

Segmentação de núcleos de células cervicais em exame de Papanicolau

Paulo Henrique Calaes Oliveira
Universidade Federal de Ouro Preto

Dissertação submetida ao
Instituto de Ciências Exatas e Biológicas
Universidade Federal de Ouro Preto
para obtenção do título de Mestre em Ciência da Computação

Dedico este trabalho a minha família, minha base.

Segmentação de núcleos de células cervicais em exame de Papanicolau

Resumo

A utilização de algoritmos que possam auxiliar no diagnóstico do exame de Papanicolau vem sendo estudada ao longo das últimas décadas devido ao aumento dos casos de câncer cervical na população e respectivos dados coletados. Uma das etapas dessa automatização do diagnóstico é a segmentação automática das imagens. Alguns dos maiores problemas quando se realiza a segmentação deste tipo de imagens são: a sobreposição celular, o dobramento das células e os artefatos que se confundem aos núcleos. Então é apresentada uma nova abordagem de segmentação nuclear utilizando uma heurística associadas a um algoritmo genético multi-objetivo. O processo envolve três etapas principais, que são, o pré-processamento, a calibração da heurística e a segmentação dos núcleos. Experimentos realizados com bases de dados sintéticas disponibilizadas no *Overlapping Cervical Cytology Image Segmentation Challenge - ISBI 2014* e uma nova base de imagens reais sugerem uma melhoria na detecção dos núcleos em comparação com os resultados obtidos pelos vencedores do desafio. Desse modo, esse trabalho apresenta uma interface *web* colaborativa criada para a geração de uma base de dados com imagens reais e um método para segmentação de núcleos que utiliza uma heurística associada a um algoritmo evolutivo multi-objetivo.

Nuclei segmentation on PAP-Smear images

Abstract

The use of algorithms that can assist in the diagnosis of Pap smear has been studied over the past decades due to the increasing incidence of cervical cancer in the population and their collected data. One of the steps in automating diagnosis is the automatic segmentation of images. Some of the biggest problems when performing the segmentation of such images are: cell overlap, cell folding and artifacts that are confused with nuclei. Then a new approach of nuclear segmentation using a heuristic associated with a multi-objective genetic algorithm is presented. The process involves three main steps, which are pre-processing, the heuristic calibration and segmentation of the nuclei. Experiments with synthetic databases available on Overlapping Cervical Cytology Image Segmentation Challenge - ISBI 2014 and a new base of real images suggest an improvement in the detection of the nuclei compared with the results obtained by the winners of the challenge. Thus, this work presents a collaborative web interface created for the generation of a database with real images and a method for segmentation of nuclei that uses a heuristic associated with a multi-objective evolutionary algorithm.

Declaração

Esta dissertação é resultado de meu próprio trabalho, exceto onde referência explícita é feita ao trabalho de outros, e não foi submetida para outra qualificação nesta nem em outra universidade.

Paulo Henrique Calaes Oliveira

Agradecimentos

Agradeço a todos que me ajudam direta ou indiretamente neste trabalho.

Aos meus pais, grandes exemplos.

A minhas irmãs, por serem tão cúmplices, e preencher um pouco da minha ausência junto a meus pais.

A Eleonora, por me apoiar, e sempre acreditar em mim, mesmo quando eu mesmo fui incapaz disso, te amo!

A minha orientadora Andrea, pela paciência e por transmitir todos os ensinamentos.

A todos os professores pelos ensinamentos, em especial, ao Prof. Gladston.

As coordenadoras Andrea, Cláudia, Daniela e Fátima, e os demais colegas do projeto CRIC, por terem me acolhido e me deixar fazer parte dessa grande história.

Aos meus amigos, de Caratinga, da Velha Guarda, de Ouro Preto, da República Maternidade, obrigado por serem minha válvula de escape, quando tudo parece estar errado.

A CAPES, CNPQ e FAPEMIG, pelo apoio durante toda pesquisa.

Muito Obrigado.

Sumário

Lista de Figuras	xv
Lista de Tabelas	xvii
Nomenclatura	1
1 Introdução	3
1.1 Introdução	3
1.2 Objetivos	6
1.3 Justificativa	6
1.4 Organização do Texto	7
2 Revisão Bibliográfica	9
2.1 Trabalhos Relacionados	9
2.2 Referencial Teórico	15
2.2.1 Simple Linear Iterative Clustering (SLIC)	15
2.2.2 Density Based Spatial Clustering of Applications with Noise (DBSCAN)	17
2.2.3 Medidas de avaliação	17
3 Base de Dados	19

3.1	Base de Herlev	19
3.2	Base de Hacettepe	20
3.3	Base ISBI 2014	20
4	Metodologia	23
4.1	Pré-Processamento	24
4.2	Heurística CIA	24
4.3	Calibração dos Parâmetros da Heurística	25
4.4	Segmentação dos Núcleos	28
5	Resultados	31
5.1	CRIC Searchable Database	31
5.2	Experimentos	33
6	Conclusão	41
6.1	Desenvolvimentos Futuros	42
	Referências Bibliográficas	43

Lista de Figuras

1.1	Comparação entre uma imagem do exame em meio líquido(a) e uma imagem do exame convencional(b)	4
1.2	Exemplo de imagem real do exame de convencional de Papanicolau	5
2.1	Comparação do esquema busca do k-means e SLIC	16
3.1	Exemplo de imagens das células e suas segmentações da base de dados Herlev	20
3.2	Exemplo de imagens da base de dados Hacettepe	21
3.3	Exemplo de imagens da base de dados ISBI 2014	22
4.1	Ilustração das etapas de segmentação	23
4.2	Etapa de pré-processamento:(a) ilustração do SLIC, (b) resultado após aplicação do DBSCAN e (C) superpixels gerados.	24
4.3	Esquema de funcionamento heurística CIA	25
4.4	Exemplos de candidatos a núcleos e a avaliação das regiões segundo a heurística CIA.	26
4.5	Esquema de execução do algoritmo multi-objetivo	27
4.6	Esquema de execução do método de segmentação do núcleo	29
5.1	Exemplo de imagens da base de dados CRIC Database	31
5.2	Tela da ferramenta colaborativa CRIC Database	32

5.3	(a) Segmentação manual de células, (b) identificação dos núcleos e classificação das células	33
5.4	Fronteira Pareto para base <i>isbi_train_45</i>	34
5.5	Exemplo de segmentação de núcleos com e sem sobreposição celular	37
5.6	Exemplo de segmentação de núcleos com e sem sobreposição celular em imagem real	39
5.7	Exemplo de segmentação de núcleos com e sem sobreposição celular em imagem real	39

Lista de Tabelas

5.1	Elementos da Fronteira Pareto para base <i>isbi_train_45</i>	35
5.2	Resultados da Segmentação dos Núcleos para a base <i>isbi_train_45</i>	35
5.3	Elementos da Fronteira Pareto para a base <i>isbi_test_90</i>	36
5.4	Resultados da Segmentação dos Núcleos para a base <i>isbi_test_90</i>	36
5.5	Elementos da Fronteira Pareto para base de imagens reais	38

“Tente uma, duas, três vezes e se possível tente a quarta, a quinta e quantas vezes for necessário. Só não desista nas primeiras tentativas, a persistência é amiga da conquista. Se você quer chegar a onde a maioria não chega, faça o que a maioria não faz.”

— Bill Gates

Nomenclatura

AG	Algoritmo Genético
CIA	Circularidade, Intensidade e Área
CNN	<i>Convolutional Neural Networks</i>
DBSCAN	<i>Density Based Spatial Clustering of Applications with Noise</i>
FCM	<i>Fuzzy C-means clustering</i>
FN	Falso negativo
FP	Falso positivo
INCA	Instituto Nacional do Câncer
ISBI	<i>IEEE International Symposium on Biomedical Imaging</i>
MSER	<i>Maximally Stable Extremal Region</i>
NSGA-II	<i>Non-dominated Sorting Genetic Algorithm - II</i>
RF	<i>Random Forest</i>
SLIC	<i>Simple Linear Iterative Clustering</i>
SVM	<i>Supported Vector Machine</i>
TBS	<i>The Bethesda System</i>
TP	Verdadeiro positivo
TN	Verdadeiro negativo
ZSI	Índice de Similaridade Zijdenbos

Capítulo 1

Introdução

1.1 Introdução

De acordo com Instituto Nacional do Câncer(INCA) estima-se que mais de 16 mil novos casos de câncer do colo do útero foram diagnosticados no Brasil em 2016. Uma forma de se realizar o diagnóstico é através do exame de Papanicolau. É colhida uma amostra de células do colo do útero da mulher e posteriormente é avaliada por um especialista que analisa as características das células afim de realizar o diagnóstico utilizando a metodologia proposta no The Bethesda System(TBS)(Solomon, Davey, Kurman, Moriarty, O’Connor, Prey, Raab, Sherman, Wilbur, Wright Jr et al. 2002). Atualmente são realizados mais de 10 milhões desses exames anualmente, sendo que eles são analisados manualmente por especialistas e verificados posteriormente para atestar a qualidade dos mesmos, gerando assim um grande volume de trabalho e um elevado custo de mão de obra.

O exame de Papanicolau pode ser realizado em dois formatos, o tradicional conhecido como exame de esfregaço, que consiste em coletar com uma espátula amostras celulares das paredes vaginais e da superfície cervical, e com uma escova para aquisição de material endocervical. Essas amostras são esfregadas em lâminas microscópicas para serem analisadas por especialistas. A segunda opção é o exame em meio líquido que utiliza a mesma técnica de coleta das amostras, mas em vez de esfregar as amostras na lâmina, são armazenadas em um frasco contendo um líquido conservante e encaminhadas para laboratório onde são processadas eletronicamente retirando assim artefatos (como hemácias, muco e etc.), colocando-as em uma lâmina de maneira clara e não amontoada.

Um exemplo dos dois modelos pode ser visto na Figura 1.1. Infelizmente o elevado custo do exame em meio líquido torna inviável sua utilização em países em desenvolvimento como é o caso do Brasil.

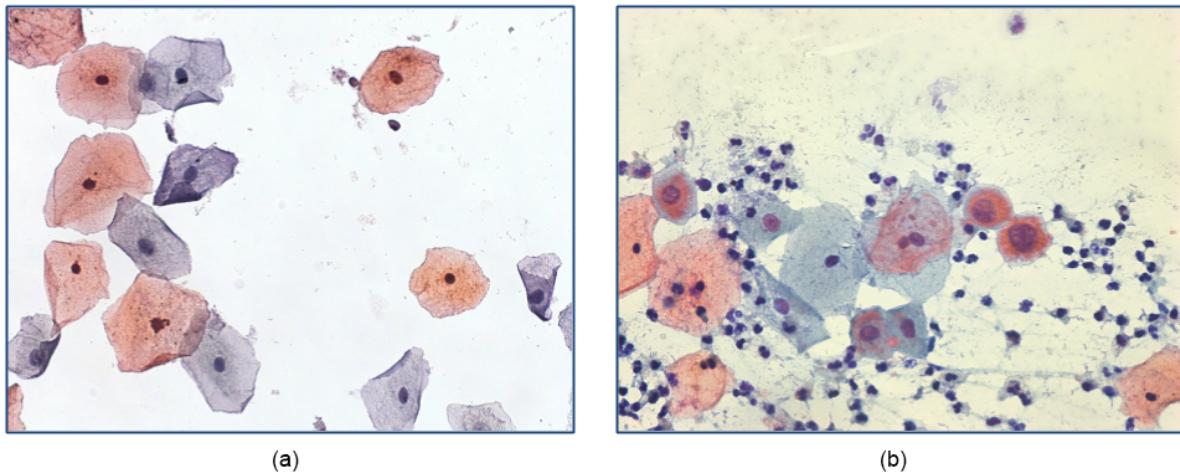


Figura 1.1: Comparação entre uma imagem do exame em meio líquido(a) e uma imagem do exame convencional(b)

Um passo importante para a automatização da análise desse exame inclui a segmentação das células cervicais, pois a individualização das mesmas auxilia no processo de classificação. Durante a inspeção das lâminas os patologistas costumam usar informações sobre a célula, sobre o seu núcleo e o seu citoplasma, assim como a razão entre eles. Entretanto, alguns autores (Plissiti & Nikou 2012a), (Lorenzo-Ginori, Curbelo-Jardines, López-Cabrera & Huergo-Suárez 2013), (Plissiti & Nikou 2012b), afirmam que apenas a informação do núcleo é suficiente para a classificação das células em suas diferentes classes de lesões. Os núcleos alterados apresentam alterações significantes como: crescimento desproporcional, irregularidade de forma e textura, hipercromasia, e irregularidade de cromatina(Plissiti, Nikou & Charchanti 2011).

Nessa dissertação é apresentado um método para segmentar núcleos de células cervicais. Como apresentado em (Bengtsson & Malm 2014), inúmeros métodos de segmentação já foram propostos para resolver esse problema, porém nenhum deles até então conseguiu resolver o problema de forma satisfatória. Em casos onde acontece sobreposição de células e dobramento das mesmas como mostrado na Figura 1.2 fica difícil a detecção das bordas e núcleos das respectivas células. A resolução desse problema resultaria em um grande passo na análise das amostras, pois uma vez segmentadas as células, podemos passar para a classificação das mesmas.

