n. 20, p. 5114, 2019.

COOMBES, L.; CULVERHOUSE, P. Pattern recognition in cervical cytological slide images. In: **Proceedings of the Fifth International Conference on Advances in Pattern Recognition (ICAPR)**. Mumbai, India: Allied Publishers, 2003. p. 1–4.

CORNIA, M.; BARALDI, L.; SERRA, G.; CUCCHIARA, R. Predicting human eye fixations via an LSTM-based saliency attentive model. **IEEE Transactions on Image Processing**, v. 27, n. 10, p. 5142–5154, 2018.

COX, S. Guidelines for Papanicolaou test screening and follow-up. **Journal of Midwifery & Women's Health**, Wiley Online Library, v. 57, n. 1, p. 86–89, 2012.

CRIC. **Center for Recognition and Inspection of Cells**. 2018. <<https://sites.google.com/view/centercric/>>.

DAI, J.; LI, Y.; HE, K.; SUN, J. R-FCN: Object detection via region-based fully convolutional networks. In: **Advances in Neural Information Processing Systems 29 (NIPS)**. Red Hook, NY, USA: Curran Associates Inc., 2016. p. 379–387.

DARGAN, S.; KUMAR, M.; AYYAGARI, M. R.; KUMAR, G. A survey of deep learning and its applications: A new paradigm to machine learning. **Archives of Computational Methods in Engineering**, Springer, p. 1–22, 2019.

DEMAY, R. M. Common problems in Papanicolaou smear interpretation. **Archives of Pathology & Laboratory Medicine**, College of American Pathologists, v. 121, n. 3, p. 229–238, 1997.

DOLZ, J.; DESROSIERS, C.; WANG, L.; YUAN, J.; SHEN, D.; AYED, I. B. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. **Computerized Medical Imaging and Graphics**, Elsevier, v. 79, p. 101660, 2020.

DUAN, L.; WU, C.; MIAO, J.; QING, L.; FU, Y. Visual saliency detection by spatially weighted dissimilarity. In: **Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Washington, DC, USA: IEEE, 2011. p. 473–480.

DUCATMAN, B. S.; WANG, H. H. **The Pap smear: Controversies in Practice**. London: Arnold, 2002. ISBN 0340759283.

DUCHOWSKI, A. T. **Eye Tracking Methodology: Theory and Practice**. 3. ed. London: Springer, 2017. ISBN 9783319578811.

EFRON, B. Estimating the error rate of a prediction rule: improvement on cross-validation. **Journal of the American Statistical Association**, Taylor & Francis, v. 78, n. 382, p. 316–331, 1983.

ELAZARY, L.; ITTI, L. A bayesian model for efficient visual search and recognition. **Vision Research**, Elsevier, v. 50, n. 14, p. 1338–1352, 2010.

FANG, S.; LI, J.; TIAN, Y.; HUANG, T.; CHEN, X. Learning discriminative subspaces on random contrasts for image saliency analysis. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 28, n. 5, p. 1095–1108, 2017.

FRINTROP, S. Computational visual attention. In: **Computer Analysis of Human Behavior**. London: Springer, 2011. p. 69–101.

GALLOWAY, M. M. Texture analysis using gray level run lengths. **Computer Graphics and Image Processing**, v. 4, n. 2, p. 172–179, 1975.

GAO, D.; MAHADEVAN, V.; VASCONCELOS, N. On the plausibility of the discriminant center-surround hypothesis for visual saliency. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 8, n. 7, p. 13.1–18, 2008.

GENÇTAV, A.; AKSOY, S.; ÖNDER, S. Unsupervised segmentation and classification of cervical cell images. **Pattern Recognition**, Elsevier, v. 45, n. 12, p. 4151–4168, 2012.

GHARIBA, B.; SHEHATA, M. S.; MCGUIRE, P. Visual saliency prediction based on deep learning. **Information**, Multidisciplinary Digital Publishing Institute, v. 10, n. 8, p. 257, 2019.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

GREER, B. E. The gynecologist's perspective of liability and quality issues with the Papanicolaou smear. **Archives of Pathology & Laboratory Medicine**, College of American Pathologists, v. 121, n. 3, p. 246–249, 1997.

GUAN, Q.; HUANG, Y.; ZHONG, Z.; ZHENG, Z.; ZHENG, L.; YANG, Y. Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. **arXiv preprint arXiv:1801.09927**, 2018.

GV, K. K.; REDDY, G. M. Automatic classification of whole slide Pap smear images using CNN with PCA based feature interpretation. In: **2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**. Los Alamitos, CA, USA: IEEE, 2019. p. 1074–1079.



HAFIZ, A. M.; BHAT, G. M. A survey of deep learning techniques for medical diagnosis. In: **Information and Communication Technology for Sustainable Development**. Singapore: Springer, 2020. p. 161–170.

HAHNLOSER, R. H.; SARPESHKAR, R.; MAHOWALD, M. A.; DOUGLAS, R. J.; SEUNG, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. **Nature**, Nature Publishing Group, v. 405, n. 6789, p. 947–951, 2000.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, IEEE, n. 6, p. 610–621, 1973.

HAREL, J.; KOCH, C.; PERONA, P. Graph-based visual saliency. In: **Advances in Neural Information Processing Systems 19 (NIPS)**. Cambridge, MA, USA: MIT Press, 2007. p. 545–552.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE, 2016. p. 770–778.

HELBREN, E. **Eye-tracking the moving medical image: development and investigation of a novel investigational tool for CT colonography**. Tese (Doutorado) — UCL (University College London), 2019.

HENDERSON, J. M.; BROCKMOLE, J. R.; CASTELHANO, M. S.; MACK, M. Chapter 25 - Visual saliency does not account for eye movements during visual search in real-world scenes. In: **Eye Movements**. Oxford: Elsevier, 2007. p. 537–III.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.

HOFFMAN, J. E. Hierarchical stages in the processing of visual information. **Perception & Psychophysics**, Springer, v. 18, n. 5, p. 348–354, 1975.

HOLLANDER, M.; WOLFE, D. A.; CHICKEN, E. **Nonparametric Statistical Methods**. New York, USA: John Wiley & Sons, 2013. v. 751.

HOU, X.; HAREL, J.; KOCH, C. Image signature: Highlighting sparse salient regions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 34, n. 1, p. 194–201, 2012.

HOU, X.; ZHANG, L. Saliency detection: A spectral residual approach. In: **Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Minneapolis, USA: IEEE, 2007. p. 1–8.

HUANG, X.; SHEN, C.; BOIX, X.; ZHAO, Q. SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In: **Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)**. Los Alamitos, CA, USA: IEEE, 2015. p. 262–270.

INCA. **Manual de Gestão da Qualidade para Laboratório de Citopatologia**. Rio de Janeiro: INCA, 2012.

INCA. **Detecção Precoce do Câncer de Colo do Útero**. 2019. <<https://www.inca.gov.br/en/node/1194>>. Acesso em: 17-12-2019.

INCA. **Estimativa 2020 - Incidência de Câncer no Brasil**. Rio de Janeiro: INCA, 2019.

ITTI, L.; KOCH, C. Computational modelling of visual attention. **Nature Reviews Neuroscience**, Nature Publishing Group, v. 2, n. 3, p. 194–203, 2001.

ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 20, n. 11, p. 1254–1259, 1998.

JAMPANI, V.; SIVASWAMY, J.; VAIDYA, V. *et al.* Assessment of computational visual attention models on medical images. In: **Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing**. New York, NY, USA: Association for Computing Machinery, 2012. p. 1–8.

JANTZEN, J.; DOUNIAS, G. Analysis of Pap-smear image data. In: **Proceedings of the Nature-Inspired Smart Information Systems 2nd Annual Symposium (NiSIS)**. Tenerife, Spain: NiSIS, 2006. p. 1–10.

JANTZEN, J.; NORUP, J.; DOUNIAS, G.; BJERREGAARD, B. Pap-smear benchmark data for pattern classification. In: **Proceedings of the Nature-Inspired Smart Information Systems Symposium (NiSIS)**. Albufeira, Portugal: NiSIS, 2005. p. 1–9.

JIANG, M.; HUANG, S.; DUAN, J.; ZHAO, Q. SALICON: Saliency in context. In: **Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE, 2015. p. 1072–1080.

JITH, O. N.; HARINARAYANAN, K.; GAUTAM, S.; BHAVSAR, A.; SAO, A. K. Deepcerv: Deep neural network for segmentation free robust cervical cell classification. In: **Computational Pathology and Ophthalmic Medical Image Analysis**. Berlin: Springer, 2018. p. 86–94.

JUDD, T. **Understanding and Predicting where People Look in Images**. Tese (Doutorado) — Massachusetts Institute of Technology, 2011.

JUDD, T.; EHINGER, K.; DURAND, F.; TORRALBA, A. Learning to predict where humans look. In: **Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV)**. Los Alamitos, CA, USA: IEEE, 2009. p. 2106–2113.

KANAN, C.; TONG, M. H.; ZHANG, L.; COTTRELL, G. W. Sun: Top-down saliency using natural statistics. **Visual Cognition**, Taylor & Francis, v. 17, n. 6-7, p. 979–1003, 2009.

KE, J.; JIANG, Z.; LIU, C.; BEDNARZ, T.; SOWMYA, A.; LIANG, X. Selective detection and segmentation of cervical cells. In: **Proceedings of the 11th International Conference on Bioinformatics and Biomedical Technology (ICBBT)**. New York, NY, USA: Association for Computing Machinery, 2019. p. 55–61.

KETKAR, N. Introduction to keras. In: **Deep learning with Python**. Berkeley, CA, USA: Apress, 2017. p. 97–111. ISBN 9781484227657.

KLINKE, T. S. The Papanicolaou smear: A brief historical perspective and where we are today. **Archives of Pathology & Laboratory Medicine**, College of American Pathologists, v. 121, n. 3, p. 205–209, 1997.

KOCH, C. Shifts in selective visual attention: Towards the underlying neural circuitry. **Human Neurobiology**, v. 4, p. 219–227, 1985.

KOSS, L. G. The Papanicolaou test for cervical cancer detection: a triumph and a tragedy. **Jama**, American Medical Association, v. 261, n. 5, p. 737–743, 1989.

