

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

JÚLIA BEATRIZ AZEVEDO TEIXEIRA  
Orientadora: Prof. Dra. Andrea Gomes Campos Bianchi  
Coorientadora: Prof. Dra. Cláudia Martins Carneiro

**SEGMENTAÇÃO DE NÚCLEOS REAIS DE CITOLOGIA CERVICAL  
COM REDES NEURAIAS CONVOLUCIONAIS**

Ouro Preto, MG  
2021

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

JÚLIA BEATRIZ AZEVEDO TEIXEIRA

**SEGMENTAÇÃO DE NÚCLEOS REAIS DE CITOLOGIA CERVICAL COM REDES  
NEURAI CONVOLUCIONAIS**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientadora:** Prof. Dra. Andrea Gomes Campos Bianchi

**Coorientadora:** Prof. Dra. Cláudia Martins Carneiro

Ouro Preto, MG  
2021



## FOLHA DE APROVAÇÃO

**Júlia Beatriz Azevedo Teixeira**

### **Segmentação de núcleos reais de citologia cervical com Redes Neurais Convolucionais**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 24 de Agosto de 2021.

#### Membros da banca

Andrea Gomes Campos Bianchi (Orientadora) - Doutora - Universidade Federal de Ouro Preto  
Cláudia Martins Carneiro (Coorientadora) - Doutora - Universidade Federal de Ouro Preto  
Eduardo José da Silva Luz (Examinador) - Doutora - Universidade Federal de Ouro Preto  
Débora Nasser Diniz (Examinadora) - Mestre - Universidade Federal de Ouro Preto

Andrea Gomes Campos Bianchi, Orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 24/08/2021.



Documento assinado eletronicamente por **Andrea Gomes Campos Bianchi, PROFESSOR DE MAGISTERIO SUPERIOR**, em 24/08/2021, às 21:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0210287** e o código CRC **D1CB90E6**.

# Agradecimentos

O presente trabalho foi realizado com apoio do CNPq, da Universidade Federal de Ouro Preto, do laboratório XR4GOOD, da Fapemig e do Laboratório Multiusuários de Microscopia Avançada e Microanálise do Núcleo de Pesquisas em Ciências Biológicas que forneceu as imagens.

# Resumo

Segundo o Instituto Nacional do Câncer, o câncer do colo do útero apresenta 570 mil novos casos no mundo a cada ano e é responsável pela morte de 311 mil mulheres por ano, sendo considerada a quarta causa de morte mais comum de câncer em mulheres. Comumente causado por subtipos de alto risco do HPV (*Human papillomavirus*), ocorrem alterações celulares que podem ser diagnosticadas pelo exame de Papanicolaou. Embora a implementação do exame tenha diminuído drasticamente os índices de mortalidade, a incidência de falsos positivos caracteriza a alta subjetividade atrelada ao diagnóstico. Esta, se dá pela relação direta da qualidade da análise a experiência do profissional, que também pode ser afetada por diversos fatores, tais como fadiga física e mental. Nesse aspecto, uma alternativa é a inserção da citologia digital no monitoramento de qualidade da rotina laboratorial, de modo a tornar a análise menos suscetível as falhas humanas. O desafio de classificar as células cervicais envolve a pré-segmentação de seus núcleos, visto que diante de anormalidades apresenta variações morfológicas e texturais significativas. Neste trabalho são apresentadas metodologias de segmentação de células que envolvem Redes neurais convolucionais, as arquiteturas U-net e SegNet. Através das bases de dados Herlev, ISBI2014 e *Cric Cervix*, foram realizados experimentos com diferentes funções de ativação, tamanhos de lote e subconjuntos de dados. Os modelos atingiram 0,9695 para a base de dados Herlev, 0,9783 para ISBI2014 e 0,9429 para *Cric Cervix* em coeficiente de Dice. A partir desses resultados, conclui-se que a metodologia proposta se mostrou capaz de segmentar núcleos cervicais reais com qualidade, mesmo em situações de sobreposição e diferentes artefatos.

**Palavras-chave:** Redes Neurais Convolucionais, Células Cervicais, Segmentação.

# Abstract

According to the National Cancer Institute, cervical cancer has 570,000 new cases worldwide each year, and is responsible for the deaths of 311,000 women each year, and is considered the fourth most common cause of cancer death in women. Commonly caused by high-risk subtypes of HPV (Human papillomavirus), cellular changes occur that can be diagnosed by the Pap test. Although the implementation of the exam has drastically reduced mortality rates, the incidence of false positives characterizes the high subjectivity associated with the diagnosis. This is due to the direct relation between the quality of the analysis and professional experience, which can also be affected by several factors, such as physical and mental fatigue. In this aspect, an alternative is the insertion of digital cytology in the quality monitoring of the laboratory routine to make the analysis less susceptible to human failures. The challenge of classifying cervical cells involves the pre-segmentation of their nuclei since in the abnormalities they present significant morphological and textural variations. This paper presents cell segmentation methodologies involving Convolutional Neural Networks, the U-net, and SegNet architectures. Using the Herlev, ISBI2014, and Cric Cervix databases, experiments were performed with different activation functions, batch sizes, and data subsets. The models achieved 0.9695 for the Herlev database, 0.9783 for ISBI2014, and 0.9429 for Cric Cervix in Dice coefficient. From these results, it is concluded that the proposed methodology is able to segment real cervical nuclei with quality, even in situations of overlap and different artifacts.

**Keywords:** Convolutional neural network, Cervical cells, Segmentation.

# Lista de Ilustrações

Figura 2.1 – Na Figura (a) cada píxel apresenta uma classificação. Já a Figura da (b), cada objeto possui uma classificação. (VARATHARASAN et al., 2019) . . . . .	14
Figura 2.2 – Exemplo de uma rede neural convolucional (VARGAS; PAES; VASCONCELOS, 2016) . . . . .	16
Figura 2.3 – Operação de convolução com preenchimento de zero as bordas e salto igual a 1 (PAVLOVSKY, 2019) . . . . .	16
Figura 2.4 – Exemplo de aplicação do <i>Max Pooling</i> tamanho 2x2 com salto 2 sobre uma mapa de características (ALBAWI; MOHAMMED; AL-ZAWI, 2017) . . . . .	20
Figura 2.5 – (a) Modelo original de rede neural (b) Modelo após a utilização da técnica Dropout (SRIVASTAVA et al., 2014) . . . . .	20
Figura 2.6 – Comparação da acurácia de teste e número de etapas do treinamento para modelo com e sem Normalização em lote. É possível notar que a presença de normalização em lote induz o modelo a convergir com menos etapas de treinamento. (IOFFE; SZEGEDY, 2015) . . . . .	21
Figura 2.7 – Operação de convolução transposta com salto 1 e sem adição de zeros nas bordas ( <i>padding</i> ) (MISHRA, 2020) . . . . .	21
Figura 2.8 – Operação de sobreamostragem com proporção (2,2) (CHAUBEY, 2020) . . . . .	21
Figura 2.9 – Análise da taxa de aprendizado através da função de perda (JORDAN, 2020) . . . . .	22
Figura 2.10–Aprendizado por transferência . . . . .	23
Figura 2.11–Número de trabalhos que utilizaram cada métrica (AMORIM et al., 2020) . . . . .	24
Figura 3.1 – Classificações da base de dados Herlev . . . . .	27
Figura 3.2 – Fluxograma da metodologia para a base de dados Herlev . . . . .	28
Figura 3.3 – Verdade terrestre dos núcleos . . . . .	28
Figura 3.4 – Fluxograma da metodologia para a ISBI2014 . . . . .	29
Figura 3.5 – Classificações da base de dados <i>CRIC Cervix Cervix</i> . . . . .	29
Figura 3.6 – Imagem com núcleos segmentados . . . . .	30
Figura 3.7 – Exemplo de imagem particionada da base de dados CRIC - os destaques em verde apresentam recortes que serão aproveitados para a base de segmentação. . . . .	31
Figura 3.8 – Fluxograma da metodologia para a base de dados <i>CRIC Cervix</i> . . . . .	31
Figura 3.9 – Exemplo de aplicação da técnica de aumento de dados (RAJ, 2018) . . . . .	32
Figura 3.10–Número de trabalhos por ano envolvendo Aprendizado Profundo para segmentação e classificação de imagens de citologia cervical (RAHAMAN et al., 2020) . . . . .	33
Figura 3.11–Arquitetura U-net (RONNEBERGER; FISCHER; BROX, 2015) . . . . .	34
Figura 3.12–Demonstração da técnica de reconstrução da resolução inicial com os índices do <i>MaxPooling</i> (PAILLASSA; BERTIN; BOUY, 2019) . . . . .	35

Figura 3.13–Arquitetura SegNet (KENDALL; BADRINARAYANAN; CIPOLLA, 2015)	35
Figura 3.14–Etapas do treinamento dos modelos . . . . .	37
Figura 4.1 – Comparação de performance da U-net com e sem Normalização em Lote para a base de dados Herlev e função de ativação Swish . . . . .	40
Figura 4.2 – Comparação de performance da SegNet com e sem Normalização em Lote para a base de dados Herlev e função de ativação ReLU . . . . .	40
Figura 4.3 – Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados Herlev . . . . .	41
Figura 4.4 – Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados Herlev . . . . .	41
Figura 4.5 – Exemplo de predição da arquitetura SegNet com normalização em lote a partir da base de dados Herlev . . . . .	41
Figura 4.6 – Exemplo de predição da arquitetura SegNet sem normalização em lote a partir da base de dados Herlev . . . . .	41
Figura 4.7 – Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados Herlev . . . . .	42
Figura 4.8 – Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados ISBI2014 . . . . .	43
Figura 4.9 – Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados ISBI2014 . . . . .	43
Figura 4.10–Performance da U-net sem e com Normalização em lote . . . . .	43
Figura 4.11–Performance da SegNet sem e com recortes . . . . .	44
Figura 4.12–Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados ISBI2014 . . . . .	44
Figura 4.13–Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura U-net sem Normalização em lote a partir da base de dados ISBI2014 . . . . .	45
Figura 4.14–Comparação de performance da U-net com e sem Normalização em Lote para a base de dados <i>CRIC Cervix</i> . . . . .	46
Figura 4.15–Comparação de performance da SegNet com e sem Normalização em Lote para a base de dados <i>CRIC Cervix</i> . . . . .	46
Figura 4.16–Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados <i>CRIC Cervix</i> . . . . .	47
Figura 4.17–Comparação de performance da U-net com e sem Dropout para a base de dados <i>CRIC Cervix</i> e função de ativação Mish . . . . .	47

Figura 4.18–Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados <i>CRIC Cervix</i> . . . . .	48
Figura 4.19–Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados <i>CRIC Cervix</i> . . . . .	48
Figura 4.20–Exemplo de predição da arquitetura SegNet com normalização em lote a partir da base de dados <i>CRIC Cervix</i> . . . . .	48
Figura 4.21–Exemplo de predição da arquitetura SegNet sem normalização em lote a partir da base de dados <i>CRIC Cervix</i> . . . . .	48
Figura 4.22–Predição com os pesos do melhor resultado da base ISBI2014 aplicado uma imagem real . . . . .	49
Figura 4.23–Predição com os pesos do melhor resultado da base Herlev aplicado uma imagem real . . . . .	50
Figura 4.24–Predição com os pesos do melhor resultado da base <i>CRIC Cervix</i> aplicado uma imagem real . . . . .	50

# Lista de Tabelas

Tabela 2.1 – Resumo dos trabalhos relacionados a segmentação . . . . .	13
Tabela 3.1 – Distribuição de classes Herlev . . . . .	27
Tabela 3.2 – Distribuição de classes <i>CRIC Cervix</i> . . . . .	30
Tabela 3.3 – Operações aplicadas para aumento de dados . . . . .	32
Tabela 4.1 – Especificação do <i>Hardware</i> da máquina virtual do <i>Google Colaboratory</i> utilizada para os experimentos . . . . .	38
Tabela 4.2 – Parâmetros variados para o treinamento . . . . .	39
Tabela 4.3 – Comparação de resultados da metodologia abordada nesse trabalho e o restante da literatura para a base de dados Herlev . . . . .	40
Tabela 4.4 – Comparação de resultados da metodologia abordada nesse trabalho e o restante da literatura para a base de dados ISBI2014 . . . . .	42
Tabela 4.5 – Resultados da metodologia abordada nesse trabalho para a base de dados <i>CRIC Cervix</i> . . . . .	45
Tabela A.1 – Resultados da arquitetura U-net com a base de dados Herlev e Normalização em Lote. . . . .	62
Tabela A.2 – Resultados da arquitetura U-net com a base de dados Herlev, sem Normalização em Lote. . . . .	63
Tabela A.3 – Resultados da arquitetura SegNet com a base de dados Herlev e Normalização em Lote. . . . .	63
Tabela A.4 – Resultados da arquitetura SegNet com a base de dados Herlev, sem Normalização em Lote. . . . .	64
Tabela A.5 – Resultados da arquitetura U-net com a base de dados ISBI2014 e com Normalização em Lote. . . . .	64
Tabela A.6 – Resultados da arquitetura U-net com a base de dados ISBI2014 e sem Normalização em Lote. . . . .	65
Tabela A.7 – Resultados da arquitetura SegNet com a base de dados ISBI2014 e com Normalização em Lote. . . . .	65
Tabela A.8 – Resultados da arquitetura SegNet com a base de dados ISBI2014 e sem Normalização em Lote. . . . .	66
Tabela A.9 – Resultados da arquitetura U-net com a base de dados <i>CRIC Cervix</i> e Normalização em Lote. . . . .	66
Tabela A.10 – Resultados da arquitetura U-net com a base de dados <i>CRIC Cervix</i> , sem Normalização em Lote. . . . .	67
Tabela A.11 – Resultados da arquitetura SegNet com a base de dados <i>CRIC Cervix</i> , com Normalização em Lote. . . . .	67

Tabela A.12–Resultados da arquitetura SegNet com a base de dados CRIC Cervix, sem Normalização em Lote. . . . .	68
---	----

# Lista de Abreviaturas e Siglas

ADLS	<i>Adaptive radius morphological dilation</i>
ASC-H	Células escamosas atípicas não podendo excluir HSIL
ASC-US	Células escamosas atípicas de significado indeterminado
ASPP	<i>Atrous spatial pyramid pooling</i>
CNN	<i>Convolutional neural network</i>
CRIC	<i>Center for recognition and inspection of cells</i>
CRF	<i>Conditional random fields</i>
DECOM	Departamento de Computação
DRLSE	<i>Distance regularized level set evolution</i>
ELU	<i>Exponential linear unit</i>
FCM	<i>Fuzzy c-means</i>
FCN	<i>Fully convolutional network</i>
GGCMM	<i>Generalized gaussian and cauchy mixture model</i>
GLSZM	<i>Gray level size zone matrix</i>
HSIL	Lesão intraepitelial escamosa de alto grau
ISBI	<i>International symposium on biomedical imaging</i>
KNN	<i>K-nearest neighbor</i>
LD	<i>Linear discriminant</i>
TFA	<i>Two-path fusion attention</i>
LFCCRF	<i>Fully connected conditional random field</i>
LSIL	Lesão intraepitelial escamosa de baixo grau
NILM	Negativo para lesão intraepitelial ou malignidade
ReLU	<i>Rectified linear unit</i>
SCC	Carcinoma de células escamosas

SRM	<i>Shape representation model</i>
SVM	<i>Support vector machine</i>
UFOP	Universidade Federal de Ouro Preto

# Lista de Símbolos

$\alpha$  Letra grega Alfa

$\beta$  Letra grega Beta

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa	3
1.2	Objetivos	4
1.3	Organização do trabalho	4
<b>2</b>	<b>Revisão bibliográfica</b>	<b>5</b>
2.1	Trabalhos relacionados	5
2.1.1	CNNs	5
2.1.2	Métodos tradicionais	9
2.1.3	Discussão	11
2.2	Fundamentação Teórica	12
2.2.1	Câncer de colo de útero	13
2.2.2	Segmentação semântica	14
2.2.3	CNNs	15
2.2.3.1	Camada de convolução	16
2.2.3.2	Função de ativação	17
2.2.3.3	Função de perda	19
2.2.3.4	Camada de subamostragem	19
2.2.3.5	<i>Dropout</i>	20
2.2.3.6	Normalização em lote	20
2.2.3.7	Camada de sobreamostragem	21
2.2.3.8	Época	22
2.2.3.9	Tamanho do lote	22
2.2.3.10	Taxa de aprendizado	22
2.2.4	Aprendizado por transferência	23
2.2.5	Métricas	23
2.2.5.1	Acurácia	24
2.2.5.2	Coefficiente de Dice	24
2.2.5.3	Revocação	24
2.2.5.4	Precisão	25
<b>3</b>	<b>Desenvolvimento</b>	<b>26</b>
3.1	Base de dados e metodologia	26
3.1.1	Herlev	26
3.1.2	ISBI2014	28
3.1.3	CRIC Cervix	29
3.2	Aumento de dados	31
3.3	Arquiteturas	32

3.3.1	U-net	33
3.3.2	SegNet	34
3.4	Treinamento	35
<b>4</b>	<b>Resultados e Discussões</b>	<b>38</b>
4.1	Herlev	39
4.2	ISBI2014	42
4.3	CRIC Cervix	45
4.4	Discussões	48
<b>5</b>	<b>Considerações finais</b>	<b>51</b>
<b>6</b>	<b>Trabalhos futuros</b>	<b>52</b>
 <b>Referências</b>		 <b>53</b>
 <b>Apêndices</b>		 <b>61</b>
<b>APÊNDICE A Resultados dos Experimentos</b>		<b>62</b>
A.1	Herlev	62
A.2	ISBI2014	64
A.3	CRIC Cervix	66

# 1 Introdução

O câncer de colo de útero é caracterizado como a quarta causa mais frequente de morte por câncer em mulheres no mundo, sendo responsável pelo óbito de 311 mil mulheres ao ano (INCA, 2020). Desde a implementação do exame de Papanicolaou nos anos 60 para rastreio de anormalidades em células cervicais, os índices de mortalidade e incidência diminuíram drasticamente (LIANG et al., 2021). O câncer cervical é um problema de saúde pública global que pode ser prevenido por meio do diagnóstico precoce através de triagem e tratamento de lesões pré-cancerosas (MABOTJA; LEVIN; KAWONGA, 2021).

O exame de Papanicolaou é um método simples que permite detectar células pré-neoplásicas e neoclássicas a partir da extração de amostras do colo uterino (BRENNAN et al., 2001). Este exame permitiu reduzir as taxas de mortalidade por câncer de colo de útero em diversos países após programas de rastreamento (FERNANDES et al., 2009). Em contraste, o número de células em cada amostra pode chegar até 300 000, o que torna a análise densa e exaustiva (RAHAMAN et al., 2020). Para proporcionar um auxílio na avaliação das amostras, diminuir a fadiga dos profissionais e melhorar a assertividade do exame, a inserção da citologia digital no monitoramento da rotina laboratorial torna-se uma alternativa.

O primeiro passo para a identificação de alterações malignas nas células é a detecção do núcleo. O núcleo, na presença de anomalias traz informações imprescindíveis ao diagnóstico, através de variações morfológicas e texturais (DINIZ, 2019). Em síntese, considera-se que destacando o núcleo cervical é possível trazer mais assertividade ao diagnóstico manual e automatizado.

Em 1988, foi introduzido o sistema Bethesda para classificação das lesões cervicais. Este, foi apresentado com o intuito de padronizar a classificação morfológica das lesões intraepiteliais cervicais (FIGUEIREDO et al., 2015). Revisado novamente em 2001 e novamente em 2014, os termos atuais para a classificação das células escamosas são: Células escamosas atípicas de significado indeterminado, possivelmente não neoplásicas (ASC-US); Células escamosas atípicas, não se pode excluir lesão de alto grau (ASC-H); Lesão intraepitelial escamosa de baixo grau (LSIL); Lesão intraepitelial escamosa de alto grau (HSIL); Lesão intraepitelial escamosa de alto grau não podendo excluir micro-invasão; Carcinoma de células escamosas (SCC) e Negativo para lesão intraepitelial ou malignidade (NILM), para células sem lesão (NAYAR; WILBUR, 2015).

Novas abordagens ao desafio da segmentação de núcleos de citologia cervical têm surgido nos últimos anos, devido à possibilidade de sua aplicação em situações reais. Em contraste, a escassez de bases de dados públicas convencionais, classificadas de acordo com o sistema Bethesda (NAYAR; WILBUR, 2017) e compostas de múltiplas células, é um impasse para avaliar

a aplicabilidade das abordagens em imagens reais. Os trabalhos Zhang et al. (2017), Allehaibi et al. (2019), Hussain et al. (2020), Win et al. (2020), Sharma e Mangat (2016), Gautam et al. (2018), Liu et al. (2018), Zhao et al. (2019), Sabeena, Gopakumar e Thampi (2020), Yang, Wu e Yin (2020) abordam métodos de segmentação validados nos conjuntos de dados ISBI2014(LU; CARNEIRO; BRADLEY, 2015), ISBI2015(LU; CARNEIRO; BRADLEY, 2015) e Herlev (MARINAKIS; DOUNIAS; JANTZEN, 2009). As bases de dados ISBI2014 (*International symposium on biomedical imaging - 2014*) e ISBI2015, constituem-se quase predominantemente de imagens sintéticas geradas com um grau de sobreposição. Já a base de dados Herlev, é composta por imagens reais células únicas. Embora utilizem imagens reais, os trabalhos Zhang et al. (2017), Zhang et al. (2019), Araújo et al. (2019), Allehaibi et al. (2019), Wan et al. (2019), Zou et al. (2020) e Xiang et al. (2020) envolvem bases em meio líquido que são resultantes do método de diluição das amostras, sendo esse menos utilizado e de alto custo comparado ao convencional.

A base de dados *CRIC Cervix* (REZENDE; BIANCHI; CARNEIRO, 2021) é uma coleção de imagens reais de esfregaços do colo do útero representando uma gama de diferentes lesões em células cervicais. Interpretadas por citopatologistas independentes dentro da plataforma CRIC (*Center for Recognition and Inspection of Cells*), a base de dados conta com mais de 3200 núcleos celulares segmentados obtidos a partir de 400 imagens, dispondo de núcleos de células com seis classificações distintas de acordo com o sistema Bethesda.

Na visão computacional, as CNNs (*Convolutional neural networks*) se destacam como estado da arte em diversos campos relacionados ao reconhecimento de padrões (ZHIQIANG; JUN, 2017). Estas, são algoritmos de aprendizado profundo que se sobressaem entre outros métodos por serem capazes de aprender filtros e características baseados em imagens de entrada sem processamento prévio. Em uma Rede neural convolucional, as imagens de entrada percorrem diversas camadas de convolução, onde cada neurônio corresponde a um píxel e cada camada é responsável pela extração de uma série de características dentro do conjunto de dados. As CNNs também se destacam por serem predominantemente as ganhadoras desde 2012 do Desafio de Reconhecimento em Larga Escala da ImageNet (RUSSAKOVSKY et al., 2015), um dos eventos mais importantes envolvendo reconhecimento visual no mundo.

A segmentação surge para facilitar a análise de imagens, através da extração de informações significativas (SULTANA; SUFIAN; DUTTA, 2020). Nesse trabalho, utilizamos a segmentação semântica, visto que visa classificar cada pixel da imagem (WEI et al., 2016). Diversos métodos de CNNs influenciaram o campo da segmentação e nos trouxeram vários modelos de sucesso. Neste trabalho, destaca-se as CNNs U-net (RONNEBERGER; FISCHER; BROX, 2015) e SegNet(KENDALL; BADRINARAYANAN; CIPOLLA, 2015). Com arquitetura *encoder-decoder*, os métodos apresentam novas abordagens para as Redes Totalmente Convolucionais. A metodologia *copy-and-copy* da U-net, consiste em alimentar o *decoder* com os mapas de características recortados provenientes das camadas do *encoder*. Já a SegNet, introduziu a técnica de reconstrução da proporção dos mapas de características, através do reaproveitamento

dos índices de *pooling*.

Baseado no que foi citado anteriormente, pretende-se treinar as arquiteturas de CNNs U-net e SegNet para segmentação semântica do núcleo de imagens reais de esfregaços de Papanicolaou através das bases de dados *CRIC Cervix*, Herlev e ISBI2014. Além disso, pretende-se avaliar a assertividade dos modelos escolhidos através das métricas acurácia, coeficiente de Dice, revocação e precisão.

## 1.1 Justificativa

O desenvolvimento do câncer de colo de útero se inicia quando as células saudáveis do cérvix reproduzem mutações (STAFF, 2019). As células com mutações se multiplicam de maneira desordenada e diferentemente das células saudáveis, não tem seu ciclo de vida finalizado. Por conseguinte, o acúmulo de células anormais forma o tumor. Quando diagnosticado em fase não invasiva, tem a possibilidade de cura entre 80% e 90%, no entanto, as chances diminuem gradativamente em estágios posteriores (OKUNADE, 2020).

Segundo Torre et al. (2017), a incidência da enfermidade diminuiu em 80% desde a implementação o exame de Papanicolaou. Contudo, Li e Najarian (2001) vincularam à análise manual altas taxas de falsos negativos, podendo variar de 5% à 55%, o que pode culminar no agravamento da doença, visto que a mulher não será submetida ao tratamento. O diagnóstico citopatológico tradicional é baseado na observação e na avaliação morfológica de células em microscópio óptico e podem ser associados a pareceres clínicos, mas é subjetivo, manual e extenso. Logo, pode ser afetado pela fadiga do operador após períodos prolongados de avaliação de lâminas, pela habituação visual, pouca exposição do citologista à casos diferenciados, e pela variabilidade entre os citopatologistas.

