



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO (CTC)
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE AUTOMAÇÃO E
SISTEMAS

Luiz Antonio Buschetto Macarini

**Utilização de Redes Neurais Convolucionais na detecção, segmentação e
classificação de núcleos celulares para identificação de aneuploidia**

Florianópolis
2021

Luiz Antonio Buschetto Macarini

Utilização de Redes Neurais Convolucionais na detecção, segmentação e classificação de núcleos celulares para identificação de aneuploidia

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Automação e Sistemas da Universidade Federal de Santa Catarina para a obtenção do título de mestre em Engenharia de Automação e Sistemas.

Orientador: Prof. Marcelo Ricardo Stemmer, Dr.

Coorientador: Prof. Aldo von Wangenheim, Dr. rer. nat.

Florianópolis

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Macarini, Luiz Antonio Buschetto

Utilização de Redes Neurais Convolucionais na detecção,
segmentação e classificação de núcleos celulares para
identificação de aneuploidia / Luiz Antonio Buschetto
Macarini ; orientador, Marcelo Ricardo Stemmer,
coorientador, Aldo von Wangenheim, 2021.

77 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia de Automação e Sistemas, Florianópolis, 2021.

Inclui referências.

1. Engenharia de Automação e Sistemas. 2. Inteligência
Artificial. 3. Aprendizado de Máquina. 4. Deep Learning.
5. Processamento de Imagens. I. Stemmer, Marcelo Ricardo.
II. von Wangenheim, Aldo. III. Universidade Federal de
Santa Catarina. Programa de Pós-Graduação em Engenharia de
Automação e Sistemas. IV. Título.

Luiz Antonio Buschetto Macarini

Utilização de Redes Neurais Convolucionais na detecção, segmentação e classificação de núcleos celulares para identificação de aneuploidia

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Jomi Fred Hübner, Dr.
Universidade Federal de Santa Catarina

Prof. Eric Aislan Antonelo, Dr.
Universidade Federal de Santa Catarina

Prof.(a) Maria Inês Meurer, Dr(a).
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia de Automação e Sistemas.

Coordenação do Programa de
Pós-Graduação

Prof. Marcelo Ricardo Stemmer, Dr.
Orientador

Florianópolis, 2021.

Este trabalho é dedicado à minha família, amigos e à todos que me ajudaram nesta trajetória.

AGRADECIMENTOS

Durante este período, muitas pessoas contribuíram para a realização deste trabalho, as quais sou profundamente grato.

À minha esposa, Thayane, pelo seu apoio incondicional, por compartilhar comigo a alegria das conquistas e me dar suporte nos momentos difíceis.

Aos meus pais, Joacir e Arlete, por sempre me incentivarem a estudar e por me ensinarem o valor da educação.

Aos meus colegas de universidade e, agora, amigos: Felipe Vieira, Luan Casagrande, Roger Florzino e Yuri Crotti.

Aos meus amigos de infância Eddie, Fernando, Luiz Filipe, Yuri, por estarem comigo desde sempre.

Aos meus colegas de laboratório, Allan, André e João Gustavo pela ajuda durante os anos juntos no laboratório.

Deixo também um agradecimento especial ao professor Antonio Carlos Sobieranski por toda a ajuda dispensada.

Aos meus orientadores, professores Marcelo Ricardo Stemmer e Aldo von Wangenheim, pela dedicação durante estes anos de muito trabalho.

Gostaria de agradecer também aos membros da banca pela disponibilidade e pelas valiosas contribuições à este trabalho.

Ao pessoal do INCoD e do LAPIX, em especial à Nathalie, por fornecerem o necessário para o avanço desta pesquisa.

À Ane Francyne, por revisar todo o conjunto de dados.

Aos professores Alexandre Onofre e Fabiana Onofre pela colaboração.

Por fim, agradeço a todas as pessoas que colaboraram, direta ou indiretamente, na elaboração deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

O câncer de colo de útero é o quarto tipo de câncer mais comum entre as mulheres. No Brasil, esta doença foi a causa da morte de 6596 mulheres em 2019. Uma das maneiras de realizar o diagnóstico é através do teste de Papanicolau. A biópsia ainda é a maneira mais comum para a detecção precoce, apesar de ser um método invasivo. Neste sentido, pesquisadores vêm buscando maneiras não-invasivas de auxiliar no diagnóstico deste tipo de doença, e os estudos na área de Citologia tem se mostrado promissores. Nos dias atuais, utiliza-se a Citometria de DNA como principal abordagem para a detecção da quantidade anormal de DNA nas células, processo conhecido como Aneuploidia. Este pode ser um indicativo de lesões pré-cancerosas. Porém, ainda é necessário que esta análise celular seja feita por profissionais treinados. O processo pode ser cansativo e está sujeito a erros. No entanto, esta tarefa é passível de automação, auxiliando na diminuição da subjetividade do processo, dos possíveis erros e do tempo necessário para realizar a análise. Nos últimos anos, as técnicas de *Deep Learning* têm obtido sucesso em resolver problemas que resistiram às melhores tentativas da comunidade de Inteligência Artificial. Atualmente, as Redes Neurais Convolucionais vêm sendo utilizadas em diversas aplicações, nas quais podemos incluir também a área de saúde. Neste contexto, este trabalho tem como objetivo apresentar um *pipeline* completo para automação da classificação de núcleos celulares preparados com o corante de Feulgen, visando encontrar aneuploidia de maneira não-invasiva. Será apresentada uma comparação de diferentes abordagens (segmentação semântica, detecção de objetos e classificação de imagens) com onze arquiteturas de redes neurais diferentes. Os dados utilizados são provenientes de pacientes do Hospital Universitário da UFSC, onde foi construído um conjunto de dados anotados e avaliados por profissionais. Os resultados obtidos mostram que, no geral, as redes neurais apresentam bom desempenho em relação a detecção/segmentação (localização) dos núcleos, porém a diferenciação entre as classes apresentou métricas ruins. Dentre as abordagens comparadas neste trabalho, a classificação de imagens foi a que apresentou os melhores resultados. Concluiu-se que estes algoritmos ainda não estão prontos para serem implementados na rotina clínica. No entanto, este trabalho aponta um possível caminho a ser seguido para investigação de uma abordagem objetiva e não-invasiva para detecção precoce de aneuploidia.

Palavras-chave: Feulgen. Citologia. Deep Learning. Detecção. Segmentação. Classificação.

ABSTRACT

Cervical cancer is the fourth most recurrent type of cancer among women. In Brazil, this disease was the cause of death of 6596 women in 2019. One of the ways to make the diagnosis is through the Pap test. The biopsy is still the most recurrent way of early detection, despite being an invasive method. Researchers have been looking for non-invasive methods to help in the diagnosis of this type of disease, and studies in the Cytology field have shown promising results. Nowadays, DNA cytometry is used as the main approach for detecting the abnormal amount of DNA in cells, a process known as Aneuploidy. This can be an indicative of precancerous lesions. However, this cell analysis still needs to be performed by trained professionals. The process can be exhaustive and error-prone. However, this task can be automated, helping on reducing the subjectivity of the process, possible errors, and the time needed to finish the analysis. In the last few years, Deep Learning techniques have been successful in solving problems that have resisted the best attempts of the Artificial Intelligence community. Currently, Convolutional Neural Networks are being used in several applications, in which we can also include the medical area. In this context, this work aims to present a complete pipeline for the automation of the classification of cell nuclei prepared with Feulgen stain, aiming to find aneuploidy in a non-invasive way. A comparison of different approaches (semantic segmentation, object detection, and image classification) with eleven different neural network architectures will be presented. The data used in this work was gathered from patients at UFSC's University Hospital, where a dataset annotated and evaluated by professionals was built. The obtained results show that, in general, neural networks present a good performance on nuclei detection/segmentation (localization), but the differentiation between classes presented poor metrics. Among the approaches compared in this work, image classification presented the best results. It was concluded that these algorithms are not yet ready to be implemented in a clinical routine. However, this work points to a possible way for the investigation of an objective and non-invasive approach for the early detection of aneuploidy.

Keywords: Feulgen. Cytology. Deep Learning. Detection. Segmentation. Classification.

LISTA DE FIGURAS

Figura 1 – Procedimento atual para atendimento de pacientes com suspeita de câncer.	22
Figura 2 – Possível cenário com a adoção da solução proposta.	23
Figura 3 – Rede Neural Artificial - <i>Multilayer Perceptron</i>	29
Figura 4 – Arquitetura típica de uma Rede Neural Convolucional.	31
Figura 5 – Visão geral da metodologia adotada.	34
Figura 6 – Demonstração dos diferentes tipos de <i>ground-truths</i> e suas respectivas abordagens.	37
Figura 7 – Distribuição de amostras (por classe) dentro do conjunto de dados.	38
Figura 8 – Amostra de cada classe presente no conjunto de dados.	38
Figura 9 – Representação esquemática da estratégia adotada no processo de treinamento da rede (HOWARD, 2019).	40
Figura 10 – Métricas médias para as redes neurais de segmentação semântica.	45
Figura 11 – Resultado das Segmentações para a U-Net + ResNet18. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	48
Figura 12 – Resultado das Segmentações para a U-Net + ResNet34. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	49
Figura 13 – Resultado das Segmentações para a <i>Weighted</i> U-Net + ResNet18. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	50
Figura 14 – Resultado das Segmentações para a <i>Weighted</i> U-Net + ResNet34. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	51
Figura 15 – Métricas médias para as redes neurais de detecção de objetos.	52
Figura 16 – Falsos Negativos (FN), Falsos Positivos (FP) e Verdadeiros Positivos (TP) - Faster R-CNN + ResNet50.	53
Figura 17 – Resultado das detecções para a Faster R-CNN + ResNet50. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	56
Figura 18 – Resultado das detecções para a Faster R-CNN + ResNet101. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	57
Figura 19 – Resultado das detecções para a RetinaNet + ResNet50. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	58

Figura 20 – Resultado das detecções para a RetinaNet + ResNet101. Linha superior - esquerda: imagem de entrada; Linha superior - direita: <i>ground-truth</i> ; Linha inferior: predição da rede neural.	59
Figura 21 – Métricas obtidas com as redes neurais de classificação de imagem.	60
Figura 22 – Exemplos de resultados para classificação de imagens com a ResNet34. Na legenda, a classe na linha superior representa o <i>ground-truth</i> e na inferior a predição da rede neural.	62
Figura 23 – Exemplos de resultados para classificação de imagens com a ResNet50. Na legenda, a classe na linha superior representa o <i>ground-truth</i> e na inferior a predição da rede neural.	63
Figura 24 – Exemplos de resultados para classificação de imagens com a ResNet101. Na legenda, a classe na linha superior representa o <i>ground-truth</i> e na inferior a predição da rede neural.	64

LISTA DE TABELAS

Tabela 1 – Critérios de Inclusão e Exclusão da SLR	18
Tabela 2 – Definições de Busca - IEEEExplore	18
Tabela 3 – Definições de Busca - Science Direct	19
Tabela 4 – Definições de Busca - Springer Link	19
Tabela 5 – Definições de Busca - ACM Digital Library	19
Tabela 6 – Definições de Busca - Wiley Online	20
Tabela 7 – Definições de Busca - PubMed	20
Tabela 8 – Equações das métricas utilizadas no presente trabalho.	43
Tabela 9 – <i>Learning rates</i> utilizados nos treinamentos das redes neurais de segmentação semântica.	44
Tabela 10 – Pesos utilizados (para cada classe) durante o treinamento das <i>Weighted U-Nets</i>	45
Tabela 11 – Métricas - Segmentação Semântica com U-Net + ResNet18.	46
Tabela 12 – Métricas - Segmentação Semântica com U-Net + ResNet34.	46
Tabela 13 – Métricas - Segmentação Semântica com <i>Weighted U-Net</i> + ResNet18.	47
Tabela 14 – Métricas - Segmentação Semântica com <i>Weighted U-Net</i> + ResNet34.	47
Tabela 15 – Métricas - Detecção de Objetos com Faster R-CNN + ResNet50	54
Tabela 16 – Métricas - Detecção de Objetos com Faster R-CNN + ResNet101	54
Tabela 17 – Métricas - Detecção de Objetos com RetinaNet + ResNet50	54
Tabela 18 – Métricas - Detecção de Objetos com RetinaNet + ResNet101	55
Tabela 19 – <i>Learning rates</i> para as redes neurais de classificação de imagens.	59
Tabela 20 – Métricas - Classificação de Imagens - ResNet34	61
Tabela 21 – Métricas - Classificação de Imagens - ResNet50	61
Tabela 22 – Métricas - Classificação de Imagens - ResNet101	61

LISTA DE ABREVIATURAS E SIGLAS

AgNOR	<i>Argyropholic Nucleolar Organizer Region</i>
AM	Aprendizado de Máquina
CEPSH	Comitê de Ética em Pesquisa da UFSC
CNN	<i>Convolutional Neural Network</i>
DNA	Ácido Desoxirribonucleico
EBSERH	Empresa Brasileira de Serviços Hospitalares
FN	<i>False Negative</i>
FP	<i>False Positive</i>
GPU	<i>Graphics Processing Units</i>
HPV	<i>Human Papillomavirus</i>
HU-UFSC	Hospital Universitário Professor Polydoro Ernani de São Thiago da Universidade Federal de Santa Catarina
IA	Inteligência Artificial
IoU	<i>Intersection Over Union</i>
mAP	<i>Mean Average Precision</i>
MLP	<i>Multilayer Perceptron</i>
PDI	Processamento Digital de Imagens
R-CNN	<i>Region-Based Convolutional Neural Networks</i>
RGB	<i>Red, Green, Blue</i>
RNAs	Redes Neurais Artificiais
SLR	Revisão Sistemática da Literatura
SSD	<i>Single Shot Detector</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
YOLO	<i>You Only Look Once</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	16
1.1.1	Objetivo Geral	16
1.1.2	Objetivos Específicos	16
1.2	REVISÃO DA LITERATURA	17
1.3	CONTRIBUIÇÕES	21
1.4	LIMITAÇÕES	22
1.5	CENÁRIO DE APLICAÇÃO	22
1.6	ESTRUTURA DO DOCUMENTO	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	IMAGEM DIGITAL	25
2.1.1	Visão Computacional	25
2.1.2	Processamento Digital de Imagens	26
2.2	INTELIGÊNCIA ARTIFICIAL	26
2.2.1	Aprendizado de Máquina	27
2.2.1.1	Redes Neurais Artificiais	28
2.2.1.2	<i>Multilayer Perceptron</i>	28
2.3	<i>DEEP LEARNING</i>	29
2.3.1	Redes Neurais Convolucionais	30
2.3.1.1	Abordagens	31
2.4	CITOLOGIA	33
2.4.1	Aplicação da Visão Computacional na área de Citologia	33
3	MATERIAIS E MÉTODOS	34
3.1	GERAÇÃO DO CONJUNTO DE DADOS	34
3.2	TREINAMENTO DAS REDES NEURAIS	39
3.2.1	Segmentação Semântica	39
3.2.2	Deteccção de Objetos	41
3.2.3	Classificação de Imagens	41
3.3	CRITÉRIOS DE AVALIAÇÃO/COMPARAÇÃO	42
4	RESULTADOS	44
4.1	SEGMENTAÇÃO SEMÂNTICA	44
4.1.1	Resultados por classe	46
4.2	DETECCÃO DE OBJETOS	51
4.2.1	Resultados por classe	53
4.3	CLASSIFICAÇÃO DE IMAGENS	59
4.3.1	Resultados por classe	60
4.4	ARTIGOS PUBLICADOS	64

5	CONCLUSÃO	66
5.1	TRABALHOS FUTUROS	67
	REFERÊNCIAS	69

1 INTRODUÇÃO

O câncer de colo de útero é o quarto tipo de câncer mais comum entre as mulheres no mundo. A cada ano, mais de 570 mil pacientes são diagnosticadas com essa doença. Em 2019, o câncer de colo de útero foi a causa da morte de 6596 mulheres no Brasil (INCA, 2021). Na maioria dos casos, este câncer é causado devido a infecção persistente pelo *Human Papillomavirus* (HPV) (COHEN *et al.*, 2019). Este tipo de câncer ainda é detectado, em muitos dos casos, em estágios avançados devido a triagem inadequada. Desta maneira, as opções de tratamento comuns são inexistentes ou inacessíveis (COGLIANO *et al.*, 2005; MUÑOZ *et al.*, 2006). O tratamento depende do estágio da doença no momento do diagnóstico e dos recursos disponíveis. Para mulheres com infecção de recorrência ou metástase, o prognóstico ainda é ruim (COHEN *et al.*, 2019). Este tipo de câncer é curável se detectado no estágio inicial da doença e tratado de maneira adequada. Porém, a detecção precoce ainda é um grande desafio (SONG *et al.*, 2016).

Uma das maneiras de realizar o diagnóstico é através do teste de Papanicolau (LOPES; RIBEIRO, 2019). Quando os resultados apontam que há evidências de lesão intraepitelial escamosa de alto grau, células escamosas atípicas, células glandulares atípicas ou adenocarcinoma, a paciente é encaminhada para exame colposcópico, onde posteriormente é realizada a biópsia para a confirmação da lesão (INCA, 2016). O HPV é o responsável pelas alterações observadas em exames citológicos, tendo ligação com o câncer cervical (GAO; SMITH, D. I., 2016). A infecção pelo HPV é necessária para que haja transformação tecidual. Porém, ela não é fator suficiente para que haja progressão das alterações para um caso de câncer (BURD, 2003). O grande desafio está em detectar quais lesões vão progredir (BÖCKING, Alfred; NGUYEN, 2004). As displasias cervicais são lesões heterogêneas, particularmente em relação ao seu comportamento clínico. Nenhum método citológico ou histológico é capaz de prever se estas células displásicas (lesões pré-câncer) vão evoluir a ponto de se tornar um carcinoma (JONES; NOVIS, 1996; LONKY *et al.*, 1999).

As anormalidades no conteúdo do Ácido Desoxirribonucleico (DNA) das células vêm sendo associadas a tumorigênese (DANIELSEN *et al.*, 2016) e a presença destes tipos de células tem sido utilizada como marcador de neoplasia (BÖCKING, Alfred; NGUYEN, 2004). A Aneuploidia é o termo técnico utilizado para denotar a quantidade anormal de DNA na célula (DANIELSEN *et al.*, 2016). Esse fenômeno pode ser quantificado e detectado utilizando Citometria de DNA (GROTE *et al.*, 2004). Diversos estudos demonstraram que a aneuploidia indica ou um carcinoma invasivo ou, prospectivamente, o desenvolvimento neoplástico em displasias cervicais (BÖCKING *et al.*, 1986; CHATELAIN *et al.*, 1989; KASHYAP *et al.*, 1990; BOLLMANN *et al.*, 2001). Em outras palavras, a aneuploidia é um marcador de prognóstico independente para

alguns tipos de câncer. Ela pode indicar a tendência a progressão das lesões para um possível caso de câncer (DANIELSEN *et al.*, 2016).

Atualmente a Citometria de DNA é a principal abordagem não-invasiva utilizada para realizar a análise celular visando a detecção da aneuploidia. Esse processo envolve corar o DNA no núcleo celular de uma amostra de tecido. É geralmente seguido pelo processo de citometria de fluxo ou análise microscópica. A partir das amostras coletadas nestes exames citológicos pode-se realizar também uma análise de histograma. Este método é utilizado para descrever a frequência populacional de um núcleo contendo diferentes quantidades de DNA no contexto do ciclo celular de células diploides (DANIELSEN *et al.*, 2016). Parte do processo de análise ainda é feita manualmente por profissionais treinados (GROTE *et al.*, 2004). Sendo assim, esse processo de diagnóstico utilizando amostras citológicas consome muito tempo e está sujeito a erros, já que há variações na percepção humana. Para evitar estes problemas, análises por meio de métodos computacionais podem ser realizadas visando detectar alterações no *status* celular (MEHROTRA *et al.*, 2011).

Para a preparação das lâminas, diferentes corantes podem ser utilizados. O mais comum é o Papanicolau (*Pap Smear*). Outros exemplos são a Coloração Hematoxilina-Eosina (H&E) e o corante de Feulgen (MATIAS *et al.*, 2021). A reação de Feulgen é uma reação cito-histoquímica muito utilizada, onde a concentração de DNA presente na célula é proporcional a intensidade do corante. É mais utilizada para quantificação de DNA nos núcleos celulares através de citometria de imagem, com o intuito de fazer a análise de ploidia (CHIECO; DERENZINI, 1999).

Devido aos avanços em software e hardware nos últimos anos, o Diagnóstico Auxiliado por Computador (CAD, do inglês *Computer-aided Diagnosis*) vem crescendo em ritmo acelerado. Essa evolução tem como principais motivações a diminuição de erros e a melhoria na interpretação de imagens (LU *et al.*, 2017). As técnicas de *Deep Learning* estão fazendo grandes avanços na solução de problemas que resistiram às melhores tentativas da comunidade de inteligência artificial por muitos anos. Estes métodos mostraram uma excelente capacidade para descobrir estruturas intrínsecas, conseguindo processar dados em sua forma bruta. Além disso, trouxeram grandes avanços nas áreas de processamento de imagens, vídeos, fala e áudio, quebrando recordes por conseguir melhorar, por uma grande margem, os resultados obtidos com os métodos clássicos de aprendizado de máquina (LECUN *et al.*, 2015).

A partir dos recentes avanços trazidos pelos métodos de *deep learning*, problemas envolvendo detecção de objetos, classificação de imagem e segmentação semântica que não possuíam solução estão sendo resolvidos. Isto também acontece na área de citologia quantitativa (AMORIM *et al.*, 2020). Problemas que resistiram por anos às tentativas dos pesquisadores da área de aprendizado de máquina e inteligência artificial foram resolvidos utilizando Redes Neurais Convolucionais, cujo nome vem

do inglês, *Convolutional Neural Network* (CNN). Os resultados alcançados utilizando CNNs superam os obtidos através de métodos clássicos por uma diferença significativa. As redes neurais convolucionais possuem algumas vantagens sobre os métodos clássicos de aprendizado de máquina. Um exemplo é o fato de serem capazes de processar os dados em sua forma bruta. Ou seja, não é necessário desenvolver um algoritmo para a extração de características das imagens. Isso significa que as CNNs são capazes de processar uma imagem sem a necessidade de aplicação de qualquer técnica de pré-processamento. As redes neurais convolucionais podem ser utilizadas como um extrator geral de características (LECUN *et al.*, 2015).

Neste contexto, este trabalho tem como objetivo apresentar um *pipeline* completo para automação do processo de segmentação, detecção e/ou classificação de núcleos em lâminas de exames citológicos preparadas com o corante de Feulgen, visando encontrar aneuploidias. As imagens foram extraídas de lâminas escaneadas em laboratório, de onde foram gerados os campos (sub-imagens). Essas imagens passaram pelo processo de marcação (*labeling*) e criou-se um conjunto de dados (*dataset*). Utilizando este conjunto de imagens, foram treinadas e testadas diferentes modelos de redes neurais convolucionais utilizando as abordagens de segmentação semântica, detecção de objetos e classificação de imagens. O objetivo é verificar qual combinação apresenta o melhor resultado em relação a classificação correta dos núcleos visando encontrar aneuploidia, sendo assim possível realizar o prognóstico de lesões com potencial de malignização. Os dados utilizados são provenientes de pacientes do Hospital Universitário Professor Polydoro Ernani de São Thiago da Universidade Federal de Santa Catarina (HU-UFSC), que faz parte da Empresa Brasileira de Serviços Hospitalares (EBSERH).

1.1 OBJETIVOS

Nesta subseção serão descritos o objetivo geral e os objetivos específicos.

1.1.1 Objetivo Geral

Apresentar um *pipeline* completo para a automação do processo de segmentação, detecção e/ou classificação de núcleos em lâminas preparadas com corante de Feulgen, utilizando CNNs, a fim de encontrar aneuploidia em exames de citometria de DNA.

1.1.2 Objetivos Específicos

Considerando o desenvolvimento do trabalho e o objetivo geral apresentado, destacam-se os seguintes objetivos específicos:

- Criar um conjunto de dados (imagens) para treinamento dos modelos de redes neurais;
- Testar/comparar diferentes modelos de redes neurais visando encontrar aquela que traz o melhor resultado em relação a segmentação semântica, detecção de objetos e/ou classificação de imagens;
- Analisar a viabilidade da utilização de redes neurais convolucionais na análise de núcleos celulares visando encontrar aneuploidia.

