

Universidade Federal do Pará
Instituto de Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica

Redes Neurais Convolucionais Para Auxiliar no Diagnóstico de Exames Preventivo de Colo de Útero

Edriane do Socorro Silva Costa

DM 21/2024

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2024

Universidade Federal do Pará
Instituto de Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica

Edriane do Socorro Silva Costa

Redes Neurais Convolucionais Para Auxiliar no Diagnóstico de Exames Preventivo de Colo de Útero

Dissertação de Mestrado submetida à avaliação da Banca Examinadora aprovada pelo colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará e julgada adequada para a obtenção do Grau de Mestre em Engenharia Elétrica na área de Computação Aplicada.

DM 21/2024

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

C837r Costa, Edriane do Socorro Silva.
 Redes Neurais Convolucionais Para Auxiliar no Diagnóstico de
 Exames Preventivo de Colo de Útero / Edriane do Socorro Silva
 Costa. — 2024.
 78 f. : il. color.

 Orientador(a): Prof. Dr. Carlos Renato Lisboa Francês
 Dissertação (Mestrado) - Universidade Federal do Pará,
 Instituto de Tecnologia, Programa de Pós-Graduação em
 Engenharia Elétrica, Belém, 2024.

 1. Visão computacional. 2. CNN. 3. Papanicolau. 4.
 Classificação Automatizada. I. Título.

CDD 006.3

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**Redes Neurais Convolucionais Para Auxiliar no
Diagnóstico de Exames Preventivo de Colo de Útero**

Autor: Edriane do Socorro Silva Costa

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA
APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM EN-
GENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO
GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO
APLICADA.

APROVADA EM: 18/09/2024

Banca Examinadora:

Carlos Renato Lisboa Francês

Orientador

Prof.^a Dra. Jasmine Priscyla Leite Araújo

(Avaliadora Interna – PPGEE/ITEC/UFGA)

Prof.^a Dra. Evelin Helena Silva Cardoso

(Avaliadora Externa ao Programa – CAMPUS
CASTANHAL/UFGA)

Visto:

Prof. Dr. Diego Lisboa Cardoso

(Coordenador do PPGEE/ITEC/UFGA)

Dedico este trabalho aos meus queridos avós maternos, Antônia Silva e José Silva (in memoriam), cujo legado de amor, sabedoria e dedicação ao estudo moldou meu caminho. Suas inspiradoras lições e apoio foram fundamentais para que eu pudesse alcançar grandes conquistas e perseguir meus objetivos com determinação. Que suas memórias continuem a iluminar meu percurso, sempre guiando-me rumo ao sucesso.

Agradecimentos

A Deus, pelo dom da vida.

À dedicada orientação do Professor Renato Francês, expresso minha sincera gratidão. Suas orientações e discussões foram essenciais para o desenvolvimento deste trabalho.

À Professora Jasmine Araújo, agradeço calorosamente por suas contribuições valiosas e apoio durante minha trajetória na pós-graduação.

À Professora Evelin Helena Silva Cardoso, expresso minha profunda gratidão pelo incentivo, apoio e dedicação oferecidos ao longo dos últimos dois anos durante minha jornada de pós-graduação. Seu comprometimento e orientação foram fundamentais para meu crescimento e aprendizado.

Aos estimados colegas do Laboratório de Planejamento de Redes de Alto Desempenho – LPRAD, especialmente ao Romário Silva, manifesto minha gratidão pelas contribuições substanciais que enriqueceram este trabalho. Também aos meus queridos amigos de vida que compreenderam minhas ausências e sempre estiveram presentes, agradeço de coração.

À minha mãe, Ana Cláudia, expresso minha profunda gratidão por seu constante apoio, por me lembrar de meus sonhos e por me amar incondicionalmente. Seus conselhos sábios e sacrifícios são a essência do meu sucesso.

À minha irmã, Edilane Costa, meu reconhecimento a todas as ocasiões acadêmicas e pessoais que esteve comigo manifestando seu apoio amor e companheirismo.

Ao meu companheiro de vida, Ítalo Melo, expresso minha gratulação por acreditar em mim nos momentos mais desafiadores e me acompanhar em todas as escolhas com resiliência e carinho.

Aos meus tios Edielson, Edinaldo, Elba e Cláudia por todo apoio durante este percurso e por terem acreditado em minha capacidade acadêmica.

Aos professores da graduação, Kátia Liége, Carla Tavares, Valdelírio Silva e Fábio Nogueira expresso minha sincera gratidão pelo incentivo constante e pelo exemplo inspirador que cada um de vocês representou. Suas orientações e dedicação ao longo do curso despertaram em mim a vontade de continuar buscando novos desafios e me motivaram a seguir para a pós-graduação. Muito obrigada por terem sido parte essencial desse caminho.