KOSS, L. G.; MELAMED, M. R. **Koss' diagnostic Cytology and Its Histopathologic Bases**. New York, NY, USA: Lippincott Williams & Wilkins, 2006. v. 1. ISBN 0781719283.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems 25 (NIPS)**. Red Hook, NY, USA: Curran Associates, Inc., 2012. p. 1097–1105.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, Taylor & Francis, v. 47, n. 260, p. 583–621, 1952.

LEE, H.; KIM, J. Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**. Los Alamitos, CA, USA: IEEE, 2016. p. 63–69.

LI, J.; LI, W.; GERTYCH, A.; KNUDSEN, B. S.; SPEIER, W.; ARNOLD, C. W. An attention-based multi-resolution model for prostate whole slide image classification and localization. **arXiv preprint arXiv:1905.13208**, 2019.

LI, R.; SHI, P.; PELZ, J.; ALM, C. O.; HAAKE, A. R. Modeling eye movement patterns to characterize perceptual skill in image-based diagnostic reasoning processes. **Computer Vision and Image Understanding**, Elsevier, v. 151, p. 138–152, 2016.

LI, Y.; CHEN, Z.; CAI, Y.; HUANG, D.; LI, Q. Accelerating convolutional neural networks using fine-tuned backpropagation progress. In: **Database Systems for Advanced Applications**. Cham: Springer International Publishing, 2017. p. 256–266. ISBN 978-3-319-55705-2.

LIANG, Y.; TANG, Z.; YAN, M.; CHEN, J.; XIANG, Y. Comparison detector: Convolutional neural networks for cervical cell detection. **arXiv preprint arXiv:1810.05952**, 2019.

LIN, T.-Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature pyramid networks for object detection. In: **Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE, 2017. p. 2117–2125.

LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFOORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. **Medical Image Analysis**, Elsevier, v. 42, p. 60–88, 2017.

LOUKAS, C.; VARYTIMIDIS, C.; RAPANTZIKOS, K.; KANAKIS, M. A. Keyframe extraction from laparoscopic videos based on visual saliency detection. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 165, p. 13–23, 2018.

LU, Z.; CARNEIRO, G.; BRADLEY, A. P.; USHIZIMA, D.; NOSRATI, M. S.; BIANCHI, A. G. C.; CARNEIRO, C. M.; HAMARNEH, G. Evaluation of three algorithms for the segmentation of overlapping cervical cells. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 21, n. 2, p. 441–450, 2016.

- MAATEN, L. van der; HINTON, G. Visualizing data using t-SNE. **Journal of Machine Learning Research**, v. 9, n. November, p. 2579–2605, 2008.
- MALM, P.; BALAKRISHNAN, B. N.; SUJATHAN, V. K.; KUMAR, R.; BENGTTSSON, E. Debris removal in Pap-smear images. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 111, n. 1, p. 128–138, 2013.
- MARIARPUTHAM, E. J.; STEPHEN, A. Nominated texture based cervical cancer classification. **Computational and Mathematical Methods in Medicine**, Hindawi, v. 2015, p. 586928, 2015.
- MARINAKIS, Y.; DOUNIAS, G.; JANTZEN, J. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. **Computers in Biology and Medicine**, Elsevier, v. 39, n. 1, p. 69–78, 2009.
- MATSUMOTO, H.; TERAOKA, Y.; YUGETA, A.; FUKUDA, H.; EMOTO, M.; FURUBAYASHI, T.; OKANO, T.; HANAJIMA, R.; UGAWA, Y. Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases. **PloS One**, Public Library of Science, v. 6, n. 12, p. e28928, 2011.
- MEHNERT, A. J. H. **Image analysis for the study of chromatin distribution in cell nuclei**. Tese (Doutorado) — School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia, 2003.
- MEUR, O. L.; BACCINO, T. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. **Behavior Research Methods**, Springer, v. 45, n. 1, p. 251–266, 2013.
- MURRAY, N.; VANRELL, M.; OTAZU, X.; PARRAGA, C. A. Saliency estimation using a non-parametric low-level vision model. In: **Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Washington, DC, USA: IEEE, 2011. p. 433–440.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: **Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)**. New York, NY, USA: Association for Computing Machinery, 2010. p. 807–814.
- NAYAR, R.; WILBUR, D. C. **The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes**. London: Springer, 2015. ISBN 9783319110738.
- NAYAR, R.; WILBUR, D. C. The Pap test and Bethesda 2014. **Acta Cytologica**, Karger Publishers, v. 59, n. 2, p. 121–132, 2015.
- NEISSER, U. **Cognitive Psychology**. New York, NY, USA: Appleton-Century-Crofts, 1967. ISBN 9780390665096.
- NGUYEN, T. V.; ZHAO, Q.; YAN, S. Attentive systems: A survey. **International Journal of Computer Vision**, Springer, v. 126, n. 1, p. 86–110, 2018.
- OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. **Pattern Recognition**, Elsevier, v. 29, n. 1, p. 51–59, 1996.

OTAZU, X.; PARRAGA, C. A.; VANRELL, M. Toward a unified chromatic induction model. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 10, n. 12, p. 1–24, 2010.

PAPANICOLAOU, G. N.; TRAUT, H. F. The diagnostic value of vaginal smears in carcinoma of the uterus. **American Journal of Obstetrics and Gynecology**, Mosby, v. 42, n. 2, p. 193–206, 1941.

PAPANICOLAOU, G. N.; TRAUT, H. F. **Diagnosis of Uterine Cancer by the Vaginal Smear**. New York: The Commonwealth Fund, 1943.

PAULA, A. C. de; SOUZA, N. G. de; PRADO, T. C. do; RIBEIRO, A. A. Indicadores do monitoramento interno da qualidade dos exames citopatológicos do laboratório clínico da pontifícia universidade católica de goias (PUC-GO). **Revista Brasileira de Análises Clínicas (RBAC)**, v. 49, n. 2, p. 200–5, 2017.

PETERS, R. J.; IYER, A.; ITTI, L.; KOCH, C. Components of bottom-up gaze allocation in natural images. **Vision Research**, Elsevier, v. 45, n. 18, p. 2397–2416, 2005.

PHOULADY, H. A.; MOUTON, P. R. A new cervical cytology dataset for nucleus detection and image classification (cervix93) and methods for cervical nucleus detection. **arXiv preprint arXiv:1811.09651**, 2018.

PLISSITI, M. E.; NIKOU, C. Cervical cell classification based exclusively on nucleus features. In: **Proceedings of the 9th International Conference on Image and Recognition (ICIAR)**. Berlin: Springer, 2012. p. 483–490.

PONS, O. Bootstrap of means under stratified sampling. **Electronic Journal of Statistics**, The Institute of Mathematical Statistics and the Bernoulli Society, v. 1, p. 381–391, 2007.

POWERS, D. M. Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. **Journal of Machine Learning Technologies**, Bioinfo Publications, v. 2, n. 1, p. 37–63, 2011.

RAMALHO, G. L. B.; FERREIRA, D. S.; BIANCHI, A. G. C.; CARNEIRO, C. M.; MEDEIROS, F. N. S.; USHIZIMA, D. M. Cell reconstruction under voronoi and enclosing ellipses from 3D microscopy. In: **ISBI. Overlapping Cervical Cytology Image Segmentation Challenge - ISBI 2015**. Brooklyn, NY, USA, 2015. p. 1–2.

REDMON, J.; FARHADI, A. Yolo9000: better, faster, stronger. In: **Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE, 2017. p. 7263–7271.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: Towards real-time object detection with region proposal networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 6, p. 1137–1149, 2017.

RIBLI, D.; HORVÁTH, A.; UNGER, Z.; POLLNER, P.; CSABAI, I. Detecting and classifying lesions in mammograms with deep learning. **Scientific Reports**, Nature Publishing Group, v. 8, n. 1, p. 4165, 2018.

RUBNER, Y.; TOMASI, C.; GUIBAS, L. J. The earth mover's distance as a metric for image retrieval. **International Journal of Computer Vision**, Springer, v. 40, n. 2, p. 99–121, 2000.

- RUDIN, L. I.; OSHER, S.; FATEMI, E. Nonlinear total variation based noise removal algorithms. **Physica D: Nonlinear Phenomena**, North-Holland, v. 60, n. 1-4, p. 259–268, 1992.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.
- SARWAR, A.; SHARMA, V.; GUPTA, R. Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis. **Personalized Medicine Universe**, Elsevier, v. 4, p. 54–62, 2015.
- SCHLEMPER, J.; OKTAY, O.; SCHAAP, M.; HEINRICH, M.; KAINZ, B.; GLOCKER, B.; RUECKERT, D. Attention gated networks: Learning to leverage salient regions in medical images. **Medical Image Analysis**, Elsevier, v. 53, p. 197–207, 2019.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. **Annual Review of Biomedical Engineering**, Annual Reviews, v. 19, p. 221–248, 2017.
- SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. **IEEE Transactions on Medical Imaging**, IEEE, v. 35, n. 5, p. 1285–1298, 2016.
- SILVA, M. do P. S. **Alcances e limites do exame citopatológico com coloração do Papanicolaou no diagnóstico das cérvico-vaginites: um estudo citológico e microbiológico de 2169 casos de um total de 10.064 exames citopatológicos**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2004.
- SILVA, R. R. V.; ARAUJO, F. H. D.; USHIZIMA, D. M.; BIANCHI, A. G. C.; CARNEIRO, C. M.; MEDEIROS, F. N. S. Radial feature descriptors for cell classification and recommendation. **Journal of Visual Communication and Image Representation**, Elsevier, v. 62, p. 105–116, 2019.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
- SOILLE, P. **Morphological Image Analysis: Principles and Applications**. 2nd. ed. Berlin: Springer, 2013. ISBN 3662039397.
- STENESH, J. **Dictionary of Biochemistry and Molecular Biology**. New York: John Wiley & Sons, 1989. ISBN 0471840890.
- STOCKMEYER, L.; CHANDRA, A. Intrinsically difficult problems. **Scientific American Trends in Computing**, New York, v. 1, p. 88–97, 1988.
- SU, J.; XU, X.; HE, Y.; SONG, J. Automatic detection of cervical cancer cells by a two-level cascade classification system. **Analytical Cellular Pathology**, Hindawi, v. 2016, p. 9535027, 2016.
- SUTSKEVER, I.; MARTENS, J.; DAHL, G.; HINTON, G. On the importance of initialization and momentum in deep learning. In: **Proceedings of the 30th International Conference on Machine Learning (ICML)**. Atlanta: JMLR.org, 2013. p. 1139–1147.
- SWAIN, M. J.; BALLARD, D. H. Color indexing. **International journal of computer vision**, Springer, v. 7, n. 1, p. 11–32, 1991.