1.2 REVISÃO DA LITERATURA

A Revisão Sistemática da Literatura (SLR) (do inglês, *Systematic Literature Review*) tem como objetivo reunir trabalhos relevantes sobre uma questão de pesquisa formulada. Desta maneira é possível realizar uma revisão crítica e abrangente da literatura. Procura-se alcançar, de maneira sistemática, o maior número de trabalhos que sejam relevantes ao tema pesquisado. A pesquisa nas bases de dados deve ser feita de maneira sistemática, explicando-se cada etapa, possibilitando que outro pesquisador consiga reproduzi-la e obter os mesmos resultados (KITCHENHAM, 2004).

A partir do resultado da SLR, selecionou-se os algoritmos e abordagens utilizadas neste trabalho. Após a realização da revisão percebeu-se a inexistência de uma base de dados com as características desejadas. Desta maneira, optou-se por construir uma, com o auxílio de especialistas, utilizando exames realizados no Hospital Universitário da UFSC.

A SLR foi conduzida de acordo com os protocolos elaborados por Kitchenham (2004). A revisão foi realizada em Setembro de 2020 (AMORIM *et al.*, 2020) e publicada em 2021 (MATIAS *et al.*, 2021). Teve-se como objetivo identificar os trabalhos que possuem abordagens computacionais ou baseadas em inteligência artificial para auxiliar na análise quantitativa ou no diagnóstico de câncer em exames citológicos. Estudos publicados entre 2016 e 2020 foram analisados. Procurou-se por abordagens de segmentação, detecção e classificação de imagens utilizando métodos de visão computacional em imagens provenientes de exames citológicos. A busca resultou em 431 artigos, dos quais 13 são duplicados. Restaram 157 artigos após a aplicação dos critérios de inclusão e exclusão, que são apresentados na Tabela 1.

Tabela 1 – Critérios de Inclusão e Exclusão da SLR

Critérios de Inclusão
<ul style="list-style-type: none"> - Artigos que utilizam visão computacional para segmentar, detectar e/ou classificar células em imagens de exames de citologia; - Estudos realizados em humanos;
Critérios de Exclusão
<ul style="list-style-type: none"> - Artigos que não utilizam técnicas de visão computacional para analisar as imagens provenientes de exames de citologia; - Estudos que foram realizados utilizando células não-humanas; <ul style="list-style-type: none"> - Artigos que focam no melhoramento das imagens; - <i>Short papers</i>, resumos ou resumos expandidos; - Artigos de revisão; - Artigos que não estavam completos na base de dados; - Artigos que não foram escritos em inglês.

A busca foi realizada nas bases IEEEExplore, Science Direct, Springer Link, ACM Digital Library, Wiley Online Library e PubMed. A Tabela 2 apresenta as definições de busca para a IEEEExplore.

Tabela 2 – Definições de Busca - IEEEExplore

Base	IEEEExplore
Data	2016 a 2020
Filtros	Qualquer Formato
Quantidade de Trabalhos	89
Quantidade de Trabalhos Selecionados	68
String de Busca	<pre> ((((("Abstract":cytology OR "Abstract":cytopathology OR "Abstract":Papanicolaou OR "Abstract":"Pap smear") AND ("Abstract":image OR "Abstract":vision) AND ("Abstract":classification OR "Abstract":segmentation OR "Abstract":detection)) OR (("Document Title":cytology OR "Document Title":cytopathology OR "Document Title":Papanicolaou OR "Document Title":"Pap smear") AND ("Document Title":image OR "Document Title":vision) AND ("Document Title":classification OR "Document Title":segmentation OR "Document Title":detection)) OR (("Author Keywords":cytology OR "Author Keywords":cytopathology OR "Author Keywords":Papanicolaou OR "Author Keywords":"Pap smear") AND ("Author Keywords":image OR "Author Keywords":vision) AND ("Author Keywords":classification OR "Author Keywords":segmentation OR "Author Keywords":detection)))))) </pre>

Na Tabela 3 são mostradas as definições de busca para a base Science Direct.

Tabela 3 – Definições de Busca - Science Direct

Base	Science Direct
Data	2016 a 2020
Filtros	Qualquer formato
Quantidade de Trabalhos	135
Quantidade de Trabalhos Selecionados	30
String de busca	((cytology OR cytopathology OR Papanicolaou OR "Pap smear") AND (image OR vision) AND (classification OR segmentation OR detection))

Na Tabela 4, as definições para a busca na Springer Link são apresentadas.

Tabela 4 – Definições de Busca - Springer Link

Base	Springer Link
Data	2016 a 2020
Filtros	Qualquer formato; Disciplina: <i>Computer Science</i>
Quantidade de Trabalhos	115
Quantidade de Trabalhos Selecionados	28
String de busca	("cytology" OR "cytopathology" OR "Papanicolaou" OR "Pap smear") AND ("image" OR "vision") AND ("classification" OR "segmentation" OR "detection")

As definições de busca para a ACM Digital Library são apresentadas na Tabela 5.

Tabela 5 – Definições de Busca - ACM Digital Library

Base	ACM Digital Library
Data	2016 a 2020
Filtros	Qualquer formato
Quantidade de Trabalhos	15
Quantidade de Trabalhos Selecionados	10
String de busca	[[[Abstract: cytology] OR [Abstract: cytopathology] OR [Abstract: papanicolaou] OR [Abstract: "pap smear"]] AND [[Abstract: image] OR [Abstract: vision]] AND [[Abstract: classification] OR [Abstract: segmentation] OR [Abstract: detection]]] OR [[[Publication Title: cytology] OR [Publication Title: cytopathology] OR [Publication Title: papanicolaou] OR [Publication Title: "pap smear"]] AND [[Publication Title: image] OR [Publication Title: vision]] AND [[Publication Title: classification] OR [Publication Title: segmentation] OR [Publication Title: detection]]] OR [[[Keywords: cytology] OR [Keywords: cytopathology] OR [Keywords: papanicolaou] OR [Keywords: "pap smear"]] AND [[Keywords: image] OR [Keywords: vision]] AND [[Keywords: classification] OR [Keywords: segmentation] OR [Keywords: detection]]] AND [Publication Date: (01/01/2016 TO 12/31/2020)]

A Tabela 6 apresenta as definições de busca para a Wiley Online.

Tabela 6 – Definições de Busca - Wiley Online

Base	Wiley Online Library
Data	2016 a 2020
Filtros	Qualquer formato
Quantidade de Trabalhos	45
Quantidade de Trabalhos Selecionados	7
String de Busca	[[Keywords: 'cytology'] OR [Keywords: 'cytopathology'] OR [Keywords: 'pap smear'] OR [Keywords: 'papanicolaou'] AND [[All: 'computer vision'] OR [All: 'image analysis']] AND [[All: 'classification'] OR [All: 'detection'] OR [All: 'segmentation']] AND [Earliest: (01/01/2016 TO 03/31/2020)]

As definições de busca para a base Pubmed são mostradas na Tabela 7.

Tabela 7 – Definições de Busca - PubMed

Base	PubMed
Data	2016 a 2020
Filtros	Qualquer Formato
Quantidade de Trabalhos	32
Quantidade de Trabalhos Selecionados	14
String de busca	("Papanicolaou" OR "Pap smear" OR "conventional cytology" OR "liquid-based cytology" OR "brush biopsy" OR "brush cytology") AND ("Convolutional Neural Networks" OR "Computer Vision" OR "Deep Learning" OR "Image Processing") AND ("classification" OR "segmentation" OR "detection")

Durante a revisão, foram identificados dois trabalhos que utilizaram *deep learning* com imagens coradas com Feulgen em exames cervicais. Em (XIANG *et al.*, 2020) os autores utilizaram a YOLOv3 como modelo principal de rede neural e adicionaram um classificador específico, visando melhorar a classificação dos núcleos. Os autores atingiram 89,3% de acurácia, 97,5% de sensibilidade, 67,8% de especificidade e *Mean Average Precision* (mAP) de 60,2. Em um artigo publicado como parte desta dissertação (MACARINI *et al.*, 2021), utilizou-se a U-Net visando resolver um problema de segmentação semântica binária. O objetivo era diferenciar entre núcleo e *background*. Com a utilização deste modelo de rede neural atingiu-se um *Intersection Over Union* (IoU) de 0,78. Apesar de terem sido encontrados dois artigos semelhantes ao presente trabalho, é preciso ter cuidado nas comparações diretas de resultados pois os trabalhos utilizam diferentes métodos, conjuntos de dados e métricas. Foram encontrados mais quatro trabalhos que utilizaram o corante de Feulgen, porém com métodos clássicos ou células provenientes de exames de outras partes do corpo. Em

um deles, os autores utilizaram um software comercial para extrair 18 características morfométricas dos núcleos e os classificaram utilizando métodos de aprendizado de máquina (BÖCKING, A. H. *et al.*, 2019). Em outro trabalho foi utilizado um conjunto de equações parciais com grafos ponderados como metodologia para implementar um método de auxílio ao diagnóstico em citologia (TOUTAIN *et al.*, 2015). Métodos de *deep learning* (Inception V3) foram utilizados para extrair as características e classificar as células. Porém, foram processadas células do pulmão (GONZALEZ, D. *et al.*, 2020). Por fim, outro estudo utilizou métodos de segmentação para encontrar os núcleos e métodos de processamento de imagens para extrair 20 características morfológicas e 8 características de textura. Para a etapa de classificação foram utilizados dois métodos de aprendizado de máquina em cascata (SU *et al.*, 2016).

Baseado na análise destes artigos foi possível determinar que os métodos de *deep learning* passaram a ser mais utilizados que os métodos clássicos de aprendizado de máquina a partir de 2019. Porém, abordagens clássicas (*Support Vector Machine*, Redes Neurais Artificiais, *K-Nearest Neighbor*, *Random Forest*, entre outros) ainda são utilizadas em alguns dos trabalhos recentes. Desta maneira, sugere-se uma tendência onde os pesquisadores estão diminuindo o uso destes métodos clássicos e migrando para as abordagens baseadas em *deep learning*. Falando de técnicas de coloração, o Papanicolau (do inglês, *Pap Smear*) é o corante mais utilizado nos experimentos. Outros exemplos são os corantes de Feulgen e o H&E.

1.3 CONTRIBUIÇÕES

A maior contribuição deste trabalho é um *pipeline* completo para a automação do processo de detecção de aneuploidia utilizando métodos de *deep learning*. Além disso, é apresentada uma comparação extensiva entre arquiteturas, *backbones*, abordagens (segmentação semântica, detecção de objetos e classificação de imagens) na análise morfométrica de amostras coradas com Feulgen. Apesar de já existirem diversos trabalhos onde foram realizadas análises quantitativas baseadas em aprendizado de máquina, alguns datando de 20 anos atrás (KOLLES *et al.*, 1993), a maioria das abordagens recentes utilizando Redes Neurais Convolucionais são focadas no corante de Papanicolau. Outra contribuição deste trabalho é a criação e publicação do *dataset*¹ utilizado nos experimentos, ajudando a mitigar a falta de uma base de dados com corante de Feulgen (AMORIM *et al.*, 2020). Os códigos-fonte desenvolvidos e utilizados neste trabalho também estão disponíveis publicamente², auxiliando na reprodutibilidade dos resultados ou na aplicação destes métodos em outras bases de dados com corante de Feulgen (ou outros corantes).

¹ O conjunto de dados pode ser baixado em: <https://arquivos.ufsc.br/d/7e7ac2f498df4cf9aa7d/>

² O código-fonte pode ser encontrado em: <https://codigos.ufsc.br/lapix/feulgen-segmentation>.

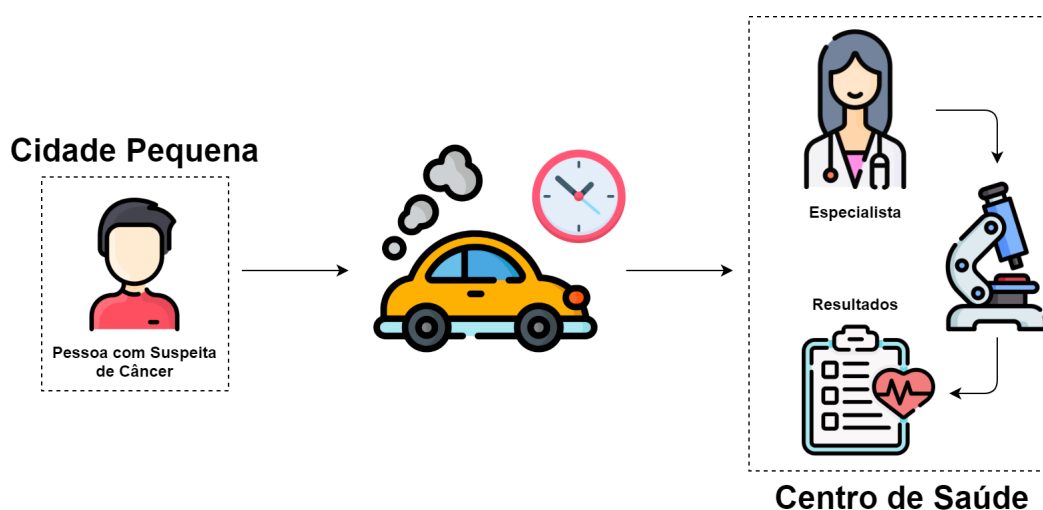
1.4 LIMITAÇÕES

Mesmo que o *dataset* seja composto por 1955 imagens, elas foram geradas a partir de cinco lâminas provenientes de exames de cinco pacientes, o que pode ser uma limitação. A utilização deste número limitado de lâminas se deu pelo fato de que o processo de anotação é extremamente custoso (em relação ao tempo necessário) e deve ser feito por profissionais treinados. No entanto, é necessário deixar claro que este trabalho é resultado da investigação de uma possível abordagem para detecção precoce de câncer. Desta maneira, quanto mais o estudo progredir, mais lâminas podem ser obtidas, refinando os resultados obtidos neste trabalho.

1.5 CENÁRIO DE APLICAÇÃO

Atualmente, são poucos os laboratórios localizados em cidades do interior com a capacidade de realizar exames de média e alta complexidade. Desta maneira, os profissionais que trabalham nestes locais têm pouca ou nenhuma experiência em dar laudos para exames de citologia, por exemplo. Isso acontece principalmente se houver algum tipo de alteração nos resultados. Geralmente os profissionais têm experiência com algumas tarefas da rotina laboratorial, como fixar as lâminas. Porém, elas são enviadas para os grandes centros de saúde para realização das análises e posterior diagnóstico. Desta maneira, a pessoa com suspeita de câncer ainda precisa viajar longas distâncias até o centro de saúde mais próximo. Estes pacientes são então atendidos pelos profissionais, onde realizam-se os exames e faz-se o diagnóstico. A Figura 1 apresenta o fluxo nos dias atuais.

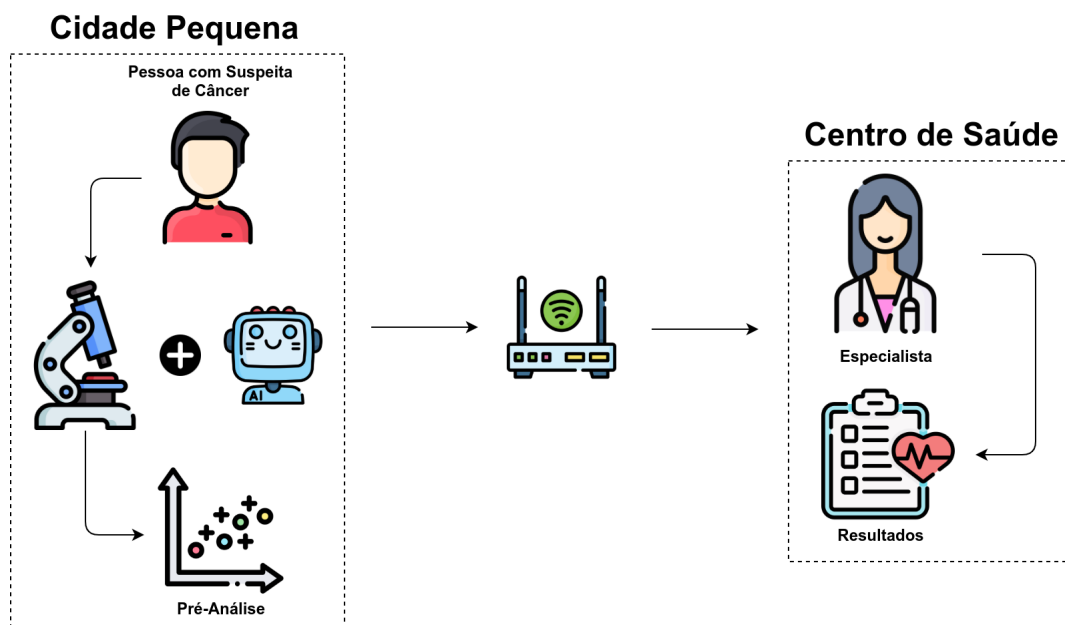
Figura 1 – Procedimento atual para atendimento de pacientes com suspeita de câncer.



Fonte – Fonte: Do autor.

A solução proposta neste trabalho pode tornar possível a tele-citologia em larga escala. A Figura 2 apresenta um possível novo fluxo.

Figura 2 – Possível cenário com a adoção da solução proposta.



Fonte – Fonte: Do autor.

Para que este paradigma funcione de maneira adequada, os profissionais nas unidades de saúde em cidades menores devem ser treinados para coletar as amostras, preparar as lâminas e digitalizá-las. Estas lâminas podem ser processadas utilizando métodos de processamento de imagens visando realizar uma análise quantitativa e automatizada. Essa análise pode ser feita localmente ou na nuvem. Porém, o software precisaria ser manipulado pelos profissionais da unidade de saúde. Os relatórios seriam gerados de maneira automatizada e enviados junto com os exames para um sistema de telemedicina, via internet. Neste sistema, o especialista de algum centro de saúde poderia analisar os relatórios, juntamente com o próprio exame, dando o diagnóstico final. Estes métodos de visão computacional ajudariam a tornar possível a criação de um sistema que é mais distribuído e rápido, focando no diagnóstico. Entretanto, para que isso aconteça, é necessário que as técnicas de visão computacional utilizadas sejam robustas o suficiente para que se possa lidar com as diferentes variações na aquisição das imagens e preparação das amostras. Outro obstáculo deste cenário seria a dependência da utilização de internet para envio dos relatórios e dos exames. É importante deixar claro que este é um sistema de auxílio ao diagnóstico, em que não há o intuito de substituir o médico.

1.6 ESTRUTURA DO DOCUMENTO

O Capítulo 2 apresenta a fundamentação teórica, onde estão presentes os conceitos necessários para um melhor entendimento do trabalho. No Capítulo 3 são apresentados os materiais e métodos utilizados no desenvolvimento dos experimentos. Neste capítulo é explicado como foram gerados os conjuntos de dados, quais foram as redes neurais utilizadas e como foram treinadas e como os resultados foram avaliados. No Capítulo 4 os resultados são apresentados, juntamente com as análises de cada abordagem e rede neural utilizada. No Capítulo 5 são apresentadas as conclusões do trabalho, assim como as propostas/possibilidades para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados alguns conceitos necessários para compreensão do trabalho. Inicialmente, serão apresentados os conceitos básicos inerentes à área de visão computacional, processamento de imagens, aprendizado de máquina, entre outros. Depois, serão abordados aqueles que possuem um pouco mais de complexidade (redes neurais, *deep learning*, redes neurais convolucionais, etc) e que, eventualmente, serão citados no decorrer do trabalho.

2.1 IMAGEM DIGITAL

Uma imagem digital é a representação discreta de dados possuindo informações espaciais (sobre o formato) e de intensidade (SOLOMON; BRECKON, 2011). É uma função bidimensional, $f(x, y)$, em que x e y são coordenadas espaciais e a amplitude de f em qualquer posição deste plano é chamada de intensidade ou nível de cinza. Quando x , y e os valores de nível de cinza de f são valores finitos e discretos, chamamos de imagem digital (GONZALEZ, R. C.; WOODS, 2009). Em outras palavras, a imagem digital é a representação do mundo real em um computador, feita por meio de amostragem, onde as informações são guardadas em uma matriz. Os elementos desta matriz são chamados de *pixels* (ANUP; XIAOBO, 1993).

2.1.1 Visão Computacional

Na ciência, a "observação" teve e tem um papel muito importante. Há poucos anos atrás, a descrição verbal e os desenhos manuais eram as únicas maneiras de documentar os resultados. Isso mudou com a invenção da fotografia. Tem-se como grande exemplo a astronomia, onde os profissionais foram capazes de medir as posições e os tamanhos das estrelas. Porém, este procedimento demandava muito tempo. Nos dias de hoje, os computadores pessoais tem poder suficiente para processar imagens, sendo capazes de lidar com uma sequência e até modelos 3D (JAHNE, 1997).

A Visão Computacional foi chave para o desenvolvimento dessas tecnologias. Um sistema de Visão Computacional é aquele que executa a mesma tarefa que a visão humana para descobrir o que está presente no mundo e onde estes itens se localizam. O objetivo geral da Visão Computacional é obter informações sobre uma cena através da análise computacional. Isso acontece de acordo com as entradas recebidas (MARR, 2010; ANUP; XIAOBO, 1993). Já o termo "Visão de Máquina" descreve os sistemas que executam tarefas como checar tamanho e integridade de peças no ambiente industrial, por exemplo (JAHNE, 1997).

2.1.2 Processamento Digital de Imagens

As primeiras aplicações na área de processamento de imagens visavam processar dados provenientes da lua. Elas foram desenvolvidas no *Jet Propulsion Laboratory* em 1964. Desta maneira, iniciou-se um novo ramo da ciência chamado de Processamento Digital de Imagens (PDI) (PITAS, 2000). O PDI se refere ao processamento de imagens digitais por um computador digital. Existe uma linha difusa que separa o processamento de imagens de outras áreas como a análise de imagens, por exemplo. Por muitas vezes define-se o PDI como uma disciplina na qual tanto a entrada quanto a saída de um processo são imagens digitais. Porém, como dito anteriormente, existem exceções à regra e um exemplo disso é o cálculo da intensidade média de uma imagem. Este procedimento resulta em um único número. Pela definição utilizada anteriormente, esta operação não seria considerada como sendo de processamento de imagens (GONZALEZ, R. C.; WOODS, 2009).

Mesmo que os limites entre as áreas não sejam claros, é importante que haja a diferenciação entre três tipos de processo: os de baixo, médio e alto nível. Os processos de baixo nível são aqueles onde há operações mais primitivas, como de pré-processamento (redução de ruídos, por exemplo). Seus processos são caracterizados por ter uma imagem como entrada e saída. As operações de nível médio são aquelas que envolvem tarefas como a segmentação (separação da imagem em regiões ou objetos). Como entrada tem-se uma imagem e como saída, os atributos extraídos. Caracteriza-se como processos de alto nível aqueles que tem como característica tentar "dar sentido" a um conjunto de objetos reconhecidos. Em outras palavras, seria algo como tentar realizar funções cognitivas normalmente associadas à visão humana (GONZALEZ, R. C.; WOODS, 2009).

2.2 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial (IA) é uma área relativamente recente. As pesquisas nesta disciplina iniciaram-se após a Segunda Guerra Mundial e o próprio nome foi dado em 1956. A IA não tenta apenas compreender as entidades inteligentes. Ela vai além, tentando construí-las (RUSSELL; NORVIG, 2004). Em muitos casos, técnicas de IA são utilizadas para solucionar problemas relativamente simples ou até aqueles que são complexos e/ou fazem parte de sistemas maiores. Em outras palavras, a "Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos" (COPPIN, 2010).

Há poucos anos atrás, o conhecimento envolvendo a inteligência artificial tinha pouco valor prático, sendo utilizado para resolver alguns problemas curiosos. Estes geralmente eram resolvidos através de algoritmos codificados especialmente para aquela situação. Para adquirir o conhecimento necessário, era preciso realizar entrevistas com

os profissionais da área visando entender quais eram as regras utilizadas nas tomadas de decisão. Desta maneira o programa era construído. Dava-se então o nome de Sistemas Especialistas ou Sistemas Baseados em Conhecimento (LORENA *et al.*, 2000).