Por fim, agradeço a todos que, de uma forma ou de outra, contribuíram para o desenvolvimento deste trabalho. Suas influências foram fundamentais para alcançar este marco em minha jornada acadêmica.

“Se cheguei até aqui foi porque me apoiei no ombro dos gigantes. ”

Isaac Newton

Resumo

O exame preventivo de colo de útero é um método de triagem amplamente utilizado para detectar câncer cervical e lesões pré-cancerosas. A classificação automatizada dos resultados pode auxiliar os profissionais de saúde a identificar com precisão padrões de citologia anormais, aumentando a acurácia e a consistência na detecção de anomalias. Além disso, a sistematização dessa solução pode reduzir o tempo de análise e os custos associados, possibilitando a disponibilização de um pré-diagnóstico imediato, especialmente em áreas remotas. Essa abordagem também possui potencial para integração em sistemas de saúde pública, contribuindo para um atendimento mais eficiente e acessível. Assim, este estudo propõe a aplicação dos modelos pré-treinados de rede neurais convolucionais VGG16 e VGG19 para classificação das imagens resultantes da técnica de citologia em base líquida, fazendo um comparativo de desempenho entre a classificação de 4 classes ou de 2 classes com dados balanceados e desbalanceados. Testou-se várias arquiteturas e como resultados foram obtidas acurácias de até 98% obtendo também boas métricas de classificação, mostrando-se como potencial solução para auxílio de profissionais da saúde em uma classificação mais assertiva desses resultados.

Palavras-chaves: Visão computacional. CNN. Papanicolau. Classificação Automatizada.

Abstract

The cervical screening exam is a widely used method to detect cervical cancer and precancerous lesions. Automated classification of the results can assist healthcare professionals in accurately identifying abnormal cytology patterns, increasing accuracy and consistency in detecting anomalies. Furthermore, systematizing this solution can reduce analysis time and associated costs, enabling the provision of an immediate pre-diagnosis, especially in remote areas. This approach also has the potential for integration into public health systems, contributing to more efficient and accessible care. Therefore, this study proposes the application of pre-trained convolutional neural network models VGG16 and VGG19 for classifying images resulting from the liquid-based cytology technique, comparing the performance of 4-class versus 2-class classification with balanced and unbalanced data. Several architectures were tested, and accuracies of up to 98% were achieved, along with good classification metrics, showing potential as a solution to assist healthcare professionals in more assertive classification of these results.

Keywords: Computer Vision. CNN. Pap Smear. Automated Classification.

Lista de Figuras

Figura 1 – Comportamento Celular NILM	8
Figura 2 – Comportamento Celular LSIL	9
Figura 3 – Comportamento Celular HSIL	10
Figura 4 – Comportamento Celular Carcinoma	11
Figura 5 – Estrutura de Sistemas de Classificação baseados em ML	12
Figura 6 – Técnicas e Aplicações Baseadas em ML	13
Figura 7 – Estrutura de Sistemas de Classificação baseados em DL	14
Figura 8 – Relação entre os conceitos de Inteligência Artificial, Aprendizado de Máquina e Aprendizagem Profunda	15
Figura 9 – Arquitetura de uma Rede Neural Artificial Simples	18
Figura 10 – Arquitetura de uma Rede Neural Convolucional	19
Figura 11 – Exemplo de Aplicação de uma Camada de Convolução	20
Figura 12 – Exemplo de Aplicação de uma Camada de <i>Pooling</i>	21
Figura 13 – Processo de Transferência de Aprendizagem	23
Figura 14 – Desempenho de Redes com Processo de Transferência de Aprendizagem	24
Figura 15 – Exemplos de Transformações Aplicadas em imagens Histopatológicas .	24
Figura 16 – Estrutura da CNN VGG16	26
Figura 17 – Estrutura da CNN VGG19	27
Figura 18 – Exemplo de matriz de confusão para classificadores binários	28
Figura 19 – Rede Bibliométrica do Levantamento de Literatura	32
Figura 20 – Fluxograma da Metodologia de Análise de Resultados do Exame Papa- nicolau Utilizando CNNs	39
Figura 21 – Representação Visual das Classes de Imagens no Dataset	40
Figura 22 – Relação de Topologias testadas para a CNN VGG16	45
Figura 23 – Relação de Topologias testadas para a CNN VGG19	45
Figura 24 – Resultado para classificação de 4 Classes para Dados Desbalanceados .	52
Figura 25 – Resultado para classificação de 2 Classes para Dados Desbalanceados .	53
Figura 26 – Resultado para classificação de 4 Classes para Dados Balanceados . . .	53
Figura 27 – Resultado para classificação de 2 Classes para Dados Balanceados . . .	54