TAKEDA, Y.; YAGI, A. Inhibitory tagging in visual search can be found if search stimuli remain visible. **Perception & Psychophysics**, Springer, v. 62, n. 5, p. 927–934, 2000.

TANG, X. Texture information in run-length matrices. **IEEE Transactions on Image Processing**, IEEE, v. 7, n. 11, p. 1602–1609, 1998.

TATLER, B. W. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 7, n. 14, p. 1–17, 2007.

TAVAKOLI, H. R.; LAAKSONEN, J. Bottom-up fixation prediction using unsupervised hierarchical models. In: **Proceedings of the 13th Asian Conference on Computer Vision (ACCV)**. Berlin: Springer, 2016. p. 287–302.

TAVAKOLI, H. R.; RAHTU, E.; HEIKKILÄ, J. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: **Proceedings of the 16th Scandinavian Conference on Image Analysis (SCIA)**. Berlin: Springer, 2011. p. 666–675.

TOBIAS, A. H. G.; AMARAL, R. G.; DINIZ, E. M.; CARNEIRO, C. M. Quality indicators of cervical cytopathology tests in the public service in Minas Gerais, Brasil. **Revista Brasileira de Ginecologia e Obstetrícia/RBGO Gynecology and Obstetrics**, Thieme Publicações Ltda, v. 38, n. 02, p. 065–070, 2016.

TORRALBA, A.; EFROS, A. A. Unbiased look at dataset bias. In: **Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Washington, DC, USA: IEEE, 2011. p. 1521–1528.

TORRALBA, A.; OLIVA, A.; CASTELHANO, M. S.; HENDERSON, J. M. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. **Psychological Review**, American Psychological Association, v. 113, n. 4, p. 766–786, 2006.

TREISMAN, A. M.; GELADE, G. A feature-integration theory of attention. **Cognitive psychology**, Elsevier, v. 12, n. 1, p. 97–136, 1980.

TSOTSOS, J. K. **A computational perspective on visual attention**. Cambridge, MA, USA: MIT Press, 2011. ISBN 9780262015417.

VOLOKITIN, A.; GYGLI, M.; BOIX, X. Predicting when saliency maps are accurate and eye fixations consistent. In: **Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE, 2016. p. 544–552.

WANG, W.; SHEN, J. Deep visual attention prediction. **IEEE Transactions on Image Processing**, IEEE, v. 27, n. 5, p. 2368–2378, 2017.

WATANABE, S.; IWASAKA, T.; YOKOYAMA, M.; UCHIYAMA, M.; KAKU, T.; MATSUYAMA, T. Analysis of nuclear chromatin distribution in cervical glandular abnormalities. **Acta Cytologica**, Karger Publishers, v. 48, n. 4, p. 505–513, 2004.

WEBB, G. I. Overfitting. In: \_\_\_\_\_. **Encyclopedia of Machine Learning**. Boston, MA: Springer, 2010. ISBN 9780387301648.

WEN, G.; RODRIGUEZ-NIÑO, B.; PECEN, F. Y.; VINING, D. J.; GARG, N.; MARKEY, M. K. Comparative study of computational visual attention models on two-dimensional medical images. **Journal of Medical Imaging**, International Society for Optics and Photonics, v. 4, n. 2, p. 025503, 2017.

WESOŁA, M.; LIPÍŃSKI, A.; JELEŃ, M. Morphometry in the cytological diagnosis of cervical smears. **Advances in Clinical and Experimental Medicine: Official Organ Wrocław Medical University**, v. 23, n. 2, p. 289–293, 2014.

WESSELS, S.; HAAR, D. van der. Applying deep learning for the detection of abnormalities in mammograms. In: **Proceedings of the 10th International Conference on Information Science and Applications (ICISA)**. Singapore: Springer, 2019. p. 201–210.

WHO. **Cervical Cancer**. 2019. <<https://www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/>>. Acesso em: 17-12-2019.

WHO; IARC. **Cytopathology of the Uterine Cervix - Digital Atlas**. 2004. Acesso em: 04-01-2020. Disponível em: <<https://screening.iarc.fr/atlascyto.php>>.

WHO; IARC. Cervix cancer screening. **IARC Handbooks of Cancer Prevention**, International Agency for Research on Cancer, v. 10, 2005.

WILLIAM, W.; WARE, A.; BASAZA-EJIRI, A. H.; OBUNGOLOCH, J. A review of image analysis and machine learning techniques for automated cervical cancer screening from Pap-smear images. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 164, p. 15–22, 2018.

WILLIAM, W.; WARE, A.; BASAZA-EJIRI, A. H.; OBUNGOLOCH, J. A Pap-smear analysis tool (PAT) for detection of cervical cancer from Pap-smear images. **Biomedical Engineering Online**, BioMed Central, v. 18, n. 1, p. 16, 2019.

WINKLER, S.; SUBRAMANIAN, R. Overview of eye tracking datasets. In: **Proceedings of the Fifth International Workshop on Quality of Multimedia Experience (QoMEX)**. Los Alamitos, CA, USA: IEEE, 2013. p. 212–217.

WOLFE, J. M.; CAVE, K. R.; FRANZEL, S. L. Guided search: An alternative to the feature integration model for visual search. **Journal of Experimental Psychology: Human perception and performance**, American Psychological Association, v. 15, n. 3, p. 419–433, 1989.

YARBUS, A. L. **Eye Movement and Vision (translated from the russian edition by Basil Haigh)**. New York, NY, USA: Plenum Press, 1967.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? In: **Advances in Neural Information Processing Systems 27 (NIPS)**. Red Hook, NY, USA: Curran Associates, Inc., 2014. p. 3320–3328.

ZHANG, C.; LIU, D.; WANG, L.; LI, Y.; CHEN, X.; LUO, R.; CHE, S.; LIANG, H.-S.; LI, Y.; LIU, S.; TU, D.; QI, G.-J.; LUO, P.; LUO, J. DCCL: A benchmark for cervical cytology analysis. In: **Proceedings of the 10th International Workshop on Machine Learning in Medical Imaging (MLMI)**. Berlin: Springer, 2019. p. 63–72.



ZHANG, J.; HE, J.; CHEN, T.; LIU, Z.; CHEN, D. Abnormal region detection in cervical smear images based on fully convolutional network. **IET Image Processing**, IET, v. 13, n. 4, p. 583–590, 2018.

ZHANG, J.; SCLAROFF, S. Saliency detection: A boolean map approach. In: **Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV)**. Los Alamitos, CA, USA: IEEE, 2013. p. 153–160.

ZHANG, J.-W.; LIAN, M.-C.; WANG, W.-P.; ZHU, L. Detection of abnormal nuclei in cervical smear images based on visual attention model. In: IEEE. **IEEE International Conference on Machine Learning and Cybernetics (ICMLC)**. Tianjin, China, 2013. v. 2, p. 920–924.

ZHANG, L.; KONG, H.; CHIN, C. T.; LIU, S.; CHEN, Z.; WANG, T.; CHEN, S. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. **Computerized Medical Imaging and Graphics**, Elsevier, v. 38, n. 5, p. 369–380, 2014.

ZHANG, L.; LIN, W. **Selective Visual Attention: Computational Models and Applications**. 1. ed. Singapore: Wiley-IEEE Press, 2013. ISBN 9780470828120.

ZHANG, L.; LU, L.; NOGUES, I.; SUMMERS, R. M.; LIU, S.; YAO, J. DeepPap: Deep convolutional networks for cervical cell classification. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 21, n. 6, p. 1633–1643, 2017.

ZHANG, L.; TONG, M. H.; MARKS, T. K.; SHAN, H.; COTTRELL, G. W. SUN: A Bayesian framework for saliency using natural statistics. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 8, n. 7, p. 1–20, 2008.

ZHU, J.; NORMAN, I.; ELFGREN, K.; GABERI, V.; HAGMAR, B.; HJERPE, A.; ANDERSSON, S. A comparison of liquid-based cytology and Pap smear as a screening method for cervical cancer. **Oncology Reports**, Spandidos Publications, v. 18, n. 1, p. 157–160, 2007.

**APÊNDICE A – FORMULÁRIO DE CONSENTIMENTO**

## Atenção Visual em Imagens Cervicais

### Formulário de Consentimento

#### **IMPORTANTE: ESTE FORMULÁRIO DEVE SER ASSINADO ANTES DE PARTICIPAR**

Nós, abaixo-assinados, concordamos em participar deste trabalho de pesquisa relativo ao projeto de doutorado do estudante DANIEL SILVA FERREIRA, orientado pela profa. Dra. Fátima Nelsizeuma Sombra de Medeiros (PPGETI/UFC). Nós afirmamos ter recebido respostas satisfatórias para as nossas questões relativas ao estudo, com o pleno conhecimento de que temos o direito de nos recusar a participar. Nós também entendemos que todos os esforços serão feitos para proteger o anonimato de nossas respostas.

	Data	Nome	Assinatura
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Nós certificamos que os nomes, datas e assinaturas nesta folha são autênticas.

Fortaleza, \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

---

Daniel Silva Ferreira  
(estudante)

---

Dra. Fátima Nelsizeuma Sombra de Medeiros  
(orientadora)

## APÊNDICE B – REGION-BASED STRUCTURAL CO-OCCURRENCE MATRIX (RSCM) APLICADA À CARACTERIZAÇÃO DE CÉLULAS CERVICAIS

Este apêndice apresenta a matriz rSCM como um novo descritor para células cervicais. Esta proposta fundamenta-se na tendência das células anormais apresentarem núcleo com textura heterogênea, de aspecto granular, devido à distribuição irregular de cromatina (DUCATMAN; WANG, 2002; CHANKONG *et al.*, 2014) e na perda de definição na borda de núcleos anormais devido ao depósito de cromatina na membrana nuclear (SILVA *et al.*, 2019).

