



**UNIVERSIDADE FEDERAL DE SÃO PAULO**  
**INSTITUTO DE CIÊNCIA E TECNOLOGIA**  
**PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA**

**Avaliação do Potencial de Redes Neurais Convolucionais  
em classificar e graduar severidade em imagens  
patológicas de biópsias de próstata**

**Autor:** Maíra Suzuka Kudo

**Orientador:** Prof. Dr. Matheus Cardoso Moraes

**Coorientadora:** Profa. Dra. Katia Ramos Moreira Leite

São José dos Campos

2022

**UNIVERSIDADE FEDERAL DE SÃO PAULO  
INSTITUTO DE CIÊNCIA E TECNOLOGIA  
PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA**

**Avaliação do Potencial de Redes Neurais Convolucionais  
de classificar e graduar severidade imagens patológicas de  
biópsias de próstata**

**Autora:** Maíra Suzuka Kudo

**Orientador:** Prof. Dr. Matheus Cardoso Moraes

**Coorientadora:** Profa. Dra. Katia Ramos Moreira Leite

Monografia apresentada ao Instituto de Ciência e Tecnologia da Universidade Federal de São Paulo como requisito parcial para obtenção do título de Mestre em Engenharia Biomédica.

São José dos Campos

2022

Elaborado por sistema de geração automática com os dados fornecidos pelo(a) autor(a).

Kudo, Maíra Suzuka

Avaliação do Potencial de Redes Neurais Convolucionais em classificar e graduar severidade em imagens patológicas de biópsias de próstata / Maíra Suzuka Kudo

Orientador(a) Matheus Cardoso Moraes; Coorientador(a) Katia Ramos Moreira Leite. - São José dos Campos, 2022.

56 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Mestrado em Engenharia Biomédica) - Universidade Federal de São Paulo - Instituto de Ciência e Tecnologia, 2022.

1. Rede Neural Convolucional. 2. Inteligência Artificial. 3. Câncer de Próstata. 4. Imagem de Biópsia. I. Cardoso Moraes, Matheus, orientador(a). II. Ramos Moreira Leite, Katia, coorientador(a). III. , coorientador(a). III. Título.

**MAÍRA SUZUKA KUDO**

**Avaliação do Potencial de Redes Neurais Convolucionais  
de classificar e graduar severidade imagens patológicas de  
biópsias de próstata**

**Presidente da banca:**

Prof. Dr. Matheus Cardoso Moraes

**Banca examinadora:**

Prof. Dr. Diego Armando Cardona Cárdenas

Prof. Dr. Adenauer Girardi Casali

Prof. Dr. Thiago Martini Pereira

## **AGRADECIMENTOS**

Ao Professor Dr. Matheus Cardoso Moraes por toda a orientação, paciência, incentivo ao longo deste projeto e um exemplo de pessoa que gostaria de me espelhar.

Aos meus pais, Eduardo e Rosa, por terem me proporcionado estar aqui, por acreditarem em mim e apoiarem minhas escolhas.

Aos meus amigos e colegas pelo apoio e companheirismo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fomento no desenvolvimento.

Ao Laboratório de Processamento de Sinais e Imagens (ICT - UNIFESP) pela estrutura e por ser um excelente ambiente de aprendizagem. E ao Laboratório de Investigações Médicas (FMUSP) e Profa. Dra. Katia Ramos Moreira Leite e o colega Vinicius Meneguette Gomes de Souza pela disponibilização de imagens e apoio neste projeto.

## SUMÁRIO

AGRADECIMENTOS .....	iv
LISTA DE FIGURAS .....	vi
LISTA DE TABELAS .....	viii
RESUMO .....	9
ABSTRACT .....	11
1 INTRODUÇÃO.....	13
1.1 Câncer de próstata.....	13
1.2 O diagnóstico .....	15
1.2.1 Padrão de Gleason, Gleason Score e ISUP.....	15
1.3 Inteligência Artificial.....	18
1.3.1 Perceptron.....	18
1.3.2 Multilayer Perceptron (MLP) .....	19
1.3.3 Aprendizado Profundo (Deep Learning).....	20
1.3.4 Redes Neurais Convolucionais (Convolutional Neural Network - CNN).....	20
1.4 O Estado da Arte.....	26
2 OBJETIVOS.....	28
3 MATERIAL E MÉTODOS.....	29
3.1 Abordagem Clínica.....	29
3.2 Abordagem Computacional .....	30
3.2.1 Pré-Processamento .....	31
3.2.2 Aprendizado Profundo – Topologias Propostas .....	35
4 RESULTADOS .....	40
4.1 Resultados topologia 1 .....	40
4.2 Resultados topologia 2.....	41
5 DISCUSSÃO E CONCLUSÃO .....	48
6 REFERÊNCIAS BIBLIOGRÁFICAS .....	51

## LISTA DE FIGURAS

<b>Figura 1:</b> Localização da próstata e suas subdivisões em Apex, base, posterior, anterior. ....	14
<b>Figura 2:</b> Exemplos de padrões de Gleason. (a) Padrão de Gleason 3. (b) Padrão de Gleason 4. (c) Padrão de Gleason 5.....	16
<b>Figura 3:</b> Representação dos diferentes padrões de Gleason, desde o padrão 1 ao 5....	17
<b>Figura 4:</b> Representação do Perceptron. X representa as entradas, o w os pesos, neurônio com pesos multiplicados pela entrada e somados, e função de ativação.....	19
<b>Figura 5:</b> Representação do MLP com as camadas de input layer, Hidden Layers e output Layer.....	20
<b>Figura 6:</b> Estrutura interna da CNN. Divisão em camadas convolucionais compostas por convolução ativação, pooling e flatenning e Fully Connected representado pela MLP.	21
<b>Figura 9:</b> Representação do funcionamento do de múltiplas entradas e múltiplas saídas. ....	24
<b>Figura 10:</b> Esquema do funcionamento do MaxPooling de tamanho 2x2. ....	24
<b>Figura 11:</b> Esquema do funcionamento do Dropout. ....	25
<b>Figura 12:</b> Diagrama de blocos da metodologia. A metodologia foi dividida em abordagem clínica que consiste na demarcação das lâminas de prostatectomia pela especialista, e abordagem computacional, que realiza a extração dos patches e constrói a topologia das CNNs.....	29
<b>Figura 13:</b> Diagrama de blocos da abordagem Clínica. As imagens de espécimes de prostatectomia são adquiridas e a especialista realiza a marcação das regiões referentes aos padrões de Gleason.....	30
<b>Figura 14:</b> Fluxograma da Abordagem Computacional: No pré-processamento, os patches são extraídos e divididos em (1) classes de não câncer e de câncer, compostas pelos padrões de Gleason 3, 4 e 5 que servem de input para a topologia. (2) Para a topologia 2 separa os padrões de Gleason 3, 4 e 5, graduando a severidade. ....	31
<b>Figura 15:</b> Fluxograma do pré processamento: Primeiro foi realizada a conversão para png, depois a extração de patches, separação do dataset em câncer e não câncer. ....	33
<b>Figura 16:</b> Fluxograma do pré-processamento: Primeiro foi realizada a conversão para png, depois a extração de patches, separação do dataset em treinamento, teste e validação. ....	34
<b>Figura 17:</b> Diagrama de blocos do pré-processamento: As regiões demarcadas são isoladas e os patches são extraídos. ....	35
<b>Figura 18:</b> Fluxograma da topologia 1. ....	37
<b>Figura 19:</b> Fluxograma da topologia 2 .....	39
<b>Figura 21:</b> Separação dos grupos de treinamento, validação e teste para cada k-fold, alternando as lâminas SA, SB e SC nos devidos grupos. ....	42
<b>Figura 22:</b> Curvas de treinamento de acurácia de validação de cada <i>kfold-cross validation</i> . (a) Curva de acurácia dos dados de treinamento e validação do primeiro <i>cross-validation</i> . (b) Curva de <i>loss</i> dos dados de treinamento e validação do primeiro <i>cross-validation</i> . (c) Curva de acurácia dos dados de treinamento e validação do segundo <i>cross</i>	

*validation*. (d) Curva de *loss* dos dados de treinamento e validação do segundo *cross-validation*. (e) Curva de acurácia dos dados de treinamento e validação do terceiro *cross-validation*. (f) Curva de *loss* dos dados de treinamento e validação do terceiro *cross-validation*..... 43

**Figura 23:** Matriz de confusão para avaliação de 3 classes. (a) Ilustração do cálculo da acurácia, computada somando os valores da diagonal principal, verdadeiros positivo e negativo, e dividido pelo número total de dados. (b), (c), e (d) Ilustração dos verdadeiros positivo e negativo, e dos falsos positivo e negativo, para o cálculo de Precisão, Sensibilidade e Especificidade para cada um dos Gleasons em foco..... 45

**Figura 24:** Matriz de confusão não normalizada e normalizada dos dados de teste para cada k-fold correspondente. (a), (b), (c) matriz de confusão para k = 1,2,3, respectivamente. (d) matriz de confusão dos dados de teste com o resultado composto dos 3 k-folds..... 45



## LISTA DE TABELAS

<b>Tabela 1:</b> Precisão, Sensibilidade e Especificidade, e Acurácia para cada k correspondente.....	46
<b>Tabela 2:</b> Tabela comparativa entre os artigos citados no estado da arte.....	49

## RESUMO

De acordo com o Instituto Nacional do Câncer (INCA), mais de 68 mil novos casos de câncer de próstata foram estimados em 2018, representando o segundo tipo mais frequente de câncer entre homens no Brasil, ficando atrás somente do câncer de pele não melanoma. Existem dois tipos de tumores, o benigno também conhecido como hiperplasia benigna da próstata, e o maligno, chamado de adenocarcinoma. As chances de cura superam 90% quando diagnosticado precocemente. Para realizar o rastreamento do câncer de próstata, é realizado o exame retal digital e o teste de PSA, sendo recomendado para todos os homens acima de 50 anos. Porém, se o paciente estiver no grupo de risco, a recomendação passa ser a partir dos 45 anos. Ambos os métodos apresentam limitações de acurácia na previsão das chances de diagnóstico do câncer de próstata. Assim, caso haja alguma anormalidade nesses exames, o paciente é encaminhado para a biópsia, sendo este o padrão ouro. As amostras de biópsia são analisadas por um patologista especialista que assinala se há ou não presença de câncer. Essa análise é realizada por meio dos critérios de Gleason, estabelecidos nos anos 60 a fim padronizar e classificar lesões utilizando características visuais. A discordância entre patologistas especialistas pode chegar a 20%, criando um obstáculo significativo para as melhores práticas do paciente com câncer. Dados recentes mostram o potencial da implementação de procedimentos computacionais e métodos de processamento de imagem para a qualificação e quantificação do câncer, mas ainda há espaço para avaliações e crescimento em acurácia e cobertura diagnóstica. Considerando este vasto panorama, o objetivo geral deste projeto é implementar um método computacional utilizando Redes Neurais Convolucionais (CNNs) em imagens patológicas de biópsias de próstata. O projeto tem como propósito classificar patches de imagens obtidos por intervenções cirúrgicas em pacientes com câncer e não câncer, e posteriormente, classificar as graduações de severidade, nos padrões Gleason 3, 4, e 5. Assim, foram obtidas 32 imagens de biópsia de câncer de próstata, sendo estas revisadas por uma uropatologista especialista. Essas 32 imagens foram então transformadas em milhares de patches para alimentar as topologias propostas. Adicionalmente, mais 13 imagens retiradas do conjunto de dados aberto PANDAS foram utilizadas para complementar os dados de classificação dos padrões de Gleason 3,4 e 5. A metodologia foi dividida em abordagens clínicas para extrair patches, e abordagens computacionais, para a

implementação das CNNs. Na avaliação foi utilizado o método *k-fold cross validation*, resultando em uma acurácia de 98,3% na distinção entre amostras com câncer e não câncer. Em uma segunda topologia de classificação dos 3 padrões de Gleason, foi alcançado 85%, 93% e 96% de taxa de verdadeiro positivo para os Gleasons 3, 4 e 5, respectivamente. Conclusão: Os métodos apresentados mostraram-se eficientes na classificação de câncer de próstata e com alta acurácia na graduação de severidade.

Palavras-chave: Rede Neural Convolucional, inteligência artificial, câncer de próstata, imagem de biópsia.

## ABSTRACT

According to the National Cancer Institute (INCA, Brazil), over 68 thousand new prostate cancer cases were issued in 2018, placing as the second most common cancer to occur in men – leave the top rank for non-melanoma skin cancer. Tumors may be classified between benign and malign, whereas the prior is known as benign prostatic hyperplasia, and the latter is known as adenocarcinoma. Once it is diagnosed in early stages, cure rates reach levels above 90%. In order to track the prostate cancer, it is performed both digital rectal exam, as well as the PSA. This health check is currently recommended for people above 50 years old, unless the patient lies within a risk group – in which case it is highly recommended to start the exams by the age of 45. Both methods present accuracy limitations for the prostate correct diagnosis. Therefore, the gold standard process is to drive the patient to a biopsy procedure in case of anomalies. The biopsy samples were analyzed by a pathology specialist, who later signaled the presence – or absence – of cancer. This analysis is performed supported by the Gleason criteria methodology, established in the 60s to standardize and classify bruises making use of visual characteristics. Subjectivity is an issue for this kind of diagnosis, resulting in up to 20% accuracy disagreement between specialists. Recent data show the potential of the employment of computational methods of image processing to quantify and qualify cancer, but there is still room for improvement in accuracy and diagnostic coverage. Considering all the above said, the general goal of this project is to implement a computational method making use of Convolutional Neural Networks (CNNs) in pathological prostate biopsy images to classify patches acquired from surgical intervention images in cancer and non-cancer patients and, finally, classify the graduations into the Gleason 3, 4 and 5 standards. To perform the above cited method, 32 prostate cancer images were acquired, and then reviewed by a uropathology specialist. These images were, then, transformed into thousands of patches to feed the proposed topologies. Additionally, another 13 images were acquired from the open database “PANDAS”, which were later used to complement the classifying data into the Gleason 3, 4 and 5 standards. The methodology was segmented into clinical approaches to extract patches, and computational approaches to implement the CNNs. The k-fold cross validation method was used in the assessment stage, resulting in an accuracy of 96% in the distinction between cancer and non-cancer samples task. In a second classification topology for classifying the three Gleason standards, the method managed to reach 85%, 93% e 96% for the Gleason 3, 4 and 5,

respectively. Conclusion: The presented methods have been shown to be efficient in prostate cancer classification tasks, demonstrating great accuracy in distinguishing severity levels.

**Keywords:** Convolutional Neural Network, artificial intelligence, prostate cancer, biopsy image.

# 1 INTRODUÇÃO

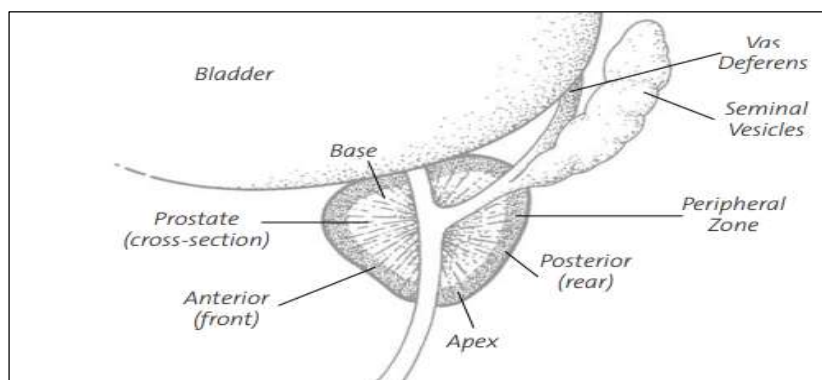
O câncer de próstata (CP) representa o quinto tipo de câncer com a maior mortalidade no mundo e segundo mais frequente entre homens. Segundo o Globocan (Global Cancer Observatory-WHO), estima-se que em 2018 cerca de 1.276.106 novos casos foram notificados no mundo todo e desses novos casos, 358.989 foram a óbito (Bray, et al., 2018). Segundo o INCA (Instituto Nacional de Câncer), estudos apontam cerca de 65.840 novos casos por ano no Brasil, entre 2020 e 2022 (Instituto Nacional de Câncer, 2020). Além disso, o CP representa 29% dos tipos de câncer no país, representando o mais comum entre homens (Instituto Nacional de Câncer, 2020).