Nesse contexto, para diminuir a fadiga física e mental dos profissionais, diminuir o tempo de espera para obtenção de resultados e melhorar a assertividade do exame, uma alternativa é a utilização de métodos para automatização deste trabalho de alta complexidade. Um primeiro passo para a identificação de alterações malignas nas células é a detecção do núcleo. Na presença de uma anomalia, os núcleos apresentam variações morfológicas e texturais significativas, responsáveis por trazer informações essenciais ao diagnóstico (DINIZ, 2019). Em síntese, destacando-se o núcleo cervical pode tornar possível melhorar a assertividade do diagnóstico manual e automatizado.

Sendo assim, justifica-se o desenvolvimento de técnicas para rastreamento da região a ser analisada. Vale ressaltar que por se tratar de um problema de visão computacional, a utilização de soluções baseadas em CNNs se sobressaem a outros métodos por dispensarem pré-processamento. Isso se deve a sua capacidade de aprender filtros e características através das imagens de entrada. Além disso, as imagens de entrada percorrem diversas camadas de convolução, onde cada camada é responsável pela extração de um conjunto de características. Em suma, o desempenho das

CNNs tem representado grande parte dos métodos estado-da-arte no reconhecimento visual (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

## 1.2 Objetivos

Este trabalho tem como objetivo verificar o desempenho de arquiteturas de CNNs quando aplicadas segmentação de núcleos reais de citologia cervical. São objetivos específicos deste trabalho:

- Implementação das arquiteturas U-net e SegNet;
- Proposição de um treinamento das arquiteturas adaptado para segmentação de núcleos reais de citologia cervical;
- Validação da metodologia proposta por meio das bases de dados *CRIC Cervix*, Herlev e ISBI2014;
- Análise da metodologia escolhida por meio dos resultados obtidos considerando as métricas acurácia, precisão, revocação e coeficiente de Dice;

## 1.3 Organização do trabalho

Nesta seção está disposta a organização desse trabalho. O Capítulo 2 aborda conceitos importantes e trabalhos com temas relacionados. No Capítulo 3 é apresentada a metodologia escolhida. No Capítulo 4 é feita a análise dos resultados e discussões. Nele, são feitas comparações dos resultados com a literatura. Por fim, o Capítulo 5 apresenta as considerações finais sobre o trabalho desenvolvido e as propostas para trabalhos futuros.

## 2 Revisão bibliográfica

Novas abordagens que tratam o problema de segmentação de núcleos de citologia cervical tem surgido nos últimos anos, devido à possibilidade de sua aplicação em situações reais. Dessa forma, este capítulo apresentará trabalhos relacionados ao tema, encontrados na literatura e conceitos importantes para o entendimento do leitor.

### 2.1 Trabalhos relacionados

Nesta seção serão apresentados alguns trabalhos relacionados a segmentação do núcleo em células. Desse modo, serão abordados importantes métodos de segmentação e resultados para fim de comparação e embasamento teórico. A seção foi dividida em duas subseções de acordo com a solução proposta pelos autores - metodologias que abordam CNNs e métodos tradicionais de aprendizado de máquina. Por fim, será apresentada uma discussão a respeito das abordagens propostas pelos autores.

#### 2.1.1 CNNs

O sucesso da AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) transformou o campo de visão computacional em direção ao aprendizado profundo. Outro marco importante envolvendo CNNs, foi o surgimento das FCNs (*Fully convolutional network*) para segmentação. Propostas em 2015 por Long, Shelhamer e Darrell (2015), sua principal característica é a inexistência de camadas completamente conectadas, melhorando assim seu tempo de execução. As arquiteturas AlexNet, VGG-net (SIMONYAN; ZISSERMAN, 2014) e GoogLeNet (SZEGEDY et al., 2014) foram adaptadas para segmentação, tendo suas camadas densas de saída substituídas por camadas de convolução. O autor executa vários experimentos com concatenação entre os resultados das camadas de subamostragem, haja vista que permitem as camadas finais aprenderem características que são capturadas em camadas anteriores, deste modo, permitem a segmentação com adição de detalhes. A abordagem bateu o estado-da-arte para as bases de dados PASCAL VOC 2011 (EVERINGHAM et al., ), NYUDv2 (SILBERMAN DEREK HOIEM; FERGUS, 2012) e SIFT Flow (LIU; YUEN; TORRALBA, 2010), além de embasar a criação de arquiteturas como a U-Net (RONNEBERGER; FISCHER; BROX, 2015), com os melhores resultados para a segmentação de imagens biomédicas.

A arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015) consiste em 28 camadas de convolução, divididas em 9 blocos convolucionais para o *encoder* e o *decoder*. Os blocos convolucionais do *encoder* são seguidos de *MaxPooling*, na qual são selecionadas as informações mais importantes para captura de contexto. Já no *decoder*, são seguidos de camadas de convolução

transposta para restaurar a resolução da imagem e capturar informações de localização. Ademais, tendo em vista que o desempenho das Redes neurais está fortemente ligado ao conjunto de entrada fornecidos, com auxílio de técnicas de aumento de dados no treinamento, o modelo alcançou uma boa performance através da base do desafio de segmentação de células do ISBI2012 que possui um pequeno número de amostras.

Outro marco para a segmentação, foi apresentado em 2015 pela arquitetura SegNet (KEN-DALL; BADRINARAYANAN; CIPOLLA, 2015). Composta por uma rede VGG16 pré-treinada no *encoder* da arquitetura, permitiu a redução significativa no número de parâmetros treináveis. A principal novidade do modelo é a reutilização dos índices de *pooling* na reestruturação espacial da imagem. A arquitetura obteve resultados competitivos em base de dados CamVid (BROSTOW; FAUQUEUR; CIPOLLA, 2009) e menor consumo de memória, ultrapassando métodos estado da arte para segmentação de prédios, árvores, entre outras estruturas.

A partir dos métodos FCM e *Back propagation neural network* (BPNN) (SHARMA; MANGAT, 2016), propuseram um sistema de segmentação de núcleos celulares. As imagens da base de dados Herlev (MARINAKIS; DOUNIAS; JANTZEN, 2009) foram separadas em aglomerados através do algoritmo FCM e, posteriormente classificadas em núcleo e não núcleo pela arquitetura BPNN. Em comparação a trabalhos anteriores, os resultados superaram por 1%, 7% e 5% respectivamente, as métricas precisão, revocação e coeficiente de Dice.

Propondo contornar a segmentação, Zhang et al. (2017) utilizaram a arquitetura ConvNet para classificar imagens de Papanicolaou. A rede foi pré-treinada com a base de dados da ImageNet (RUSSAKOVSKY et al., 2015) e as bases de células isoladas Herlev e a base de dados coletada em meio líquido HEMLBC (ZHANG et al., 2017). Para evitar o sobre-ajuste da rede, foram aplicadas operações de translação e rotação para aumento da base de dados. Embora os resultados tenham superado trabalhos anteriores, os autores pontuam o fato de ambas as bases de dados não demonstrarem situações reais - imagens com alto grau de ruído, artefatos e sobreposição de células.

Com base na diferença de contraste entre núcleo e fundo, Gautam et al. (2018), propuseram uma metodologia de 3 passos para segmentação de núcleos da base Herlev. O primeiro passo é o pré-processamento seletivo, que consiste na divisão da base em imagens que possuem ou não padrão de cromatina relativamente homogêneo. Segundos os autores, essa divisão é necessária para aplicação de aumento de contraste nas imagens homogêneas para facilitar a segmentação do núcleo. Em seguida, as divisões são treinadas separadamente em um modelo baseado na VGGNet. Por fim, algumas operações morfológicas pós processamento são aplicadas para gerar a segmentação final. A proposta superou resultados anteriores, com revocação de 0,91 e f-score de 0,90. Liu et al. (2018) apresentaram uma metodologia de segmentação de núcleos, através do modelo Mask-RCNN para predição e *Fully connected conditional random field* (LFCCRF) para refinamento das bordas dos resultados. A avaliação dos métodos foi realizada através da base de dados Herlev e se demonstrou superior a metodologias previamente apresentadas.

Baseado na arquitetura U-net, a PGU-net+ (ZHAO et al., 2019) se difere da arquitetura original pela adição de módulos residuais e pelo modelo de treinamento progressivo. O treinamento progressivo, com a base de dados Herlev, consiste na divisão do treinamento por etapas na qual o modelo e a resolução da imagem são gradativamente incrementados. Inicialmente o modelo é treinado com poucas camadas e imagens de menor resolução, até que de forma gradual o modelo atinja seu tamanho original. A metodologia obteve resultados superiores ao estado-da-arte.

Por meio da arquitetura ResNeXt (XIE et al., 2016) em topologia de árvore e integrando pares de camadas adjacentes, Zhang et al. (2019) propuseram um novo método para melhorar a captura de informações semânticas e espaciais na segmentação. Os mapas de características das camadas são combinados a partir de operações de sobre-amostragem e convolução para garantir a mesma dimensionalidade. Após combinados, os mapas passam por uma operação de convolução. Em seguida, a rede obterá dois conjuntos de mapas de características de baixo e alto nível, na qual serão concatenados na aplicação do algoritmo TFA (*Two-path fusion attention*). Através da geração de mapas com peso, designa a fusão de características em comum. Por fim, a rede combina os mapas de características da camada de entrada com a saída do algoritmo, para entrada do classificador Softmax. O método proposto obteve melhor resultados em comparação aos métodos U-Net, FCN e DeepLabv3+ (CHEN et al., 2018) para precisão, revocação e média do coeficiente de Dice nas bases de dados ISBI2014 (LU; CARNEIRO; BRADLEY, 2015) e na base de dados coletada em meio líquido da Universidade chinesa de medicina Guangzhou (ZHANG et al., 2019).

A proposta principal do artigo proposto por Araújo et al. (2019) é uma ferramenta baseada em aprendizado profundo para auxiliar o citopatologista, na qual dada às imagens de entrada coloca em um *ranking* de probabilidade dessa imagem conter células anormais. Primeiramente, utilizando imagens reais da base de dados *CRIC Cervix* (REZENDE; BIANCHI; CARNEIRO, 2021), foram criados aleatoriamente os conjuntos de treino e de teste. Para o pré processamento, as imagens com pouca informação e apenas fundo ruidoso foram removidas e as restantes particionadas em janelas de 52 x 52. Em seguida, uma Rede neural convolucional de arquitetura LeNet (LECUN et al., 2015) pré-treinada aplicou uma classificação pixel por pixel da probabilidade de conter uma célula anormal. Um processo pós segmentação para remover estruturas ruidosas e não pertencentes a célula foi determinado através do raio do elemento estruturante e área mínima maior que o limiar definido - tendo em vista o processo aplicado de estimação de parâmetros. Com o objetivo final de segmentar regiões com apenas células anormais, os resultados obtidos demonstraram 73% de precisão e *Mean average precision* de 0,936% para classificação de células anormais.

Através da utilização da arquitetura Mask R-CNN(HE et al., 2017) pré-treinada com a base de dados COCO, Allehaibi et al. (2019) abordaram a segmentação de células e a classificação em uma arquitetura VGG mais compacta, com 7 camadas e filtros menores. Além disso, por

transformações geométricas para aumento de dados na base Herlev tornou-se possível alcançar a precisão média de 92,0% para segmentação da estrutura celular.

A segmentação precisa de células com sobreposição foi o principal fator considerado por Wan et al. (2019) ao proporem a metodologia voltada a CNNs. O primeiro passo foi a utilização da rede TernausNet (IGLOVIKOV; SHVETS, 2018) baseada na arquitetura U-Net e com os pesos da VGG11 pré-treinada pela base de dados da ImageNet com a última camada totalmente conectada dando lugar a uma camada de convolução com 512 filtros para classificar entre núcleo, citoplasma e fundo. Para filtrar os núcleos classificados erroneamente foi utilizado o classificador *Adaboost* recebendo como entrada duas características: o Histograma de Gradientes Orientados (HOG) e o descritor GLSZM (*Gray Level Size Zone Matrix*). Baseado na localização do centro do núcleo para localização da célula foi abordado um método de duas janelas para capturar a região de interesse, visto que delimitaria a porção da imagem na qual o citoplasma está presente. Em seguida, para segmentar o citoplasma, os autores apostaram no modelo DeepLabV2 (CHEN et al., 2017). A principal vantagem do modelo é a presença de ASPP (*Atrous spatial pyramid pooling*) – o método se baseia em várias camadas de dilatação paralelas com taxas diferentes aplicadas no mapa de características e no final ocorre a junção das informações, a qual auxilia a detecção objetos da mesma classe de tamanhos diferentes, incorporando assim um contexto maior. Por fim, o método CRF (*Conditional random fields*) e DRLSE (*Distance regularized level set evolution*) analisa a relação entre a vizinhança de píxeis para preservar as bordas e refinar as bordas, respectivamente. Ademais, através da utilização das bases de dados ISBI2015, ISBI2014 e INHouse demonstraram 15% de melhora para coeficiente de Dice comparado a trabalhos anteriores.

Através do modelo FC-DenseNet56, Sabeena, Gopakumar e Thampi (2020) abordaram a importância da segmentação de núcleos para a automatização da análise de células cervicais. O modelo foi treinado no conjunto de células únicas Herlev. Com 0,99 de precisão, o resultado foi superior a trabalhos anteriores. Ainda, o trabalho atingiu 0,93 de coeficiente de Dice, o que caracteriza uma segmentação próxima a verdade terrestre.

Baseando-se no fato de que os métodos estado-da-arte para classificação de imagens de citologia cervical obtém como entrada as células já segmentadas, Zou et al. (2020) apresentaram um modelo de Rede neural convolucional para segmentação e posterior classificação das células. A abordagem consiste na segmentação do núcleo utilizando o modelo Mask R-CNN, as bases de dados 2018 *Kaggle Science Bowl* e *Becton-Dickinson* para treinamento e validação, respectivamente. Ainda, através de 15 arquiteturas de CNNs pré-treinadas com a base da ImageNet, foi obtida a classificação centrada ao núcleo da célula recém segmentada. Por fim, o autor conclui que todos os modelos atingiram bons resultados e a abordagem do núcleo como principal agente na classificação é promissora. Embora o modelo tenha sido treinado e validado com bases distintas visando a generalização, as amostras coletadas em meio líquido são um desafio menor, visto que apresentam um número de sobreposição inferior ao método convencional.

A partir do algoritmo de detecção YOLOv3, Xiang et al. (2020) apresentaram resultados para a classificação de lesões através da arquitetura InceptionV3. A metodologia tornou-se possível pela reutilização da saída do método de detecção, caixas delimitadoras da região da célula, como entrada para o classificador. O método YOLOv3, consiste em uma rede Darknet-53 pré-treinada com a base de dados da ImageNet para extração de atributos em alto nível, em seguida a utilização de sub-amostragem para obter a predição em três escalas diferentes. A detecção foi necessária pela utilização de uma base de dados própria de múltiplas células em meio aquoso. Por fim, obteve-se mAP de 60,2% para detecção, e 89,3% de acurácia, 97,5% de sensibilidade e 67,8% de especificidade para classificação.

Fundamentada na eficácia de FCNs, Hussain et al. (2020) apresentaram uma abordagem para segmentação de núcleos cervicais. A metodologia consiste no modelo pré-treinado SRM (*Shape representation model*) para aprender as informações dos núcleos baseado na verdade-terrestre da segmentação dada como entrada. Em seguida, a saída do modelo é utilizado como entrada para a arquitetura proposta. Baseada no modelo U-net, a arquitetura consiste na metodologia *encoder-decoder*, apenas com a adição de conexões residuais e blocos densamente conectados. Além da segmentação binária, foram apresentados resultados para segmentação de instâncias contendo núcleos normais e anormais. Os resultados foram obtidos em uma nova base gerada, a partir de três bases de dados distintas - Herlev, e duas bases privadas, coletada em meio líquido e convencional. O método proposto obteve os melhores resultados em Dice, precisão, revocação e acurácia comparados ao modelo U-net e Mask R-CNN para a base de dados gerada.

Através de um modelo similar a U-net, Yang, Wu e Yin (2020) introduziram uma metodologia que consiste na utilização dos métodos ICM (*Interacting convolutional module*) e IPRCM (*Internal pyramid resolution complementing module*). O ICM é um módulo convolucional que agrega novos detalhes através 2 tamanhos de filtros. Já o IPRCM foi aplicado para combinar mapas de características de diferentes resoluções, tendo em vista que as informações seriam complementares entre si. Com precisão de 0.972, esse trabalho superou resultados anteriores para a base de dados Herlev.

### 2.1.2 Métodos tradicionais

Além de metodologias baseadas em Redes neurais convolucionais, muitos autores também realizaram propostas de segmentação de células que envolveram outras metodologias em aprendizado de máquina. Nesta seção serão abordados trabalhos que não englobaram aprendizado profundo. Através de um método não supervisionado, Ushizima, Bianchi e Carneiro (2015) descrevem uma abordagem para detecção de células individuais, incluindo a distinção entre núcleo e citoplasma. A abordagem consiste em três passos: cálculo de massa celular, super-pixéis para detecção do núcleo e repartição das imagens em polígonos convexos para detecção de citoplasma. Foram utilizadas as imagens do conjunto de dados ISBI2014, sintéticas. Os testes apresentaram resultados

maiores que 86% para coeficiente de Dice.

Saha, Bajger e Lee (2017) estabelecem uma metodologia de segmentação de núcleos através do algoritmo de agrupamento nebuloso com base em formas circulares CiscFC (*Circular shape constrained fuzzy clustering*). Superando resultados anteriores, os autores atingiram 0,938 em coeficiente de Dice na base de dados ISBI2014. Foram calculadas as características tamanho da região, intensidade média e solidez para determinar as regiões segmentadas. Por fim, foram apontadas as restrições da base de dados que, além de sintética, apresenta poucas amostras, o que dificulta a generalização do modelo.

Nos estudos de Phoulady et al. (2017) é retratada a detecção e segmentação de núcleos e citoplasmas. A metodologia é dividida em 3 etapas: a segmentação do núcleo, a segmentação do citoplasma através das informações de localização do núcleo e a segmentação de citoplasmas sobrepostos. Primeiro, as amostras sofrem aplicação de um filtro de remoção de ruído adaptativo. Em seguida, o algoritmo binariza iterativamente a imagem, começando com um limiar baixo para encontrar pontos de diferentes núcleos. Regiões muito pequenas ou muito côncavas são removidas após cada binarização e as regiões restantes são adicionadas a uma máscara de núcleo. A máscara do núcleo mantém as regiões dos núcleos segmentadas em cada fase de execução do algoritmo. Por fim, o algoritmo atingiu coeficiente de Dice de 0,8730 para a base de dados ISBI2015 e 0,9380 para a base de dados ISBI2014.

Huang et al. (2020) propuseram a segmentação de células cervicais baseados no algoritmo *Multi-scale fuzzy clustering* aplicado a base de dados ISBI2014 e ISBI2015. A primeira etapa do método consiste na separação do fundo das células em primeiro plano com o algoritmo, identificando assim as aglomerações. O segundo passo é a utilização *Multi-scale fuzzy clustering* para adquirir o citoplasma e os candidatos a núcleo para a construção de uma árvore hierárquica com as regiões segmentadas. Por fim, no terceiro passo a partir do método de contorno ativo DRLSE foi possível refinar as bordas do núcleo, e em seguida, a detecção do ponto côncavo para obtenção dos núcleos candidatos. Ainda, foi aplicado um limiar baseado em características morfológicas para filtrar e remover regiões entre os candidatos. Comparados com outros métodos presentes na literatura, obteve precisão e revocação superiores ao estado da arte.

Através do modelo assimétrico GCM (Generalized gaussian and cauchy mixture model), Huang e Zhu (2020) obtiveram como resultado o contorno dos núcleos e aglomerados de células. Ainda com a premissa de que cada núcleo pertence a uma célula, o centróide do núcleo foi calculado. Em seguida, os falsos núcleos foram removidos através de operações morfológicas, baseado-se no seu tamanho. Por meio do conjunto de pontos que determina o contorno da aglomeração, cada ponto foi mapeado ao centróide mais próximo, e com a mesma distância entre dois centróides foram considerados pontos de junção. A partir deles foi possível estimar a curva inicial da área de sobreposição. Para refinar o contorno, foi utilizado o método de contorno ativo DRLSE. Por fim, a abordagem obteve 0,94 para o coeficiente de Dice na base de dados ISBI2014.

Por meio da segmentação, Win et al. (2020) apresentaram um sistema de triagem assistida

por computador para classificação de células cervicais. A metodologia se baseia em 6 passos: aquisição e aprimoramento de imagens, segmentação de células, extração, seleção e classificação de características. Quanto as imagens, as bases SIPaKMeD (PLISSITI et al., 2018) - base de dados real de múltiplas células - e Herlev foram utilizadas. Para o pré-processamento das imagens, foram aplicadas operações de redução de ruído e aumento de contraste. Na segmentação, a aplicação do método *Watershed transformation* tornou necessária a conversão das imagens em escala de cinza, visto que o método trata a imagem como um mapa topográfico, onde a intensidade de cada píxel representa a altura no mapa. As regiões detectadas, compostas de inúmeras células, foram recortadas e classificadas utilizando características baseadas em intensidade. Na etapa de seleção de características, o algoritmo *Random forest* foi utilizado. Para classificação, foi aplicado um *Ensemble*, na qual combina o resultado dos classificadores: LD (*Linear discriminant*), SVM (*Support vector machine*), KNN (*k-nearest neighbor*), *Boosted trees* e *Bagged trees*. Por fim, o método proposto atingiu 0,862 na segmentação do núcleo para o coeficiente de Dice, acurácia de 98,27% na classificação de duas classes (normais e anormais) e 94,09 % em cinco classes.

Devido à necessidade da segmentação de núcleos para o diagnóstico precoce de câncer cervical, Wang et al. (2020) propuseram a segmentação baseada em profundidade. Primeiramente, para capturar a região com os aglomerados de células, núcleos e fundo foi utilizado o algoritmo de clusterização FCM. Em seguida, foi construída uma árvore com as regiões candidatas, e sucessivamente a busca em profundidade para remoção de regiões e minimização da estrutura. Ainda, aplicaram o algoritmo ADLS (*Adaptive radius morphological dilation*) para refinar a segmentação e remover falsos positivos. Os resultados obtidos foram superiores em precisão e revocação, e inferiores sob o coeficiente de Dice, quando comparados aos trabalhos citados anteriormente para a base de dados ISBI2015.

A tarefa de segmentar imagens de citologia cervical é desafiadora devido a presença de sobreposições, artefatos, neutrófilos e células com bordas espúrias. Com uma metodologia baseada em propriedades de contorno, Hoque et al. (2021) atingiram resultados superiores ao estado-da-arte na segmentação de núcleos celulares da base de dados ISBI2014. A hipótese consistia na aplicação do filtro Gaussiano para diminuir ruídos, e em seguida um algoritmo de limiar adaptativo. Por fim, foram calculadas características baseadas no contorno para remoção falsos positivos. A metodologia também foi aplicada ao conjunto de dados privado em meio líquido BSMMU, da Universidade de Medicina Bangabandhu Sheikh Mujib.

### 2.1.3 Discussão

Os trabalhos de Ushizima, Bianchi e Carneiro (2015), Sharma e Mangat (2016), Saha, Bajger e Lee (2017), Phoulady et al. (2017), Zhang et al. (2017), Gautam et al. (2018), Liu et al. (2018), Zhao et al. (2019), Allehaibi et al. (2019), Hussain et al. (2020), Win et al. (2020), Sabeena, Gopakumar e Thampi (2020) e Yang, Wu e Yin (2020) foram realizados envolvendo bases de dados contendo células únicas (recortadas) ou sintéticas, logo, não remetem a situações reais,

que apresentam um maior nível de complexidade quanto a sobreposição e artefatos na imagem. Tivemos ainda trabalhos envolvendo bases de dados com amostras coletadas em meio líquido (ZHANG et al., 2017), (ZHANG et al., 2019), (ARAÚJO et al., 2019), (ALLEHAIBI et al., 2019), (WAN et al., 2019), (ZOU et al., 2020), (XIANG et al., 2020) e (HOQUE et al., 2021). Embora sejam imagens reais, são coletadas em meio líquido, que não é um método tão acessível pois é de alto custo quando comparado ao convencional.

Em contraste ao trabalhos anteriores, Araújo et al. (2019) abordaram a segmentação do núcleo de imagens convencionais, classificadas de acordo com o sistema Bethesda (NAYAR; WILBUR, 2017). Apesar de o método proposto ter obtido uma segmentação de 98% dos núcleos corretamente (verdadeiros positivos), também abrangeu outras estruturas resultando em um número elevado de falsos positivos. Além disso, a abordagem proposta conseguiu segmentar 73% das células anormais. Em síntese, o trabalho destinou-se a ordenar imagens pela probabilidade da presença de células anormais. Desse modo, não há uma análise detalhada voltada aos resultados da segmentação do núcleo.