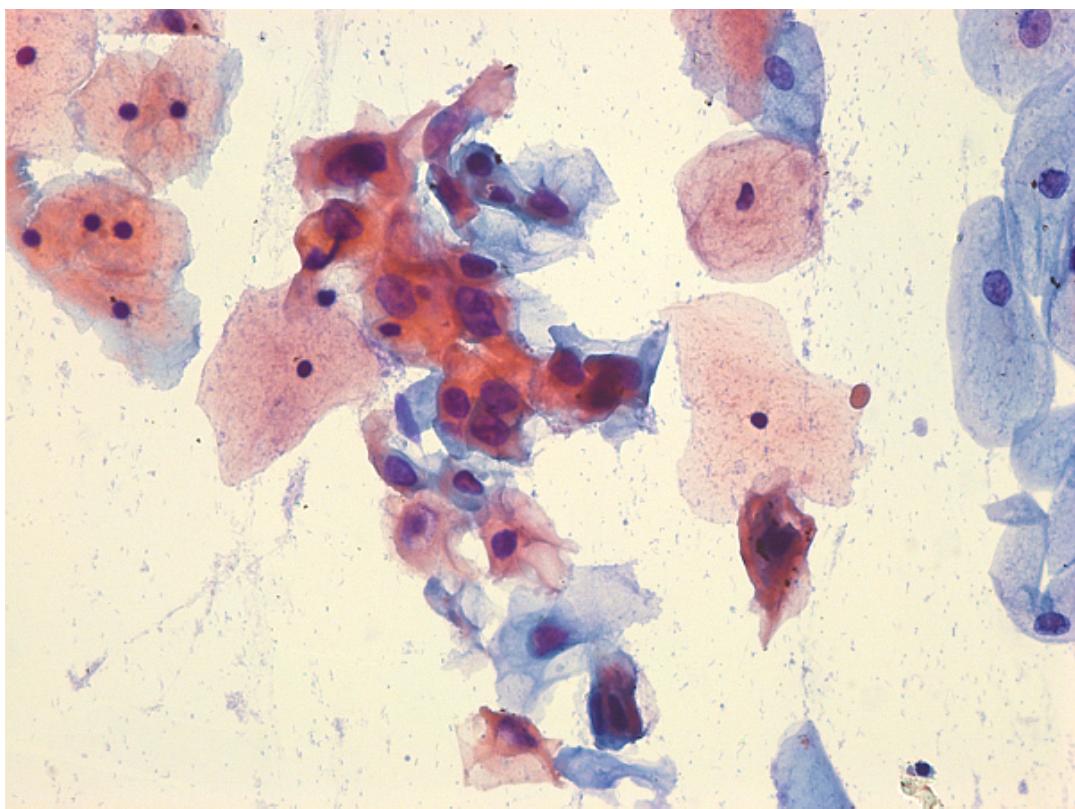


Figura 1.2: Exemplo de imagem real do exame de convencional de Papanicolau

Foi feita uma revisão das bases de dados disponíveis para realizar os experimentos, onde foram encontradas 3 bases de dados de imagens de células cervicais. Maiores detalhes podem ser vistos no Capítulo 3. A partir dessa revisão foi detectada a necessidade de uma base de dados de imagens reais resultantes de exame de esfregaço, visto que, nenhuma das 3 bases estudadas são desse tipo.

Os testes foram realizados na base de dados fornecida pelo primeiro Overlapping Cervical Cytology Image Segmentation Challenge que aconteceu no IEEE International Symposium on Biomedical Imaging (ISBI) 2014 (Lu, Carneiro, Bradley, Ushizima, Nosrati, Bianchi, Carneiro & Hamarneh 2017a) e em 22 imagens reais que foram utilizadas para a criação de uma nova base de dados. As imagens foram rotuladas utilizando uma ferramenta web colaborativa desenvolvida como parte desse trabalho e será melhor detalhada na Seção 5.1.

O método proposto para a segmentação dos núcleos é dividido em três etapas, (i) o pré-processamento, (ii) a calibração dos parâmetros da heurística e (iii) a segmentação dos núcleos. No pré-processamento é realizada a geração de *superpixels*, que são regiões con-

tendo grupos de *pixels*, geralmente menores que estruturas celulares, e que são avaliadas posteriormente nos outros dois passos. Nos dois passos restantes, uma heurística baseada em circularidade, intensidade e área é utilizada em conjunto com um algoritmo baseado em NSGA-II (Deb, Agrawal, Pratap & Meyarivan 2000).

1.2 Objetivos

O trabalho tem como objetivo principal apresentar um método de segmentação de núcleos de células cervicais e a comparação dos resultados com os métodos utilizados no *Overlapping Cervical Cytology Image Segmentation Challenge - ISBI 2014*.

A publicação de uma base de dados de imagens de lâminas resultantes do exame de Papanicolau é um objetivo secundário desse trabalho, visto que, hoje na literatura o conjunto de bases de imagens de células cervicais é bem limitado.

Espera-se ainda que o processo de segmentação realizado permita a classificação de células cervicais de modo que possam ser identificadas as células alteradas na amostra.

1.3 Justificativa

Segundo (Li & Najarian 2001) os erros de diagnósticos variam de 5% e 55%, portanto a criação de uma ferramenta que auxilie no diagnóstico e minimize essa taxa de erro seria um grande avanço para a diminuição da mortalidade causada pela doença, visto que, em casos onde as alterações celulares são detectadas de forma prematura, as chances de cura chegam próximo de 100%.

A detecção do núcleo para a classificação das imagens é um fator muito importante, pois muitos algoritmos utilizam sua localização como ponto de partida para a segmentação do citoplasma, como veremos com maiores detalhes na Seção 2.1. Muitos autores como (Plissiti & Nikou 2012a), (Lorenzo-Ginori, Curbelo-Jardines, López-Cabrera & Huergo-Suárez 2013), (Plissiti & Nikou 2012b), acreditam que apenas a informação contida no núcleo é suficiente para a classificação das células, eles utilizam informações de tamanho, forma, textura, borda, entre outras.

A criação de uma base de imagens reais se faz necessária para que os métodos possam ser testados e melhor avaliados, pois os trabalhos até então apresentados utilizam bases

de imagens simuladas, ou contendo apenas uma célula, ou em imagens de preparação em meio líquido. Em imagens reais de esfregaço a complexidade para a realização da segmentação é bem mais elevada, pois como podemos visualizar na Figura 1.2, a sobreposição celular é frequente, bem como o dobramento da mesma, além da ocorrência de artefatos que se confundem aos núcleos como as hemácias.

1.4 Organização do Texto

O texto desta dissertação encontra-se organizado da seguinte forma:

No Capítulo 1 são apresentadas uma introdução, os objetivos e as justificativas do trabalho de mestrado.

No capítulo 2 alguns aspectos importantes sobre segmentação de células cervicais, trabalhos relacionados e métricas de avaliação são apresentados.

No Capítulo 3 é realizada uma revisão das bases de dados utilizadas para experimentos de segmentação e classificação de células cervicais.

No Capítulo 4 são discutidas as metodologias utilizadas para o desenvolvimento do trabalho. São abordadas separadamente as três etapas do método proposto.

No Capítulo 5 estão os resultados obtidos com o método desenvolvido. Foram realizados experimentos de forma independente e conjunta entre as etapas do método. Ainda é apresentado uma ferramenta para rotulação e segmentação de imagens desenvolvida para gerar uma nova base de imagens reais.

No Capítulo 6, conclusões sobre o assunto são discutidas e uma série de possíveis trabalhos futuros são definidos.

Capítulo 2

Revisão Bibliográfica

2.1 Trabalhos Relacionados

O *Cytoanalyzer* foi a primeira tentativa de criar um sistema de análise automatizado do esfregaço de Papanicolau (Tolles & Bostrom 1956). A premissa do projeto, desenvolvido nos Estados Unidos, era que as células cancerígenas se diferenciariam das demais pelo tamanho e densidade ótica do seu núcleo. Em testes realizados e apresentados em (Spencer & Bostrom 1962) verificou-se que muitos falsos positivos foram encontrados, pois muitas estruturas foram classificadas como células alteradas, tais como, células sanguíneas, fios de tecido e muco. Nos anos sessenta o projeto foi descontinuado devido a esses problemas.

Em 1966 Prewitt e Mendelsohn apresentaram uma técnica para a segmentação de células baseada na localização dos mínimos no histograma de densidade ótica da imagem digitalizada (Prewitt & Mendelsohn 1966), sendo que os pontos de mínimos correspondem as fronteiras entre fundo e citoplasma, e citoplasma e núcleo. Dois problemas aparecem frequentemente nessa metodologia, o primeiro é que pode haver mais picos e vales no histograma do que os necessários para isolar citoplasma e núcleo, o segundo é que existem outros artefatos que comumente são confundidos com as células.

No Reino Unido um projeto baseado no tamanho do núcleo foi desenvolvido nos anos 60 (Spriggs, Diamond & Meyer 1968), porém por problemas análogos aos citados no projeto *Cytoanalyzer*, ele também foi descontinuado.

CYBEST (Watanabe, Group et al. 1974) foi uma tentativa japonesa desenvolvida

pela TOSHIBA nos anos 70, a tecnologia era baseada nas medidas de área e densidade nuclear, área citoplasmática e relação núcleo/citoplasma. Os pesquisadores ainda perceberam que a forma celular e o padrão da cromatina eram elementos importantes, porém não conseguiram utilizar tais recursos para a classificação das células devido a problemas com o foco automático que não conseguia, de forma confiável, imagens dos núcleos celulares. Num período de 15 anos foram desenvolvidas 4 gerações de protótipos, sendo que a última versão apresentou resultados promissores segundo (Tanaka, Ueno, Ikeda, Ishikawa, Yamauchi, Okamoto & Hosoi 1987), conseguindo classificar todas as células anormais de forma correta, entretanto apresentou 40% de falsos positivos provenientes células agrupadas e leucócitos polimorfonucleares. Mesmo com bons resultados, nenhum dos protótipos veio a se tornar um produto.

Em 1977, Cahn , Poulsen e Toussaint, propõem um método que combina limiarização e a detecção de bordas com heurísticas para realizar a segmentação e classificação das células cervicais (Cahn, Poulsen & Toussaint 1977). Seus resultados apresentam 90% de sucesso na classificação de células como normal e alterada. Entretanto, segundo seus autores, o algoritmo perde acurácia quando trabalha com grupos de células que se tocam ou que se encontram sobrepostas.

Ainda em 1977 (Aggarwal & Bacus 1977) utilizam uma combinação de duas técnicas, a limiarização e a clusterização para segmentação das células. A limiarização é utilizada para o núcleo e a clusterização para o citoplasma, defini-se um valor de pico inicial no histograma da imagem completa, então um *cluster* é inicializado para cada pixel que possui esse valor, feito isso o limite é reduzido e é feito a conexão com os *pixels* vizinhos de mesmo valor, se o pixel for conectado a mais de um *cluster* ele é atribuído ao *cluster* com centro mais próximo da massa, caso o pixel não possa ser atribuído a um *cluster* existente, um novo *cluster* é iniciado. Para a segmentação dos núcleos foi alcançada uma taxa de sucesso de 87,6% e de 88,1% para o citoplasma, foram utilizadas 233 imagens reais cedidas pelo *Cytology Laboratory, University of Illinois Medical Center*.

BioPEPR (Zahniser, Oud, Raaijmakers, Vooys & Van de Walle 1979), FAZYTAN (Erhardt, Reinhardt, Schlipf & Bloss 1979), Cerviscan (Tucker & Husain 1981), LEYTAS (Ploem, van Driel-Kulker & Verwoerd 1987), e Diascanner (Nordin 1989) foram metodologias similares ao projeto CYBEST, diferenciando apenas no que diz respeito às características extraídas. Um fator importante foi a melhoria na resolução e no foco durante a aquisição das imagens. O projeto Diascanner, visando melhoria de tempo, utilizou uma abordagem de digitalização em baixa resolução, seguida de outra digitalização de alta resolução apenas nas regiões de interesse. A maioria desses sistemas geraram

protótipos nos anos 80 com acurácia semelhante à conseguida nos exames convencionais, porém nenhum deles resultou em produto devido ao alto custo, a falta de microscópios automatizados e o custo elevado dos computadores com grande poder de processamento.

O PAPNET da Neuromedical Systems foi o primeiro projeto a utilizar interação com usuário no seu processo (Mango 1994). Assim como o Diascanner ele fazia a análise em duas etapas primeiramente em baixa resolução, e, nas regiões de interesse em alta resolução, além disso utilizava um algoritmo classificador e uma rede neural. Como saída dos classificadores ele retornava os 64 candidatos mais anormais, que eram analisados por um técnico em uma etapa de revisão. As células classificadas como anormais eram ainda analisadas por um citologista que fornecia a análise final.

O AutoPap 300 da NeoPath era semelhante ao PAPNET no uso de classificadores de redes neurais (Bernstein, Sanchez-Ramos & Ndubisi 2001). Analogamente ao CYBEST preocupou-se com o aumento na resolução da aquisição da imagem utilizando iluminação estroboscópica. Análogo ao Diascanner utilizou dois níveis de resolução no mapeamento das imagens, o primeiro com baixa resolução e o segundo com alta resolução nas regiões de interesse. A maior parte do processamento era baseado em morfologia matemática e resultou em 68 diferentes características a serem enviadas para os classificadores. Como resultado final ele classificava as células como "normal" ou "requer inspeção visual". O método foi avaliado em (Stevens, Milne, James, Brancheau, Ellison & Kuan 1997), onde pode ser visto que sua eficiência foi muito parecida com o analisador humano. O autor ainda sugere que se façam mais testes pois a base de imagens era limitada.

Em (Bamford & Lovell 1998) é apresentado um método de segmentação de núcleos baseado em contornos ativos. O método se mostrou bem comportado com relação aos artefatos que normalmente são confundidos com os núcleos, tanto quanto aos ruídos e problemas de inicialização. No entanto, 2 falhas foram reconhecidas: quando ocorre a presença de um grande gradiente na fronteira do núcleo, o algoritmo não se comporta bem. Outro problema é a definição do parâmetro λ , que se mal escolhido pode gerar falsos positivos confundindo os artefatos com núcleos. Num conjunto de 20130 imagens de células reais o algoritmo teve um índice de acerto de 99,64%, contudo esse conjunto de imagens possuía uma aproximação do núcleo de forma que o citoplasma não pode ser visualizado por completo.