Os principais fatores de risco do câncer de próstata estão relacionados a características genéticas, assim como ao histórico familiar (Hemminki, 2012) (Randazzo, Marco, et al., 2016), etnia, mutações hereditárias como BRCA (Bancroft, Elizabeth K., et al., 2014), (Quinn, M & Babb, P, 2002), (Lynch, Henry T., et al., 2016), (Castro, Elena, et al., 2015), síndrome metabólica e peso corporal (Esposito, K., et al., 2013) (Vidal, et al., 2014). Além disso, os hábitos alimentares e a faixa etária são importantes fatores, uma vez que estudos demonstram mais casos em obesos acima de 55 anos. O artigo de Bell et al teve como objetivo estimar a prevalência de câncer incidental por idade e mostrou que a prevalência de CP em uma população de 30 anos foi de 5%, enquanto octogenários apresentaram 60% de positividade (Bell, Katy J.L., Del Mar, Chris, Wright, Gordon, Dickinson, James, & Glasziou, Paul, 2015). Em relação à hereditariedade, o risco de CP é até 10 vezes superior em parentes de primeiro grau, quando comparado a homens com histórico familiar negativo (Santos, 2018).

## 1.1 Câncer de próstata

A próstata é uma glândula localizada abaixo da bexiga, em frente ao reto e circunscrita a uretra, pertencendo ao sistema reprodutor masculino. Sua função é nutrir e compor os fluidos do sêmen, sendo fundamental para a fertilidade (Marks, 2009). Assim, o fluido produzido na próstata promove uma menor viscosidade ao sêmen permitindo a maior locomoção dos espermatozoides, além de produzir várias enzimas como por exemplo o *Prostate Specific Antigen* (PSA) (Marks, 2009). A próstata possui dois lobos, o direito e esquerdo, assim como a porção posterior e anterior. A região mais distante da

bexiga chama-se ápex, sendo a parte mais próxima a bexiga denominada base (Marks, 2009), como é mostrado na figura 1.



Fonte: Marks, 2009.

**Figura 1:** Localização da próstata e suas subdivisões em Apex, base, posterior, anterior.

A cascata fisiopatológica do câncer de próstata ocorre através de mutações no genoma, ocasionando danos no DNA e a proliferação glandular atípica descontrolada. Essa anomalia faz com que haja a perda de aderências ultra estruturais, que pode causar metástases celulares, resultando em morte devido à progressão do câncer. O CP em estágios iniciais é geralmente assintomático e até mesmo no estágio avançado pode ter curso clínico silencioso, manifestando sintomas principalmente relacionados à doença metastática. Os sintomas mais comuns envolvem fluxo vazio fraco, dor ao urinar, disfunção erétil e presença de sangue no sêmen e na urina. Porém, todos esses achados são inespecíficos e podem ser encontrados em características benignas, como Hiperplasia Prostática Benigna (HPB) e prostatite.

As chances de cura do CP, quando diagnosticado previamente, chega a 90% (Santos, 2018). Estimular o diagnóstico precoce e assim evitar os avanços da doença metastática, são os principais motivos para manter o rastreamento de CP em todo o mundo. Os principais métodos de rastreamento incluem exame retal digital e PSA em amostras de sangue. Além disso, a triagem para o CP varia entre as recomendações das sociedades de oncologia e urologia mundiais. Considerando os benefícios de buscá-la, o debate é amplo entre especialistas em oncologia médica e urologia versus provedores de saúde pública. Dois dos estudos randomizados de rastreamento de CP mais relevantes obtiveram resultados conflitantes. O Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, apresentou aumento no diagnóstico em cerca de 12%, mas não resultou em redução geral ou específica da mortalidade por câncer (Andriole, Gerald L.,

et al., 2012). Por outro lado, a European Randomized Study of Screening for Prostate Cancer (ERSPC) demonstrou uma redução de mortalidade de 21% (Schröder, Fritz H., et al., 2012). O exame retal digital sozinho apresenta cerca de 51% de sensibilidade (Naji, et al., 2018). Além disso, deve-se prestar atenção às possíveis causas do aumento do PSA além do câncer, como a hiperplasia benigna da próstata, prostatite, infecções do trato urinário e cateterismo uretral.

As recomendações da sociedade urológica brasileira sobre triagem de CP estimulam todos os homens acima de 50 anos - e os pertencentes ao grupo de risco acima de 45 anos -, realizar uma consulta anual ao urologista com exame retal digital e coleta de sangue PSA até atingir a idade de 75 anos (Eliney F. Faria, et al., 2010).

## **1.2 O diagnóstico**

Com um rastreamento positivo, um teste confirmatório deve ser feito para detectar o CP ou excluir o diagnóstico. Atualmente, o debate envolve o uso de biópsia e meios para evitar seu risco. A ressonância magnética multiparamétrica da próstata surge como teste-chave para evitar uma biópsia desnecessária (Kasivisvanathan, Emberton, & Moore, 2018) (Rouvière, et al., 2018), entretanto a biópsia ainda é fundamental para o diagnóstico, prognóstico e tratamento da doença. As amostras de biópsia da próstata são guiadas e obtidas por meio de um ultrassom do órgão e seu protocolo envolve amostragem aleatória das áreas mais propensas a desenvolver CP (zona periférica, geralmente 12 fragmentos), além da representação da lesão quando a ressonância magnética mostra lesões suspeitas. Todas as amostras são avaliadas por um especialista em patologia para distinguir o câncer de achados benignos.

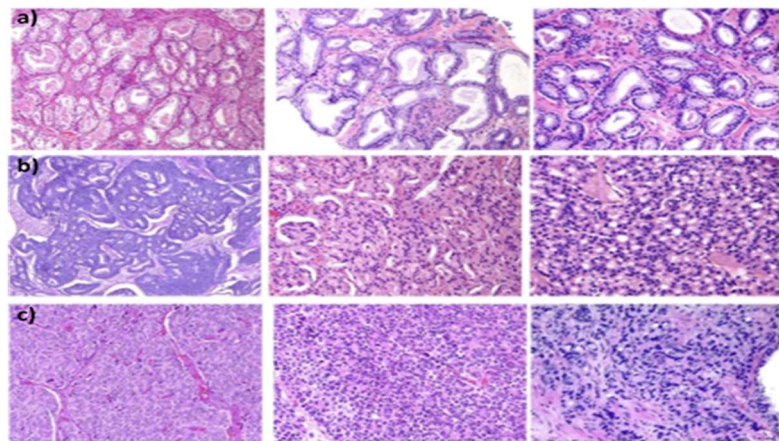
### *1.2.1 Padrão de Gleason, Gleason Score e ISUP*

O *Gleason score* surgiu em 1966 criado por Donald Gleason. Tornou-se uma escala inovadora ao analisar que o padrão arquitetural do tumor era analisado e oferecer a classificação do tumor, assim como o tratamento adequado ao paciente (Epstein, 2018). Esse primeiro modelo de classificação passou por várias modificações ao longo do tempo, até que em 2014 sofreu a última mudança a fim de que patologistas chegassem em um consenso sobre a classificação da arquitetura. Assim, a International Urology Society (ISUP) reuniu cerca de 80 profissionais qualificados do mundo todo para



atualizarem o método e entrarem em consenso baseando-se nas pesquisas mais recentes. As amostras de câncer são avaliadas pelos critérios de Gleason (Gleason, D. F. & Mellinger, G. T, 1974) (Epstein, J. I., et al.), o qual classifica a agressividade do câncer de nível inferior como o Gleason 3 até o nível superior Gleason 5. Logo, foi definido que o Gleason *Score* 1+1 = 2, assim como o *Gleason Score* 3 e 4 não devem ser classificados como câncer, pois a grande maioria das ocorrências correspondiam a casos benignos. Portanto, o Gleason *Score* foi atualizado e atualmente utiliza-se o ISUP para a graduação final.

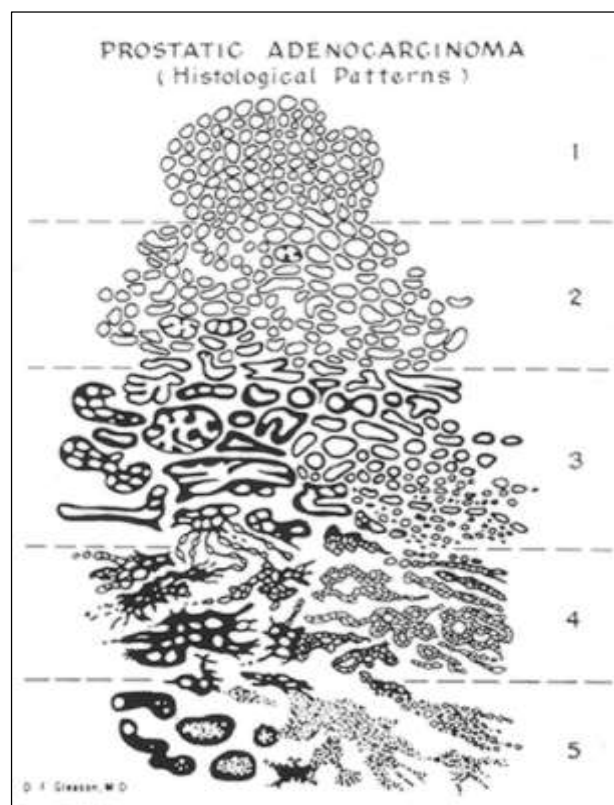
Para chegar à classificação de ISUP, o especialista determina regiões de padrões de Gleason de 3 a 5, que correspondem a diferentes estruturas histológicas. O padrão de Gleason 3 mostrado nas figuras 2(a) e 3 apresenta células de tamanho uniforme, grandes com formatos distintos, podendo apresentar dobramentos internos. O padrão de *Gleason* 4 nas figuras 2(b) e 3, possui formato cribiforme irregular, grande, com a presença de lúmen bem visualizado, ou lúmen em formato de fenda, estruturas glomeruloides e junção de glândulas. Por fim, o padrão 5 presente nas figuras 2(c) e 3, apresenta células com alta desordem, glândulas cribiformes com seu centro necrosado, formações de rosetas, vacúolos claros e espalhados e comedonecrose (Epstein, 2018).



Fonte: Epstein, 2018.

**Figura 2:** Exemplos de padrões de Gleason. (a) Padrão de Gleason 3. (b) Padrão de Gleason 4. (c) Padrão de Gleason 5.

A classificação final do Score de Gleason inclui o padrão mais frequente adicionado ao segundo mais frequente (exemplo Gleason 6 (3 + 3), *Gleason 7* (4 + 3), etc.) e simplifica a graduação de uma maneira mais inteligível para os níveis de ISUP: (*Gleason 6* (3 + 3) = ISUP 1; *Gleason 7* (3 + 4) = ISUP 2; *Gleason 7* (4 + 3) = ISUP 3; *Gleason 8* = ISUP 4 e *Gleason 9* e 10 = ISUP 5) (Epstein, J. I., et al.).



Fonte: Epstein, J. I., et al.

**Figura 3:** Representação dos diferentes padrões de Gleason, desde o padrão 1 ao 5.

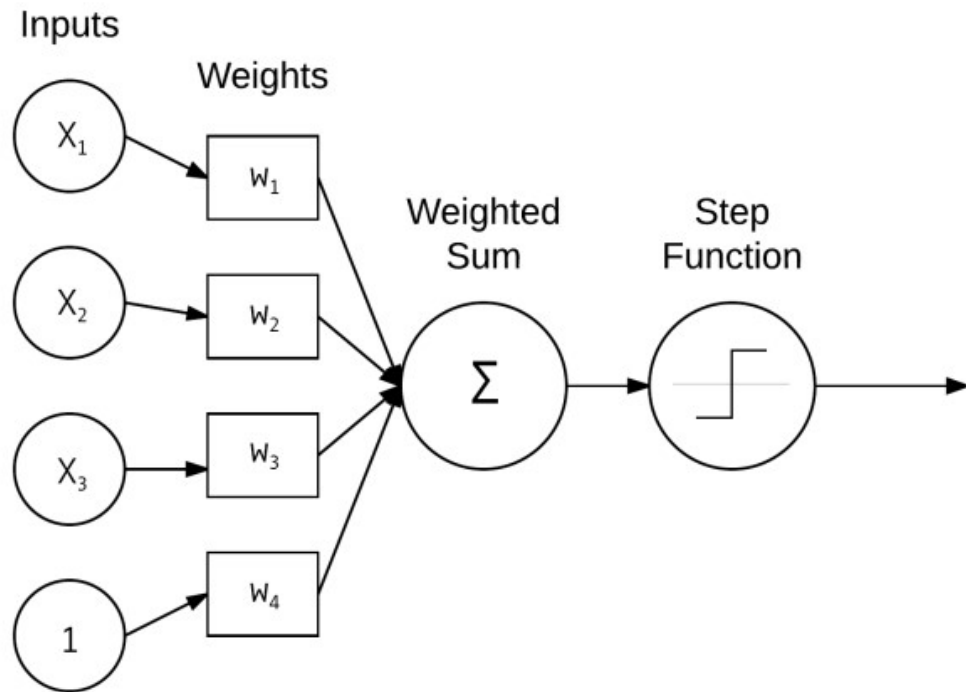
Os padrões de Gleason exigem identificações visuais de padrões e a variabilidade entre observadores é um obstáculo desafiador. A classificação subjetiva não quantitativa coloca pacientes e médicos em risco na prática médica diária. A literatura demonstrou concordância entre os uropatologistas na graduação de Gleason variando de 51 a 78% (Ozkan, T. A., et al., 2016); (Meliti, A., Sadimin, E., Diolombi, M., Khani, F., & Epstein, J. I., 2017). A fim de superar este problema, a implementação de métodos computacionais que combinam operações de processamento de imagem, extração de recursos e análise de redes neurais artificiais tem sido cada vez mais o foco de grupos de pesquisadores para homogeneizar a classificação do câncer de próstata otimizando os resultados clínicos e o tratamento do câncer.

### 1.3 Inteligência Artificial

A inteligência artificial consiste em um sistema qualificado a fim de perceber os dados externos, ser capaz de aprender corretamente e se transformar através dos erros, e assim melhorar seu aprendizado (Zhang, Lipton, Li, & Smola, 2020). Por volta de 1942, Alan Turing na Inglaterra, desenvolveu o que foi considerado o primeiro computador eletromecânico para solucionar um enigma utilizado pelos alemães na Segunda Guerra Mundial. Apesar disso, o termo IA surgiu anos depois, em 1956 por Marvin Minsky e John McCarthy que promoveram um workshop em Dartmouth, reunindo os fundadores da IA. Durante duas décadas, houve importantes contribuições de sucesso nessa área, o que estimulou o interesse em financiamentos em pesquisas de IA. Entretanto, em 1973 o congresso americano atribuiu fortes críticas aos altos gastos com as pesquisas e assim como o governo britânico retirou grande parte dos investimentos. Um dos motivos da estagnação da IA foi a criação de sistemas especialistas que consistiam em uma combinação de ideias baseadas em condicionais, “*if-else*”, que não ofereciam plasticidade de aprendizado. Em 1940 o canadense Donald Hebb já havia criado uma teoria que imitava o processo de aprendizado humano, porém havia o conhecimento de que os computadores não possuíam esse poder de processamento. Na década de 1950, surgiu o modelo perceptron que era capaz de classificar utilizando pesos sem a necessidade de intervenção humana. Uma consequente evolução do modelo perceptron foi a criação do *multilayer perceptron* (MLP), esta é composta por combinações de múltiplos *perceptrons* permitindo maior capacidade de classificação. Já em 1998 Yann LeCun revoluciona o campo da *deep learning* com a LeNet, modelo que serviu como inspiração para outras CNNs criadas posteriormente. A estrutura da LeNet era composta por camada de convolução, *pooling* e camada não linear em sequência, sendo a convolução utilizada para extrair características espaciais. Além disso, empregou-se o uso da retro propagação, utilizou as funções *tanh* e sigmoide na camada não linear e *multilayer perceptron* como classificador (LeCun, Y., Bengio, Y., & Hinton, G., 2015).