Lista de Tabelas

Tabela 1 – Resumo dos Trabalhos Correlatos	36
Tabela 2 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 4 Classes Desbalanceadas	41
Tabela 3 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 2 Classes Desbalanceadas	42
Tabela 4 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 4 Classes após Data Augmentation	43
Tabela 5 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 2 Classes após Data Augmentation	43
Tabela 6 – Resultados de desempenho para a VGG16 com 4 Classes Desbalanceadas	47
Tabela 7 – Resultados de desempenho para a VGG19 com 4 Classes Desbalanceadas	48
Tabela 8 – Resultados de desempenho para a VGG16 com 2 Classes Desbalanceadas	48
Tabela 9 – Resultados de desempenho para a VGG19 com 2 Classes Desbalanceadas	49
Tabela 10 – Resultados de desempenho para a VGG16 com 4 Classes Balanceadas .	49
Tabela 11 – Resultados de desempenho para a VGG19 com 4 Classes Balanceadas .	50
Tabela 12 – Resultados de desempenho para a VGG16 com 2 Classes Balanceadas .	50
Tabela 13 – Resultados de desempenho para a VGG19 com 2 Classes Balanceadas .	51
Tabela 14 – Resultados de Desempenho Sintetizados	55

Lista de abreviaturas e siglas

INCA	Instituto Nacional de Câncer
IA	Inteligência Artificial
CNN	Convolutional Neural Network
HPV	Papiloma Vírus Humano
NILM	Negativo para Lesão Intraepitelial ou Malignidade
LSIL	Lesão Intraepitelial de Baixo Grau
HSIL	Lesão Intraepitelial de Alto Grau
DNA	Ácido Desoxirribonucleico
ML	Machine Learning
DL	Deep Learning
PCA	Análise de Componentes Principais
RNN	Rede Neural Recorrente
SSD	Single Shot MultiBox Detection
ReLU	Rectified Linear Activation
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
GPU	Unidade de Processamento Gráfico
ResNet	Redes Residuais
VGG	Visual Geometry Group
YOLO	You Only Look Once

Sumário

1	INTRODUÇÃO	1
1.1	Contextualização	1
1.2	Motivação e Desafios	2
1.3	Objetivos	4
1.4	Reconhecimento da Pesquisa	4
1.5	Organização do Trabalho	5
2	FUNDAMENTAÇÃO TEÓRICA	6
2.1	Considerações Iniciais	6
2.2	Histórico do Papanicolau	6
2.3	Classes de Diagnóstico no Papanicolau	7
2.3.1	Classe NILM	7
2.3.2	Classe LSIL	8
2.3.3	Classe HSIL	9
2.3.4	Classe Carcinoma	10
2.4	Inteligência Artificial	11
2.4.1	Aprendizado de Máquina	12
2.4.2	Aprendizagem Profunda	14
2.4.3	Visão Computacional	15
2.4.3.1	Classificação em Visão Computacional	16
2.5	Inteligência Artificial na Medicina	16
2.6	Redes Neurais	17
2.6.1	Redes Neurais Convolucionais	18
2.6.1.1	Arquitetura de uma Rede Neural Convolucional	19
2.6.1.2	Camada Convolucional	20
2.6.1.3	Camada de Pooling	21
2.6.1.4	Camada Totalmente Conectada	22
2.7	Transferência de Aprendizado	22
2.8	<i>Data Augmentation</i>	24
2.9	Rede Neural Convolucional VGG 16	25
2.10	Rede Neural Convolucional VGG 19	26
2.11	Métricas de Avaliação	27
2.11.1	Matriz de Confusão	27
2.11.2	Acurácia	28
2.11.3	Precisão	28