Por fim, dois trabalhos usaram bases reais convencionais Win et al. (2020) e Hussain et al. (2020), são elas, a SIPaKMeD (PLISSITI et al., 2018) e uma base de dados privada. A base de dados SIPaKMeD ainda que apresente imagens reais, não é classificada a cerca do sistema Bethesda usado em citologia convencional para o diagnóstico de lesões celulares. Não possui classificações de lesões essenciais ao diagnóstico. A base de dados utilizada por Hussain et al. (2020) é composta pela junção de três bases de dados distintas - Herlev, e bases de dados coletada em meio líquido e convencional. Embora uma das bases de dados utilizadas seja convencional, os autores não demonstraram os resultados para cada uma das bases separadamente. Desse modo, não foi possível avaliar o desempenho do método em imagens convencionais.

A Tabela 2.1 constitui um resumo dos trabalhos abordados na Seção 2. São apresentados na primeira coluna os autores dos trabalhos, na segunda coluna as bases de dados e na última coluna a solução proposta pelos autores para segmentação.

Devido ao seu destaque na literatura, o presente trabalho utiliza-se como referência as metodologias propostas por Ronneberger, Fischer e Brox (2015) e Kendall, Badrinarayanan e Cipolla (2015). Além disso, através das bases de dados públicas ISBI2015 e Herlev presentes na literatura pretende-se validar a eficácia da metodologia escolhida. Ainda, destaca-se a escassez de base de dados fiéis às amostras recolhidas de forma tradicional na literatura, por conseguinte a aplicação da metodologia a base de dados *CRIC Cervix* difere-se dos demais trabalhos.

## 2.2 Fundamentação Teórica

Nesta seção serão apresentados algumas definições relacionados ao tema em questão. Desse modo, serão apresentados a formalização de métodos importantes. Ainda, a seção foi dividida em quatro subseções: 2.2.1 Câncer de Colo de Útero e Segmentação 2.2.2 para definição do

Trabalhos	Bases de dados	Método para segmentação
Ushizima, Bianchi e Carneiro (2015)	ISBI2014	Super-píxeis e diagrama de Voronoi
Long, Shelhamer e Darrell (2015)	Pascal VOC 2011, NYUDv2, SIFT Flow	FNC
Ronneberger, Fischer e Brox (2015)	ISBI2012	U-net
Kendall, Badrinarayanan e Cipolla (2015)	CamVid	SegNet
Sharma e Mangat (2016)	Herlev	FCM e BPNN
Zhang et al. (2017)	HEMLBC base coletada em meio líquido	-
Saha, Bajger e Lee (2017)	ISBI2014	CiscFC
Phoulady et al. (2017)	ISBI2014	Método iterativo de binarização
Gautam et al. (2018)	Herlev	CNN com pré-processamento seletivo
Liu et al. (2018)	Herlev	Mask-RCNN LFC-CRF
Zhao et al. (2019)	Herlev	PGU-net+
Zhang et al. (2019)	ISBI 2014 e base de dados do hospital chinês Guangzhou	Arquitetura com topologia em rede baseada na ResNeXt e algoritmo TFA
Aratijo et al. (2019)	CRIC Cervix	LeNet
Allehaibi et al. (2019)	Herlev	Mask R-CNN
Wan et al. (2019)	ISBI2015, ISBI2014 e INHouse	TernausNet
Sabeena, Gopakumar e Thampi (2020)	Herlev	FC DenseNet56
Zou et al. (2020)	2018 Kaggle Science Bowl e Becton-Dickinson (BD)	Mask R-CNN
Xiang et al. (2020)	Base de dados privada coletada em meio líquido	-
Hussain et al. (2020)	Herlev, bases de dados privadas coletadas em meio líquido e convencional	Arquitetura baseada na U-net
Yang, Wu e Yin (2020)	Herlev	Interacting convolution with pyramid structure network
Huang et al. (2020)	ISBI2014 e ISBI15	FCM
Huang e Zhu (2020)	ISBI2014	GGCMM
Win et al. (2020)	SIPaKMeD e Herlev	Watershed transformation
Wang et al. (2020)	ISBI2015	FCM e ADLS
Hoque et al. (2021)	ISBI2014 e BSMMU	Método baseado em propriedade de contorno

Tabela 2.1 – Resumo dos trabalhos relacionados a segmentação

problema, 2.2.3 CNNs para apresentação de abordagens do método e 2.2.5 métricas para fim de comparação dos resultados.

### 2.2.1 Câncer de colo de útero

O Câncer de Colo de Útero é uma lesão invasiva intrauterina ocasionada principalmente por subtipos de alto risco do HPV, o papilomavírus humano (COHEN et al., 2019). Por causa do vírus, ocorrem alterações celulares que podem ser diagnosticadas pelo exame criado por Georgios Papanikolaou (GREENWOOD; MACHADO; SAMPAIO, 2006). Comumente conhecido como exame de Papanicolaou, permite a coleta de células no colo uterino que são enviadas para a análise microscópica em laboratórios especializados. É considerado um exame rápido e de baixo custo (MARTINS; THULER; VALENTE, 2005).

A obtenção dos esfregaços utilizando o método convencional a partir do exame baseia-se em seis etapas (REZENDE, 2018). A primeira etapa consiste na coleta através de uma espátula para coletar células da ectocérvice do órgão e da escova endocervical para coletar células da endocérvice. Em seguida, o material é disposto em uma lâmina e fixado em álcool. Ainda, as amostras passam pelas etapas de coloração e montagem para, por fim serem classificadas através da análise visual microscópica. As células são classificadas de acordo com o sistema Bethesda, que classifica as células de acordo com o nível das lesões, desde normais até lesões de alto grau.

Segundo Instituto Nacional de Câncer (INCA, 2020), o câncer de colo do útero apresenta a cada ano 570 mil novos casos no mundo. Segundo Ferlay et al. (2015), mais de 80% de todos os casos de câncer cervical são encontrados em países com poucos recursos, devido à falta de conscientização e dificuldade em executar programas de rastreamento. Em países desenvolvidos,

a incidência foi significativamente reduzida (SACHAN et al., 2018). Ainda, a estimativa para cada ano do triênio de 2020-2022 no Brasil é 16590 casos.

Bray et al. (2018) apontam que os índices de incidência e mortalidade por essa neoplasia maligna do trato genital estão fortemente ligados ao Índice de Desenvolvimento Humano (IDH) dos países. Ainda, segundo Torre et al. (2017) quase 90% das mortes por câncer cervical ocorrem em países em desenvolvimento. Desse modo, as taxas são um reflexo do acesso à detecção precoce e tratamento adequado.

Um dos erros mais comuns atrelados ao método de rastreamento é a incidência de falsos negativos (MITTELDORF, 2016). Isso se deve a subjetividade atrelada a interpretação do exame, visto que cada amostra pode conter até 300 000 células com diferentes orientações e sobreposições (RAHAMAN et al., 2020). Dessa forma, a qualidade da análise é diretamente relacionada a experiência do profissional. Nesse aspecto, uma alternativa é a inserção da citologia digital para monitoramento de qualidade da rotina laboratorial, de modo a proporcionar um auxílio na avaliação das amostras e diminuir a fadiga física e mental do especialista.

## 2.2.2 Segmentação semântica

A segmentação de imagens é uma das tarefas mais indispensáveis na visão computacional. A partir da segmentação é possível segregar várias regiões de acordo com propriedades presentes em cada pixel. Para facilitar a análise, o principal objetivo da segmentação é extrair informações significativas (SULTANA; SUFIAN; DUTTA, 2020).

A segmentação semântica pode ser definida como a tarefa de determinar uma classe a cada píxel (WEI et al., 2016). Em contraste a segmentação de instâncias, a segmentação semântica não difere instâncias distintas da mesma estrutura (MATCHA, 2021). A Figura 2.1 exibe um exemplo dessa diferenciação.

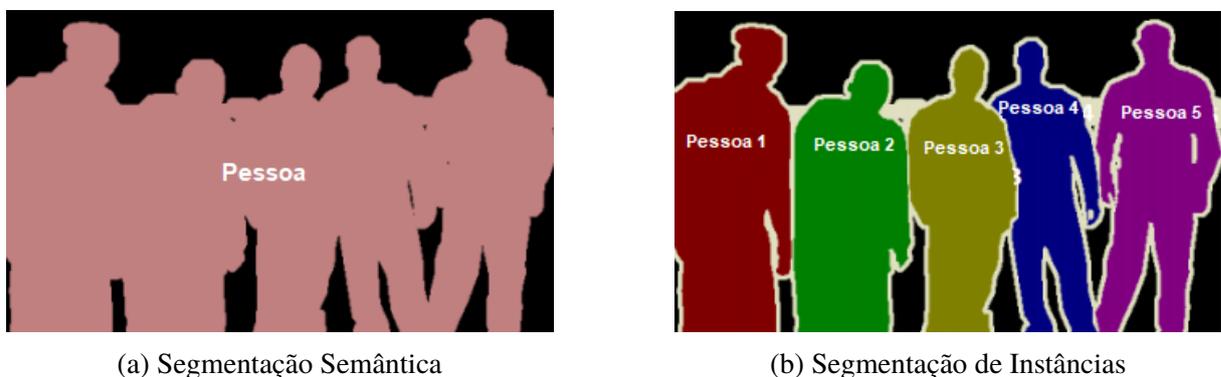


Figura 2.1 – Na Figura (a) cada píxel apresenta uma classificação. Já a Figura da (b), cada objeto possui uma classificação. (VARATHARASAN et al., 2019)

Para o rastreamento de lesões no colo do útero, determinar a região da célula possui alta relevância para o diagnóstico. Ainda, como mencionado na Seção 1.1, o núcleo de uma célula

anormal apresenta variações morfológicas e texturais (DINIZ, 2019), tornando sua segmentação uma peça chave para a automatização da detecção de lesões.

### 2.2.3 CNNs

As CNNs são algoritmos de aprendizado profundo que se sobressaem entre outros métodos por serem capazes de construir filtros baseados nas características das imagens de entrada, dispensando pré processamento. Além disso, em uma Rede neural convolucional as imagens de entrada percorrem diversas camadas de convolução, onde cada camada é responsável pela extração de um conjunto de características das amostras da base de dados. As CNNs podem ser treinadas de ponta a ponta pelo método de gradiente descendente, além de manter a invariância de translação e escala (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Desse modo, o desempenho das CNNs no reconhecimento de imagens está fortemente ligado a sua capacidade de processamento e aprendizado (ZHIQIANG; JUN, 2017).

Historicamente, o conceito de Rede neural convolucional existe desde 1980 através do trabalho de Fukushima e Miyake (1982), no qual foi apresentado um modelo de reconhecimento visual, que aprenderia com a repetição de padrões de estímulos. Em 1998, LeCun et al. (2015) propuseram a arquitetura LeNet. Baseado no problema de identificação de dígitos de códigos postais manuscritos, este trabalho foi marcante na aplicação real de algoritmos de retro-propagação. Contudo, com a capacidade limitada do equipamento de *hardware*, especialmente de Unidades Gráficas de Processamento (*Graphics Processing Unit - GPU*) as Redes neurais como método para visão computacional foram deixadas de lado.

Anos mais tarde, o surgimento da AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) foi uma marco para a Inteligencia Artificial. Tomando o primeiro lugar no desafio de reconhecimento visual em grande escala da ImageNet em 2012, a arquitetura conta com 8 camadas e 60 milhões de parâmetros. Ainda, utilizou a função Unidade Linear Rectificada (ReLU) como função de ativação e a técnica de *Dropout* para conter o sobre-ajuste da rede.

Posteriormente, as arquiteturas VGGNet (SIMONYAN; ZISSERMAN, 2014), Inception (SZEGEDY et al., 2015) e ResNet (HE et al., 2016) se destacaram pela performance durante o desafio anual de reconhecimento visual em grande escala da ImageNet. Desse modo, novas técnicas tem surgido com o objetivo de aumentar a performance das CNNs.

A Figura 2.2 apresenta um exemplo de rede neural convolucional para classificação com Camadas de *Pooling* e Convolução para extração de características. Para melhor entendimento a respeito dos conceitos presentes neste trabalho, este capítulo apresenta a formalização teórica das camadas de uma Rede neural convolucional e principais definições envolvidas.

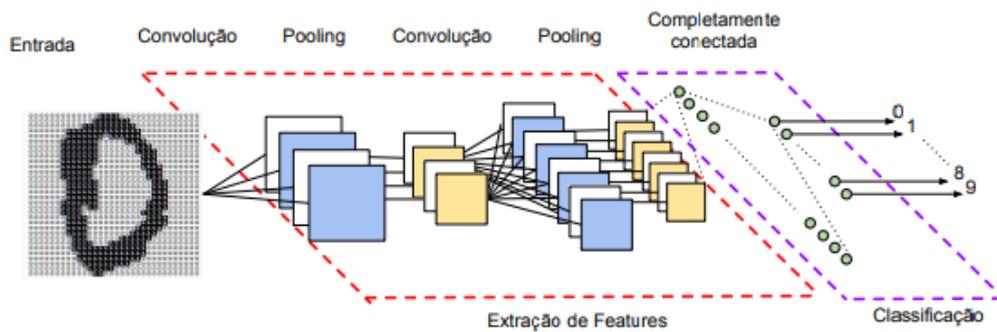


Figura 2.2 – Exemplo de uma rede neural convolucional (VARGAS; PAES; VASCONCELOS, 2016)

### 2.2.3.1 Camada de convolução

Uma imagem digital é uma matriz de largura, altura e profundidade, na qual cada célula representa a intensidade dos pixels, com seus valores no intervalo de 0 a 255 (KARPATHY, 2015). A profundidade determina a quantidade de canais de cores, visto que se a imagem estiver em escala de cinza possui apenas um canal, e se tratando de imagens RGB (*Red, Green, Blue*) três canais. Para que seja possível a extração de características de uma imagem é necessário a aplicação de filtros ou *kernels*, através da convolução. Matematicamente, uma convolução é a medida integral da área de sobreposição entre duas funções (JANGIR, 2019). A operação de convolução em uma imagem digital é representada pela soma de matrizes ponderada por um filtro, na qual resulta na similaridade entre elas. A Figura 2.3 mostra a operação de convolução com um filtro detector de diagonais secundárias.

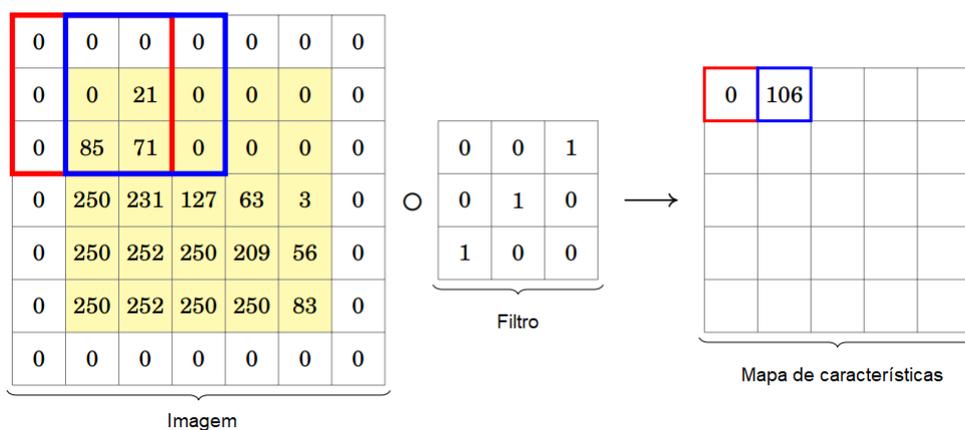


Figura 2.3 – Operação de convolução com preenchimento de zero as bordas e salto igual a 1 (PAVLOVSKY, 2019)

O neurônio em uma Rede neural convolucional é representado pelos filtros que se comportam como detectores de características. Este, é inicializado aleatoriamente e tem seus pesos atualizados à medida que o aprendizado é propagado pela rede. Essa operação constrói o conceito

de campo receptivo local, visto que dada a representação numérica do objeto a ser analisado, seleciona subconjuntos com determinado atributo (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

Dado o volume de entrada  $L_1 * A_1 * D_1$  onde:

- Largura da imagem  $L_1$
- Altura da imagem  $A_1$
- Número de canais de cores  $D_1$

Segundo Karpathy (2015), na camada de convolução temos um conjunto de filtros aplicados ao volume de entrada. Ainda, a sequência de mapas de características resultante da operação decorre da configuração dos principais hiper-parâmetros:

- Número de filtros ou *kernels*  $K$
- Extensão espacial dos filtros  $F$
- Espaçamento para aplicação dos filtros sucessivos, conhecido como salto (ou *stride*)  $S$
- Preenchimento de zeros as bordas ou *padding*  $P$

Realizando a convolução sobre os volumes  $A_1$ ,  $L_1$  e  $D_1$  usando os hiper-parâmetros citados anteriormente, temos como saída da camada de convolução o volume  $L_2$ ,  $A_2$ , e  $D_2$  apresentados na Equação 2.1.

$$L_2 = (L_1 - F + 2 * P) / S + 1$$

$$A_2 = (A_1 - F + 2 * P) / S + 1$$

$$D_2 = K$$

(2.1)

Por fim, o volume de saída  $L_2 * A_2 * D_2$ , onde  $L_2$ ,  $A_2$  e  $D_2$  representam respectivamente, a largura, altura e quantidade de mapas de características resultantes da operação.

### 2.2.3.2 Função de ativação

Responsáveis pela não-linearidade da rede, as funções de ativação determinam como ocorrerá o aprendizado, visto que seu resultado indica se o neurônio será ou não ativado (ALVES, 2018). Ademais, a função é aplicada como passo seguinte a convolução, e ao final de um ciclo de treinamento a atualização dos pesos é feita através da direção do gradiente da função escolhida.

Na Equação 2.2  $\hat{y}$ , indica a saída da operação,  $\sigma$  indica a aplicação da função de ativação,  $w_i$  os pesos,  $x_i$  a entrada da rede e  $b$  o bias.

$$\hat{y} = \sigma\left(\sum (w_i * x_i) + b\right) \quad (2.2)$$

Existe uma variedade de funções que podem ser utilizadas como função de ativação, a principal delas atualmente é a ReLU. Aplicada a imagem, a função *ReLU* atribui o maior valor entre zero e o valor do píxel para cada um dos canais de cores, conforme a Equação 2.3. Pela sua não-linearidade, permite a rede a melhora na detecção de padrões e evita a dissipação do gradiente com o aumento da complexidade da rede (WANG, 2019).

$$ReLU(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.3)$$

Embora tenha um papel importante no controle da dissipação de gradiente, a função ReLU tem como principal desvantagem a morte dos neurônios. Se um neurônio tem seu valor sempre negativo sua saída sempre será zerada. Deste modo, para tentar contornar essa desvantagem serão apresentadas funções de ativação alternativas a ReLU.

Apesar de muito semelhante a ReLU, a proposta da função de ativação LeakyRelu (XU et al., 2015) é contornar a morte dos neurônios por meio de uma constante  $\alpha$ , em oposição a zerar a saída do neurônio torna o valor menor mantendo a proporcionalidade da entrada. A função está presente na Equação 2.4.

$$LeakyReLU(x) = \begin{cases} x & x > 0 \\ \alpha x & x \leq 0 \end{cases} \quad (2.4)$$

A função ELU (*Exponential Linear Unit*) (CLEVERT; UNTERTHINER; HOCHREITER, 2015), para valores positivos se comporta da mesma forma que a função ReLU. A diferença aparece para valores negativos que assumirão um valor muito próximo e inferior a zero. O parâmetro  $\alpha$  pertence a um intervalo de [0.1,0.3]. Ainda, pela função aproximar o valor a zero, tem efeitos parecidos a normalização em lote. Contudo, a função apresenta custo computacional maior. A função está descrita na Equação 2.5

$$ELU(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases} \quad (2.5)$$

Outra variação da ReLU, a Swish (RAMACHANDRAN; ZOPH; LE, 2017), evita a saturação em um valor máximo, visto que a função é ilimitada superiormente. Através da base de dados MNIST (LECUN; CORTES, 2010) a função Swish apresentou melhor performance para arquiteturas mais profundas (NERD, 2018). Similar a ELU, possui maior custo computacional devido ao

exponencial. A função está disponível na Equação 2.6. O valor  $\beta$  representa um parâmetro treinável.

$$Swish(x) = x \cdot \text{sigmoid}(\beta x) \quad (2.6)$$

Inspirada na Swish, a função de ativação Mish (MISRA, 2019) apresenta vantagens e desvantagens similares. Além disso, a função Mish demonstrou maior acurácia com o modelo Squeeze Excite Net-18 (HU; SHEN; SUN, 2017) para a base de dados CIFAR-100 (KRIZHEVSKY; HINTON et al., 2009). A função Mish está presente na Equação 2.7.

$$Mish(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (2.7)$$

Para a última camada da rede é comum a aplicação de uma função que mapeie os valores em um intervalo de  $[0,1]$  (SHARMA, 2019). A função Sigmoid, presente na Equação 2.8, é usada principalmente em casos na qual é necessário prever a probabilidade do dado pertencer a determinada classe (NWANKPA et al., 2018).

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.8)$$

### 2.2.3.3 Função de perda

De acordo com Wu (2017), a função de perda ou *loss* é empregada a última camada da rede para medir a discrepância entre uma predição  $\hat{y}_i$  e a verdade terrestre  $y_i$  para a  $i$ -ésima. Para classificação binária, a função entropia cruzada binária ou *Binary crossentropy* é comumente empregada - função disponível na equação 2.9, a variável  $n$  representa a dimensão do volume de entrada.

$$\text{Loss}(y) = -\frac{1}{n} \sum_{i=1}^n y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i) \quad (2.9)$$

### 2.2.3.4 Camada de subamostragem

A Camada de subamostragem ou *Pooling* consiste na redução da dimensionalidade do mapa de características para diminuir o número de parâmetros e consequentemente, a complexidade do modelo (LOY, 2019). Para CNNs uma escolha comum é a utilização da função *Max Pool*, esta operação tem como objetivo capturar os pixels mais significativos da imagem, visto que seleciona o maior valor de uma vizinhança pré-definida. A Figura 2.4 apresenta o antes e depois da operação de *Max Pool* sobre o mapa de características. Outras opções de operações são a *Mean Pool* e *Min Pool* que capturam respectivamente, a média e o mínimo dentro de uma janela de pixels.

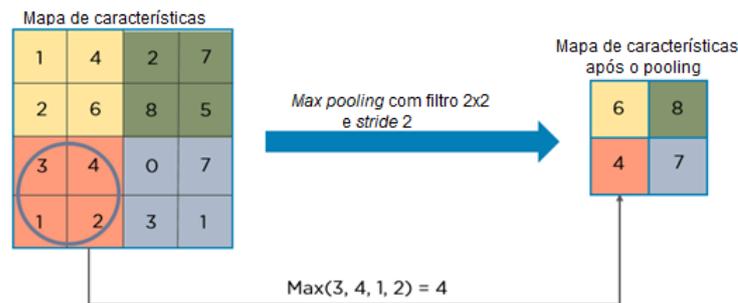


Figura 2.4 – Exemplo de aplicação do *Max Pooling* tamanho 2x2 com salto 2 sobre uma mapa de características (ALBAWI; MOHAMMED; AL-ZAWI, 2017)

### 2.2.3.5 Dropout

Para prevenir o sobreajuste, a técnica *Dropout* em aprendizagem de máquina consiste na remoção temporária de unidades, dada uma probabilidade  $P$  obtida durante o treinamento (SRIVASTAVA et al., 2014). A remoção ocorre por meio da atribuição do valor zero a saída do neurônio selecionado aleatoriamente. A principal vantagem dessa técnica, é o aumento da generalização do modelo, haja vista que força a rede a extrair características mais robustas. Em contraste, uma desvantagem é a perda do aprendizado, resultante da morte do neurônio. A Figura 2.5 mostra uma rede antes e depois da aplicação do *Dropout*.

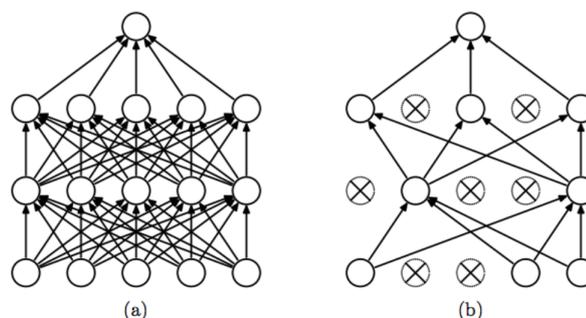


Figura 2.5 – (a) Modelo original de rede neural (b) Modelo após a utilização da técnica *Dropout* (SRIVASTAVA et al., 2014)

### 2.2.3.6 Normalização em lote

Assim como o *Dropout*, a Normalização em lote (IOFFE; SZEGEDY, 2015) também possui como principal objetivo a generalização do modelo. Além disso, a Normalização em lotes aumenta a velocidade de aprendizado da rede, visto que mapeia os valores em um intervalo, o que torna conjunto de dados mais simétrico e permite ao modelo ser treinado com taxa de aprendizagem maior. Ainda, diminui a importância dos pesos iniciais, haja vista que o intervalo na qual os pesos serão inicializados é inferior. A gráfico da Figura 2.6 demonstra a comparação do desempenho da rede com e sem Normalização em lote.