#### 1.3.1 Perceptron

O algoritmo perceptron foi definido como um modelo de aprendizado supervisionado que classifica as suas respectivas saídas em 0 ou 1. A figura 4 representa o perceptron, onde existem  $n$  inputs, e cada um será multiplicado por um peso  $w$ . Assim, passa por um nó que realiza a soma ponderada das entradas e emprega uma função de ativação que determina a classe de saída.

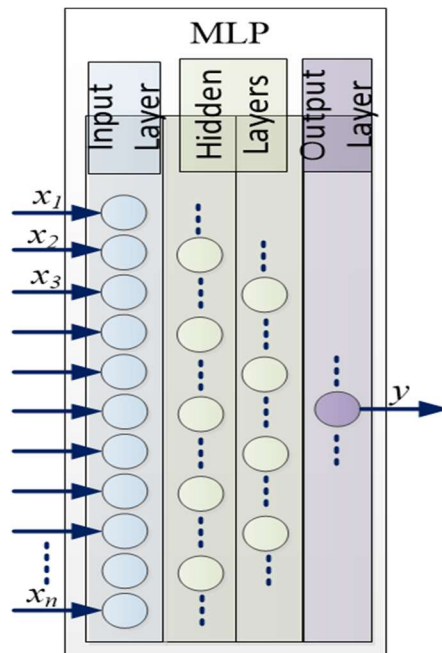


Fonte: Epstein, J. I., et al.

**Figura 4:** Representação do Perceptron. X representa as entradas, o w os pesos, neurônio com pesos multiplicados pela entrada e somados, e função de ativação.

### 1.3.2 Multilayer Perceptron (MLP)

A estrutura da MLP observada na Figura 5 combina múltiplos perceptrons em camada de entrada (Input layer), camada escondida (hidden layer) e camada de saída (output layer). A camada de entrada corresponde ao vetor de características. Já a camada escondida garante a plasticidade do classificador, contém uma quantidade de neurônios que está diretamente relacionada a construção de equações complexas que separam as classes. O número de subcamadas aumenta o grau da função de separação e assim define como unir os hiperplanos responsáveis pela separação das classes. A camada de saída também pode interferir na união dos hiperplanos, mas essa camada é dedicada para definir o número de classes da saída.



Fonte: Autoria Própria

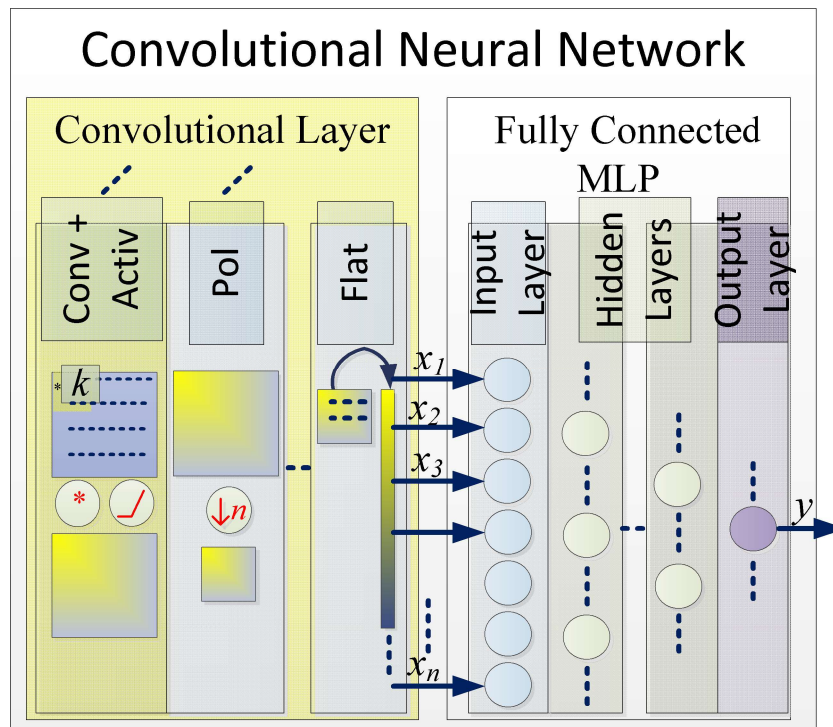
**Figura 5:** Representação do MLP com as camadas de input layer, Hidden Layers e output Layer.

### 1.3.3 Aprendizado Profundo (Deep Learning)

*Deep Learning* ou Aprendizado Profundo consiste em modelos computacionais repletos de camadas que têm a capacidade de extrair atributos necessários e utilizam a retro propagação para atualizar seus aprendizados. As redes de *Deep Learning* conseguiram oferecer um avanço significativo nas áreas de processamento de imagens, áudio e visão computacional (LeCun, Y., Bengio, Y., & Hinton, G., 2015).

### 1.3.4 Redes Neurais Convolucionais (Convolutional Neural Network - CNN)

A CNN é um tipo de estrutura pertencente a métodos de Aprendizado Profundo. Esta utiliza camadas convolucionais em sua estrutura para extrair atributos e parâmetros do meio externo (Zhang, Lipton, Li, & Smola, 2020). A camada convolucional detém 3 tipos de subcamadas, a convolução, *Pooling* e *Flattening*. Já a camada densa, que consiste no classificador, é denominada *Fully Connected* onde encontra-se o classificador *Multilayer Perceptron* (MLP). Essas camadas são dispostas e configuradas pelo desenvolvedor, visando melhor estrutura para aplicação como ilustra a figura 6.

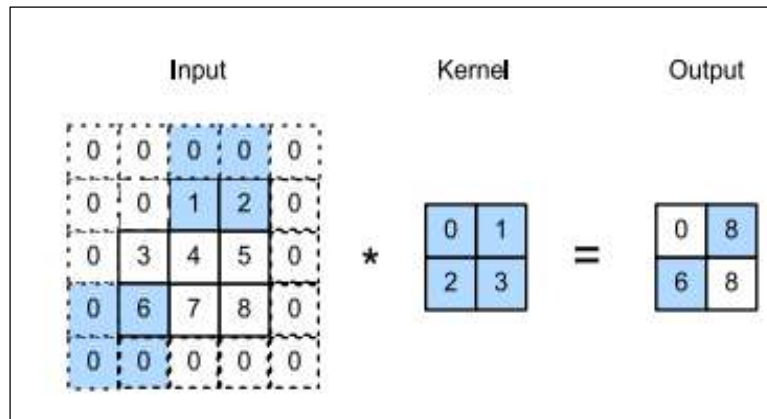


Fonte: Autoria Própria

**Figura 6:** Estrutura interna da CNN. Divisão em camadas convolucionais compostas por convolução ativação, pooling e flatenning e Fully Connected representado pela MLP.

A camada convolucional produz mapas de ativação resultantes de convoluções entre kernels e as imagens de entrada. Contudo, a rigor a operação que ocorre de fato é uma correlação (Zhang, Lipton, Li, & Smola, 2020). Logo, o mapa de ativação, consequência da convolução, representa a similaridade entre o *kernel* e a imagem de entrada, e, portanto, tem a função de potencializar a diferença entre as amostras. Além disso, alguns parâmetros são definidos pelo desenvolvedor, tais como tamanho e número de *kernels* (filtros) e o tamanho do passo para operação (*stride*). A dimensão e quantidade de *kernels*, assim como a variação dos passos, devem ser pensados baseados no conjunto de características que definem e discriminam as classes a serem separadas. Neste sentido, quanto mais filtros, mais atributos específicos serão extraídos e consequentemente maior a acurácia aparente. Contudo, esta superespecialização pode ser prejudicial em alguns aspectos. O primeiro ponto negativo é a baixa capacidade de generalizar o modelo para ser aplicado a dados levemente diferentes, devido à alta especialidade das características pontuais do conjunto de dados usados, afetando a robustez do modelo. O segundo ponto é a elevada quantidade de atributos que podem ser desnecessários, tornando o custo computacional desnecessariamente alto.

Por fim, o *stride*, representado na figura 7, corresponde ao passo em que o kernel desloca-se dentro da imagem, alterando seu tamanho e resolução, o que influencia diretamente nas particularidades das informações extraídas. Assim, o tamanho do *kernel* e *stride* devem também serem refletidos baseados no tamanho e separação das características que diferenciam as classes dentro de uma imagem.



Fonte: Zhang, Lipton, Li, & Smola, 2020

**Figura 1:** Representação do funcionamento do stride.

Assim como no *perceptron*, dentro das cadeias de camadas convolucionais a função de ativação realiza uma modulação da magnitude de informações desejáveis e não desejáveis. Na neurociência, esta consiste na taxa de disparo do neurônio que é modulada dependendo da quantidade de estímulo que foi dado. Existem vários tipos de função de ativação, mas a mais utilizada dentro das camadas convolucionais por seu desempenho é a *Rectified Linear Unit (ReLU)* definida como:

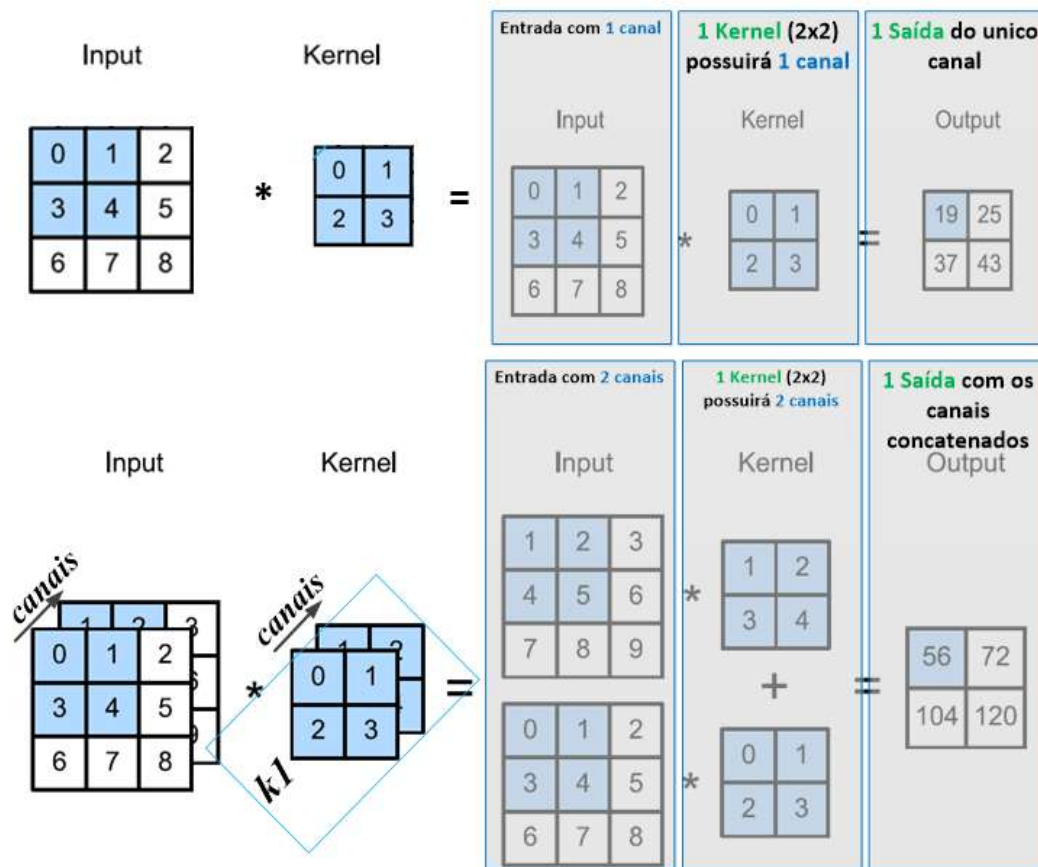
$$f(x) = x^+ = \max(0, x)$$

em que  $x$  é o *input* da função que anula valores negativos, transformando-os em 0, e mantém os valores de entrada  $x$  maiores que zero. Ou seja, é uma função linear para valores maiores que zero (Glorot, X., Bordes, A., & Bengio, Y., 2011).

*Múltiplas entradas e Saídas* – No caso da entrada da camada convolucional possuir mais de um canal, o *kernel* treinado também terá a mesma quantidade de canais do *input*. Assim, o mesmo *kernel* irá realizar convoluções paralelas em todos os canais e, em cada um deles, vai expressar a característica desejada para a classificação. Como exemplo temos o caso da figura 8, em que há uma entrada com um canal, um kernel de tamanho 2x2 também com um canal, devido a entrada. Assim a saída será o resultado da convolução desse único canal. Quando existe uma entrada com dois canais, o *kernel* também possuirá 2 canais, que vai produzir a concatenação das convoluções desses 2



canais.



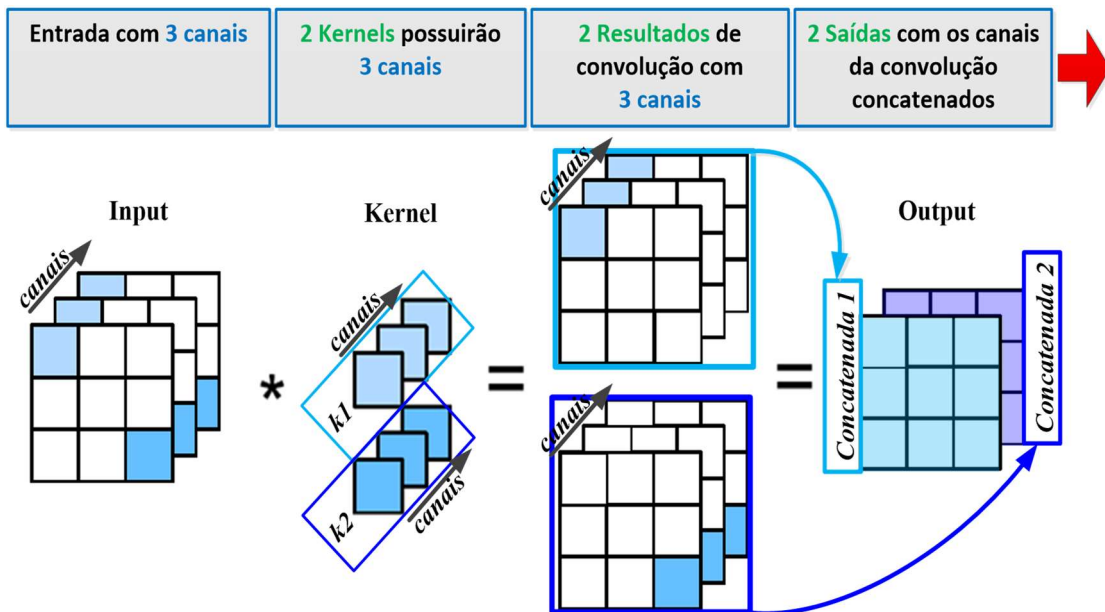
Fonte: Adaptação de Zhang, Lipton, Li, & Smola, 2020

**Figura 2:** Representação do funcionamento do de múltiplas entradas e múltiplas saídas.

Como foram gerados *kernels* com o mesmo número de canais da entrada, assim como a convolução em cada canal, a saída será correspondente ao número de *kernels* e ocorrerá a concatenação dos canais de convolução. Utilizar mais de uma entrada é interessante pois neste caso cada canal pode deter padrões distintos que, quando combinados, auxiliam na construção do vetor de características.

Já no caso em que há vários kernels, como na figura 9, O número de saídas será igual ao número de kernels definidos. Assim, o kernel irá realizar convoluções paralelas em todos os canais e, em cada um deles, vai expressar a característica desejada para a classificação. Para uma entrada com 3 canais e 2 kernels, cada kernel também possuirá 3 canais, e assim, a saída terá 2 resultados de convolução dos 3 canais cada. Cada uma das 2 saídas com 3 canais será concatenada, gerando uma saída com 2 matrizes (figura 9).

