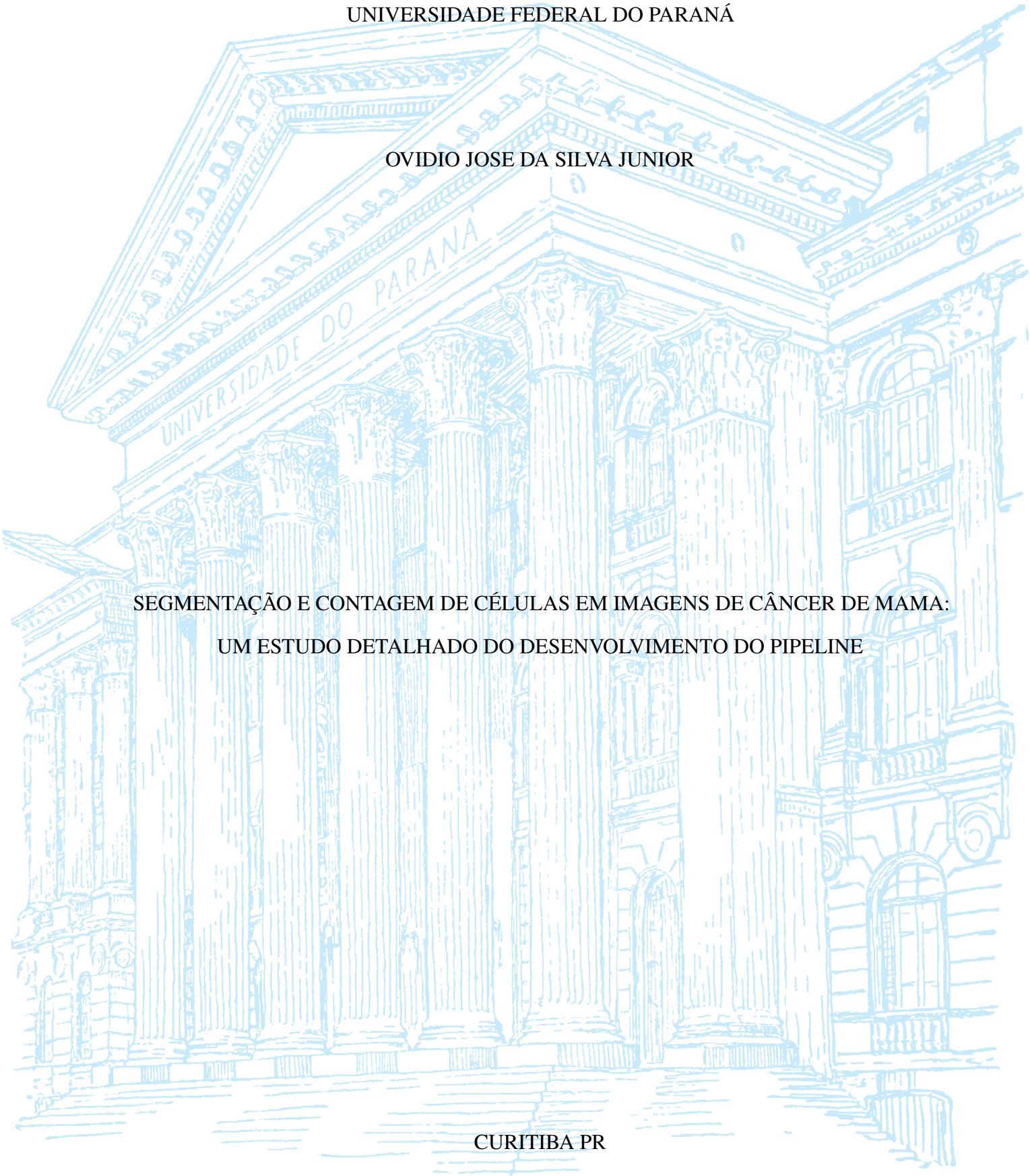
UNIVERSIDADE FEDERAL DO PARANÁ

OVIDIO JOSE DA SILVA JUNIOR



SEGMENTAÇÃO E CONTAGEM DE CÉLULAS EM IMAGENS DE CÂNCER DE MAMA: UM ESTUDO DETALHADO DO DESENVOLVIMENTO DO PIPELINE

CURITIBA PR

2018

OVIDIO JOSE DA SILVA JUNIOR

SEGMENTAÇÃO E CONTAGEM DE CÉLULAS EM IMAGENS DE CÂNCER DE MAMA: UM ESTUDO DETALHADO DO DESENVOLVIMENTO DO PIPELINE

Trabalho apresentado ao Departamento de Informática, pertencente ao Setor de Exatas da Universidade Federal do Paraná, como requisito parcial para a conclusão do curso Bacharelado em Ciência da Computação.

Área de concentração: *Ciência da Computação*.

Orientador: Lucas Ferrari de Oliveira.

CURITIBA PR

2018

Universidade Federal do Paraná
Setor de Ciências Exatas
Curso de Ciência da Computação

Ata de Apresentação de Trabalho de Conclusão de Curso 2

Título do Trabalho: SEGMENTAÇÃO E CONTAGEM DE CÉLULAS EM IMAGENS DE CÂNCER DE MAMA: UM ESTUDO DETALHADO DO DESENVOLVIMENTO DO PIPELINE

Autor(es):

GRR 20182667 Nome: Ovidio José da Silva Júnior

GRR _____ Nome: _____

Apresentação: Data: 12/12/23 Hora: 14h Local: Auditório Alexandre Direne

Orientador: Lucas Ferrari de Oliveira

Membro 1: Sérgio Ossamu Ioshii

Membro 2: Glenda Proença Train

(nome) (assinatura)

AVALIAÇÃO – Produto escrito		ORIENTADOR	MEMBRO 1	MEMBRO 2	MÉDIA
Conteúdo	(00-40)				
Referência Bibliográfica	(00-10)				
Formato	(00-05)				
AVALIAÇÃO – Apresentação Oral					
Domínio do Assunto	(00-15)				
Desenvolvimento do Assunto	(00-05)				
Técnica de Apresentação	(00-03)				
Uso do Tempo	(00-02)				
AVALIAÇÃO – Desenvolvimento					
Nota do Orientador	(00-20)		**********	**********	
NOTA FINAL		**********	**********	**********	90

Os pesos indicados são sugestões.

Conforme decisão do colegiado do curso de Ciência da Computação, a entrega dos documentos comprobatório de trabalho de Conclusão de Curso 2 deve deve respeitar os seguintes procedimentos: o orientador deve abrir um processo no Sistema Eletrônico de Informações (SEI – UFPR); Selecionar o tipo: *Graduação: Trabalho Conclusão de Curso*; informar os interessados: nome do aluno e o nome do orientador; anexar esta ata escaneada e a versão final do PDF da monografia do aluno; Tramitar o processo para CCOMP (Coordenação de Ciência da Computação).

Dedico isso a todos os que amo, aos que amei, e a mim mesmo, por não ter desistido deste sonho.

AGRADECIMENTOS

Nunca pensei que, depois de escrever páginas e mais páginas, o mais difícil seria conseguir resumir em uma só página o quanto sou grato por ter chegado onde cheguei. Agradeço, primeiramente, à minha mãe Adriana e ao meu pai Ovidio, que acreditaram em mim quando decidi cursar uma universidade pública a centenas de quilômetros de distância de casa. Agradeço também ao meu irmão Ismael, que, além de me inscrever em todos os concursos que abriam, também me ajudou nas etapas do vestibular.

Um agradecimento especial e um abraço apertado a Cristina Moreira, que ajudou a convencer que um garoto, filho de pais faxineiros, caberia em um lugar como a UFPR.

Retomando a alguns anos atrás, quando cheguei em Curitiba, quero agradecer à minha querida mãe de consideração, Gessica, que me acolheu. Eu tinha 17 anos e estava muito assustado. Obrigado por tudo; foram os momentos mais difíceis que passei na minha vida, e você esteve aqui.

Obrigado aos meus amigos de longa data, Bruna, Daniel e Giseli. Desde 2019, mantivemos contato com a maior frequência possível, algo que achei que seria impossível. Mesmo longe, vocês me visitavam e não se esqueciam de mim. Bons tempos e obrigado por serem quem vocês são.

Dedico este parágrafo para agradecer ao meu amor. Foi um prazer compartilhar momentos tão divertidos e felizes, e, mesmo nos momentos não tão felizes, ter você me ajudando nas provas e trabalhos e abdicando do seu tempo para ficar comigo me confortava. Meu companheiro de vida.

Agradeço ao meu orientador, Lucas Ferrari. Sempre te admirei muito e foi um prazer estar trabalhando com você nestes últimos meses, sua paciência e maneira de lidar com todo meu nervosismo foram essenciais para a conclusão deste trabalho

Por fim, quero agradecer a todos os amigos que fiz nessa caminhada, mesmo os que não são tão próximos. Cada um teve uma contribuição importante nesta realização.

Por fim, mas não menos importante, agradeço aos meus filhotes, Fadão e Mel. Senti muito a falta de vocês, mas tive a oportunidade de vê-los felizes em todas as visitas que fiz à minha antiga casa, e agora posso ver essa felicidade todos os dias com a finalização deste trabalho.

RESUMO

Com aumento do numero de incidências nos últimos anos, o câncer de mama se tornou um grande fomentador de pesquisa na área de visão computacional. As projeções do surgimento de novos casos da doença para os próximos anos mostram a importância e a relevância do debate no âmbito de saúde publica. Os esforços nas pesquisas tem tido como principal enfoque o diagnóstico precoce, em sua grande parte com processos automáticos e de grande poder computacional. Uma etapa importante nesses processos é a fase de segmentação que tem como grande desafio sua automatização devido à complexidade inerente das células e à variabilidade das imagens. Este trabalho propõe um pipeline semi-automático no processo de segmentação em imagens de imuno-histoquímica com intuito de reduzir o tempo e esforço dos profissionais na etapa de contagem de células sem a necessidade do uso de tecnologias de alto poder computacional tal como as redes neurais convolucionais, se mostrando ser uma alternativa viável para o diagnóstico.