Em 2002, N. Lassouaoui and L. Hamami, apresentaram um método para segmentação de células cervicais onde um algoritmo multifractal que era baseado no cálculo de expoentes de singularidade em cada um dos pontos da imagem. Logo em seguida, em

2003, foi proposta uma melhoria desse método utilizando algoritmos genéticos(AG) em (Lassouaoui & Hamami 2003). A utilização dos AG vem como um passo complementar a fim de otimizar o resultado, utilizando como função objetivo a qualidade da segmentação. Porém um problema apresentado foi a lentidão na execução do algoritmo segundo (Duanggate, Uyyanonvara & Koanantakul 2008).

Diferentemente dos métodos anteriores em (Plissiti, Charchanti, Krikoni & Fotiadis 2006), a base de dados utilizada para a realização dos testes possuía imagens complexas, onde se pode encontrar várias células por imagem, enquanto, nos trabalhos anteriores utilizava imagens com apenas uma célula. O método é composto por três etapas, o pré-processamento em que são utilizadas técnicas de equalização do histograma, limiarização e erosão morfológica. A segunda etapa consiste em estimar a fronteira do núcleo, através da aproximação do centroide do núcleo. Por fim na terceira etapa é realizada a aplicação do modelo deformável, também utilizado no trabalho proposto anteriormente. O método foi testado em várias imagens contendo um total de 462 núcleos de células, onde foram encontrados 94,15% dos núcleos e uma sensitividade de 95,94%. A desvantagem da metodologia está relacionada aos vários parâmetros que devem ser definidos para se chegar nesses resultados.

Os algoritmos apresentados até então utilizam base de dados onde não ocorre sobreposição celular e problemas de coloração das células, já em (Kale & Aksoy 2010) é apresentado um algoritmo genérico de segmentação das células que pode ser utilizado tanto em imagens que contém uma única célula quanto imagens que contém mais células. O algoritmo trabalha em duas etapas, na primeira ele utiliza a técnica de watershed multi-escala para criar regiões segmentadas em cada escala formando uma hierarquia de segmentos. Na segunda etapa os segmentos são classificados através de uma heurística onde são avaliados dois fatores,a homogeneidade e a circularidade; o primeiro deles é medido através do F-Statistic, e, o segundo fator é medido pela excentricidade do segmento. Para os experimentos foram utilizadas as bases Herlev contendo 917 imagens e a base Hacettepe contendo 82 imagens. A avaliação da segmentação das imagens da base Herlev foi feita utilizando o Índice de Similaridade Zijdenbos(ZSI), e obteve média acima de 0,9 para todas as classes. A avaliação da classificação da base Hacettepe segundo os autores obteve um índice de 96% de acertos, considerando núcleo e citoplasma.

O método apresentado por (Plissiti, Nikou & Charchanti 2011) também pode ser utilizado tanto em imagens que contém uma única célula, como imagens onde encontramos várias células. O método é dividido em quatro fases, sendo a primeira o pré-processamento, onde é eliminado o fundo diminuindo a região de interesse da imagem. Na segunda fase

é realizada a detecção dos candidatos a centroides dos núcleos através da técnica de reconstrução morfológica em escala de cinza (Vincent 1993) combinada à técnica de detecção da mínima regional apresentada por (Breen & Jones 1996). A terceira fase é onde acontece o refinamento dos candidatos, o método utiliza a informação de gradiente nas bordas dos núcleos para determinar seu centroide. A decisão é a quarta e última fase, é composta de duas etapas: aplicação da regra de Distância Dependente e a aplicação de técnicas de classificação. Os experimentos foram realizados numa base de 38 imagens que continham um total de 5617 núcleos que foram identificados por especialistas. A porcentagem de verdadeiros positivos após a segunda fase no melhor conjunto de parâmetros foi de 99,25% e acima de 75% de falsos positivos. Após a quarta fase foi obtido através do *Supported Vector Machine*(SVM) 69,86% de sensibilidade e 92,02% de especificidade, e 90,57% de sensibilidade e 75,28% de especificidade para o *Fuzzy C-means clustering*(FCM).

O método apresentado por (Bergmeir, Silvente & Benítez 2012) introduz um algoritmo de detecção de núcleos celulares que pode ser utilizado tanto em imagens com uma célula ou várias células. O método inicializa a fase de pré-processamento convertendo a imagem em escala de cinza, em seguida utiliza o método de Otsu (Otsu 1975) para segmentar o fundo da imagem, em seguida é aplicado o filtro Mean-Shift (Comaniciu & Meer 1999) para eliminar os ruídos da imagem, preservando as bordas e finalizando essa fase é utilizado o algoritmo de Canny para a detecção das bordas. A segunda fase consiste na localização dos núcleos, as regiões são testadas por diversos critérios que envolvem basicamente tamanho do diâmetro da elipse, posição do centro da elipse, excentricidade e direção do gradiente e caso seja reprovada em algum dos testes é descartada. A terceira fase consiste na adequação de uma elipse à forma real do núcleo, e para isso é utilizado o método Level Set. O método foi testado em 207 imagens contendo 549 núcleos. O método proposto alcançou uma taxa de 95,63% de verdadeiros positivos e F-measure de 96,15%.

(Gençtav, Aksoy & Önder 2012) apresenta um algoritmo genérico e livre de parâmetros capaz de segmentar núcleos e citoplasmas de imagens com coloração inconsistente, de baixo contraste e com células sobrepostas. Em um primeiro momento o algoritmo separa o fundo das regiões com células utilizando operações morfológicas e limiares automáticos. No segundo momento é construída uma árvore de segmentação hierárquica utilizando o algoritmo de Watershed multiescala(Olsen 1997), que automaticamente seleciona as regiões de núcleo através da medida de homogeneidade e circularidade. No terceiro momento finaliza a separação de núcleo e citoplasma utilizando um classificador binário.

Para a realização dos experimentos foi utilizada a base de dados Herlev com 917 imagens e a base Hacettepe com 82 imagens, sendo que o primeiro conjunto são de imagens que contém uma única célula e a segunda que contém diversas células por imagem. Para a base Hacettepe foi obtida uma precisão de 74,71%, revocação de 93,53% e ZSI de 89%.

(Song, Zhang, Chen, Ni, Li, Zhou, Lei & Wang 2014) apresenta uma abordagem baseada em SLIC e Redes Neurais de Convolução(CNN). Na primeira etapa é realizada um pré-processamento com o filtro de médias aparadas, visando a eliminação do ruído. Na segunda etapa é criada a máscara do citoplasma utilizando o método de limiar de Otsu. A terceira etapa é responsável pela criação dos *superpixels* através do SLIC. CNN é utilizado na quarta etapa para extrair características dos *superpixels*. Na quinta etapa é realizada uma segmentação grosseira do núcleo, usando o método de limiar adaptativo. Na etapa seguinte é realizada a correção do núcleo encontrado onde candidatos a núcleo são validados por uma rede neural e são expandidos em 5 *pixels*. Os testes foram realizados em 21 imagens, um total de 1400 células foram coletadas, sendo 1200 para treinamento da CNN e 200 para testes. Para detecção dos núcleos foi obtido uma acurácia de 94,5%, sensibilidade de 94,06%, especificidade de 94,95% e F1 de 94,53%.

Em 2014, ocorreu o primeiro desafio de segmentação de células cervicais sobrepostas dentro do International Symposium on Biomedical Imaging(ISBI), o desafio consistia em segmentar os citoplasmas e os núcleos das células cervicais. Um fator importante do desafio, é que, utilizava um banco de imagens composto por células reais, que eram aglomeradas por um algoritmo de modo a gerar imagens sintéticas com células sobrepostas.

O algoritmo proposto por (Nosrati & Hamarneh 2014) obteve a segunda colocação no desafio, e apresenta uma técnica que num primeiro momento é realizado a detecção do núcleo através do *Maximally Stable Extremal Region*(MSER) (Matas, Chum, Urban & Pajdla 2004) aliado a um classificador *Random Forest*(RF). Para a detecção do citoplasma, o núcleo é utilizado como centro de um modelo onde são calculadas as distâncias das bordas, o RF é utilizado para decidir se o pixel faz parte ou não daquele citoplasma. O método foi testado utilizando a base de dados fornecida pelo ISBI 2014 e obteve 98% de revocação e 99% de precisão para os núcleos na base de testes, no resultado oficial do ISBI 2014 obteve para detecção dos núcleos precisão de 97,7% e revocação de 88,3% e para segmentação do citoplasma obteve Dice 87,1%, FN 11,1%, TP 87,5% e FP 0,4%.

No mesmo desafio foi apresentado o algoritmo proposto por (Ushizima, Bianchi & Carneiro 2014) que obteve a primeira colocação. Para a detecção do núcleo é utilizado o método proposto por (Phansalkar, More, Sabale & Joshi 2011). A partir do núcleo

detectado são criados polígonos convexos com bordas equidistantes dos núcleos mais próximos, chamados polígonos de Voronoi. Os dados também foram testados utilizando a base de dados fornecida pelo ISBI 2014, e segundo o resultado oficial do desafio obteve para detecção dos núcleos precisão de 95,9% e revocação de 89,5% e para segmentação do citoplasma obteve Dice 87,2%, FN 26,7%, TP 84,1% e FP 0,2%.

(Nosrati & Hamarneh 2015), apresenta em 2015 seu novo algoritmo para segmentação de células cervicais. A detecção do núcleo agora utiliza o filtro bilateral seguido do método apresentado por (Phansalkar, More, Sabale & Joshi 2011) assim como (Ushizima, Bianchi & Carneiro 2014) e a transformada de Hough para selecionar os candidatos, utiliza ainda uma RF para determinar quais os candidatos foram selecionados como núcleo. O método foi testado no conjunto de imagens fornecido pelo ISBI 2014. Para a detecção dos núcleos foi encontrado o F1-value de 0,99 e para a segmentação do citoplasma foi obtido um Dice 88%, TP 93%, FP 0,05% e FN 11%.

Apesar das diversas tentativas que vem sendo feitas de segmentação de núcleos celulares, muitos desafios ainda persistem, principalmente no que diz respeito a imagens com muitas células e com sobreposição e dobramento das mesmas. São poucos os trabalhos que utilizam imagens reais, com muitas células por imagem, onde o nível de sobreposição e ocorrência de artefatos que se confundem aos núcleos é bem superior. Percebe-se que um complicador é o fato de não existir esse tipo de base disponível publicamente.

2.2 Referencial Teórico

Durante o desenvolvimento da proposta de solução, várias abordagens foram utilizadas e testadas, esta Seção apresenta o referencial teórico útil ao entendimento da abordagem utilizada.

2.2.1 Simple Linear Iterative Clustering (SLIC)

O *Simple Linear Iterative Clustering* (SLIC) (Achanta, Shaji, Smith, Lucchi, Fua & Susstrunk 2012) é uma adaptação do *K-means* para a geração de superpixels com duas importantes diferenças, o número de cálculos é otimizado, limitando o espaço de busca para uma região proporcional ao *superpixel*, como pode ser visto na Figura 2.1, e a medida da distância ponderada combina cor e proximidade espacial, simultaneamente,

proporcionando controle sobre o tamanho e compactação dos *superpixels*.

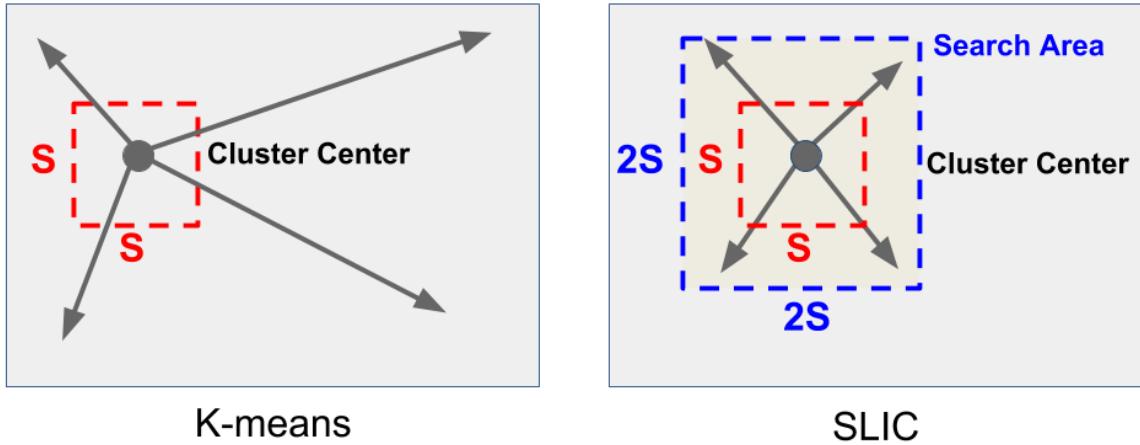


Figura 2.1: Comparação do esquema busca do k-means e SLIC

A equação utilizada para o cálculo do peso das distâncias é apresentado na Equação 2.1, onde d_c é a distância de cor e d_s é a distância espacial. Para calcular d_c o sistema de cores da imagem é convertido no formato *LAB*, onde L é a luminância (representada como uma escala que varia de 0 para preto e 100 para branco), A é a escala de cores que varia de verde a vermelho (-120 a +120), e B é outra escala de cor que varia de azul até amarelo (-120 a +120). Em d_s , as coordenadas x e y dos pixels são utilizadas. Em D' uma normalização de d_c e d_s é realizada, onde N_c e N_s são as distâncias máximas entre um cluster para cor e espaço. $N_s = S = \sqrt{N/k}$ é a distância espacial máxima esperada, onde N é o tamanho da imagem e k é o número desejado de superpixels. Entretanto, determinar N_c não é simples, pois as distâncias de cores podem variar de cluster para cluster e imagem para imagem. Assumindo N_c como uma constante m o problema é resolvido. A Equação 2.5 apresenta a média ponderada das distâncias representadas por D .