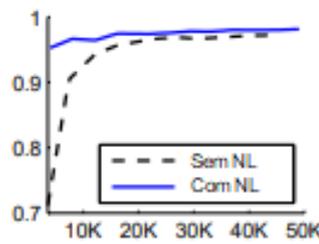


Figura 2.6 – Comparação da acurácia de teste e número de etapas do treinamento para modelo com e sem Normalização em lote. É possível notar que a presença de normalização em lote induz o modelo a convergir com menos etapas de treinamento. (IOFFE; SZEGEDY, 2015)

### 2.2.3.7 Camada de sobreamostragem

A Camada de sobreamostragem é responsável por aumentar a resolução espacial da imagem. Na segmentação semântica, é muito utilizada para retornar a imagem de entrada a sua resolução inicial, após varias operações de subamostragem e convolução. Entre as técnicas de sobreamostragem um exemplo é a Camada de convolução Transposta, utilizada pela sua capacidade de aproveitar o aprendizado (MISHRA, 2020). Nela, cada elemento da matriz de entrada é multiplicada pelo *kernel* e os valores sobrepostos são somados. A Figura 2.7 mostra o resultado da convolução transposta. Outra técnica é a camada de escalonamento simples, que copia os valores dos pixels mais próximos, conforme a Figura 2.8.

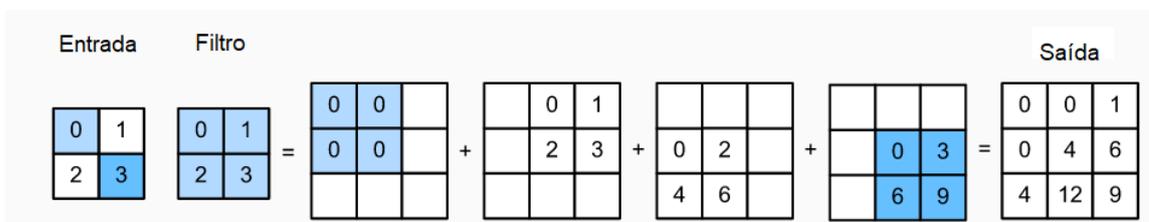


Figura 2.7 – Operação de convolução transposta com salto 1 e sem adição de zeros nas bordas (*padding*) (MISHRA, 2020)

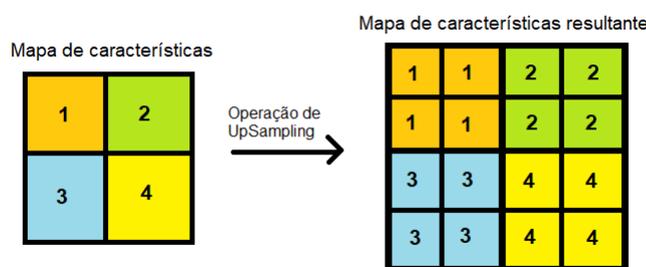


Figura 2.8 – Operação de sobreamostragem com proporção (2,2) (CHAUBEY, 2020)

### 2.2.3.8 Época

O hiper-parâmetro época determina o número de ciclos de aprendizado que o método executará em toda a base de dados de treinamento. Desse modo, quando definido apenas uma época, os parâmetros internos serão atualizados com cada amostra de treinamento (BROWNLEE, 2019).

### 2.2.3.9 Tamanho do lote

O tamanho do lote é o número de amostras utilizadas no treinamento antes da atualização dos pesos (BROWNLEE, 2019). Tamanhos de lote pequenos tem efeito na generalização do modelo, a medida que os dados são interpretados como mais ruidosos (MASTERS; LUSCHI, 2018).

Segundo Bengio (2012), quanto maior o tamanho do lote mais rápido é o treinamento. Contudo, devido ao número de atualizações por época ser menor requer visitas a mais exemplos para chegar ao mesmo erro. Além disso, tamanho do lote fica limitado também a memória disponível, devido ao número de amostras sendo propagadas pela rede. Por fim, o autor recomenda o tamanho de lote com 32 amostras como valor padrão.

### 2.2.3.10 Taxa de aprendizado

A taxa de aprendizado ou *learning rate*, dimensiona a magnitude das atualizações dos pesos para minimizar a função de perda da rede (JORDAN, 2020). A Figura 2.9 apresenta a análise do hiper-parâmetro, através da curva de aprendizado.

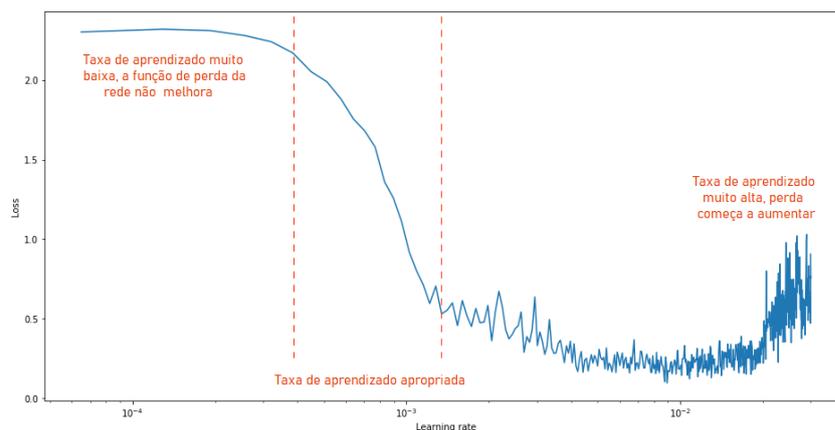


Figura 2.9 – Análise da taxa de aprendizado através da função de perda (JORDAN, 2020)

Segundo Goodfellow, Bengio e Courville (2016), para definir a taxa de aprendizado inicial deve-se analisar o desempenho da rede após um número maior ou igual a 100 épocas. Ainda, a taxa de aprendizado ideal consiste em um número maior a que produz o melhor desempenho, mas não tão alta a ponto de causar instabilidade.

## 2.2.4 Aprendizado por transferência

Em comparação a métodos tradicionais de Aprendizado de Máquina, o aprendizado profundo está vinculado a dependência de uma quantidade em larga escala de dados para treinamento. Construir uma base de dados com milhares de amostras é um trabalho extremamente complexo, além de estar suscetível ao elevado custo da rotulação (TAN et al., 2018).

O Aprendizado por transferência explora o conhecimento previamente adquirido para facilitar a modelagem preditiva (LU et al., 2015). Primeiramente, a rede é treinada com uma base de dados em larga escala rotulada. Em seguida, os valores dos parâmetros das camadas internas são transferidos para uma nova rede. Para compensar a diferença entre as imagens das duas bases de dados, são adicionadas camadas de adaptação para que novas características da nova base de dados possam ser aprendidas (OQUAB et al., 2014). Congelando-se as camadas que receberam os parâmetros aprendidos, é possível diminuir o custo computacional da rede, haja vista que não precisarão ser atualizados durante o treinamento. A Figura 2.10 demonstra essa abordagem.

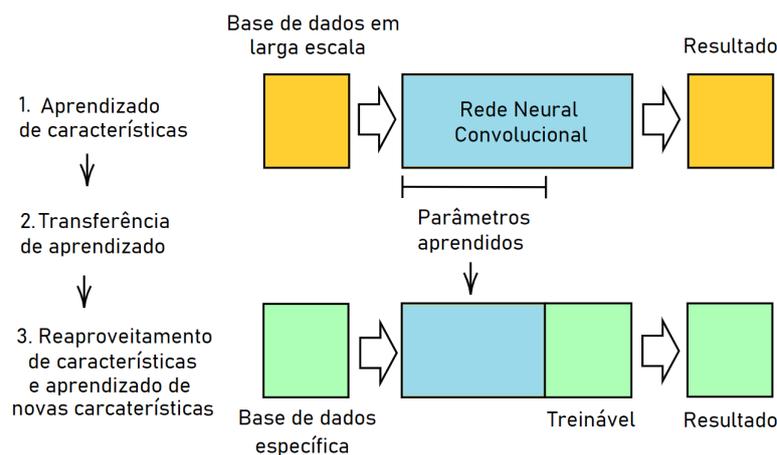


Figura 2.10 – Aprendizado por transferência

Os trabalhos de Allehaibi et al. (2019), Wan et al. (2019), e Xiang et al. (2020) presentes na Seção 2.1 apostaram em modelos pré-treinados para contornar, além do número de amostras celulares disponíveis, o impacto da inicialização dos pesos.

## 2.2.5 Métricas

Para determinar a qualidade da segmentação, será mensurado o número de verdadeiros positivos (VP), visto que apontam os píxeis classificados corretamente como pertencentes ao núcleo, verdadeiros negativos (VN), que denotam os píxeis não pertencentes aos núcleos classificados corretamente como fundo, falsos positivos (FP), que se referem a píxeis do fundo classificados erroneamente como núcleo e falsos negativos (FN), píxeis pertencentes ao núcleo classificados erroneamente como pertencentes ao fundo. As medidas estatísticas são descritas a seguir.

Segundo Amorim et al. (2020) entre as métricas mais recorrentes para segmentação de esfregaços de Papanicolaou, no período de 2016-2020 estão acurácia e coeficiente de Dice, seguidas de revocação e precisão. O resultado desse estudo está presente na Figura 2.11.

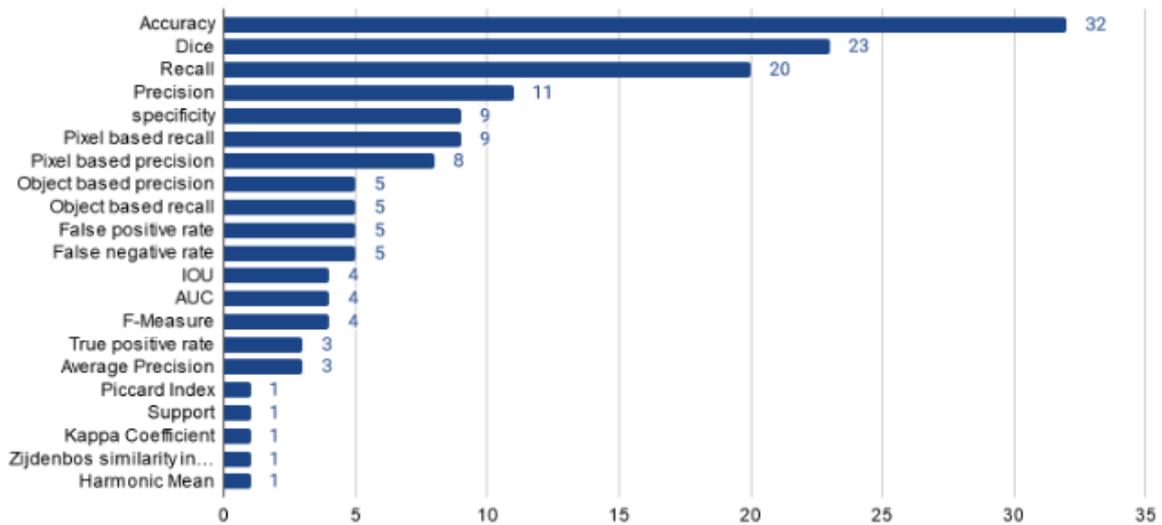


Figura 2.11 – Número de trabalhos que utilizaram cada métrica (AMORIM et al., 2020)

### 2.2.5.1 Acurácia

A acurácia aponta a performance geral do modelo. Deste modo, o seu resultado indica a porcentagem de classificações verdadeiramente negativas e positivas dentre toda a classificação do conjunto de dados. É determinada pela Equação 2.10.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.10)$$

### 2.2.5.2 Coeficiente de Dice

A métrica visa a comparação entre atributos comuns a dois conjuntos de dados. O cálculo da similaridade é dado através da comparação de regiões segmentadas pela nossa proposta e pela verdade terrestre, ou *Ground Truth*. A fórmula é igual a duas vezes o número de elementos comuns a ambas as imagens (classificado corretamente) dividido pela soma do número total de pixels em cada uma, conforme Equação 2.11.

$$\text{Dice} = \frac{2VP}{2VP + FP + FN} \quad (2.11)$$

### 2.2.5.3 Revocação

A revocação ou *Recall*, é a porcentagem de dados positiva que foram corretamente classificados com positivos. Ainda, é conveniente em situações na qual os falsos negativos são mais prejudiciais que os falsos positivos (RODRIGUES, 2020). É definido pela Equação 2.12.

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (2.12)$$

#### 2.2.5.4 Precisão

A precisão tem o intuito de expor dentre as classificações positivas do resultado da nossa proposta, quantas estão corretas. É uma opção favorável quando os falsos positivos são mais prejudiciais comparados aos falsos negativos (RODRIGUES, 2020). É definido pela Equação 2.13.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.13)$$

## 3 Desenvolvimento

Neste capítulo são especificadas as fases do desenvolvimento deste projeto. Na Seção 3.1 são descritas as bases de dados e suas respectivas metodologias, na Seção 3.2 são apresentadas as operações para aumento de dados, na Seção 3.3 são especificadas as arquiteturas de rede utilizadas, e por fim na Seção 3.4 o treinamento.

### 3.1 Base de dados e metodologia

Nessa seção serão descritas as bases de dados de utilizadas nesse trabalho. Na literatura, existe uma gama de trabalhos que abordam a segmentação de núcleos de citologia cervical. Contudo, os conjuntos de dados disponíveis apresentam escassez de amostras voltadas ao diagnóstico e a variabilidade das células.

A base de dados Herlev (MARINAKIS; DOUNIAS; JANTZEN, 2009), descrita na Seção 3.1.1, composta de 917 núcleos e citoplasmas segmentados contempla imagens reais de células individualizadas e classificadas acerca de seus tipos celulares. O conjunto de dados ISBI2014(LU; CARNEIRO; BRADLEY, 2015) apresentado na Seção 3.1.2, tem em sua composição imagens sintéticas de células cervicais geradas a partir de um algoritmo que reorganiza células segmentadas através de uma taxa de sobreposição. O conjunto conta com 45 imagens para treino e 90 imagens para validação e 810 para teste.

Por fim, a base de dados *CRIC Cervix* (REZENDE; BIANCHI; CARNEIRO, 2021) descrita na Seção 3.1.3 conta com 3322 núcleos e citoplasmas segmentados e seus tipos de lesões classificadas de acordo com o sistema Bethesda (NAYAR; WILBUR, 2015), nomenclatura utilizada para diagnósticos médicos. Ademais, a principal vantagem deste conjunto de dados além da quantidade de núcleos segmentados é a proximidade de situações reais, tendo em vista a composição de imagens sem pré-processamento, extraídas diretamente de amostras. Desse modo, através da aplicação dos métodos sob a base de dados obteremos a aplicabilidade deste método para automatização do diagnóstico.

Cada base de dados se distingue em dimensionalidade e quantidade de informações sobre as regiões segmentadas. Desse forma, nessa seção serão apresentadas as bases de dados e suas diferentes abordagens para cada conjunto de dados de acordo com suas características. Em seguida, as imagens serão submetidas ao aumento de dados.

#### 3.1.1 Herlev

A base de dados Herlev (MARINAKIS; DOUNIAS; JANTZEN, 2009) contém 917 imagens de células de Papanicolaou, classificadas por cito-técnicos e médicos no Hospital

Universitário Herlev, na Dinamarca. Cada célula é descrita por 20 características numéricas, e as células se dividem em 7 classes, conforme a Tabela 3.1. As classes 1 – 3 (242 células) representam células normais e 4 – 7 (675 células) são anormais como mostra a Figura 3.1. A base é composta por imagens contendo apenas uma célula, e possui verdade terrestre para segmentação - regiões do núcleo, citoplasma e fundo delimitadas.

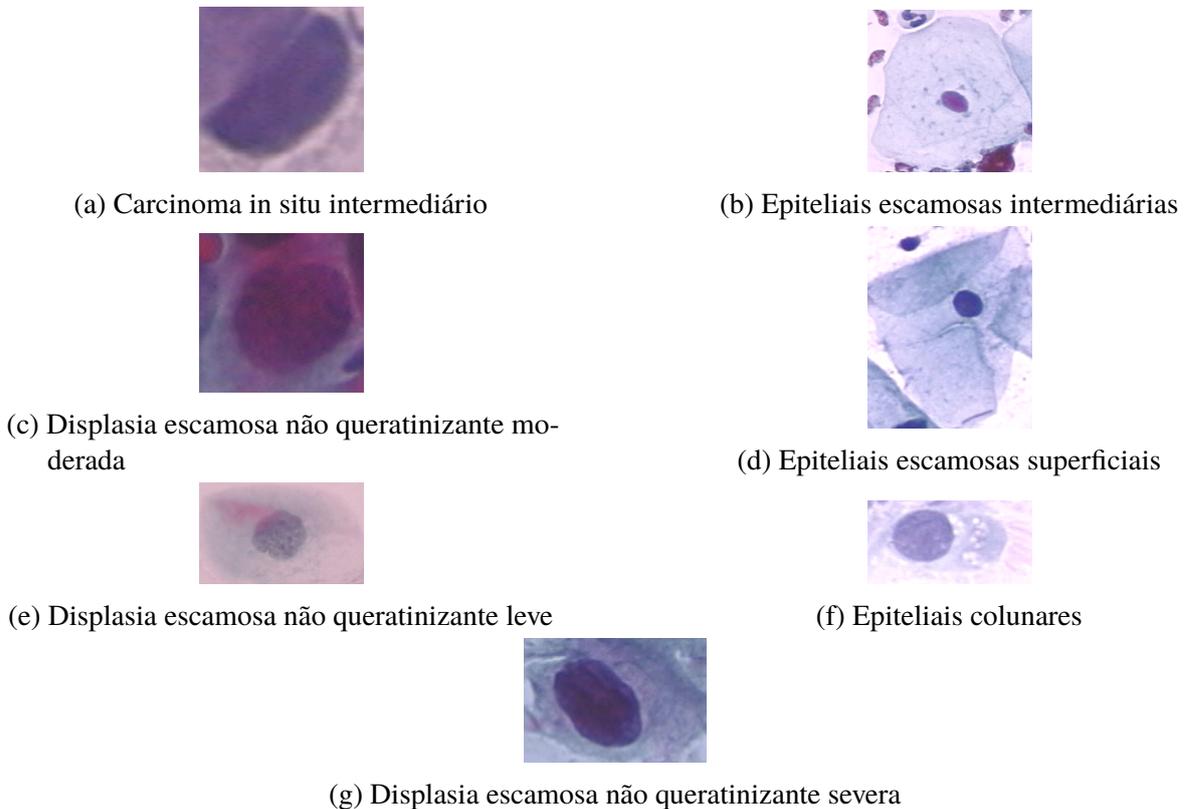


Figura 3.1 – Classificações da base de dados Herlev

Classe	Categoria	Tipo de célula	Total de núcleos segmentados
1	Normal	Epiteliais escamosas superficiais	74
2	Normal	Epiteliais escamosas intermediárias	70
3	Normal	Epiteliais colunares	98
4	Anormal	Displasia escamosa não queratinizante leve	182
5	Anormal	Displasia escamosa não queratinizante moderada	146
6	Anormal	Displasia escamosa não queratinizante severa	197
7	Anormal	Carcinoma espinocelular in situ intermediário	150

Tabela 3.1 – Distribuição de classes Herlev

Tendo em vista que a base de dados Herlev já possui as regiões de núcleo e citoplasma delimitadas, foi necessário redimensionar as imagens para o padrão 128 x 128 pixels devido as diferentes dimensões apresentadas pelas células. Embora essa operação possa impactar na classificação posterior, é a técnica mais utilizada na literatura para contornar a diferença de dimensionalidade entre as imagens. Em seguida, base de dados foi dividida de  $k$  maneiras distintas em treinamento e teste. A partir da base de treinamento, são criadas duas partições:

treino e validação. Por fim, as amostras são submetidas a operações de aumento de dados. Na Figura 3.2, é exibido um resumo da metodologia aplicada a base de dados.

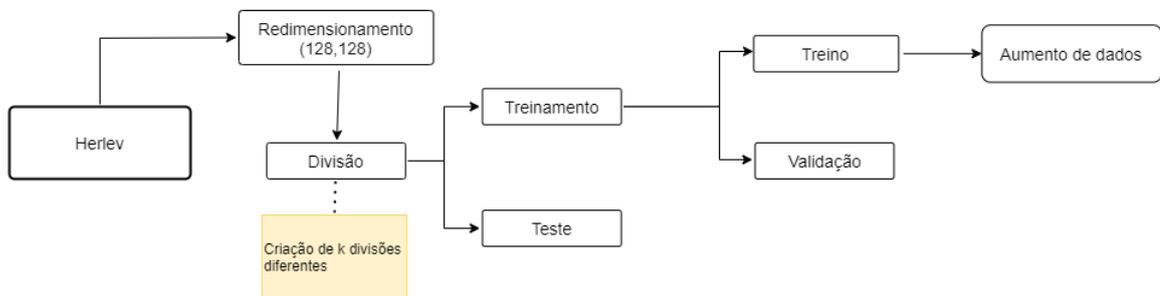
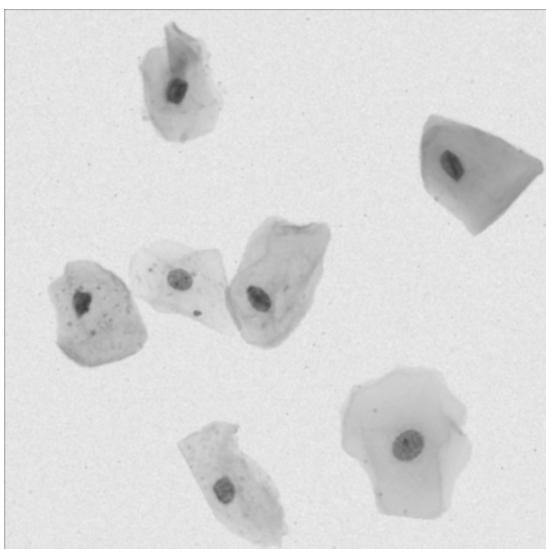


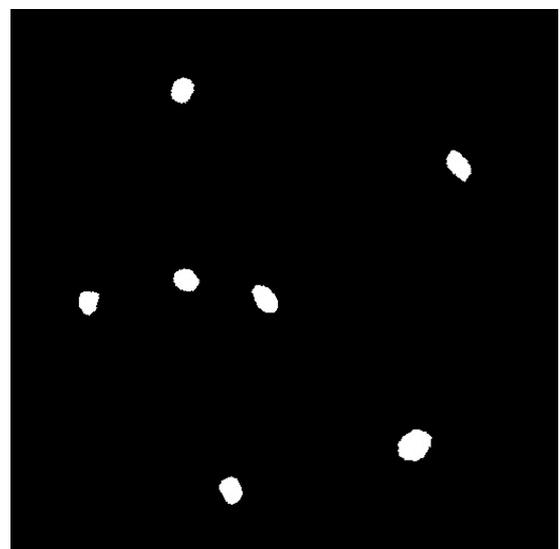
Figura 3.2 – Fluxograma da metodologia para a base de dados Herlev

### 3.1.2 ISBI2014

O conjunto de dados ISBI2014 (LU; CARNEIRO; BRADLEY, 2015) é composto por imagens com profundidade de foco estendida. A base de dados é constituída por 45 imagens para treinamento, 90 para validação e 810 para teste. A Figura 3.3 apresenta um exemplo de imagem produzida sinteticamente, presente na base de dados e sua verdade terrestre para segmentação.



(a) Amostra da base de dados ISBI2014



(b) Verdade Terrestre

Figura 3.3 – Verdade terrestre dos núcleos

Para adequar as imagens da base de dados ISBI2014 à entrada esperada pelas arquiteturas, foi aplicado o redimensionamento das imagens. A desproporcionalidade entre a quantidade de pixels com núcleo e fundo, pode afetar o desempenho dos modelos. Para contornar essa situação, a base de dados de treino foi sucedida à recortes, como uma forma de aumento de dados. Como a base já está dividida entre treino e teste, mantivemos as divisões já existentes. Sobre a base de treino fizemos o aumento de dados nas imagens inteiras e recortes de 128 x 128. Nesses recortes

eram excluídos os que não possuíam informação de núcleo. Nesse sentido, além das imagens completas e aumento de dados a base de treino também apresenta recortes aleatórios de 128 x 128. Um fluxograma de toda a metodologia aplicada a base de dados ISBI2014 está presente na Figura 3.4.

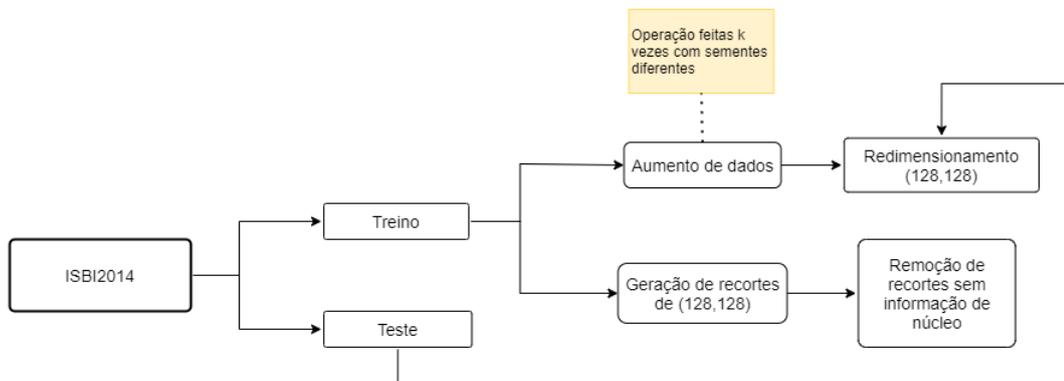


Figura 3.4 – Fluxograma da metodologia para a ISBI2014

### 3.1.3 CRIC Cervix

A plataforma de banco de dados de imagens pesquisáveis CRIC (REZENDE; BIANCHI; CARNEIRO, 2021), inclui a *CRIC Cervix*, uma coleção de imagens de células cervicais que foram segmentadas e classificadas por citopatologistas independentes. A foto-documentação das amostras foi realizada no Laboratório Multiusuários de Microscopia Avançada e Microanálise do Núcleo de Pesquisas em Ciências Biológicas, Universidade Federal de Ouro Preto, MG. A base de dados conta com 3322 núcleos e citoplasmas celulares segmentados obtidos a partir de 400 imagens. As células na coleção *CRIC Cervix* são classificadas em seis (6) classes: (1) negativo para lesão intraepitelial ou malignidade (NILM); (2) células escamosas atípicas de significado indeterminado, possivelmente não neoplásicas (ASC-US); (3) lesão intraepitelial escamosa de baixo grau (LSIL); (4) células escamosas atípicas, não pode excluir lesão de alto grau (ASC-H); (5) lesão intraepitelial escamosa de alto grau (HSIL); e (6) carcinoma de células escamosas (SCC) conforme a Figura 3.5. A distribuição de classes com o número de segmentações está presente da Tabela 3.2.