Realizamos experimentos de classificação binária<sup>1</sup> tendo como base o trabalho de Silva *et al.* (2019). Complementamos o estudo com testes de significância estatística com o objetivo de investigar a consistência dos achados em relação a outros algoritmos estado-da-arte. A partir dos resultados obtidos, definimos os parâmetros de ajuste da rSCM como parte do modelo AGMic (Seção 6.1.2).

### B.1 Base de Imagens

A base de dados Herlev (JANTZEN *et al.*, 2005) foi utilizada nos experimentos. Optamos por este conjunto de imagens por ser público e largamente utilizado na literatura; dispor da segmentação do núcleo como referência; apresentar diversidade de amostras para diferentes tipos de células, tanto normais quanto anormais; apresentar resolução similar a imagens da base CRIC; reunir imagens reais coletadas em hospital de referência (*Herlev University Hospital*)<sup>2</sup>.

Esta base é composta por 917 imagens de células cervicais (recortes) distribuídas em 7 classes: #{escamoso superficial: 74, escamoso intermediário: 70, colunar: 98, displasia leve: 182, displasia moderada: 146, displasia severa: 197 e *carcinoma in situ*: 150}. A Figura 48 apresenta amostras para as 7 classes disponíveis. Conforme Jantzen *et al.* (2005) e seguindo metodologia descrita em (SILVA *et al.*, 2019), combinamos as classes escamoso superficial, escamoso intermediário e colunar como classe normal, enquanto displasia leve, displasia moderada, displasia severa e *carcinoma in situ* formam a classe anormal. Assim totalizamos 242 amostras de células normais e 675 anormais. Todas as imagens possuem resolução de  $0,201\mu m$  por *pixel* e equivalem a recortes de diferentes tamanhos. Um estudo detalhado sobre como as amostras da base Herlev estão correlacionadas ao longo das diferentes classes está descrito em (JANTZEN; DOUNIAS, 2006).

<sup>1</sup> Classificação em duas classes: normal e anormal.

<sup>2</sup> <<https://www.herlevhospital.dk/>>

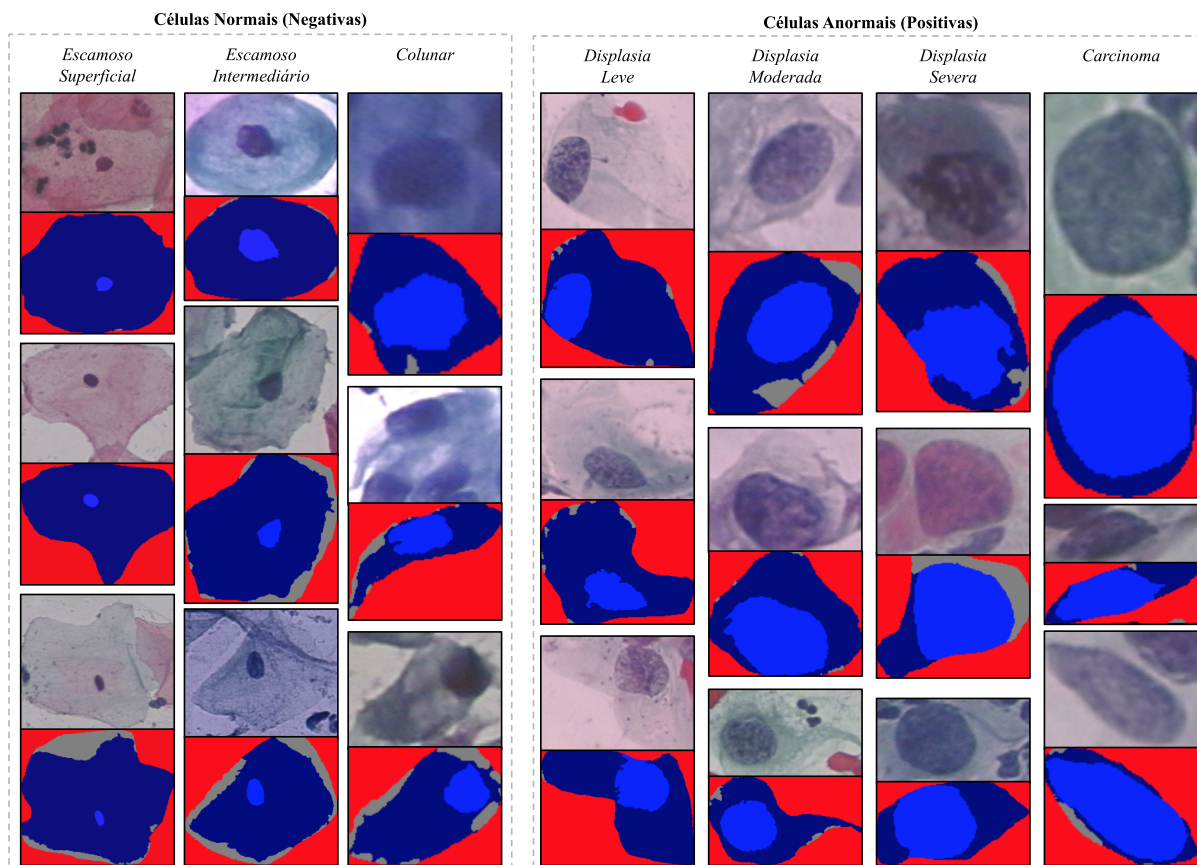


Figura 48 – Amostras de recortes de células cervicais, com suas respectivas segmentações, disponíveis na base de dados Herlev. A segmentação do núcleo está representada em azul claro. A área de citoplasma é marcada por azul escuro e as demais cores mostram o fundo (*background*). O agrupamento das 7 classes em normal e anormal segue a proposta de Jantzen *et al.* (2005). Fonte: elaboração própria a partir dos dados disponíveis na base Herlev (JANTZEN *et al.*, 2005).

## B.2 Medidas de Avaliação

Reportamos valores para duas medidas comumente empregadas em problemas de classificação: Acurácia e coeficiente *Cohen's Kappa* (GENÇTAV *et al.*, 2012). A Acurácia (ACC) corresponde à porcentagem de dados que são corretamente classificados e pode ser descrita para um problema em  $N_c$  classes como:

$$ACC = \frac{\sum_{i=1}^{N_c} \text{Número de imagens classificadas corretamente na classe } i}{\sum_{i=1}^{N_c} \text{Total de imagens na classe } i}. \quad (\text{B.1})$$

O coeficiente *Cohen's Kappa*, ou simplesmente Kappa ( $K$ ), representa a concordância entre a classe desejada e a predita, ajustando esses valores para os acordos que poderiam

ser esperados devido apenas ao acaso. Foi introduzida na área de psicologia para medir a concordância entre especialistas sobre pacientes e vem sendo bastante utilizada para avaliar o desempenho de algoritmos de classificação (CHANKONG *et al.*, 2014; GENÇTAV *et al.*, 2012; SILVA *et al.*, 2019). Formalmente, temos:

$$K = \frac{ACC - P_e}{1 - P_e}, \quad (\text{B.2})$$

onde  $P_e$  é a probabilidade hipotética de um acordo por acaso, usando os valores da matriz de confusão  $C$  para estimar as probabilidades de escolhas aleatórias em cada classe, ou seja,  $P_e = \sum_{i=1}^{N_c} \frac{C_i \times C_j}{T^2}$ , onde  $C_i = \sum_{j=1}^{N_c} C_{ij}$  (somatório da linha  $i$ ),  $C_j = \sum_{l=1}^{N_c} C_{lj}$  (somatório da coluna  $j$ ) e  $T = \sum C$ .

O valor máximo de Kappa é 1, com esse valor implicando em perfeita concordância entre os rótulos esperados e preditos. À medida que se aproxima de zero, esta concordância é reduzida. Em algumas circunstâncias, Kappa pode assumir valores negativos, sinalizando que as chances de uma predição ao acaso superam a taxa de acerto (ACC) observada. A Tabela 7 apresenta a interpretação de Kappa adotada nesta tese, conforme abordagem utilizada previamente por Silva *et al.* (2019).

Tabela 7 – Interpretação adotada para os valores do coeficiente Kappa. Fonte: elaboração própria a partir de (SILVA *et al.*, 2019).

Qualidade da Concordância	Faixa de Valores
Concordância pobre	inferior a 0,20
Concordância justa	de 0,20 a 0,40
Concordância moderada	de 0,40 a 0,60
Boa concordância	0,60 a 0,80
Muito boa concordância	0,80 a 1,00

### B.3 Metodologia

A Figura 49 apresenta a metodologia empregada na avaliação do desempenho da matriz rSCM como descritor de células cervicais. Como a base Herlev é desbalanceada para as classes de células normais e anormais (Seção B.1), aleatoriamente subamostramos a maior classe para balanceá-la. Inspirados pelo algoritmo *stratified bootstrap* (PONS, 2007) e (EFRON, 1983), executamos esse procedimento selecionando, com reposição,  $q$  amostras de ambas as classes, onde  $q$  representa o número de amostras da classe menor. As imagens selecionadas em

ambas classes foram utilizadas para treinar um classificador, enquanto as remanescentes para teste. Limitamos o conjunto de teste da classe majoritária pela quantidade de amostras de teste presentes na classe minoritária. As etapas de divisão treino-teste, extração de características e classificação foram realizadas em 1000 iterações independentes. Os testes foram executados para a rSCM e o conjunto de descritores abordados por Silva *et al.* (2019).