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (2.1)$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2.2)$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \quad (2.3)$$

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad (2.4)$$

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \quad (2.5)$$

2.2.2 Density Based Spatial Clustering of Applications with Noise (DBSCAN)

O método proposto por (Ester, Kriegel, Sander & Xu 1996) é utilizado para o agrupamento de regiões semelhantes. O algoritmo utiliza um ponto de partida, onde são avaliados os vizinhos daquele ponto e caso sejam similares estes passam a constituir um agrupamento. O processo é repetido várias vezes até que se encontrem todos os grupos ou *clusters*.

2.2.3 Medidas de avaliação

Para verificarmos a precisão dos resultados podemos fazer uma avaliação quantitativa e qualitativa, onde a primeira seria a contagem de verdadeiros positivos (TP), que são os objetos que deveriam ser encontrados e foram encontrados, dos falsos positivos (FP), que são aqueles que foram encontrados mas na realidade não deveriam, dos falsos negativos (FN), que deveriam ser encontrados e não foram, e dos verdadeiros negativos, que não deveriam ser encontrados e não foram. A avaliação qualitativa é feita a nível de *pixels*, comparando cada um deles com a verdade terrestre.

Derivados dessas medidas ainda temos a sensibilidade, que é a razão dos TP pelo conjunto total de indivíduos, e a especificidade que é a razão entre os TN pelos elementos negativos, e podem ser visualizado respectivamente nas equações 2.6 e 2.7.

$$sensibilidade = \frac{TP}{(TP + FN)} \quad (2.6)$$

$$especificidade = \frac{TN}{(TN + FP)} \quad (2.7)$$

Precisão e revocação são ainda medidas derivadas, onde a primeira avalia a capacidade do método em evitar a seleção de resultados não desejados como os FP, e a segunda, mede a capacidade de encontrar itens desejados como os TP e são obtidos, respectivamente, pelas equações apresentadas em 2.8 e 2.9.

$$Precisão = \frac{VP}{(VP + FP)} \quad (2.8)$$

$$Revocação = \frac{VP}{(VP + FN)} \quad (2.9)$$

A média harmônica da precisão e revocação é denominada valor F e também é utilizada na literatura para avaliação dos métodos e pode ser calculada através da equação 2.10.

$$F1 = 2 \frac{(Precisão \times Revocação)}{(Precisão + Revocação)} \quad (2.10)$$

Capítulo 3

Base de Dados

Para a realização do trabalho, foi feita uma revisão a respeito das bases de dados de imagens de células cervicais disponíveis na literatura, para que os testes fossem realizados. As bases encontradas foram a base Herlev(Jantzen, Norup, Dounias & Bjerregaard 2005), a base Hacetepe(Kale & Aksoy 2010) e as do *Overlapping Cervical Cytology Image Segmentation Challenge - ISBI 2014 e 2015*(Lu, Carneiro, Bradley, Ushizima, Nosrati, Bianchi, Carneiro & Hamarneh 2017a).

3.1 Base de Herlev

A base de imagens de Herlev é composta de 917 imagens que contém uma única célula cervical por imagem. As imagens foram divididas em 7 classes por especialistas. Sendo uma classe das células normais e mais 6 com diferentes níveis de alteração. A base ainda disponibiliza a máscara de segmentação para núcleo e citoplasma. Um exemplo de imagens das células da base juntamente com suas segmentações pode ser visto na Figura 3.1. Uma desvantagem encontrada na base é que não possui imagens com mais de uma célula, como são as imagens reais obtidas por microscópio, dessa forma, os algoritmos de segmentação não podem ser avaliados pela sua capacidade de isolar a célula do ambiente da lâmina, fato que se caracteriza como principal desafio no caso de exames de Papanicolau.

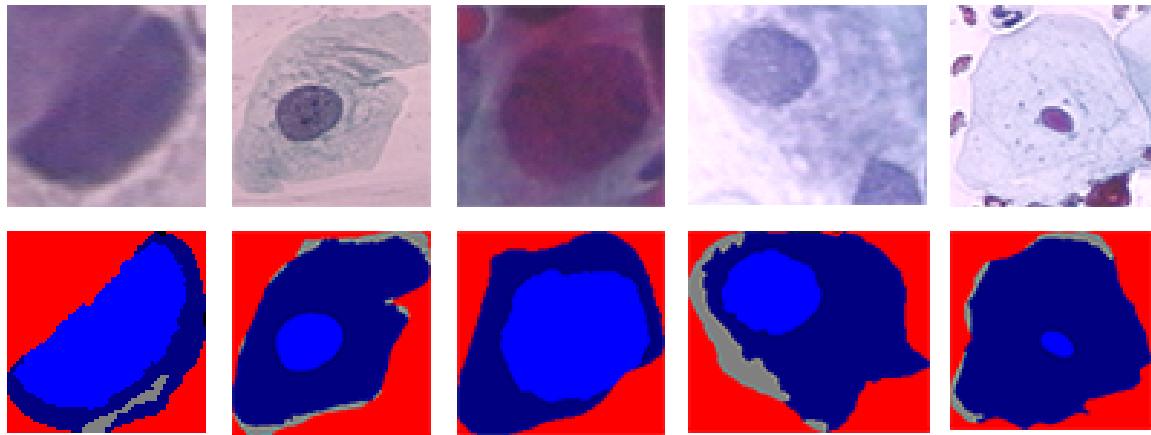


Figura 3.1: Exemplo de imagens das células e suas segmentações da base de dados Herlev

3.2 Base de Hacettepe

A base de imagens Hacettepe consiste em 82 imagens de 2048 x 2048 *pixels* contendo diversas células por imagem. Um exemplo de imagens da referida base pode ser visto na Figura 3.2 (Kale & Aksoy 2010) . A base não possui muita informação na literatura e não é disponibilizada publicamente o que dificulta a utilização da mesma. A base ainda aparenta ser de imagens de exame em meio líquido. Como a base não está disponível não é possível utilizá-la para comparações de diferentes metodologias de segmentação e classificação das células.

3.3 Base ISBI 2014

A base de imagens do ISBI composta por imagens sintéticas derivadas de imagens reais. As células foram recortadas manualmente de imagens reais e através de um algoritmo são reorganizadas a fim de gerar novas imagens com diversos níveis de sobreposição. A versão de 2014 contém 45 imagens de treinamento, 90 imagens para testes e 810 imagens para avaliação final do desafio, um exemplo de imagens da base pode ser visualizado na Figura 3.3. A grande vantagem da base ISBI é que além de disponibilizar as imagens de célula, a base também fornece as máscaras para seus núcleos e citoplasmas .

Como neste trabalho iremos realizar segmentações de núcleos em imagens de exames

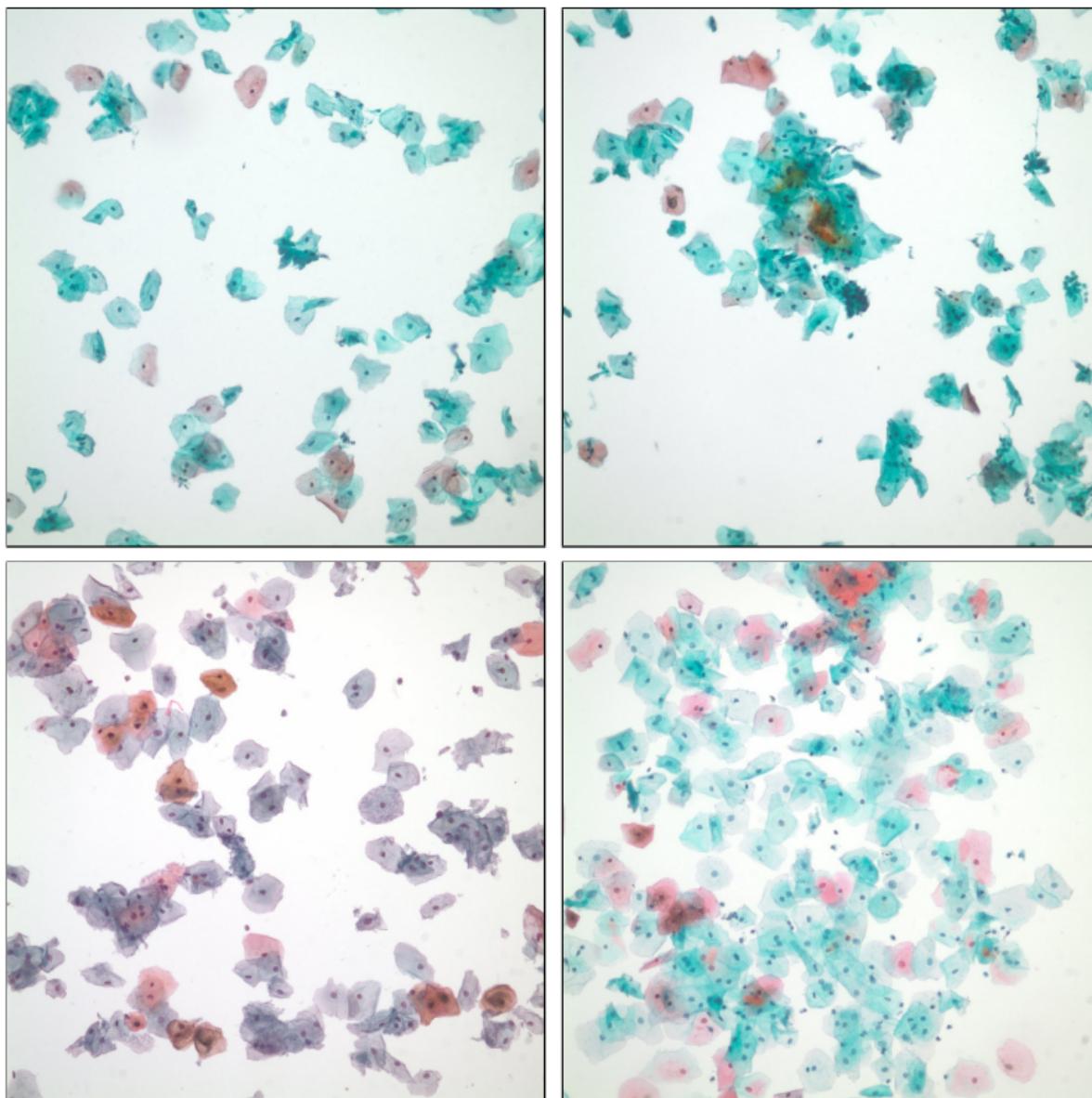


Figura 3.2: Exemplo de imagens da base de dados Hacettepe

de Papanicolaу, foi escolhida a base de dados do ISBI 2014 para a realização dos testes por ser uma base mais próxima das imagens reais e estar disponível publicamente, tanto as imagens quanto suas segmentações.

Como podemos observar, o número de base de dados é bastante limitado, não existe na literatura uma base que represente de forma satisfatória as imagens de esfregaço de Papanicolaу e que possuam informações sobre suas segmentações e classificações. Desse modo, como parte dos resultados obtidos neste trabalho está a criação de uma ferramenta web para a criação, segmentação, classificação e disponibilização de uma nova base de

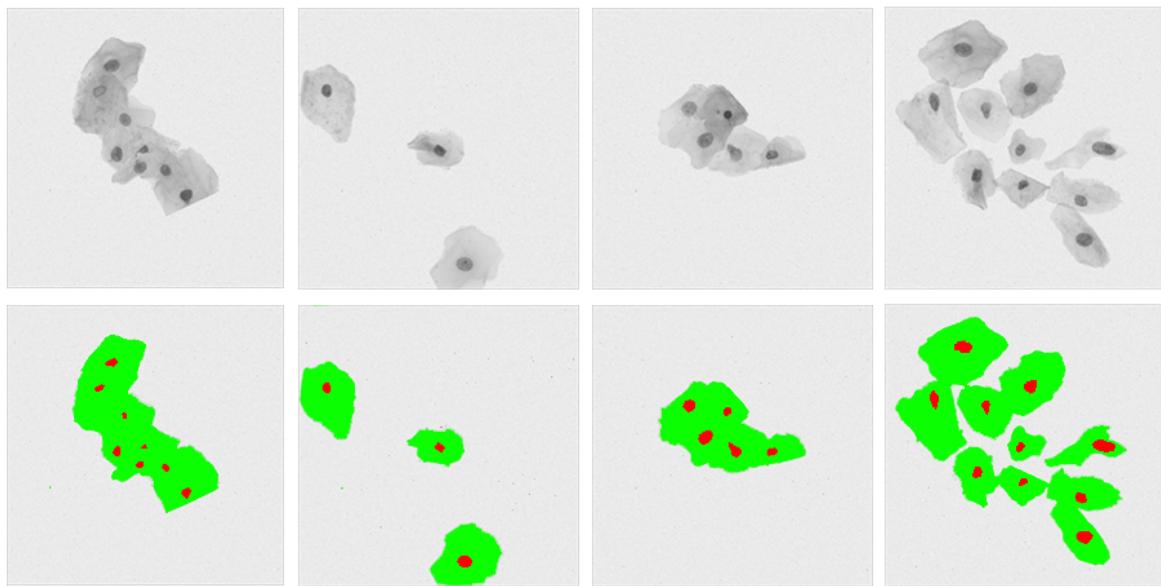


Figura 3.3: Exemplo de imagens da base de dados ISBI 2014

dados contendo imagens reais provenientes de exames de Papanicolau. A apresentação da base e suas funcionalidades pode ser visto em maiores detalhes na Seção 5.1 que apresenta os Resultados dessa dissertação.

Capítulo 4

Metodologia

A proposta principal da metodologia é que o núcleo celular pode ser definido por meio de informações relacionadas a sua morfologia, tais como área, circularidade e variações da intensidade, vários estudos (Plissiti & Nikou 2012a), (Lorenzo-Ginori, Curbelo-Jardines, López-Cabrera & Huergo-Suárez 2013), (Plissiti & Nikou 2012b) e também o contato com especialistas salientaram a sua importância para a definição de lesões durante um exame convencional. De tal forma que a identificação do núcleo era feita em etapas, com a definição de possíveis candidatos e posterior delimitação real dos núcleos.