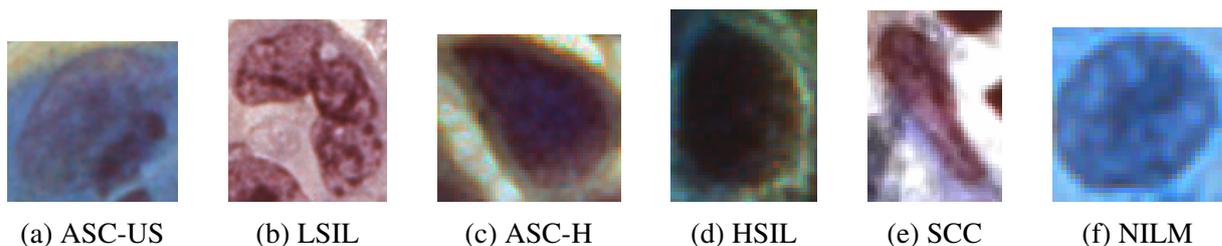


Figura 3.5 – Classificações da base de dados *CRIC Cervix Cervix*

Classes	Categoria	Tipo de célula	Total de núcleos segmentados
1	Anormal	ASC-US	286
2	Anormal	LSIL	598
3	Anormal	ASC-H	536
4	Anormal	HSIL	874
5	Anormal	SCC	77
6	Normal	NILM	862

Tabela 3.2 – Distribuição de classes *CRIC Cervix*

A base é composta por imagens contendo um número variável de células segmentadas. As imagens possuem verdade terrestre para segmentação - regiões do núcleo e citoplasma de algumas das células presentes na imagem, como demonstrado na Figura 3.6.

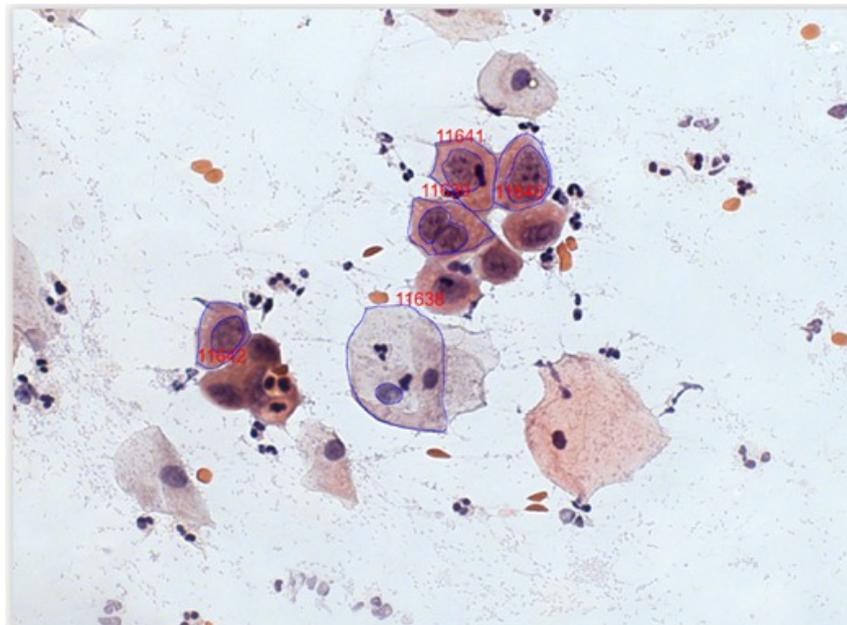


Figura 3.6 – Imagem com núcleos segmentados

A consistência das informações passadas a rede tem um papel crucial no desempenho do modelo. A base de dados *CRIC Cervix* é composta por imagens total e parcialmente segmentadas por citopatologistas. Para aproveitar as imagens parcialmente segmentadas, foi necessário a criação de uma base de dados específica para segmentação. Essa base de dados foi obtida por meio de uma ferramenta que submetia as imagens e recortes para a aprovação de um citopatologista por meio de inspeção visual conforme a Figura 3.7. A inspeção segue os seguintes passos: divisão das imagens em marcações, seleção visual e manual dos recortes por um especialista em citologia, eram mantidos apenas os recortes que possuíam informações de todos os núcleos segmentados, armazenagem dos recortes e respectivas segmentações, e descarte dos recortes sem informação de núcleos segmentados.

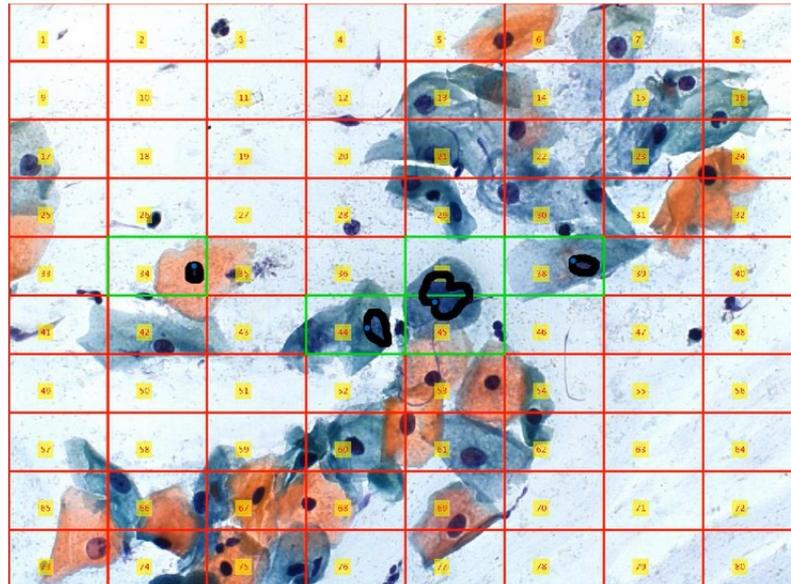


Figura 3.7 – Exemplo de imagem particionada da base de dados CRIC - os destaques em verde apresentam recortes que serão aproveitados para a base de segmentação.

Como as imagens da base de segmentação apresentam as dimensões 102 x 102, para adequá-las a entrada da rede foram redimensionados para 128 x 128 para que as proporções do núcleo real fossem mantidas. Por se tratar de um recorte quadrado, a operação não ocasionará distorção aos núcleos, preservando características morfológicas essenciais ao diagnóstico. Em seguida, as imagens foram divididas para treinamento (treino e validação) e teste. Por fim a base de treino sofreu aumento de dados. A Figura 3.8 apresenta a metodologia de preparação da base de dados para os experimentos.

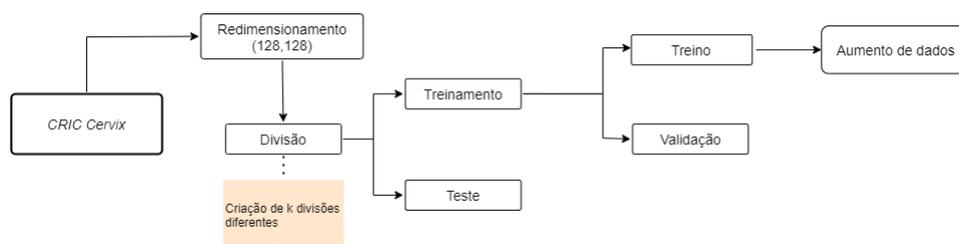


Figura 3.8 – Fluxograma da metodologia para a base de dados *CRIC Cervix*

## 3.2 Aumento de dados

Para apresentar uma boa performance as redes precisam de uma grande quantidade de imagens para treino, quando não existem imagens suficientes realiza-se o aumento de dados (SHORTEN; KHOSHGOFTAAR, 2019). O método de aumento de dados consiste em um conjunto de técnicas para gerar outras imagens a partir de uma base de dados pequena. As transformações incluem rotação, inversão na horizontal e vertical, entre outros - como exemplo demonstrado na Figura 3.9. Para realizar o aumento nas três bases de dados é necessário levar em

consideração o impacto em características morfológicas do núcleo celular, na qual as alterações podem impactar no diagnóstico. Nesse trabalho foram realizadas operações de rotação, inversão e modificação no brilho e contraste da imagem de acordo com um conjunto de parâmetros de modificação, de tal forma que para cada imagem eram geradas de 20 à 100 imagens. As operações e suas respectivas variações estão descritas na Tabela 3.3.

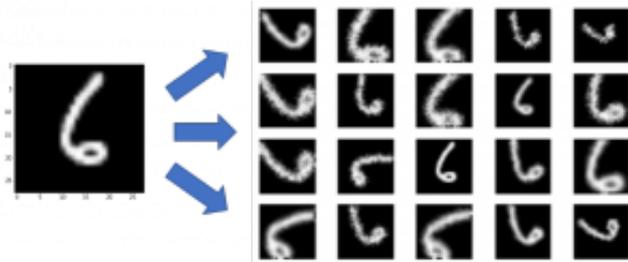


Figura 3.9 – Exemplo de aplicação da técnica de aumento de dados (RAJ, 2018)

Operação	Valor atribuído
Faixa de rotação	180°
Inversão Horizontal	Verdadeiro
Inversão Vertical	Verdadeiro
Faixa de brilho	[0,2 - 0,9]
Faixa de contraste	[0,2 - 0,9]

Tabela 3.3 – Operações aplicadas para aumento de dados

### 3.3 Arquiteturas

Como mencionado na Seção 2.1, as CNNs são algoritmos que vem ganhando espaço no campo da Visão Computacional a partir do momento que a arquitetura AlexNet(KRIZHEVSKY; SUTSKEVER; HINTON, 2012) foi ganhadora do Desafio de Reconhecimento em Larga Escala da ImageNet(RUSSAKOVSKY et al., 2015). Desde então, um número maior de trabalhos envolvendo aprendizado profundo tem surgido na literatura. Isso não se distingue para a segmentação e classificação de imagens de citologia cervical, reforçado pelo gráfico da Figura 3.10.

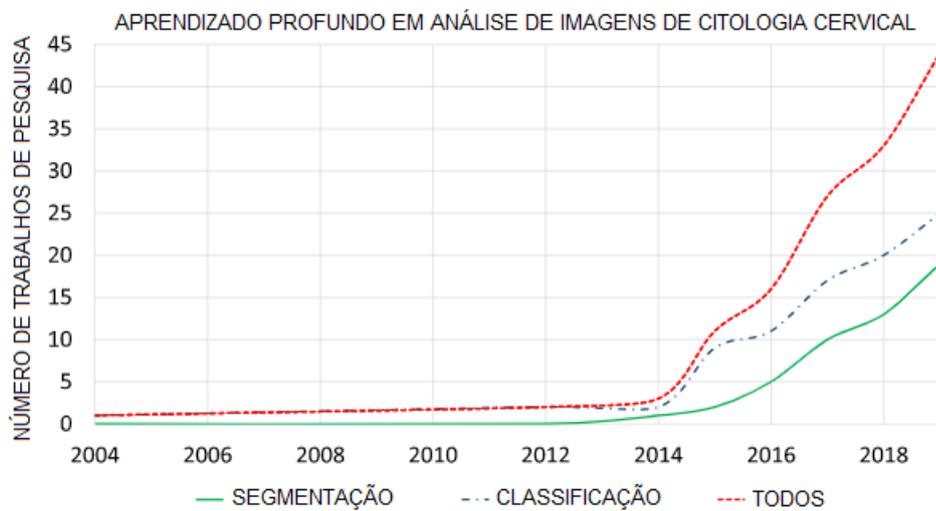


Figura 3.10 – Número de trabalhos por ano envolvendo Aprendizado Profundo para segmentação e classificação de imagens de citologia cervical (RAHAMAN et al., 2020)

Ainda que não existam trabalhos na literatura que envolvam a SegNet para segmentação de núcleos de citologia cervical, a arquitetura se destacou na retenção de detalhes através do conjunto de dados CamVid (BROSTOW; FAUQUEUR; CIPOLLA, 2009) de imagens de ruas e a redução no número de parâmetros por meio da exclusão de camadas totalmente conectadas da VGG-16. Ressalta-se também, a nova abordagem para reconstrução dos mapas de características, através dos índices do subamostragem.

A U-net, proposta para a segmentação de imagens biomédicas e apresentada no Desafio de segmentação ISBI2015, se destacou pela abordagem para reaproveitar os mapas de características do *encoder* no *decoder* da rede. Com essa proposta, a arquitetura ocupou o primeiro lugar no desafio. Nesse sentido, essa seção apresentará as duas arquiteturas usadas nesse trabalho para a segmentação de núcleos em células cervicais, a U-net 3.3.1 e a SegNet 3.3.2.

### 3.3.1 U-net

A U-net (RONNEBERGER; FISCHER; BROX, 2015) é uma arquitetura totalmente convolucional voltada a segmentação semântica. Dividida em um caminho de contração, o *encoder* da rede e um caminho expansivo, o *decoder* da rede. No *encoder*, as características são extraídas e a dimensão espacial toma proporções menores. Na U-net cada bloco de convolução consiste em 3 camadas de convolução cada um seguida da função de ativação ReLU, e ao final do bloco a operação de *MaxPooling*. Para restaurar a resolução imagem e capturar de informações localização, o *decoder* consiste em blocos de convolução seguidos de convoluções transpostas. A arquitetura está presente na Figura 3.11.

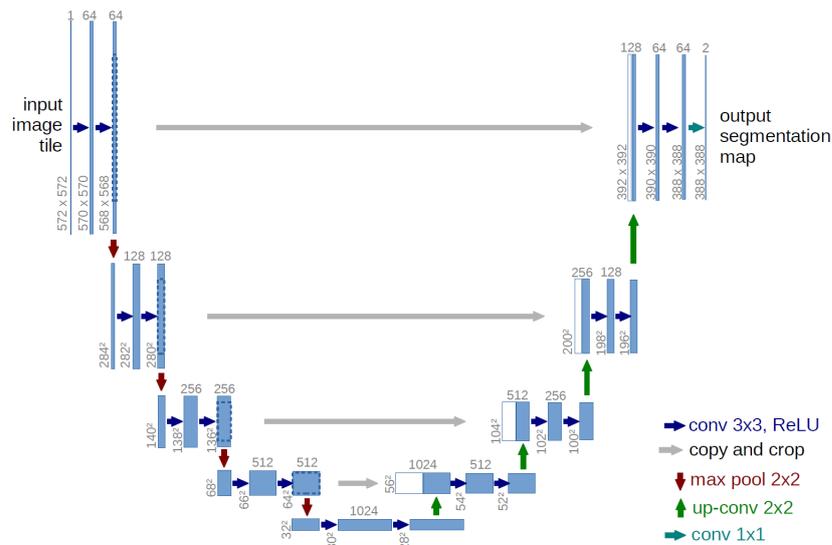


Figura 3.11 – Arquitetura U-net (RONNEBERGER; FISCHER; BROX, 2015)

A abordagem *copy-and-crop* proposta pela U-Net consiste em recortar os mapas de características para a dimensão de entrada das camadas que reconstróem a proporção original da imagem. Desse modo, a U-net reutiliza as saídas das camadas do *encoder* como entrada para as camadas do *decoder*, haja vista que assim a arquitetura permite reaproveitar as características extraídas anteriormente. Ademais, o modelo pode ser treinado de ponta-a-ponta.

Embora a técnica *copy and crop* da U-net demande menos processamento e memória, a proporção da máscara gerada pela rede é menor do que as imagens originais. Desse modo, após a previsão é necessária a aplicação de redimensionamento. Para dispensar essa operação, a técnica foi substituída por conexões residuais, visto que o *decoder* ainda pode ser alimentado com as imagens do *encoder* sem exigir o recorte dos mapas de características.

### 3.3.2 SegNet

Em distinção à U-net, a arquitetura SegNet (KENDALL; BADRINARAYANAN; CIPOLLA, 2015) trouxe a novidade da reconstrução dos mapas de características através dos índices de subamostragem armazenados. Dessa forma, são mantidas as posições dos píxeis mais significativos dos mapas de características para que a sobreamostragem possa reconstruir a resolução original da imagem. A técnica é demonstrada na Figura 3.12.

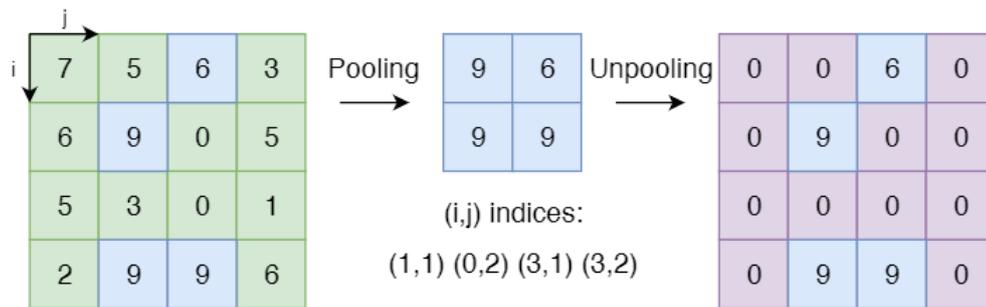


Figura 3.12 – Demonstração da técnica de reconstrução da resolução inicial com os índices do *MaxPooling* (PAILLASSA; BERTIN; BOUY, 2019)

O *encoder* da rede consiste em uma arquitetura VGG16 pré-treinada pela base de dados da ImageNet. Através da remoção das camadas totalmente conectadas da arquitetura original, atingiu-se a diminuição de 134M para 14,7M em número de parâmetros. Além disso, cada bloco de convolução é seguido de normalização em lote, ReLU e *MaxPooling*, como mostra a Figura 3.13. Ainda, com adição de zeros as bordas dos mapas de características, a informações contidas nas bordas são mantidas na convolução.

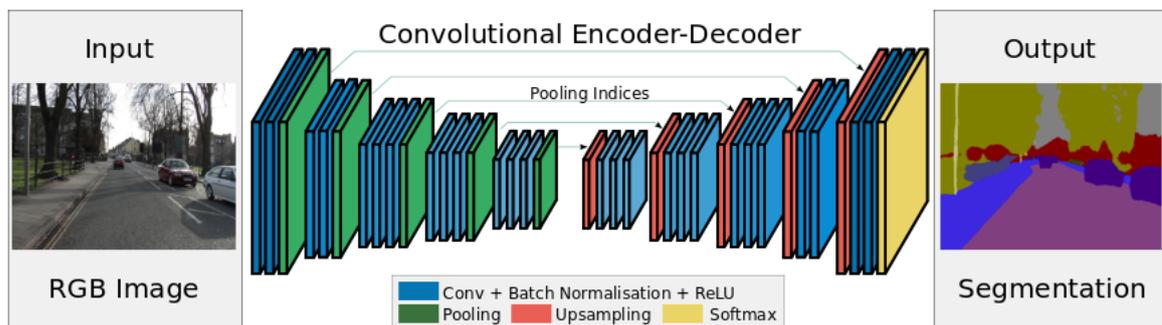


Figura 3.13 – Arquitetura SegNet (KENDALL; BADRINARAYANAN; CIPOLLA, 2015)

### 3.4 Treinamento

As técnicas e parâmetros empregados no treinamento de uma Rede neural convolucional são fundamentais para o aprendizado da arquitetura. Dessa modo, nesta seção é descrita a metodologia aplicada no treinamento.

A primeira etapa para o treinamento de uma Rede Neural Convolucional é a divisão de sua base. Essa etapa é impactante na generalização do modelo tendo em vista o nível de dificuldade em cada amostra na divisão dos conjuntos. Para a segmentação, é comum dividir os dados aleatoriamente.

As bases *CRIC Cervix* e *Herlev* foram divididas em treino e teste. Em seguida, a base de treino é separada em treinamento e validação. Já a *ISBI2014* é disponibilizada com a divisão.

Por fim, o aumento de dados é feito apenas na base de treino, para que o modelo seja avaliado somente em amostras reais.

Para garantir a generalização do modelo, cada base foi separada em  $k$  subconjuntos distintos de treino, teste e validação. Em seguida, essas partições foram treinadas e avaliadas separadamente. Por fim, o resultado final é uma média dos resultados de cada bases para cada respectivo experimento. As Figuras 3.2 e 3.8 apresentam um diagrama dessa operação.

Para contornar a inicialização aleatória dos pesos, a arquitetura Segnet apresenta a possibilidade de reaproveitamento dos pesos de arquiteturas pré-treinadas. Nesse sentido, foi aplicada a técnica de Transferência de aprendizado apresentada na Seção 2.2.4, devido ao seu *encoder* ser composto por uma VGG16 (SIMONYAN; ZISSERMAN, 2014).

O treinamento das redes foi dividido em ciclos de treinamento. Após o modelo completar uma época de treinamento, é verificado se houve ou não melhora em relação ao resultado da época anterior. A melhora é medida através da função de perda (descrita na Seção 2.2.3.3) aplicada as máscaras do modelo geradas pela base de validação. Se a perda da validação apresentar diminuição, os pesos da rede são salvos. Nesse aspecto, isso garante que os valores dos parâmetros treináveis sejam os melhores encontrados durante o treinamento. Ainda, a aplicação da técnica de parada antecipada determina que aquele ciclo de treinamento será interrompido após um número  $\pi_{pa}$  de épocas sem melhora.

Se o número de épocas sem melhora e o limite superior de épocas ainda não foi atingido, o modelo prosseguirá com o ciclo de treinamento. Assim que o ciclo for interrompido é feita a verificação se foi atingido o limite inferior  $\pi_{lr}$  da taxa de aprendizado (*Learning rate*). Se atingido, o treinamento é finalizado. Em contrapartida, a taxa de aprendizado é diminuída. Muitas vezes essa operação melhora a precisão da predição da rede, visto que a taxa de aprendizado é responsável pelo tamanho da etapa na descida do gradiente. Ainda, segundo You et al. (2019) começar com uma taxa de aprendizado grande impede que a rede memorize dados ruidosos, enquanto a diminuição gradual melhora a aprendizagem de padrões complexos.

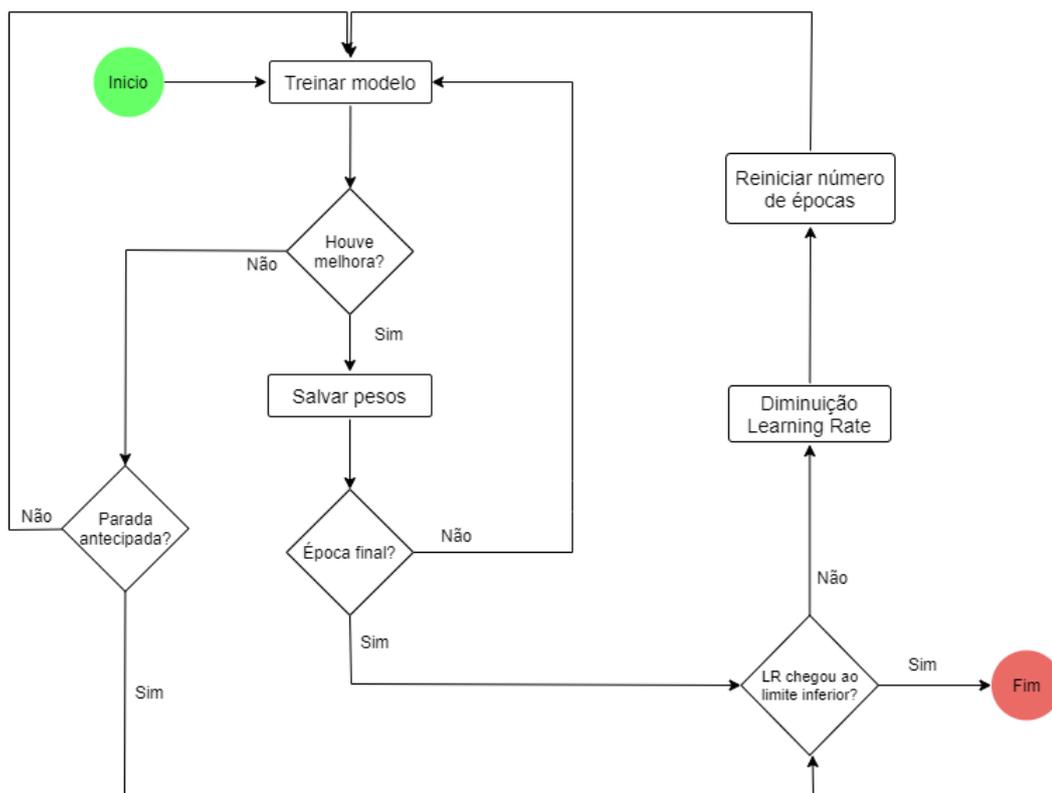


Figura 3.14 – Etapas do treinamento dos modelos

## 4 Resultados e Discussões

Neste capítulo são apresentados os resultados da metodologia de segmentação escolhida para as bases de dados **Herlev**, **CRIC Cervix** e **ISBI2014**, descritas na seção 3.1.1, 3.1.2 e 3.1.3, respectivamente. As bases de dados Herlev e *CRIC Cervix* foram divididas em 80% para Treinamento e 20% para Teste. Em seguida, 20% da base de Treinamento foi destinada para a validação. Os experimentos foram executados através da máquina virtual do *Google Colaboratory*<sup>1</sup>. As especificações de *Hardware* estão presentes na Tabela 4.1. A construção das CNNs foi feita através da utilização a biblioteca em *python Tensorflow/Keras*<sup>2</sup>.