Palavras-chave: Cancêr de Mama. IHQ. Processamento de Imagens. Segmentação

ABSTRACT

With an increase in the number of incidences in recent years, breast cancer has become has become a major promoter of research in the area of computer vision. The projections of the emergence of new cases of the disease in the coming years shows the importance and relevance of the debate in the field of public health. Research efforts have had as main focus on early diagnosis, for the most part using automatic and highly efficient processes. computational power. An important step in these processes is the segmentation phase, which automation is a major challenge due to the inherent complexity of cells and the image variability. This work proposes a semi-automatic pipeline in the process of segmentation in immunohistochemistry images in order to reduce the time and effort of professionals in the cell counting stage without the need for the use of high-quality technologies computational power such as convolutional neural networks, proving to be an alternative viable for diagnosis.

Keywords: Breast Cancer. IHQ. Image Processing. Segmentation

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ACRÔNIMOS

WHO	World Health Organization
IARC	International Agency for Research on Cancer
FIOCRUZ	Fundação Oswaldo Cruz
HSV	Hue-Saturation-Value
RGB	Red-Green-Blue
IS	Positive cells Intensity Score
PS	Positive cells Proportion Score
WSI	Whole Slide Image
ER	Receptor de Estrogênio
PR	Receptor de Progesterona
IHQ	Imuno-Histoquímica
FLD-MNN	Fisher Linear Discriminant Preprocessing

SUMÁRIO

1 INTRODUÇÃO

Dados da *International Agency for Research on Cancer* indicam que em 2020 o câncer de mama ocupou o primeiro lugar dos índices de ocorrência internacionais de cânceres (IARC, 2023), e as estimativas para 2025 são de um aumento de 10,6% (INCA, 2020). No Brasil, o câncer de mama está em segundo lugar entre os tumores malignos mais comuns, com uma taxa de incidência de 10,5%, de acordo com dados fornecidos pelo Instituto Nacional de Câncer (INCA, 2020). No artigo Teixeira e Araújo (2020) escrito por pesquisadores do Fiocruz descreve um panorama geral da evolução das pesquisas sobre o assunto no território nacional e aborda as controvérsias envolvendo a disputa das tecnologias nesse setor, enfatizando o câncer de mama como um tema de pesquisa significativo no âmbito de saúde publica. Um dos exemplos é adoção da campanha Outubro Rosa como um reflexo da conscientização e relevância social tanto para a população quanto para as autoridades de saúde. Devido à importância crucial de estudos que contribuam para o diagnóstico precoce do câncer de mama, este estudo concentra-se na proposição e análise de métodos modernos de segmentação e contagem de células no diagnóstico da doença, adotando uma abordagem de *pipelene* semi-automática em um processo de diagnóstico assistido por computador.

1.1 PROPOSTA

Utilizando técnicas de segmentação clássicas e da proposto de (Mouelhi et al., 2013), este trabalho sugere uma metodologia de *pipelene* para identificar receptores de estrogênio (RE), essenciais no diagnóstico de câncer de mama. Espera-se que este *pipelene* seja parte da etapa de *Positive cells Intensity Score*, automatizando a contagem de células, afim de fornecer de maneira quantitativa a expressão de RE nas células. Dentro dos objetivos esperados com este trabalho estão:

- Execução do algoritmo de segmentação com técnicas clássicas;
- Utilização de algoritmo de divisão de células sobrepostas;
- Definição de um *pipelene* semi-automático que se utiliza de informações dispostas do processo de IS.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais que são essenciais para a elaboração deste estudo. Ele discute e revisa determinados assuntos sob uma perspectiva computacional, enquanto outros são examinados considerando o processo diagnóstico do cancer de mama.

2.1 CÂNCER DE MAMA

O câncer é uma doença caracterizada pelo crescimento descontrolado e a disseminação anormal de células. No câncer de mama esse fenômeno ocorre no tecido mamário, geralmente começa nas células dos ductos mamários (câncer ductal) ou nas glândulas que produzem leite (câncer lobular). O câncer de mama é comumente detectado por meio de exames de rastreamento ou pela identificação de um nódulo. A análise microscópica do tecido mamário é essencial para determinar a extensão e o tipo da doença. Testes de IHC, descritos na seção anterior, são utilizados para identificar biomarcadores cruciais, tais como os Receptores Hormonais (Estrogênio e Progesterona) e o Receptor 2 do Fator de Crescimento Epidérmico Humano (HER-2). Esses biomarcadores são avaliados para orientar o tratamento, fornecendo informações cruciais para uma abordagem mais personalizada e eficaz no tratamento do câncer de mama (Oncoguia, 2023).

2.2 IMUNO-HISTOQUÍMICA

A imuno-histoquímica (IHQ) é uma técnica utilizada na biologia e na patologia para identificar a distribuição e a presença de proteínas em amostras de tecidos, a qual se baseia em princípios de histologia e imunologia. Ela funciona detectando proteínas específicas dentro das células de uma amostra de tecido, usando anticorpos que se ligam a essas proteínas. Esses anticorpos são marcados de forma que possam ser vistos ao microscópio. A técnica é muito útil para identificar diferentes tipos de células em doenças como o câncer ou para detectar a presença de agentes infecciosos (Diagnósticos do Brasil, 2021).

2.3 ALLRED

Allred é um sistema de pontuação utilizado para avaliar o status dos receptores de estrogênio em células de câncer de mama. Ele foi desenvolvido para fornecer uma avaliação mais precisa e padronizada do status ER, o que é crucial para o planejamento do tratamento do câncer de mama, especialmente em relação à terapia hormonal. Esse sistema é dividido em duas etapas:

- *Positive cells Proportion Score* (PS): Esta parte da pontuação varia de 0 a 5, com base na porcentagem de células tumorais que mostram positividade para o receptor hormonal.
- *Positive cells Intensity Score* (IS): Esta parte da pontuação varia de 0 a 3, dependendo da intensidade da coloração observada nas células positivas.

Por fim, a pontuação total do *Allred* é a soma dessas duas partes e pode variar de 0, sendo nenhuma expressão, até 8 representando alta expressão. Um *score* mais alto indica uma maior expressão dos receptores hormonais. O sistema de pontuação de *Allred* é importante porque ajuda a determinar mais claramente se um tumor é positivo ou negativo para ER, o que pode influenciar as decisões sobre o uso de terapias hormonais (Qureshi e Pervez, 2010).

2.4 DIAGNÓSTICO ASSISTIDO POR COMPUTADOR

O Diagnóstico Auxiliado por Computador (CAD) é uma tecnologia revolucionária na área de diagnóstico médico por imagem, que serve como uma ferramenta de suporte para os radiologistas. Neste sistema, análises quantitativas e automatizadas de imagens radiográficas são realizadas por computador, fornecendo aos médicos uma "segunda opinião"na interpretação dos resultados(Azevedo Marques, 2001).

O impacto do CAD é significativo ao melhorar a capacidade dos médicos de detectar lesões sutis e aumentar a confiança nos diagnósticos, mesmo com a presença de falsos positivos. O CAD não busca igualar ou superar o desempenho dos médicos, mas sim complementá-lo, oferecendo benefícios adicionais na interpretação das imagens médicas (Doi, 2007).

O CAD representa um avanço notável no campo do diagnóstico médico por imagem, e a medida que continua a evoluir, espera-se que desempenhe um papel cada vez mais vital na medicina diagnóstica, beneficiando pacientes e profissionais de saúde ao redor do mundo (Doi, 2007).

2.5 TIPO DAS IMAGENS

O conjunto de dados deste estudo consiste em Imagens de Lâmina Inteira (Whole Slide Images, WSIs), imagens digitais de alta resolução obtidas de lâminas de microscópio. Essas WSIs, formadas por várias pequenas imagens consecutivas (*patches*), representam campos de visão microscópicos e, por sua natureza, são grandes e necessitam de compressão. O conjunto específico abrange 270 WSIs de imuno-histoquímica de 135 pacientes, avaliados para os biomarcadores ER e PR. Ampliadas 40x e classificadas pelos *Intensity Score* (IS) e *Proportion Score* (PS), as WSIs foram rotuladas em duas categorias por biomarcador para cada paciente, com IS variando de 0 a 3+ e PS de 0 a 5.

2.6 SEGMENTAÇÃO

O processo de segmentação no cenário de visão computacional pode ser definido como um processo de decomposição de uma imagem digital em regiões ou objetos (Gonzalez e Woods, 2007). Por meio dessas técnicas é possível extrair informações e trabalhar com elas individualmente. Dentre as principais técnicas existentes duas terão destaque neste trabalho e serão apresentados abaixo.

2.6.1 Limiarização

Limiarização é uma técnica que se baseia na análise do histograma de uma imagem onde ocorre uma redistribuição dos dados em n-modal, esses dados são informações de cor em tons de cinza que serão divididos em grupos de valores, ou de outra forma, intervalos de valores. Esses intervalos ou grupos na técnica de limiarização passam a ser entendidos como semelhantes (Gonzalez e Woods, 2007).

Um exemplo desse processo é ilustrado na figura 2.1, onde a função de limiarização é parametrizada com zero. Esta configuração ajusta os valores de modo que, se forem maiores que zero, recebam o valor 1; caso contrário, recebam 0.