A metodologia proposta para a segmentação de núcleos em imagens de células cervicais é dividida em três etapas, o pré-processamento, a calibração dos parâmetros da heurística e a segmentação do núcleo, como representado na Figura 4.1. Cada uma das etapas serão detalhadas nas seções posteriores.

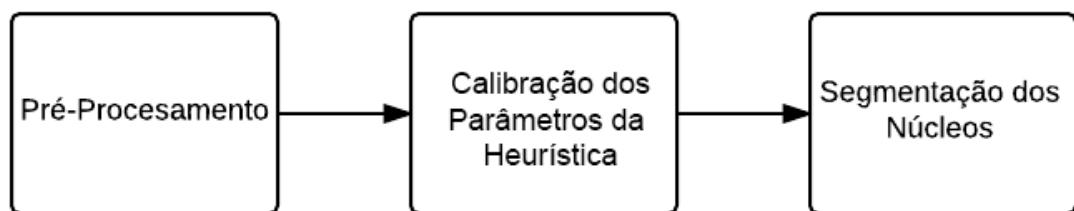


Figura 4.1: Ilustração das etapas de segmentação

4.1 Pré-Processamento

A ideia de isolar o núcleo passa por uma fase que denominamos de pré-processamento e consiste em isolar regiões que são possíveis candidatos a núcleo, nessa etapa são utilizados os algoritmos SLIC para a geração de *superpixels* e o DBSCAN para agrupar os *superpixels* similares, visando uma separação da imagem em regiões menores de características similares.

Um exemplo de imagem após a execução do pré-processamento pode ser visto na Figura 4.2. Em (a) podemos ver o resultado da divisão da imagem em *superpixels* após a execução do SLIC, em (b) podemos ver o resultado da junção dos *superpixels* feito pelo DBSCAN, e em (c) podemos ver as regiões finais, candidatas a núcleos, geradas pelo pre-processamento e que vão ser avaliadas pela Heurística CIA.

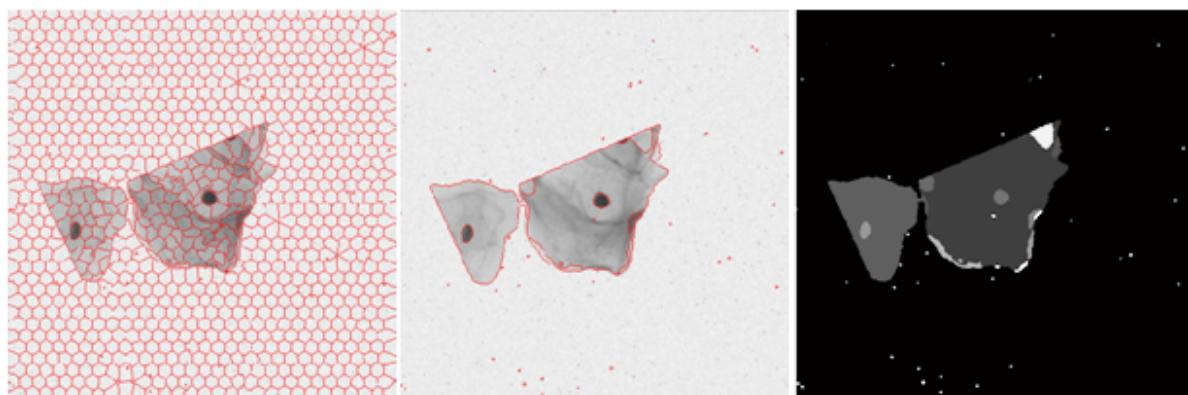


Figura 4.2: Etapa de pré-processamento:(a) ilustração do SLIC, (b) resultado após aplicação do DBSCAN e (C) superpixels gerados.

4.2 Heurística CIA

Como ilustrado na Figura 4.3, o método recebe um conjunto de candidatos a núcleo, e cada uma dessas regiões candidatas passa por uma análise composta por três etapas, onde são avaliadas circularidade, intensidade e área, de acordo com os valores encontrados na etapa de calibração.

A medida de circularidade é dada pela Equação 4.1, e o resultado dessa equação é avaliado dentro de um limite de mínimo e máximo que são definidos na calibração e passados como parâmetro da função.

$$\text{circularidade} = \frac{4\pi \text{Area}}{\text{Perímetro}^2} \quad (4.1)$$

Para realizar a avaliação da intensidade, é calculada a média das intensidades dos *pixels* pertencentes a região, sendo que o valor deve ser menor que o limite máximo passado como parâmetro para o método.

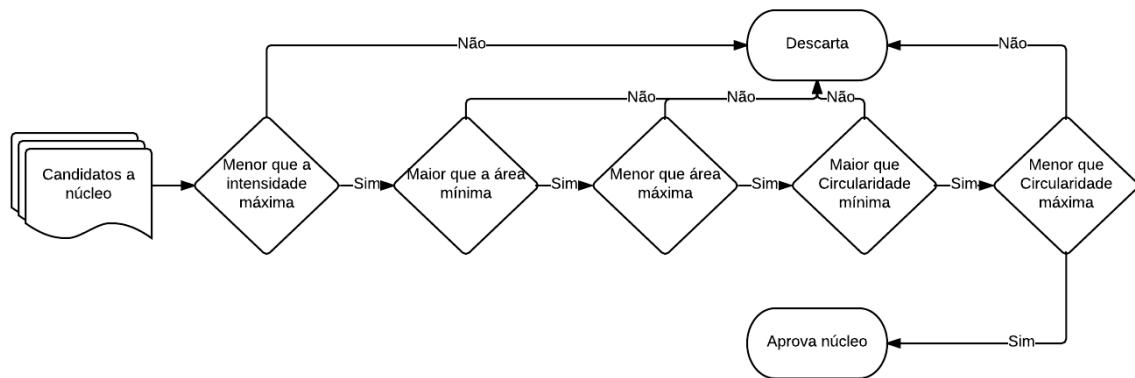


Figura 4.3: Esquema de funcionamento heurística CIA

Por fim é avaliada a área da região que deve estar contida dentro de um limite mínimo e máximo também estabelecidos como parâmetros da heurística.

Caso a região avaliada passe por todos esses testes ela é classificada como núcleo, caso contrário ela é descartada.

Na Figura 4.4(a) é apresentado um exemplo de candidatos a núcleo, onde dois deles foram aprovados em verde, e dois foram descartados, em vermelho, ambos na Figura 4.4(b). O Capítulo 5 apresenta os resultados quantitativos de detecção do núcleo pela metodologia proposta.

4.3 Calibração dos Parâmetros da Heurística

Na etapa de calibragem dos parâmetros é utilizado o algoritmo proposto por (Oliveira, Moreira, Ushizima, Carneiro, Medeiros, de Araujo, e Silva & Bianchi 2017), que é uma

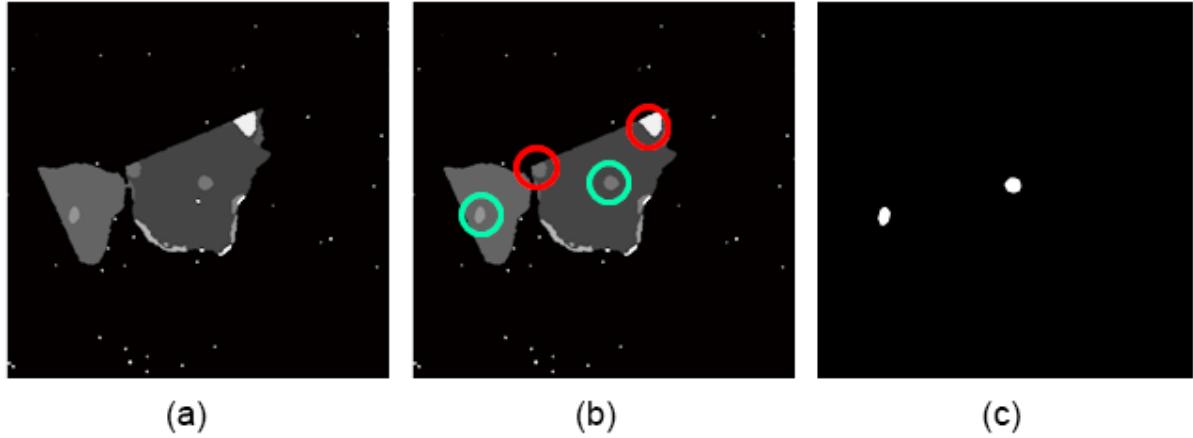


Figura 4.4: Exemplos de candidatos a núcleos e a avaliação das regiões segundo a heurística CIA.

abordagem multi-objetivo que visa otimizar os parâmetros de detecção dos núcleos utilizados na heurística de segmentação dos núcleos e que foi apresentada na Seção 4.2.

O algoritmo tem como **objetivos a maximização dos verdadeiros positivos(TP)** e **minimização dos falsos positivos(FP)**.

Na Figura 4.5 podemos observar melhor o funcionamento do algoritmo. No seu primeiro passo é criada a população inicial contendo n indivíduos, onde cada indivíduo possui 5 genes, sendo que cada gene recebe um valor inicial aleatório dentro de um intervalo permitido. Os intervalos foram estabelecidos baseando-se em dados de treinamento do núcleo. Cada indivíduo possui os seguintes 5 genes:

- Intensidade máxima, um valor inteiro no intervalo [10,200];
- Área mínima, um valor inteiro no intervalo [10,100];
- Área máxima, um valor inteiro no intervalo [200,1500];
- Circularidade mínima, um valor real no intervalo [0.50,1.00];
- Circularidade máxima, um valor real no intervalo [1.00,1.50].

Depois da geração da população inicial é feita sua avaliação, onde cada indivíduo é utilizado como parâmetro de entrada para a segmentação dos núcleos através da heurística. Os núcleos segmentados são analisados, retornando as taxas de TP e FP.

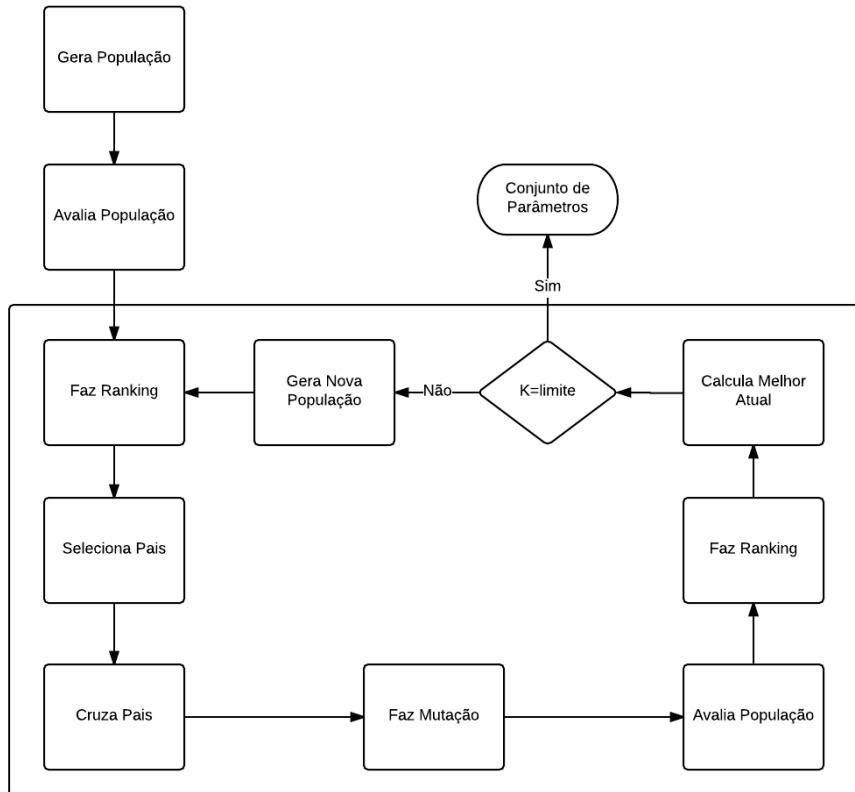


Figura 4.5: Esquema de execução do algoritmo multi-objetivo

Através dos valores de TP e FP, é feito um ranking dos indivíduos, os elementos que estão na Fronteira Pareto recebem valor 1, depois são retirados e calcula-se uma nova fronteira que recebe valor 2 e assim sucessivamente até todos os elementos serem ranqueados.

São selecionados então n pais, escolhendo aleatoriamente dois candidatos, comparando-os através de sua posição no ranking. O cruzamento dos pais é realizado gene a gene, dada uma taxa de cruzamento, o gene escolhido pode ser tanto do pai 1 ou pai 2.

Baseado em uma taxa de mutação, também gene a gene, é realizada o próximo passo, sendo que para cada gene é somado o valor $(\pm \delta)$, onde δ é definido de forma aleatória e pode assumir os seguintes valores:

- Intensidade máxima, $\delta \in [0, 5]$;
- Área mínima, $\delta \in [0, 5]$;

- Área máxima, $\delta \in [0, 5]$;
- Circularidade mínima, $\delta \in [0.0, 0.1]$;
- Circularidade máxima, $\delta \in [0.0, 0.1]$.

Após a mutação a nova população contendo os pais e filhos são reavaliados e é gerado um novo ranking. Recalculado o ranking, os elementos que aparecem na posição 1 são os elementos ditos incomparáveis, e estão presentes na Fronteira Pareto daquela geração. O processo é repetido até que seja atingido o limite de gerações dado como entrada do método.

O resultado do algoritmo é um grupo de indivíduos que contém ranking 1 ao final das gerações.

4.4 Segmentação dos Núcleos

Na etapa de segmentação dos núcleos, os *superpixels* resultantes da etapa de pré-processamento são avaliados afim de classificar ou não o *superpixel* como núcleo. Para realizar essa avaliação é utilizada a Heurística CIA e apresentada por (Oliveira, Moreira, Ushizima, Carneiro, Medeiros, de Araujo, e Silva & Bianchi 2017), onde cada uma dessas características possui valores limites que são passados como parâmetros para a função.