A inteligência artificial é uma área muito próspera, com muitas aplicações práticas e uma comunidade de pesquisa muito ativa. Nos primeiros anos as soluções desenvolvidas foram capazes de resolver problemas que eram intelectualmente difíceis para os seres humanos, mas relativamente fáceis para os computadores. Estes eram que problemas que, geralmente, podiam ser descritos de maneira formal, por regras matemáticas. O grande desafio da IA se provou ser a resolução de problemas que podem ser resolvidos de maneira intuitiva, como reconhecer a fala ou faces em imagens (GOODFELLOW *et al.*, 2016).

As soluções desenvolvidas pela comunidade de inteligência artificial permitem que os computadores aprendam por experiência e entendam o mundo em termos de uma hierarquia de conceitos. Cada conceito é definido por uma relação de conceitos mais simples. Ganhando conhecimento por experiência, essa abordagem evita a necessidade de um operador humano precisar formalmente especificar todo o conhecimento que o computador precisa. Essa hierarquia permite que o computador aprenda conceitos complexos, construindo-os através de conceitos mais simples (GOODFELLOW *et al.*, 2016).

2.2.1 Aprendizado de Máquina

Desde a sua invenção, imagina-se o que o computador pode aprender. Se for descoberto um método que faça com que o computador aprenda através de experiência, assim como o ser humano, o impacto seria enorme. Isso elevaria o computador a um novo nível (MITCHELL, 1997). Nos dias de hoje, os algoritmos criados fazem com que certas tarefas sejam bem executadas. Algumas das aplicações que são facilmente realizadas pelos seres humanos, como identificar pessoas pelo rosto, ainda é uma tarefa muito complexa para ser realizada por um computador (CARVALHO *et al.*, 2011).

A medida que o conhecimento nesta área vai se tornando mais maduro, mais o papel do Aprendizado de Máquina (AM) vai ganhando importância dentro da ciência da computação (MITCHELL, 1997). A partir da década de 1970, a inteligência artificial começou a ser utilizada para resolver problemas práticos. Como houve um aumento na complexidade dos problemas e um aumento também na geração de dados, foi necessário criar soluções mais sofisticadas, que fossem capazes de aprender por experiências passadas. Estas deveriam ser capazes de construir uma hipótese/função que fosse capaz de resolver o problema que se deseja tratar. A este processo dá-se o nome de Aprendizado de Máquina. Esta capacidade de aprendizado é essencial para

que haja um comportamento inteligente (CARVALHO *et al.*, 2011).

De outra maneira, o AM pode ser definido como "a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência". Esta área é naturalmente associada a inteligência artificial. Porém, muitas outras disciplinas como a Probabilidade e Estatística, a Teoria da Informação e até a Neurociência tiveram uma parcela de contribuição no desenvolvimento do aprendizado de máquina (MITCHELL, 1997). Esta é uma das áreas que mais vem crescendo nos últimos anos (CARVALHO *et al.*, 2011).

2.2.1.1 Redes Neurais Artificiais

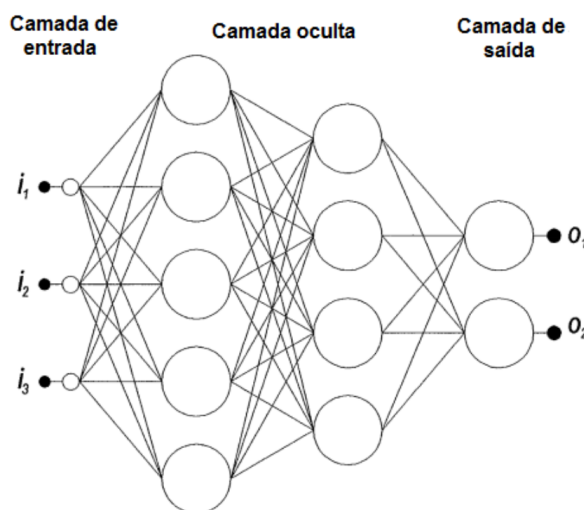
O sistema de aprendizado biológico é construído através de redes de neurônios. Através da observação deste sistema obteve-se a motivação para o estudo das Redes Neurais Artificiais (RNAs) (MITCHELL, 1997). A RNA é uma maneira distribuída e paralela de processamento de informações. Ela é formada pelos "neurônios" (elementos de processamento) e eles são ligados por um canal de sinal chamado de "conexão". Cada neurônio possui uma saída única, a qual pode ser distribuída em quantas direções forem desejadas. O sinal de saída pode ser de qualquer tipo matemático (HECHT-NIELSEN, 1992).

2.2.1.2 *Multilayer Perceptron*

Uma rede neural do tipo *Multilayer Perceptron* (MLP) consiste em um sistema de neurônios interconectados. Este representa o mapeamento não-linear entre a entrada e a saída, onde os neurônios estão conectados por ligações ponderadas. Ou seja, cada ligação possui um valor multiplicador que dará um peso a conexão. A saída de cada neurônio é uma função da soma das entradas modificadas por uma função de ativação, também conhecida como função de transferência (GARDNER; DORLING, 1998).

A saída em cada um dos neurônios é uma função ponderada pelo peso da conexão com a camada anterior. O valor de saída de cada neurônio representa a entrada para a próxima camada da rede. Esta característica demonstra em que direção ocorre o processamento da informação, de maneira que as MLPs são conhecidas como redes neurais *feed forward*. A informação é encaminhada sempre em uma direção (GARDNER; DORLING, 1998). A estrutura básica de uma Rede Neural *Multilayer Perceptron* é apresentada na Figura 3.

Antes da camada de saída, as MLPs podem ter uma ou mais camadas ocultas. Uma rede neural é "completamente conectada" quando cada neurônio é conectado com todos os outros da camada anterior e da próxima. As redes neurais são métodos de aprendizado supervisionado, ou seja, ela "aprende" através de treinamento, necessitando-se de um conjunto de dados "anotados" (*labeled data*). Durante a fase

Figura 3 – Rede Neural Artificial - *Multilayer Perceptron*.

Fonte – Adaptada de (GARDNER; DORLING, 1998).

de treinamento, a rede é alimentada repetidamente com os dados e os pesos das conexões são ajustados até que o conjunto de entradas e saídas estejam de acordo com o esperado (GARDNER; DORLING, 1998).

Durante a fase de treinamento, os valores gerados nas saídas da rede diferem dos desejados. Assim, o erro é definido pela diferença entre a saída atual e a desejada. Durante esta fase, a magnitude do erro é utilizada para determinar quanto os pesos das conexões devem ser ajustados para que este valor de erro possa ser diminuído. Desta maneira, existem vários algoritmos que podem ser utilizados para realizar o treinamento de uma rede neural (GARDNER; DORLING, 1998).

2.3 DEEP LEARNING

A escolha da representação de dados (*features*) adequada influencia fortemente o desempenho dos algoritmos de aprendizado de máquina. Sendo assim, existe um grande esforço em desenvolver métodos de pré-processamento e transformação de dados que melhore a representação dos mesmos, criando um aprendizado efetivo. Neste contexto define-se *Representation Learning* como o aprendizado de representação dos dados que faça com que seja mais fácil extrair informações úteis quando se está construindo classificadores ou preditores. Uma boa representação dos dados é útil como entrada de um algoritmo de aprendizado supervisionado (BENGIO *et al.*, 2013).

Os métodos de *deep learning* são abordagens de *representation learning* com múltiplos níveis de representação. Estes são obtidos pela composição de módulos não-lineares que transformam as representações em cada nível, iniciando pela entrada

bruta, até o nível mais alto, aumentando o nível de abstração (BENGIO *et al.*, 2013). As técnicas de *deep learning* estão fazendo grandes avanços ao resolver problemas que resistiram a muitas das tentativas da comunidade de inteligência artificial durante os anos. Estes métodos se tornaram muito bons em descobrir estruturas intrínsecas em conjuntos de dados complexos (LECUN *et al.*, 2015).

Quando comparadas com as abordagens de *deep learning*, as técnicas clássicas de aprendizado de máquina são limitadas pelo fato de não conseguirem processar os dados em sua forma bruta. Durante décadas, construir um sistema de reconhecimento de padrões ou aprendizado de máquina exigia certo trabalho, já que era necessário desenvolver um algoritmo de extração de características que conseguisse representar de maneira adequada a estrutura intrínseca dos dados. Além disso, normalmente precisava-se de um classificador capaz de processar adequadamente as características extraídas. O grupo de métodos de *representation learning* permite que as máquinas sejam alimentadas com dados brutos e automaticamente descubram a representação necessária para realizar a detecção ou classificação (LECUN *et al.*, 2015).

2.3.1 Redes Neurais Convolucionais

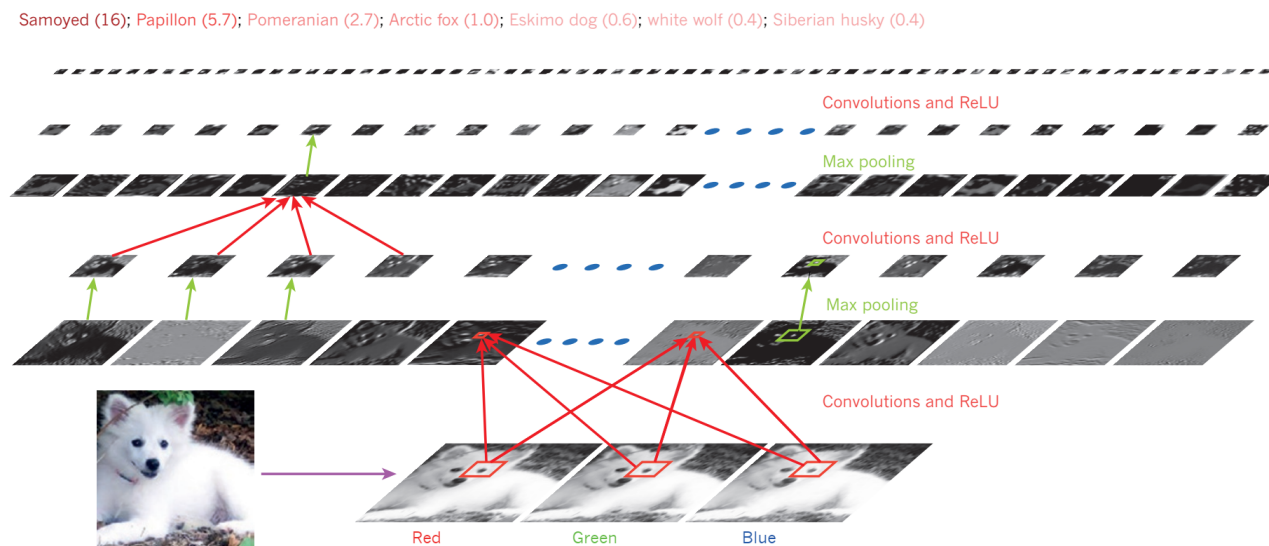
As Redes Neurais Convolucionais são um tipo particular de redes neurais *feed-forward*. Por serem mais fáceis de treinar e generalizar melhor por serem totalmente conectadas entre camadas adjacentes, estes algoritmos vêm sendo adotados cada vez mais pela comunidade de visão computacional. As CNNs são desenvolvidas para processar os dados de entrada em forma de múltiplos *arrays*. Um exemplo disso é uma imagem com os canais *Red*, *Green*, *Blue* (RGB), que é composta por três matrizes 2D, onde cada posição da matriz contém a intensidade do pixel para cada um dos canais (LECUN *et al.*, 2015).

Existem quatro ideias-chave por trás das CNNs: conexões locais, pesos compartilhados, *pooling* e a utilização de várias camadas. A arquitetura típica de uma CNN é apresentada na Figura 4. As redes neurais convolucionais exploram a propriedade de que muitos sinais são compostos por hierarquias, na qual as características dos níveis mais altos são obtidas pela composição dos níveis mais baixos. Por exemplo, em imagens, combinações locais de bordas formam faces, as faces formam partes e as partes formam objetos (LECUN *et al.*, 2015).

Em uma rede neural convolucional, o extrator de características (*backbone*) se refere às camadas (geralmente uma rede neural completa como a ResNet50, por exemplo) que são utilizadas para extrair as características das imagens. Já a arquitetura define as funcionalidades “em volta” deste extrator de características. Desta maneira, o *backbone* pode ser trocado e o modelo (arquitetura) pode ser escolhido de acordo com cada problema visando encontrar o melhor resultado em termos de acurácia, tempo

de processamento e outros requisitos (AMJOURD; AMROUCH, 2020).

Figura 4 – Arquitetura típica de uma Rede Neural Convolutacional.



Fonte – Retirada de (LECUN *et al.*, 2015).

Existem aplicações de CNNs que datam do início da década de 1990. Desde o começo dos anos 2000, estas vêm sendo aplicadas com grande sucesso na detecção, segmentação e reconhecimento de objetos e regiões em imagens. Apesar do seu sucesso atual, as CNNs foram "esquecidas" pela comunidade de visão computacional até a competição *ImageNet* de 2012. Nesta ocasião estas foram utilizadas para processar um *dataset* de aproximadamente um milhão de imagens e mil classes, onde atingiram resultados espetaculares. Esse sucesso trouxe uma revolução no campo da visão computacional (LECUN *et al.*, 2015).

2.3.1.1 Abordagens

Tratando-se de CNNs, existem diferentes abordagens que podem ser adotadas para resolver um problema. As mais comuns são: classificação de imagem, detecção de objetos, segmentação semântica e segmentação de instância (*instance segmentation*). A utilização de cada uma depende do problema e de como deseja-se apresentar a solução. Neste trabalho, optou-se por utilizar as abordagens de segmentação semântica, detecção de objetos e classificação de imagens. Elas serão apresentadas em mais detalhes à seguir.

A Segmentação Semântica se refere ao processo de classificar cada pixel em uma imagem de acordo com um conjunto de classes. Em outras palavras, dada uma imagem, um algoritmo de segmentação semântica deve apresentar em sua saída quais pixels pertencem a um mesmo grupo, "semanticamente" falando. Os métodos recen-

tes de segmentação semântica podem ser divididos em três categorias. A primeira engloba os algoritmos que realizam a segmentação baseada em regiões, os quais extraem suas formas, as descrevem e as classificam. A segunda classe se refere aos métodos conhecidos como *FCN-based Semantic Segmentation Algorithms*. Estes, por sua vez, são baseados em algoritmos FCN (*Fully Convolutional Networks*) e fazem um mapeamento de pixel para pixel, sem extrair as *region proposals*. A terceira classe é conhecida como *Weakly Supervised Segmentation* e são métodos que tem como objetivo realizar o processo de segmentação semântica utilizando anotações na forma de *bounding boxes* ou classificação em nível de imagem (*image-level labels*) (GUO *et al.*, 2018).

Na abordagem de Detecção de Objetos precisa-se localizar um objeto na imagem e posteriormente, classificá-lo. Antigamente, os detectores de objetos eram baseados em extrator de características desenvolvidos "à mão". Para encontrar os objetos, utilizava-se a técnica conhecida como janela deslizante (*sliding window*), o que tornava esta abordagem computacionalmente ineficiente e menos precisa. Os detectores de objetos modernos podem ser classificados em duas categorias. A primeira é baseada em *region proposals*, onde pode-se citar as arquiteturas *Region-Based Convolutional Neural Networks* (R-CNN) (GIRSHICK *et al.*, 2014) e *Faster R-CNN* (REN *et al.*, 2016). A segunda categoria é chamada de *classification-based methods*, na qual podemos citar a *You Only Look Once* (YOLO) (REDMON *et al.*, 2016), *Single Shot Detector* (SSD) (LIU *et al.*, 2016), entre outras (DHILLON; VERMA, 2020).

A Classificação de Imagens é uma das tarefas fundamentais da visão computacional. Este processo tem como objetivo categorizar as imagens em uma ou mais classes, podendo ser considerada a base para outras tarefas de visão computacional como a localização, detecção e a segmentação (RAWAT; WANG, Z., 2017). Mesmo sendo uma tarefa trivial para os seres humanos, é desafiadora para os sistemas automatizados (CIRESAN *et al.*, 2011). Anteriormente, os desenvolvedores adotavam uma *pipeline* de duas etapas para resolver os problemas. Primeiro, era desenvolvido um algoritmo para extrair as características das imagens. Depois, essas características eram utilizadas para treinar um classificador (LECUN *et al.*, 1998). Porém, existem muitas complicações quando se utiliza esta abordagem. Um exemplo disso é o fato de esta ser dependente do ponto de vista do observador. Também há grande variabilidade intra-classe, já que existem diferentes tipos de um mesmo objeto (CIRESAN *et al.*, 2011). Outro problema a ser citado é que a acurácia da tarefa de classificação é profundamente dependente do desenvolvimento e robustez do algoritmo de extração de características (LECUN *et al.*, 1998). Nos dias atuais os modelos de *deep learning* estão sendo utilizados para realizar a extração de características integrada à transformação das mesmas, assim como análise de padrões e classificação. Desta maneira todos estes passos estão integrados em um único *pipeline* que possui uma etapa de

classificação realizada por uma rede neural convolucional. Essa abordagem vem se mostrando capaz de resolver os problemas atuais (RAWAT; WANG, Z., 2017).

2.4 CITOLOGIA

A Citologia é uma técnica de diagnóstico onde as células são colhidas de tecidos através de aspiração ou raspagem. É uma ferramenta que vem sendo utilizada para diagnosticar uma grande quantidade de patologias. Porém, só ficou popular em 1941 (apesar de existir desde 1837), quando foi utilizada para diagnosticar células malignas em amostras coletadas do útero (DIAMANTIS; MAGIORKINIS, 2014).

Nas décadas seguintes, a citologia foi aplicada com sucesso para diagnosticar vários tipos de câncer do corpo humano. Essas técnicas também foram utilizadas para diagnosticar uma variedade complexa de doenças como inflamações e patógenos como fungos, bactérias e vírus (IVANOVIC, 2014). Coletar células utilizando este método apresenta várias vantagens. Dentre elas podemos citar o baixo custo, é menos invasiva e é mais fácil de ser realizada por técnicos de saúde quando comparada com outras maneiras de coletar amostras, como a biópsia (BEDROSSIAN, 2007).

As células coletadas em exames citológicos podem ser analisadas através de diferentes abordagens. Podemos citar a quantificação de DNA, a identificação de proteínas, a presença de patógenos e também a morfologia celular. Cada método requer diferentes corantes e diferentes tipos de processamento. Porém, independentemente do método, a análise final geralmente é feita através de um microscópio (IVANOVIC, 2014).

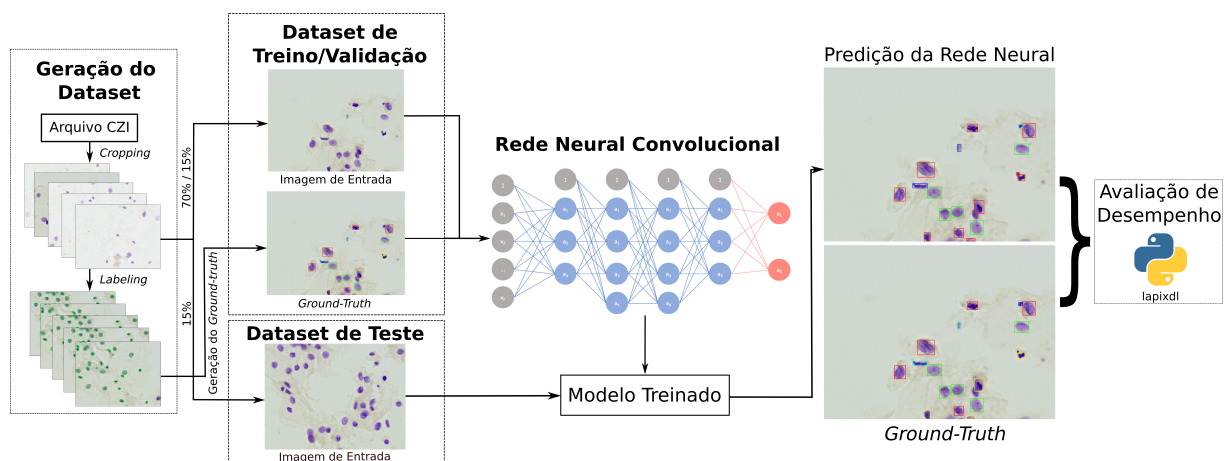
2.4.1 Aplicação da Visão Computacional na área de Citologia

De acordo com os trabalhos analisados na revisão de literatura realizada (AMORIM *et al.*, 2020; MATIAS *et al.*, 2021), pode-se perceber que grande parte dos estudos que utilizam visão computacional na área de citologia ainda estão em uma fase experimental. Ou seja, ainda não estão prontos para serem aplicados na rotina clínica. Uma grande parte deles empregam métodos de *deep learning*. Observando estes trabalhos e comparando com as tendências gerais que são observadas na comunidade de visão computacional, pode-se inferir que nos próximos anos haverá uma maior adoção destes métodos de *deep learning*. Porém, na área de citologia com processamento de imagens ainda existem trabalhos que utilizam métodos clássicos como solução para a automação da análise citológica.

3 MATERIAIS E MÉTODOS

A visão geral da metodologia adotada é apresentada na Figura 5. O primeiro passo foi a geração do conjunto de dados. Nesta etapa as imagens contendo as lâminas escaneadas provenientes de exames citológicos foram utilizadas para gerar os campos (*fields*), que nada mais são do que imagens com a resolução de 1200×1600 pixels. Na etapa seguinte fez-se a anotação destes dados e a exportação das mesmas para os diferentes formatos necessários em cada abordagem. Logo após esta etapa, foi feita a divisão do *dataset* nos subconjuntos de treino (70%), validação (15%) e teste (15%). As imagens do conjunto de treino e validação foram utilizadas para treinar os modelos de redes neurais. Utilizando o *dataset* de teste e o modelo treinado, avaliou-se o desempenho das redes. Estas etapas serão explicadas em detalhes no decorrer desta seção.

Figura 5 – Visão geral da metodologia adotada.



Fonte – Do autor.

3.1 GERAÇÃO DO CONJUNTO DE DADOS

Como citado anteriormente, na literatura não foi encontrado nenhum *dataset* com as características necessárias e que estivesse publicamente disponível. Desta maneira, foi necessário construir um.

O corante de Feulgen é uma reação cito-histoquímica muito utilizada, onde a intensidade do corante é proporcional à concentração de DNA presente na célula. É mais utilizada para quantificação de DNA nos núcleos celulares através de citometria de imagem, com o intuito de fazer a análise de ploidia (CHIECO; DERENZINI, 1999). Escolheu-se por utilizar o corante de Feulgen porque a análise de material citológico utilizando o corante de Papanicolau não é um método confiável para quantificação de

DNA (GURLEY *et al.*, 1990). Um estudo comparativo entre os corantes de Papanicolau, Hematoxilina e Feulgen mostrou que os histogramas gerados através da análise de células epiteliais apresentou uma distribuição regular quando utilizado Feulgen. Já os corantes de Papanicolau e Hematoxilina mostraram uma ampla dispersão de valores, além de uma quantidade maior de valores abaixo de 4c. Sendo assim, a análise de DNA proveniente de amostras coradas com Papanicolau ou Hematoxilina não são adequadas para conclusões diagnósticas ou prognósticas. Isso acontece porque não podem ser interpretados de maneira confiável devido à falta de estequiometria de coloração nuclear (BIESTERFELD *et al.*, 2011).

As imagens utilizadas para construir o conjunto de dados utilizado neste trabalho são provenientes de exames de mulheres que foram à clínica de ginecologia e colposcopia do Hospital Universitário da UFSC (HU-UFSC). Estas mulheres, apresentando alterações citológicas nos exames, foram encaminhadas ao HU-UFSC para realização de exame ginecológico, colposcopia e biópsia. Este projeto de pesquisa foi aprovado pelo Comitê de Ética em Pesquisa da UFSC (CEPSH), número de protocolo 57423616.3.0000.0121. Todas as pacientes que participaram foram devidamente avisadas sobre os objetivos do estudo. Aquelas que aceitaram participar assinaram um termo de consentimento.

No momento do exame ginecológico, foram coletadas amostras citológicas com espátula de Ayre e escova endocervical para a execução de citologia em base líquida e da citometria do DNA. As amostras foram armazenadas em solução ThinPrep® PreservCyt® (Hologic Inc. Massachusetts, EUA), enviadas ao Laboratório de Citologia Clínica da UFSC. As lâminas coradas pelo método de Papanicolaou foram examinadas por meio de microscopia de luz para verificar a celularidade da amostra e a presença de células atípicas para aplicação da citometria de DNA. Após análise citológica, as lâminas foram imersas em xilol até a completa remoção das lamínulas para posterior coloração de Feulgen, sendo processadas para a confecção das lâminas. O processo completo da preparação das amostras e citologia está descrito em (COSTA, 2018).