Parâmetros	Especificação
GPU	T4 e P100
Memória da GPU	25GB
Número de núcleos da GPU	4
Memória RAM	26.75 GB
Memória em Disco	358 GB

Tabela 4.1 – Especificação do *Hardware* da máquina virtual do *Google Colaboratory* utilizada para os experimentos

Para a execução dos experimentos foram geradas  $k$  bases de dados, com  $k$  apresentando o valor 5 a partir de cada uma das bases. Os conjuntos de dados Herlev e *CRIC Cervix* foram divididos aleatoriamente 5 vezes de acordo com as proporções mencionadas anteriormente. Para a ISBI2014, como os dados já são disponibilizados divididos, para gerar as 5 conjuntos distintos foram aplicados aumentos de dados com sementes diferentes. Ainda, cada conjunto de treinamento foi incrementado com o auxílio das técnicas de aumento de dados mencionadas na Seção 3.2. Essas divisões implicam em 5 partições distintas, que serão treinadas e avaliadas separadamente. Por fim, o resultado apresentado é uma média dos resultados de cada base, para cada experimento.

Os parâmetros variados foram **tamanho de lote**, **função de ativação** e **normalização em lote**. O tamanho de lote foi variado de acordo com as limitações do hardware e variaram de 16, 32 e 64. A aplicação da normalização em lote visa avaliar a capacidade de generalização da técnica, pode ser usada ou não. Por fim, as funções de ativação tem o intuito de evitar a morte dos neurônios ocasionada pela aplicação da função de ReLU, empregada nas arquiteturas originais. A Tabela 4.2 apresenta a variação de parâmetros para o treinamento.

<sup>1</sup> colab.research.google.com

<sup>2</sup> www.tensorflow.org

Parâmetros	Valores atribuídos
Tamanho do lote	[16,32,64]
Normalização em lote	[Verdadeiro, Falso]
Função de ativação	[Relu, Mish, Swish, Elu, LeakyRelu]

Tabela 4.2 – Parâmetros variados para o treinamento

Todas os experimentos foram executados em 4 vezes de 300 épocas (ou ciclos de treinamento, conforme a Seção 3.4). A taxa de aprendizado inicial foi obtida através da observação do treinamento das arquiteturas em um intervalo de  $[10^{-7}, 10^{-2}]$  em 300 épocas. Ainda, a taxa de aprendizado inicial foi definida como  $10^{-3}$  e dividida por 10 quando necessário, até atingir o limite inferior de  $\pi_{lr} = 10^{-6}$ . Por fim, a Parada Antecipada ocorria no ciclo de treinamento quando o modelo atingia  $\pi_{pa}$  épocas sem melhora, com  $\pi_{pa} = 25$ .

As métricas precisão, revocação e Dice são avaliadas a partir de um limiar. Para coeficiente de Dice, o limiar 0,7 é o mais empregado na literatura, na qual a segmentação é considerada adequada quando mais de 70% do núcleo é segmentado corretamente. Para precisão e revocação o valor mais empregado é 0,6 (SAHA; BAJGER; LEE, 2017). Por fim, as predições das arquiteturas são feitas através da função de ativação *sigmoid* (Seção 2.8) em um intervalo de probabilidade de 0 à 1. Nos resultados de predições a seguir, as variações de cores nas predições indicam a possibilidade do pixel pertencer a região do núcleo, variando de amarelo a roxo (passando pelo azul), onde amarelo é o mais provável e o roxo o pixel menos provável.

Ainda, esse capítulo foi dividido em 4 seções, para análise dos resultados de cada base de dados separadamente. As Seções 4.1, 4.2, 4.3 e 4.4 apresentam os resultados dos experimentos nas bases Herlev, ISBI2014 e CRIC Cervix, respectivamente, além de discussões gerais. .

## 4.1 Herlev

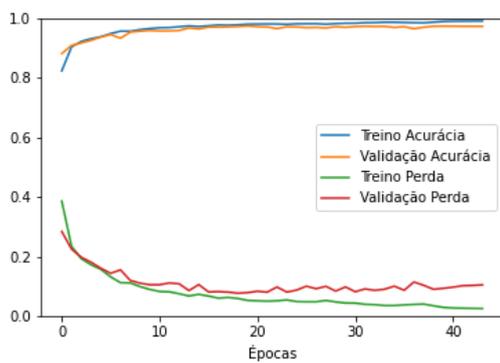
Essa seção apresenta os resultados da metodologia escolhida aplicada a base de dados Herlev. Como mencionado na Seção 3.1.1, a base de dados Herlev é composta por imagens de células únicas. Quando comparada as outras bases de dados desse trabalho, a Herlev representa um desafio menor a segmentação na qual o núcleo está centralizado nas amostras. Foi observado no treinamento, que como resultado a rede direciona o aprendizado mais em definição das bordas do que em informações de localização.

Os resultados para cada umas das metodologias propostas estão dispostos nas Tabelas A.1, A.2, A.3 e A.3 apresentadas no Apêndice A. Com 0,9744 para revocação, 0,9695 para coeficiente Dice com a método SegNet com Normalização em lote, e 0,9708 para acurácia com U-net com Normalização em lote os resultados superaram trabalhos anteriores, como demonstrado na Tabela 4.3. Ainda, obtivemos 0,9667 para precisão com a SegNet com Normalização em lote. Exemplos de predições das redes estão presentes nas Figuras 4.3, 4.4, 4.5 e 4.6 para cada uma das metodologias abordadas.

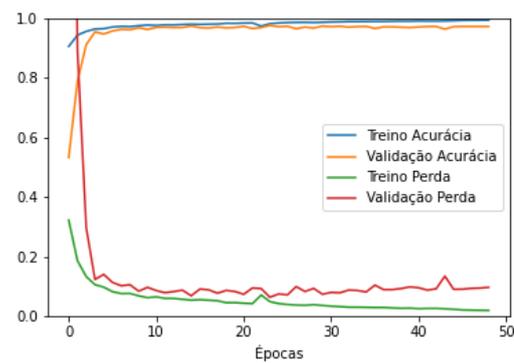
Trabalhos	Método	Precisão	Revocação	Dice	Acurácia
Sharma e Mangat (2016)	FCM e BPNN	0,8600	0,9000	-	-
Gautam et al. (2018)	CNN com pré-processamento seletivo	0,8900	0,9100	0,9000	-
Liu et al. (2018)	Mask-RCNN LFC-CRF	0,9600	0,9600	-	-
Zhao et al. (2019)	PGU-net+	0,9010	0,9680	0,9260	-
Sabeena, Gopakumar e Thampi (2020)	FC DenseNet56	<b>0,9996</b>	0,8767	0,9342	0,9500
Yang, Wu e Yin (2020)	<i>Interacting convolution with pyramid structure network</i>	0,9720	0,9730	-	-
Nossa abordagem	U-net com Normalização em lote	0,9634	0,9678	0,9671	<b>0,9708</b>
Nossa abordagem	U-net sem Normalização em lote	0,9665	0,9636	0,9648	0,9683
Nossa abordagem	SegNet com Normalização em lote	0,9667	<b>0,9744</b>	<b>0,9695</b>	0,9664
Nossa abordagem	SegNet sem Normalização em lote	0,9102	0,8866	0,8981	0,9008

Tabela 4.3 – Comparação de resultados da metodologia abordada nesse trabalho e o restante da literatura para a base de dados Herlev

A arquitetura U-net modificada de acordo com a Seção 3.3.1 obteve resultados semelhantes com e sem Normalização em lote, como mostra a Figura 4.1. Já a SegNet apresentou instabilidade e dificuldade no aprendizado, apresentado na Figura 4.2. Em vista disso, é possível notar que aproximadamente a partir da época 18, a SegNet sem a Normalização em Lote não consegue prosseguir com o aprendizado.

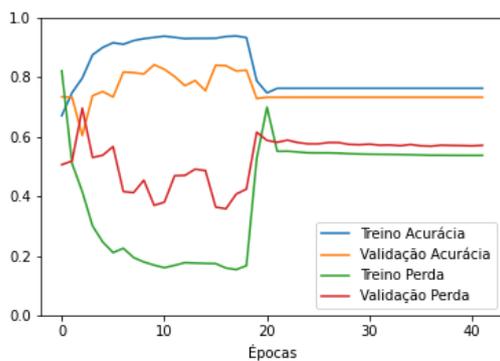


(a) U-net sem Normalização em Lote

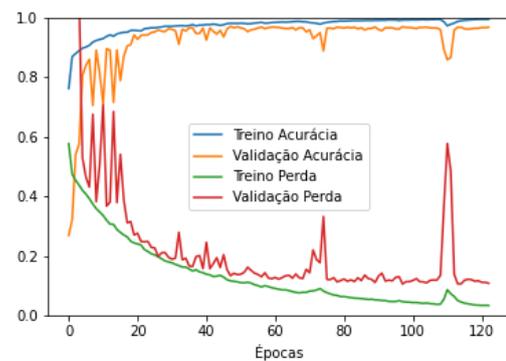


(b) U-net com Normalização em Lote

Figura 4.1 – Comparação de performance da U-net com e sem Normalização em Lote para a base de dados Herlev e função de ativação Swish



(a) SegNet sem Normalização em Lote



(b) SegNet com Normalização em Lote

Figura 4.2 – Comparação de performance da SegNet com e sem Normalização em Lote para a base de dados Herlev e função de ativação ReLU



Figura 4.3 – Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados Herlev



Figura 4.4 – Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados Herlev



Figura 4.5 – Exemplo de predição da arquitetura SegNet com normalização em lote a partir da base de dados Herlev



Figura 4.6 – Exemplo de predição da arquitetura SegNet sem normalização em lote a partir da base de dados Herlev

As funções de ativação obtiveram variação de resultados inferiores a 0,1 para os modelos U-net com Normalização em lote, U-net sem Normalização em lote e SegNet com Normalização em lote. Com a U-net com Normalização em lote, variando de 0,9614 à 0,9671, o modelo U-net sem Normalização variando de 0,9020 à 0,9648 e o modelo SegNet com Normalização em lote com valores de 0,9613 a 0,9695 para coeficiente de Dice conforme as Tabelas A.1, A.2, e A.3 no Apêndice A. Por outro lado, o modelo SegNet sem Normalização em lote sofreu muita variação de acordo com as funções de ativação empregadas, a Figura 4.7 apresenta essa comparação. Nela,

é possível notar que os melhores resultados obtidos para essa abordagem foram com a função ReLU e os piores com a função Swish e ELU. Ademais, os tamanhos de lote geraram pouca variação.

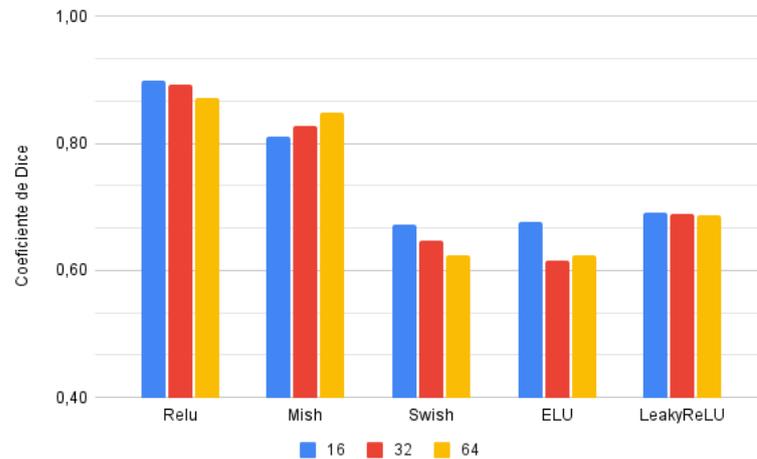


Figura 4.7 – Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados Herlev

## 4.2 ISBI2014

Os resultados para cada uma das abordagens deste trabalho estão dispostos nas Tabelas A.5, A.6, A.7 e A.7 presentes no Apêndice A. Com 0,9933 para precisão e 0,9783 para coeficiente Dice com o método SegNet sem Normalização em lote e 0,9714 para revocação com a U-net com Normalização em lote, os resultados superaram trabalhos anteriores, como demonstrado na Tabela 4.4. Ademais, alcançamos 0,9961 para acurácia com a SegNet com Normalização em lote.

Trabalhos	Método	Precisão	Revocação	Dice	Acurácia
Ushizima, Bianchi e Carneiro (2015)	Super-píxeis e diagrama de Voronoi	0,9590	0,8950	0,9140	-
Saha, Bajger e Lee (2017)	CiscFC	0,9680	0,8820	0,9380	-
Phoulady et al. (2017)	Método iterativo de binarização	0,9610	0,9330	0,9380	-
Huang et al. (2020)	<i>Multi-scale fuzzy clustering</i>	0,9810	0,9390	0,9360	-
Hoque et al. (2021)	Método baseado em propriedades de contorno	0,9780	0,9333	-	-
Nossa abordagem	U-net com Normalização em lote	0,9794	<b>0,9714</b>	0,9743	0,9957
Nossa abordagem	U-net sem Normalização em lote	0,9767	0,9603	0,9684	0,9948
Nossa abordagem	SegNet com Normalização em lote	0,9725	0,9706	0,9726	<b>0,9961</b>
Nossa abordagem	SegNet sem Normalização em lote	<b>0,9933</b>	0,9654	<b>0,9783</b>	0,9929

Tabela 4.4 – Comparação de resultados da metodologia abordada nesse trabalho e o restante da literatura para a base de dados ISBI2014

Em distinção aos resultados na base de dados Herlev, para o conjunto ISBI2014 a instabilidade ocorreu com o modelo U-net sem normalização em lote. Nas imagens 4.8 e 4.9 é possível observar a diferença entre a predição do modelo com e sem Normalização em lote. Na predição sem Normalização em lote, o modelo apresentou um número maior de falsos negativos,

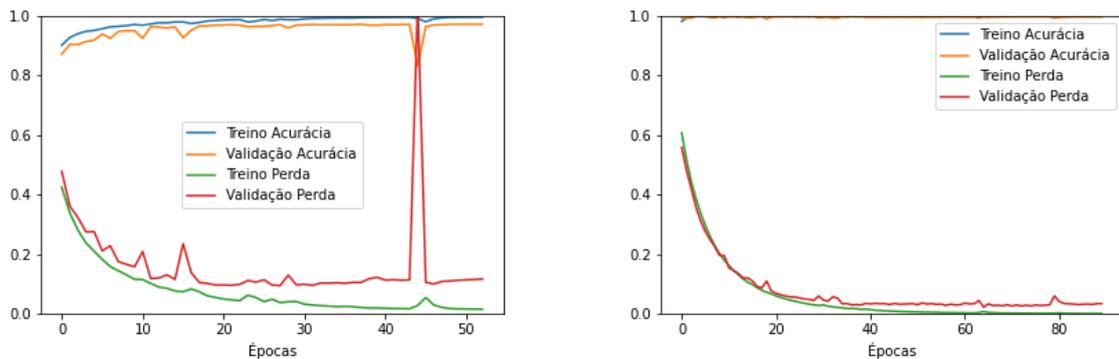
por classificar núcleos como fundo. Pode-se notar também a diferença nos gráficos da Figura 4.10 do treinamento.



Figura 4.8 – Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados ISBI2014



Figura 4.9 – Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados ISBI2014



(a) U-net sem Normalização em lote

(b) U-net com Normalização em lote

Figura 4.10 – Performance da U-net sem e com Normalização em lote

A SegNet demonstrou dificuldade no aprendizado, provavelmente devido desproporcionalidade das amostras entre regiões de núcleo e fundo. Para contornar essa situação, a base de dados de treino foi sucedida à recortes de (128,128) como uma forma de aumento de dados. Nesse sentido, além das imagens completas e aumento de dados a base de treino também apresenta recortes de (128,128). A comparação do treinamento com e sem recortes está presente na Figura 4.11.

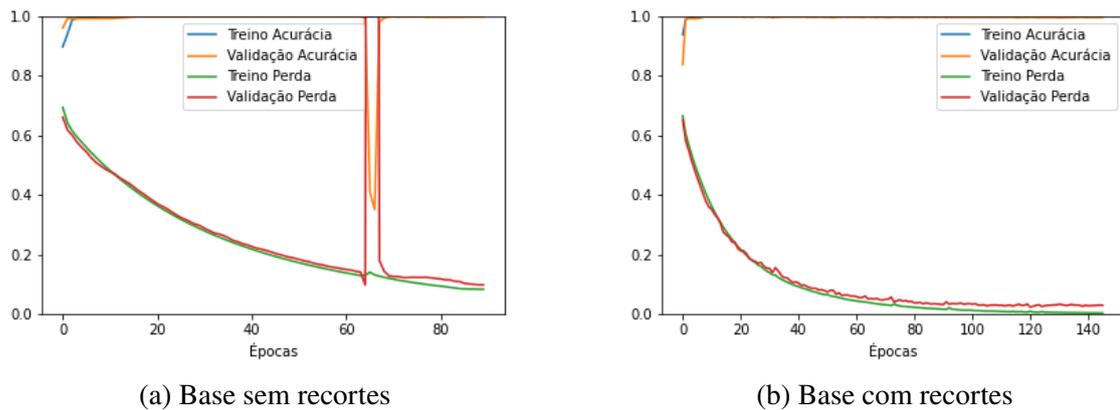


Figura 4.11 – Performance da SegNet sem e com recortes

Dentre a variação das funções de ativação e tamanhos de lote, a abordagem dos modelos sem Normalização em lote apresentaram intervalos maiores conforme as Figuras 4.12 e 4.13. A função de ativação ReLU obteve os melhores resultados para todos os tamanhos de lote. Os tamanhos de lotes empregados atingiram resultados com variação menores que 0,1 para todas as funções. Para os demais modelos, a variação ficou entre 0,9631 para a U-net com Normalização em lote e 0,8982 a 0,9726 com a SegNet com Normalização em lote para coeficiente de Dice.

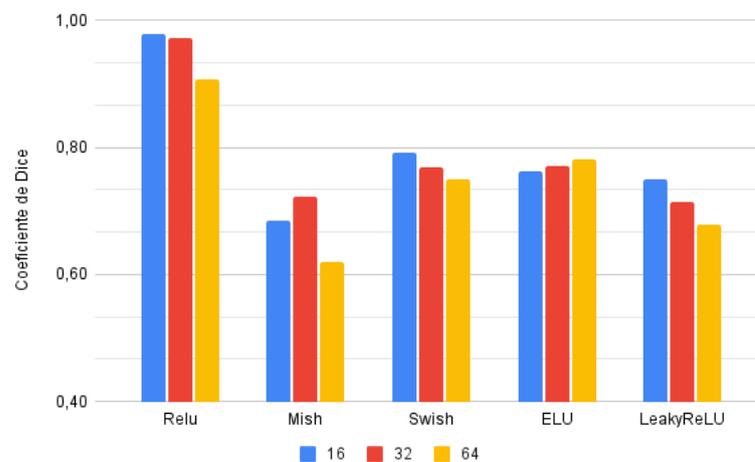


Figura 4.12 – Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados ISBI2014

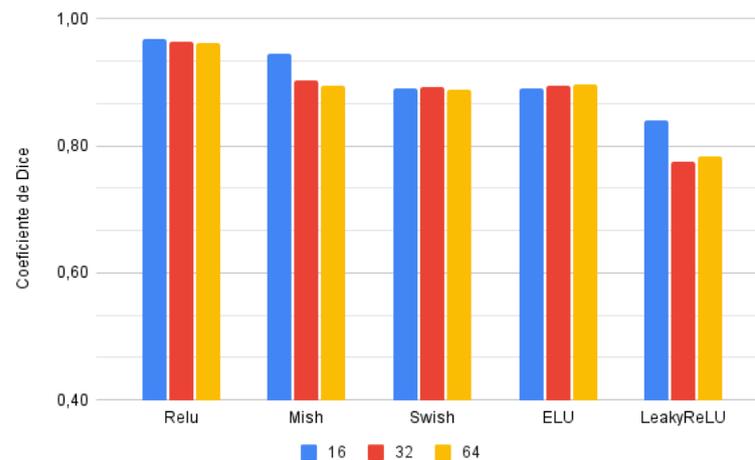


Figura 4.13 – Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura U-net sem Normalização em lote a partir da base de dados ISBI2014

### 4.3 CRIC Cervix

A base de dados de múltiplas células **CRIC Cervix** para segmentação é composta por partições de imagens reais, podendo conter uma ou mais células a cada partição. Além de aglomerados de células, as imagens apresentam diversos ruídos, neutrófilos, entre outros artefatos. Em vista disso, o principal desafio para segmentação de núcleos celulares é encontrar métodos que consigam distinguir núcleos de outras estruturas semelhantes, que possam compor as imagens. Essa seção, apresenta os resultados e discussões sobre a metodologia escolhida nesse trabalho para a base de dados *CRIC Cervix*.

Através de resultados de 0,9524 para precisão, 0,9336 para revocação, 0,9429 para coeficiente de Dice com a metodologia U-net sem Normalização em lote e 0,9582 para acurácia com a U-net com Normalização em lote, as metodologias se mostraram capazes de segmentar núcleos cervicais com qualidade. Por se tratar de uma base de dados ainda não publicada, ainda não possuem outros trabalhos presentes na literatura para segmentação. A Tabela 4.5 apresenta os melhores resultados de acordo com as métricas empregadas. Por fim, exemplos de predições das redes de acordo com os melhores resultados estão presentes nas Figuras 4.18, 4.19, 4.20 e 4.21 para cada uma das metodologias abordadas.

Método	Precisão	Revocação	Dice	Acurácia
U-net com Normalização em lote	0,9418	0,9307	0,9352	<b>0,9582</b>
U-net sem Normalização em lote	<b>0,9524</b>	<b>0,9336</b>	<b>0,9429</b>	0,9560
SegNet com Normalização em lote	0,9027	0,9166	0,9096	0,9552
SegNet sem Normalização em lote	0,8026	0,6544	0,7130	0,9042

Tabela 4.5 – Resultados da metodologia abordada nesse trabalho para a base de dados *CRIC Cervix*

As Tabelas A.11 e A.12 presentes no Apêndice A demonstram os resultados de precisão, revocação, Dice e acurácia para a SegNet, com e sem normalização em lote, respectivamente; foram usados diferentes tamanhos de lotes e funções de ativação. A arquitetura U-net demonstrou resultados similares com e sem Normalização em Lote, conforme a Figura 4.14 e as Tabelas A.9 e A.10. Em contraste, sem as camadas de Normalização em Lote a SegNet demonstrou mais sobreajuste e instabilidade no treinamento conforme a Figura 4.15. Os melhores resultados para cada métrica estão destacados em negrito nas tabelas.