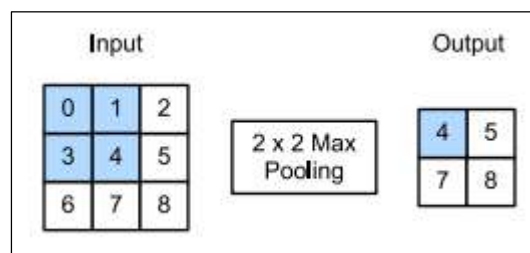




Fonte: Adaptação de Zhang, Lipton, Li, & Smola, 2020

**Figura 7:** Representação do funcionamento do de múltiplas entradas e múltiplas saídas.

O *Pooling* tem a função de diminuir a dimensão do mapa de ativação, restando apenas as informações mais relevantes. Essa diminuição deve ser realizada de forma a minimizar efeitos de translação, escala e rotação da informação encontrada na operação de convolução. Existem vários critérios de *Pooling*, como selecionar o valor máximo, menor valor ou a média em uma determinada janela, o que minimiza os efeitos de translação, rotação e deslocamento. Neste trabalho foi utilizado o *MaxPooling* (figura 10), o qual se refere ao maior valor encontrado na janela definida na amostra.



Fonte: Zhang, Lipton, Li, & Smola, 2020

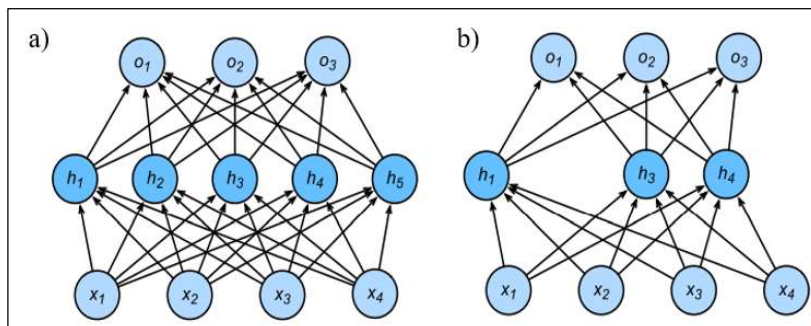
**Figura 8:** Esquema do funcionamento do MaxPooling de tamanho 2x2.

Já, o *Flattening* transforma o mapa de atributos 2D em um vetor 1D, sendo que cada posição desse vetor corresponde a um atributo da entrada da *Multilayer Perceptron*. Assim, o tamanho do vetor 1D é relacionado ao mapa de ativação final, considerando os

redimensionamentos com os *strides* e *poolings*. Esse vetor, produto das etapas convolucionais, trará diferentes disposições de valores relacionados às diferentes características extraídas. Essa disposição de valores corresponde à entrada da camada de *Fully Connected (MLP)*.

Por fim, tendo como entrada as diferentes disposições de valores, relacionados com cada classe, as camadas de neurônios da MLP são treinadas a fim de encontrar a melhor combinação de pesos para a classificação das classes. Esses valores são ajustados durante o treinamento, gerando a melhor combinação entre si antes de entrarem para a última camada de classificação. Os números de neurônios da camada de entrada são relacionados ao tamanho do vetor 1D resultante, enquanto o número de neurônio de saída é relacionado ao número de classes que se deseja classificar. A quantidade de subcamadas e neurônios da *hidden layer* são investigados e proporcionais ao conjunto e a complexidade de classificação das classes.

*Dropout* – É uma técnica que consiste em aplicar ruído ao longo da propagação direta enquanto processa cada camada, resultando em uma redução do *overfitting*. O *overfitting* ocorre quando uma camada dispara um padrão específico de ativações para a próxima camada, tornando esta última dependente deste padrão. Esse estado é chamado também de co-adaptação. O *dropout* tem esse nome pois não utiliza uma porcentagem pré-definida de neurônios durante o treinamento, interrompendo essa co-adaptação. Assim, há a remoção de viés de cada camada devido a normalização pela fração de nós retidos. Na figura 11(a) temos um classificador MLP com apenas uma *hidden layer*. Retirando os neurônios  $h_2$  e  $h_5$  (figura 11(b)), o cálculo da saída não será dependente de todos os elementos do *hidden layer*.



Fonte: Zhang, Lipton, Li, & Smola, 2020

**Figura 9:** Esquema do funcionamento do Dropout.

*Batch Normalization* (BN) – É uma técnica aplicada a cada uma das camadas a cada iteração. Após a ativação ser calculada, uma normalização é realizada em cada nó através da subtração da média e divisão pelo desvio padrão. Logo, é realizada uma estatística das quantidades baseadas nesse mini *batch* atual. No contexto das camadas convolucionais, o BN é fundamental para que haja uma normalização de cada saída dos canais, pois cada canal possui parâmetros de deslocamento e sua própria escala.

#### 1.4 O Estado da Arte

O estado da arte tem mostrado o potencial da implementação de métodos e procedimentos computacionais para a qualificação e quantificação do câncer em seus diversos tipos, graus e classificações. Propostas mais antigas apresentam métodos baseados em processamento e/ou aprendizado de máquina fazendo uso de classificadores não supervisionados como k-means, ou supervisionados como *Support Vector Machine* (SVM). No entanto, com o advento das capacidades computacionais propostas de topologias utilizando redes convolucionais, surgiram inúmeras aplicações em visão computacional.

Há diversas pesquisas publicadas tanto em áreas *de machine learning* quanto em CNNs recentes. No artigo de Nagpal e colegas (Nagpal, et al., 2019), há a classificação dos padrões de Gleason regionais por meio de uma CNN e depois por um classificador fundamentado em *k-nearest-neighbor*. Além da classificação, foi realizada uma avaliação e análise profunda do método proposto, mostrando que seu desempenho foi superior ao de especialistas (Nagpal, et al., 2019). No estudo de Li e colegas (Li, et al., 2020), foram utilizadas imagens de *microarray* de câncer de próstata em uma rede de estrutura paralela e linear a fim de segmentar o *microarray* de tecido nos 3 padrões de ISUP (Li, et al., 2020). No trabalho de Bulten e colegas, foram utilizadas *Unet* para a segmentação de regiões de interesse, e *cycleGANs* para o treinamento não supervisionado, propondo-se uma solução para o problema de falta de imagens (Bray, et al., 2018). Neste trabalho foram utilizadas 5759 biopsias coletadas de 1243 pacientes, alcançando um quadrático Cohens's Kappa de 0.918. Já em Pantanowitz, et al., (Pantanowitz, et al., 2020), 512 slides de biopsia de câncer de próstata foram transformados em 1 357 480 patches. Utilizaram um *gradient boosting* para distinguir o fundo do slide e a área de tecido para então processar um conjunto de 3 modelos diferentes de CNN. Logo implementaram as

topologias Inception V1, Inception V3 e Resnet. Ao final do processo, são gerados 18 mapas de calor que combinados computam as pontuações de Gleason 5, Gleason 7+, invasão perineural, câncer e porcentagem de câncer. Alcançaram o potencial de sensibilidade de 96% e especificidade de 89% no conjunto de validação (Pantanowitz, et al., 2020). No trabalho de Yechan, Paik, *et al.*, foram utilizados 7600 slides inteiros de biopsia de câncer de próstata providas de duas instituições distintas a fim de encontrar as regiões de interesse por meio de uma CNN combinada com o aprendizado de múltiplas instâncias. Obtiveram a precisão de 77,5% com um kappa score de 0,65 (Yechan, Paik, Shin, Kwak, & Chang, 2021). No artigo de Tolkach, Dohmgörge e colegas, foram utilizados *patches* de tamanhos variados de lâminas de câncer de próstata e a NASNetLarge uma *convnet* recente e pouco explorada na área da histopatologia. Na primeira etapa, foi proposto a classificação de *patches* com tecido benigno e tecido tumoral, conseguindo a acurácia geral do modelo de 98%. Já na segunda etapa a mesma *convnet* foi treinada para classificar as 3 classes de padrão de Gleason 3, 4 e 5 (Tolkach, Dohmgörge, Toma, & Kristiansen, 2020).

Os trabalhos mencionados trouxeram informações e auxiliaram no avanço da compreensão da aplicação de métodos de inteligência artificial focados no auxílio ao diagnóstico por imagens patológicas. Mesmo apresentando bons resultados, essas investigações são recentes e apresentam carência na abrangência do contexto que os estudos estão inseridos. Além disso, em algumas pesquisas nem sempre está claro a descrição do método para reprodução, ou não conseguem boa acurácia, ou detém um elevado custo computacional. Assim, há necessidade de novas investigações e aperfeiçoamento para melhorias em relação à abrangência, acurácia, precisão e custo computacional da classificação.

## **2 OBJETIVOS**

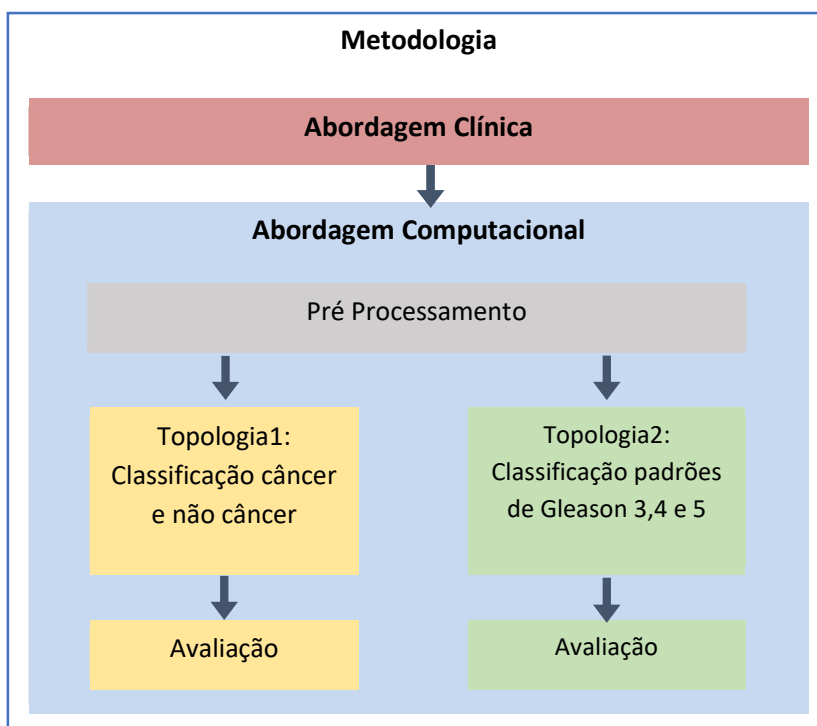
Sendo assim, o objetivo do presente trabalho consiste em desenvolver e avaliar o potencial de redes neurais convolucionais de classificar e graduar a severidade de imagens patológicas de biópsias de próstata. Especificamente, foram propostas, desenvolvidas, e avaliadas duas topologias de CNN e encontrou-se seus melhores parâmetros visando a robustez e a elevada acurácia:

1º Objetivo específico: Criar e avaliar uma topologia de CNN para classificação de amostras entre câncer e não-câncer.

2º Objetivo específico: Criar e avaliar uma topologia de CNN para graduação de severidade, e assim classificar amostras de Gleasons 3, 4, 5.

### 3 MATERIAL E MÉTODOS

A metodologia foi dividida em duas seções: Abordagem Clínica e Abordagem Computacional (Figura 12). As ações clínicas consistem na aquisição das imagens e contam com uma patologista especialista que realiza a seleção das regiões específicas do câncer. Já as ações computacionais consistem no pré-processamento para a extração de *patches* e no desenvolvimento de uma CNN para a classificação dos *patches* em câncer e não câncer e outra CNN para a classificação nos padrões de *Gleason*.



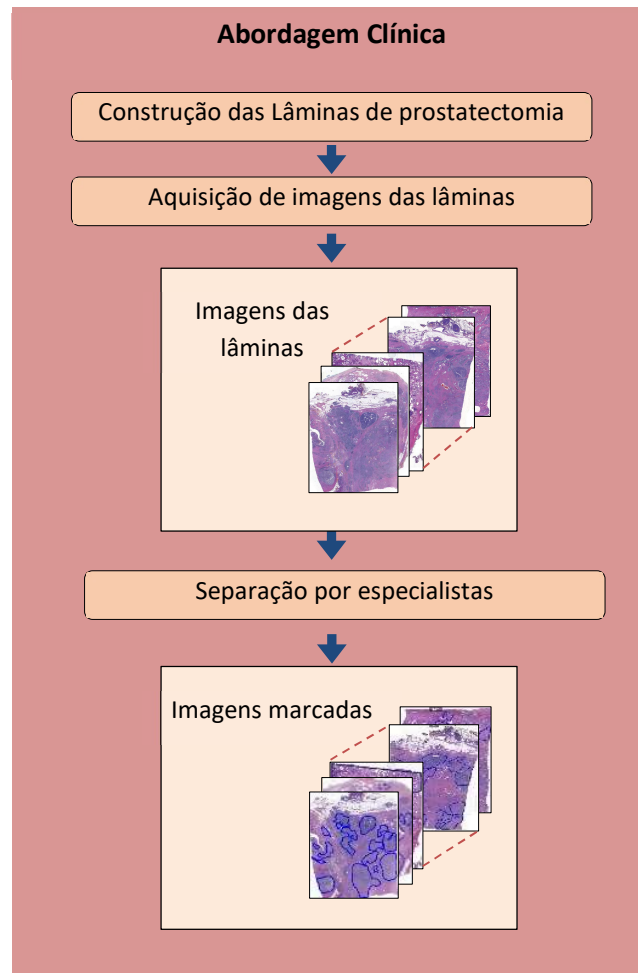
Fonte: Autoria Própria

**Figura 10:** Diagrama de blocos da metodologia. A metodologia foi dividida em abordagem clínica que consiste na demarcação das lâminas de prostatectomia pela especialista, e abordagem computacional, que realiza a extração dos *patches* e constrói a topologia das CNNs.

#### 3.1 Abordagem Clínica

Na primeira seção, como pode ser visto na figura 13, as lâminas coradas com hematoxilina e eosina (H&E) de espécimes de prostatectomia foram digitalizadas pelo sistema Aperio®, com um zoom aproximado de 20 vezes, em seguida, examinadas por uma uropatologista especialista, sendo indentificadas as áreas de tumor e as áreas em que não havia câncer. Denominou-se não câncer as regiões que não haviam tecido canceroso,

mas que poderiam haver outro tipo de patologia benigna, e também foram incluídas as áreas de tecido saudável. Um padrão de amostragem altamente confiável foi realizado para aumentar a qualidade do conjunto de dados, o que deu suporte aos resultados da CNN para um padrão clinicamente eficiente.