Cada conjunto de imagens possui características próprias e necessita de uma calibração ideal para esses parâmetros, e é utilizado o algoritmo genético multiobjetivo apresentado na Seção 4.3 para realizar esse ajuste. Um resumo do processo é apresentado na Figura 4.6, onde podemos ver que os candidatos a núcleo, gerados na primeira etapa, passam por uma avaliação realizada pela Heurística CIA, que utiliza os parâmetros calibrados na segunda etapa. Os candidatos aprovados nessa avaliação da heurística são rotulados como núcleos e os não aprovados são descartados.

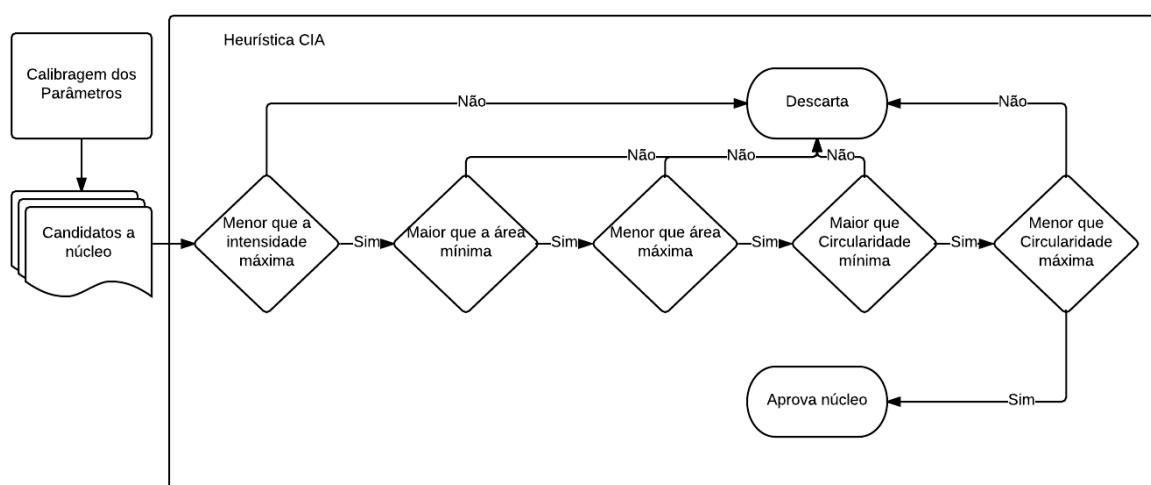


Figura 4.6: Esquema de execução do método de segmentação do núcleo

Capítulo 5

Resultados

5.1 CRIC Searchable Database

Devido a dificuldade de obtenção de informações sobre imagens reais de Papanicolau para este trabalho possibilitou a criação de uma ferramenta web capaz de armazenar, realizar segmentações e classificações em imagens reais de exames de Papanicolau.

A nova base de imagens foi cedida pelo Núcleo de Pesquisas Biológicas da Universidade Federal de Ouro Preto. As imagens foram selecionadas por especialistas, que ainda realizaram na primeira versão a marcação de todos os núcleos das imagens. Exemplos de imagens da base são apresentados na Figura 5.1, a base foi denominada *CRIC Searchable Database*.

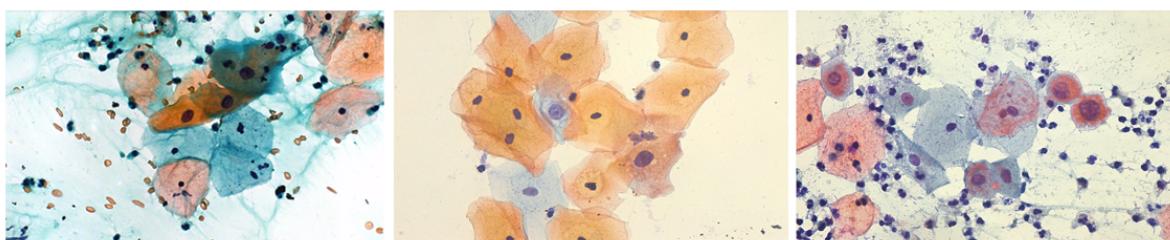


Figura 5.1: Exemplo de imagens da base de dados CRIC Database

A primeira versão da base contém 400 imagens reais. Essas imagens contém 11.480 núcleos que foram marcados e classificados em 6 classes distintas: 6792 normais, 627 ASC-US, 1336 LSIL, 915 ASC-H, 1687 HSIL e 123 Carcinoma. Todos os núcleos tiveram seus centros marcados e classificados por dois diferentes especialistas e conferido por mais

um.

Essas informações preliminares serão úteis para identificação de núcleos e classificação das células.

A ferramenta utilizada foi desenvolvida para ser um ambiente colaborativo onde os especialistas pudessem cadastrar imagens e realizar a segmentação e classificação das mesmas. Ela se encontra hospedada no endereço www.cricdatabase.com.br. O acesso a base ainda é restrito pois estamos realizando uma publicação, no futuro espera-se divulgar as informações da base de imagens. Uma tela da ferramenta que mostra as imagens listadas para classificação pode ser vista na Figura 5.2.

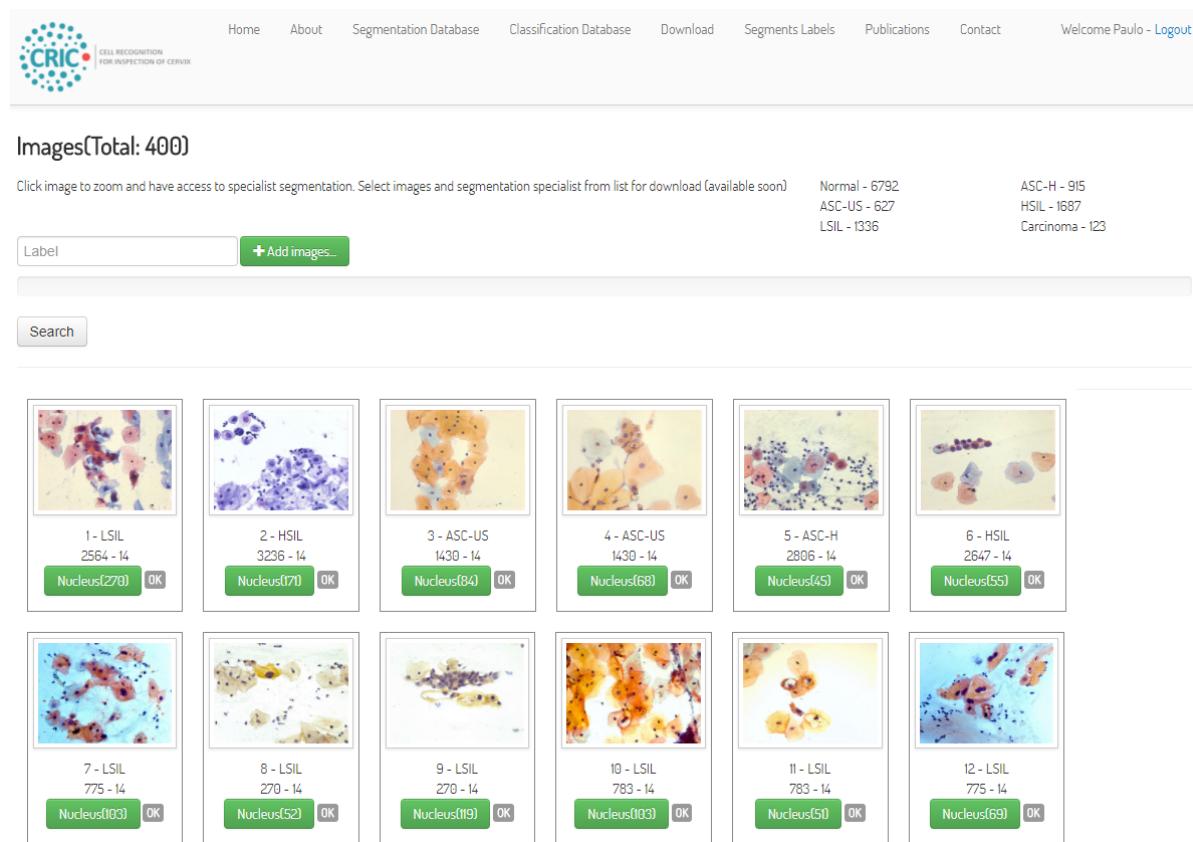


Figura 5.2: Tela da ferramenta colaborativa CRIC Database

Com a ferramenta o especialista pode tanto segmentar manualmente as células e seus núcleos como visto na Figura 5.3 (a), quanto localizar os núcleos e classificá-los como visto na Figura 5.3 (b).

A ferramenta ainda encontra-se em evolução e a ideia é que os algoritmos de segmentação e classificação de núcleos e citoplasmas sejam adicionados como ferramentas

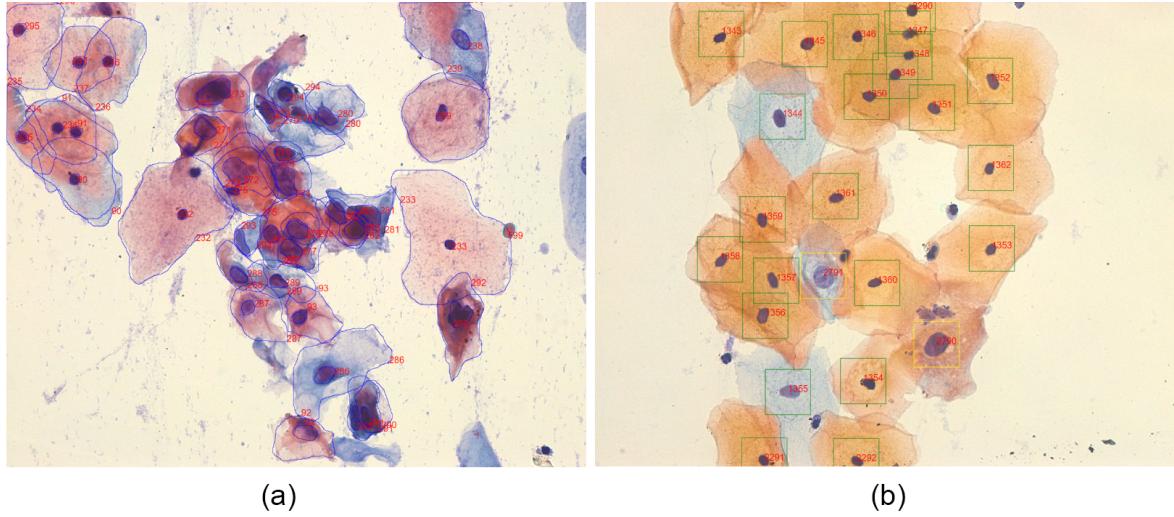


Figura 5.3: (a) Segmentação manual de células, (b) identificação dos núcleos e classificação das células

da aplicação.

Como a segmentação manual das células é extremamente custosa e vem sendo feita de forma cautelosa pelos especialistas, essa primeira versão a ser disponibilizada em breve contém apenas marcações e classificações dos núcleos.

5.2 Experiments

Para a execução dos experimentos foi utilizado um servidor modelo Dell Power Edge R720 com 10 processadores Intel Xeon E5-2660 2.20GHz x40 e 284GB de memória RAM.

Para a execução dos testes foram utilizadas as base de dados fornecidas pelo ISBI 2014, sendo que, o conjunto escolhido é composto da base de treino contendo 45 imagens sintéticas(*isbi_train_45*) e da base de testes contendo 90 imagens(*isbi_test_90*) e uma base de dados contendo 22 imagens reais. Para a base *isbi_train_45*, 40 foram utilizadas como treinamento para a fase de calibragem dos parâmetros e as outras 5 foram utilizadas para os testes. De forma similar, para a base *isbi_test_90* foram utilizadas 81 imagens para treino e 9 para testes. Para as imagens reais foram utilizadas 20 para treino e 2 para testes.

Na fase de pré-processamento o SLIC utilizou como parâmetros $k=1000$ que é o

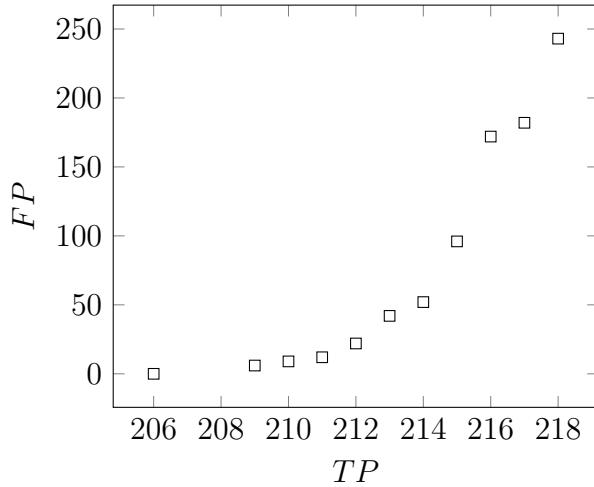


Figura 5.4: Fronteira Pareto para base *isbi_train_45*

número desejado de *superpixels*, $m=5$ que é um fator de ponderação entre as distâncias espaciais e de cores, *seRadius*=1 que é o parâmetro que regula se os pixels serão agrupados, e *colopt*='median' que define qual o método para indicar o centro de cor do *cluster*. O DBSCAN utilizou como parâmetros *l* que é a imagem rotulada com os *superpixels*, *Cp* que contém os atributos dos *superpixels* e *Am* que é a matriz de adjacências, todos retornados do SLIC, e *E*=5 que é a tolerância para a junção dos *superpixels*.