2.11.4	Sensibilidade ou <i>Recall</i>	29
2.11.5	<i>F1-score</i>	29
2.12	Considerações Finais	30
3	TRABALHOS CORRELATOS	31
3.1	Considerações Iniciais	31
3.2	Levantamento do Estado da Arte	31
3.3	Considerações Finais	37
4	MATERIAIS E MÉTODOS	38
4.1	Considerações Iniciais	38
4.2	Conjunto de dados	39
4.2.1	Primeiro Momento- Classes Desbalanceadas	41
4.2.2	Segundo Momento- Classes Balanceadas	42
4.3	Treinamento dos modelos	43
4.3.1	Classificadores com CNNs	44
4.4	Avaliação de Perfomance	45
4.5	Considerações Finais	46
5	RESULTADOS E DISCUSSÃO	47
5.1	Resultados para a Classificação com 4 Classes Desbalanceadas	47
5.2	Resultados para a Classificação com 2 Classes Desbalanceadas	48
5.3	Resultados para a Classificação com 4 Classes Balanceadas . . .	49
5.4	Resultados para a Classificação com 2 Classes Balanceadas . . .	50
5.5	Resultados Finais	51
6	CONCLUSÕES	56
6.1	Limitações e Dificuldades Encontradas	57
6.2	Trabalhos Futuros	57
6.3	Trabalhos Publicados	58
	REFERÊNCIAS	59

1 Introdução

Este capítulo serve como ponto de partida para o desenvolvimento do estudo, fornece uma introdução inicial ao tema, que inclui a contextualização, descrição do problema em questão, objetivos gerais e específicos delineados e, por fim, aborda a estruturação da dissertação.

1.1 Contextualização

O carcinoma de colo de útero também conhecido como câncer cervical, é uma neoplasia maligna que se desenvolve nas células do colo do útero. Esta condição é uma das principais causas de morte por câncer entre mulheres, especialmente em regiões em desenvolvimento onde o acesso aos serviços de rastreamento e tratamento são limitados (HULL et al., 2020). Apresenta-se como uma das patologias que mais matam mulheres em todo o mundo, ocupando no Brasil o terceiro lugar em mortalidades advindas do câncer, representando 6,1% do total de óbitos, de acordo com dados do Instituto Nacional de Câncer (INCA) ((INCA), 2023).

Quando se trata das regiões com maior mortalidade por esta causa, algumas áreas do país se destacam. Por exemplo, regiões Norte e Nordeste geralmente têm taxas mais altas de mortalidade por câncer cervical devido a uma combinação de fatores, incluindo acesso limitado a serviços de saúde e desafios socioeconômicos (SOUSA et al., 2022). Estados como Maranhão, Piauí, Pará e Bahia costumam apresentar algumas das taxas mais elevadas de mortalidade por câncer cervical no Brasil (HULL et al., 2020).

Diante disso, enfrentar e controlar essa patologia representa um desafio complexo e necessário. Detectar o carcinoma do colo do útero precocemente é crucial para um tratamento eficaz e melhores resultados para os pacientes (BANERJEE et al., 2022). Uma das ferramentas fundamentais nesse processo é o teste de Papanicolau ou preventivo, que envolve a coleta de células do colo do útero e sua posterior análise laboratorial para identificar possíveis anormalidades celulares que podem indicar a presença de câncer cervical (KALBHOR et al., 2023).

Embora essencial, a disponibilidade limitada de especialistas e a falta de infraestrutura adequada, pode, muitas vezes, resultar em atrasos nos resultados e na tomada de decisões clínicas, prejudicando ainda mais o processo de diagnóstico e tratamento.

Além disso, a abordagem atual de triagem do câncer cervical está sujeita à geração de resultados falso-positivos devido a erros individuais (MACIOS; NOWAKOWSKI, 2022). Isso ocorre porque, como se trata de uma análise laboratorial, ela depende diretamente da

infraestrutura e disponibilidade de patologistas para sua execução. Esses falsos positivos podem levar a uma preocupação adicional e, em alguns casos, a procedimentos invasivos desnecessários, impactando negativamente tanto o paciente quanto o sistema de saúde (KALBHOR et al., 2023).

Nesse contexto, novas abordagens tecnológicas, como o uso de redes neurais e visão computacional, estão sendo exploradas como alternativas promissoras para aumentar a precisão e a eficiência da triagem do câncer cervical. Essas tecnologias podem processar grandes volumes de dados de maneira rápida e automatizada, identificando padrões e anormalidades que podem escapar da detecção humana, reduzindo assim o risco de resultados falso-positivos e aprimorando a qualidade da triagem (ALLAHYARI; MOODI; TAHERGORABI, 2022).

Tais avanços tecnológicos não apenas podem melhorar a precisão do diagnóstico, mas também têm o potencial de democratizar o acesso aos serviços de saúde, especialmente em áreas remotas ou com escassez de recursos. A implementação dessas ferramentas pode não apenas reduzir o ônus sobre os sistemas de saúde, mas também salvar vidas ao permitir diagnósticos mais precisos e tratamentos oportunos para o câncer cervical.