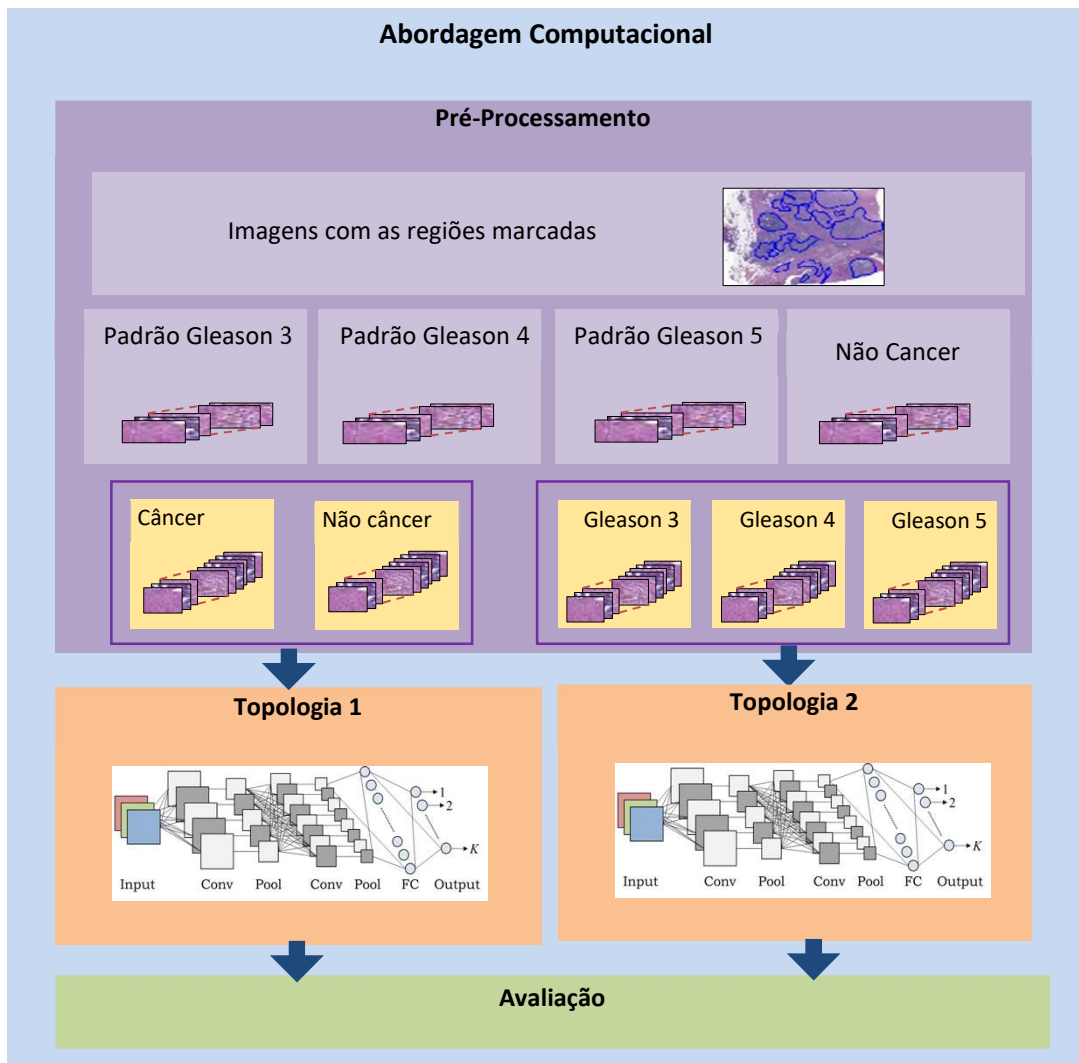
Fonte: Autoria Própria

**Figura 11:** Diagrama de blocos da abordagem Clínica. As imagens de espécimes de prostatectomia são adquiridas e a especialista realiza a marcação das regiões referentes aos padrões de Gleason.

### 3.2 Abordagem Computacional

A presente seção, conforme mostrado na figura 14, consiste em preparar as imagens de slides das biópsias para torná-las o *input* da CNN. Neste caso, foi utilizado o *software* Case Viewer para visualização dos slides e para a conversão para a extensão de imagem png. Na sequência, foi utilizado o python – com as bibliotecas Keras e Tensorflow -, com o objetivo de estruturar o escopo do CNN. As abordagens

computacionais foram divididas em três etapas: Pré-Processamento, Aprendizado Profundo (Topologia 1 e Topologia 2), e Avaliação.



Fonte: Autoria Própria

**Figura 12:** Fluxograma da Abordagem Computacional: No pré-processamento, os patches são extraídos e divididos em (1) classes de não câncer e de câncer, compostas pelos padrões de Gleason 3, 4 e 5 que servem de input para a topologia. (2) Para a topologia 2 separa os padrões de Gleason 3, 4 e 5, graduando a severidade.

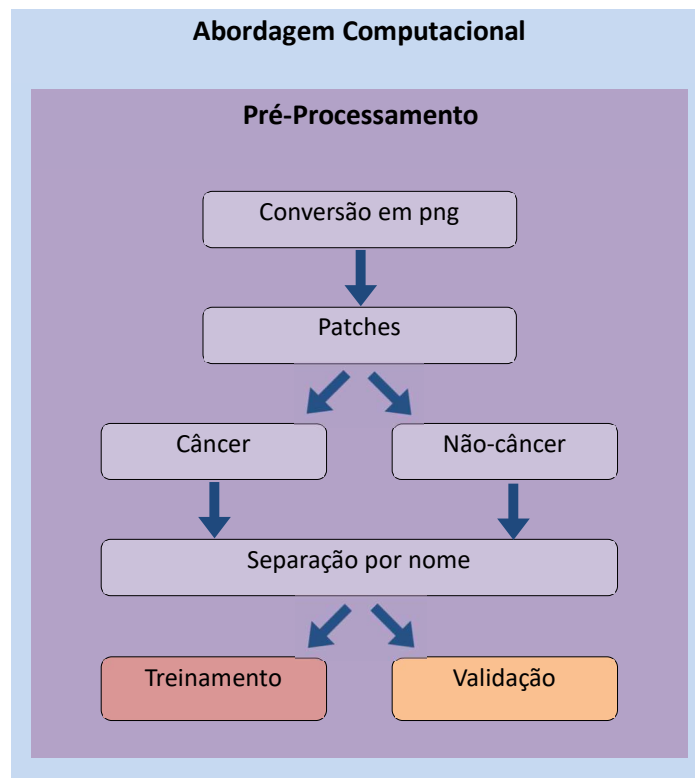
### 3.2.1 Pré-Processamento

O estágio de pré-processamento é dedicado a adquirir, organizar e assim formar um conjunto de dados contendo as regiões relevantes da biópsia. Com as amostras clínicas disponibilizadas foram construídos dois *datasets*, um para a classificação de *patches* câncer e não câncer, e outro para a classificação nos padrões de *Gleason*, realizados nos estágios de pré-processamento.



Inicialmente, o formato de imagem adquirido do fabricante é o *mrxs*; portanto, foi necessário convertê-lo em *png* para manipulá-lo (Figuras 13 e 17a). As regiões que não foram delimitadas na etapa de Abordagens Clínicas foram retiradas das imagens, restando apenas as regiões de interesse (Figura 17b). Consequentemente, conjuntos de amostras, *patches* de 256 por 256 pixels, foram extraídos das áreas marcadas restantes (Figura 17c). O tamanho da janela foi escolhido considerando a resolução do scanner utilizado; assim, este tamanho contém os elementos clínicos essenciais para fornecer informações suficientes para processar e discriminar cada tipo diferente de tecido. Especificamente, é importante que este tamanho de *patch* contemple núcleos, lúmens e ácinos. A característica de disposição e conjunto morfológico da combinação dos elementos mencionados é a espinha dorsal para classificar a amostra.

*Dataset usado na Topologia 1* - O primeiro *dataset*, detém um número total de 2594 *patches*, igualmente dividido em duas categorias, câncer e não câncer. A extração desses *patches* foi realizada mais próxima ao centro das regiões demarcadas. Classes com mesmos números de dados definem um peso igual no treinamento do modelo. Em seguida, um aumento no *dataset* com *data augmentation* foi realizado rotacionando os 2594 *patches* em 90, 180 e 270 graus, ampliando o número de amostra para 10.376. Este aumento, melhora a robustez, minimizando efeitos de translação e rotação. Após essa etapa, realizou-se a separação por nome da lâmina para evitar que dados da mesma lâmina fossem tanto para o grupo de treinamento quanto de validação. O conjunto de dados foi então separado pelo nome e randomizado para evitar vieses de seleção na próxima etapa (Figura 15). Nesta separação aleatória, foram criados 2 grupos, o primeiro com 6918 conjuntos de imagens de treinamento, e o segundo com 3458 para teste.



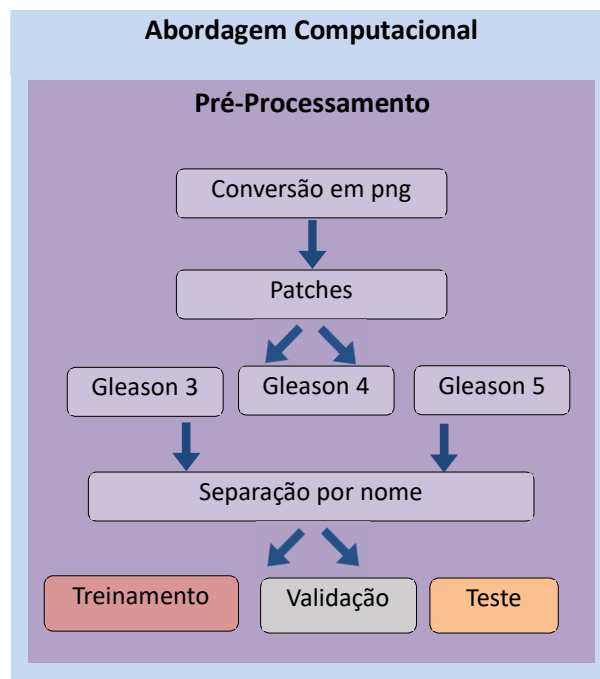
Fonte: Autoria Própria

**Figura 13:** Fluxograma do pré processamento: Primeiro foi realizada a conversão para png, depois a extração de patches, separação do dataset em câncer e não câncer.

*Dataset usado na Topologia 2* - Para a formação do segundo dataset (figura 16) utilizou-se, além das 32 imagens demarcadas pela especialista, também dados do Prostate Câncer Grade Assessment (PANDA) Challenge o qual detém o maior conjunto de dados aberto para fins de pesquisa de câncer de próstata. O PANDA é composto pelos dados do Radboud University Medical Center, com diversas imagens demarcadas em *background*, estroma, tecido benigno, padrão Gleason 3, padrão Gleason 4 e padrão Gleason 5, todas coradas com H&E. Assim, de todas as imagens disponíveis foram utilizadas 13 a fim de complementar o dataset formado anteriormente e implementar um método robusto capaz de generalizar os dados disponíveis. O *dataset* foi construído com *patches* de tamanho 200x200, de imagens foram adquiridas com zoom de 20x, suficiente para englobar estruturas e padrões vistos pela análise de especialistas

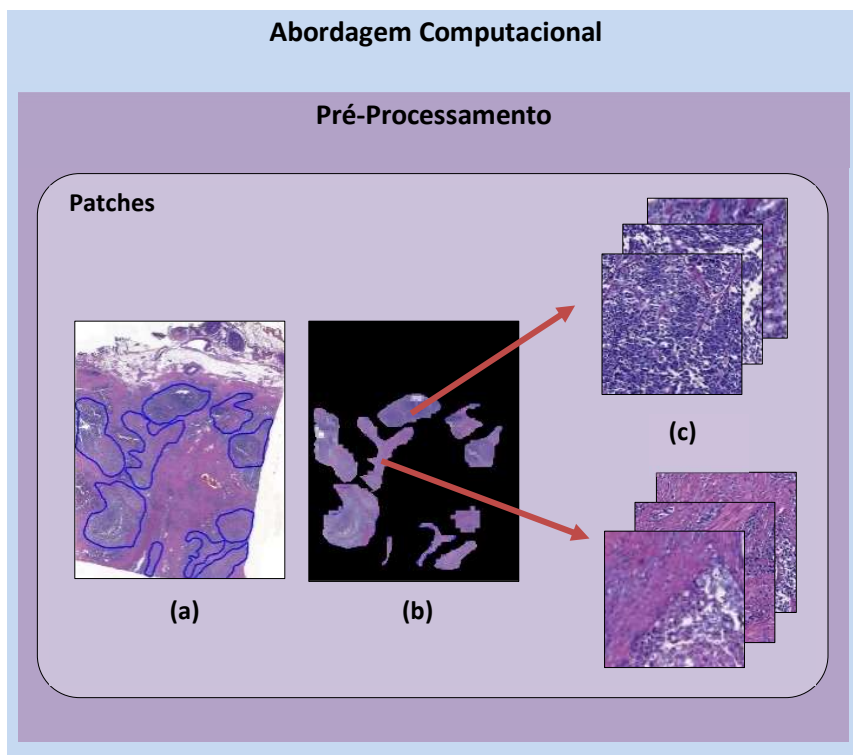
O número total de *patches* extraídos foi de 6982, sendo 5036 do *dataset* do nosso colaborador e o restante do PANDAS *Challenge*. Cada *patch* foi nomeado pela região e lâmina em que se encontrava para facilitar sua identificação. Foi realizada a separação do

treinamento, teste e validação (80%-10%-10%) através do nome da lâmina, assim *patches* da mesma região ficaram no mesmo grupo de treinamento, teste ou validação para evitar falhas na acurácia. E após essa etapa os *datasets* foram randomizados para ser o *input* da rede. Além disso, também foi realizado o *data augmentation* a fim de reduzir a especificidade do modelo e auxiliar na aprendizagem, com rotações, brilho e zoom aleatórios.



Fonte: Autoria Própria

**Figura 14:** Fluxograma do pré-processamento: Primeiro foi realizada a conversão para png, depois a extração de patches, separação do dataset em treinamento, teste e validação.



Fonte: Autoria Própria

**Figura 15:** Diagrama de blocos do pré-processamento: As regiões demarcadas são isoladas e os patches são extraídos.

### 3.2.2 Aprendizado Profundo – Topologias Propostas

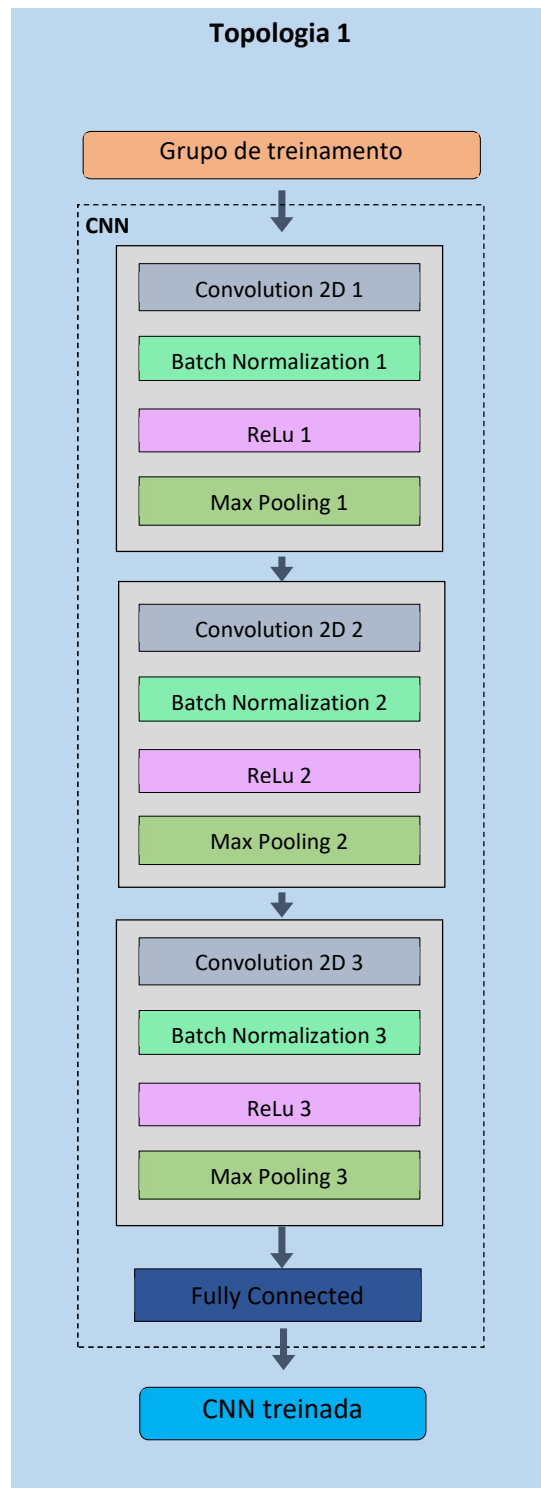
O estágio de aprendizado profundo é dedicado à aplicação de um conjunto organizado de dados para configurar e treinar um modelo de CNN para uma aplicação. A topologia de rede neural aplicada foi a *ConvNet* tradicional, uma estrutura CNN que aplica *kernels* convolucionais de tamanhos específicos seguidos de operações relacionadas de forma serial. Assim, promove-se a extração dos melhores atributos para classificação em uma aplicação específica. Neste projeto foram desenvolvidas duas CNNs, a primeira com a finalidade de classificar *patches* em câncer e não câncer, e a segunda para graduar a severidade do câncer, e assim classificar os padrões de *Gleason* 3, 4 e 5.