O sistema de análise de imagem utilizado foi o MotiCyt DNA Cytology Work Station (Motic, Xiamen, China), que consiste em um computador acoplado a um microscópio Motic BA410™, com câmera de vídeo colorida Moticam Pro 285A™ (1.4 megapixels, 1360 x 1024 pixels). Para este trabalho foram utilizadas cinco lâminas ($n = 5$). Antes do processo de anotação, os campos (*fields*) são gerados a partir de um processo de recorte (*cropping*) da lâmina usando o *Zen Software 2.6 (Blue Edition)* e um *script* desenvolvido em Python. Essa geração de campos é necessária porque a resolução da imagem que representa a lâmina é muito grande para ser usada no treinamento da rede neural (cerca de 209.000×148.000 pixels). O processo consiste exportar (gerar) os campos a partir do arquivo original gerado pelo *scanner* de lâminas. Este processo gera sub-imagens com resolução de 1200×1600 .

Esses núcleos precisam ser rotulados (anotados) para o treinamento da rede neural. Esse processo de anotação consiste em selecionar cada um dos núcleos presentes nas imagens e indicar a qual classe (descritas no fim desta subseção) ele pertence. Desta maneira, os especialistas utilizaram os softwares *labelme* (WADA, 2016) e *LabelBox*¹ para realizar as anotações. Estes softwares geram um arquivo de saída contendo os pontos do polígono circundante que foi criado. Tendo essas anotações prontas, é possível gerar diferentes *ground-truths* (anotações) para diferentes abordagens.

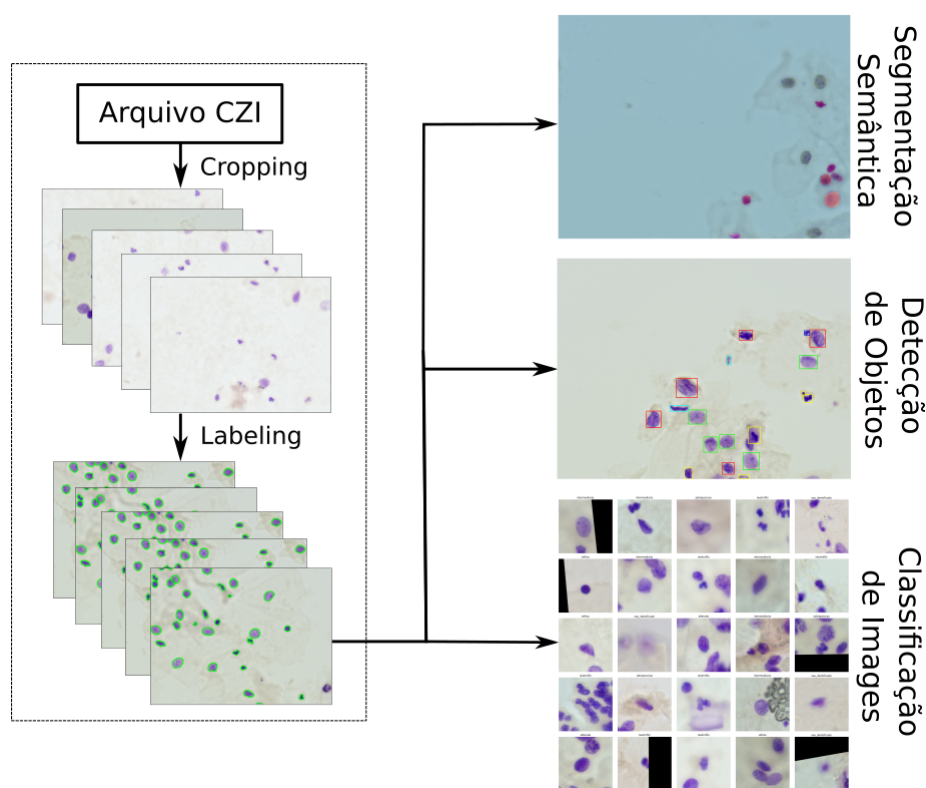
Para gerar as anotações para os testes com segmentação semântica foi desenvolvido um *script* em Python que interpreta os arquivos de saída dos softwares de anotação. Para a geração das anotações para os experimentos de detecção de objetos, interpretou-se as marcações de cada núcleo e gerou-se o quadrado mínimo que possa envolver este polígono. Para a geração das amostras para o treinamento das redes de classificação de imagem, utilizou-se as marcações de cada núcleo e selecionou-se um campo de 256x256 partindo do centro de cada núcleo. Quando alguma anotação se encontrava nas bordas das imagens, "completou-se" as mesmas com pixels pretos. A Figura 6 apresenta os tipos de abordagens testadas neste trabalho.

O conjunto de dados foi dividido entre os subconjuntos de treinamento, validação e teste. O conjunto de treinamento consiste em um grupo de exemplos que são utilizados pelo classificador para "aprender". Ou seja, para ajustar seus parâmetros internos. O conjunto de validação contém exemplos que são utilizados para ajustar os hiper-parâmetros do modelo (para ajustar a quantidade de camadas internas de uma rede neural, por exemplo). O conjunto de teste é utilizado para verificação de performance do classificador treinado. Em outras palavras, o conjunto de teste é a parte dos dados que é utilizada para obter uma avaliação sem vieses do modelo final, que foi treinado utilizando os dados do conjunto de treinamento (RIPLEY, 2007). O *dataset* utilizado neste trabalho contém 1955 imagens no total e foi dividido aleatoriamente em 70% para o treinamento (1565 imagens), 15% (195 imagens) para validação e 15% (195 imagens) para teste.

Dentre as classes presentes no conjunto de dados, além do *background* (fundo da lâmina), os núcleos foram classificados em "Intermediário" (núcleo de referência), aquele que não possui uma quantidade anormal de DNA. A segunda classe é o núcleo "Alterado". Este possui alterações na quantificação de DNA. A classe "Artefato" representa alguns artefatos de técnica que podem aparecer nas lâminas (geralmente uma mancha preta). A classe "Não Identificada" representa os núcleos que geralmente estão fora de foco ou possuem uma quantidade baixa de corante. Os núcleo das células "Epiteliais" são aqueles de células epiteliais que estão se separando da mucosa. A classe "Neutrófilo" representa um neutrófilo (um tipo específico de célula de defesa). O

¹ <https://labelbox.com/>

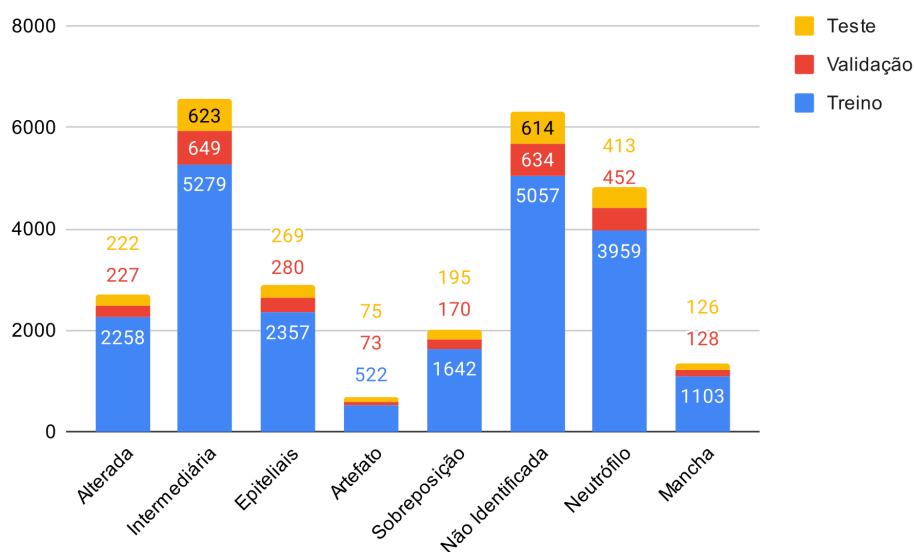
Figura 6 – Demonstração dos diferentes tipos de *ground-truths* e suas respectivas abordagens.



Fonte – Do autor.

conjunto de dados também contem a classe "Sobreposição", que demonstra a sobreposição entre dois ou mais núcleos. Finalmente tem-se a classe "Mancha", que denota um agrupamento contendo um grande número de células e/ou excesso de corante. A Figura 7 demonstra a distribuição das classes dentro do conjunto de dados.

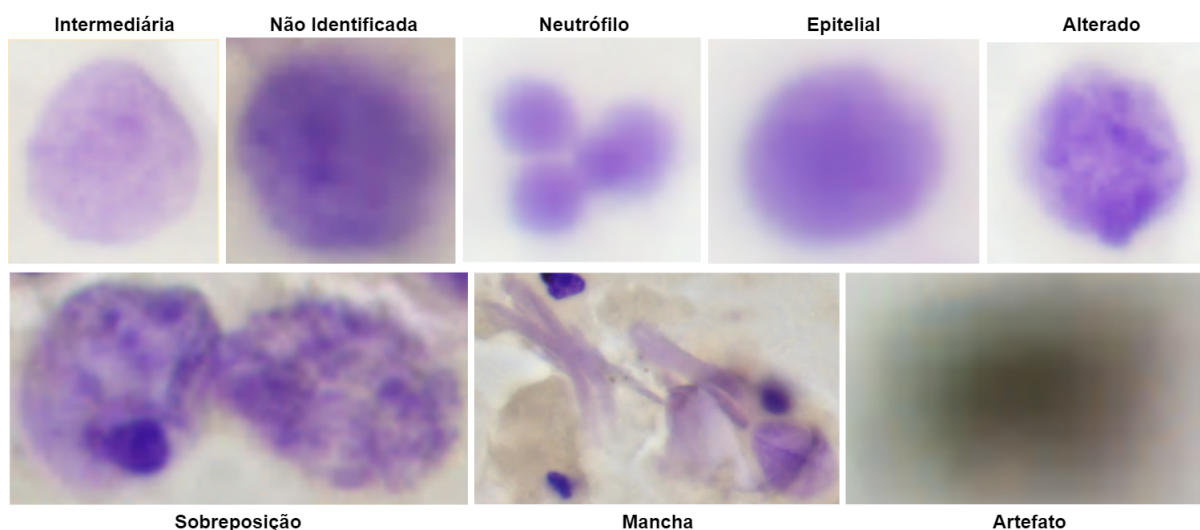
Figura 7 – Distribuição de amostras (por classe) dentro do conjunto de dados.



Fonte – Do autor.

A Figura 8 apresenta um exemplo visual de cada classe de núcleo presente no conjunto de dados.

Figura 8 – Amostra de cada classe presente no conjunto de dados.



Fonte – Do autor.

3.2 TREINAMENTO DAS REDES NEURAIIS

Para verificar qual é a melhor combinação entre rede neural e abordagem para detecção precoce de câncer, foram comparados diferentes modelos de CNNs e abordagens. Por isso, além das redes neurais de segmentação semântica, serão realizados experimentos com redes de detecção de objetos e classificação de imagens.

3.2.1 Segmentação Semântica

Para treinar as redes neurais de segmentação semântica foi utilizada a biblioteca *fast.ai*². A *U-Net* foi escolhida por ter sido criada para segmentação de imagens biomédicas (RONNEBERGER *et al.*, 2015), apresentando bons resultados em aplicações descritas na literatura. Visando mitigar o problema de desbalanceamento entre classes no conjunto de dados, optou-se por testar a *Weighted U-Net*. A ideia é dar pesos diferentes para cada classe de maneira inversamente proporcional a quantidade de pixels de cada classe presente no *dataset*. Em outras palavras, desenvolveu-se um *script* para verificar quantos pixels cada classe possuía dentro do conjunto de dados. Aquele com a maior quantidade (excluindo-se o *background*) receberia o menor peso.

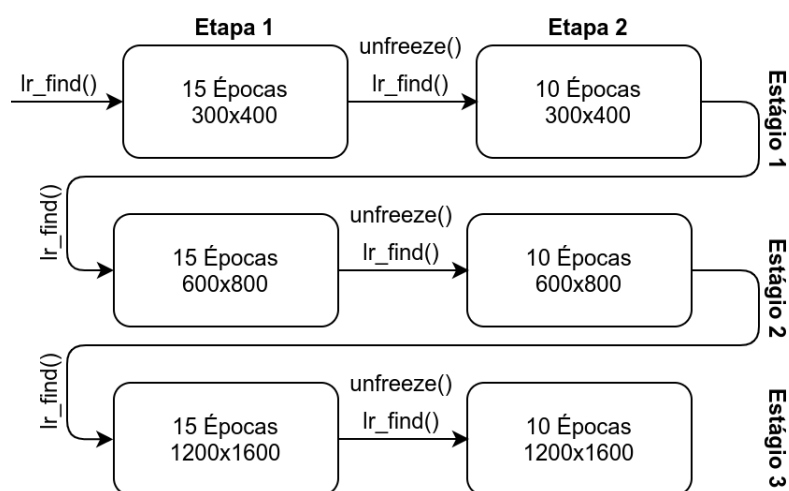
Por exemplo, verificou-se que a classe "Mancha" possuía 35591241 pixels dentro do *dataset*, a maior quantidade dentre as classes. Desta maneira, ela recebeu peso 1 (um), o valor mais baixo. A classe "Alterada" possui 11038699 pixels. Dividindo a maior quantidade de pixels do *dataset* (da classe "Mancha") pela quantidade de pixels da presente classe obteve-se um peso de 3,22422425. Isso foi feito para todas as classes do conjunto de dados. Para o treinamento da rede neural os pesos foram normalizados entre zero e um. Esses valores são utilizados na *loss function* (*Cross Entropy Loss*), no qual cada classe irá receber pesos diferentes no momento de calcular o erro para cada uma delas.

Quando lidamos com segmentação de dados médicos, a disponibilidade destes geralmente é um problema, levando em conta que é necessário anotá-los. *Data Augmentation* ajuda a aumentar a quantidade de amostras, auxiliando também a prevenir *overfitting*. Além disso, geralmente há melhora no desempenho da rede neural quando se está lidando com dados de fora do conjunto de treino. É um processo importante quando se está desenvolvendo um *pipeline* de *deep learning* (EATON-ROSEN *et al.*, 2018).

Os processos de *data augmentation* aplicados no treinamento das redes de segmentação semântica foram o *vertical flip* com probabilidade de 0,5, o *horizontal flip* com probabilidade de 0,5, a rotação randômica entre -10 e 10 graus com probabilidade de 0,75, zoom randômico entre 1 e 1,1 com probabilidade de 0,75 e *affine* randômica com *shear* de 0,2, que também foi aplicada com probabilidade de 0,75.

² <https://www.fast.ai/>

Figura 9 – Representação esquemática da estratégia adotada no processo de treinamento da rede (HOWARD, 2019).



Fonte – Do autor.

A estratégia de treinamento foi dividida em três estágios e cada estágio foi dividido em duas etapas, como mostrado na Figura 9. De um estágio para o outro a resolução da imagem foi aumentada por um fator de 2 (dobrou-se a resolução). Em outras palavras, no primeiro estágio, a rede foi treinada com as imagens possuindo 1/4 da sua resolução original. No segundo estágio, o treinamento continuou com as imagens tendo 1/2 da sua resolução original, e finalmente o treino foi concluído utilizando as imagens em sua resolução original. Esta estratégia é baseada na ideia de que, ao aumentar a resolução das imagens, nós temos um novo conjunto de dados. Ou seja, esse incremento na resolução é similar ao processo de *data augmentation*, ajudando a evitar *overfitting* (HOWARD, 2019).

Em cada etapa a rede neural foi treinada por quinze ou dez épocas utilizando o *1cycle policy* (SMITH, L. N., 2018). Este processo consiste em treinar a rede aumentando o *learning rate* de um valor muito baixo para um muito alto, parando quando a *loss* de treino e validação começam a divergir. As *losses* são plotadas com os *learning rates* e o valor que está um pouco antes da *loss* mínima é escolhido, porque assim ela ainda pode ser melhorada. A implementação provida pelo *fast.ai* foi feita com os mesmos métodos de *data augmentation* do artigo original, mas com uma pequena modificação. Os *padded pixels* em preto não foram colorados. Assim, foi utilizado um *reflection padding*.

O *batch size* em cada estágio foi escolhido por limitações técnicas. Foram tentados vários valores, começando em um relativamente alto, sendo diminuídos por um fator 2^n (32, 16, 8...), até que os dados pudessem ser totalmente alocados na memória

da placa de vídeo³. O valor de taxa de aprendizado (*learning rate*) em cada etapa e estágio foi escolhido utilizando o método discutido em (SMITH, L. N., 2015). Este método consiste em testar valores de $1e^{-7}$ até $1e^{-10}$ e parar quando a *loss* diverge. O valor da taxa de aprendizado deve ser escolhido logo antes de a melhor *loss* ser atingida, já que permite que ela seja melhorada durante o treino (HOWARD, 2019). O processo de definição da taxa de aprendizado é realizado antes de cada etapa do treinamento (Figura 9).

Na primeira etapa de cada estágio apenas as camadas finais da rede neural foram treinadas. O restante da rede foi congelada (*frozen*). Na segunda etapa de cada estágio todas as camadas foram treinadas (descongelou-se a CNN) e foi realizado um *fine-tuning*. Esse processo foi repetido três vezes (os três estágios) com a resolução da imagem crescente. No último estágio a rede neural foi treinada com as imagens em sua resolução original. Na primeira etapa dos estágios 1 e 2 o *learning rate* foi fixo. Utilizou-se métodos de transferência de aprendizado (*transfer learning*) nos experimentos realizados. Em outras palavras, o extrator de características foi pré-treinado no ImageNet.

3.2.2 Detecção de Objetos

Para treinar as redes de detecção de objetos optou-se por utilizar a biblioteca *detectron2* (WU *et al.*, 2019), do *Facebook AI Research*. Esta possui diversos modelos de redes neurais convolucionais implementadas, facilitando os testes neste tipo de pesquisa aplicada. A biblioteca foi implementada utilizando o *PyTorch*⁴. Utilizou-se modelos pré-treinados com o *COCO Dataset*⁵. Optou-se por comparar dois modelos de rede neural: *Faster R-CNN* (REN *et al.*, 2016) e *RetinaNet* (LIN *et al.*, 2017). Esses modelos foram testados com os *backbones ResNet50* e *ResNet101*, resultando em quatro comparações. Neste experimento as redes foram pré-treinadas com o ImageNet.

3.2.3 Classificação de Imagens

Para os experimentos da abordagem de classificação de imagem optou-se por treinar e testar três redes neurais diferentes: a *ResNet34*, *ResNet50* e *ResNet101*. Estas foram implementadas utilizando o *fast.ai*. A grande diferença entre estas três arquiteturas de redes neurais é a quantidade de camadas. Em outras palavras, a *ResNet34* tem 34 camadas, a *ResNet50* possui 50 e a *ResNet101* contém 101 camadas.

O processo de treinamento foi dividido em duas etapas. Antes de cada uma, foi necessário encontrar o *learning rate*. Optou-se por usar o mesmo método (*lr_find*) utilizado nos treinamentos das redes de segmentação semântica. Na primeiras etapas

³ Do inglês, *Graphics Processing Units* (GPU).

⁴ <https://pytorch.org/>

⁵ <https://cocodataset.org/>

os modelos foram treinados por 15 épocas com as camadas iniciais congeladas. Na segunda etapa, descongelou-se a rede inteira e treinou-se por mais 15 épocas. As redes neurais utilizadas foram pré-treinadas utilizando o ImageNet.

3.3 CRITÉRIOS DE AVALIAÇÃO/COMPARAÇÃO

As redes neurais serão avaliadas de acordo com a sua assertividade em relação a segmentação, detecção e classificação de cada tipo de núcleo. A avaliação será feita utilizando a biblioteca *LAPiX DL*⁶, implementada pelo grupo de pesquisa do LAPiX⁷, laboratório da UFSC. Para a avaliação dos modelos serão utilizadas as seguintes métricas:

- Verdadeiro Positivo - *True Positive* (TP): é a quantidade de amostras corretamente classificadas. No contexto de detecção de objetos, considera-se um TP quando o IoU entre o objeto predito e o *ground-truth* é maior do que um determinado limiar;
- Verdadeiro Negativo - *True Negative* (TN): denota aquelas amostras que são consideradas negativas e o algoritmo prediz como negativa;
- Falso Positivo - *False Positive* (FP): quando a amostra não está presente no *ground-truth*, mas o algoritmo a detecta. Em outras palavras, quando o modelo detecta um objeto que não está presente no *ground-truth* ou foi detectado com uma confiança muito baixa;
- Falso Negativo - *False Negative* (FN): quando a amostra está presente no *ground-truth* e o algoritmo não foi capaz de detectá-la;
- IoU - *Intersection Over Union*: é uma métrica baseada no *Jaccard Index* (JACCARD, 1901) que avalia a sobreposição entre duas regiões na imagem: a região do *ground-truth* e a região predita pela rede neural (ZHAO *et al.*, 2018; ULKU; AKAGUNDUZ, 2019);
- Acurácia: é a razão entre os verdadeiros positivos mais os verdadeiros negativos e o total de amostras. Indica a efetividade geral do modelo (FAWCETT, 2006);
- *Recall* (ou Sensitividade): é a razão entre os verdadeiros positivos e a soma dos falsos negativos com os verdadeiros positivos. Essa métrica indica a probabilidade de o modelo classificar casos positivos corretamente (FAWCETT, 2006; POWERS, 2011);
- *Precision*: é a razão entre os verdadeiros positivos e a soma dos falsos positivos com os verdadeiros positivos. Essa métrica também é conhecida como *Positive*

⁶ <https://github.com/lapix-ufsc/lapixdl>

⁷ <http://www.lapix.ufsc.br/>

Predictive Value (PPV) e demonstra o quão preciso é o modelo (FAWCETT, 2006; POWERS, 2011);

- *F-Score* (ou *F-Measure*): é a medida de similaridade normalizada, consistindo na união de classes normalizada (CRUM *et al.*, 2006). Também pode ser definida como a média harmônica da *precision* e do *recall* para um dado limiar (ULKU; AKAGUNDUZ, 2019; POWERS, 2011). É uma métrica utilizada para medir o desempenho quando o modelo é treinado com dados desbalanceados (HE; MA, 2013);
- Especificidade: é a razão entre os verdadeiros negativos e a soma dos falsos positivos e verdadeiros negativos. Essa métrica demonstra a probabilidade de o modelo classificar os casos negativos corretamente (FAWCETT, 2006; POWERS, 2011).

As fórmulas para determinar cada uma das métricas são apresentadas na Tabela 8.

Tabela 8 – Equações das métricas utilizadas no presente trabalho.

Métrica	Fórmula
IoU	$\frac{TP}{FP+FN+TP} = \sum_{j=1}^k \frac{n_{jj}}{n_{ij}+n_{ji}+n_{jj}}, \quad i \neq j$
Acurácia	$\frac{TP+TN}{FP+FN+TP+TN} = \sum_{j,f=1}^k \frac{n_{ff}+n_{jj}}{n_{jj}+n_{ff}+n_{ji}+n_{ij}}, \quad i \neq j, i \neq f$
Recall	$\frac{TP}{FN+TP} = \sum_{j=1}^k \frac{n_{jj}}{n_{ij}+n_{jj}}, \quad i \neq j$
Precision	$\frac{TP}{FP+TP} = \sum_{j=1}^k \frac{n_{jj}}{n_{ij}+n_{jj}}, \quad i \neq j$
F-Score	$\frac{2 \times (Precision \times Recall)}{Precision + Recall} = \frac{2 \times TP}{FP+FN+2TP} = \sum_{j=1}^k \frac{2 \times n_{jj}}{n_{ij}+n_{ji}+2n_{jj}}, \quad i \neq j$
Especificidade	$\frac{TN}{FP+TN} = 1 - FPR = \sum_{j,f=1}^k \frac{n_{ff}}{n_{ij}+n_{ff}}, \quad i \neq j, i \neq f$

4 RESULTADOS

Os treinamentos das redes neurais de segmentação semântica e classificação de imagens foram realizados na plataforma Google Colab Pro¹, onde é disponibilizada uma das seguintes GPUs: Nvidia Tesla K80 (12 GiB VRAM), T4 (16 GiB VRAM), P100 (16 GiB VRAM) ou Tesla V100 (16 GiB VRAM). O processador é um Intel(R) Xeon(R) CPU @ 2.20GHz × 4, com 26 GiB de memória RAM e Ubuntu 18.04.3 LTS 64-bit.