A calibragem de parâmetros utilizou como entrada para o algoritmo multi-objetivo, geração de tamanho 50, população de tamanho 20, probabilidade de cruzamento de 50%, e probabilidade de mutação de 30%. Essa combinação foi executada 30 vezes, os valores de TP e FP dos indivíduos encontrados na fronteira Pareto, ao final da execução, para a base *isbi_train_45* são apresentados na Figura 5.4.

Nas 40 imagens em que a calibragem foi executada existiam 236 núcleos sendo que para o indivíduo 1 da Tabela 5.1, foi encontrado um TP=206, FP = 0 e FN = 30. Como o objetivo posterior a detecção dos núcleos é a classificação das células como normais ou alteradas, o conjunto de parâmetros encontrados no indivíduo 1 é utilizado pois tem maiores chances de não gerar falsos positivos na etapa de classificação, pois apresentou FP = 0.

Após a escolha do conjunto de parâmetros foi executada a Heurística CIA nas 5 imagens restantes, sendo que as imagens possuíam um total de 34 núcleos, onde foi obtido um TP=29, FP = 0 e FN = 5. A etapa de segmentação do núcleo pode ser comparada, como visto na Tabela 5.2, com o trabalho apresentado por (Ushizima, Bianchi & Carneiro

Tabela 5.1: Elementos da Fronteira Pareto para base *isbi_train_45*

i	Intensidade	Área Mínima	Área Máxima	Circularidade Mínima	Circularidade Máxima	TP	FN	FP
1	140	95	919	0,67	1,31	206	30	0
2	136	92	920	0,6073	1,26	209	27	6
3	153	63	626	0,55	1,49	215	21	96
4	140	95	1359	0,6073	1,49	211	25	12
5	140	95	626	0,5422	1,49	212	24	22
6	153	95	626	0,55	1,49	214	22	52
7	153	95	626	0,6073	1,3	213	23	42
8	136	95	626	0,5422	1,49	210	26	9
9	153	98	625	0,4651	1,25	217	19	182
10	153	92	626	0,476	1,26	216	20	172
11	153	63	626	0,4651	1,3	218	18	243

2014) onde foi obtido TP=23, FP = 0 e FN = 11. Foram realizadas ainda outras medidas em que foi encontrado um valor de revocação de 85,29%, precisão de 100% e Valor-F de 92,06%, onde podemos observar uma melhoria de 17,65% na revocação, 11,36% no Valor-F, mantendo a precisão em 100% e mantendo ainda a taxa de FP em 0.

Tabela 5.2: Resultados da Segmentação dos Núcleos para a base *isbi_train_45*

Método	Daniela et al.(2014)	Método Proposto
TP	23	29
FP	0	0
FN	11	5
% Revocação	67,65	85,29
% Precisão	100,00	100,00
% Valor-F	80,70	92,06

Para a base *isbi_test_90*, que contém 80 imagens de treino, a calibragem foi executada existiam 472 núcleos sendo que para o indivíduo 1 da Tabela 5.1, foi encontrado um TP=433, FP = 0 e FN = 33.

Assim como feito na base *isbi_train_45*, após a escolha do conjunto de parâmetros foi

executada a Heurística CIA nas 10 imagens restantes, sendo que as imagens possuíam um total de 68 núcleos, onde foi obtido um TP=66, FP = 0 e FN = 2. A etapa de segmentação do núcleo pode ser comparada, como visto na Tabela 5.4, com o trabalho apresentado por (Ushizima, Bianchi & Carneiro 2014) onde foi obtido TP=40, FP = 0 e FN = 28. Foram realizadas ainda outras medidas em que foi encontrado um valor de revocação de 97,06%, precisão de 100% e Valor-F de 98,51%, onde podemos observar uma melhoria de 38,24% na revocação, 24,44% no Valor-F, mantendo a precisão em 100% e mantendo ainda a taxa de FP em 0.

Tabela 5.3: Elementos da Fronteira Pareto para a base *isbi_test_90*

i	Intensidade	Área Mínima	Área Máxima	Circularidade Mínima	Circularidade Máxima	TP	FN	FP
1	139	45	914	0,3756	1,22	433	39	0
2	142	57	914	0,3756	1,4	439	33	7
3	142	46	906	0,3544	1,22	440	32	21
4	144	15	913	0,3648	1,399	442	30	74

Tabela 5.4: Resultados da Segmentação dos Núcleos para a base *isbi_test_90*

Método	Daniela et al.(2014)	Método Proposto
TP	40	66
FP	0	0
FN	28	2
% Revocação	58,82	97,06
% Precisão	100,00	100,00
% Valor-F	74,07	98,51

Um exemplo do resultado final da segmentação dos núcleos em células com e sem sobreposição derivadas da base do ISBI 2014 pode ser visto na Figura 5.5. Como pode ser visto a partir dos resultados, o método é capaz de detectar com precisão núcleos de células cervicais. No entanto, como o número de imagens que o método foi aplicado é limitado, consideramos esses resultados como promissores, embora preliminares, e temos a intenção de desenvolver um maior número de imagens de núcleos, a fim de estimar um

melhor comportamento do nosso método. Outra desvantagem foi relacionada ao tempo de processamento para a abordagem de calibração multi-objetivo, cerca de 13 horas para cada uma das 30 execuções.

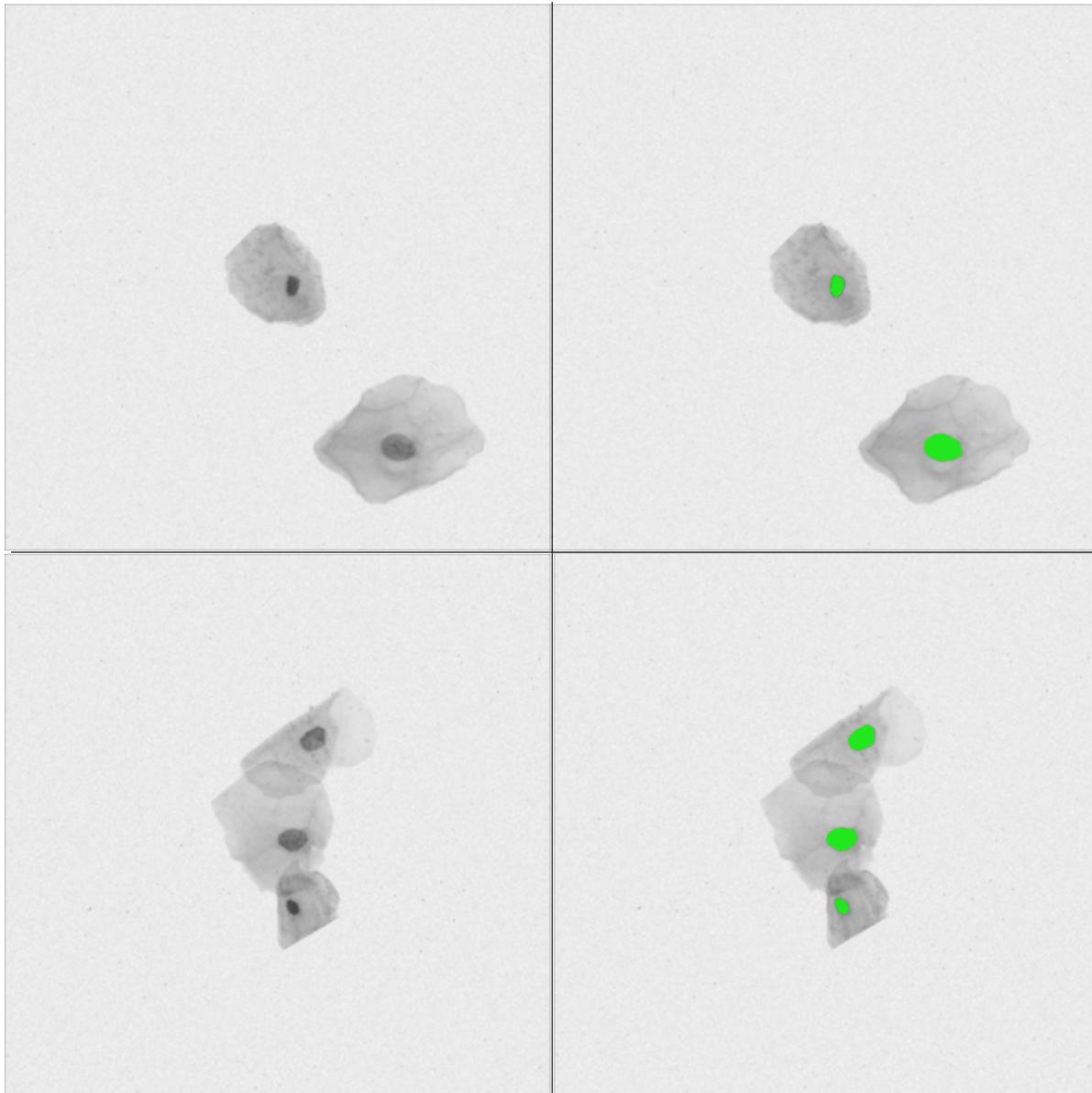


Figura 5.5: Exemplo de segmentação de núcleos com e sem sobreposição celular

Para a base de imagens reais contendo 22 imagens o período de treinamento para uma execução demorou aproximadamente 24 horas para 30 gerações. Os parâmetros foram os mesmos utilizados na base do ISBI 2014. O elementos da fronteira Pareto ao final da execução podem ser vistos na Tabela 5.5. Como pode ser visualizado os valores encontrados para FP são muito elevados, mas seguindo a mesma metodologia,

foi escolhido o elemento com menor valor de FP da linha 1.

Tabela 5.5: Elementos da Fronteira Pareto para base de imagens reais

i	Intensidade	Área Mínima	Área Máxima	Circularidade Mínima	Circularidade Máxima	TP	FN	FP
1	168	253	1451	0,4173	1,28	544	287	1286
2	168	216	1451	0,4441	1,2800	549	282	1356
3	184	253	1341	0,4173	1,1455	551	280	1370
4	184	253	1455	0,4173	1,2800	563	268	1393
5	182	217	1450	0,4441	1,1000	566	265	1449
6	185	217	1455	0,4441	1,1000	569	262	1476
7	187	216	1451	0,4441	1,2800	571	260	1503
8	187	214	1451	0,4441	1,2287	573	258	1517
9	189	216	1451	0,4441	1,2800	574	257	1524
10	189	212	1346	0,4441	1,1503	575	256	1533
11	189	214	1451	0,4441	1,2287	576	255	1538
12	189	212	1451	0,4441	1,2800	580	251	1550
13	184	216	1455	0,4173	1,2800	601	230	1627

Assim como feito nas bases anteriores, após a escolha do conjunto de parâmetros foi executada a Heurística CIA nas 2 imagens restantes, sendo que as imagens possuíam um total de 39 núcleos, onde foi obtido um TP=29, FP = 114 e FN = 10. Foram realizadas ainda outras medidas em que foi encontrado um valor de revocação de 74,36%, precisão de 20,28% e Valor-F de 31,87%. Podemos ver com os resultados que a metodologia não se comportou tão bem em imagens reais. As duas imagens usadas para testes da segmentação podem ser visualizadas nas Figuras 5.6 e 5.7, onde as marcações em verde representam os núcleos encontrados e em vermelho os candidatos classificados de forma equivocada.

É importante ressaltar que os valores escolhidos para os parâmetros foram baseados nos resultados obtidos nas células simuladas, notamos que as imagens reais tem uma maior variabilidade dos núcleos, entre normais e lesionados, e que, um maior número de gerações poderia gerar resultados mais satisfatórios.

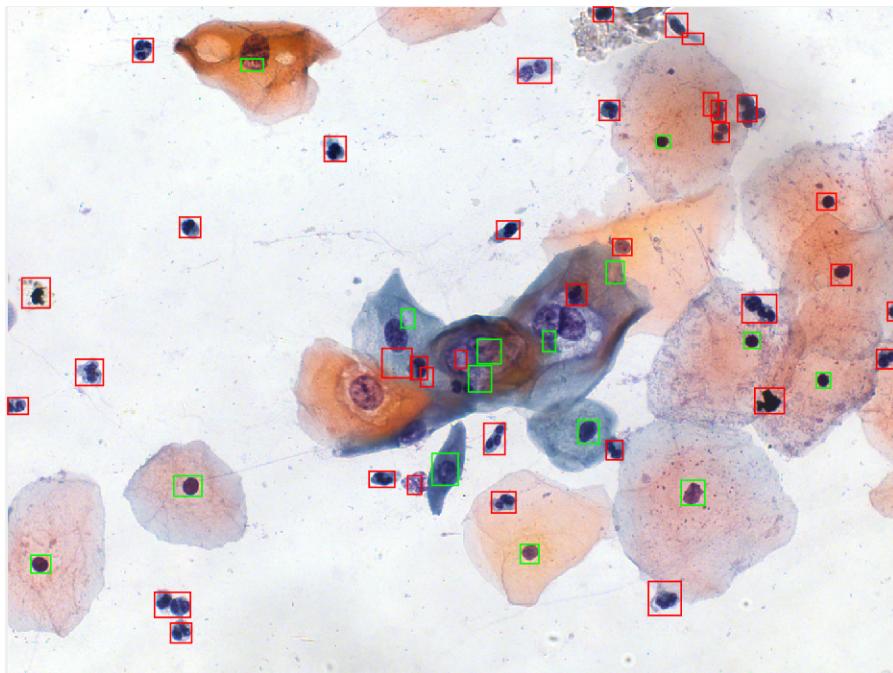


Figura 5.6: Exemplo de segmentação de núcleos com e sem sobreposição celular em imagem real

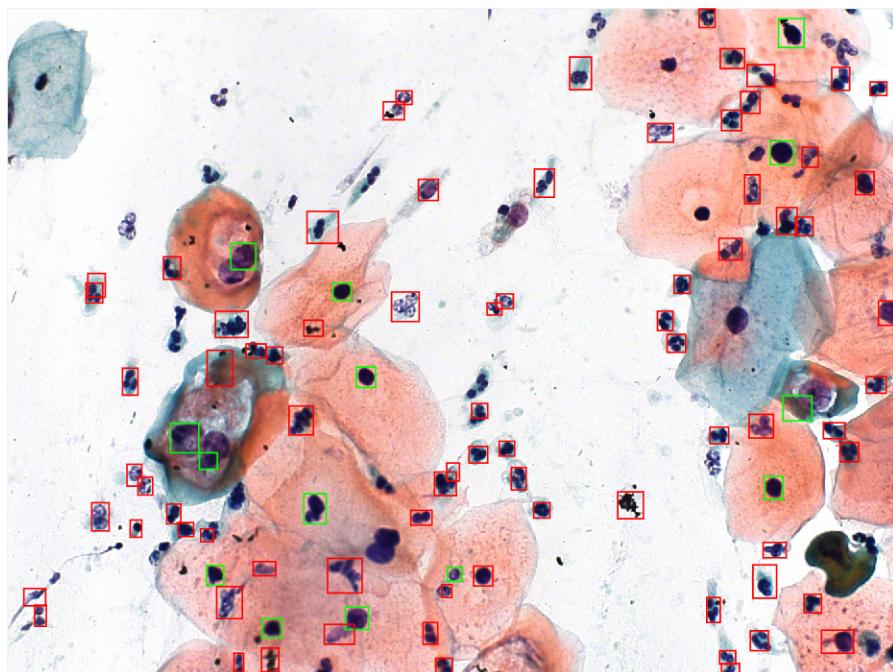


Figura 5.7: Exemplo de segmentação de núcleos com e sem sobreposição celular em imagem real

Capítulo 6

Conclusão

Esse trabalho caracteriza-se como uma importante contribuição aos trabalhos que vêm sendo desenvolvidos na área de identificação e classificações de lesões em células cervicais nos últimos anos. Neste sentido essa dissertação contribui apresentando tanto técnicas de segmentação de núcleos celulares, quanto uma nova base de dados segmentada e classificada manualmente por especialistas.