**Proposta de Topologia 1** - A topologia proposta para a primeira aplicação foi dividida em 3 blocos convolucionais com *Batch Normalization* e *MaxPooling*, e classificando com a camada *Fully Connected* (Figura 18).

Inicialmente, os *patches* do *dataset* 1 com tamanho de 200 por 200 pixels são

usados como *input* do modelo. Assim, as convoluções são realizadas para produzir os respectivos mapas de ativação. Os mapas de ativação são fonte primária de informação para construir atributos para a devida aplicação. Especificamente, foram definidas uma sequência de 3 blocos de convoluções com 16, 32, 64 filtros respectivamente, usando a função de ativação ReLu, com os seguintes tamanhos distintos para os *kernels* correspondentes: 5x5, 7x7 e 15x15. Posterior a cada uma das 3 operações de convolução, foram inseridas as operações de *Batch Normalization* e *Max Pooling*. O *Batch Normalization* foi aplicado para normalizar as entradas e padronizá-las, ou seja, diminuir a covariância das amostras. Já o *Max Pooling* foi configurado com kernel de 2x2 e *stride* de 2, assim há a diminuição das dimensões da matriz pela sua metade ao final de cada um dos blocos convolucionais. Os tamanhos dos kernels das camadas convolucionais e *Max Pooling* foram estimados empiricamente para proporcionar a busca de informação em 3 diferentes escalas. Por tanto, informações de distribuição espacial dos núcleos celulares e morfologia de elementos combinados são identificados para a classificação. Além disso, há a potencialização do afinamento entre o conjunto de *kernels* escolhidos e as informações da lâmina em diferentes escalas.

Como resultado das operações anteriores, o mapa de ativação resultante possui tamanho de 25 linhas e 25 colunas. Consequentemente, este é transformado em 1D, resultando em um vetor de tamanho 625 posições para servir de entrada para a camada *Fully Connected*. Portanto, os parâmetros utilizados neste bloco foram: 625 neurônios de entradas intrínsecos, 64 neurônios na camada oculta, e 1 neurônio de saída para gerar os valores 0 ou 1 correspondente a cada classe. Enquanto as funções de ativação utilizadas foram: linear para a de entrada e sigmoide para as duas últimas. A *learning rate* foi de 0,001, uma vez que obteve a menor oscilação na descida do gradiente. O número de camadas de entrada, ocultas e de saída, bem como, o número de neurônios para boa classificação, foram pré-configurados e empiricamente calibrados durante o estágio de treinamento.



Fonte: Aatoria Própria

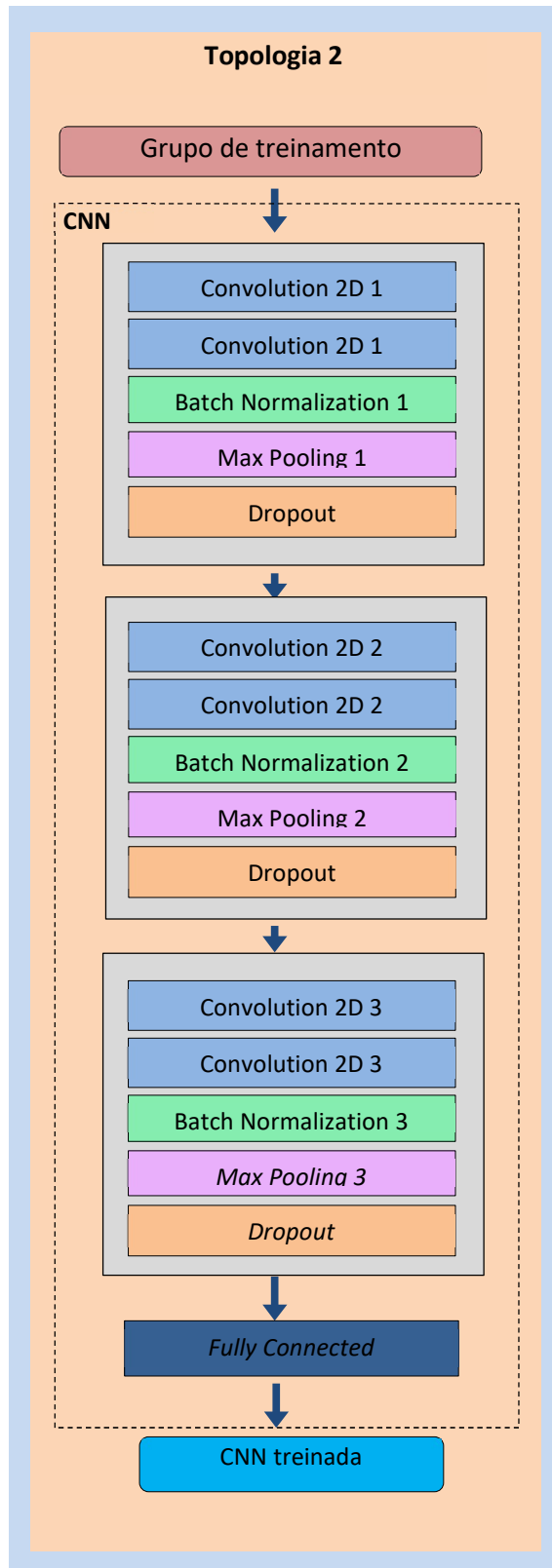
**Figura 16:** Fluxograma da topologia 1.

**Proposta de Topologia 2** – Para a topologia proposta 2, inicialmente considerou-se experimentos com *transfer learning* e redes complexas. Porém, após adaptações para melhor relação entre performance e complexidade, a estrutura final obtida esta apresentada na figura 19.

Inicialmente, como *input*, temos *patches* com tamanho de 256x256 os quais passam por uma sequência de 3 pares de convoluções. O primeiro par possui duas subcamadas convolucionais; a primeira com 32 filtros de *kernel* 5x5 e a segunda com 64 filtros de kernel 5x5. O segundo par possui 64 e 128 filtros, respectivamente, e ambas com kernel de 3x3. O terceiro par possui 128 e 256 filtros, ambos também de tamanho 3x3. Os *kernels* de tamanhos menores utilizados, buscam e armazenam características de baixa escala; tais como, orientação, borda, e principalmente, de padrão de cores. Assim, há discriminação das classes pela identificação dos núcleos celulares e suas disposições espaciais, diretamente relacionadas com o padrão de Gleason. A quantidade de filtros relaciona-se com a capacidade de diversificação de extração de características dentro do conjunto da imagem, sendo que quanto maior a quantidade de filtros, maior a capacidade de extração de características específicas. Por fim, camadas convolucionais em série permitem uma extração de características combinadas, na qual a segunda refina e potencializa as informações adquiridas pela primeira subcamada.

Cada um dos pares de convolução é seguido por operações de *batch normalization*, *maxpooling* com tamanho de 2x2, e *dropout* de 0.25. O *batch normalization*, para normalização, o *Max Pooling*, para redução de dimensão da matriz de característica e o *dropout*, com valor de 0.25 para auxiliar na regularização. Considerando as sequências acima, o mapa de ativação resultante da camada convolucional será de 32x32x256, totalizando 262144 posições de entrada para a camada *Fully Connected*.

Por fim, a *Fully Connected*, além do número de entradas mencionadas, emprega 512 neurônios na camada escondida, e 3 neurônios na camada de saída. Adicionalmente, também foram empregados *batch normalization*, *dropout* de 0.5. A função de ativação utilizada foi a *softmax*. Conseqüentemente, a configuração acima resulta na classificação de cada *patch* nos padrões de *Gleason* 3, 4, e 5.



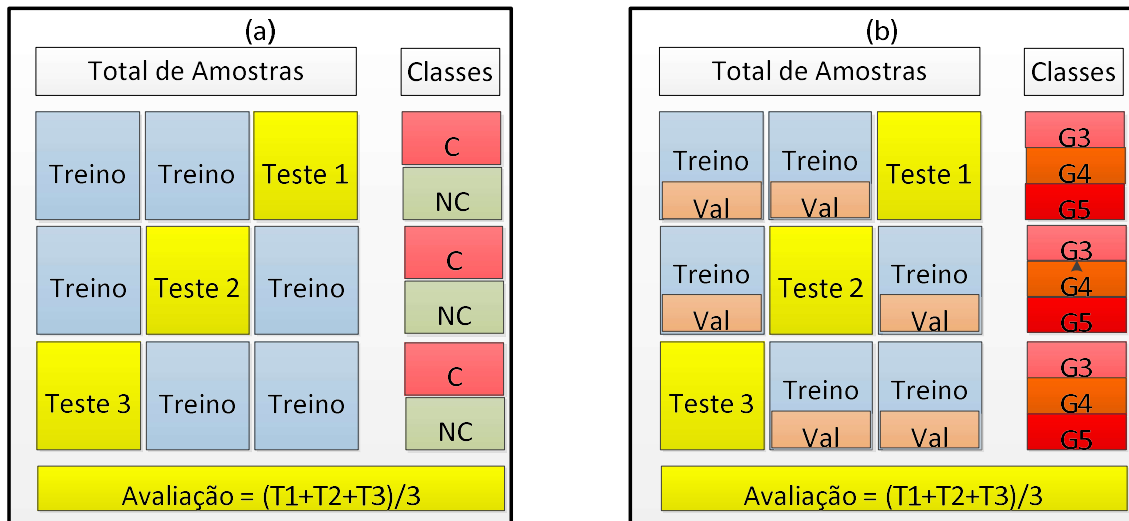
Fonte: Autoria Própria.

**Figura 17:** Fluxograma da topologia 2



## 4 RESULTADOS

Ambas as topologias foram avaliadas usando *k-fold cross validation*, como mostra a figura 20 a seguir. Escolheu-se um *k fold* de 3 vezes devido haver apenas 3 pacientes distintos com regiões de padrão de *Gleason* 3. Assim, para evitar que *patches* do mesmo paciente sejam deslocados para diferentes grupos de teste, treinamento e validação, optou-se por rodar 3 vezes. Para a topologia 1 proposta, os dados correspondentes foram separados em treino e teste para classificar duas classes, câncer (C), e não câncer (NC). Para a topologia 2 proposta, os dados correspondentes foram separados em treinamento, validação e teste para classificar 3 classes, padrões de Gleasons 3, 4, e 5 (G3), (G4), (G5). A opção por utilizar dados de validação na segunda topologia foi para obter melhor entendimento do desempenho da rede para esta aplicação.



Fonte: Autoria Própria.

**Figura 4:** Esquemas das etapas do *k-fold three-way cross-validation*. (a) Divisão das 3 etapas para avaliação da topologia 1 proposta, com dados separados em treino e teste para 2 classes câncer (C), e não câncer (NC). (b) Divisão das 3 etapas para avaliação da topologia 2 proposta, com dados separados em treino, validação e teste para 3 classes: Gleasons 3 (G3), 4 (G4), e 5 (G5).

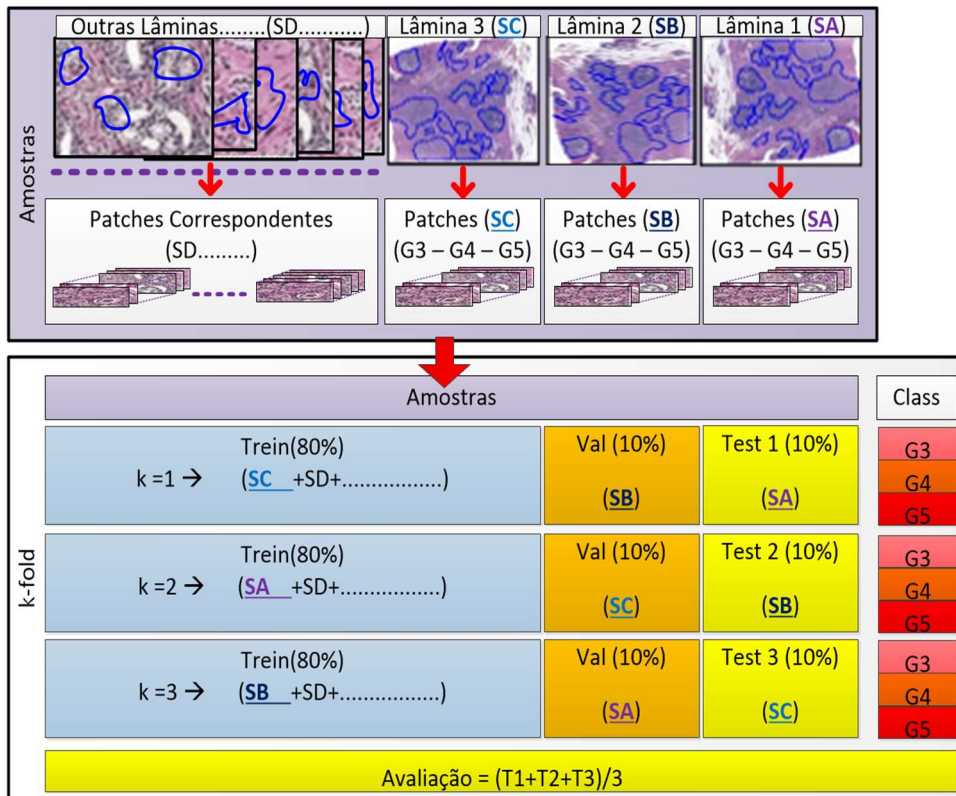
### 4.1 Resultados topologia 1

Para avaliar o modelo CNN da Topologia 1 proposta (figura 17), foi implementado o método *k-fold three-way cross-validation*. Neste, utilizou-se o *dataset* descrito na subseção - *Dataset usado na Topologia 1* -, a cada *k-fold* de avaliação foram construídas 2 divisões de dados, 6918 *patches* destinados para treinamento, e 3458 destinados para

teste. O tempo de treinamento total atingiu cerca de 86.400 segundos. Tal desempenho foi obtido em um Intel Core i7 3,50 GHz, 16 GB de memória RAM, Disco Rígido de 2 TB, patrimônio do Laboratório de Processamento de Sinais e Imagens do Instituto de Ciência e Tecnologia. A taxa de verdadeiro positivo foi calculada considerando os 3 resultados de testes T1, T2 e T3 (figura 17(a)), sendo o resultado desta avaliação a média de 98,3% de verdadeiro positivo. Com quase 100% de verdadeiro positivo, a topologia proposta se demonstrou alto potencial de classificação para amostras de câncer e não câncer.