O tempo de treinamento para a U-Net + ResNet18 foi de, aproximadamente, 11h50min. Já a U-Net + ResNet34 teve um tempo de treinamento de 13h25min. A Weighted U-Net + ResNet18 levou 12h30min e a Weighted U-Net + ResNet34, aproximadamente, 7h45min². Para os experimentos de classificação de imagem, a ResNet34 foi treinada em aproximadamente 45 min. A ResNet50 levou cerca de 50 minutos e a ResNet101, 1h30min. Para o treinamento das redes de detecção de objetos utilizou-se uma máquina com Intel(R) Core(R) i9-10850K CPU @ 3.60GHz × 20, com 32 GiB de memória RAM, NVIDIA GeForce RTX3070 e Ubuntu 18.04.5 LTS 64-bit. O treinamento da Faster R-CNN e da RetinaNet com ResNet50 levou aproximadamente 1h40min. Já o treinamento destas duas redes neurais utilizando a ResNet101 como *backbone* levou aproximadamente 2h40min. Os resultados obtidos com cada abordagem serão apresentados nas subseções a seguir.

4.1 SEGMENTAÇÃO SEMÂNTICA

Para todas as redes neurais desta abordagem utilizou-se o valor de *weight decay* de $1e^{-3}$. A Tabela 9 apresenta os *learning rates* para cada etapa e estágio do treinamento das redes neurais de segmentação semântica.

Tabela 9 – *Learning rates* utilizados nos treinamentos das redes neurais de segmentação semântica.

Classe	U-Net + R18	U-Net + R34	W. U-Net + R18	W. U-Net + R34
Etapa 1 / Estágio 1	$1e^{-3}$	$1e^{-3}$	$1e^{-3}$	$1e^{-3}$
Etapa 2 / Estágio 1	$1e^{-4}$ a $1e^{-3}$	$1e^{-4}$ a $1e^{-3}$	$1e^{-4}$ a $1e^{-3}$	$1e^{-4}$ a $1e^{-3}$
Etapa 1 / Estágio 2	$1e^{-4}$	$1e^{-3}$	$1e^{-3}$	$1e^{-4}$
Etapa 2 / Estágio 2	$1e^{-4}$ a $1e^{-3}$	$1e^{-3}$	$1e^{-4}$ a $1e^{-3}$	$1e^{-5}$
Etapa 1 / Estágio 3	$1e^{-3}$	$1e^{-3}$	$1e^{-4}$	$1e^{-5}$
Etapa 2 / Estágio 3	$1e^{-3}$	$1e^{-4}$ a $1e^{-3}$	$1e^{-3}$	$1e^{-5}$

A Tabela 10 apresenta os pesos utilizados (para cada classe) durante o treinamento das *Weighted U-Nets*.

¹ <https://colab.research.google.com/>

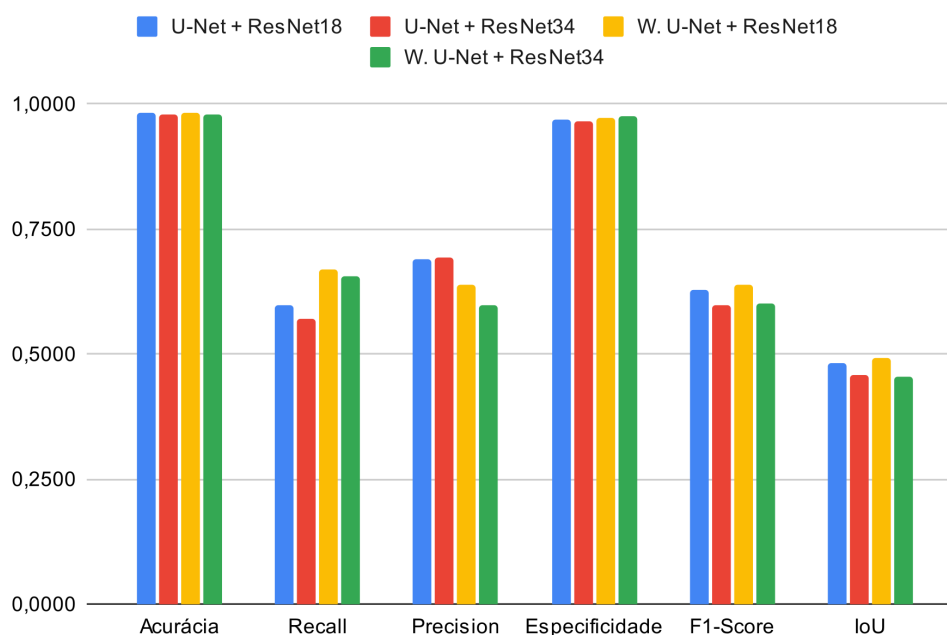
² Os tempos de treinamento variam de acordo com a disponibilização dos modelos de placa de vídeo pela plataforma da Google.

Tabela 10 – Pesos utilizados (para cada classe) durante o treinamento das *Weighted U-Nets*.

Classe	Peso
Background	0,1278772378516624
Alterada	0,411764705882353
Intermediária	0,159846547314578
Epitelial	1,0
Artefato	0,4092071611253197
Sobreposição	0,3414322250639386
Não Identificada	0,4092071611253197
Neutrófilo	0,5588235294117647
Mancha	0,1278772378516624

A Figura 10 apresenta os resultados obtidos após o treinamento das quatro redes neurais utilizadas neste trabalho: U-Net + ResNet18, U-Net + ResNet34, *Weighted U-Net* + ResNet18 e *Weighted U-Net* + ResNet34.

Figura 10 – Métricas médias para as redes neurais de segmentação semântica.



Fonte – Do autor.

De acordo com os dados presentes na Figura 10, pode-se observar que a *Weighted U-Net* com a ResNet18 obteve o melhor IoU médio dentre os quatro modelos. O segundo melhor resultado foi obtido com a U-Net + ResNet18, novamente com o menor *backbone* dentre os dois comparados. Isso mostra que, para esta aplicação, o *backbone* menor acabou apresentando o melhor resultado. Além disso, a utilização

dos pesos para diminuir o desbalanceamento entre as classes apresentou melhorias nos resultados. Além dos dois melhores IoU, as redes com ResNet18 apresentaram o melhor *F-Score*, mostrando que a rede teve boa acurácia processando os dados do conjunto de teste.

Em relação a *Precision*, as duas *Weighted* U-Nets apresentaram resultados inferiores quando comparadas com as redes sem a diferenciação de pesos, mostrando uma queda significativa nesta métrica. Em relação ao *Recall*, as *Weighted* U-Nets tiveram os melhores resultados, selecionando uma quantidade maior de amostras relevantes. Em relação a especificidade, as redes anteriormente citadas também apresentaram resultados melhores, refletindo em uma quantidade menor de falsos negativos.

4.1.1 Resultados por classe

A Tabela 11 apresenta as métricas obtidas com a U-Net e ResNet18 como *backbone*.

Tabela 11 – Métricas - Segmentação Semântica com U-Net + ResNet18.

Classe	IoU	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.3344	0.9980	0.4361	0.5892	0.9993	0.5012
Intermediária	0.6623	0.9970	0.8267	0.7691	0.9982	0.7968
Epitelial	0.3768	0.9989	0.5327	0.5628	0.9995	0.5473
Artefato	0.2696	0.9970	0.2947	0.7595	0.9996	0.4247
Sobreposição	0.4195	0.9972	0.5216	0.6820	0.9991	0.5911
Não Identificada	0.2969	0.9970	0.4364	0.4814	0.9986	0.4578
Neutrófilo	0.5367	0.9987	0.7841	0.6298	0.9991	0.6985
Mancha	0.4667	0.9937	0.5592	0.7383	0.9980	0.6364

As métricas para a U-Net + ResNet34 podem ser encontradas na Tabela 12.

Tabela 12 – Métricas - Segmentação Semântica com U-Net + ResNet34.

Classe	IoU	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.3475	0.9982	0.4151	0.6808	0.9995	0.5158
Intermediária	0.6525	0.9969	0.8045	0.7754	0.9983	0.7897
Epitelial	0.3553	0.9989	0.4937	0.5589	0.9995	0.5243
Artefato	0.2115	0.9970	0.2173	0.8885	0.9999	0.3492
Sobreposição	0.4334	0.9974	0.5273	0.7086	0.9992	0.6047
Não Identificada	0.2656	0.9971	0.3651	0.4938	0.9989	0.4198
Neutrófilo	0.5106	0.9985	0.8406	0.5654	0.9988	0.6761
Mancha	0.3505	0.9915	0.4682	0.5825	0.9967	0.5191

As métricas obtidas com a *Weighted* U-Net e ResNet18 como *backbone* são apresentadas na Tabela 13.

Tabela 13 – Métricas - Segmentação Semântica com *Weighted* U-Net + ResNet18.

Classe	IoU	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.3942	0.9980	0.5669	0.5642	0.9990	0.5655
Intermediária	0.6339	0.9971	0.7009	0.8690	0.9992	0.7759
Epitelial	0.3159	0.9982	0.7064	0.3636	0.9985	0.4801
Artefato	0.4069	0.9970	0.5489	0.6113	0.9987	0.5784
Sobreposição	0.4865	0.9971	0.7059	0.6101	0.9983	0.6545
Não Identificada	0.3161	0.9967	0.5224	0.4445	0.9981	0.4803
Neutrófilo	0.5163	0.9985	0.8632	0.5624	0.9987	0.6810
Mancha	0.3622	0.9927	0.4213	0.7211	0.9983	0.5318

Na Tabela 14 são apresentados os resultados obtidos com a *Weighted* U-Net + ResNet34.

Tabela 14 – Métricas - Segmentação Semântica com *Weighted* U-Net + ResNet34.

Classe	IoU	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.2711	0.9977	0.3616	0.5202	0.9992	0.4266
Intermediária	0.6204	0.9967	0.7571	0.7745	0.9984	0.7657
Epitelial	0.3140	0.9983	0.6283	0.3856	0.9988	0.4779
Artefato	0.4692	0.9963	0.8651	0.5062	0.9968	0.6387
Sobreposição	0.3963	0.9970	0.5119	0.6370	0.9989	0.5677
Não Identificada	0.2792	0.9957	0.5740	0.3522	0.9969	0.4366
Neutrófilo	0.4463	0.9980	0.8626	0.4805	0.9983	0.6171
Mancha	0.3146	0.9924	0.3591	0.7174	0.9986	0.4786

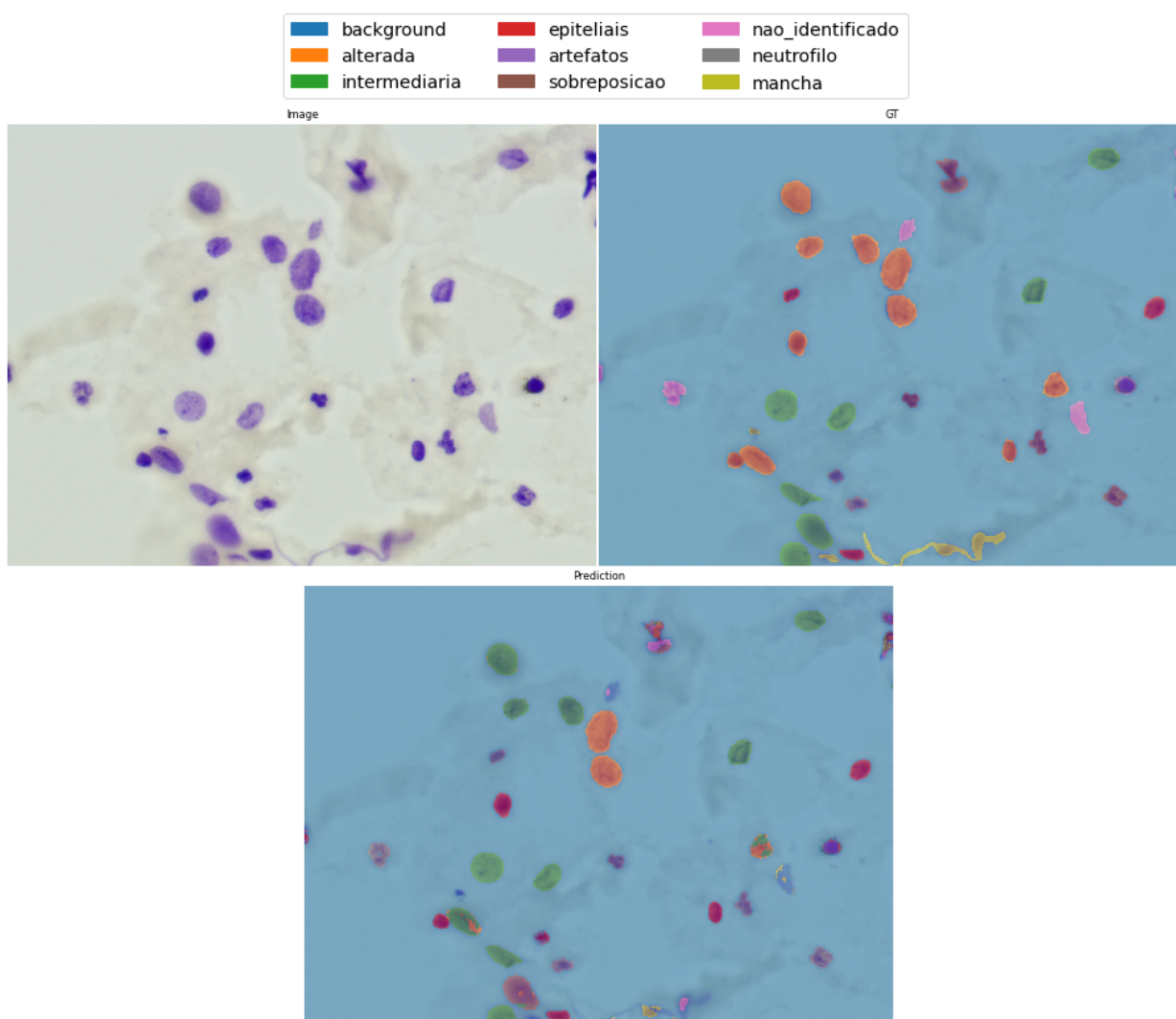
Observando os resultados das U-Nets em relação ao IoU, é possível verificar que a ResNet18 apresentou melhores métricas para a maioria das classes, exceto para os núcleos alterados e as sobreposições. Analisando o *Recall*, o resultado foi similar, onde a ResNet18 apresentou métricas maiores para todas as classes, exceto para Sobreposição (por uma diferença muito pequena) e para os neutrófilos. Já para os valores de IoU das *Weighted* U-Nets, o resultado foi diferente. A rede com ResNet34 apresentou melhores resultados para quase todas as classes, exceto os núcleos intermediários e aqueles das células epiteliais. Em relação ao *Recall*, o resultado foi misto, onde as melhores métricas variaram entre as classes.

Os valores de *F-Score* obtidos com a U-Net + ResNet18 foram melhores para todas as classes, exceto para a Alterada e a Sobreposição. Para as *Weighted* U-Nets os resultados foram similares, onde a ResNet18 apresentou as melhores métricas para todas as classes, exceto para a classe Artefato, aquela que possui a menor quantidade de amostras. Em relação a *Precision*, os resultados foram mistos para as U-Nets, onde o melhor resultado para cada classe variou entre os dois *backbones*. Para as *Weighted* U-Nets, aquela com a ResNet18 apresentou os melhores resultados para seis das oito classes, sendo que a ResNet34 apresentou as melhores métricas para os núcleos das células epiteliais e a classe Sobreposição.

Quando fazemos uma comparação entre as redes com e sem o balanceamento por pesos, a *Weighted* ResNet18 melhorou os resultados em 5 das 8 classes. Para as classes Intermediária, Neutrófilo e Velhas, a U-Net com ResNet18 apresentou os melhores resultados, sendo estas as classes que possuem as maiores quantidades de amostras (excluindo-se os núcleos não identificados). Já para as ResNets34, fazendo uma comparação classe a classe, houve melhora apenas para a classe Artefato, aquela que possui menor quantidade de amostras. Para as demais, o resultado foi similar ou pior.

A Figura 11 apresenta alguns resultados obtidos com a U-Net + ResNet18.

Figura 11 – Resultado das Segmentações para a U-Net + ResNet18. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.

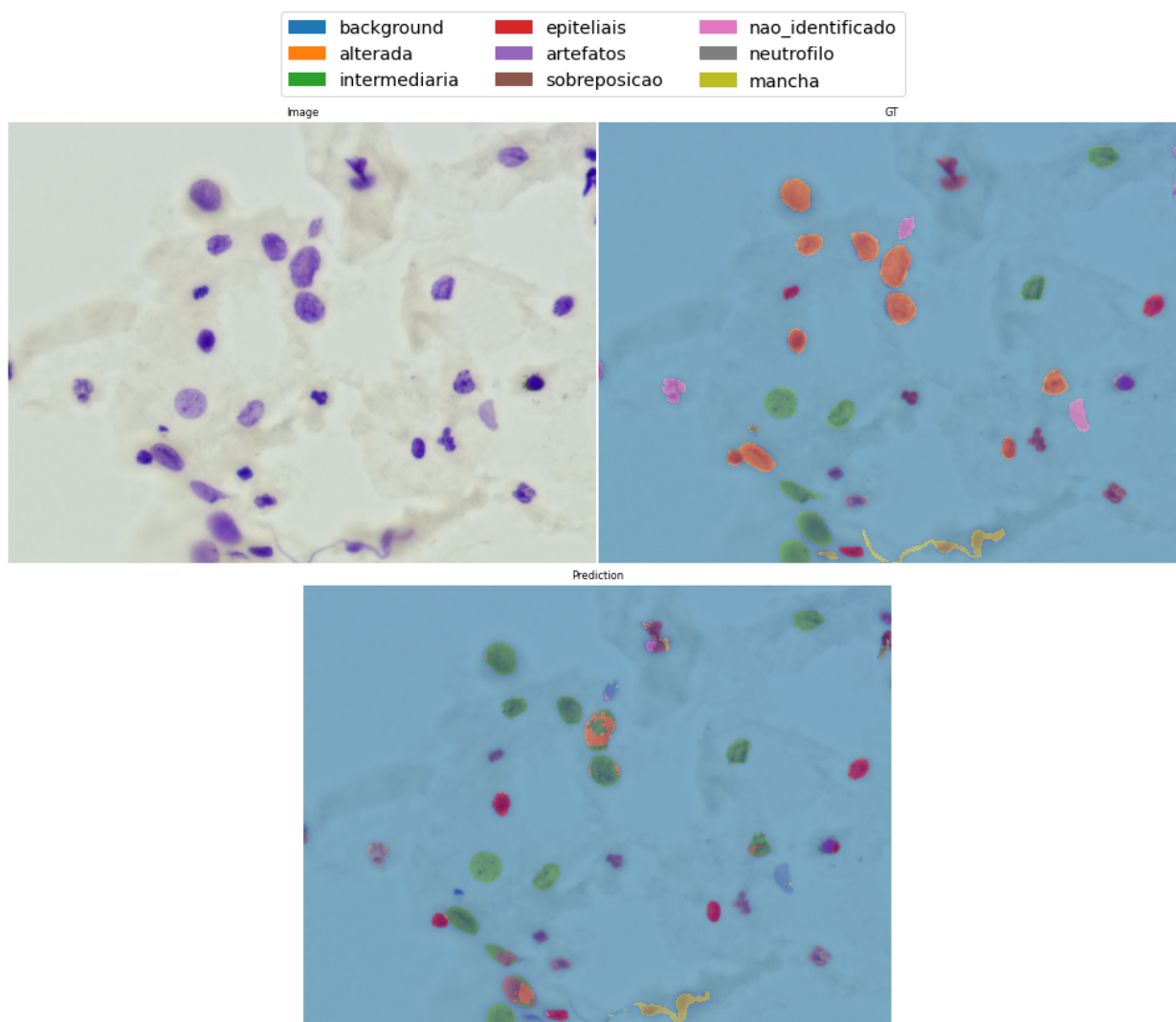


Fonte – Do autor.

Na Figura 12 são apresentados alguns dos resultados obtidos com a U-Net e a

ResNet50 como *backbone*.

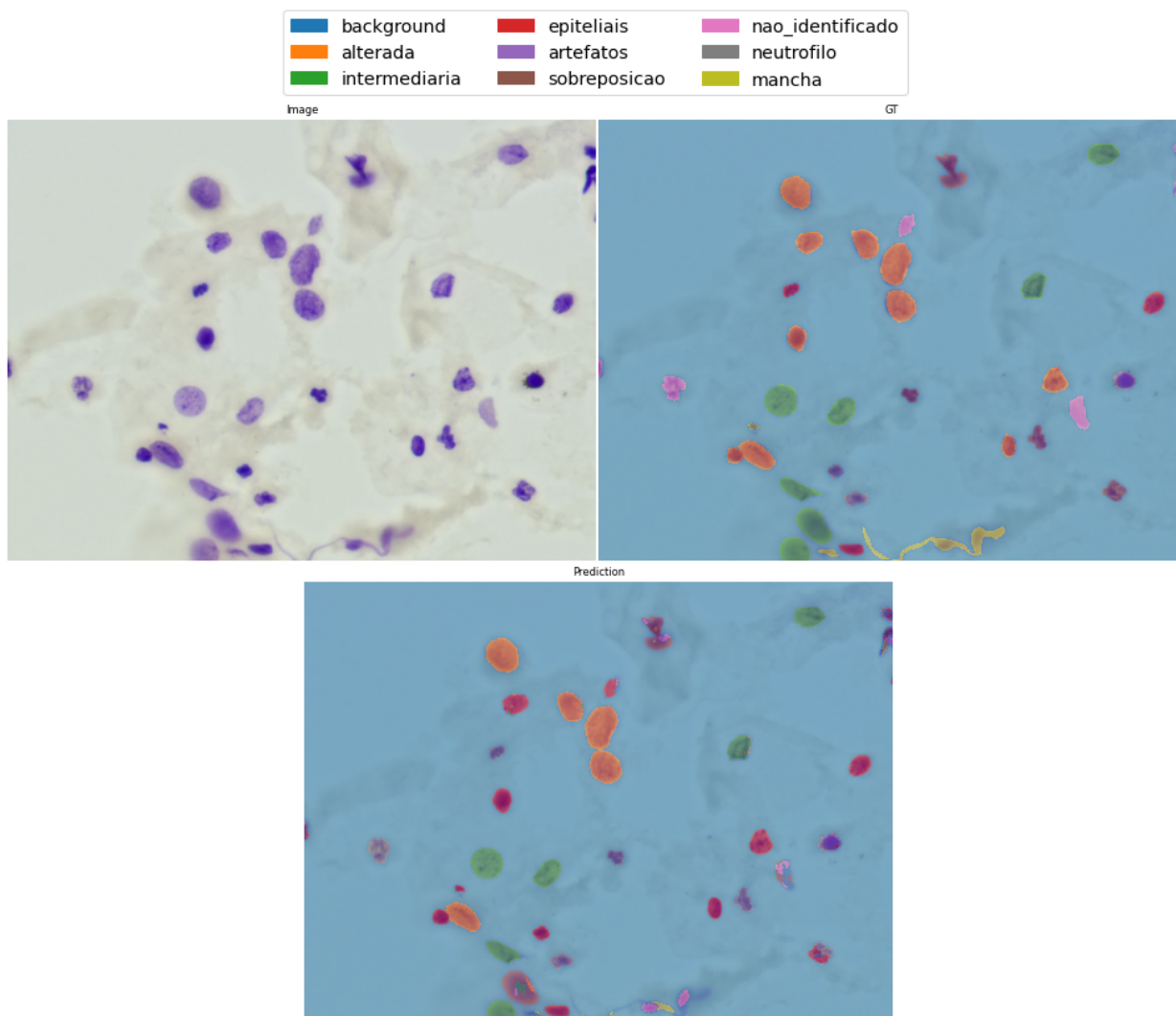
Figura 12 – Resultado das Segmentações para a U-Net + ResNet34. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

Alguns resultados obtidos com a *Weighted* U-Net + ResNet18 são apresentados na Figura 13.