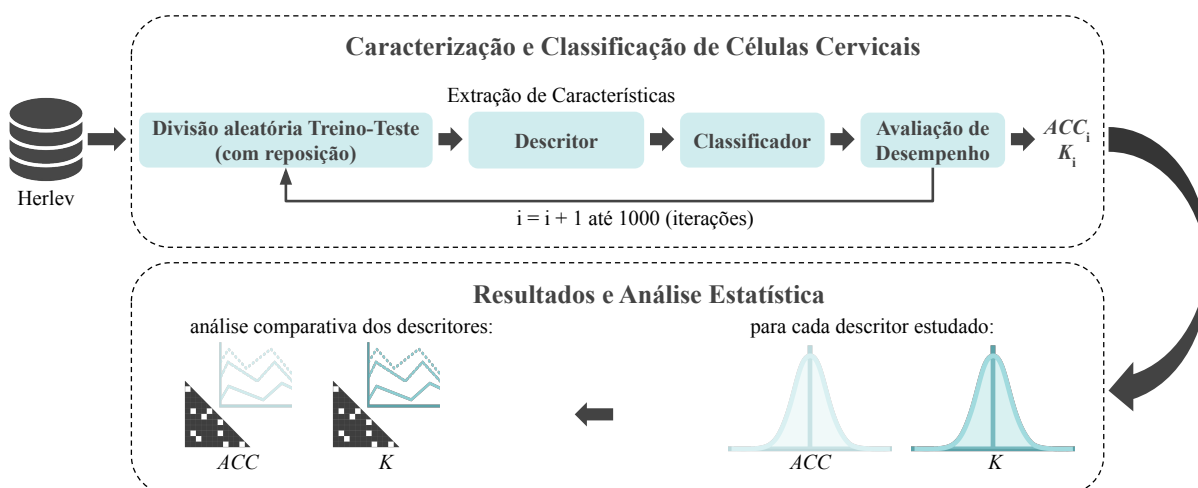


Figura 49 – Metodologia para análise do descritor rSCM como extrator de características de células cervicais normais e anormais. Fonte: elaboração própria.

Para cada descritor, reunimos um conjunto de valores de medidas, os quais foram avaliados estatisticamente pelo teste de Kruskal-Wallis, em abordagem similar à discutida na Seção 7.1.4.1. Em todos os experimentos, utilizamos o classificador *Random Forest*, com a mesma configuração de parâmetros proposta por Silva *et al.* (2019). Uma vez que esta tese emprega a matriz rSCM como parte do modelo proposto na Seção 6.1.2.2, destacando partes da imagem com anormalidade, limitamos o estudo ao problema da classificação em duas classes: normal e anormal.

## B.4 Resultados

Conforme discutido na Seção 6.1.2, adotamos o filtro passa-baixa com *kernel* gaussiano para tornar a rSCM sensível à textura heterogênea, a qual é sugestiva de lesão celular. Projetamos o algoritmo para modelar atributos do núcleo e do seu entorno, incluindo a membrana nuclear. Inicialmente, realizamos uma busca por *grid search*<sup>3</sup> para encontrar o desvio padrão do *kernel* gaussiano ( $\sigma$ ) e quanto da região ao redor do núcleo ( $rm$ ) é necessária para maximizar

<sup>3</sup> *Grid search* é uma técnica que tenta estimar os valores ótimos de hiperparâmetros de um modelo a partir de uma busca exaustiva dentro de um conjunto de valores.

$ACC$  na classificação binária<sup>4</sup>. A partir das máscaras de segmentação do núcleo disponíveis na base Herlev, atribuímos  $rm$  como raio de um elemento estruturante circular para uma operação morfológica de dilatação (SOILLE, 2013). A opção por essa forma de elemento estruturante deve-se à morfologia esférica ou ovóide do núcleo. Definimos os campos de busca para  $\sigma$  e  $rm$  com valores inteiros positivos conforme os seguintes intervalos:  $\sigma \in [1, 6]$  e  $rm \in [0, 7]$ , onde  $rm = 0$  sinaliza que apenas a área do núcleo é considerada. Assumimos  $N_Q = 8$  níveis de quantização para a função  $Q$  (Seção 6.1.2), resultando em um vetor de características de 64 elementos para a rSCM (vetorização da matriz). Realizamos experimentos com  $N_Q \in \{8, 16\}$ . Por simplificação, reportamos apenas os resultados para  $N_Q = 8$ , para o qual o algoritmo apresentou melhor desempenho.

A Figura 50a apresenta a melhor Acurácia em cada  $rm$  avaliado. Encontramos que  $rm = 2$ , com  $\sigma = 4$ , maximizam o desempenho do descritor proposto. Embora o teste estatístico de Kruskal-Wallis, combinado com o teste de Nemenyi, não aponte diferenças ( $\alpha = 0,05$ ) no intervalo de  $rm \geq 2$  e  $rm \leq 6$ ,  $rm = 2$  possui o menor desvio padrão ( $SD = 0,018$ ) entre as 1000 execuções de classificação. Assim, experimentos realizados mostraram que o par ( $rm = 2, \sigma = 4$ ) fornece ao método melhor consistência de resultados. Esse achado confirma que a inclusão da membrana nuclear e partes do citoplasma que estão no entorno do núcleo (Figura 51) é relevante para detectar células cervicais anormais. No entanto, áreas em que  $rm > 6$  levam à inclusão de regiões do citoplasma com textura não discriminativa entre células normais e anormais, resultando em decréscimo de desempenho do descritor proposto. Detalhamos os resultados para cada  $\sigma$  em  $rm = 2$  na Figura 50b, e apresentamos o histograma para o par ( $rm = 2, \sigma = 4$ ) na Figura 50c. É importante notar que existe um valor de  $\sigma$  que maximiza  $ACC$  em cada  $rm$ , valores superiores ou inferiores a este também levam a perda de desempenho.

Analizamos a matriz rSCM comparando seus resultados aos de descritores estado-da-arte. Para cada descritor, utilizamos as mesmas configurações definidas em (SILVA *et al.*, 2019). Dividimos esses algoritmos em três categorias de acordo com sua entrada: - Categoria 1 (C#1): necessitam apenas da segmentação do núcleo; - Categoria 2 (C#2): necessitam da segmentação do núcleo e do citoplasma e Categoria 3 (C#3): são independentes de segmentação, necessitando apenas da definição de uma ROI retangular. Representamos C#3 pela versão de propósito geral do descritor proposto, *Structural Co-occurrence Matrix* (SCM), no qual a coocorrência estrutural é computada em toda área da ROI. Para a SCM, estimamos a quantidade de níveis de quantização

<sup>4</sup> Nossa metodologia de análise (Figura 49) realiza balanceamento entre as classes, logo a computação de  $ACC$  (Equação B.1) não possui viés para a classe majoritária.



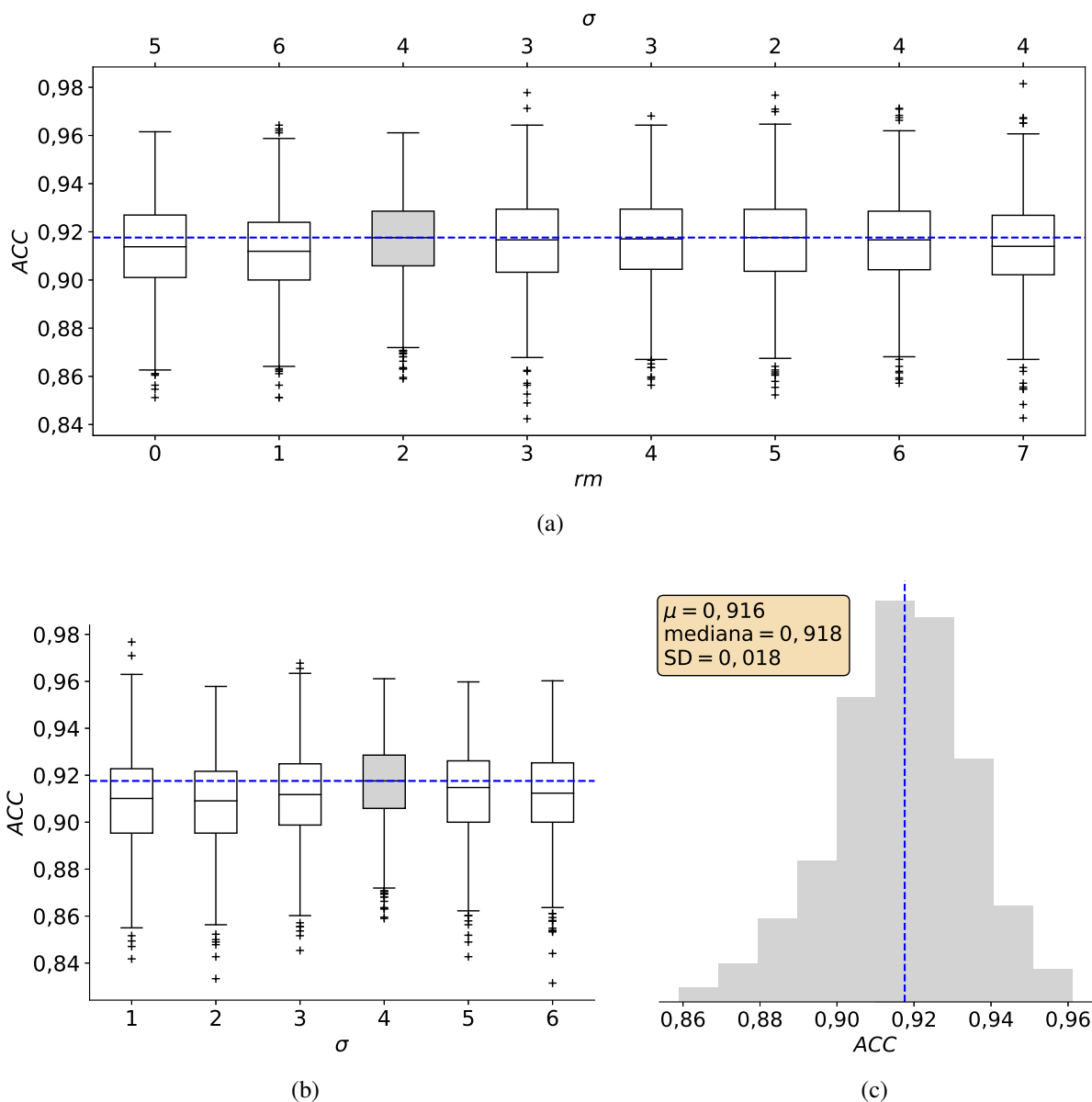


Figura 50 – Análise da matriz rSCM como descritor de células cervicais. a) Desempenho com máscara de segmentação do núcleo de diferentes tamanhos ( $rm$ ), destacando a abertura do filtro gaussiano ( $\sigma$ ) de melhor desempenho. b) Resultados para os valores de  $\sigma$  analisados, considerando  $rm = 2$  e c) histograma das 1000 rodadas de classificação para o par ( $rm = 2, \sigma = 4$ ). A linha azul destaca o melhor resultado conforme valor de mediana. Os elementos em cinza correspondem à mesma população de valores de Acurácia (ACC). Fonte: elaboração própria.