1.2 Motivação e Desafios

A utilização de Redes Neurais Convolucionais (CNNs) para o diagnóstico de exames preventivos oferece diversas motivações significativas. Isso porque essa aplicação pode aprimorar a detecção precoce do câncer cervical, resultando em diagnósticos mais rápidos e precisos, o que é determinante para o tratamento eficaz e a sobrevivência dos pacientes (DESIANI et al., 2021). Além disso, a automatização do diagnóstico reduz a possibilidade de erros humanos, proporcionando uma análise mais confiável dos exames, uma vez que elimina a variabilidade inerente às avaliações manuais (KIPTOO; NDERU; MUTANU, 2020).

Outra motivação importante é melhorar a eficiência operacional e a ampliação do acesso e a escalabilidade dos diagnósticos. Sistemas automatizados baseados em CNNs podem ser implementados em regiões carentes de especialistas médicos, democratizando o acesso ao diagnóstico preventivo. Uma vez que a tecnologia pode aumentar a eficiência dos laboratórios, permitindo o processamento de grandes volumes de exames em um curto espaço de tempo (FARUQUI et al., 2023).

Além disso, a integração dessas tecnologias com sistemas de saúde eletrônicos facilita o armazenamento e a análise de dados históricos, promovendo uma abordagem mais informada e contínua do atendimento ao paciente (SHAHIN et al., 2024). Também, a inovação tecnológica na medicina é uma motivação importante, já que o uso de CNNs para diagnósticos preventivos pode impulsionar a pesquisa e o desenvolvimento contínuos

na interface entre a inteligência artificial e a medicina. Isso não só promove avanços tecnológicos, mas também abre novas frentes de pesquisa que podem beneficiar o campo da saúde em geral (YASMIN et al., 2021). Assim, a aplicação de CNNs no diagnóstico de exames preventivos apresenta um potencial significativo para revolucionar a prática médica, tornando-a mais precisa, eficiente e acessível.

Em contrapartida, um dos principais obstáculos é a necessidade de grandes quantidades de dados de alta qualidade e variedade para treinar as CNNs. A obtenção desses dados pode ser difícil em regiões onde a infraestrutura de coleta e armazenamento de informações médicas é limitada (MAHMOOD et al., 2018). Além disso, a rotulagem precisa dos dados, essencial para o treinamento dos modelos, é uma tarefa onerosa e demorada, exigindo um esforço considerável de especialistas para garantir a acurácia dos rótulos (AMIN et al., 2021).

Outro desafio significativo é a interpretação dos resultados gerados pelas tecnologias em questão. As redes, algumas vezes, são complexas de serem entendidas o que dificulta a compreensão de como elas chegam a determinadas conclusões. Essa falta de transparência pode ser um obstáculo para a aceitação pelos profissionais de saúde, que precisam confiar nos sistemas automatizados para tomar decisões clínicas importantes, mesmo que estas sejam apenas decisões iniciais que demandem exames complementares (MUN et al., 2021). Além disso, os sistemas automatizados de diagnóstico precisam passar por rigorosos processos de aprovação regulatória antes de serem utilizados clinicamente, o que pode atrasar sua implementação (RAPARTHI; DODDA; MARUTHI, 2021).

A integração das novas tecnologias com os sistemas de saúde já existentes também representa um desafio complexo. Essa integração pode ser custosa e tecnicamente difícil, exigindo ajustes significativos nos sistemas atuais para acomodar as novas ferramentas (AMIN et al., 2021). A atualização e manutenção contínuas dos modelos de CNN são igualmente desafiadoras, pois necessitam de novos dados regularmente para manter sua acurácia e relevância (MBUNGE; BATANI, 2023).

Por fim, a segurança e a privacidade dos dados dos pacientes são desafios críticos na implementação de soluções baseadas em IA. Garantir que os dados sejam protegidos contra acessos não autorizados é fundamental para manter a confiança dos pacientes e cumprir com as regulamentações de privacidade (CHIKHAOUI; ALAJMI; LARABI-MARIE-SAINTÉ, 2022). Além disso, a desigualdade no acesso à tecnologia em diferentes regiões pode dificultar a implementação dessas soluções de forma equitativa. O custo inicial de desenvolvimento e implementação dos sistemas baseados em CNNs também pode ser elevado, representando um desafio para instituições de saúde com recursos financeiros limitados (CHINTALA, 2022).

1.3 Objetivos

GERAL

O presente trabalho busca desenvolver uma solução computacional para classificar lesões em exames preventivos utilizando técnicas de aprendizagem profunda baseadas em redes neurais convolucionais.