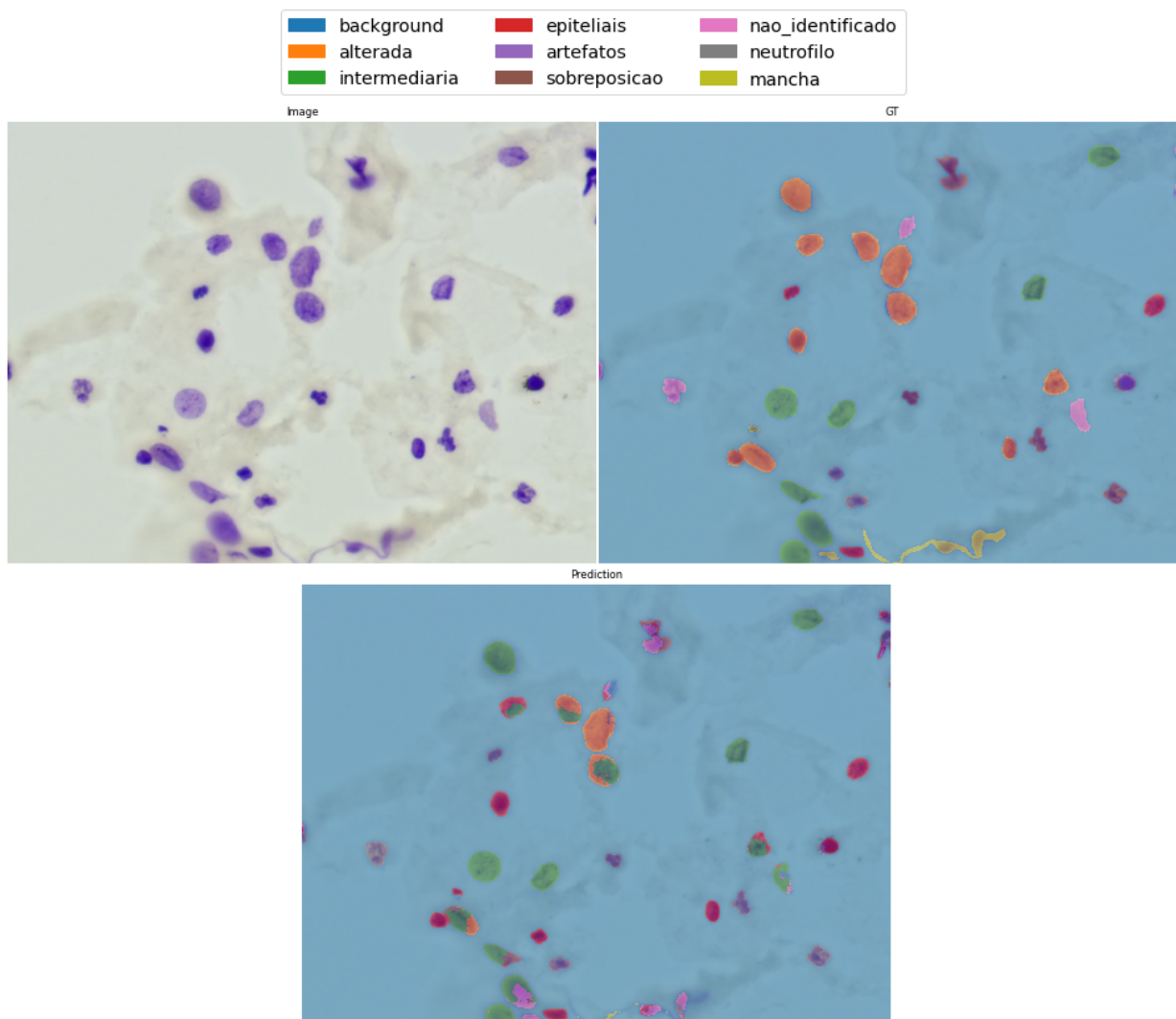
Figura 13 – Resultado das Segmentações para a *Weighted* U-Net + ResNet18. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

A Figura 14 apresenta alguns resultados obtidos com a *Weighted* U-Net e a ResNet50.

Figura 14 – Resultado das Segmentações para a *Weighted* U-Net + ResNet34. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.

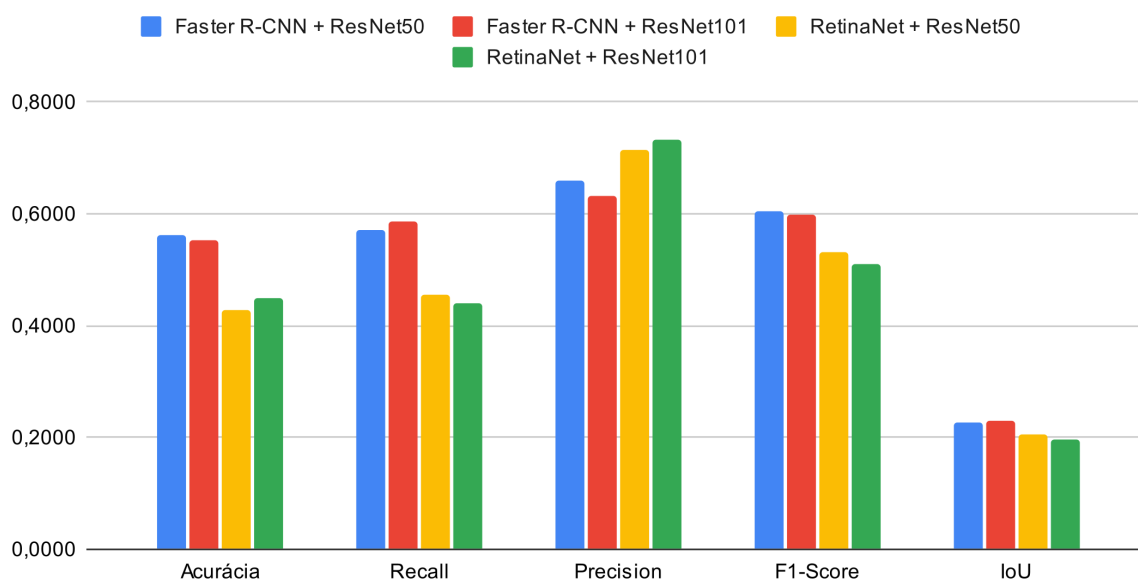


Fonte – Do autor.

4.2 DETECÇÃO DE OBJETOS

Os quatro modelos de redes neurais convolucionais para detecção de objetos testadas neste trabalho foram: Faster R-CNN + ResNet50, Faster R-CNN + ResNet101, RetinaNet + ResNet50 e RetinaNet + ResNet101. Utilizou-se o *batch size* de 4, selecionado por questões técnicas. O valor do *learning rate* foi $2,5 \times 10^{-4}$ e cada rede neural foi treinada por 15000 iterações. A Figura 15 apresenta as métricas médias para as quatro redes neurais supracitadas.

Figura 15 – Métricas médias para as redes neurais de detecção de objetos.



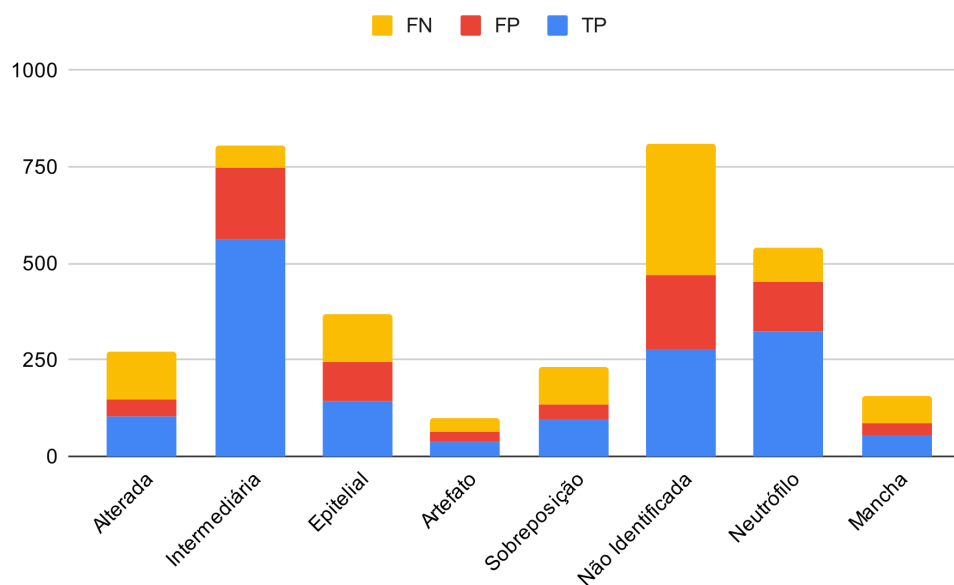
Fonte – Do autor.

O primeiro ponto que deve ser levado em conta é o fato de que o extrator de características Faster R-CNN teve um desempenho superior em relação as RetinaNets em quase todas as métricas, exceto *Precision*. Em outras palavras, a Faster R-CNN apresentou um desempenho geral superior quando comparado com a RetinaNet. Desta maneira, as RetinaNets apresentaram melhor desempenho no momento de classificar as instâncias relevantes dentro daquelas detectadas.

As Faster R-CNNs tiveram um desempenho similar quando comparamos a ResNet50 e a ResNet101. A ResNet50 apresentou um desempenho melhor em relação a *Acurácia*, *Precision* e *F-Score*. Já a ResNet101 apresentou melhores resultados em relação ao *Recall*, mostrando que consegue ter uma assertividade maior dentre as classes. Porém, o valor de *Precision* foi menor, mostrando que há também uma incidência maior de falsos positivos. O IoU foi mais alto com a ResNet101 quando comparada à ResNet50. Além da *Acurácia*, a Faster R-CNN + ResNet50 apresentou melhor resultado em relação ao *F-Score*, mostrando que generaliza melhor nos dados do conjunto de teste.

A Figura 16 apresenta os valores de Falsos Negativos, Falsos Positivos e Verdadeiros Positivos, para cada classe, referente a Faster R-CNN + ResNet50, a rede neural que apresentou o melhor resultado geral.

Figura 16 – Falsos Negativos (FN), Falsos Positivos (FP) e Verdadeiros Positivos (TP) - Faster R-CNN + ResNet50.



Fonte – Do autor.

Observando a Figura 16 foi possível verificar que a classe Intermediária apresentou a maior quantidade de verdadeiros positivos, e, proporcionalmente, a menor quantidade de falsos negativos. Isso se deve ao fato de a mesma possuir a maior quantidade de amostras (excluindo-se a classe "Não Identificada") dentro do conjunto de dados. Se for levada em conta a classe "Não Identificada", que possui a maior quantidade de amostras do *dataset*, a proporção de falsos negativos é maior quando comparada à classe Intermediária. Isso se dá pelo fato de que estes tipos de núcleo são difíceis de serem classificados visualmente (ou estão fora de foco). Em contrapartida tem-se a classe Neutrófilo que, apesar de possuir uma quantidade menor de amostras, tem uma taxa de falsos negativos e falsos positivos menores quando comparada a classe "Não Identificada". Isso se deve ao fato de que um neutrófilo tem um formato característico, sendo mais fácil distingui-los visualmente dos outros tipos de núcleos.

4.2.1 Resultados por classe

Os resultados (separados por classe) obtidos com a Faster R-CNN e a ResNet50 como *backbone* são apresentados na Tabela 15.

Tabela 15 – Métricas - Detecção de Objetos com Faster R-CNN + ResNet50

Classe	IoU	Acurácia	Recall	Precision	F-Score
Alterada	0.1817	0.3778	0.4595	0.6800	0.5484
Intermediária	0.4963	0.6985	0.9037	0.7547	0.8225
Epitelial	0.2126	0.3913	0.5353	0.5926	0.5625
Artefato	0.1094	0.3700	0.5000	0.5873	0.5401
Sobreposição	0.1388	0.4138	0.4923	0.7218	0.5854
Não Identificada	0.1459	0.3424	0.4511	0.5869	0.5101
Neutrófilo	0.4031	0.6048	0.7893	0.7212	0.7537
Mancha	0.1375	0.3544	0.4444	0.6364	0.5234

Na Tabela 16 são mostrados os resultados, por classe, para a Faster R-CNN + ResNet101.

Tabela 16 – Métricas - Detecção de Objetos com Faster R-CNN + ResNet101

Classe	IoU	Acurácia	Recall	Precision	F-Score
Alterada	0.2055	0.3622	0.6396	0.4551	0.5318
Intermediária	0.4507	0.6417	0.7159	0.8610	0.7818
Epitelial	0.1604	0.3292	0.3978	0.6564	0.4954
Artefato	0.1096	0.3461	0.4800	0.5538	0.5143
Sobreposição	0.1563	0.4380	0.5436	0.6928	0.6092
Não Identificada	0.1989	0.4039	0.6091	0.5452	0.5754
Neutrófilo	0.4048	0.6038	0.7603	0.7458	0.7530
Mancha	0.1431	0.3681	0.5360	0.5403	0.5381

Os resultados obtidos com a RetinaNet + ResNet50 são apresentados na Tabela 17.

Tabela 17 – Métricas - Detecção de Objetos com RetinaNet + ResNet50

Classe	IoU	Acurácia	Recall	Precision	F-Score
Alterada	0.2077	0.3618	0.5721	0.4961	0.5314
Intermediária	0.4078	0.5414	0.5666	0.9241	0.7025
Epitelial	0.1034	0.2166	0.2230	0.8823	0.3561
Artefato	0.1102	0.3563	0.4246	0.6889	0.5254
Sobreposição	0.1537	0.4326	0.5436	0.6795	0.6040
Não Identificada	0.1142	0.1861	0.2059	0.6597	0.3138
Neutrófilo	0.3999	0.5907	0.6780	0.8211	0.7427
Mancha	0.1394	0.3151	0.4228	0.5532	0.4793

Já na Tabela 18 são apresentados os resultados para a combinação RetinaNet + ResNet101.

Tabela 18 – Métricas - Detecção de Objetos com RetinaNet + ResNet101

Classe	IoU	Acurácia	Recall	Precision	F-Score
Alterada	0.1939	0.4007	0.5360	0.6134	0.5721
Intermediária	0.5072	0.6984	0.8363	0.8090	0.8224
Epitelial	0.0620	0.1465	0.1487	0.9091	0.2556
Artefato	0.0918	0.2588	0.2973	0.6667	0.4112
Sobreposição	0.1379	0.4167	0.4872	0.7422	0.5882
Não Identificada	0.0462	0.0694	0.0723	0.6286	0.1298
Neutrófilo	0.4009	0.5912	0.6844	0.8127	0.7431
Mancha	0.1428	0.3775	0.4634	0.6706	0.5481

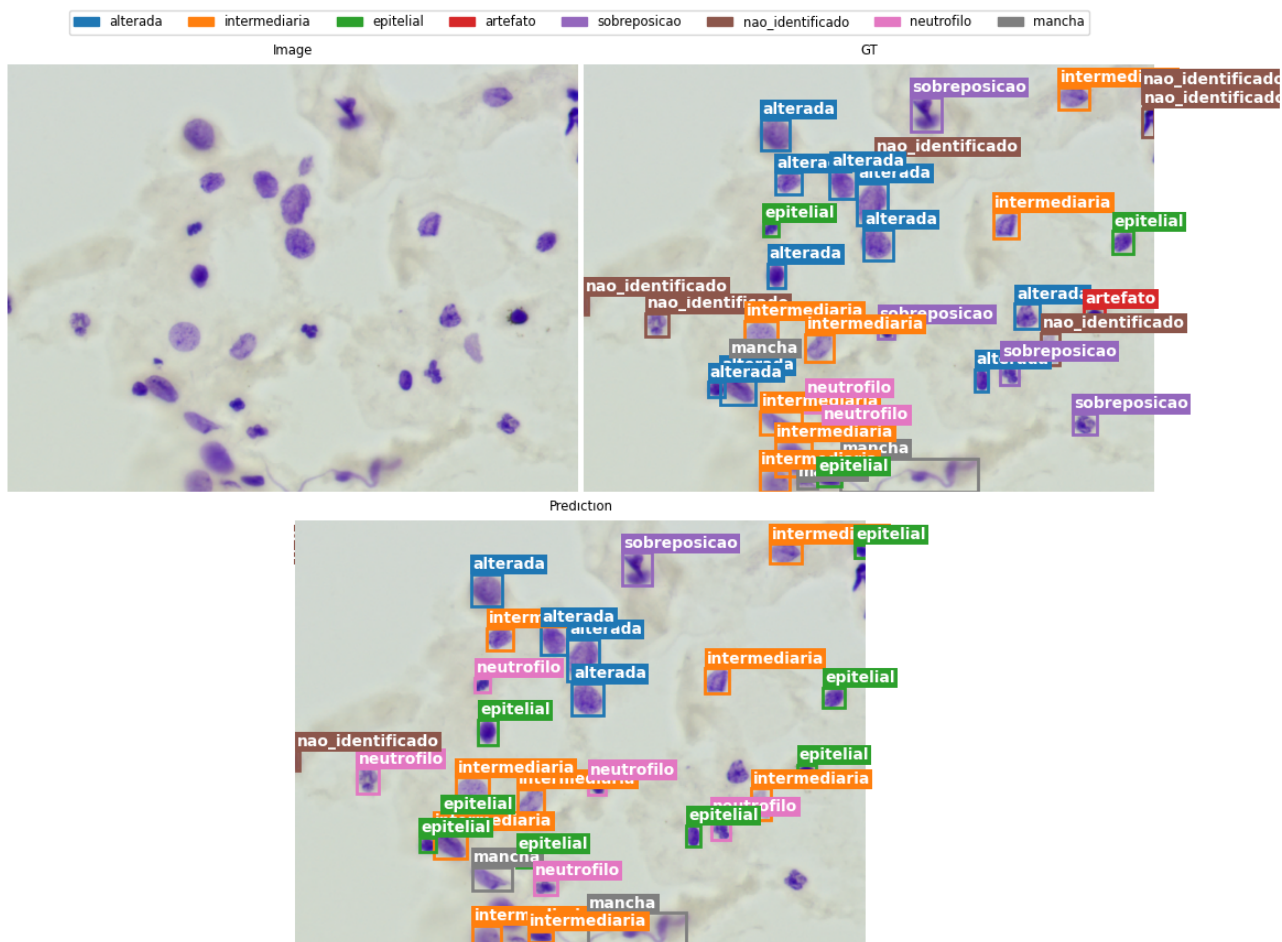
Os maiores valores de IoU, para as quatro redes neurais testadas nesta abordagem, foram obtidos nas classes Intermediária e Neutrófilo. Isso se deve ao fato de estas possuírem as maiores quantidades de amostras dentro do *dataset* (se excluirmos a classe Não Identificada). Em contrapartida, a classe Artefato, apesar de ser facilmente identificada (visualmente), não apresentou bons resultados, justamente pelo número limitado de amostras. Comparando-se as classes Sobreposição e Mancha (que são visualmente parecidas), as redes neurais apresentaram resultados similares.

A Faster R-CNN + ResNet50 apresentou os melhores resultados para as classes Intermediária e Neutrófilo (as duas com maior quantidade de amostras) em relação ao *Recall*. A mesma rede também apresentou o valor mais alto para esta métrica com a classe Artefato, aquela que possui a menor quantidade de amostras. Em contrapartida, a ResNet101 apresentou um valor de *Recall* maior na classe "Não Identificada". Já em relação às RetinaNets, o cenário foi um pouco diferente. Tanto para os núcleos Intermediários quanto para os Neutrófilos, a rede neural maior (ResNet101) apresentou os melhores resultados. Porém, a RetinaNet + ResNet50 apresentou o melhor *Recall* para os núcleos Não Identificados.

Em termos de *Precision* a Faster R-CNN + ResNet101, quando comparada à rede menor (Faster R-CNN + ResNet50) apresentou melhores resultados para as classes Intermediária e Neutrófilo. Fazendo uma comparação entre as RetinaNets para as mesmas classes, aquela com ResNet50 apresentou resultados melhores quando comparada à ResNet101. Os maiores valores de *F-Score* foram obtidos (com uma diferença de 0,0001 entre eles) com a Faster R-CNN + ResNet50 e a RetinaNet + ResNet101, ambas na classe Intermediária. Isso dá um indicativo de que essas redes conseguem generalizar melhor em um conjunto de imagens inédito, principalmente para classes que possuem uma quantidade maior de amostras.

Na Figura 17 são apresentados alguns resultados obtidos com a Faster R-CNN + ResNet50.

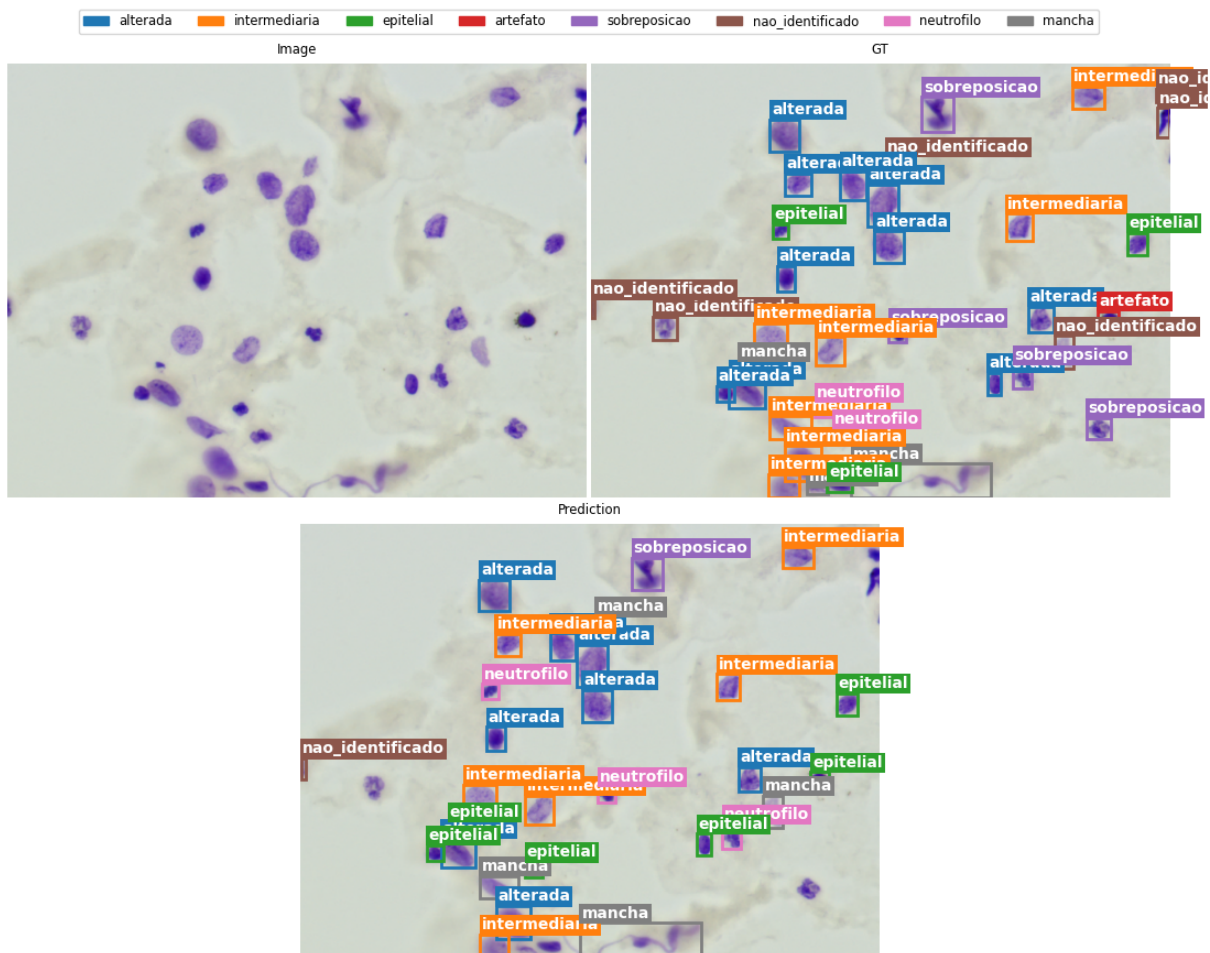
Figura 17 – Resultado das detecções para a Faster R-CNN + ResNet50. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

Já a Figura 18 mostra os resultados para a Faster R-CNN, utilizando a ResNet101 como *backbone*.