$N_Q$  da função  $Q$  (Seção 6.1.2) e  $\sigma$  que maximizam ACC por *grid search*, encontrando o par ( $N_Q = 8, \sigma = 8$ ). Abordamos os resultados da SCM como controle para avaliarmos a relevância de restringir o processamento do descritor à região do núcleo. Descritores clássicos de textura, como GLCM (HARALICK *et al.*, 1973), LPB (OJALA *et al.*, 1996) e GLRLM (GALLOWAY, 1975; TANG, 1998), não foram considerados por apresentarem desempenho inferior à RFD (SILVA *et al.*, 2019). Assim, focamos a análise apenas em descritores especializados em células

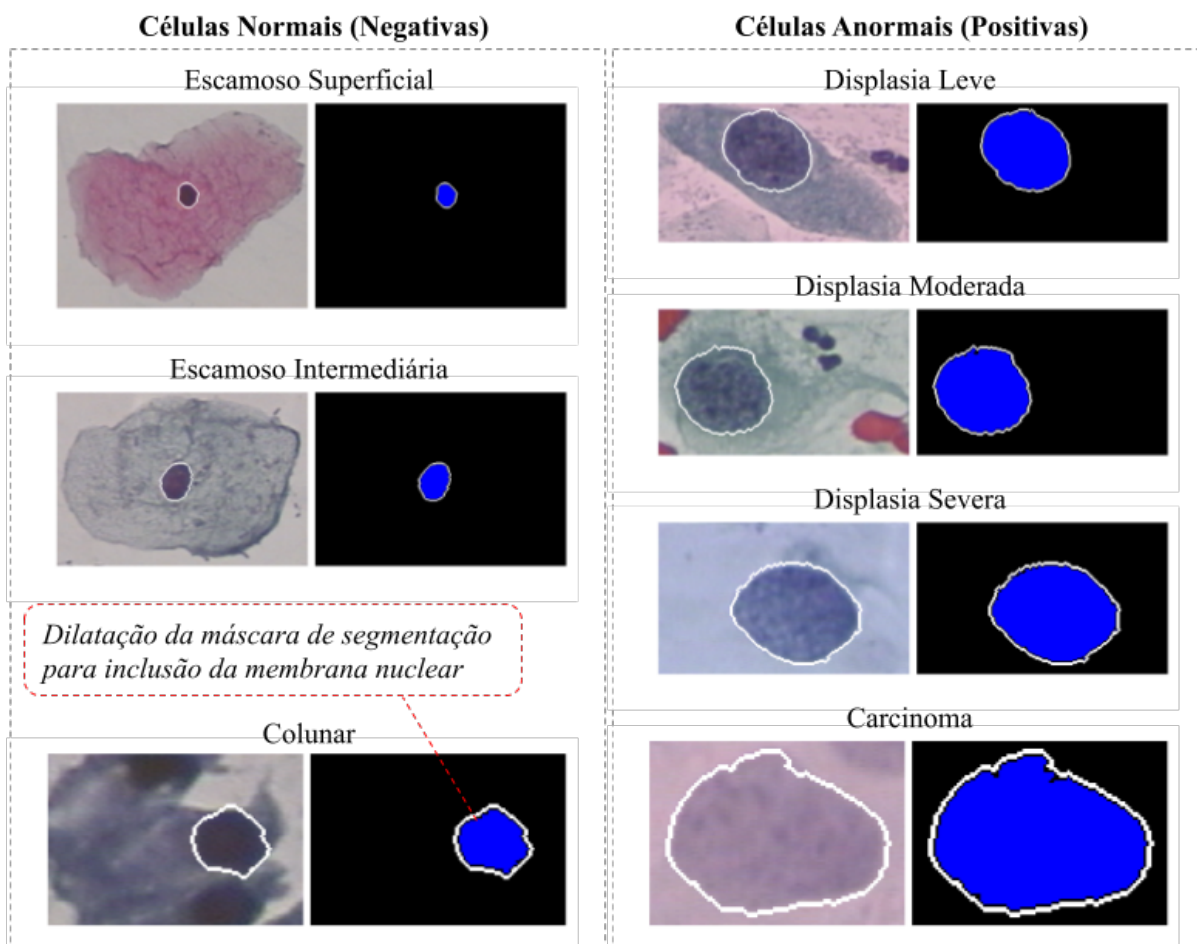


Figura 51 – Exemplos de células cervicais destacando a área nuclear periférica (bordas brancas) processada pelo algoritmo rSCM. A área em azul representa o núcleo. Fonte: elaboração própria a partir dos dados da base Herlev (JANTZEN *et al.*, 2005).

cervicais. A Tabela 8 resume esses algoritmos e os posiciona conforme a categorização adotada.

A Figura 52 apresenta os resultados de Acurácia ( $ACC$ ) e Kappa ( $K$ ), com valores detalhados na Tabela 9. Utilizamos a mediana dos resultados de 1000 rodadas de classificação (Figura 49), tornando a análise mais robusta a *outliers*. Incluímos também o intervalo de confiança (95%) de cada descritor. Os algoritmos da Categoria 2 apresentaram desempenho superior aos demais, destacando o método proposto por GençTav *et al.* (2012) com os melhores resultados. O algoritmo rSCM superou todos descritores de sua categoria (ou seja, C#1), além da metade dos algoritmos da Categoria 2. Também observamos ganhos de  $ACC$  e  $K$ , respectivamente, de 17,87% e 49,97% em relação à SCM (controle), onde o descritor é computado para toda área da ROI. A justificativa que temos é que a textura do citoplasma celular não contém informação relevante para detectar amostras anormais. Como a SCM (controle) compacta a informação de textura como uma matriz de coocorrência, informação relevante e não relevante são dispostas como soma de ocorrências na matriz, contribuindo para a confusão entre as classes. Além disso

Tabela 8 – Relação dos descritores avaliados e sua respectiva categorização quanto à entrada.  
Fonte: elaboração própria.

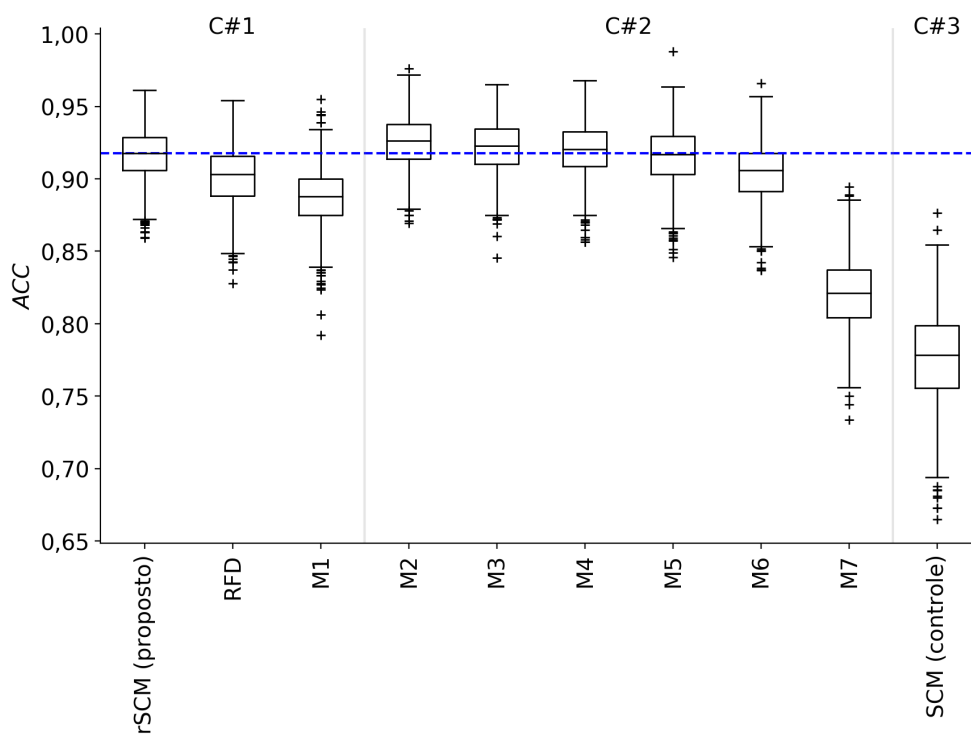
	<b>Método</b>	<b>Descreve:</b>
Segmentação do Núcleo (C#1)	rSCM (Proposta)	Informação de textura da região nuclear e entorno por meio de uma matriz de coocorrência. O vetor de características possui 64 elementos, correspondendo à vetorização da matriz. Detalhes na Seção 6.1.2.
	RFD (SILVA <i>et al.</i> , 2019)	Textura do núcleo e do citoplasma por meio de um Histograma Radial (RH), o qual modela a fronteira núcleo-citoplasma. Avaliamos RFD como 1) RH associado à matriz GLRLM (GALLOWAY, 1975) que computa as variações de intensidade no núcleo e 2) RH apenas. Não observamos diferença estatística relevante ( $\alpha = 0,05$ ) entre (1) e (2), logo reportamos (2), o qual obteve melhor desempenho. Vetor de características com 256 elementos. Os resultados apresentados correspondem ao melhor conjunto de parâmetro nas 1000 rodadas de classificação.
	(PLISSITI; NIKOU, 2012)	Conjunto de 9 características, incluindo área, diâmetro, alongamento, brilho, máximo e mínimo e intensidade do núcleo.
Segmentação do núcleo e do citoplasma (C#2)	(GENÇTAV <i>et al.</i> , 2012)	Conjunto de 14 características relacionadas a tamanho, cor, forma e textura do núcleo e do citoplasma. Estes atributos incluem maior e menor diâmetro do núcleo, perímetro da região nuclear, relação de tamanho entre núcleo e citoplasma, intensidade média no núcleo e citoplasma, dentre outros.
	(SARWAR <i>et al.</i> , 2015)	Conjunto com 20 características morfológicas, como o tamanho do núcleo, tamanho do citoplasma, além de intensidade de <i>pixels</i> do núcleo e do citoplasma, dentre outras.
	(MARINAKIS <i>et al.</i> , 2009)	Atributos de forma, como área, diâmetro e alongamento, além da intensidade, brilho, máximo e mínimo, computados dentro de uma vizinhança 3x3. Vetor de características com 20 elementos.
	(CHEN <i>et al.</i> , 2014)	Conjunto de 13 características que incluem atributos morfológicos, como perímetro, área, relação núcleo-citoplasma, e de textura, como contraste, entropia e <i>coarseness</i> .
	(CHANKONG <i>et al.</i> , 2014)	Conjunto de 9 atributos, sendo 6 calculados apenas no núcleo e 3 com informação do citoplasma. Dentre as características, temos: área do núcleo, diâmetros de uma elipse que envolve o núcleo, homogeneidade do núcleo, área total da célula.
	(MARIARPUTHAM; STEPHEN, 2015)	Atributos de forma e textura. Combina um conjunto de 7 descritores, incluindo o tamanho relativo de núcleo e citoplasma, momentos estatísticos quanto à intensidade dos <i>pixels</i> no núcleo e citoplasma, posicionamento do núcleo no citoplasma, além dos descritores GLCM, LBP, dentre outros.
ROI (C#3)	SCM (Controle)	Textura da imagem, similar à rSCM, mas aplicado em toda área da ROI. Compacta a informação de textura do núcleo, do citoplasma e eventualmente de parte do fundo da imagem na matriz de coocorrência. Assumimos a SCM como <i>controle</i> para verificar a importância de restringir o processamento à região do núcleo.