ESPECÍFICOS:

- Realizar um levantamento sobre soluções computacionais voltadas para o diagnóstico de câncer cervical;
- Verificar o comportamento de um conjunto público de dados (dataset) com imagens de citologias cervicais utilizando tecnologias recentes na área de visão computacional;
- Realizar a execução de testes dos modelos para análise e validação dos resultados obtidos;
- Avaliar os modelos de classificação de Citologias Cervicais através das métricas de Acurácia, Precisão, Sensitividade e *F-Score*.

1.4 Reconhecimento da Pesquisa

É relevante destacar o papel fundamental do Núcleo de Pesquisa em Oncologia (NPO) do Hospital Universitário João de Barros Barreto (HUJBB) da Universidade Federal do Pará (UFPA) na concepção deste trabalho. O tema abordado nesta pesquisa foi diretamente sugerido por essa equipe, que se destaca como referência na área de oncologia. O Núcleo, além de propor o tema, ofereceu um suporte essencial, compartilhando seu vasto conhecimento e expertise, o que foi importante para o desenvolvimento deste estudo. O presente trabalho reflete o compromisso desse grupo com a inovação científica e a excelência no combate ao câncer cervical.

Em especial, agradece-se ao Dr. Paulo Pimentel de Assumpção, ao Dr. João Farias Guerreiro e à biomédica Josiellem Damasceno de Souza, cujas contribuições científicas e apoio técnico foram essenciais para o direcionamento e sucesso deste projeto. Seus conhecimentos e disposição em compartilhar suas experiências trouxeram solidez às diversas etapas da pesquisa.

Assim, a presente pesquisa alinha-se aos esforços contínuos do Núcleo de Pesquisa em Oncologia na busca por inovações que aprimorem o diagnóstico e a detecção precoce do câncer cervical. A dedicação e o empenho da equipe foram essenciais para a realização deste estudo, e sua colaboração representa um pilar significativo na condução da pesquisa. Assim, este trabalho expressa um sincero reconhecimento pelo suporte prestado e pela

contribuição científica proporcionada pelo núcleo, além de reconhecer que o tema central aqui tratado é fruto de seu direcionamento estratégico.

1.5 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

Capítulo 1: Apresenta uma introdução ao contexto do câncer cervical, justificando a relevância desta pesquisa e delineando seus objetivos, destacando a importância de novas abordagens para a detecção e tratamento precoce;

Capítulo 2: Fornece uma revisão bibliográfica sobre os principais conceitos e técnicas utilizadas no desenvolvimento da pesquisa;

Capítulo 3: Apresenta estudos correlatos que orientaram a elaboração deste trabalho, destacando suas contribuições e limitações, e como essas pesquisas anteriores fundamentaram a abordagem adotada nesta dissertação;

Capítulo 4: Explica a metodologia aplicada e as etapas necessárias para alcançar os objetivos propostos nesta dissertação, enfatizando as técnicas e os métodos de análise utilizados;

Capítulo 5: Descreve os resultados obtidos para os modelos desenvolvidos, discutindo as métricas desses modelos;

Capítulo 6: Apresenta as considerações finais do trabalho, sugestões para pesquisas futuras, trabalhos publicados referente ao tema e desafios da pesquisa.

2 Fundamentação Teórica

2.1 Considerações Iniciais

2.2 Histórico do Papanicolau

O exame de Papanicolau, também conhecido como citologia cervical, foi desenvolvido na década de 1940 pelo médico grego-americano *George Papanicolaou* (SHAW, 2000). Ele descobriu que células cancerígenas podiam ser detectadas em amostras do colo do útero e da vagina através da microscopia. Este exame revolucionou a ginecologia e a oncologia, permitindo a detecção precoce de lesões pré-cancerosas e cancerosas no colo do útero. Antes do desenvolvimento do Papanicolau, a maioria dos casos de câncer cervical era diagnosticada em estágios avançados, com poucas opções de tratamento eficazes (KOSS, 1989).

Segundo Bonfiglio (2005), nos anos seguintes ao seu desenvolvimento, o exame de Papanicolau foi amplamente adotado em todo o mundo. Nos Estados Unidos, campanhas de saúde pública na década de 1950 incentivaram as mulheres a realizarem o teste regularmente. Como resultado, houve uma queda significativa na incidência e mortalidade por câncer cervical. Em muitos países, programas nacionais de rastreamento foram implementados, recomendando que mulheres em idade reprodutiva fizessem o exame anualmente ou a cada três anos, dependendo de fatores de risco e diretrizes de saúde.