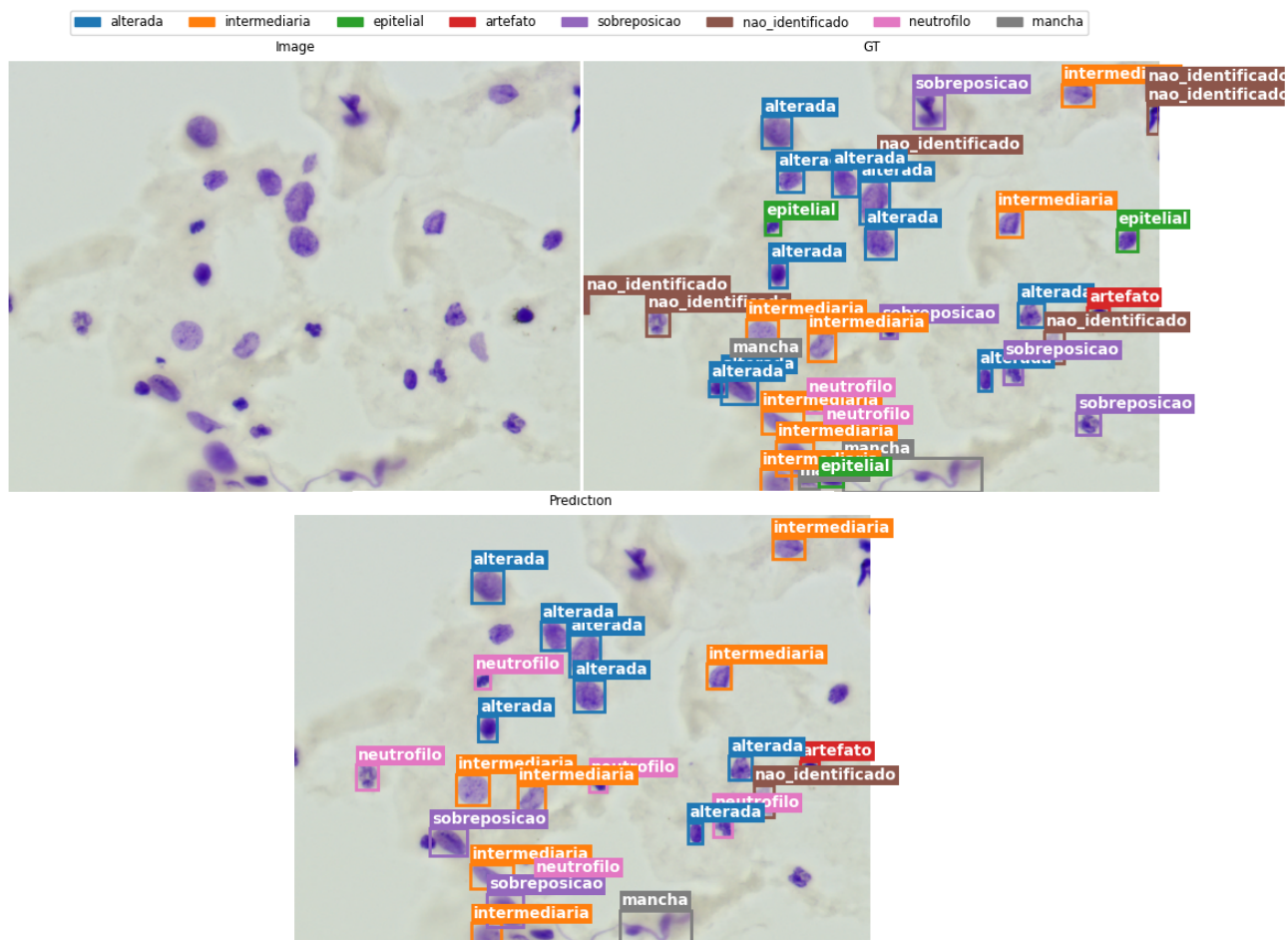
Figura 18 – Resultado das detecções para a Faster R-CNN + ResNet101. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

Exemplos de resultados obtidos com a RetinaNet e ResNet50 são apresentados na Figura 19.

Figura 19 – Resultado das detecções para a RetinaNet + ResNet50. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

Na Figura 20 são apresentados alguns resultados gerados com a RetinaNet + ResNet101.

Figura 20 – Resultado das detecções para a RetinaNet + ResNet101. Linha superior - esquerda: imagem de entrada; Linha superior - direita: *ground-truth*; Linha inferior: predição da rede neural.



Fonte – Do autor.

4.3 CLASSIFICAÇÃO DE IMAGENS

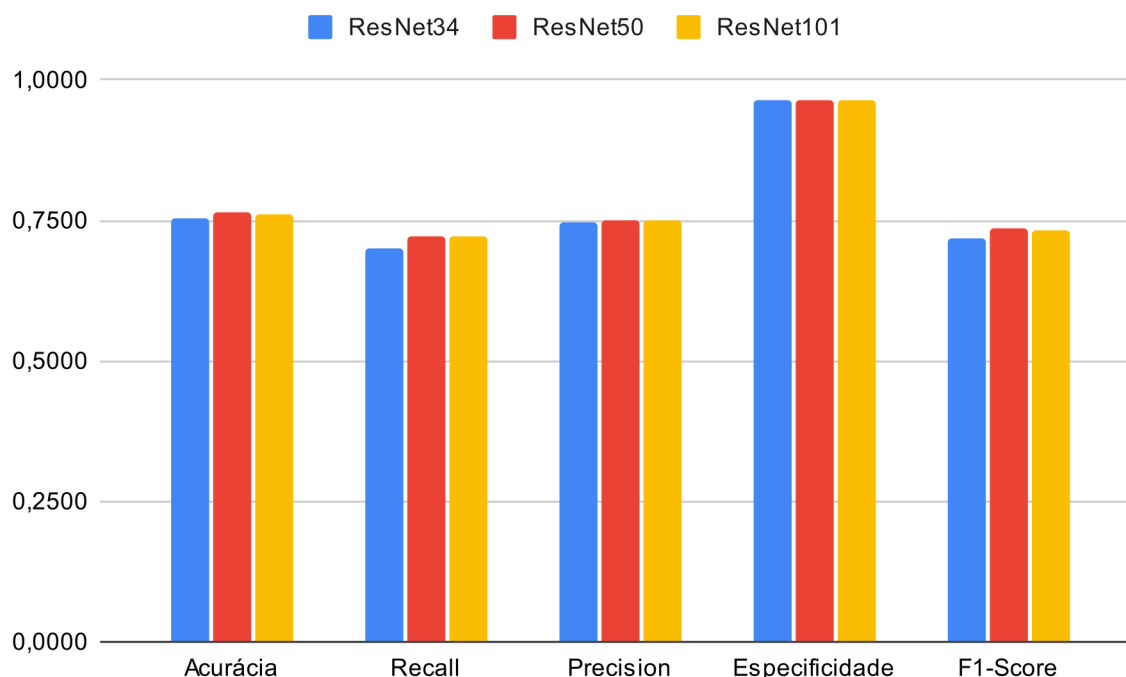
Para os experimentos de classificação de imagens foram testadas três arquiteturas de redes neurais: a ResNet34, ResNet50 e a ResNet101. Utilizou-se o valor de *weight decay* de 1×10^{-2} . A Tabela 19 apresenta os *learning rates* para cada etapa do treinamento.

Tabela 19 – *Learning rates* para as redes neurais de classificação de imagens.

Classe	Etapa 1	Etapa 2
ResNet34	$1e^{-2}$	$1e^{-7}$
ResNet50	$1e^{-2}$	$1e^{-6}$ a $1e^{-5}$
ResNet101	$1e^{-2}$	$1e^{-7}$

A Figura 21 apresenta as métricas para os três modelos de rede neurais testadas.

Figura 21 – Métricas obtidas com as redes neurais de classificação de imagem.



Fonte – Do autor.

Observando os gráficos da Figura 21 podemos verificar que a ResNet50 apresentou os melhores resultados gerais entre os três modelos testados. Em outras palavras, para este experimento, a rede neural com tamanho intermediário (se comparada aos outros dois modelos) apresentou as melhores métricas. Assim como visto em outros experimentos, nem sempre a maior das redes neurais vai se apresentar como a melhor solução, em relação a assertividade, para o problema. Os resultados podem variar com a complexidade dos dados. Observando os resultados obtidos com a ResNet34, é possível verificar que foram levemente inferiores em relação as outras duas arquiteturas maiores. Em um cenário onde a aplicação precise ser executada em tempo real ela pode ser uma boa escolha.

4.3.1 Resultados por classe

A Tabela 20 apresenta os resultados obtidos com a ResNet34.

Tabela 20 – Métricas - Classificação de Imagens - ResNet34

Classe	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.9338	0.4595	0.6800	0.9793	0.5484
Intermediária	0.9283	0.8876	0.8316	0.9415	0.8587
Epitelial	0.9200	0.6059	0.6269	0.9572	0.6162
Artefato	0.9901	0.8133	0.8472	0.9955	0.8299
Sobreposição	0.9507	0.5487	0.7430	0.9842	0.6313
Não Identificada	0.8865	0.8176	0.7404	0.9085	0.7771
Neutrófilo	0.9271	0.8353	0.7467	0.9449	0.7886
Mancha	0.9724	0.6429	0.7641	0.9896	0.6983

Os resultados obtidos com a ResNet50 são apresentados na Tabela 21.

Tabela 21 – Métricas - Classificação de Imagens - ResNet50

Classe	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.9334	0.5360	0.6432	0.9715	0.5848
Intermediária	0.9235	0.8876	0.8168	0.9352	0.8508
Epitelial	0.9231	0.5539	0.6652	0.9669	0.6045
Artefato	0.9901	0.8267	0.8378	0.9951	0.8322
Sobreposição	0.9531	0.6205	0.7289	0.9808	0.6704
Não Identificada	0.8916	0.7948	0.7661	0.9225	0.7802
Neutrófilo	0.9361	0.8596	0.7734	0.9510	0.8142
Mancha	0.9752	0.7063	0.7739	0.9892	0.7386

Já na Tabela 22 são apresentadas as métricas obtidas com a ResNet101.

Tabela 22 – Métricas - Classificação de Imagens - ResNet101

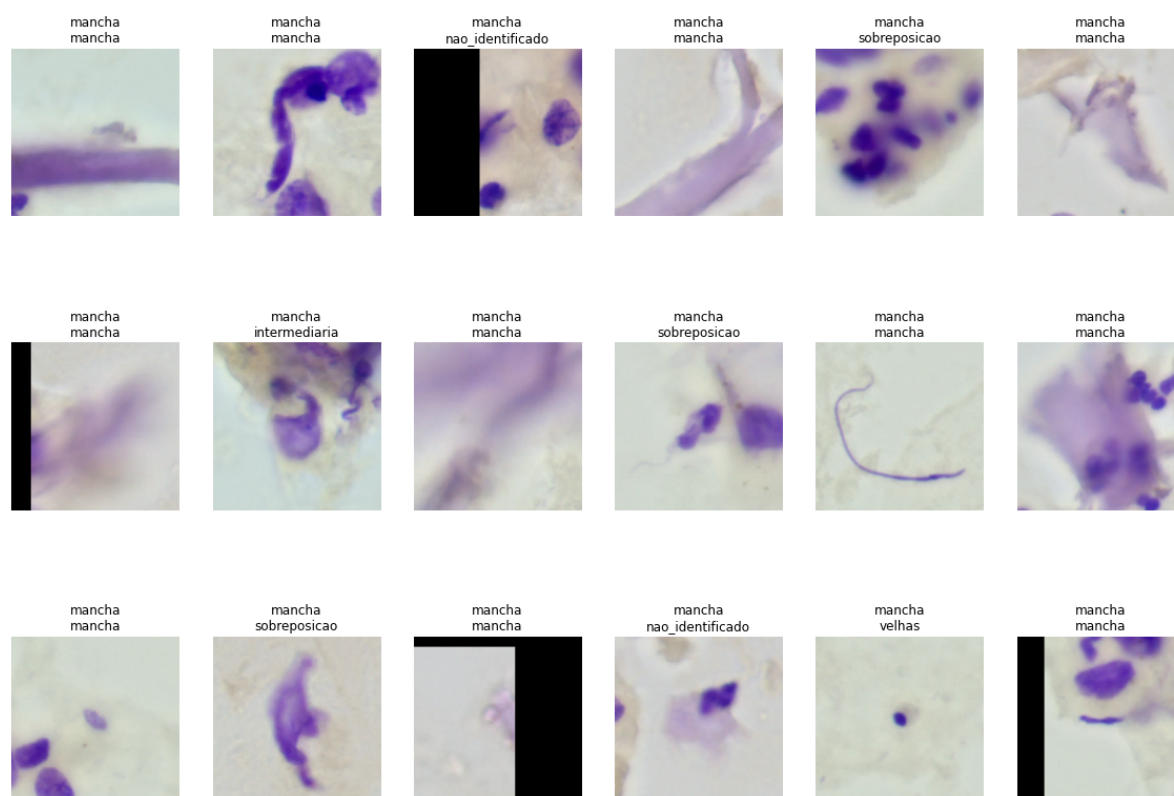
Classe	Acurácia	Recall	Precision	Especificidade	F-Score
Alterada	0.9342	0.5901	0.6328	0.9672	0.6107
Intermediária	0.9259	0.8620	0.8404	0.9467	0.8510
Epitelial	0.9283	0.5762	0.6951	0.9700	0.6301
Artefato	0.9913	0.8000	0.8955	0.9972	0.8451
Sobreposição	0.9495	0.6410	0.6831	0.9752	0.6614
Não Identificada	0.8857	0.8225	0.7361	0.9059	0.7769
Neutrófilo	0.9381	0.8039	0.8137	0.9642	0.8088
Mancha	0.9692	0.6825	0.6935	0.9842	0.6880

Levando em conta a Acurácia e fazendo uma comparação classe a classe, pode-se observar que não houveram melhoras significativas processando os dados com as redes maiores. O melhor resultado para determinada classe variou entre os três modelos, onde a ResNet50 apresentou a melhor acurácia geral. Porém, se observarmos a classe Intermediária, aquela que possui mais amostras, a ResNet34 teve um desempenho melhor (em relação a Acurácia). A classe Não Identificada apresentou a menor acurácia para todos os três modelos de rede neural.

Os valores de Especificidade foram altos para todas as classes dentre as três arquiteturas. Isso mostra que houve uma quantidade baixa de falsos negativos. O *F-Score* para a classe Intermediária variou pouco dentre as três redes. O resultado para a métrica *Precision* foi similar, mostrando bons resultados para classes como a Intermediária, que possui grande quantidade de amostras, ou Sujeira, que pode ser facilmente distinguida (visualmente) dentre as classes. De maneira semelhante às métricas citadas anteriormente, a variação para o *Recall* também foi grande. Essas variações nas métricas acaba evidenciando o efeito do desbalanceamento entre as classes dentro do conjunto de dados.

Na Figura 22 são apresentados alguns resultados para classificação de imagens utilizando a ResNet34.

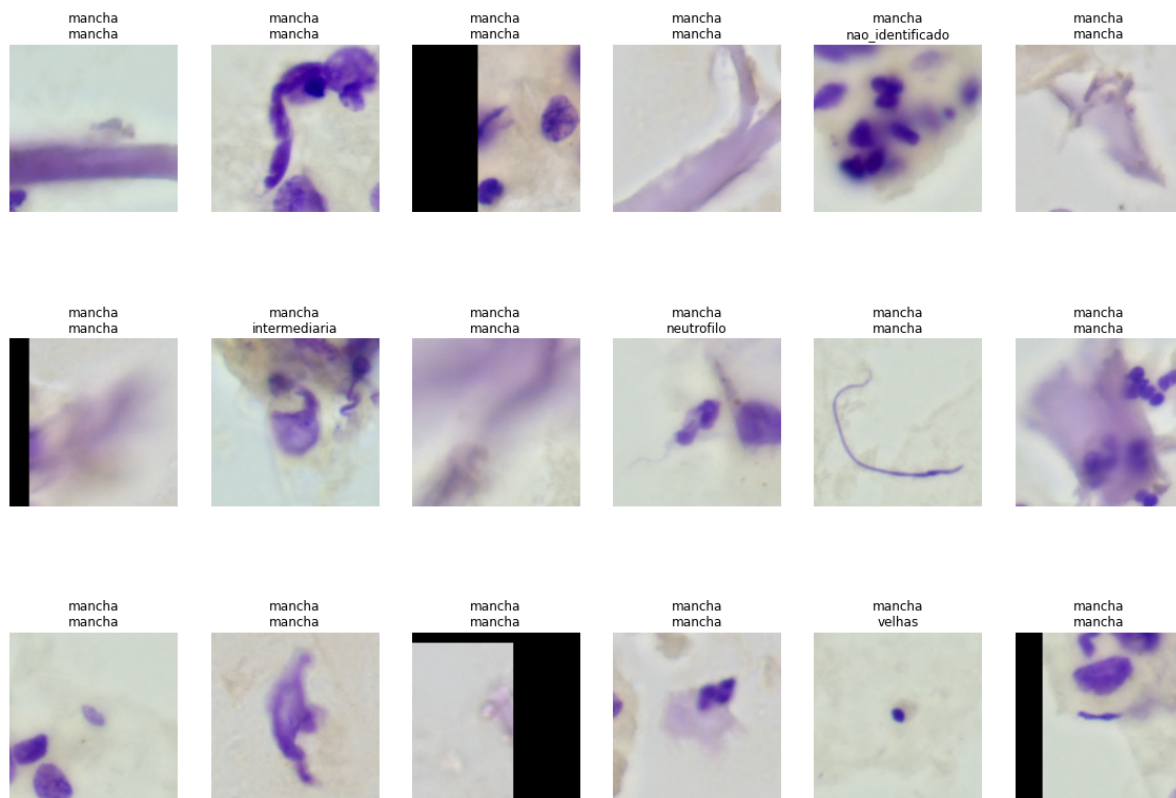
Figura 22 – Exemplos de resultados para classificação de imagens com a ResNet34. Na legenda, a classe na linha superior representa o *ground-truth* e na inferior a predição da rede neural.



Fonte – Do autor.

A Figura 23 apresenta alguns exemplos visuais de resultado obtidos com a ResNet50.

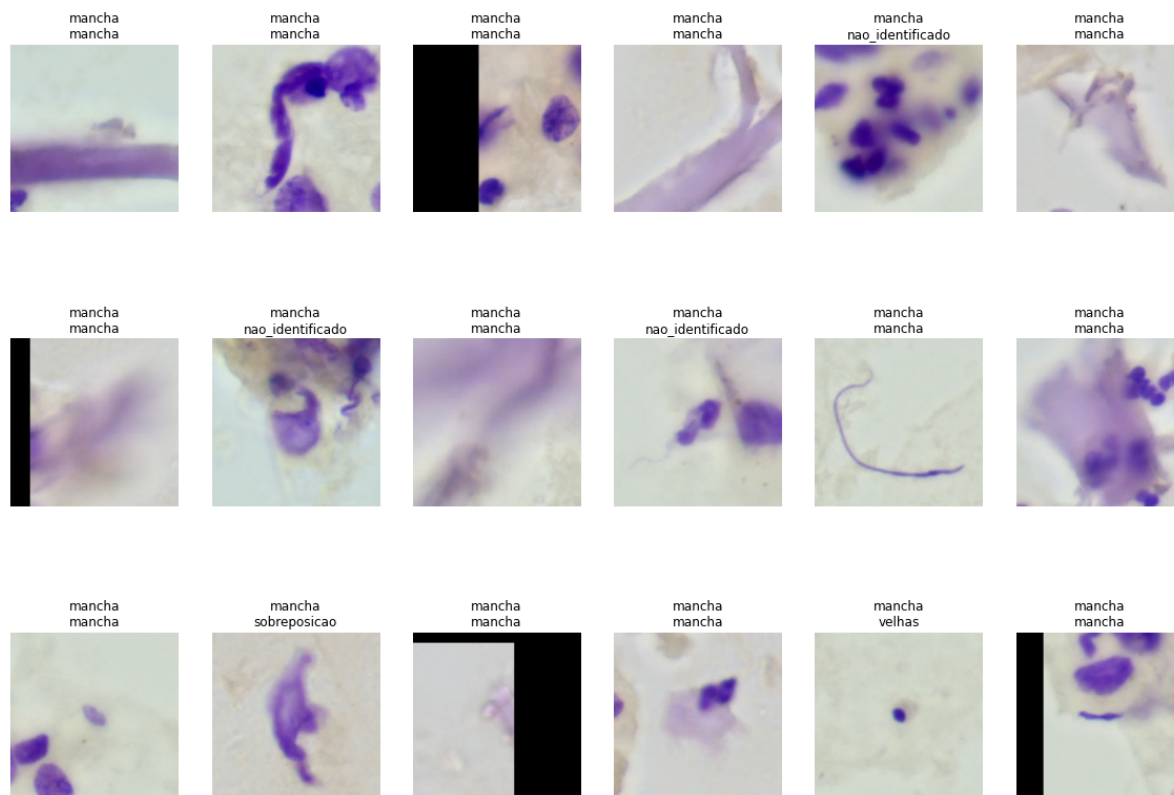
Figura 23 – Exemplos de resultados para classificação de imagens com a ResNet50. Na legenda, a classe na linha superior representa o *ground-truth* e na inferior a predição da rede neural.



Fonte – Do autor.

Os resultados apresentados na Figura 24 foram obtidos com a ResNet101.

Figura 24 – Exemplos de resultados para classificação de imagens com a ResNet101. Na legenda, a classe na linha superior representa o *ground-truth* e na inferior a predição da rede neural.



Fonte – Do autor.

4.4 ARTIGOS PUBLICADOS

Durante a execução deste trabalho, alguns dos experimentos foram publicados como artigos. Os dois trabalhos diretamente ligados a este projeto de mestrado são citados abaixo:

- MATIAS, André Victória et al. What is the State of the Art of Computer Vision-Assisted Cytology? A Systematic Literature Review. **Computerized Medical Imaging and Graphics**, p. 101934, 2021.
- MACARINI, L. A. B.; VON WANGENHEIM, A.; DALTOE, F. P.; ONOFRE, A. S. C.; ONOFRE, F. B. M.; STEMMER, M. R. Towards a Complete Pipeline for Segmenting Nuclei in Feulgen-Stained Images. XI Computer on the Beach, Balneário Camboriú. **Anais do XI Computer on the Beach**. 2020.

Outros trabalhos foram publicados utilizando uma metodologia similar, mas com corantes diferentes. São eles:

- MATIAS, A. V.; CERENTINI, A.; MACARINI, L. A. B.; AMORIM, J. G. A.; DALTOE, F. P.; VON WANGENHEIM, A. Segmentation, Detection and Classification of Cell Nuclei on Oral Cytology Samples Stained with Papanicolaou. **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. IEEE, 2020. p. 552-557.
- AMORIM, J. G. A., MACARINI, L. A. B., MATIAS, A. V., CERENTINI, A., ONOFRE, F. B. D. M., ONOFRE, A. S. C., VON WANGENHEIM, A. A Novel Approach on Segmentation of AgNOR-Stained Cytology Images Using Deep Learning. **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. IEEE, 2020. p. 552-557.
- MATIAS, A. V., CERENTINI, A., MACARINI, L. A. B., AMORIM, J. G. A., PEREIRA, R., DALTOÉ, F. P., VON WANGENHEIM, A. Localização de Núcleos Celulares em Citologia Oral usando Métodos de Deep Learning. **Journal of Health Informatics**, v. 12, 2021.
- ATKINSON, J. G.; MACARINI, L. A. B.; MATIAS, A. V.; CERENTINI, A.; ONOFRE, F. B. M.; ONOFRE, A. S. C.; VON WANGENHEIM, A. Detecção de núcleos em imagens citológicas de AgNOR utilizando Aprendizado Profundo. **Journal of Health Informatics**, v. 12, 2021.

Trabalhos aceitos para publicação:

- MACARINI, L. A. B.; SOBIERANSKI, A. C.; VON WANGENHEIM, A.; ONOFRE, A. S. C.; ONOFRE, F. B. D. M.; STEMMER, M. R. Analysis of Semantic Segmentation Approaches for the Morphometric Analysis of Feulgen-Stained Cytological Samples. **Analytical and Quantitative Cytopathology and Histopathology**. Aceito para publicação.

5 CONCLUSÃO

Neste trabalho foi apresentado um estudo comparativo entre diferentes abordagens envolvendo métodos de *Deep Learning* para análise de aneuploidia em núcleos corados com Feulgen. As três abordagens utilizadas neste trabalho foram: segmentação semântica, detecção de objetos e classificação de imagens. Foram testadas onze redes neurais diferentes, sendo quatro para segmentação semântica, quatro para detecção de objetos e três para classificação de imagens. Em todos os experimentos optou-se por analisar os resultados levando em conta as diferentes classes dos núcleos celulares (diferentemente de uma abordagem binária). As classes contidas no conjunto de dados são: *background*, núcleo de células Intermediárias, Alteradas, Epiteliais, Artefato, Sobreposição, Não Identificada, Neutrófilo e Mancha.