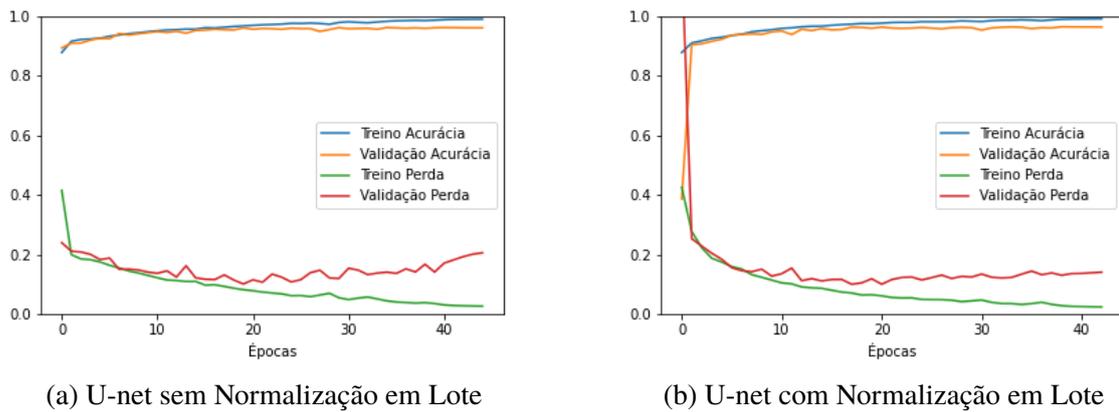


Figura 4.14 – Comparação de performance da U-net com e sem Normalização em Lote para a base de dados *CRIC Cervix*

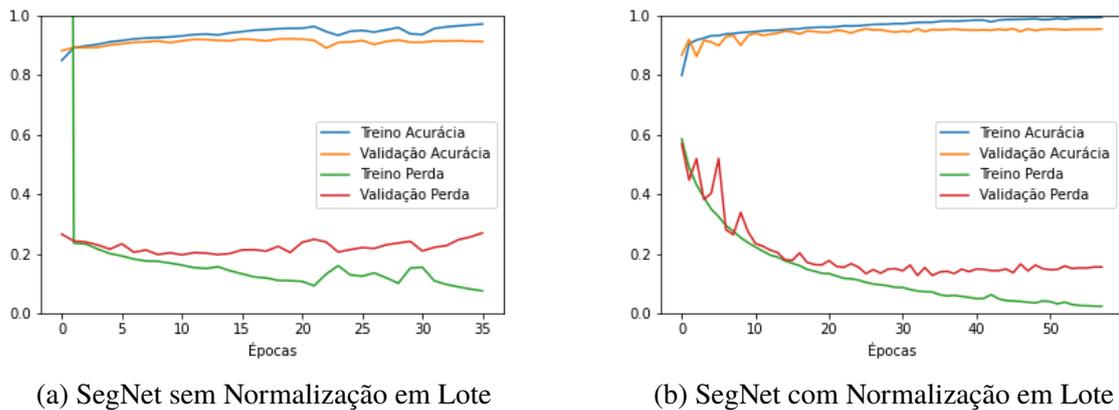


Figura 4.15 – Comparação de performance da SegNet com e sem Normalização em Lote para a base de dados *CRIC Cervix*

Entre a variação de parâmetros, os modelos obtiveram resultados distintos conforme as Tabelas A.9, A.10, A.11 e A.12 presentes no Apêndice A. Com exceção da U-net com Normalização em lote, a função LeakyReLU obteve o pior desempenho. Por outro lado, as funções ReLU e Swish obtiveram os melhores resultados. A abordagem com maior variação foi a SegNet sem Normalização em lote, conforme o gráfico da Figura 4.16. Nela, a maioria

dos resultados ficaram abaixo de 0,7, o que caracteriza uma segmentação distante da verdade terrestre.

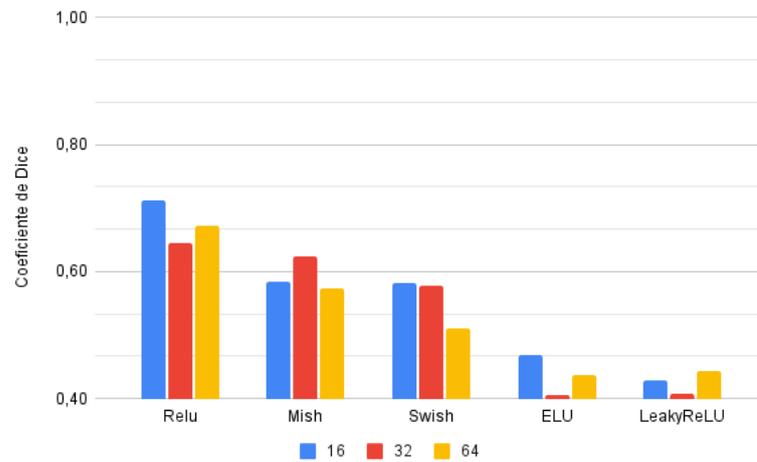


Figura 4.16 – Comparação entre tamanhos de lote e funções de ativação para os resultados da arquitetura SegNet sem Normalização em lote a partir da base de dados CRIC Cervix

Para tentar contornar o sobreajuste foram feitos experimentos com a incorporação de *Dropout* nas arquiteturas. A Figura 4.17 apresenta a comparação da U-net com e sem *Dropout*. Embora a técnica tenha diminuído o sobreajuste, após sua aplicação as arquiteturas apresentaram dificuldade no aprendizado. Portanto, a alternativa encontrada foi salvar os pesos dos modelos antes que o sobreajuste acontecesse.

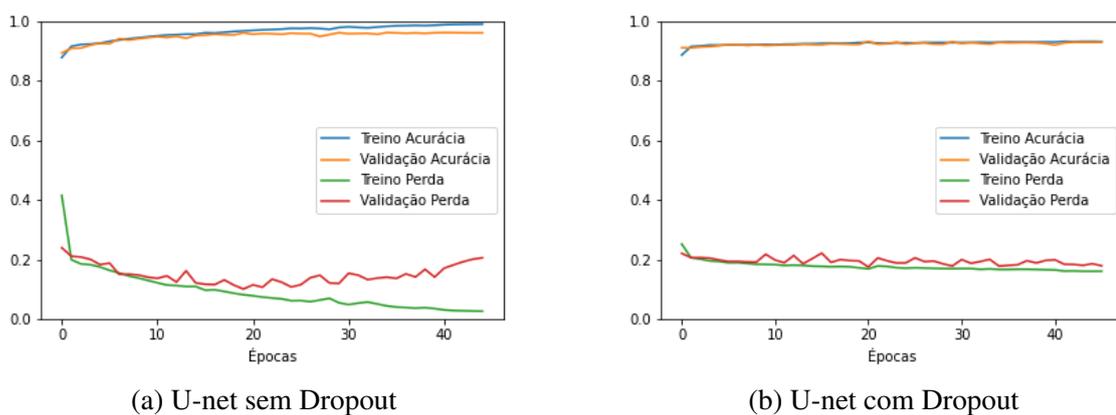


Figura 4.17 – Comparação de performance da U-net com e sem Dropout para a base de dados CRIC Cervix e função de ativação Mish



Figura 4.18 – Exemplo de predição da arquitetura U-net com normalização em lote a partir da base de dados *CRIC Cervix*



Figura 4.19 – Exemplo de predição da arquitetura U-net sem normalização em lote a partir da base de dados *CRIC Cervix*



Figura 4.20 – Exemplo de predição da arquitetura SegNet com normalização em lote a partir da base de dados *CRIC Cervix*



Figura 4.21 – Exemplo de predição da arquitetura SegNet sem normalização em lote a partir da base de dados *CRIC Cervix*

## 4.4 Discussões

Em muitos trabalhos da literatura, os autores obtiveram resultados com alta precisão e revocação menor, isso se dá pelo número elevado de falsos negativos. Nas predições dos métodos, o desequilíbrio em quantidade de pixels de fundo e núcleo torna essa assimetria ainda mais evidente. Nesse aspecto, a principal métrica a ser analisada é o coeficiente de Dice, ele demonstra o equilíbrio entre precisão e revocação. Em síntese, o coeficiente de Dice foi utilizado para selecionar os melhores resultados.

A seguir apresentamos resultados de predição para uma imagem real a partir dos três melhores modelos gerados pelas diferentes base de dados. Foram selecionados os pesos dos resultados com maior valor para coeficiente de Dice, conforme as Tabelas 4.3, 4.4, 4.5 e os parâmetros destacados nas Tabelas A.3, A.8, e A.10 dispostas no Apêndice A. Foi utilizada uma imagem real totalmente segmentada que possuísse estruturas com núcleo/citoplasma, neutrófilos e diferentes colorações.

Como comentado, as imagens da base de dados ISBI2014 não remetem a situações reais, pois apresentam poucas sobreposições, e não são compostas de outros artefatos. A Figura 4.22 apresenta o resultado de uma predição do modelo treinado pela base de dados do ISBI2014 para uma imagem real com artefatos. É possível notar que a rede não consegue discernir entre neutrófilos e núcleos, visto que as imagens da base, sintéticas, não apresentam outras estruturas além da célula. Ainda, a predição não encontra os núcleos celulares que podem conter diversos tamanhos e formas. A falta de variabilidade entre as células da base também a distancia de situações reais. Em síntese, a predição apresenta um número alto de falsos positivos e negativos.

A base de dados Herlev, embora seja uma base de dados real é composta apenas de células únicas. A Figura 4.23 apresenta o resultado da predição do modelo treinado que gerou o melhor resultado para o conjunto de dados Herlev. A base de dados Herlev não possui aglomerados de células e artefatos, desse modo é possível notar que o modelo segmenta núcleos distantes e muitos neutrófilos. Embora melhores que os resultados obtidos em relação aos gerados através da base de dados ISBI2014, a predição ainda apresenta um número alto de falsos positivos.

Por fim, treinados através de recortes da base de dados reais de múltiplas células *CRIC Cervix* o modelo demonstrou boa diferenciação entre núcleos e outros artefatos. A imagem apresentou poucos ruídos, que podem ser removidos com operações elementares. Na predição, é possível observar que os núcleos estão próximos ou atrás de neutrófilos foram parcialmente ou não segmentados. Outro ponto importante, é a capacidade do modelo treinado com recortes na predição de recortes sem informação, o desempenho do modelo gerou resultados com alto número de verdadeiros negativos. Em suma, o modelo conseguiu segmentar grande parte dos núcleos com qualidade.

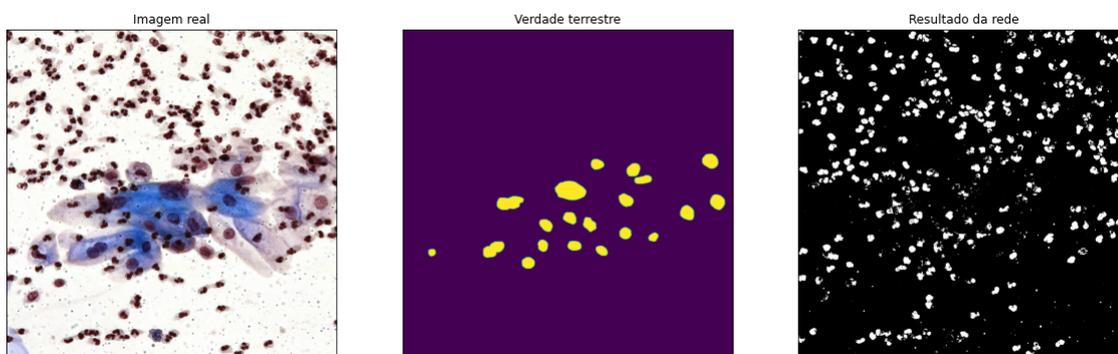


Figura 4.22 – Predição com os pesos do melhor resultado da base ISBI2014 aplicado uma imagem real

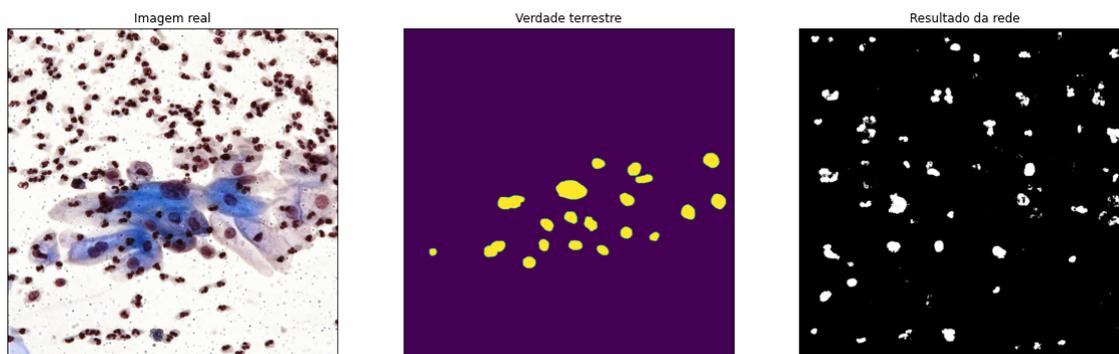


Figura 4.23 – Predição com os pesos do melhor resultado da base Herlev aplicado uma imagem real

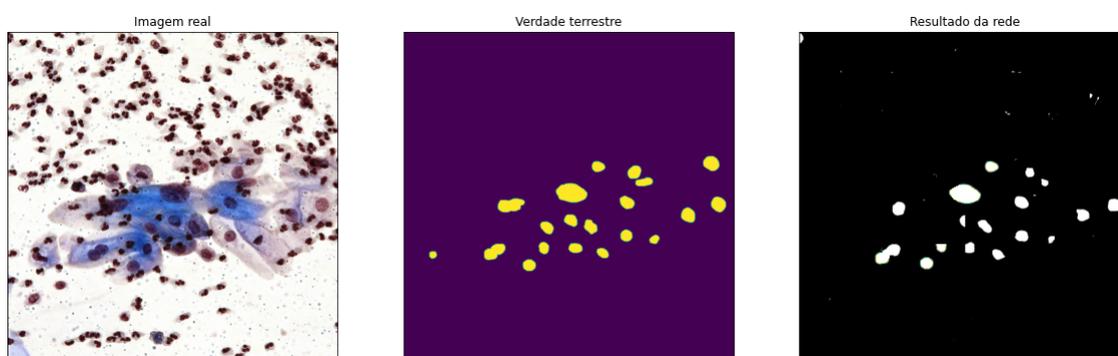


Figura 4.24 – Predição com os pesos do melhor resultado da base *CRIC Cervix* aplicado uma imagem real

## 5 Considerações finais

A análise de células cervicais é considerada uma tarefa de alta subjetividade. Sujeita a experiência e fadiga do profissional, o resultado pode ser muitas vezes, comprometido. Para tornar a análise menos suscetível à falhas, uma alternativa é a inserção da citologia digital no monitoramento de qualidade da rotina laboratorial. Ainda, o diagnóstico automatizado através da citologia digital pode ser dividido em duas etapas: a segmentação e a classificação do núcleo cervical.

A segmentação de núcleos cervicais é crucial para a classificação, visto que o núcleo apresenta informações essenciais ao diagnóstico. Nesse aspecto, o principal objetivo desse trabalho foi verificar o desempenho de arquiteturas de CNNs quando aplicadas a segmentação de núcleos reais de citologia cervical. Por meio das métricas acurácia, coeficiente de Dice, revocação e precisão, foram avaliados os modelo U-net e SegNet propostos por [Ronneberger, Fischer e Brox \(2015\)](#) e [Kendall, Badrinarayanan e Cipolla \(2015\)](#), respectivamente.

Para validar a metodologia de segmentação proposta neste trabalho, foram realizados experimentos considerando as bases de dados Herlev e ISBI2014 muito presentes na literatura, além da base de dados *CRIC Cervix*. Com 0,9744 para revocação, 0,9695 para coeficiente Dice os resultados superaram trabalhos anteriores, além disso, obtivemos 0,9708 para acurácia e 0,9667 para precisão na base de dados Herlev. Ainda, superamos o estado-da-arte para base de dados ISBI2014, com 0,9933 para precisão, 0,9714 para revocação, 0,9783 para coeficiente Dice e 0,9961 para acurácia. O uso das metodologias em base de dados conhecidas da literatura serviu para nos direcionar quanto a confiabilidade dos resultados obtidos.

Por fim, com resultados de 0,9524 para precisão, 0,9336 para revocação, 0,9429 para coeficiente de Dice e 0,9582 para acurácia na base de dados *CRIC Cervix*, a metodologia proposta se mostrou capaz de segmentar núcleos cervicais reais, mesmo em situações de sobreposição e diferentes artefatos.

## 6 Trabalhos futuros

Uma vez que definimos um conjunto de abordagens e modelo de segmentação para as imagens reais, pretendemos evoluir para uma ferramenta de segmentação e classificação dos núcleos cervicais em imagens reais, assim como submeter esses resultados a avaliação de especialistas da área de saúde e avaliar o impacto dessa tecnologia na rotina e apoio nos laboratórios.

# Referências

- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. *2017 International Conference on Engineering and Technology (ICET)*. [S.l.], 2017. p. 1–6.
- ALLEHAIBI, K. H. S.; NUGROHO, L. E.; LAZUARDI, L.; PRABUWONO, A. S.; MANTORO, T. et al. Segmentation and classification of cervical cells using deep learning. *IEEE Access*, IEEE, v. 7, p. 116925–116941, 2019.
- ALVES, G. *Entendendo Redes Convolucionais (CNNs)*. 2018. Disponível em: <<https://medium.com/neuronio-br/entendendo-redes-convolucionais-cnns-d10359f21184>>.
- AMORIM, J. G.; CERENTINI, A.; MACARINI, L. A. B.; MATIAS, A. V.; WANGENHEIM, A. von. Systematic literature review of computer vision-aided cytology. 2020.
- ARAÚJO, F. H.; SILVA, R. R.; USHIZIMA, D. M.; REZENDE, M. T.; CARNEIRO, C. M.; BIANCHI, A. G. C.; MEDEIROS, F. N. Deep learning for cell image segmentation and ranking. *Computerized Medical Imaging and Graphics*, Elsevier, v. 72, p. 13–21, 2019.
- BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: *Neural networks: Tricks of the trade*. [S.l.]: Springer, 2012. p. 437–478.
- BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, Wiley Online Library, v. 68, n. 6, p. 394–424, 2018.
- BRENNA, S. M. F.; HARDY, E.; ZEFERINO, L. C.; NAMURA, I. Conhecimento, atitude e pratica do exame de Papanicolaou em mulheres com de colo uterino. *Cadernos de Sao Paulo*, scielo, v. 17, p. 909 – 914, 08 2001. ISSN 0102-311X. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-311X2001000400024&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2001000400024&nrm=iso)>.
- BROSTOW, G. J.; FAUQUEUR, J.; CIPOLLA, R. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, Elsevier, v. 30, n. 2, p. 88–97, 2009.
- BROWNLEE, J. *Difference between a batch and an epoch in a neural network*. 2019. Disponível em: <<https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>>.
- CHAUBEY, A. *Downsampling and upsampling of images*. 2020. Disponível em: <<https://medium.com/analytics-vidhya/downsampling-and-upsampling-of-images-demystifying-the-theory-4ca7e21db24a>>.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 4, p. 834–848, 2017.

- CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 801–818.
- CLEVERT, D.-A.; UNTERTHINER, T.; HOCHREITER, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.
- COHEN, P. A.; JHINGRAN, A.; OAKNIN, A.; DENNY, L. Cervical cancer. *The Lancet*, Elsevier, v. 393, n. 10167, p. 169–182, 2019.
- DINIZ, D. N. *Segmentação de núcleos em células cervicais obtidas em exames de Papanicolaou*. Dissertação (Mestrado) — Universidade Federal de Ouro Preto, aug 2019.
- EVERINGHAM, M.; GOOL, L. V.; WILLIAMS, C. K. I.; WINN, J.; ZISSERMAN, A. *The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results*. [Http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html](http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html).
- FERLAY, J.; SOERJOMATARAM, I.; DIKSHIT, R.; ESER, S.; MATHERS, C.; REBELO, M.; PARKIN, D. M.; FORMAN, D.; BRAY, F. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer*, Wiley Online Library, v. 136, n. 5, p. E359–E386, 2015.
- FERNANDES, J. V.; RODRIGUES, S. H. L.; COSTA, Y. G. A. S. d.; SILVA, L. C. M. d.; BRITO, A. M. L. d.; AZEVEDO, J. W. V. d.; NASCIMENTO, E. D. d.; AZEVEDO, P. R. M. d.; FERNANDES, T. A. A. d. M. Conhecimentos, atitudes e prática do exame de papanicolaou por mulheres, nordeste do brasil. *Revista de Saúde Pública*, SciELO Public Health, v. 43, p. 851–858, 2009.
- FIGUEIREDO, A. C. C. et al. *Caracterização imunocitoquímica da expressão da proteína RAPI em blocos de células escamosas provenientes de citologia cervical em meio líquido*. Tese (Doutorado), 2015.
- FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: *Competition and cooperation in neural nets*. [S.l.]: Springer, 1982. p. 267–285.
- GAUTAM, S.; BHAVSAR, A.; SAO, A. K.; HARINARAYAN, K. Cnn based segmentation of nuclei in pap-smear images with selective pre-processing. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2018: Digital Pathology*. [S.l.], 2018. v. 10581, p. 105810X.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning (adaptive computation and machine learning series). *DOI*, v. 10, p. 1762–1766, 2016.
- GREENWOOD, S. d. A.; MACHADO, M. d. F. A. S.; SAMPAIO, N. M. V. Motivos que levam mulheres a não retornarem para receber o resultado de exame papanicolau. *Revista latino-americana de Enfermagem*, SciELO Brasil, v. 14, n. 4, p. 503–509, 2006.
- HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. B. Mask R-CNN. *CoRR*, abs/1703.06870, 2017. Disponível em: <<http://arxiv.org/abs/1703.06870>>.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

HOQUE, I. T.; IBTEHAZ, N.; CHAKRAVARTY, S.; RAHMAN, M. S.; RAHMAN, M. S. A contour property based approach to segment nuclei in cervical cytology images. *BMC Medical Imaging*, BioMed Central, v. 21, n. 1, p. 1–12, 2021.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. *CoRR*, abs/1709.01507, 2017. Disponível em: <<http://arxiv.org/abs/1709.01507>>.

HUANG, J.; WANG, T.; ZHENG, D.; HE, Y. Nucleus segmentation of cervical cytology images based on multi-scale fuzzy clustering algorithm. *Bioengineered*, Taylor & Francis, v. 11, n. 1, p. 484–501, 2020.

HUANG, Y.; ZHU, H. Segmentation of overlapped cervical cells using asymmetric mixture model and shape constraint level set method. *Mathematical Problems in Engineering*, Hindawi, v. 2020, 2020.

HUSSAIN, E.; MAHANTA, L. B.; DAS, C. R.; CHOUDHURY, M.; CHOWDHURY, M. A shape context fully convolutional neural network for segmentation and classification of cervical nuclei in pap smear images. *Artificial Intelligence in Medicine*, Elsevier, p. 101897, 2020.

IGLOVIKOV, V.; SHVETS, A. Terausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation. *CoRR*, abs/1801.05746, 2018. Disponível em: <<http://arxiv.org/abs/1801.05746>>.

INCA. Câncer no brasil: dados dos registros de base populacional. 2020.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

JANGIR, D. S. k. *Beginner's Guide for Convolutional Neural Network (CNN / ConvNets)*. Towards Data Science, 2019. Disponível em: <<https://towardsdatascience.com/beginners-guide-for-convolutional-neural-network-cnn-convnets-5a5e725ea581>>.

JORDAN, J. *Setting the learning rate of your neural network*. Jeremy Jordan, 2020. Disponível em: <<https://www.jeremyjordan.me/nn-learning-rate/>>.

KARPATY, A. *Convolutional Neural Networks (CNNs / ConvNets)*. 2015. Disponível em: <<https://cs231n.github.io/convolutional-networks/>>.

KENDALL, A.; BADRINARAYANAN, V.; CIPOLLA, R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*, 2015.

KRIZHEVSKY, A.; HINTON, G. et al. Learning multiple layers of features from tiny images. Citeseer, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LECUN, Y.; CORTES, C. MNIST handwritten digit database. 2010. Disponível em: <<http://yann.lecun.com/exdb/mnist/>>.

LECUN, Y. et al. Lenet-5, convolutional neural networks. URL: <http://yann.lecun.com/exdb/lenet>, v. 20, p. 5, 2015.

- LI, Z.; NAJARIAN, K. Automated classification of pap smear tests using neural networks. In: IEEE. *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*. [S.l.], 2001. v. 4, p. 2899–2901.
- LIANG, L. A.; EINZMANN, T.; FRANZEN, A.; SCHWARZER, K.; SCHAUBERGER, G.; SCHRIEFER, D.; RADDE, K.; ZEISSIG, S. R.; IKENBERG, H.; MEIJER, C. J. et al. Cervical cancer screening: Comparison of conventional pap smear test, liquid-based cytology, and human papillomavirus testing as stand-alone or cotesting strategies. *Cancer Epidemiology and Prevention Biomarkers*, AACR, v. 30, n. 3, p. 474–484, 2021.
- LIU, C.; YUEN, J.; TORRALBA, A. Sift flow: Dense correspondence across scenes and its applications. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 33, n. 5, p. 978–994, 2010.
- LIU, Y.; ZHANG, P.; SONG, Q.; LI, A.; ZHANG, P.; GUI, Z. Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. *IEEE Access*, IEEE, v. 6, p. 53709–53721, 2018.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440.
- LOY, J. *Neural Network Projects with Python*. [S.l.]: Packt Publishin, 2019.
- LU, J.; BEHBOOD, V.; HAO, P.; ZUO, H.; XUE, S.; ZHANG, G. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, Elsevier, v. 80, p. 14–23, 2015.
- LU, Z.; CARNEIRO, G.; BRADLEY, A. P. An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. *IEEE Transactions on Image Processing*, IEEE, v. 24, n. 4, p. 1261–1272, 2015.
- MABOTJA, M. C.; LEVIN, J.; KAWONGA, M. Beliefs and perceptions regarding cervical cancer and screening associated with pap smear uptake in johannesburg: A cross-sectional study. *PloS one*, Public Library of Science San Francisco, CA USA, v. 16, n. 2, p. e0246574, 2021.
- MARINAKIS, Y.; DOUNIAS, G.; JANTZEN, J. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Computers in Biology and Medicine*, Elsevier, v. 39, n. 1, p. 69–78, 2009.
- MARTINS, L. A.-s. F. L.; THULER, L. C. S.; VALENTE, J. G. A. Cobertura do exame de Papanicolaou no Brasil e seus fatores determinantes: uma revisao sistem pratica da literatura. *Revista Brasileira de Ginecologia e Obstetracia*, scielo, v. 27, p. 485 – 492, 08 2005. ISSN 0100-7203. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-72032005000800009&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-72032005000800009&nrm=iso)>.
- MASTERS, D.; LUSCHI, C. Revisiting small batch training for deep neural networks. *CoRR*, abs/1804.07612, 2018. Disponível em: <<http://arxiv.org/abs/1804.07612>>.
- MATCHA, A. C. N. *A 2021 guide to Semantic Segmentation*. AI & Machine Learning Blog, 2021. Disponível em: <<https://nanonets.com/blog/semantic-image-segmentation-2020/>>.
- MISHRA, D. *Transposed Convolution Demystified*. Towards Data Science, 2020. Disponível em: <<https://towardsdatascience.com/transposed-convolution-demystified-84ca81b4baba>>.