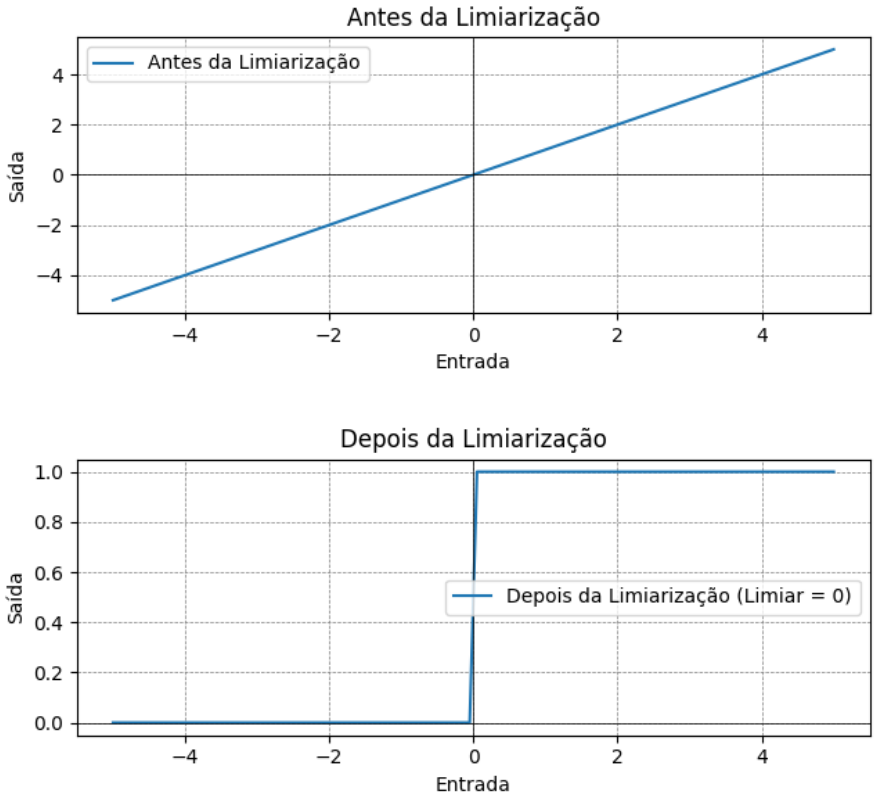


Figura 2.1: Exemplo limiarização com limiar zero.

2.6.2 Clustering

O termo *clustering* refere-se a um conjunto de algoritmos destinados ao agrupamento de objetos com um grau específico de similaridade (Kleinberg e Tardos, 2005). Essa técnica baseia-se em coeficientes de afinidade, tais como densidade, centróides, distância, entre outros. Um exemplo prático que ilustra a aplicação do algoritmo K-means pode ser observado na figura 2.2. Um dos algoritmos de *clustering* é o K-Means tem como objetivo agrupar dados, tentando distribuir as amostras em n grupos, de forma que a variância seja aproximadamente igual entre esses grupos, ao mesmo tempo em que minimiza um critério amplamente reconhecido chamado inércia ou soma dos quadrados intra-cluster (Arthur e Vassilvitskii, 2007), assim como foi observado na figura 2.2.

2.6.2.1 Watershed

A técnica conhecida como transformada *Watershed* adota uma abordagem morfológica para resolver o desafio de segmentar imagens. Nessa abordagem, as imagens são tratadas como superfícies, onde cada pixel é considerado uma coordenada, e os diferentes níveis de cinza simulam altitudes. A ideia principal é identificar áreas que se assemelham a bacias hidrográficas, as quais são definidas por mínimos locais e as áreas circundantes sob sua influência (Beucher e Meyer, 1993). Pode-se enxergar este método com uma analogia na qual podemos pensar em elevar gradualmente o nível de água na imagem, começando nos mínimos locais. Conforme o nível da água aumenta, áreas de diferentes mínimos regionais se encontram, formando uma barreira, que efetivamente delineia as bordas e regiões na imagem. Em um contexto de visão

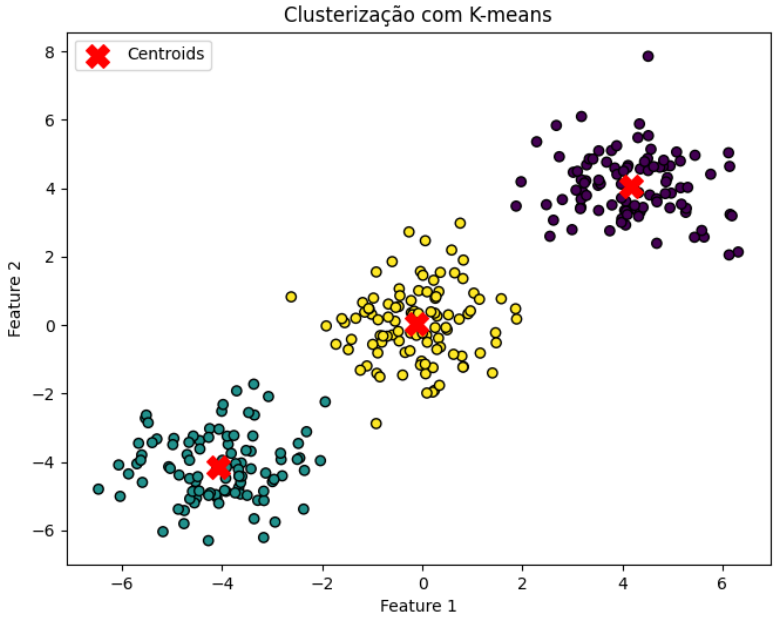


Figura 2.2: Exemplo Clustering em um gráfico com três grupos utilizando o algoritmo K-means.

computacional conseguimos simular esse evento onde os pontos mínimos são pontos de objetos acoplados que temos intenção de separar.

2.7 ESPAÇO DE COR

Em representações digitais das imagens conseguimos representá-las de diversas maneiras, duas formas muito comuns são HSV (*Hue-Saturation-Value*) e RGB (*Red-Green-Blue*). No espaço de cor HSV o H é o responsável por representar as cores, e é ajustado com valor angular de 0 a 360, sendo a demonstração pura ou essencial da cor, como o vermelho, azul e verde (Nishad, 2013). Na camada S definimos a pureza da cor, logo estamos variando a a relação com o branco e sua vividade. já no *Value*, definimos o brilho, onde quando menor o seu valor, maior sua relação com a cor preta, como podemos ver na figura 2.3.

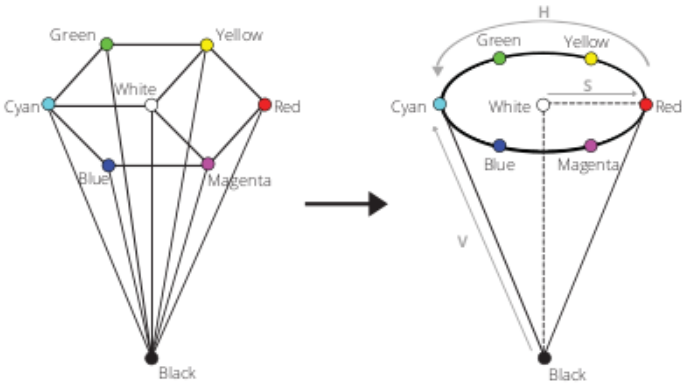


Figura 2.3: Representação do espaço de cor HSV.
(ROGALSKY, 2021)

No espaço RGB, a representação ocorre de outra maneira, são 3 canais que representam as cores primarias Vermelho, Verde e Azul. Para esse espaço, a representação é no intervalo de 0 a 255 para cada segmento que, juntos, compõe as cores. O modelo RGB é descrito como um cubo tridimensional dentro deste espaço, cada ponto é definido pela combinação das intensidades máximas e mínimas de cada uma das cores primárias. A cor preta é obtida quando o vermelho, o verde e o azul estão ajustados para os níveis mais baixos, enquanto a cor branca surge do ajuste dessas três cores primárias aos seus níveis máximos. A mistura de cores no cubo RGB segue esta regra fundamental, uma representação deste cubo tridimensional, pode ser visto na figura 2.4

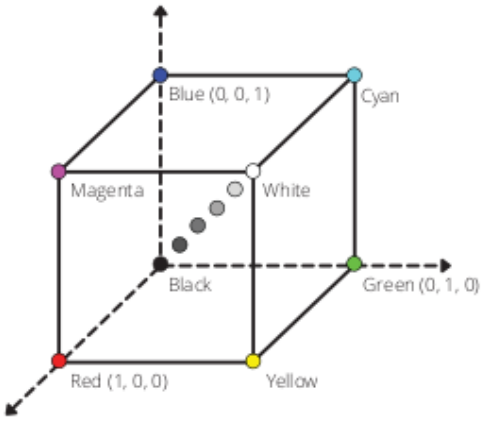


Figura 2.4: Representação do espaço de cor RGB.
(ROGALSKY, 2021)

2.8 MOMENTOS DE HU

A teoria desenvolvida por Ming-Kuei Hu, conforme detalhado em sua publicação (Hu, 1962), introduz o conceito de invariantes de momento bidimensionais aplicados a figuras planas. Esta teoria é significativa por estabelecer uma relação direta entre os invariantes de momento de uma figura e as propriedades invariantes de formas geométricas planas. Essencialmente, este relacionamento permite a identificação e o reconhecimento de padrões geométricos de maneira eficaz, independentemente de variações na posição, escala ou orientação dessas figuras no espaço. Eles são definidos como:

- $\phi_1 = \eta_{20} + \eta_{02}$, que é invariante à translação e escala, capturando a dispersão geral da forma.
- $\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$, que oferece invariância adicional à rotação, medindo a simetria da forma.
- $\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$, sensível à assimetria da forma.
- $\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$, que também detecta assimetria, mas com ênfase em características distintas.
- $\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$, captura a complexidade da forma.

- $\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$, que é útil para análise de inclinação.
- $\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$, que diferencia espelhamento de formas.

Cada momento de Hu é projetado para capturar propriedades únicas de uma forma, proporcionando uma análise robusta de características geométricas que são consistentes sob transformações de escala, translação e rotação.

2.9 MORFOLOGIA MATEMÁTICA

A morfologia matemática, conforme abordada em (OpenCV, 2023), são operações simples realizadas sobre imagens binarias para processamento de imagens em tons de cinza ou binarias. Essa operações tem como entradas a imagem binaria e um kernel que atua como elemento estruturante. Algumas das principais operações são:

- **Dilatação:** A dilatação é uma operação que expande as regiões de interesse em uma imagem. Essa expansão é realizada pela inclusão de pixels vizinhos aos já existentes, aumentando assim a área ocupada pelas regiões de interesse.
- **Erosão:** A erosão, por sua vez, é uma operação que contrai as regiões de interesse na imagem. Ela é realizada removendo pixels da borda das regiões, o que pode resultar na quebra de conexões entre componentes e na redução do tamanho das regiões.
- **Abertura:** A abertura é uma operação composta que combina a erosão seguida pela dilatação. A erosão é aplicada inicialmente para remover pequenos detalhes e regiões indesejadas, seguida pela dilatação para restaurar a forma geral e a estrutura das regiões remanescentes (Tsesmelis, 2023).