Tabela 9 – Ranque de resultados dos descritores quanto à classificação das células cervicais em normais e anormais. O valor da mediana para Acurácia (ACC) e Kappa (K) são reportados. O resultado do descritor proposto está destacado em cinza. IC corresponde ao intervalo de confiança (95%) para cada decritor. Os métodos estão listados na sequência do melhor para o pior desempenho. Cat: Categoria do método. Fonte: elaboração própria.

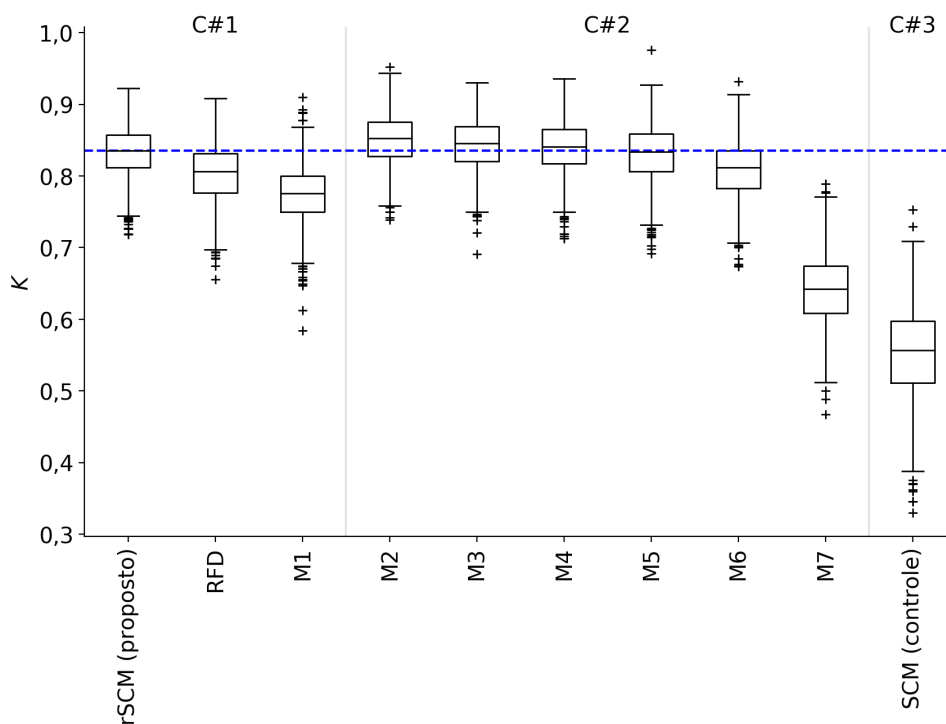
<b>Métodos</b>	<b>ACC Mediano (SD) IC* (%)</b>	<b>K Mediano (SD) IC* (%)</b>	<b>Cat.</b>
(GENÇTAV <i>et al.</i> , 2012)	0,9261 (± 0,0178) 88,82 a 95,73	0,8523 (± 0,0357) 77,64 a 91,46	2
(SARWAR <i>et al.</i> , 2015)	0,9229 (± 0,0184) 88,37 a 95,57	0,8458 (± 0,0367) 76,74 a 91,14	2
(MARINAKIS <i>et al.</i> , 2009)	0,9203 (± 0,0180) 88,17 a 95,21	0,8407 (± 0,0359) 76,34 a 90,43	2
rSCM (proposta)	0,9176 (± 0,0178) 87,82 a 94,89	0,8352 (± 0,0355) 75,64 a 89,78	1
(CHEN <i>et al.</i> , 2014)	0,9167 (± 0,0198) 87,50 a 95,00	0,8333 (± 0,0395) 74,99 a 90,00	2
(CHANKONG <i>et al.</i> , 2014)	0,9059 (± 0,0196) 86,47 a 94,32	0,8118 (± 0,0392) 72,94 a 88,64	2
RFD	0,9032 (± 0,0199) 86,25 a 94,05	0,8065 (± 0,0399) 72,50 a 88,10	1
(PLISSITI; NIKOU, 2012)	0,8876 (± 0,0202) 84,42 a 92,53	0,7753 (± 0,0404) 68,83 a 85,06	1
(MARIARPUTHAM; STEPHEN, 2015)	0,8212 (± 0,0255) 76,88 a 87,20	0,6425 (± 0,0511) 53,76 a 74,39	2
SCM (controle)	0,7784 (± 0,0314) 71,11 a 83,33%	0,5569 (± 0,0629) 42,22 a 66,67	3

a matriz SCM (controle) não considera informação quanto ao tamanho do núcleo, o qual é uma característica importante para detectar lesão. Assim, a inserção da máscara de segmentação do núcleo para o algoritmo rSCM foi valiosa nessa proposta. É importante ressaltar que as posições de ranqueamento dos descritores são constantes em relação às métricas estudadas, sugerindo a não existência de viés na computação das mesmas pelo balanceamento de classes.

O teste de Kruskal-Wallis, combinado com o teste de Nemenyi, não revelou diferença estatística ( $\alpha = 0,05$ ) entre o algoritmo rSCM e os métodos propostos por Marinakis *et al.* (2009) e Chen *et al.* (2014) da Categoria 2 (Figura 53). Esses testes também apontaram diferença



(a)



(b)

Figura 52 – Resultados de a) Acurácia (ACC) e b) Kappa (K) para os descritores analisados na caracterização de células cervicais normais e anormais. A linha tracejada destaca o desempenho do algoritmo proposto. C#{1,2,3} identifica a categoria por grupo de descritores. M1: (PLISSITI; NIKOU, 2012), M2: (GENÇTAV *et al.*, 2012), M3: (SARWAR *et al.*, 2015), M4: (MARINAKIS *et al.*, 2009), M5: (CHEN *et al.*, 2014), M6: (CHANKONG *et al.*, 2014) e M7: (MARIARPUTHAM; STEPHEN, 2015). Fonte: elaboração própria.

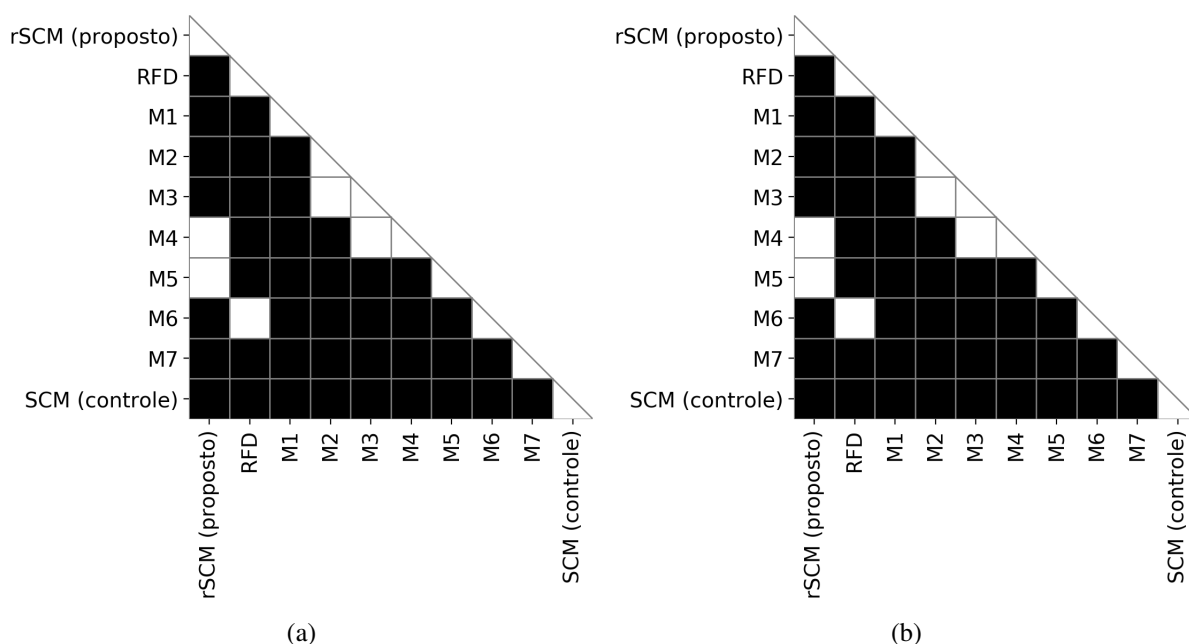


Figura 53 – Análise em pares dos descritores usando o teste estatístico de Kruskal-Wallis com o teste de Nemenyi. a) Acurácia ( $ACC$ ) e b) Kappa ( $K$ ). Os quadrados pretos representam os pares com significativa diferença na caracterização das células cervicais ( $\alpha = 0,05$ ). M1: (PLISSITI; NIKOU, 2012), M2: (GENÇTAV *et al.*, 2012), M3: (SARWAR *et al.*, 2015), M4: (MARINAKIS *et al.*, 2009), M5: (CHEN *et al.*, 2014), M6: (CHANKONG *et al.*, 2014) e M7: (MARIARPUTHAM; STEPHEN, 2015). Fonte: elaboração própria.