A introdução do exame de Papanicolau não só ajudou a detectar o câncer cervical em estágios iniciais, mas também promoveu a investigação e a compreensão das doenças do colo do útero. Pesquisas subsequentes revelaram a relação entre o vírus do papiloma humano (HPV) e o desenvolvimento de câncer cervical (MUÑOZ; BOSCH, 1996). Isso levou ao desenvolvimento de vacinas contra o HPV, que agora são recomendadas para meninas e meninos para prevenir infecções que poderiam resultar em câncer cervical e outros tipos de câncer associados ao HPV.

Hoje, o exame de Papanicolau continua a ser uma ferramenta vital na prevenção do câncer cervical, embora tenha evoluído com a introdução de testes de HPV e tecnologias mais avançadas de citologia (ORGANIZATION et al., 2022). Nesse viés, o autor Rodríguez et al. (2024), destaca que em combinação com a vacinação contra o HPV, o rastreamento regular com o Papanicolau tem o potencial de eliminar virtualmente o câncer cervical em populações com bom acesso a cuidados de saúde. No entanto, desafios persistem em áreas com recursos limitados, onde a implementação de programas de rastreamento e vacinação continua a ser uma prioridade de saúde pública.

2.3 Classes de Diagnóstico no Papanicolau

As classes de diagnóstico no exame de Papanicolau foram desenvolvidas para padronizar a interpretação e o relatório dos resultados das citologias cervicais. Essa padronização surgiu da necessidade de uniformizar a comunicação entre laboratórios e profissionais de saúde, facilitando a comparação de dados e a tomada de decisões clínicas (NJUE et al., 2022). As primeiras classificações foram estabelecidas na década de 1950, mas desde então, vários sistemas evoluíram para refletir avanços no entendimento e na tecnologia.

A mais amplamente utilizada hoje é o *Sistema Bethesda*, introduzido em 1988, que aprimorou as categorias de diagnóstico e incluiu terminologia mais descritivas e orientadas para a conduta clínica (PANGARKAR, 2022). Ele inclui várias categorias de diagnóstico, que vão desde Negativo para Lesão Intraepitelial ou Malignidade até Carcinoma. As principais categorias são: NILM, LSIL, HSIL e Carcinoma.

Cada classe reflete uma progressão de alterações celulares que vai desde um estado normal e saudável até a presença de células cancerígenas invasivas. Essas mudanças têm implicações profundas para o diagnóstico, monitoramento e tratamento clínico. Em estágios iniciais, como os de células normais e LSIL, a detecção precoce pode permitir intervenções menos invasivas e mais eficazes, prevenindo a progressão da doença. À medida que as células evoluem para HSIL e carcinoma, a identificação dessas alterações se torna crucial para a implementação de tratamentos mais agressivos e específicos. Assim, a compreensão detalhada de cada estágio é fundamental para estratégias de saúde pública, protocolos de rastreamento e abordagens terapêuticas que visam reduzir a mortalidade e melhorar a qualidade de vida dos pacientes (PUNHANI; AHLUWALIA, 2024).

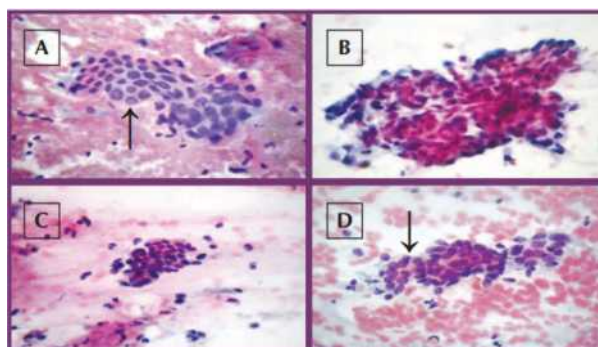
2.3.1 Classe NILM

A classe Normal no *Sistema Bethesda*, tecnicamente referida como Negativo para Lesão Intraepitelial ou Malignidade (NILM), indica que as células coletadas durante o exame de Papanicolau não apresentam grandes anormalidades. Nesse caso, não há sinais de infecções, inflamações graves, lesões pré-cancerosas ou células cancerígenas. Esse resultado sugere que o colo do útero está saudável e que não há necessidade imediata de intervenções adicionais além do acompanhamento regular (MERRICK et al., 2022).

Mulheres com resultados NILM são geralmente aconselhadas a continuar com os exames de rotina conforme as diretrizes de saúde, que recomendam a repetição do teste a cada três a cinco anos, dependendo da idade e do histórico clínico, já que a manutenção regular dos exames de Papanicolau ajudam a garantir que qualquer alteração celular futura seja detectada e tratada precocemente, aumentando significativamente as chances de prevenção e cura do câncer cervical.