Essas operações são executadas em conjuntos de pixels, utilizando o elementos estruturantes que define a forma e o tamanho das operações.

2.10 TRABALHOS RELACIONADOS

Para o desenvolvimento deste trabalho, foram revisados artigos e publicações teóricas que possuem relação com a segmentação e contagem de células, e que contavam com técnicas no âmbito de visão computacional. Portanto, mesmo que alguns destes tenham foco em alternativas como redes neurais, focamos em entender os processos de processamento de imagem e afins.

Na dissertação ROGALSKY (2021), a autora propõe uma alternativa semiautomática no processo de score para o conjunto de PR e PS. As imagens utilizadas no trabalho utilizaram o formato *Whole Slide Images* (WSI), e em sua etapa de pré-processamento das imagens, foram utilizadas técnicas como equalização de histograma, limiarização, separação de cores utilizando o espaço HSV, histograma CLBP e estatísticas GLCM. Já para a etapa de classificação, é mencionado o uso do classificador SVM. Com a junção dessas técnicas, os resultados apresentados alcançam mais de 76% na taxa de F1-score em uma abordagem bloco a bloco.

Em Mouelhi et al. (2013), o autor propõe uma alternativa à segmentação de células de ER em biópsias de câncer de mama. Para lidar com a variedade de características, o autor utiliza cor e tecnologia de clusterização *Fuzzy* para detectar o núcleo das células. Além disso, propõe uma nova etapa que identifica sobreposição e núcleos unidos para separá-los, utilizando um

algoritmo de *Watershed* aprimorado baseado em um grafo de vértices côncavos, em um pipeline finalizado com um classificador FLD-MNN (Fisher Linear Discriminant Preprocessing), que respondeu com acurácias acima de 97,8%. Em outro trabalho (Mouelhi et al., 2018), o mesmo autor desenvolveu um algoritmo que realiza a classificação automática do escore *Allred*, utilizado para avaliar a proporção e intensidade de células positivas para receptores de estrogênio em amostras. Os resultados obtidos indicaram uma precisão de mais de 98% na detecção de núcleos e classificação do score de câncer *Allred* em um conjunto de dados composto por 84 imagens. O método de segmentação empregado no estudo se baseou em diversas técnicas, incluindo limiares locais adaptativos, operações morfológicas, um filtro Laplaciano modificado e uma melhoria no método de segmentação de imagens *Watershed*.

Outra abordagem foi a proposta em (Silva, 2015), neste trabalho, em que o autor utiliza uma abordagem de morfologia para contagem de células em imagens de placas de Petri no formato TIFF (*Tagged Image File Format*). Nesta abordagem, foi proposto um software com interface que obteve 89% de sucesso para estimar o valor médio de núcleos nas amostras. Outras soluções propostas utilizando a estratégia de software são a QuickCount (Tiong et al., 2018), IMAGEJ (Schneider et al., 2012) e CellProfiler (Carpenter et al., 2006), entre outras. A contagem de células é um tema de pesquisas de âmbito patológico tanto na área de pesquisa quanto nos setores do mercado/industria.

2.11 CONCLUSÃO

Nesta seção revisamos conceitos importantes para o desenvolvimento deste trabalho e alguns conceitos que foram utilizados na implementação das técnicas, além disso, foi possível ter um aparato de tecnologias e experimentos presentes no assunto abordado neste trabalho.

3 MATERIAIS E MÉTODOS

3.1 CONJUNTO DE DADOS

Os dados empregados na validação do pipeline constituíram um subconjunto do HistoBC-HR. Dentre as classes 0, 1+, 2+, e 3+ associadas ao grupo ER, 100 imagens foram aleatoriamente selecionadas. No entanto, é crucial ressaltar a notável disparidade na distribuição dessas classes. A classe 3+, que representa mais da metade das amostras disponíveis, contrasta fortemente com a classe 1+, que abrange apenas 8,3% do conjunto. Essa desproporção evidencia de maneira expressiva o desbalanceamento do banco de dados, uma consideração vital para a interpretação e confiabilidade dos resultados obtidos, dado que as classes com maior quantidade de amostra consegue representar melhor a diversidade encontrada na classe. A seguir, apresenta-se a distribuição das classes:

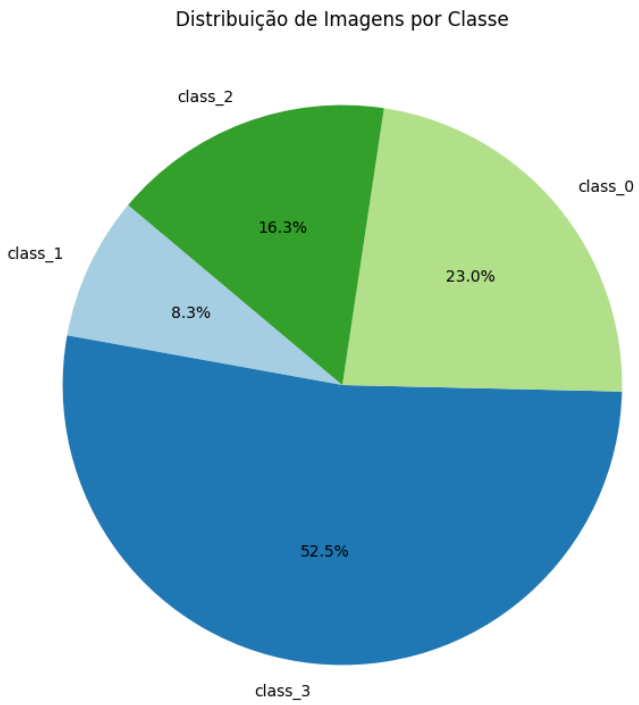


Figura 3.1: Distribuição das classes no *dataset*.

Para proporcionar uma análise mais aprofundada de cada classe, apresenta-se na figura 3.2 uma amostra de cada uma das classes.

3.2 PIPELINE

Tendo em vista que o calculo de IS antecede a etapa de PS, foi proposto um pipeline que se adapte a potencial classe identificada na primeira etapa, isso se deu ao fato de que as

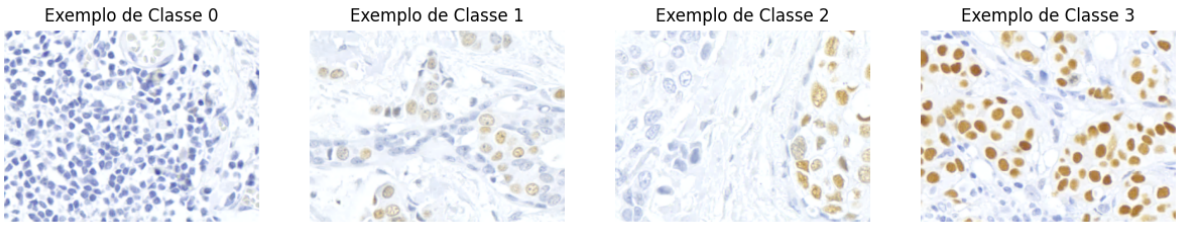


Figura 3.2: Exemplo de cada classe, onde classe 0 (negativa), classe 1+ (fraca/leve), classe 2+ (moderada), classe 3+ (forte/intensa).

características predominantes em cada classe como, por exemplo, as cores servem como uma referência para cada classe. Desta forma para cada classe IS temos um algoritmo específico, o que automatiza as escolhas dos valores de *threshold* e melhora os resultados. Foi realizado uma análise individual de cada classe IS para ter um fluxo que conseguisse extrair o máximo de características das imagens. O pipeline resultante sera descrito nessa subseção e pode ser visualizado na figura 3.3.

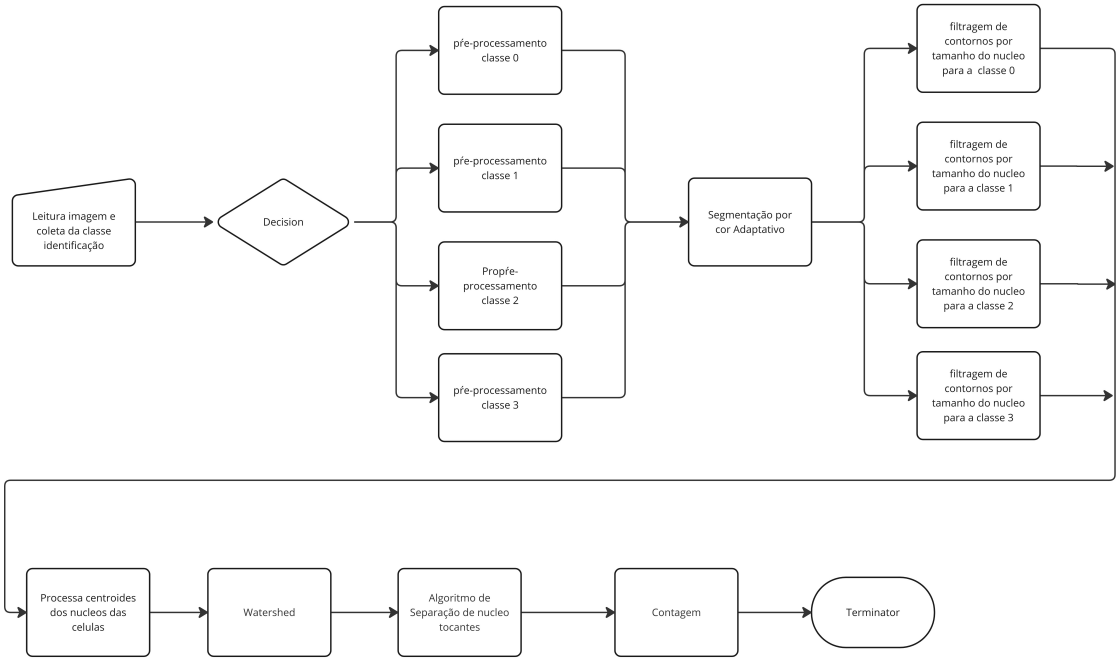


Figura 3.3: Etapas do pipeline proposto

3.2.1 Pré-processamento

O conjunto de dados foi submetido a um pipeline que consistiu inicialmente em um pré processamento, com a técnica de ajuste de contraste chamado Kernel Laplaciano. Este kernel age realçando as áreas da imagem que possuem uma mudança abrupta nos valores dos pixel, onde normalmente essas mudanças indicam uma potencial borda, aplicando uma convolução foi possível obter imagens com bordas realçadas e com alto contraste como mostrado em 3.6. Esta técnica foi aplicada em todas as classes.