significativa para o descritor RFD, mostrando que o rSCM tem ganhos dentre os métodos da Categoria 1. Assim o algoritmo rSCM revelou-se promissor para classificação binária de células cervicais, permitindo alcançar resultados similares ao de algoritmos estado-da-arte a partir da segmentação apenas do núcleo. Esse achado torna-se relevante do ponto de vista computacional, desde que segmentar o núcleo é comumente menos complexo que segmentar a estrutura do citoplasma, devido a maior chance desta estar contaminada por artefatos e/ou sobreposição celular (Figura 2). Os resultados da análise estatística apresentados na Figura 53 foram iguais para os valores de  $ACC$  e  $K$  devido a metodologia de balanceamento descrita na Seção B.3.

## B.5 Conclusões

A matriz rSCM proposta tem potencial para descrever os padrões de textura que caracterizam as células cervicais em normais e anormais. As diferenças estruturais nos padrões de cromatina do núcleo, na membrana nuclear e nas transições núcleo-citoplasma, codificadas na matriz, são discriminantes valiosos para a identificação de células cervicais anormais. Esse achado permitiu que a matriz proposta descrevesse células com resultados superiores a métodos

da literatura, dispondo apenas da segmentação do núcleo. Adicionalmente, mostramos que a inserção de grandes áreas de textura do citoplasma leva à redução da discriminação entre as classes normal e anormal. Dessa forma, esta tese contribui com uma solução alternativa a algoritmos dependentes da segmentação do citoplasma para extração de características.

**ANEXO A - PROTOCOLO DE APROVAÇÃO NO COMITÊ DE ÉTICA DA UFC**



**PARECER CONSUBSTANCIADO DO CEP**

**DADOS DO PROJETO DE PESQUISA**

**Título da Pesquisa:** Estudo da Atenção Visual Seletiva de Estudantes e Profissionais em Citologia na Análise de Exames de Papanicolaou

**Pesquisador:** DANIEL SILVA FERREIRA

**Área Temática:**

**Versão:** 1

**CAAE:** 81055617.5.0000.5054

**Instituição Proponente:** UNIVERSIDADE FEDERAL DO CEARÁ

**Patrocinador Principal:** Financiamento Próprio

**DADOS DO PARECER**

**Número do Parecer:** 2.439.252

**Apresentação do Projeto:**

O Exame de Papanicolaou é a mais popular técnica de rastreamento para prevenção e detecção precoce do câncer de colo do útero, no entanto, sua avaliação é qualitativa e dependente da proficiência do especialista. Devido a essa subjetividade, a acurácia desse exame torna-se crítica quando as amostras de células são coletadas pelo método convencional, largamente utilizado pelo SUS, pois a presença de artefatos prejudica a inspeção

visual do especialista no microscópio. Nos últimos anos, vários métodos computacionais foram desenvolvidos para auxiliar o citologista no diagnóstico de lesões em exames de Papanicolaou. No entanto, devido à característica ruidosa das imagens, os resultados obtidos ainda não são satisfatórios. Para eliminar regiões não importantes em uma imagem e reduzir a complexidade no reconhecimento de objetos, cientistas tem

estudado a atenção seletiva humana para aprimorar/desenvolver sistemas de visão computacional. Dessa forma, é possível fornecer informações

aos algoritmos que os permitem simular a atenção humana, eliminando objetos desinteressantes para o entendimento da cena. Estudos de atenção também tem contribuído com pesquisadores que buscam explicar causas que levam a falha de diagnóstico em exames médicos. Neste contexto, nós propomos coletar dados de atenção visual de estudantes e profissionais em citologia para desenvolver sistemas computacionais de auxílio ao diagnóstico de câncer cervical. Propomos também realizar um estudo sobre os padrões de leitura de estudantes e citologistas

**Endereço:** Rua Cel. Nunes de Melo, 1000

**Bairro:** Rodolfo Teófilo

**CEP:** 60.430-275

**UF:** CE

**Município:** FORTALEZA

**Telefone:** (85)3366-8344

**E-mail:** comepe@ufc.br

Continuação do Parecer: 2.439.252

com diferentes níveis de experiência e avaliar o impacto dessas estratégias em situações de erro no diagnóstico. Além disso, pretendemos disponibilizar os dados coletados, bem como os resultados obtidos, à comunidade científica. Acreditamos que esse projeto tem potencial para contribuir com pesquisadores de várias áreas, em especial, com aqueles que desenvolvem metodologias para redução da mortalidade por câncer do colo do útero utilizando imagens similares as do SUS.

### **Objetivo da Pesquisa:**

Objetivo Primário:

Aprimorar/desenvolver métodos de visão computacional que utilizem a atenção seletiva de especialistas para auxiliar no diagnóstico precoce do câncer do colo do útero.

Objetivo Secundário:

1. Avaliar como as estratégias de leitura dos participantes se relacionam com casos de erro no diagnóstico.
2. Modelar protótipos de softwares que se utilizem da simulação de mapas de atenção de especialistas para formação e capacitação de profissionais em citologia.
3. Publicação de uma base de dados sobre atenção de citopatologistas em exames de Papanicolaou.

### **Avaliação dos Riscos e Benefícios:**

Riscos:

1. Leve desconforto nos olhos após algum tempo de experimento, devido à manutenção do foco no monitor para observar as imagens.
2. Leve desconforto nas costas, devido o participante permanecer sentado e sem poder realizar grandes movimentos.

Benefícios:

1. Certificado de participação em experimento de pesquisa acadêmica emitido por Grupo de Pesquisa devidamente reconhecido pela UFC e pelo CNPq.
2. Possibilidade do participante contribuir com trabalhos que visam reduzir a mortalidade por câncer do colo do útero.

### **Comentários e Considerações sobre a Pesquisa:**

Pesquisa de relevância na medida em que vai permitir maior conforto ao examinador e um melhor protocolo para os exames de Papanicolaou.

**Endereço:** Rua Cel. Nunes de Melo, 1000

**Bairro:** Rodolfo Teófilo

**CEP:** 60.430-275

**UF:** CE

**Município:** FORTALEZA

**Telefone:** (85)3366-8344

**E-mail:** comepe@ufc.br

Continuação do Parecer: 2.439.252

**Considerações sobre os Termos de apresentação obrigatória:**

Todos os termos foram devidamente apresentados.

**Recomendações:**

Sem recomendações.

**Conclusões ou Pendências e Lista de Inadequações:**

Sem pendências.

**Considerações Finais a critério do CEP:**

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_P ROJETO_1049888.pdf	12/12/2017 17:08:10		Aceito
Projeto Detalhado / Brochura Investigador	projetoComiteEtica_DANIEL_CORRIGID O.pdf	12/12/2017 17:05:00	DANIEL SILVA FERREIRA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_DANIEL_CORRIGIDO.pdf	12/12/2017 17:03:32	DANIEL SILVA FERREIRA	Aceito
Outros	Parecer_CEP_base_de_imagens.pdf	07/12/2017 19:47:20	DANIEL SILVA FERREIRA	Aceito
Outros	lattes_DANIEL.pdf	07/12/2017 19:40:37	DANIEL SILVA FERREIRA	Aceito
Outros	carta_apresentacao_DANIEL.pdf	07/12/2017 19:38:37	DANIEL SILVA FERREIRA	Aceito
Orçamento	orcamento_DANIEL.pdf	07/12/2017 19:34:01	DANIEL SILVA FERREIRA	Aceito
Declaração de Pesquisadores	declaracao_pesquisadores_DANIEL.pdf	07/12/2017 19:33:26	DANIEL SILVA FERREIRA	Aceito
Outros	autorizacao_ppgeti_DANIEL.pdf	07/12/2017 19:31:11	DANIEL SILVA FERREIRA	Aceito
Outros	autorizacao_deti_DANIEL.pdf	07/12/2017 19:30:40	DANIEL SILVA FERREIRA	Aceito
Cronograma	cronograma_DANIEL.pdf	07/12/2017 19:25:44	DANIEL SILVA FERREIRA	Aceito
Folha de Rosto	folha_rosto_DANIEL.pdf	07/12/2017 19:07:05	DANIEL SILVA FERREIRA	Aceito

**Situação do Parecer:**

**Endereço:** Rua Cel. Nunes de Melo, 1000

**Bairro:** Rodolfo Teófilo

**CEP:** 60.430-275

**UF:** CE

**Município:** FORTALEZA

**Telefone:** (85)3366-8344

**E-mail:** comepe@ufc.br

Continuação do Parecer: 2.439.252

Aprovado

**Necessita Apreciação da CONEP:**

Não

FORTALEZA, 14 de Dezembro de 2017

---

**Assinado por:**  
**FERNANDO ANTONIO FROTA BEZERRA**  
**(Coordenador)**

**Endereço:** Rua Cel. Nunes de Melo, 1000

**Bairro:** Rodolfo Teófilo

**UF:** CE

**Município:** FORTALEZA

**CEP:** 60.430-275

**Telefone:** (85)3366-8344

**E-mail:** comepe@ufc.br