Foi proposta uma nova abordagem para o uso de um algoritmo genético multi-objetivo associado a uma heurística de segmentação. A metodologia foi testada utilizando uma base de imagens de células sintéticas do desafio ISBI e em algumas imagens reais.

Os resultados obtidos através dos experimentos apresentados no Capítulo 5 demonstram uma grande melhoria na taxa de TP em relação ao estado da arte(Lu, Carneiro, Bradley, Ushizima, Nosrati, Bianchi, Carneiro & Hamarneh 2017b), mantendo a taxa de FP em 0. Os experimentos mostraram que o uso da abordagem multi-objetivo pode melhorar o resultado de algoritmos onde é necessário o ajuste de parâmetros. A técnica de segmentação permitiu qualidade e precisão nas identificações dos núcleos, e consequentemente às classificações das lesões, feitas futuramente.

Como pontos a serem melhorados, podemos citar o tempo gasto na fase de pré-processamento e o tempo gasto na fase de calibragem dos parâmetros. Maiores investigações devem ser realizadas nesse sentido, o uso de programação paralela não foi abordado e pode ser uma alternativa.

Outra importante contribuição do trabalho foi a ferramenta criada para a geração

da base de imagens reais, que possibilitou a disponibilização de uma base de imagens contendo a princípio 22 imagens rotuladas, que foram utilizadas nesse trabalho. E em breve a primeira versão oficial contendo 400 imagens rotuladas e conferidas por especialistas. É importante ressaltar que não existe na literatura nenhuma base como essa. A grande vantagem dos resultados da base de dados será permitir a extração de conhecimento e informações em uma base real de imagens, contrário ao que se tem até o momento.

A aplicação do método nessa base de imagens reais deve ser melhor estudado também, visto que, as características da imagem são bem diferentes das imagens sintéticas, e deve ser investigado se um melhor ajuste dos parâmetros do método pode trazer melhores resultados.

6.1 Desenvolvimentos Futuros

Os resultados apresentados nesta dissertação abrem vários caminhos para estudos mais abrangentes e completos em células cervicais e lesões de câncer. Dentre os trabalhos a serem desenvolvidos na continuação deste projeto incluímos: a investigação de outros parâmetros que possam vir a ser relevantes, e possam ser incluídos na heurística; o uso de processamento paralelo para melhorar o tempo de execução do algoritmo, bem como analisar outros algoritmos de geração de superpixels e melhor avaliar os parâmetros utilizados no algoritmo, para ver seu comportamento na base de imagens reais.

A ferramenta utilizada para a geração da base de dados também pode evoluir para uma plataforma onde os algoritmos propostos seja implementados e se tornem uma funcionalidade da mesma, de forma que o usuário possa submeter imagens e ter o resultado da segmentação e classificação da imagem.

Referências Bibliográficas

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**(11): 2274–2282.
- Aggarwal, R. & Bacus, J. (1977). A multi-spectral approach for scene analysis of cervical cytology smears., *Journal of Histochemistry & Cytochemistry* **25**(7): 668–680.
- Bamford, P. & Lovell, B. (1998). Unsupervised cell nucleus segmentation with active contours, *Signal Processing* **71**(2): 203–213.
- Bengtsson, E. & Malm, P. (2014). Screening for cervical cancer using automated analysis of pap-smears, *Computational and mathematical methods in medicine* **2014**.
- Bergmeir, C., Silvente, M. G. & Benítez, J. M. (2012). Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework, *Computer methods and programs in biomedicine* **107**(3): 497–512.
- Bernstein, S. J., Sanchez-Ramos, L. & Ndubisi, B. (2001). Liquid-based cervical cytologic smear study and conventional papanicolaou smears: a metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy, *American journal of obstetrics and gynecology* **185**(2): 308–317.
- Breen, E. J. & Jones, R. (1996). Attribute openings, thinnings, and granulometries, *Computer Vision and Image Understanding* **64**(3): 377–389.
- Cahn, R., Poulsen, R. & Toussaint, G. (1977). Segmentation of cervical cell images., *Journal of Histochemistry & Cytochemistry* **25**(7): 681–688.
- Comaniciu, D. & Meer, P. (1999). Mean shift analysis and applications, *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Vol. 2, IEEE, pp. 1197–1203.

- Deb, K., Agrawal, S., Pratap, A. & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii, *Parallel Problem Solving from Nature PPSN VI*, Springer, pp. 849–858.
- Duanggate, C., Uyyanonvara, B. & Koanantakul, T. (2008). A review of image analysis and pattern classification techniques for automatic pap smear screening process, *International Conference on Embedded Systems and Intelligent Technology, Thailand*, pp. 212–217.
- Erhardt, R., Reinhardt, E., Schlipf, W. & Bloss, W. (1979). Fazytan: a system for fast automated cell segmentation, cell image analysis and feature extraction based on tv-image pickup and parallel processing., *Analytical and quantitative cytology* **2**(1): 25–40.
- Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise., **96**(34): 226–231.
- Gençtav, A., Aksoy, S. & Önder, S. (2012). Unsupervised segmentation and classification of cervical cell images, *Pattern Recognition* **45**(12): 4151–4168.
- Jantzen, J., Norup, J., Dounias, G. & Bjerregaard, B. (2005). Pap-smear benchmark data for pattern classification technical University of Denmark, *Nature inspired Smart Information Systems* pp. 1–9.
- Kale, A. & Aksoy, S. (2010). Segmentation of cervical cell images, *Pattern Recognition (ICPR), 2010 20th International Conference on*, IEEE, pp. 2399–2402.
- Lassouaoui, N. & Hamami, L. (2003). Genetic algorithms and multifractal segmentation of cervical cell images, *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, Vol. 2, IEEE, pp. 1–4.
- Li, Z. & Najarian, K. (2001). Automated classification of pap smear tests using neural networks, *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, Vol. 4, IEEE, pp. 2899–2901.
- Lorenzo-Ginori, J. V., Curbelo-Jardines, W., López-Cabrera, J. D. & Huergo-Suárez, S. B. (2013). Cervical cell classification using features related to morphometry and texture of nuclei, *Iberoamerican Congress on Pattern Recognition*, Springer, pp. 222–229.

- Lu, Z., Carneiro, G., Bradley, A. P., Ushizima, D., Nosrati, M. S., Bianchi, A. G., Carneiro, C. M. & Hamarneh, G. (2017a). Evaluation of three algorithms for the segmentation of overlapping cervical cells, *IEEE journal of biomedical and health informatics* **21**(2): 441–450.
- Lu, Z., Carneiro, G., Bradley, A. P., Ushizima, D., Nosrati, M. S., Bianchi, A. G., Carneiro, C. M. & Hamarneh, G. (2017b). Evaluation of three algorithms for the segmentation of overlapping cervical cells, *IEEE journal of biomedical and health informatics* **21**(2): 441–450.
- Mango, L. J. (1994). Computer-assisted cervical cancer screening using neural networks, *Cancer Letters* **77**(2): 155–162.
- Matas, J., Chum, O., Urban, M. & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions, *Image and vision computing* **22**(10): 761–767.
- Nordin, B. (1989). The development of an automated prescreener for the early detection of cervical cancer: Algorithms and implementation.
- Nosrati, M. & Hamarneh, G. (2014). A variational approach for overlapping cell segmentation, *ISBI Overlapping Cervical Cytology Image Segmentation Challenge* pp. 1–2.
- Nosrati, M. S. & Hamarneh, G. (2015). Segmentation of overlapping cervical cells: a variational method with star-shape prior, *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, IEEE, pp. 186–189.
- Oliveira, P. H., Moreira, G., Ushizima, D. M., Carneiro, C. M., Medeiros, F. N., de Araujo, F. H., e Silva, R. R. & Bianchi, A. G. (2017). A multi-objective approach for calibration and detection of cervical cells nuclei, *Evolutionary Computation (CEC), 2017 IEEE Congress on*, IEEE, pp. 2321–2327.
- Olsen, O. F. (1997). Multi-scale watershed segmentation, *Gaussian Scale-Space Theory*, Springer, pp. 191–200.
- Otsu, N. (1975). A threshold selection method from gray-level histograms, *Automatica* **11**(285-296): 23–27.
- Phansalkar, N., More, S., Sabale, A. & Joshi, M. (2011). Adaptive local thresholding for detection of nuclei in diversity stained cytology images, *Communications and Signal Processing (ICCSP), 2011 International Conference on*, IEEE, pp. 218–220.

- Plissiti, M., Charchanti, A., Krikoni, O. & Fotiadis, D. (2006). Automated segmentation of cell nuclei in pap smear images, *Proc. IEEE International Special Topic Conference on Information Technology in Biomedicine, Greece*, Citeseer.
- Plissiti, M. E. & Nikou, C. (2012a). Cervical cell classification based exclusively on nucleus features, *International Conference Image Analysis and Recognition*, Springer, pp. 483–490.
- Plissiti, M. E. & Nikou, C. (2012b). On the importance of nucleus features in the classification of cervical cells in pap smear images, *University of Ioannina*.
- Plissiti, M. E., Nikou, C. & Charchanti, A. (2011). Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering, *IEEE Transactions on information technology in biomedicine* **15**(2): 233–241.
- Ploem, J., van Driel-Kulker, A. & Verwoerd, N. (1987). Leytas-a cytology screening system using the new modular image analysis computer (miac) from leitz, *Clinical cytometry and histometry* pp. 24–35.
- Prewitt, J. & Mendelsohn, M. L. (1966). The analysis of cell images*, *Annals of the New York Academy of Sciences* **128**(3): 1035–1053.
- Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'connor, D., Prey, M., Raab, S., Sherman, M., Wilbur, D., Wright Jr, T. et al. (2002). The 2001 bethesda system: terminology for reporting results of cervical cytology, *Jama* **287**(16): 2114–2119.
- Song, Y., Zhang, L., Chen, S., Ni, D., Li, B., Zhou, Y., Lei, B. & Wang, T. (2014). A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei, *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, IEEE, pp. 2903–2906.
- Spencer, C. C. & Bostrom, R. C. (1962). Performance of the cytoanalyzer in recent clinical trials, *Journal of the National Cancer Institute* **29**(2): 267–276.
- Spriggs, A., Diamond, R. & Meyer, E. (1968). Automated screening for cervical smears?, *The Lancet* **291**(7538): 359–360.
- Stevens, M. W., Milne, A. J., James, K. A., Brancheau, D., Ellison, D. & Kuan, L. (1997). Effectiveness of automated cervical cytology rescreening using the autopap® 300 qc system, *Diagnostic cytopathology* **16**(6): 505–512.

- Tanaka, N., Ueno, T., Ikeda, H., Ishikawa, A., Yamauchi, K., Okamoto, Y. & Hosoi, S. (1987). Cybest model 4. automated cytologic screening system for uterine cancer utilizing image analysis processing., *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **9**(5): 449–454.
- Tolles, W. E. & Bostrom, R. (1956). Automatic screening of cytological smears for cancer: the instrumentation, *Annals of the New York Academy of Sciences* **63**(6): 1211–1218.
- Tucker, J. & Husain, O. (1981). Trials with the cerviscan experimental prescreening device on polylysine-prepared slides., *Analytical and quantitative cytology* **3**(2): 117–120.
- Ushizima, D., Bianchi, A. & Carneiro, C. (2014). Segmentation of subcellular compartments combining superpixel representation with voronoi diagrams, *ISBI Overlapping Cervical Cytology Image Segmentation Challenge* pp. 1–2.
- Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms, *Image Processing, IEEE Transactions on* **2**(2): 176–201.
- Watanabe, S., Group, C. et al. (1974). An automated apparatus for cancer prescreening: Cybest, *Computer Graphics and Image Processing* **3**(4): 350–358.
- Zahniser, D., Oud, P., Raaijmakers, M., Vooys, G. & Van de Walle, R. (1979). Biopepr: a system for the automatic prescreening of cervical smears., *Journal of Histochemistry & Cytochemistry* **27**(1): 635–641.