## 4.2 Resultados topologia 2

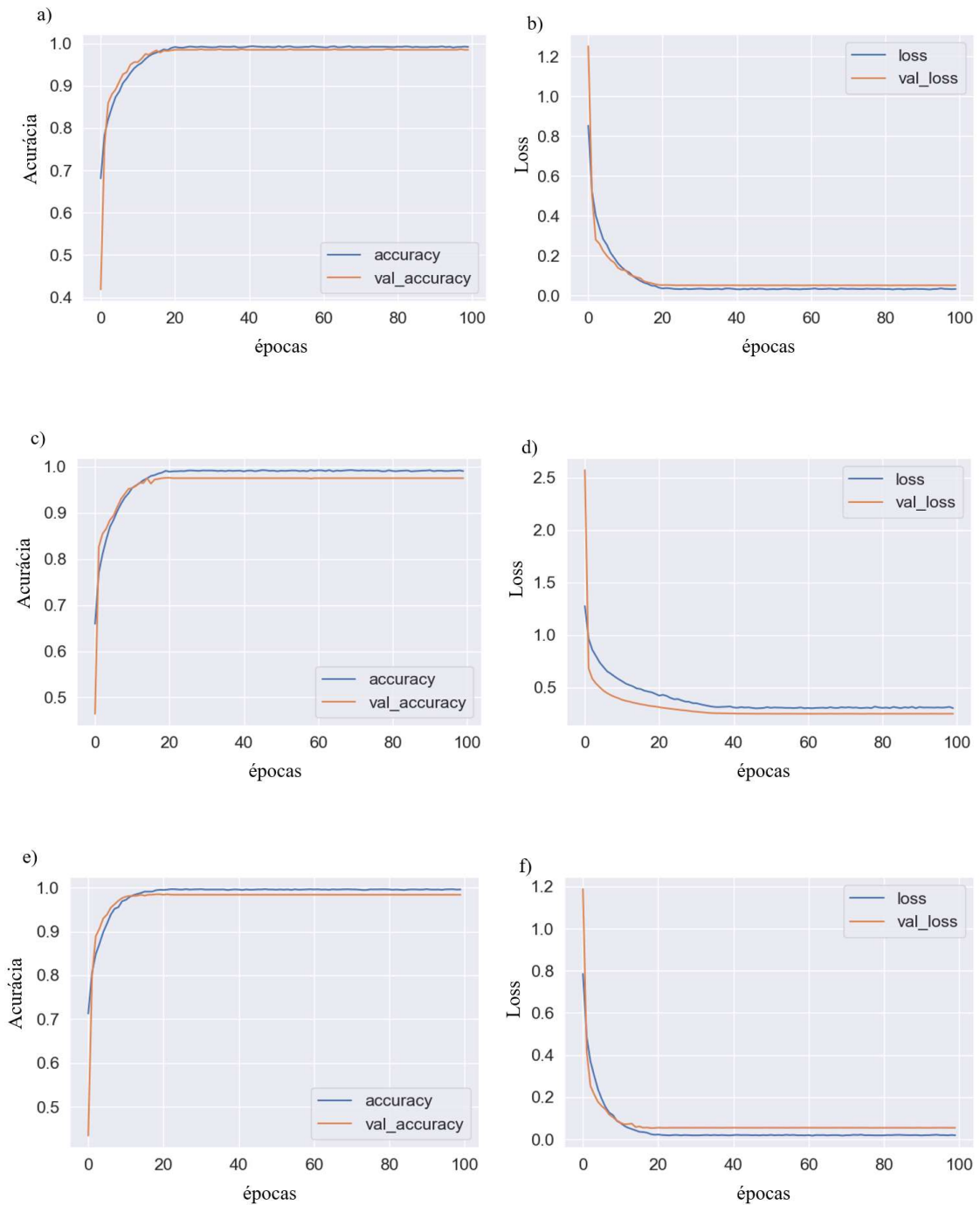
Para avaliação da topologia 2, foram considerados mais parâmetros para melhor conclusão do desempenho da rede. Assim, parâmetros de desempenho foram extraídos para diferentes interpretações durante etapas de treinamento e validação e a de teste posteriormente. Foi implementado o *k-fold three-way cross-validation*, esquematizado na figura 17. O *dataset* foi estruturado em treinamento, validação e teste. Em relação ao *dataset* do colaborador, os patches foram extraídos novamente, para aumentar sua quantidade; assim, foram retirados patches de toda região demarcada, e não só das regiões centrais das marcações como na primeira proposta. Além disso, havia poucas regiões de padrão de Gleason 3 demarcadas; assim, 3 lâminas denominadas SA, SB e SC foram separadas por possuírem número de patches com proporção semelhantes para as classes G3, G4, e G5, e obedecer os 10% destinados a etapa de teste. Para essa separação no *kfold* de 3 vezes, no primeiro *k* a lâmina SC foi destinada ao grupo de treinamento, juntamente com as outras demais lâminas, totalizando 80% dos dados, assim como a lâmina SB foi para o grupo de validação completando os 10% e a lâmina SA destinou-se ao grupo de teste, de também 10%. Já no segundo *k*, a lâmina SB juntamente com as outras lâminas foram para o grupo de treinamento, a lâmina SC foi para o grupo de validação e a lâmina SB para o grupo de teste. Por fim, no terceiro *k*, a lâmina SA foi para o treinamento, a SC para validação e SB para teste. Assim, todas as 3 lâminas de padrão de Gleason 3 foram alternadas em todos os grupos, evitando que patches da mesma lâmina sejam destinados a diferentes grupos no mesmo *k*, evitando incoerências na acurácia de treinamento e validação da rede. A figura 20 demonstra essa separação para cada *kfold*.



Fonte: Autoria Própria.

**Figura 18:** Separação dos grupos de treinamento, validação e teste para cada k-fold, alternando as lâminas SA, SB e SC nos devidos grupos.

Como pode ser visto na figura 21 abaixo, em uma primeira avaliação, as curvas e resultados da etapa de treinamento (curva azul) e validação (curva laranja) da acurácia e *loss* foram computadas para 100 épocas, valor que alcançou a estabilidade. O tempo de processamento de treinamento foi cerca de 1200 segundos em um computador com configurações Intel Core i7 3,50 GHz, Placa de vídeo NVIDIA GeForce GTX 1060, 16 GB de memória RAM, Disco Rígido de 2 TB, disponível no Laboratório de Processamento de Sinais e Imagens do Instituto de Ciência e Tecnologia. O resultando nas acurácias de validação ficaram em torno de 98%. Os 3 *k-folds* apresentaram para acurácia e *loss* curvas de treinamento e validação bem convergentes, a pequena diferença entre elas demonstra um insignificante *overfitting*. Essa insignificância é reforçada pela proporção com a alta acurácia atingida. Por fim, as curvas de treinamento e validação corroboram o potencial da topologia para as avaliações com os dados de teste.



Fonte: Autoria Própria.

**Figura 19:** Curvas de treinamento de acurácia de validação de cada *kfold-cross validation*. (a) Curva de acurácia dos dados de treinamento e validação do primeiro *cross-validation*. (b) Curva de *loss* dos dados de treinamento e validação do primeiro *cross-validation*. (c) Curva de acurácia dos dados de treinamento e validação do segundo *cross validation*. (d) Curva de *loss* dos dados de treinamento e validação do segundo *cross-validation*. (e) Curva de acurácia dos dados de treinamento e validação do terceiro *cross validation*. (f) Curva de *loss* dos dados de treinamento e validação do terceiro *cross-validation*

A cada  $k$ -fold, após treinamento e validação, a avaliação foi realizada no grupo de teste. O primeiro  $k$ , composto por 69 patches de padrão Gleason 3, 258 de padrão *Gleason* 4 e 349 de padrão Gleason 5. Já no segundo  $k$ , temos 94 patches de padrão Gleason 3, 250 de padrão *Gleason* 4 e 326 de padrão Gleason 5. E por fim, no terceiro  $k$ , temos 72 patches de padrão Gleason 3, 269 de padrão *Gleason* 4 e 353 de padrão Gleason 5. O conjunto de imagens de teste nunca foram vistos pela rede, sendo assim apropriado para avaliar capacidade de aprendizado e generalização da topologia treinada para dados de diferentes fontes. Além da matriz de confusão, outros 4 parâmetros foram computados.

$$Acurácia = \frac{(TP+TN)}{n}, \quad (1)$$

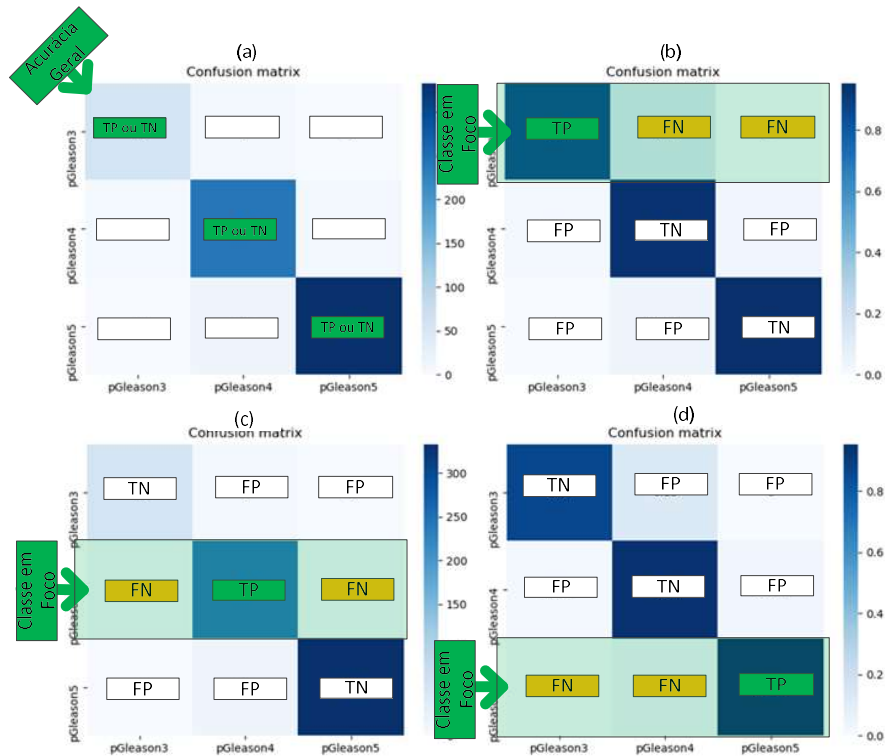
nos quais,  $n$  é o número total de amostras. E, neste contexto cálculo de acurácia para 3 classes,  $(TP + TN)$  podem ser considerados apenas  $TP$  para cada uma das 3 classes. Assim, especificamente, é a somatória dos acertos na diagonal principal, correspondendo ao azul mais escuro das matrizes de confusão (figura 22(a)). Adicionalmente, a partir das matrizes de confusão, para cada um dos 3 Gleasons, os seguintes 3 parâmetros foram calculados:

$$Precisão = \frac{TP}{TP+F}, \quad (2)$$

$$Sensibilidade = \frac{TP}{(TP+ )}, \quad (3)$$

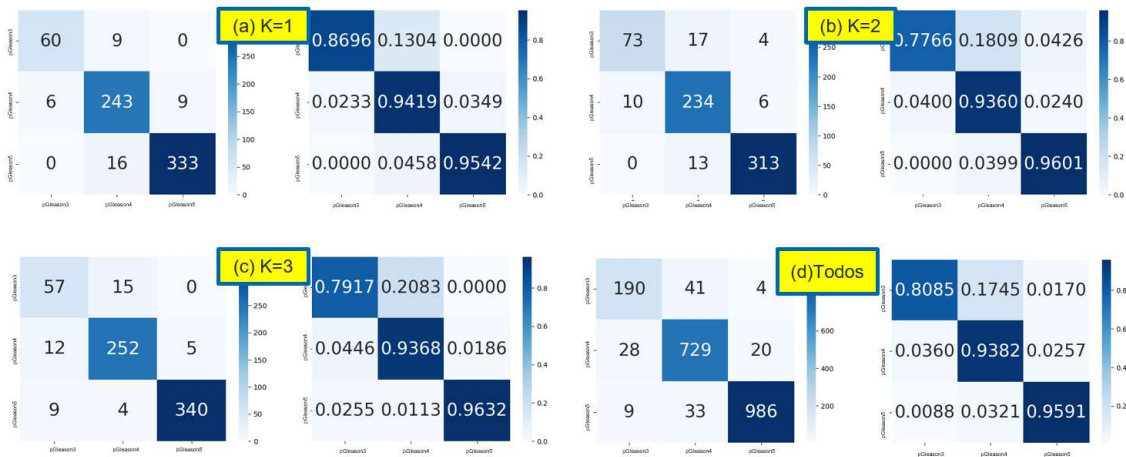
$$Especificidade = \frac{TN}{(FP+TN)}, \quad (4)$$

nos quais,  $TP$  são os valores de intersecção, acertos da classe de padrão de *Gleason* específica sendo foco de avaliação.  $TN$  são os valores de intersecção das outras 2 classes de padrões de Gleasons que não estão como foco de avaliação. Os  $FN$  são os valores da linha, sem intersecção com a classe do padrão de Gleason específico, sendo foco de avaliação. Os  $FP$  são os valores das colunas, os quais encontram-se fora da linha foco de avaliação sem intersecção entre classes correspondentes (figura 22(b), (c), (d)). Por fim, após computar os dados de testes para os 3  $k$ -folds as matrizes de confusão correspondentes foram geradas (figuras 23(a), (b), (c)). Adicionalmente, a matriz de confusão do resultado composto dos 3 grupos de  $k$ -fold podem ser vistos na figura 20(d). Finalmente, ao computar os parâmetros descritos nas equações (1), (2), (3), e (4), nas matrizes de confusão correspondentes é obtido o desempenho da topologia proposta Tabela 1.



Fonte: Autoria Própria.

**Figura 20:** Matriz de confusão para avaliação de 3 classes. (a) Ilustração do cálculo da acurácia, computada somando os valores da diagonal principal, verdadeiros positivo e negativo, e dividido pelo número total de dados. (b), (c), e (d) Ilustração dos verdadeiros positivo e negativo, e dos falsos positivo e negativo, para o cálculo de Precisão, Sensibilidade e Especificidade para cada um dos Gleasons em foco.



Fonte: Autoria Própria

**Figura 21:** Matriz de confusão não normalizada e normalizada dos dados de teste para cada k-fold correspondente. (a), (b), (c) matriz de confusão para k = 1,2,3, respectivamente. (d) matriz de confusão dos dados de teste com o resultado composto dos 3 k-folds.

**Tabela 1:** Precisão, Sensibilidade e Especificidade, e Acurácia para cada k correspondente.

<i>k-fold</i>	<i>classe</i>	<i>Precisão (%)</i>	<i>Sensibilidade (%)</i>	<i>Especificidade (%)</i>	<i>Acurácia (%)</i>	<i>Acurácia Balanceada (%)</i>
<i>k=1</i>	<i>pGleason 3</i>	90,91	86,95	99,01	94,00	92,18
	<i>pGleason 4</i>	90,67	94,18	94,02		
	<i>pGleason 5</i>	97,36	95,41	97,25		
<i>k=2</i>	<i>pGleason 3</i>	87,95	77,65	98,26	95,02	89,09
	<i>pGleason 4</i>	88,63	93,60	92,85		
	<i>pGleason 5</i>	96,90	96,01	97,09		
<i>k=3</i>	<i>pGleason 3</i>	73,07	79,16	96,62	94,00	89,72
	<i>pGleason 4</i>	92,98	93,68	95,52		
	<i>pGleason 5</i>	98,55	96,31	98,53		
<i>média</i>	<i>pGleason 3</i>	83,70	80,85	97,95	94,89	92,18
	<i>pGleason 4</i>	90,78	93,82	94,14		
	<i>pGleason 5</i>	97,62	95,91	97,62		

Fonte: Autoria Própria

Seja individualmente, ou para o resultado médio composto, os resultados apresentados nas matrizes de confusão da figura 20, e Tabela 1 são motivadores. A interpretação dos resultados acima, demonstram o potencial do modelo proposto. Primeiramente e muito importante, a proximidade dos valores de acurácia final entre a etapa de treinamento e validação 98%, com a etapa de teste 94%, demonstram a potencial capacidade de generalização e robustez para *datasets* de diferentes origens. Adicionalmente, os resultados muito próximos para os 3 *ks* e, juntamente com a insignificante diferença com a média geral (Tabela 1), indicam a estável capacidade de aprendizado do modelo. Além disso, a precisão, sensibilidade e especificidade, importantes no contexto médico para quantificar o potencial individual de separação das classes, se mostram alinhados com a alta acurácia. Como pode ser visto na Tabela 1, elas atingem valores tão altos quanto a acurácia para as classes Gleasons 4 e 5, demonstrando uma equivalência na classificação dessas classes. Contudo, como pode ser visto na Tabela

1, elas se mostraram levemente menor para a classe Gleason 3 em relação as outras, aproximadamente 81% e 87%, para precisão e sensibilidade respectivamente. A justificativa é que esta classe era a que possuía menos amostras, tanto para treinamento e validação, quanto para teste, descrito na subseção - *Dataset* usado na Topologia 2 -. Adicionalmente, por possuir menos amostras, cada *patch* classificado errado representava uma maior proporção. O efeito desta diferença de número de amostras também pode ser notado com a menor especificidade das outras duas classes em relação ao Gleason 3, Tabela 1. Como a especificidade quantifica a capacidade do modelo de classificar os “negativos”, e sendo os Gleasons 4 ou 5 focos de classificação, o menor potencial de identificação do Gleason 3, diminui os valores de especificidade das classes 4 e 5. Ao contrário, quando o foco for no próprio Gleason 3, o potencial de identificação dos “negativos”, Gleason 4 e 5, afetarão positivamente a especificidade do Gleason 3.