3.2.2 Segmentação

Cada classe IS possui suas características, destacando-se pela presença de determinadas cores e tamanhos de células. Essas características levaram ao processo de segmentação de cor. Inicialmente, foram analisadas outras alternativas, como a proposta por (Mouelhi et al., 2013), que utilizou clusterização *Fuzzy*. De forma similar, foi testada uma outra alternativa utilizando Kmeans, mas, neste caso, a abordagem que obteve os melhores resultados foi a segmentação de cor.

Para a segmentação de cor, selecionou-se a imagem processada na etapa anterior e aplicou-se uma conversão do espaço de cores de RGB para HSV, permitindo manipular e isolar melhor as cores. Após essa etapa, os canais foram separados em três vetores: H, S e V e analisados individualmente para cada classe.

Com o objetivo de isolar as cores do *background* da imagem, ou seja, o espaço além das células, calculou-se a média do canal Value. Essa estratégia foi adotada com base em uma análise inicial das presenças de cores das imagens do banco original, e o que melhor descreveu o comportamento do *background* das células foi o canal V, esse processo gerou os gráficos descritos em 3.4. Utilizando o valor encontrado, criou-se uma máscara com o intervalo de 0 até a média encontrada, e, por fim, fez-se a diferença sobre a imagem pré-processada e a máscara encontrada, gerando uma nova imagem binária que corresponde ao *background* da imagem. Como temos interesse no contorno das células, foi aplicada uma operação de inversão, resultando na segmentação das células como o exemplo 3.4.

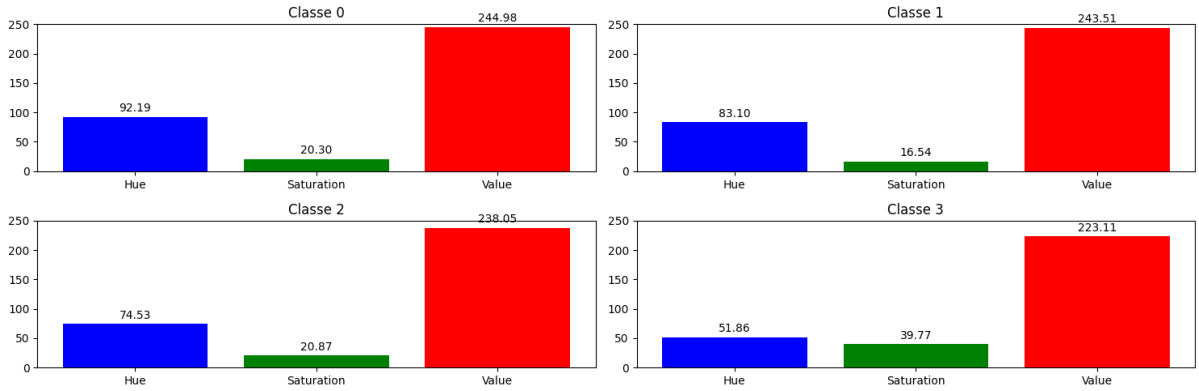


Figura 3.4: Distribuição dos canais de cores das classes utilizando a média de todas as imagens.

3.2.3 Pós-processamento da segmentação

Para o pós processamento da segmentação foram utilizadas técnicas de morfologia matemática para fazer a limpeza dos ruídos e fechar as células que possuem em sua imagem original centros com algum nível de transparência como mostrado na imagem 3.5, além de limpar os ruídos da segmentação e ajustar as cores binarias como na figura 3.6.

Essa etapa impactava consideravelmente alguns núcleos mais sensíveis, influenciando diretamente no resultado final. Portanto, as alterações realizadas nesse momento foram suavizadas, visando minimizar esse impacto.

Para um melhor resultado, o pós-processamento da segmentação foi ajustado de forma adaptativa a cada classe IS. A seguir, apresentamos uma visão sobre os ajustes realizados nas segmentações.

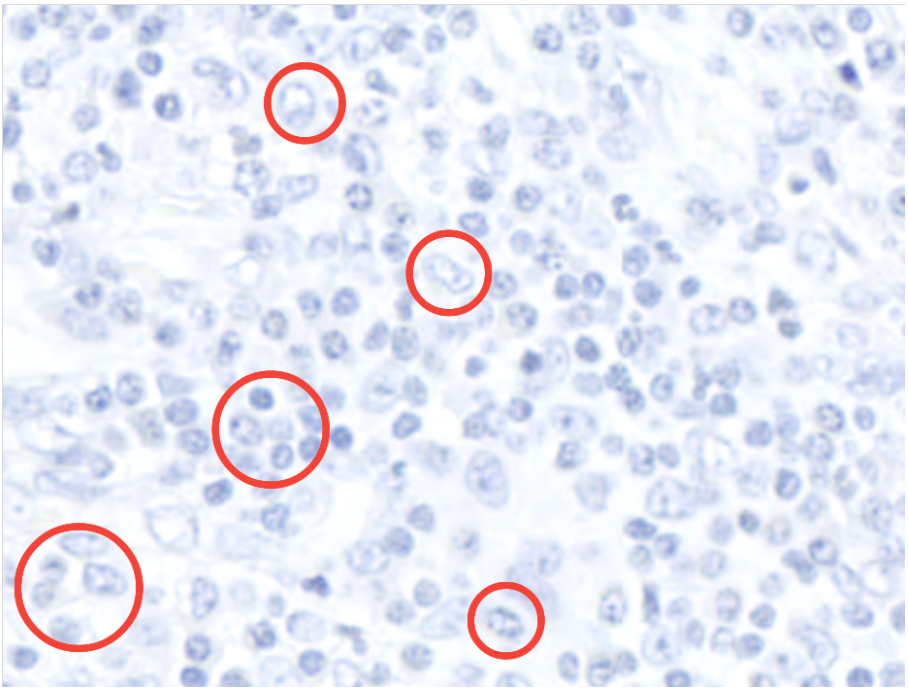


Figura 3.5: Amostra de uma imagem com excesso de células com núcleos translúcidos, onde alguns estão marcados com círculos vermelhos.

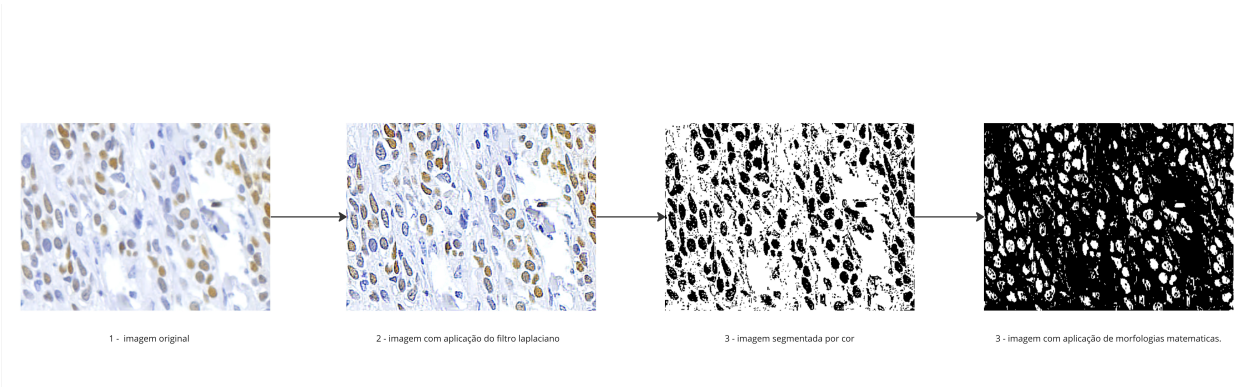


Figura 3.6: Exemplo do pipeline até a etapa de pós processamento.

3.2.3.1 Classe 0

Nesta etapa, foi necessário aplicar algumas operações adicionais. As imagens da classe 0 possuem muitas células opacas e bem menores em comparação com as outras classes, tornando-as mais sensíveis às operações morfológicas de erosão. Portanto, antes de aplicar essa técnica, foi necessário preencher as células com operações de fechamento.

3.2.3.2 Classes 1 e 2

As células presentes nas classes 1 e 2 não sofrem do problema citado na classe 0. Além disso, nessas classes, as células indicadoras da presença do câncer são mais aparentes. Portanto, para esse grupo, é possível aplicar técnicas de erosão para melhorar a separação das células.

3.2.3.3 Classe 3

Como nesta classe as células possuem grandes indicadores da presença de câncer, elas estão mais contrastadas, permitindo a operação de erosão.

3.2.4 Filtragem dos contornos

Outra etapa crucial na construção do pipeline foi a limpeza dos contornos, pois a segmentação gerava uma quantidade considerável de ruídos e contornos irrelevantes para a identificação das células na imagem. Nessa etapa, foram realizadas as seguintes ações:

- Realizamos a extração de contornos que correspondiam ao tamanho médio especificado para a classe em questão. Essa análise foi realizada de forma aleatória, examinando individualmente cada amostra para determinar esses valores;
- Remoção de contornos muito pequenos, definidos com o tamanho médio analisado nas amostras;
- Remoção de contornos que se encontravam dentro de outros contornos.
- Remoção de contornos na borda, proposta pelo autor (Mouelhi et al., 2013). Nesta técnica, analisamos individualmente cada contorno da imagem, buscando eliminar células que não viabilizam serem analisadas por completo, ou seja, células que possuem alguma parte fora das bordas da imagem.