Observando as métricas, no geral, os resultados podem ser considerados bons quando comparados com outros trabalhos encontrados na literatura (MATIAS *et al.*, 2021). Porém, os métodos de segmentação semântica e detecção de objetos apresentaram métricas ruins. Em outras palavras, a segmentação e a detecção (localização) dos núcleos foram boas. Já a etapa de classificação apresentou resultados não tão bons, onde as redes neurais apresentaram dificuldades para diferenciar os núcleos entre as classes. A abordagem de classificação de imagem foi a que apresentou as melhores métricas de desempenho. Porém, tem como grande desvantagem o fato de a análise ser feita em um recorte da imagem, não apresentando localização dos núcleos.

Com os resultados obtidos, ainda não é possível aplicar a abordagem em uma rotina clínica, de maneira que o desempenho das redes neurais precisa ser melhorado. Para que essa abordagem possa ser, de fato, implementada, é necessário possuir um método confiável que seja capaz de detectar, segmentar ou classificar os núcleos de maneira correta, em qualquer que seja o cenário. Um ponto negativo do presente trabalho é o tamanho do conjunto de dados, já que os métodos de *deep learning* precisam de uma grande quantidade de dados para serem treinados. Além disso, o desbalanceamento entre as classes impactou negativamente os resultados. Isto pode explicar as variações nas métricas de desempenho entre classes.

Uma das maiores contribuições do presente trabalho é a comparação de abordagens e métodos para detecção precoce de aneuploidia. Como citado anteriormente, estes modelos de redes neurais ainda não estão prontos para serem aplicados em uma rotina clínica. Porém, acredita-se que este trabalho poderá servir como um primeiro passo na investigação de possíveis métodos para detecção precoce (não só utilizando o corante de Feulgen) de aneuploidia de maneira não invasiva. Pode-se citar como outra contribuição a publicação do *dataset*. Observou-se que não existem dados similares aos utilizados neste trabalho e que estejam disponíveis de maneira pública. Este fato faz com que seja difícil a comparação dos resultados obtidos no presente

estudo com diferentes abordagens apresentadas na literatura. Isso porque os trabalhos analisados na SLR utilizam diferentes métodos de pré-processamento, técnicas, métricas, modelos, parâmetros e conjuntos de dados (MATIAS *et al.*, 2021).

O tempo de treinamento pode ser considerado bom quando comparado com outros trabalhos encontrados. O tempo necessário para exportar e carregar os modelos, e processar as imagens é pequeno. Também é preciso citar que neste caso as imagens não precisam ser processadas em tempo real. Uma das desvantagens dos métodos de *deep learning* é a necessidade de uma quantidade significativa de dados para treinar os modelos e fazer com que eles generalizem bem fora do conjunto de treino (O'MAHONY *et al.*, 2019). Além disso, é necessário uma grande quantidade de horas para anotar os núcleos, sendo este um dos principais gargalos do projeto. Ainda, para este caso, a pessoa responsável pela anotação de dados precisa ser um especialista em citologia.

Mesmo que os métodos de *deep learning* precisem de uma grande quantidade de tempo para treinar, principalmente se comparado aos métodos clássicos, acredita-se que é mais vantajoso utilizar este tipo de método ao invés dos métodos clássicos (pelo menos para esta aplicação). É importante considerar que as redes neurais convolucionais funcionam como um extrator geral de características. Além disso, desenvolver um algoritmo para extrair características para esta aplicação pode ser um processo difícil levando em conta a natureza e a complexidade do problema. Desta maneira, apesar de o tempo de treinamento ser alto, este processo precisa ser feito apenas uma vez.

Levando em conta todos os pontos abordados até o momento, acredita-se que o objetivo principal de apresentar um *pipeline* completo para detecção precoce e não-invasiva de aneuploidia foi atingido. Mesmo que os resultados obtidos com duas das três abordagens não tenham sido bons, a classificação de imagem apresentou boas métricas. Sabe-se que uma das principais limitações deste trabalho é a quantidade de amostras utilizadas, além da pequena representatividade das mesmas ($n = 5$). Porém, o grupo de pesquisa possui acesso a mais exames deste tipo. Desta maneira, desde que devidamente anotados, esses exames podem ser utilizados para complementar o conjunto de dados visando melhorar os resultados obtidos com a presente abordagem.

5.1 TRABALHOS FUTUROS

Para dar andamento ao presente trabalho, um ponto crucial será aumentar a quantidade de amostras do conjunto de dados, tanto em relação a quantidade de imagens quanto em amostras de pacientes. Acredita-se que esta é a principal limitação do trabalho. Este processo, porém, é custoso, pois a anotação das imagens demanda muito tempo e, neste caso, um profissional treinado. A área de *deep learning* está em pleno crescimento, havendo o desenvolvimento e a publicação de novos modelos

de CNNs de maneira muito rápida. Sendo assim, pode-se testar diversos modelos de redes neurais encontrados na literatura, como a HRNet (WANG, J. *et al.*, 2019), por exemplo. Além disso, diferentes abordagens podem ser utilizadas. Uma delas é a *Weakly Supervised Segmentation*, citada neste trabalho (Capítulo 2). Métodos clássicos de segmentação semântica podem ser utilizados para realizar uma comparação com os métodos de *deep learning*. Exemplos são o *Watershed*, *Mumford-Shah*, Filtro de Difusão Anisotrópica, etc. Métodos de pré- e pós-processamento como o ajuste de brilho, de contraste, entre outros, também podem ser utilizados para realizar o melhoramento das imagens. Outros tipos de corantes também podem ser usados. Através da SLR verificou-se a utilização recorrente do corante de Papanicolau, sendo este um exemplo de experimento futuro. Outro corante é o *Argyropholic Nucleolar Organizer Region* (AgNOR), um método histoquímico que destaca as chamadas regiões organizadoras de núcleo (NOR).

REFERÊNCIAS

AMJOURD, Ayoub Benali; AMROUCH, Mustapha. Convolutional neural networks backbones for object detection. *In: SPRINGER. INTERNATIONAL Conference on Image and Signal Processing. [S.l.: s.n.], 2020. P. 282–289.*

AMORIM, João Gustavo; CERENTINI, Allan; MACARINI, Luiz Antonio Buschetto; MATIAS, André Victória; WANGENHEIM, Aldo von. **Systematic Literature Review of Computer Vision-Aided Cytology – A Review of Classic Computer Vision and Deep Learning-Based Approaches published between January/2016 – March/2020.** [S.l.], 2020.

ANUP, Basu; XIAOBO, Li. **Computer Vision: Systems, Theory And Applications: Selected Papers From Vision Interface 1992.** [S.l.]: World Scientific, 1993. v. 38.

BEDROSSIAN, Carlos WM. Cytopathology: A global perspective. **Diagnostic cytopathology**, Wiley Online Library, v. 35, n. 12, p. 745–748, 2007.

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation learning: A review and new perspectives. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 35, n. 8, p. 1798–1828, 2013.

BIESTERFELD, Stefan; BECKERS, Sascha; CADENAS, Maria Del Carmen Villa; SCHRAMM, Martin. Feulgen staining remains the gold standard for precise DNA image cytometry. **Anticancer research**, International Institute of Anticancer Research, v. 31, n. 1, p. 53–58, 2011.

BÖCKING, A; HILGARTH, M; AUFFERMANN, W; HACK-WERDIER, Ch; FISCHER-BECKER, D; VON KALKREUTH, G. DNA-cytometric diagnosis of prospective malignancy in borderline lesions of the uterine cervix. **Acta cytologica**, v. 30, n. 6, p. 608–615, 1986.

BÖCKING, Alfred; NGUYEN, Vu Quoc Huy. Diagnostic and prognostic use of DNA image cytometry in cervical squamous intraepithelial lesions and invasive carcinoma. **Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society**, Wiley Online Library, v. 102, n. 1, p. 41–54, 2004.

BÖCKING, Alfred H; FRIEDRICH, David; MEYER-EBRECHT, Dietrich; ZHU, Chenyan; FEIDER, Anna; BIESTERFELD, Stefan. Automated detection of cancer cells in

effusion specimens by DNA karyometry. **Cancer cytopathology**, Wiley Online Library, v. 127, n. 1, p. 18–25, 2019.

BOLLMANN, Reinhard; BOLLMANN, Magdolna; HENSON, Donald E; BODO, Miklos. DNA cytometry confirms the utility of the Bethesda system for the classification of Papanicolaou smears. **Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society**, Wiley Online Library, v. 93, n. 3, p. 222–228, 2001.

BURD, Eileen M. Human papillomavirus and cervical cancer. **Clinical microbiology reviews**, Am Soc Microbiol, v. 16, n. 1, p. 1–17, 2003.

CARVALHO, ACPLF; FACELI, K; LORENA, A; GAMA, J. Inteligência Artificial—uma abordagem de aprendizado de máquina. **Rio de Janeiro: LTC**, 2011.

CHATELAIN, R; SCHUNCK, T; SCHINDLER, EM; SCHINDLER, AE; BÖCKING, A. Diagnosis of prospective malignancy in koilocytic dysplasias of the cervix with DNA cytometry. **The Journal of reproductive medicine**, v. 34, n. 8, p. 505–510, 1989.

CHIECO, Pasquale; DERENZINI, Massimo. The Feulgen reaction 75 years on. **Histochemistry and cell biology**, Springer, v. 111, n. 5, p. 345–358, 1999.

CIRESAN, Dan Claudiu; MEIER, Ueli; MASCI, Jonathan; GAMBARDELLA, Luca Maria; SCHMIDHUBER, Jürgen. Flexible, high performance convolutional neural networks for image classification. *In*: TWENTY-SECOND international joint conference on artificial intelligence. [S.l.: s.n.], 2011.

COGLIANO, Vincent; BAAN, Robert; STRAIF, Kurt; GROSSE, Yann; SECRETAN, Béatrice; EL GHISSASSI, Fatiha. Carcinogenicity of human papillomaviruses. **The lancet oncology**, Elsevier, v. 6, n. 4, p. 204, 2005.

COHEN, Paul; JHINGRAN, Anjua; OAKNIN, Ana; DENNY, Lynette. Cervical cancer. **The Lancet**, Elsevier, v. 393, n. 10167, p. 169–182, 2019.

COPPIN, B. **Inteligência Artificial**. [S.l.]: LTC, 2010. ISBN 9788521617297.

COSTA, Ane Francyne. **Análise prospectiva da ploidia do DNA em atipias e lesões intraepiteliais do colo do útero**. 2018. Diss. (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis. Dissertação de Mestrado.

CRUM, W.R.; CAMARA, O.; HILL, D.L.G. Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. **IEEE Transactions on Medical Imaging**, Institute of Electrical e Electronics Engineers (IEEE), v. 25, n. 11, p. 1451–1461, nov. 2006. DOI: 10.1109/tmi.2006.880587. Disponível em: <https://doi.org/10.1109/tmi.2006.880587>.

DANIELSEN, Håvard E; PRADHAN, Manohar; NOVELLI, Marco. Revisiting tumour aneuploidy—the place of ploidy assessment in the molecular era. **Nature Reviews Clinical Oncology**, Nature Publishing Group, v. 13, n. 5, p. 291, 2016.

DHILLON, Anamika; VERMA, Gyanendra K. Convolutional neural network: a review of models, methodologies and applications to object detection. **Progress in Artificial Intelligence**, Springer, v. 9, n. 2, p. 85–112, 2020.

DIAMANTIS, A; MAGIORKINIS, E. Pioneers of exfoliative cytology in the 19th century: the predecessors of George Papanicolaou. **Cytopathology**, Wiley Online Library, v. 25, n. 4, p. 215–224, 2014.

EATON-ROSEN, Zach; BRAGMAN, Felix; OURSELIN, Sebastien; CARDOSO, M. Jorge. Improving Data Augmentation for Medical Image Segmentation. *In: 1ST Conference on Medical Imaging with Deep Learning (MIDL 2018)*. [S.l.: s.n.], 2018.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, Elsevier BV, v. 27, n. 8, p. 861–874, jun. 2006. DOI: 10.1016/j.patrec.2005.10.010. Disponível em: <https://doi.org/10.1016/j.patrec.2005.10.010>.

GAO, Ge; SMITH, David I. Human papillomavirus and the development of different cancers. **Cytogenetic and genome research**, Karger Publishers, v. 150, n. 3-4, p. 185–193, 2016.

GARDNER, Matt W; DORLING, SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998.

GIRSHICK, Ross; DONAHUE, Jeff; DARRELL, Trevor; MALIK, Jitendra. Rich feature hierarchies for accurate object detection and semantic segmentation. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. P. 580–587.

- GONZALEZ, Daniel; DIETZ, Robin L; PANTANOWITZ, Liron. Feasibility of a deep learning algorithm to distinguish large cell neuroendocrine from small cell lung carcinoma in cytology specimens. **Cytopathology**, Wiley Online Library, v. 31, n. 5, p. 426–431, 2020.
- GONZALEZ, Rafael C; WOODS, Richard C. **Processamento digital de imagens**. [S.l.]: Pearson Educação, 2009.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron; BENGIO, Yoshua. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1.
- GROTE, Hans Juergen; NGUYEN, Huy VQ; LEICK, Anand Gilbert; BÖCKING, Alfred. Identification of progressive cervical epithelial cell abnormalities using DNA image cytometry. **Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society**, Wiley Online Library, v. 102, n. 6, p. 373–379, 2004.
- GUO, Yanming; LIU, Yu; GEORGIU, Theodoros; LEW, Michael S. A review of semantic segmentation using deep neural networks. **International journal of multimedia information retrieval**, Springer, v. 7, n. 2, p. 87–93, 2018.
- GURLEY, AM; HIDVEGI, DF; BACUS, JW; BACUS, SS. Comparison of the Papanicolaou and Feulgen staining methods for DNA quantification by image analysis. **Cytometry: The Journal of the International Society for Analytical Cytology**, Wiley Online Library, v. 11, n. 4, p. 468–474, 1990.
- HE, Haibo; MA, Yunqian. **Imbalanced learning: foundations, algorithms, and applications**. John Wiley & Sons, 2013.
- HECHT-NIELSEN, Robert. Theory of the backpropagation neural network. *In*: **NEURAL networks for perception**. [S.l.]: Elsevier, 1992. P. 65–93.
- HOWARD, Jeremy. **Lesson 3: Deep Learning 2019 - Data blocks; Multi-label classification; Segmentation**. Jan. 2019. Disponível em: <https://www.youtube.com/watch?v=MpZxV6DVsmM>.
- INCA, Instituto Nacional de Câncer. **Controle do Câncer do Colo do Útero: Conceito e Magnitude**. Jul. 2021. Disponível em: <https://www.inca.gov.br/controle-do-cancer-do-colo-do-utero/conceito-e-magnitude>.

INCA, Instituto Nacional de Câncer. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero**. 2. ed. [S.l.], 2016.

IVANOVIC, M. Overview of Cytopathology Procedures and Techniques. *In*: CYTOPATHOLOGY in Oncology. [S.l.]: Springer, 2014. P. 1–12.

JACCARD, Paul. **Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines**. Palais de Rumine, Place de la Riponne 6, 1005 Lausanne, CH: Bulletin de la Société Vaudoise des Sciences Naturelles, 1901. v. 37, p. 241–272.

JAHNE, Bernd. Digital image processing: concepts. **Algorithms, and Scientific**, 1997.

JONES, Bruce A; NOVIS, David A. Cervical biopsy-cytology correlation. **Arch Pathol Lab Med**, v. 120, p. 523–531, 1996.

KASHYAP, V; DAS, DK; LUTHRA, UK. Microphotometric nuclear DNA analysis in cervical dysplasia of the uterine cervix: its relation to the progression to malignancy and regression to normalcy. **Neoplasma**, v. 37, n. 5, p. 497–500, 1990.

KITCHENHAM, Barbara. Procedures for performing systematic reviews. **Keele, UK, Keele University**, Citeseer, v. 33, n. 2004, p. 1–26, 2004.

KOLLES, Harry; WANGENHEIM, Aldo Von; VINCE, Giles Hamilton; FEIDEN, Wolfgang. Automatic grading of gliomas in stereotactic biopsies. Comparison of the classification results of neuronal networks and discriminant analysis. **Clinical neuropathology**, v. 12, n. 5, p. 253, out. 1993.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Yann; BOTTOU, Léon; BENGIO, Yoshua; HAFFNER, Patrick. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LIN, Tsung-Yi; GOYAL, Priya; GIRSHICK, Ross; HE, Kaiming; DOLLÁR, Piotr. Focal loss for dense object detection. *In*: PROCEEDINGS of the IEEE international conference on computer vision. [S.l.: s.n.], 2017. P. 2980–2988.

LIU, Wei; ANGUELOV, Dragomir; ERHAN, Dumitru; SZEGEDY, Christian; REED, Scott; FU, Cheng-Yang; BERG, Alexander C. Ssd: Single shot multibox detector. *In: SPRINGER. EUROPEAN conference on computer vision. [S.l.: s.n.], 2016. P. 21–37.*

LONKY, Neal M; SADEGHI, Masood; TSADIK, Girma Wolde; PETITTI, Diana. The clinical significance of the poor correlation of cervical dysplasia and cervical malignancy with referral cytologic results. **American journal of obstetrics and gynecology**, Elsevier, v. 181, n. 3, p. 560–566, 1999.

LOPES, Viviane Aparecida Siqueira; RIBEIRO, Jose Mendes. Cervical cancer control limiting factors and facilitators: a literature review. **Ciencia & saude coletiva**, SciELO Public Health, v. 24, p. 3431–3442, 2019.

LORENA, Ana Carolina; GAMA, João; FACELI, Katti. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. [S.l.]: Grupo Gen-LTC, 2000.

LU, Le; ZHENG, Yefeng; CARNEIRO, Gustavo; YANG, Lin. Deep learning and convolutional neural networks for medical image computing. **Advances in Computer Vision and Pattern Recognition**, Springer, 2017.

MACARINI, Luiz Antonio Buschetto; VON WANGENHEIM, Aldo; DALTOÉ, Felipe Perozzo; ONOFRE, Alexandre Sherlley Casimiro; ONOFRE, Fabiana Botelho de Miranda; STEMMER, Marcelo Ricard. Towards a Complete Pipeline for Segmenting Nuclei in Feulgen-Stained Images. *In: ANAIS do XI Computer on the Beach. [S.l.: s.n.], 2021. v. 1.*

MARR, David. **Vision: A computational investigation into the human representation and processing of visual information**. [S.l.]: MIT press, 2010.

MATIAS, André Victória; AMORIM, João Gustavo Atkinson; MACARINI, Luiz Antonio Buschetto; CERENTINI, Allan; ONOFRE, Alexandre Sherlley Casimiro; ONOFRE, Fabiana Botelho De Miranda; DALTOÉ, Felipe Perozzo; STEMMER, Marcelo Ricardo; WANGENHEIM, Aldo von. What is the State of the Art of Computer Vision-Assisted Cytology? A Systematic Literature Review. **Computerized Medical Imaging and Graphics**, Elsevier, p. 101934, 2021.

MEHROTRA, Ravi; MISHRA, Sanjay; SINGH, Mamta; SINGH, Mangal. The efficacy of oral brush biopsy with computer-assisted analysis in identifying precancerous and

cancerous lesions. **Head & Neck Oncology**, Springer Science e Business Media LLC, v. 3, n. 1, ago. 2011. DOI: 10.1186/1758-3284-3-39. Disponível em: <https://doi.org/10.1186/1758-3284-3-39>.

MITCHELL, Tom M. **Machine Learning**. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997.

MUÑOZ, Nubia; CASTELLSAGUÉ, Xavier; GONZÁLEZ, Amy Berrington de; GISSMANN, Lutz. HPV in the etiology of human cancer. **Vaccine**, Elsevier, v. 24, s1–s10, 2006.

O'MAHONY, Niall; CAMPBELL, Sean; CARVALHO, Anderson; HARAPANAHALLI, Suman; HERNANDEZ, Gustavo Velasco; KRPALKOVA, Lenka; RIORDAN, Daniel; WALSH, Joseph. Deep learning vs. traditional computer vision. *In*: SPRINGER. SCIENCE and Information Conference. [S.l.: s.n.], 2019. P. 128–144.

PITAS, Ioannis. **Digital image processing algorithms and applications**. [S.l.]: John Wiley & Sons, 2000.

POWERS, David Martin. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. **Journal of Machine Learning Technologies**, Bioinfo Publications, 2011. DOI: 10.9735/2229-3981. Disponível em: <https://doi.org/10.9735/2229-3981>.

RAWAT, Waseem; WANG, Zenghui. Deep convolutional neural networks for image classification: A comprehensive review. **Neural computation**, MIT Press, v. 29, n. 9, p. 2352–2449, 2017.

REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali. You only look once: Unified, real-time object detection. *In*: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. P. 779–788.

REN, Shaoqing; HE, Kaiming; GIRSHICK, Ross; SUN, Jian. Faster r-cnn: Towards real-time object detection with region proposal networks. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 39, n. 6, p. 1137–1149, 2016.

RIPLEY, Brian D. **Pattern recognition and neural networks**. [S.l.]: Cambridge University Press, 2007.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. *In: SPRINGER. INTERNATIONAL Conference on Medical image computing and computer-assisted intervention*. [S.l.: s.n.], 2015. P. 234–241.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. [S.l.]: CAMPUS - RJ, 2004. ISBN 9788535211771.

SMITH, Leslie N. A disciplined approach to neural network hyper-parameters: Part 1—learning rate, batch size, momentum, and weight decay. **arXiv preprint arXiv:1803.09820**, 2018.

SMITH, Leslie N. No More Pesky Learning Rate Guessing Games. **CoRR**, abs/1506.01186, 2015. arXiv: 1506.01186. Disponível em: <http://arxiv.org/abs/1506.01186>.

SOLOMON, Chris; BRECKON, Toby. **Fundamentals of Digital Image Processing: A practical approach with examples in Matlab**. [S.l.]: John Wiley & Sons, 2011.

SONG, Youyi; TAN, Ee-Leng; JIANG, Xudong; CHENG, Jie-Zhi; NI, Dong; CHEN, Siping; LEI, Baiying; WANG, Tianfu. Accurate cervical cell segmentation from overlapping clumps in pap smear images. **IEEE transactions on medical imaging**, IEEE, v. 36, n. 1, p. 288–300, 2016.

SU, Jie; XU, Xuan; HE, Yongjun; SONG, Jinming. Automatic detection of cervical cancer cells by a two-level cascade classification system. **Analytical Cellular Pathology**, Hindawi, v. 2016, 2016.

TOUTAIN, Matthieu; ELMOATAZ, Abderrahim; DESQUESNES, Xavier; PRUVOT, Jean-Hugues. A Unified Geometric Model for Virtual Slide Image Processing and Classification. **IEEE Journal of Selected Topics in Signal Processing**, IEEE, v. 10, n. 1, p. 151–160, 2015.

ULKU, Irem; AKAGUNDUZ, Erdem. **A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images**. [S.l.: s.n.], 2019. arXiv: 1912.10230 [cs.CV]. Disponível em: <https://arxiv.org/abs/1912.10230>.

WADA, Ketaro. **labelme: Image Polygonal Annotation with Python**. [S.l.: s.n.], 2016. <https://github.com/wkentaro/labelme>.

WANG, Jingdong *et al.* Deep High-Resolution Representation Learning for Visual Recognition. **TPAMI**, 2019.

WU, Yuxin; KIRILLOV, Alexander; MASSA, Francisco; LO, Wan-Yen; GIRSHICK, Ross. **Detectron2**. [S.l.: s.n.], 2019. <https://github.com/facebookresearch/detectron2>.

XIANG, Yao; SUN, Wanxin; PAN, Changli; YAN, Meng; YIN, Zhihua; LIANG, Yixiong. A novel automation-assisted cervical cancer reading method based on convolutional neural network. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 40, n. 2, p. 611–623, 2020.

ZHAO, Zhong-Qiu; ZHENG, Peng; XU, Shou-tao; WU, Xindong. Object Detection with Deep Learning: A Review. **CoRR**, abs/1807.05511, 2018. arXiv: 1807.05511.
Disponível em: <http://arxiv.org/abs/1807.05511>.