## 5 DISCUSSÃO E CONCLUSÃO

Uma vez que o câncer de próstata utiliza a interpretação humana para fins de diagnósticos, sua eficiência está diretamente relacionada a subjetividade e expertise do profissional envolvido. Assim, métodos computacionais para reduzir essa subjetividade e auxiliar o patologista na tomada de decisão podem incrementar melhoras nos procedimentos de diagnóstico. Desta forma, propostas que visem este auxílio são importantes a serem exploradas e aperfeiçoadas.

Muitos conhecimentos foram gerados e acumulados ao longo dos anos para tal finalidade. Os mais recentes e relevantes trabalhos que apresentaram boas propostas e resultados promissores relacionados à próstata podem ser vistos na tabela 2. Em Li, et al., 2020, *et al.*, 2020 (Li, et al., 2020), utilizaram *microarrays* de tecido prostático em uma rede CNN para classificar nos 3 padrões de Gleason, atingindo um kappa de 0,77. Já em Bulten, (Bulten, et al., 2020), houve uma segmentação das regiões de interesse utilizando uma Unet e *cycleGANs*, chegando a 0,91 para a classificação dos *patches*. Em Pantanowitz, Quiroga-Garza, *et al.*, 2020, (Pantanowitz, et al., 2020), usaram Inception V1, Inception V3 e Resnet para a classificação de Gleason 5, Gleason 7+, invasão perineural, câncer e porcentagem de câncer. Assim, alcançaram uma sensibilidade de 96% e especificidade de 89% no conjunto de validação. Yechan, Paik, *et al.*, 2021 (Yechan, Paik, Shin, Kwak, & Chang, 2021), empregaram uma CNN de múltiplas instâncias para a classificação das regiões de interesse. Atingiram a precisão de 77,5% com um kappa score de 0,65. Por fim, em Tolkach, Dohmgörge, *et al.*, 2020 (Tolkach, Dohmgörge, Toma, & Kristiansen, 2020), classificaram *patches* de biopsia de câncer de próstata, por meio de uma NASNetLarge. Conseguindo a acurácia geral do modelo de 98%. Apesar dos excelentes trabalhos citados, ainda há espaço para aperfeiçoamento e otimizações. O método proposto utiliza *patches* com zoom suficiente para encontrar estruturas fundamentais para a classificação de amostras entre câncer e não câncer, assim como para diferenciar os padrões de Gleasons 4, 5, 6. No estudo de Tolkach, Dohmgörge, *et al.*, CNN é avaliada na aplicação de histopatologia de próstata. O projeto é dividido em duas etapas, a primeira para separar câncer e não câncer, e a segunda etapa é também dedicada a separar Gleasons 3, 4, 5. O modelo proposto conseguiu acurácia geral de 98% na classificação de *patches* com tecido benigno e tecido tumoral (Tolkach, Dohmgörge, Toma, & Kristiansen, 2020).

**Tabela 2:** Tabela comparativa entre os artigos citados no estado da arte

<i>Artigo</i>	<i>Classe</i>	<i>Precisão (%)</i>	<i>Sensibilidade (%)</i>	<i>Especificidade (%)</i>	<i>Kappa score</i>	<i>Acurácia (%)</i>
<i>Li, et al., 2020</i>	<i>n/a</i>				0,77	
<i>Bulten, et al., 2020</i>	<i>n/a</i>				0,91	
<i>Pantanowitz, et al., 2020</i>	<i>n/a</i>		96,00	89,00		
<i>Yechan, et al., 2021</i>	<i>n/a</i>	77,50			0,65	
<i>Tolkach, et al., 2020</i>	<i>n/a</i>					98,00
<i>Topologia Proposta 1</i>	<i>n/a</i>					98,30
<i>Topologia Proposta 2</i>	<i>pGleason3</i>	83,70	80,85	97,95		94,90
	<i>pGleason4</i>	90,78	93,82	94,14		
	<i>pGleason5</i>	97,62	95,91	97,62		

Fonte: Autoria Própria

Apesar dessas soluções propostas apresentarem um alto potencial e abranger o estudo proposto, limitações podem ser encontradas. Na maior parte dos casos, essas limitações estão relacionadas ao desempenho, tecnologia aplicada, informações da ação clínica, e características e classificação do conjunto de dados gerais, ou relacionada a classes específicas.

Sendo assim, o trabalho proposto visou contribuir com topologias inseridas no contexto disponível para complementar a compressão do potencial de classificação destas crescentes tecnologias. O método apresentou bom desempenho nas duas propostas, classificação de *patches* tumorais assim como para diferenciar os Gleasons 3, 4 e 5. Os valores atingidos estão alinhados com o da literatura, corroborando boa acurácia e robustez para os dados utilizados. Por fim, nossos esforços se concentraram em encontrar os melhores parâmetros em relação a quantidade de filtros, tamanhos de *kernels*, sequência de camadas, quantidade de *hidden layer*, melhor função de ativação e otimizador visando a acurácia de treinamento, validação e teste dos dados. Além disso, foram utilizados 2 diferentes bancos de dados, esse aumento de variabilidade torna o processo mais desafiador, contribuindo para a construção de um modelo mais robusto, melhorando aprendizado e generalização do modelo treinado.

No entanto, há a limitação de não envolver a classe de tecidos saudáveis na segunda topologia, não realizar a classificação da lâmina inteira e há a necessidade de aumentar o número de amostras. Uma outra possível limitação, consequência do número de amostras, poderia ser atribuída à menor precisão e sensibilidade do padrão de Gleason 3. Como mencionado na seção anterior, ela é justificada pelo menor número de amostras adquiridas dessa classe, além da maior variação na coloração das imagens que, em sua maioria foram providas pelo PANDAS *Challenge*. Todavia, ainda assim, como pode ser visto na figura 20(d), ainda é insignificante os equívocos entre classes em termos numéricos e percentuais. Mais ainda, a imensa maioria dos erros de interpretação, é entre classes vizinhas. Um exemplo é que apenas 1 *patch* de padrão de Gleason 3 foi interpretado como padrão de Gleason 5, demonstrando a insignificância de erros extremos. Por fim, considerando que os *patches* representam porções pequenas de uma área grande, esses poucos equívocos são insignificantes para a classificação de regiões da lâmina, o que não afeta significativamente a interpretação da área observada.

Trabalhos futuros consistem em fazer investigações paralelas com diferentes topologias, melhorar conjunto de dados com colaboradores adicionais e métodos de *augmentation*, graduar a lâmina inteira nos padrões de Gleason, realizar a quantificação de cada padrão e assim fornecer o *Gleason score* final da lâmina digital. Além disso, é importante avaliar também a variabilidade entre os especialistas patologistas, assim, é essencial 3 a 5 patologistas para realizar a marcação dos dados.

As contribuições específicas envolvem as topologias propostas avaliadas, identificação da sequência, quantidade e estrutura das camadas; assim como, melhores configurações dos parâmetros do modelo para a classificação 1º entre patches saudáveis e cancerígenos, e em 2º a graduação nos padrões de Gleason. Consequentemente, o método proposto mostrou-se com alto potencial de apoio ao patologista durante o processo de diagnóstico, fornecendo informações adicionais e extras para a decisão final e assim, afetando na intervenção e o tratamento dos pacientes.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

Andriole GL, Crawford ED, Grubb RL, III, Buys SS, Chia D, Church TR, et al. Prostate Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: Mortality Results after 13 Years of Follow-up. *J Natl Cancer Inst.* 2012; 104(2):125-132. Disponível em: <https://doi.org/10.1093/jnci/djr500>

Bancroft EK, Page EC, Castro E, Lilja H, Vickers A, Sjoberg D, et al. Targeted Prostate Cancer Screening in BRCA1 and BRCA2 Mutation Carriers: Results from the Initial Screening Round of the IMPACT Study. *European Urology.* 2014; 66(3):489-499. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24484606>

Bell KJ, Del Mar C, Wright G, Dickinson J, & Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *Int. J.* 2015; 137(7): 1749-1757. Disponível em: <https://doi.org/10.1002/ijc.29538>

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, & Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6):394-424.

Bulten W, Pinckaers H, Boven H, Vink R, de Bel T, Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *The Lancet Oncology.* 2020; 21(2):233-241.

Castro E, Goh C, Leongamornlert D, Saunders Ed, Tymrakiewicz M, Dadaev T, et al. Effect of BRCA Mutations on Metastatic Relapse and Cause-specific Survival After Radical Treatment for Localised Prostate Cancer. *Eur Urol.* 2015; 68(2):186-193. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25454609>

Faria EF, Carvalhal GF, Vieira RA, Silva TB, Mauad EC, & Carvalho AL. Program for Prostate Cancer Screening Using a Mobile Unit: Results From Brazil. *Urology.* 2010; 76(5):1052-1057. Disponível em: <https://doi.org/10.1016/j.urology.2010.02.044>

Epstein, J. Prostate cancer grading: a decade after the 2005 modified system. *Mod Pathol.* 2018; 31:47–63. Disponível em: <https://doi.org/10.1038/modpathol.2017.133>

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, & Grading Committee. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol*. 2016; 40(2):244–252. Disponível em: <https://doi.org/10.1097/PAS.0000000000000530>

Esposito K, Chiodini P, Capuano A, Bellastella G, Maiorino MI, Parretta E, et al. Effect of metabolic syndrome and its components on prostate cancer risk: Meta-analysis. *J Endocrinol Invest*. 2013; 36:132–139. Disponível em: <https://doi.org/10.1007/BF03346748>

Gleason DF & Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *The Journal of urology*. 1974; 111(1): 58–64. Disponível em: <https://www.auajournals.org/doi/abs/10.1016/S0022-5347%2817%2959889-4>

Glorot X, Bordes A, & Bengio Y. Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, in PMLR. 2011; 15:315-323. Disponível em: <http://proceedings.mlr.press/v15/glorot11a.html>

Haenlein M & Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*. 2019; 61(4):5-14. Disponível em: <https://doi.org/10.1177/0008125619864925>

Hemminki K. Familial risk and familial survival in prostate cancer. *World J Urology*. 2012; 30(2):143–148. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22116601/>

Instituto Nacional de Câncer; 2020. Fonte: [inca.gov](http://inca.gov) Disponível em: <https://www.inca.gov.br/campanhas/cancer-de-prostata/2020/saude-do-homem>

Kasivisvanathan V, Emberton M, & Moore CM. MRI-Targeted Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018; 379(6):589-590. Disponível em: <https://www.nejm.org/doi/10.1056/NEJMc1807507>

Kudo MS, de Souza VM, Amaral Gd, Melo PA, Estivallet CL, Santos ER, et al. The Potential of Convolutional Neural Network Diagnosing Prostate Cancer. *Res Biomed Eng*. 2021; 37:1-7.

LeCun Y, Bengio Y, & Hinton G. Deep learning. *Nature*. 2015; 521:436–444. Disponible em: <https://www.nature.com/articles/nature14539>

Li Y, Huang M, Zhang Y, Chen J, Xu H, Wang G, & Feng W. Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer. *IEEE*. 2020; 8:117714 - 117725. Disponible em: <https://ieeexplore.ieee.org/document/9126781>

Lynch HT, Kosoko-Lasaki O, Leslie SW, Rendell M, Shaw T, Snyder C, et al. Screening for familial and hereditary prostate can. Screening for familial and hereditary prostate cancer. *International Journal of Cancer*. 2016; 138(11):2579 - 91. Disponible em: <https://pubmed.ncbi.nlm.nih.gov/26638190>

Marks, S. *Prostate & Cancer: A Family Guide to Diagnosis, Treatment, & Survival*. 4th Ed.). Cambridge, MA: Da Capo Lifelong; 2009.

Meliti A, Sadimin E, Diolombi M, Khani F, & Epstein JI. Accuracy of Grading Gleason Score 7 Prostatic Adenocarcinoma on Needle Biopsy: Influence of Percent Pattern 4 and Other Histological Factors. *The Prostate*, 2017; 77(6):681–685. Disponible em: <https://doi.org/10.1002/pros.23314>

Nagpal K, Foote D, Liu Y, Chen P HC, Wulczyn E, Tan F, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digital Medicine* 2. 2019; 2(1). Disponible em: <https://www.nature.com/articles/s41746-019-0112-2>

Naji L, Randhawa H, Sohani Z, Dennis B, Lautenbach D, Kavanagh O, et al. Digital Rectal Examination for Prostate Cancer Screening in Primary Care: A Systematic Review and Meta-Analysis. *Annals of Family Medicine*. 2018; 16(2):149–154. Disponible em: [10.1016/S0140-6736\(14\)60525-0](https://doi.org/10.1016/S0140-6736(14)60525-0)

Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol*, 2016; 50(6): 420–424. Disponible em: <https://pubmed.ncbi.nlm.nih.gov/27416104/>

Pantanowitz L, Quiroga-Garza GM, Bien L, Heled R, Laifenfeld D, Linhart C, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of

core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digital Health*. 2020; 2(8):407-416. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33328045>

Quinn M, Babb P. Patterns and trends in prostate cancer incidence, survival, prevalence and mortality. Part I: international comparisons. *BJU International*. 2002; 90:162-173. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/12081758>

Randazzo M, Müller A, Carlsson S, Eberli D, Huber A, Grobholz R, et al. A positive family history as a risk factor for prostate cancer in a population-based study with organised prostate-specific antigen screening: results of the Swiss European Randomised Study of Screening for Prostate Cancer (ERSPC, Aarau). *BJU International*. 2016; 117(4): 576-83. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26332304>

Rosebrock A. Table of Contents: Deep Learning for Computer Vision with Python. Pyimagesearch. 1th ed; 2017

Rouvière O, Puech P, Renard-Penn R, Claudon M, Roy C, Mège-Lechevallier F, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naïve patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. 2018; 20(1):100-109. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30470502>

Santos Md. Estimativa 2018: Incidência de Câncer no Brasil. *Revista Brasileira de Cancerologia*. 2018; 64(1):119-120. Disponível em: <https://rbc.inca.gov.br/revista/index.php/revista/article/view/115/55>

Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Prostate-Cancer Mortality at 11 Years of Follow-up. *N Engl J Med*. 2012; 366(11):981-990. Disponível em: <https://www.nejm.org/doi/full/10.1056/nejmoa1113135>

Tolkach Y, Dohmgörgen T, Toma M, Kristiansen G. (2020). High-accuracy prostate cancer pathology using. *Nat Mach Intell*. 2020; 2:411-418. Disponível em: <https://www.nature.com/articles/s42256-020-0200-7>

Vidal AC, Howard LE, Moreira DM, Castro-Santamaria R, Andriole Jr GL, & Freedland SJ. Obesity Increases the Risk for High-Grade Prostate Cancer: Results from the REDUCE Study. *Cancer Epidemiol Biomarkers Prev*. 2014; 23(12):2936-2942. Disponível em: <https://doi.org/10.1158/1055-9965.EPI-14-0795globocan>

Yeohan M, Paik I, Shin S, Kwak TY, Chang H. Yet Another Automated Gleason Grading System (YAAGGS) por deep learning fracamente supervisionado. npj Digital Medicine. 2021; 4(1):99. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34127777>

Zhang A, Lipton ZC, Li M, Smola AJ. Dive into Deep Learning. 2021