Para isso, cada contorno foi avaliado individualmente, sendo aplicada uma filtragem para manter apenas os contornos que atendiam aos parâmetros específicos da classe fornecida. Essa abordagem contribuiu significativamente para a melhoria da qualidade da segmentação, eliminando elementos indesejados e preservando apenas os contornos relevantes para a identificação das células.

3.2.5 Marcação dos Centroides

Por fim, efetuamos a busca de centróides 3.7, que se refere ao centro das células. Essa informação foi necessária visando a etapa de *watershed*. Para executar a busca dos centróides, utilizamos o momento central de ordem zero de Hu, permitindo encontrar as coordenadas x e y a partir dos momentos 00 e 01.

3.2.6 Watershed

Para a etapa de *watershed*, foram propostas duas alternativas. Na primeira, uma estratégia documentada por Tsesmelis em (Tsesmelis, 2023) utiliza a função OpenCV *distance Transform* para obter a representação derivada de uma imagem binária, onde o valor de cada pixel é substituído pela sua distância ao pixel de fundo mais próximo. Nessa estratégia, o polimento das imagens no final ficou visivelmente melhor, mas o algoritmo teve dificuldades em diferenciar células que não possuíam um formato polido, algo comum nas imagens do domínio do problema. Na segunda abordagem, empregamos a imagem binária resultante da filtragem e aplicamos operações de morfologia matemática para criar imagens contendo marcadores dos pontos de interesse, cujo propósito é a subsequente separação das células.

As imagens geradas na etapa de *watershed* possuem algumas células agrupadas. Para tentar solucionar esse problema, foi utilizada uma técnica proposta por (Mouelhi et al., 2013),

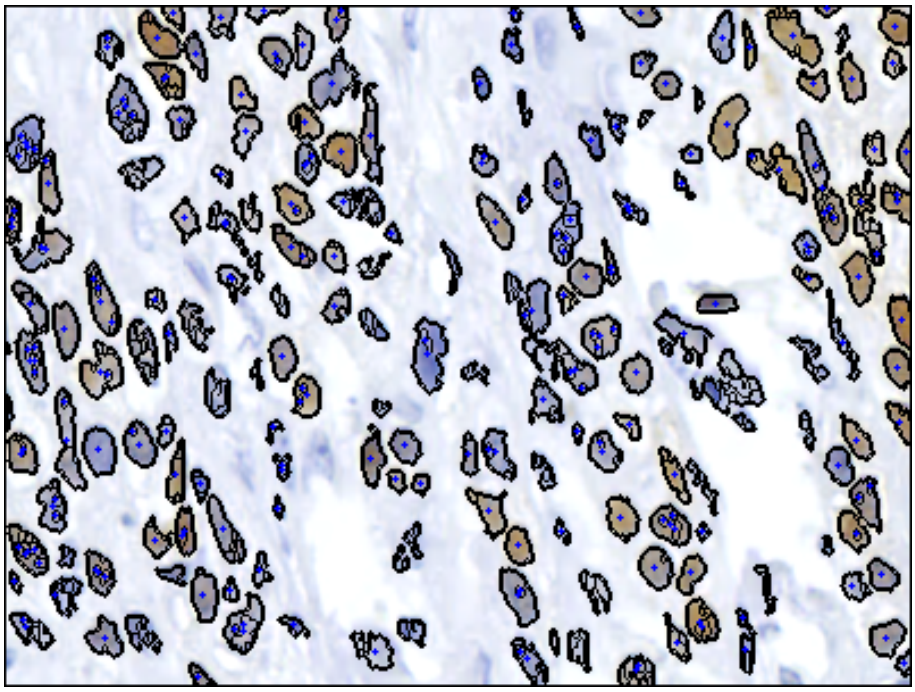


Figura 3.7: Exemplo das imagens com centroides das células em uma imagem da classe 2.

onde são buscados os pontos côncavos dos contornos e analisados individualmente. Por fim, identificamos os pontos côncavos da imagem, empregando a técnica para selecionar os pontos de separação entre células tocantes e agrupadas descritos em (Mouelhi et al., 2013). Para o uso desse algoritmo foi necessário implementar a busca dos pontos côncavos além da filtragem dos pontos côncavos que não se encaixavam nos lineares testados, essa relação entre os pontos côncavos pode ser visto na imagem 3.8.

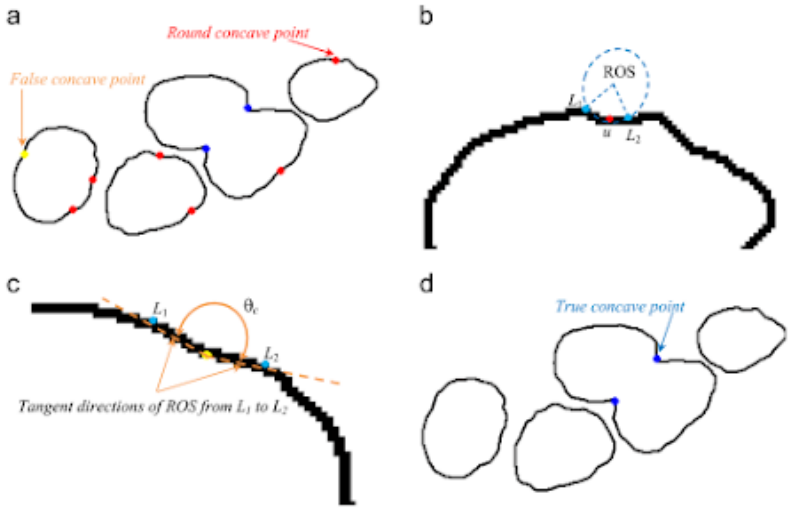


Figura 3.8: Detecção de pontos côncavos: (a) candidatos iniciais a pontos côncavos, (b) região de suporte (ROS) de um ponto côncavo redondo, (c) avaliação do ângulo de um ponto côncavo candidato e (d) pontos côncavos verdadeiros.

Mouelhi et al.

3.2.7 Contagem de Celulas

Para realizar a contagem do número de núcleos celulares, adotamos a mesma técnica detalhada no Capítulo 3.2.5. Após a execução de todas as técnicas propostas, registramos e armazenamos a quantidade de núcleos identificados. Este processo foi crucial para a obtenção de dados quantitativos que é usado na etapa de PS, permitindo uma avaliação precisa do conteúdo celular nas imagens processadas utilizando da relação entre células tumorais positivas e o total de células presentes na amostra.

3.3 EXPERIMENTO

3.3.1 SETUP

Para executar o pipeline foi utilizado a linguagem de programação python 3.8 (Van Rossum e Drake, 2009), responsável em controlar o fluxo das informações para cada etapa. Nas etapas de segmentação houve uso das bibliotecas Numpy (Harris et al., 2020) para manipulação de matrizes e a openCV (Bradski, 2000) para uso de função de manipulação de imagens e uma das alternativas além desta foi utilizando a biblioteca ScikitLearning (Pedregosa et al., 2011), Por fim os resultados foram gerados utilizando a biblioteca MatplotLib (Hunter, 2007).

3.3.2 Teste Pipeline

Para o experimento de segmentação, foi desenvolvido um *script* que recebe como entrada um *dataset* no formato:

```
dataset/
|-- class_0/
|-- class_1/
|-- class_2/
|-- class_3/
```

Com essa organização, o algoritmo é responsável por executar para cada imagem dos subdiretórios, levando em consideração o diretório pai para definir sua classe. O retorno é um *dataset* no mesmo formato com as imagens de saídas.

3.3.3 Teste de contagem

A execução gerou um CSV contento uma tabela com o nome da imagem, a classe de entrada e por fim a contagem do algoritmo para validação manual dos profissionais, um exemplo da saída pode ser observada na tabela abaixo.

Tabela 3.1: Tabela representando o CSV de saída do algoritmo, contendo o nome da entrada utilizada assim como sua classe e a contagem resultante.

Nome	**Classe**	**Contagem**
./database/class_1/499RE_s0c0x154135-1600y77167-1200m1989.png	class_1	200
./database/class_1/534RE_s0c0x136834-1600y98584-1200m3580.png	class_1	301
./database/class_1/631RE_s0c0x171032-1600y144577-1200m7324.png	class_1	302
./database/class_1/491RE_s1c0x162367-1600y50361-1200m1846.png	class_1	97
./database/class_1/051RE_s1c0x225755-1600y131807-1200m5817.png	class_1	316
./database/class_1/096RE_s2c0x189886-1600y139308-1200m9092.png	class_1	320
./database/class_1/499RE_s0c0x164141-1600y84677-1200m2271.png	class_1	209

4 RESULTADOS E DISCUSSÃO

Neste capítulo, descreveremos uma avaliação qualitativa dos resultados alcançados por meio das técnicas previamente propostas. Cada subseção será dedicada a uma análise aprofundada das características das etapas do pipeline, aliado as características das classes, visando identificar tanto os pontos positivos quanto os negativos do pipeline. O propósito subjacente é realizar uma discussão minuciosa sobre a efetividade dessas abordagens, proporcionando uma compreensão mais abrangente do desempenho do sistema. Além disso, será apresentada uma análise detalhada dos resultados gerados, permitindo-nos extrair características valiosas afim de fundamentar as considerações sobre a eficácia global do processo em questão.

Devido a restrições temporais significativas, alguns resultados deste estudo não estavam prontos para validação no momento da redação deste artigo. Embora tenhamos obtido resultados preliminares notáveis, é crucial reconhecer que a falta de validação pode impactar a robustez das conclusões. Este desafio sublinha a importância de futuras investigações que possam abordar as limitações identificadas e validar completamente os resultados apresentados.