- MISRA, D. Mish: A self regularized non-monotonic neural activation function. *CoRR*, abs/1908.08681, 2019. Disponível em: <<http://arxiv.org/abs/1908.08681>>.
- MITTELDORF, C. A. T. Cervical cancer screening: from pap smear to future strategies. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, SciELO Brasil, v. 52, n. 4, p. 238–245, 2016.
- NAYAR, R.; WILBUR, D. C. *The Bethesda system for reporting cervical cytology: definitions, criteria, and explanatory notes*. [S.l.]: Springer, 2015.
- NAYAR, R.; WILBUR, D. C. The bethesda system for reporting cervical cytology: a historical perspective. *Acta cytologica*, Karger Publishers, v. 61, n. 4-5, p. 359–372, 2017.
- NERD, R. *Swish Activation Function by Google*. Medium, 2018. Disponível em: <<https://medium.com/@neuralnets/swish-activation-function-by-google-53e1ea86f820>>.
- NWANKPA, C.; IJOMAH, W.; GACHAGAN, A.; MARSHALL, S. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018.
- OKUNADE, K. S. Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology*, Taylor & Francis, v. 40, n. 5, p. 602–608, 2020.
- OQUAB, M.; BOTTOU, L.; LAPTEV, I.; SIVIC, J. Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 1717–1724.
- PAILLASSA, M.; BERTIN, E.; BOUY, H. Maximask and maxitrack: Two new tools for identifying contaminants in astronomical images using convolutional neural networks. *Astronomy & Astrophysics*, v. 634, 12 2019.
- PAVLOVSKY, V. *Introduction To Convolutional Neural Networks*. Vojtech Pavlovsky, 2019. Disponível em: <<https://www.vojtech.net/posts/intro-convolutional-neural-networks/>>.
- PHOULADY, H. A.; GOLDFOF, D.; HALL, L. O.; MOUTON, P. R. A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images. *Computerized Medical Imaging and Graphics*, Elsevier, v. 59, p. 38–49, 2017.
- PLISSITI, M. E.; DIMITRAKOPOULOS, P.; SFIKAS, G.; NIKOU, C.; KRIKONI, O.; CHARCHANTI, A. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In: *IEEE. 2018 25th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2018. p. 3144–3148.
- RAHAMAN, M. M.; LI, C.; WU, X.; YAO, Y.; HU, Z.; JIANG, T.; LI, X.; QI, S. A survey for cervical cytopathology image analysis using deep learning. *IEEE Access*, IEEE, v. 8, p. 61687–61710, 2020.
- RAJ, B. *Data Augmentation | How to use Deep Learning when you have Limited Data — Part 2*. 2018. Disponível em: <<https://medium.com/nanonets/how-to-use-deep-learning-when-you-have-limited-data-part-2-data-augmentation-c26971dc8ced>>.
- RAMACHANDRAN, P.; ZOPH, B.; LE, Q. V. Searching for activation functions. *CoRR*, abs/1710.05941, 2017. Disponível em: <<http://arxiv.org/abs/1710.05941>>.

REZENDE, M. T. *DESENVOLVIMENTO E VALIDAÇÃO DE FERRAMENTAS COMPUTACIONAIS APLICADAS NO RASTREIO DO CÂNCER DO COLO DO ÚTERO*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE OURO PRETO, 2018.

REZENDE, M. T.; BIANCHI, A. G. C.; CARNEIRO, C. M. Cervical cancer: Automation of pap test screening. *Diagnostic Citopathology*, p. to appear, 2021.

RODRIGUES, V. *Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças?* Medium, 2020. Disponível em: <<https://medium.com/@vitorborbarodrigues/métricas-de-avaliaç~ao-acurácia-precis~ao-recall-quais-as-diferenças-c8f05e0a513c>>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015.

Sabeena, K.; Gopakumar, C.; Thampi, R. An improved deep convolutional model for segmentation of nucleus and cytoplasm from pap stained cell images. In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. [S.l.: s.n.], 2020. p. 1274–1278.

SACHAN, P. L.; SINGH, M.; PATEL, M. L.; SACHAN, R. A study on cervical cancer screening using pap smear test and clinical correlation. *Asia-Pacific journal of oncology nursing*, Wolters Kluwer–Medknow Publications, v. 5, n. 3, p. 337, 2018.

SAHA, R.; BAJGER, M.; LEE, G. Circular shape constrained fuzzy clustering (ciscfc) for nucleus segmentation in pap smear images. *Computers in Biology and Medicine*, Elsevier, v. 85, p. 13–23, 2017.

SHARMA, B.; MANGAT, K. K. An improved nucleus segmentation for cervical cell images using fcm clustering and bpnn. In: IEEE. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.], 2016. p. 1924–1929.

SHARMA, S. *Activation Functions in Neural Networks*. Towards Data Science, 2019. Disponível em: <<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6#:~:text=The%20main%20reason%20why%20we,sigmoid%20is%20the%20right%20choice.>>>

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, Springer, v. 6, n. 1, p. 1–48, 2019.

SILBERMAN DEREK HOIEM, P. K. N.; FERGUS, R. Indoor segmentation and support inference from rgbd images. In: *ECCV*. [S.l.: s.n.], 2012.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

- STAFF, M. C. *Cervical cancer*. Mayo Foundation for Medical Education and Research, 2019. Disponível em: <<https://www.mayoclinic.org/diseases-conditions/cervical-cancer/symptoms-causes/syc-20352501>>.
- SULTANA, F.; SUFIAN, A.; DUTTA, P. Evolution of image segmentation using deep convolutional neural network: a survey. *Knowledge-Based Systems*, Elsevier, v. 201, p. 106062, 2020.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S. E.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. Disponível em: <<http://arxiv.org/abs/1409.4842>>.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9.
- TAN, C.; SUN, F.; KONG, T.; ZHANG, W.; YANG, C.; LIU, C. A survey on deep transfer learning. In: SPRINGER. *International conference on artificial neural networks*. [S.l.], 2018. p. 270–279.
- TORRE, L. A.; ISLAM, F.; SIEGEL, R. L.; WARD, E. M.; JEMAL, A. *Global cancer in women: burden and trends*. [S.l.]: AACR, 2017.
- USHIZIMA, D. M.; BIANCHI, A. G.; CARNEIRO, C. M. *Segmentation of subcellular compartments combining superpixel representation with voronoi diagrams*. [S.l.], 2015.
- VARATHARASAN, V.; SHIN, H.-S.; TSOURDOS, A.; COLOSIMO, N. Improving learning effectiveness for object detection and classification in cluttered backgrounds. In: IEEE. *2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS)*. [S.l.], 2019. p. 78–85.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. *Proceedings of the xxix conference on graphics, patterns and images*. [S.l.], 2016. v. 1, n. 4.
- WAN, T.; XU, S.; SANG, C.; JIN, Y.; QIN, Z. Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks. *Neurocomputing*, Elsevier, v. 365, p. 157–170, 2019.
- WANG, C.-F. *The Vanishing Gradient Problem*. 2019. Disponível em: <<https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484>>.
- WANG, T.; HUANG, J.; ZHENG, D.; HE, Y. Nucleus segmentation of cervical cytology images based on depth information. *IEEE Access*, IEEE, 2020.
- WEI, Y.; LIANG, X.; CHEN, Y.; JIE, Z.; XIAO, Y.; ZHAO, Y.; YAN, S. Learning to segment with image-level annotations. *Pattern Recognition*, Elsevier, v. 59, p. 234–244, 2016.
- WIN, K. P.; KITJAIKURE, Y.; HAMAMOTO, K.; AUNG, T. M. Computer-assisted screening for cervical cancer using digital image processing of pap smear images. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 5, p. 1800, 2020.
- WU, J. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, v. 5, p. 23, 2017.

- XIANG, Y.; SUN, W.; PAN, C.; YAN, M.; YIN, Z.; LIANG, Y. 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.
- XIE, S.; GIRSHICK, R. B.; DOLLÁR, P.; TU, Z.; HE, K. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016. Disponível em: <<http://arxiv.org/abs/1611.05431>>.
- XU, B.; WANG, N.; CHEN, T.; LI, M. Empirical evaluation of rectified activations in convolutional network. *CoRR*, abs/1505.00853, 2015. Disponível em: <<http://arxiv.org/abs/1505.00853>>.
- YANG, X.; WU, J.; YIN, Y. Interacting convolution with pyramid structure network for automated segmentation of cervical nuclei in pap smear images. In: IEEE. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. [S.l.], 2020. p. 499–502.
- YOU, K.; LONG, M.; WANG, J.; JORDAN, M. I. *How Does Learning Rate Decay Help Modern Neural Networks?* 2019.
- ZHANG, J.; LIU, Z.; DU, B.; HE, J.; LI, G.; CHEN, D. Binary tree-like network with two-path fusion attention feature for cervical cell nucleus segmentation. *Computers in biology and medicine*, Elsevier, v. 108, p. 223–233, 2019.
- ZHANG, L.; LU, L.; NOGUES, I.; SUMMERS, R. M.; LIU, S.; YAO, J. Deeppap: deep convolutional networks for cervical cell classification. *IEEE journal of biomedical and health informatics*, IEEE, v. 21, n. 6, p. 1633–1643, 2017.
- ZHAO, J.; DAI, L.; ZHANG, M.; YU, F.; LI, M.; LI, H.; WANG, W.; ZHANG, L. Pgu-net+: progressive growing of u-net+ for automated cervical nuclei segmentation. In: SPRINGER. *International Workshop on Multiscale Multimodal Medical Imaging*. [S.l.], 2019. p. 51–58.
- ZHIQIANG, W.; JUN, L. A review of object detection based on convolutional neural network. In: IEEE. *2017 36th Chinese Control Conference (CCC)*. [S.l.], 2017. p. 11104–11109.
- ZOU, J.; XUE, Z.; BROWN, G.; LONG, R.; ANTANI, S. Deep learning for nuclei segmentation and cell classification in cervical liquid based cytology. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*. [S.l.], 2020. v. 11318, p. 1131811.

# **Apêndices**

# APÊNDICE A – Resultados dos Experimentos

Como forma de evitar muitas tabelas e informações ao longo do texto, esse apêndice irá reunir tabelas com os resultados dos experimentos realizados para as diferentes CNNs e diferentes base de dados.

As tabelas apresentam os resultados das métricas de precisão, revocação, acurácia e coeficiente de Dice para os diferentes parâmetros, tamanho de lote, função de ativação e com e sem normalização em lote. Os melhores resultados para cada métrica estão destacados de negrito. Ademais, o melhor resultado geral de cada tabela, considerando coeficiente de Dice, está destacado de amarelo.

## A.1 Herlev

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9644</b>	0,9678	<b>0,9671</b>	<b>0,9708</b>
2	32	ReLU	0,9628	0,9702	0,9665	0,9701
3	64	ReLU	0,9620	0,9701	0,9663	0,9704
4	16	Mish	0,9610	0,9701	0,9649	0,9699
5	32	Mish	0,9542	0,9703	0,9625	0,9697
6	64	Mish	0,9553	0,9705	0,9632	0,9692
7	16	Swish	0,9547	0,9708	0,9637	0,9699
8	32	Swish	0,9552	0,9709	0,9629	0,9686
9	64	Swish	0,9499	<b>0,9710</b>	0,9631	0,9700
10	16	ELU	0,9563	0,9698	0,9629	0,9682
11	32	ELU	0,9554	0,9699	0,9624	0,9683
12	64	ELU	0,9551	0,9695	0,9621	0,9677
13	16	LeakyReLU	0,9551	0,9692	0,9620	0,9679
14	32	LeakyReLU	0,9542	0,9678	0,9626	0,9689
15	64	LeakyReLU	0,9541	0,9690	0,9614	0,9681

Tabela A.1 – Resultados da arquitetura U-net com a base de dados Herlev e Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	0,9608	<b>0,9681</b>	0,9634	0,9667
2	32	ReLU	0,9583	0,9654	0,9624	0,9680
3	64	ReLU	<b>0,9665</b>	0,9636	<b>0,9648</b>	0,9677
4	16	Mish	0,9629	0,9649	0,9638	<b>0,9683</b>
5	32	Mish	0,9584	0,9648	0,9614	0,9660
6	64	Mish	0,9583	0,9635	0,9611	0,9654
7	16	Swish	0,9595	0,9637	0,9612	0,9667
8	32	Swish	0,9552	0,9636	0,9603	0,9667
9	64	Swish	0,9579	0,9620	0,9595	0,9639
10	16	ELU	0,9515	0,9625	0,9598	0,9672
11	32	ELU	0,9564	0,9624	0,9593	0,9656
12	64	ELU	0,9525	0,9556	0,9539	0,9646
13	16	LeakyReLU	0,9023	0,9003	0,9020	0,9542
14	32	LeakyReLU	0,9178	0,9149	0,9156	0,9675
15	64	LeakyReLU	0,9124	0,9140	0,9138	0,9670

Tabela A.2 – Resultados da arquitetura U-net com a base de dados Herlev, sem Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	0,9555	0,9704	0,9638	<b>0,9677</b>
2	32	ReLU	0,9566	0,9692	0,9617	0,9668
3	64	ReLU	<b>0,9667</b>	<b>0,9744</b>	<b>0,9695</b>	0,9664
4	16	Mish	0,9647	0,9689	0,9668	0,9623
5	32	Mish	0,9619	0,9694	0,9672	0,9667
6	64	Mish	0,9627	0,9689	0,9663	0,9661
7	16	Swish	0,9636	0,9649	0,9627	0,9665
8	32	Swish	0,9576	0,9651	0,9613	0,9672
9	64	Swish	0,9585	0,9690	0,9637	0,9670
10	16	ELU	0,9610	0,9657	0,9629	0,9668
11	32	ELU	0,9575	0,9671	0,9638	0,9676
12	64	ELU	0,9605	0,9678	0,9636	0,9672
13	16	LeakyReLU	0,9595	0,9634	0,9629	0,9670
14	32	LeakyReLU	0,9608	0,9673	0,9635	0,9662
15	64	LeakyReLU	0,9552	0,9674	0,9633	0,9659

Tabela A.3 – Resultados da arquitetura SegNet com a base de dados Herlev e Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9102</b>	<b>0,8866</b>	<b>0,8981</b>	0,9008
2	32	ReLU	0,9080	0,8770	0,8919	<b>0,9153</b>
3	64	ReLU	0,8970	0,8506	0,8724	0,9087
4	16	Mish	0,8710	0,7609	0,8104	0,8837
5	32	Mish	0,8941	0,7787	0,8265	0,8857
6	64	Mish	0,8934	0,8080	0,8486	0,9012
7	16	Swish	0,8804	0,5987	0,6717	0,8187
8	32	Swish	0,8810	0,5763	0,6467	0,8436
9	64	Swish	0,8773	0,5088	0,6230	0,8521
10	16	ELU	0,8780	0,5803	0,6766	0,7555
11	32	ELU	0,8256	0,5829	0,6156	0,8467
12	64	ELU	0,8292	0,5930	0,6230	0,8236
13	16	LeakyReLU	0,7658	0,6405	0,6905	0,8919
14	32	LeakyReLU	0,7816	0,6496	0,6882	0,8914
15	64	LeakyReLU	0,8021	0,6301	0,6877	0,8615

Tabela A.4 – Resultados da arquitetura SegNet com a base de dados Herlev, sem Normalização em Lote.

## A.2 ISBI2014

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	0,9710	0,9637	0,9672	0,9957
2	32	ReLU	0,9783	0,9688	0,9730	0,9966
3	64	ReLU	<b>0,9794</b>	<b>0,9714</b>	<b>0,9743</b>	0,9957
4	16	Mish	0,9764	0,9704	0,9739	0,9970
5	32	Mish	0,9755	0,9709	0,9737	0,9966
6	64	Mish	0,9657	0,9653	0,9640	0,9957
7	16	Swish	0,9634	0,9667	0,9661	0,9968
8	32	Swish	0,9656	0,9668	0,9647	0,9964
9	64	Swish	0,9632	0,9666	0,9639	0,9970
10	16	ELU	0,9636	0,9659	0,9637	<b>0,9970</b>
11	32	ELU	0,9617	0,9654	0,9631	0,9969
12	64	ELU	0,9628	0,9663	0,9635	0,9968
13	16	LeakyReLU	0,9641	0,9659	0,9637	0,9960
14	32	LeakyReLU	0,9611	0,9666	0,9638	0,9959
15	64	LeakyReLU	0,9623	0,9669	0,9636	0,9952

Tabela A.5 – Resultados da arquitetura U-net com a base de dados ISBI2014 e com Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9767</b>	<b>0,9603</b>	<b>0,9684</b>	0,9948
2	32	ReLU	0,9721	0,9606	0,9650	<b>0,9960</b>
3	64	ReLU	0,9711	0,9555	0,9625	0,9952
4	16	Mish	0,9359	0,9336	0,9451	0,9959
5	32	Mish	0,9221	0,8778	0,9037	0,9920
6	64	Mish	0,9315	0,8667	0,8949	0,9941
7	16	Swish	0,9170	0,8626	0,8907	0,9949
8	32	Swish	0,9302	0,8660	0,8936	0,9947
9	64	Swish	0,9239	0,8596	0,8884	0,9940
10	16	ELU	0,9305	0,8625	0,8906	0,9947
11	32	ELU	0,9223	0,8672	0,8943	0,9957
12	64	ELU	0,9355	0,8697	0,8967	0,9948
13	16	LeakyReLU	0,8856	0,8060	0,8393	0,9930
14	32	LeakyReLU	0,8296	0,7436	0,7752	0,9906
15	64	LeakyReLU	0,8275	0,7525	0,7837	0,9953

Tabela A.6 – Resultados da arquitetura U-net com a base de dados ISBI2014 e sem Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	0,9584	0,9573	0,9564	0,9947
2	32	ReLU	0,9586	0,9619	0,9623	0,9947
3	64	ReLU	0,9579	0,9656	0,9657	0,9941
4	16	Mish	0,9616	0,9679	0,9666	0,9960
5	32	Mish	0,9619	0,9688	0,9635	0,9963
6	64	Mish	0,9590	0,9688	0,9580	0,9959
7	16	Swish	<b>0,9776</b>	0,9594	0,9679	<b>0,9964</b>
8	32	Swish	0,9766	0,9604	0,9683	0,9940
9	64	Swish	0,8560	0,9442	0,8982	0,9962
10	16	ELU	0,9742	0,9649	0,9694	0,9961
11	32	ELU	0,9726	0,9649	0,9694	0,9944
12	64	ELU	0,9750	0,9656	0,9692	0,9950
13	16	LeakyReLU	0,9710	0,9654	0,9691	0,9962
14	32	LeakyReLU	0,9725	<b>0,9706</b>	<b>0,9726</b>	0,9961
15	64	LeakyReLU	0,9569	0,9610	0,9600	0,9914

Tabela A.7 – Resultados da arquitetura SegNet com a base de dados ISBI2014 e com Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9933</b>	<b>0,9654</b>	<b>0,9783</b>	0,9929
2	32	ReLU	0,9898	0,9566	0,9720	0,9937
3	64	ReLU	0,9583	0,8743	0,9066	0,9933
4	16	Mish	0,9369	0,5633	0,6852	0,9921
5	32	Mish	0,9360	0,6091	0,7217	0,9928
6	64	Mish	0,9295	0,4844	0,6206	0,9906
7	16	Swish	0,8803	0,7247	0,7920	<b>0,9942</b>
8	32	Swish	0,8780	0,6943	0,7694	0,9939
9	64	Swish	0,8558	0,6744	0,7504	0,9940
10	16	ELU	0,8521	0,6909	0,7627	0,9941
11	32	ELU	0,8517	0,7040	0,7704	0,9931
12	64	ELU	0,8827	0,7227	0,7823	0,9936
13	16	LeakyReLU	0,7655	0,7086	0,7495	0,9910
14	32	LeakyReLU	0,7354	0,6933	0,7133	0,9935
15	64	LeakyReLU	0,6745	0,6843	0,6783	0,9915

Tabela A.8 – Resultados da arquitetura SegNet com a base de dados ISBI2014 e sem Normalização em Lote.

### A.3 CRIC Cervix

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9420</b>	<b>0,9283</b>	<b>0,9351</b>	<b>0,9581</b>
2	32	ReLU	0,9412	0,9307	0,9352	0,9582
3	64	ReLU	0,9353	0,9249	0,9298	0,9575
4	16	Mish	0,9361	0,9242	0,9297	0,9580
5	32	Mish	0,9343	0,9237	0,9285	0,9575
6	64	Mish	0,9302	0,9190	0,9225	0,9577
7	16	Swish	0,9360	0,9245	0,9297	0,9580
8	32	Swish	0,9361	0,9245	0,9285	0,9584
9	64	Swish	0,9255	0,9210	0,9238	0,9573
10	16	ELU	0,9347	0,9183	0,9234	0,9557
11	32	ELU	0,9310	0,9184	0,9242	0,9577
12	64	ELU	0,9350	0,9127	0,9188	0,9567
13	16	LeakyReLU	0,9321	0,9193	0,9252	0,9565
14	32	LeakyReLU	0,9329	0,9203	0,9256	0,9567
15	64	LeakyReLU	0,9302	0,9180	0,9230	0,9569

Tabela A.9 – Resultados da arquitetura U-net com a base de dados CRIC Cervix e Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9524</b>	<b>0,9336</b>	<b>0,9429</b>	<b>0,9560</b>
2	32	ReLU	0,9511	0,9320	0,9414	0,9559
3	64	ReLU	0,9495	0,9296	0,9395	0,9562
4	16	Mish	0,9495	0,9303	0,9398	0,9557
5	32	Mish	0,9491	0,9300	0,9395	0,9560
6	64	Mish	0,9402	0,9155	0,9277	0,9502
7	16	Swish	0,9288	0,8992	0,9137	0,9496
8	32	Swish	0,9359	0,9096	0,9225	0,9559
9	64	Swish	0,9348	0,9068	0,9205	0,9526
10	16	ELU	0,9144	0,8724	0,8925	0,9515
11	32	ELU	0,9336	0,9054	0,9192	0,9571
12	64	ELU	0,9299	0,9003	0,9148	0,9550
13	16	LeakyReLU	0,7941	0,7232	0,7566	0,9510
14	32	LeakyReLU	0,8301	0,7698	0,7986	0,9597
15	64	LeakyReLU	0,8120	0,7535	0,7815	0,9507

Tabela A.10 – Resultados da arquitetura U-net com a base de dados CRIC Cervix, sem Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	<b>0,9029</b>	0,9104	0,9066	0,9510
2	32	ReLU	0,8897	0,9027	0,8961	0,9457
3	64	ReLU	0,8703	0,8941	0,8820	0,9402
4	16	Mish	0,8784	0,8975	0,8878	0,9545
5	32	Mish	0,8838	0,9025	0,8930	<b>0,9557</b>
6	64	Mish	0,8909	0,9093	0,9000	0,9534
7	16	Swish	0,8986	0,9122	0,9053	0,9535
8	32	Swish	0,8963	0,9076	0,9019	0,9534
9	64	Swish	0,9027	0,9166	<b>0,9096</b>	0,9552
10	16	ELU	0,8974	0,9067	0,9020	0,9532
11	32	ELU	0,8988	0,9108	0,9048	0,9535
12	64	ELU	0,8987	0,9146	0,9066	0,9535
13	16	LeakyReLU	0,8992	0,9138	0,9065	0,9543
14	32	LeakyReLU	0,9005	0,9165	0,9084	0,9540
15	64	LeakyReLU	0,8935	<b>0,9213</b>	0,9072	0,9518

Tabela A.11 – Resultados da arquitetura SegNet com a base de dados CRIC Cervix, com Normalização em Lote.

Testes	Tamanho do lote	Função de ativação	Precisão	Revocação	Dice	Acurácia
1	16	ReLU	0,8026	<b>0,6544</b>	<b>0,7130</b>	0,9042
2	32	ReLU	0,7977	0,5827	0,6456	0,9079
3	64	ReLU	0,8086	0,6108	0,6723	<b>0,9246</b>
4	16	Mish	0,7902	0,4897	0,5833	0,8908
5	32	Mish	0,8136	0,5302	0,6239	0,9127
6	64	Mish	0,8045	0,4778	0,5742	0,9016
7	16	Swish	0,8185	0,4755	0,5827	0,8994
8	32	Swish	<b>0,8379</b>	0,4699	0,5777	0,8886
9	64	Swish	0,8344	0,3825	0,5109	0,8862
10	16	ELU	0,7484	0,3700	0,4678	0,7552
11	32	ELU	0,7175	0,3082	0,4051	0,8305
12	64	ELU	0,7517	0,3362	0,4369	0,8340
13	16	LeakyReLU	0,5226	0,3824	0,4291	0,8804
14	32	LeakyReLU	0,4815	0,3712	0,4086	0,8527
15	64	LeakyReLU	0,5347	0,3886	0,4438	0,8596

Tabela A.12 – Resultados da arquitetura SegNet com a base de dados CRIC Cervix, sem Normalização em Lote.