4.0.1 Resultados

4.0.1.1 Segmentação

Para a avaliação da segmentação seria necessário os valores reais das contagens de células presentes em cada amostra no conjunto de dados, oque não foi possível obter tempo com um patologista. Uma amostra de cada classe pode ser vista na imagem 4.1, 4.2, 4.3 e 4.4. Restando a avaliação como sugestão de trabalhos futuros.

1. **Classe 0:** Para a classe 0, observou-se uma irregularidade marcante na morfologia das células. Os formatos apresentam notável ausência de padrões arredondados, fenômeno atribuído à falta de preenchimento nas células nas imagens originais. Este fato resulta na presença de centroides das células em branco quando a segmentação é aplicada, o que acarreta em significativa interferência no processo de segmentação, um exemplo de resultado pode ser visto na figura 4.1.

2. **Classe 1:** Para a classe 1, a segmentação apresentou resultados satisfatórios, como evidenciado pelo exemplo visualizado na figura 4.2. Esta classe possui características semelhantes à classe 0, notavelmente no que diz respeito às células com núcleos menos evidentes. Além disso, as células da classe 1 tendem a ter tamanhos menores quando comparadas às das classes 2 e 3. Esta similaridade na apresentação dos núcleos entre as classes 0 e 1, o que reforça a precisão da técnica de segmentação utilizada

3. **Classe 2:** Para a classe 2, o algoritmo demonstrou dificuldades em algumas imagens, conforme ilustrado na figura 4.3. Alguns pontos da imagem foram erroneamente considerados junto aos núcleos, entretanto, sem a avaliação de um especialista, não foi possível realizar ajustes no algoritmo para corrigir esses casos.

4. **Classe 3:** Assim como na classe 2, a classe 3 também apresentou marcações equivocadas em alguns locais da imagem, conforme evidenciado na figura 4.4. Além disso, devido ao método de segmentação que realiza a meédia da imagem, células com cores muito claras e distantes das células indicadoras positivas da doença também foram erroneamente desconsideradas.

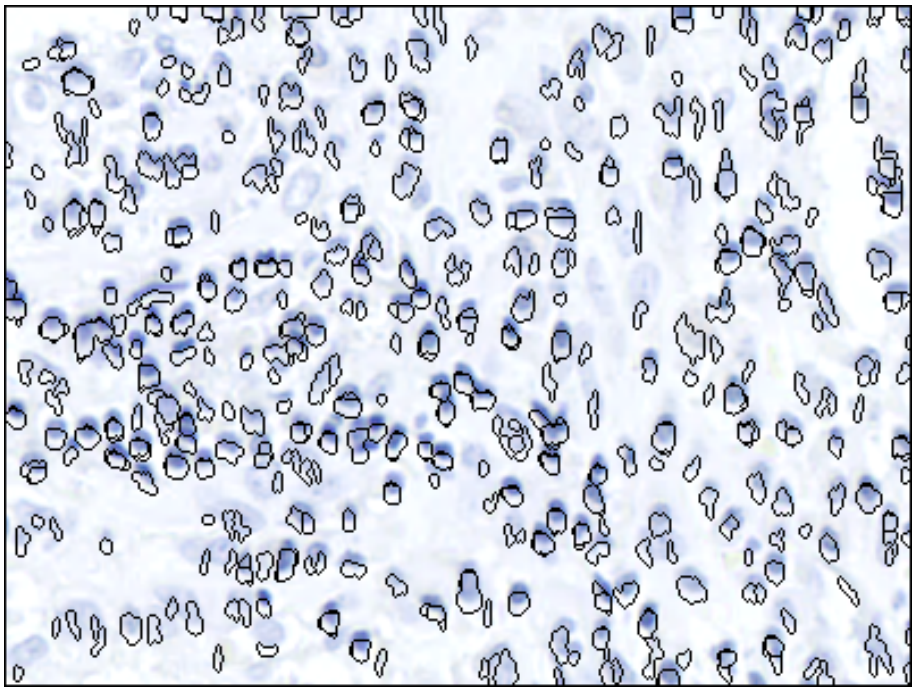


Figura 4.1: Exemplo de saída do algoritmo para a classe 0.

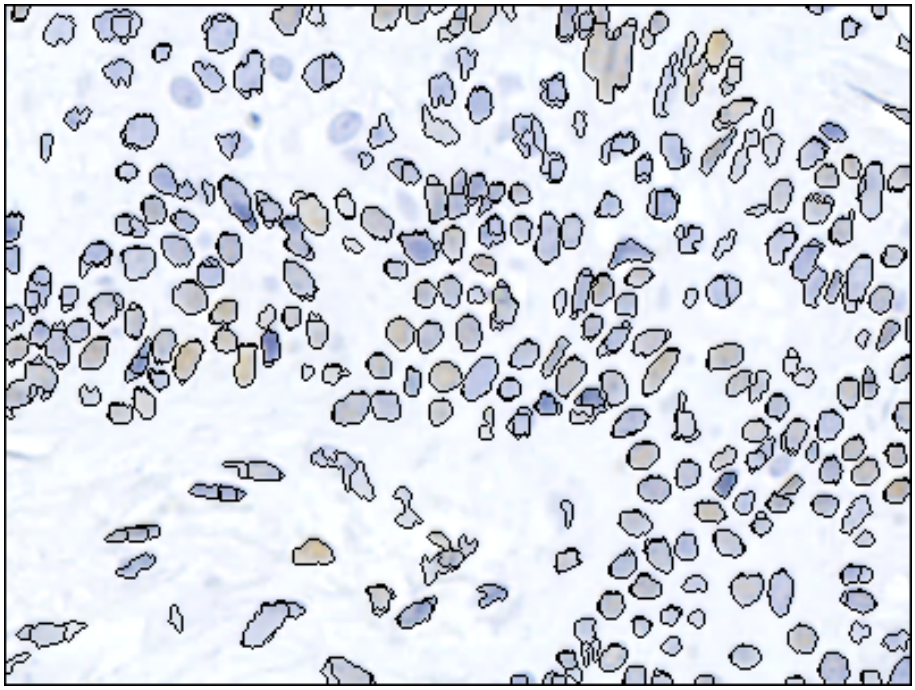


Figura 4.2: Exemplo de saída do algoritmo para a classe 1.

4.0.1.2 Algoritmo de divisão de células

Para o algoritmo proposto de divisão de núcleos tocantes, foram encontrados alguns problemas:

- **Complexidade de implementação da técnica:** A implementação da técnica em questão apresenta uma série de desafios e complexidades que, inevitavelmente, impactaram o cronograma estabelecido.

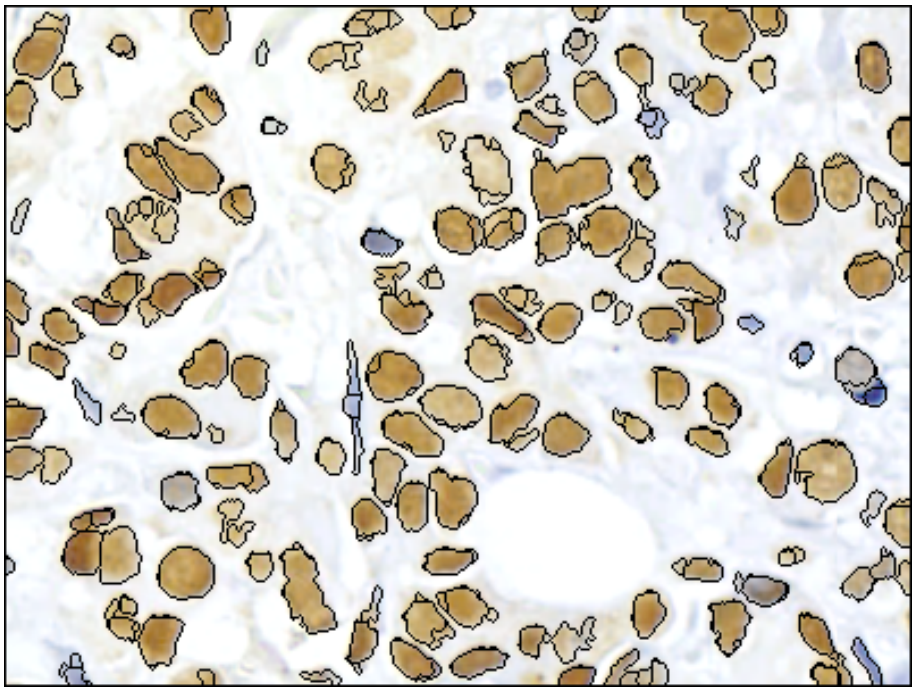


Figura 4.3: Exemplo de saída do algoritmo para a classe 2.

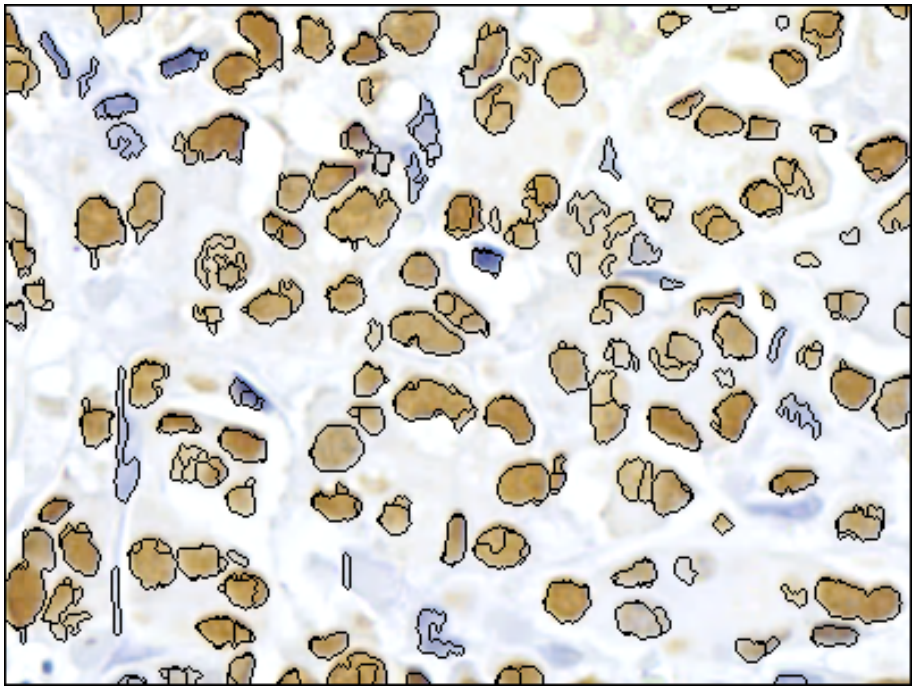


Figura 4.4: Exemplo de saída do algoritmo para a classe 3.

- **Necessidade de Segmentação Mais Precisa:** A segmentação resultante do algoritmo precisa ser mais refinada, especialmente considerando a morfologia diversificada dos núcleos nas imagens do banco de dados.
- **Desafios com Padrões de Morfologia Diversificada:** O uso de um código baseado no artigo (Mouelhi et al., 2013) não trouxe melhorias significativas à segmentação. Esse problema é atribuído à presença de muitos núcleos com morfologias diversas nas imagens do banco de dados.

- **Limitações em Casos de Pontos Côncavos:** O algoritmo enfrentou dificuldades em lidar com a presença excessiva de pontos côncavos nos núcleos como demostrado na figura 4.5. Essa limitação é documentada pelo autor em (Mouelhi et al., 2013) como uma área não eficiente do método.

Em resumo, complexidade de implementação, juntamente à falta de disponibilidade do código-fonte, juntamente com desafios relacionados à diversidade morfológica dos núcleos e à presença de pontos côncavos, impactou a eficácia do algoritmo proposto de divisão de núcleos tocantes. Esses fatores ressaltam a importância de considerar a complexidade das características das imagens de câncer de mama ao desenvolver técnicas de segmentação.

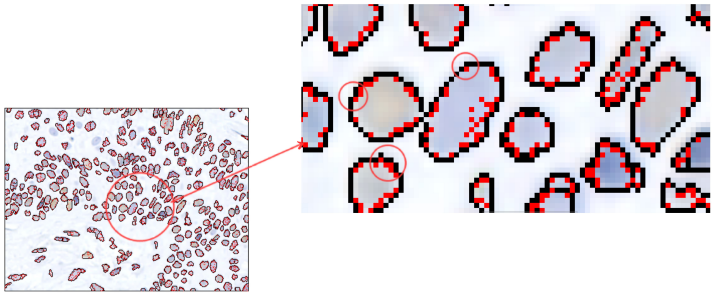


Figura 4.5: Exemplo de saída do algoritmo de separação de núcleos tocantes, sendo os pontos vermelhos a identificação incorreta dos pontos côncavos, além dos círculos vermelhos demonstrando problemas de identificação, ocasionados pelos contornos não lineares

4.0.2 Validação dos resultados

Os resultados não foram concluídos a tempo para serem validados e documentados neste artigo. Recomenda-se como trabalho futuro a realização de uma avaliação abrangente dos impactos do código, considerando os resultados obtidos.

4.0.2.1 Pipeline

Para validar o pipeline, é essencial concluir a etapa de validação dos resultados. Porém, infelizmente, essa fase não foi concluída dentro do prazo estabelecido. Essa validação envolveria a pontuação manual das imagens e a marcação manual dos núcleos celulares.

4.0.3 Limitações

A validação por especialistas é crucial para garantir a precisão dos resultados, visando evitar falsos positivos. As imagens disponíveis nas classes continham uma quantidade significativa de informações que, à primeira vista, poderiam ser interpretadas como células. No entanto, em uma análise mais aprofundada para o cálculo dos índices, essas informações podem não ter influência na contagem real. Essa complexidade se tornou uma considerável barreira no avanço do trabalho, limitando a capacidade de ajustes nos parâmetros propostos e comprometendo a validação dos resultados.

Outro desafio a ser destacado é a ausência de máscaras e contagens das células no conjunto de dados, incluindo a falta de informações sobre a quantidade de células presentes em

cada amostra, esses fatores, em conjunto, exigem abordagens cautelosas e estratégias específicas para garantir a robustez e confiabilidade do processo de segmentação e contagem de células.

5 CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS

Em conclusão, o desenvolvimento deste trabalho expôs desafios cruciais que impactaram diretamente a qualidade e interpretação dos resultados obtidos. A carência de marcações do conjunto de dados representaram obstáculos significativos, sendo a ausência de marcações particularmente prejudicial para a análise dos resultados, impossibilitando a avaliação adequada das quantidades de células em cada amostra do conjunto de dados.

Diante dessas adversidades, propomos como trabalho futuro a documentação abrangente do conjunto de dados e técnicas como as de *data augmentation*, visando não apenas facilitar investigações relacionadas ao câncer de mama, mas também fortalecer a base de dados para pesquisas futuras. Essa iniciativa não apenas beneficia a comunidade científica ao oferecer dados mais acessíveis, mas também contribui para a confiabilidade e reprodutibilidade dos estudos na área.

Ainda, para uma conclusão geral mais abrangente, reconhecemos a importância não apenas da segmentação, mas também da contagem das células. A contagem é crucial, pois com ela é possível definir o PS utilizado para o diagnóstico, um aspecto que foi implementado neste trabalho. Entretanto, sem as máscaras nas imagens e o valor total de células em cada amostra presente no conjunto de dados, torna-se inviável avaliar de maneira adequada a contagem das células. Recomenda-se, portanto, como extensão natural deste estudo, a implementação de técnicas de contagem e a obtenção das máscaras e contagens para aprimorar a eficácia do diagnóstico por meio do IS.

Em última análise, este trabalho não apenas identificou desafios, mas também delineou oportunidades significativas para aprimorar a pesquisa na área de análise de imagens para câncer de mama. A superação desses desafios, aliada às sugestões apresentadas, não só eleva a qualidade intrínseca deste estudo, mas também contribui para o avanço contínuo da pesquisa em um campo vital para a saúde pública.

REFERÊNCIAS

Arthur, D. e Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. Em *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics.

Azevedo Marques, P. M. d. (2001). Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, 34(5):285–293.

Beucher, S. e Meyer, F. (1993). The morphological approach to segmentation: The watershed transformation. Em *Mathematical Morphology in Image Processing*. M. Dekker, New York.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J. R., Golland, P., Sabatini, D. M. e Eliceiri, K. W. (2006). Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(R100).

Diagnósticos do Brasil (2021). Lâmina imuno-histoquímica. `https://www.diagnosticosdobrasil.com.br/material-tecnico/lamina-imuno-histoquimica`. Acessado em: 21 nov. 2023.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*, 31(4-5):198–211. Epub 2007 Mar 8. PMID: 17349778; PMCID: PMC1955762.

Gonzalez, R. C. e Woods, R. E. (2007). *Digital Image Processing*. Pearson Education, Inc.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. e Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

IARC (2023). Cancer today. `https://gco.iarc.fr/today`. Acessado em 22/10/2023.

INCA (2020). Inca estima 704 mil casos de câncer por ano no brasil até 2025. `https://www.gov.br/inca/pt-br/assuntos/noticias/2022/inca-estima-704-mil-casos-de-cancer-por-ano-no-brasil-ate-2025`. Acessado em 22/10/2023.

Kleinberg, J. e Tardos, E. (2005). *Algorithm Design*. Pearson Education, Inc.

Mouelhi, A., Rmili, H., Ali, J. B., Sayadi, M., Doghri, R. e Mrad, K. (2018). Fast unsupervised nuclear segmentation and classification scheme for automatic allred cancer scoring in immunohistochemical breast tissue images. *Computer Methods and Programs in Biomedicine*, 165:37–51.

Mouelhi, A., Sayadi, M., Fnaiech, F., Mrad, K. e Romdhane, K. (2013). A new automatic image analysis method for assessing estrogen receptors' status in breast tissue specimens. *ELSEVIER*, 32(4):? invited paper.

Nishad, P. (2013). Various colour spaces and colour space conversion. *Journal of Global Research in Computer Science*, 4(1):44–48.

Oncoguia (2023). Câncer de mama: Receptor de hormônio. `http://www.oncoguia.org.br/conteudo/cancer-de-mama-receptor-de-hormonio/10879/264/`. Acessado em 22/10/2023.

OpenCV (2023). Morphological transformations. `https://docs.opencv.org/3.4/d9/d61/tutorial_py_morphological_ops.html`. Acessado em 22/10/2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. e Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Qureshi, A. e Pervez, S. (2010). Allred scoring for er reporting and it's impact in clearly distinguishing er negative from er positive breast cancers. *Journal Pakistan Medical Association*, 60(5):350–353.

ROGALSKY, J. E. (2021). Semi-automatic er and pr scoring in immunohistochemistry h-dab breast cancer images. Dissertação de Mestrado, PósGraduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná., Curitiba - PR.

Schneider, C. A., Rasband, W. S. e Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9:671–675.

Silva, B. S. d. (2015). Detecção automática de células via técnicas de morfologia matemática e processamento digital de imagens. `http://www.repositorio.poli.ufrj.br/monografias/monopoli10013465.pdf`. Acessado em 22/10/2023.

Teixeira, L. e Araújo, L. (2020). Câncer de mama no brasil: medicina e saúde pública no século xx. `https://www.scielo.br/j/sausoc/a/dtTQhvkw8hzw9mSRYTQCT9v/?lang=pt`. Acessado em 22/10/2023.

Tiong, K. H., Chang, J. K., Pathmanathan, D., Fadlullah, M. Z. H., Yee, P. S., Liew, C. S., Rahman, Z. A. A., Beh, K. L. e Cheong, S. C. (2018). Quickcount: a novel automated software for rapid cell detection and quantification. *BioTechniques*, 65(06):322–330.

Tsesmelis, T. (2023). Image segmentation with distance transform and watershed algorithm. `https://docs.opencv.org/4.x/d2/dbd/tutorial_distance_transform.html`. Acessado em 22/10/2023.

Van Rossum, G. e Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.