Sobre o comportamento celular desta classe, as células cervicais normais são uniformes em tamanho e forma, com núcleos pequenos e regulares, e citoplasma claro e bem definido. Não há sinais de inflamação, infecção ou alterações anormais. Essas células desempenham suas funções típicas sem interrupção, mantendo a saúde do tecido cervical (BANERJEE et al., 2022). A uniformidade e a integridade das células são indicativas de um ambiente celular equilibrado e saudável. A Figura 1 apresenta a lâmina contendo essa classificação celular:

Figura 1 – Comportamento Celular NILM



Fonte: Retirado de (INCA) (2023)

2.3.2 Classe LSIL

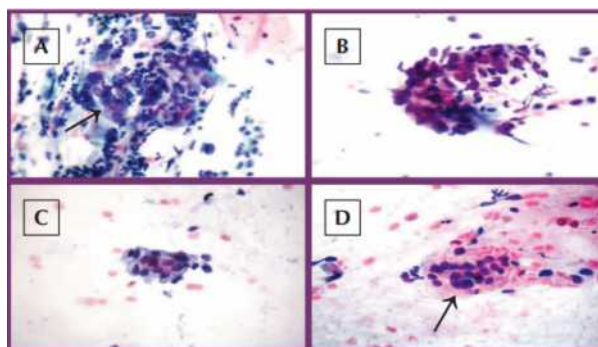
A classe LSIL no *Sistema Bethesda* é uma categoria de diagnóstico que indica a presença de alterações leves nas células do colo do útero, frequentemente associadas à infecção pelo vírus do papiloma humano (HPV) (ALRAJJAL et al., 2021). Conforme Alrajjal et al. (2021), o resultado LSIL sugere que há mudanças celulares que podem ser benignas e, muitas vezes, temporárias. Este diagnóstico não é imediatamente alarmante, mas requer acompanhamento regular para garantir que as alterações não evoluam para uma condição mais grave. A infecção pelo HPV é comum e pode ser transitória, com o sistema imunológico eliminando o vírus naturalmente em muitos casos.

Para pacientes com resultados de LSIL, o manejo clínico geralmente envolve monitoramento contínuo com exames complementares, bem como o de Papanicolau repetidos em intervalos mais curtos. Esse acompanhamento é importante para detectar qualquer progressão para lesões de alto grau ou outras anormalidades que possam necessitar de intervenções mais agressivas.

Sobre o comportamento celular desta classe, as células exibem alterações leves devido à infecção pelo vírus do papiloma humano (HPV). Essas alterações incluem núcleos maiores e ligeiramente irregulares, algumas variações no tamanho e na forma das células e um aumento na proporção núcleo-citoplasma. Essas mudanças são indicativas de uma infecção ativa que pode ou não resolver espontaneamente (ALREFAI; ALHEJAILI; HADDAD, 2024). O comportamento celular ainda é relativamente ordenado, mas as

alterações indicam uma resposta do tecido à infecção viral, que pode levar a uma proliferação celular anormal se não for monitorada. A Figura 2 apresenta a lâmina contendo essa classificação celular:

Figura 2 – Comportamento Celular LSIL



Fonte: Retirado de (INCA) (2023)

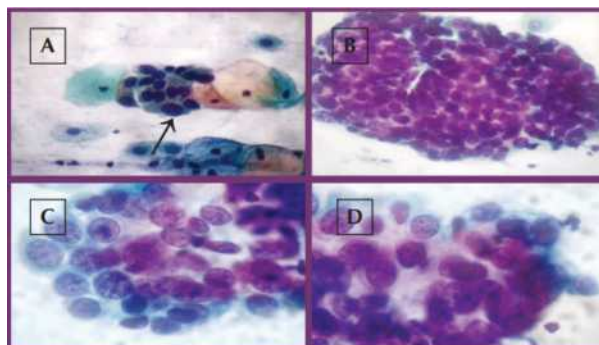
2.3.3 Classe HSIL

A classe HSIL no *Sistema Bethesda* representa alterações celulares mais graves no colo do útero, indicando um maior risco de progressão para câncer cervical se não tratadas. Esse diagnóstico sugere que as células do colo do útero estão significativamente anormais, o que pode ser precursora de um carcinoma invasivo (JUG RACHEL M.B.B.CH.; BEAN, 2023). Sua detecção é importante porque, ao contrário das lesões LSIL, essas lesões têm uma probabilidade muito maior de evoluir para uma condição cancerígena. Portanto, o diagnóstico requer uma avaliação e intervenção mais imediatas (PERKINS et al., 2020).

O autor Jug Rachel M.B.B.Ch. e Bean (2023), destaca que o manejo clínico das pacientes com HSIL geralmente inclui procedimentos adicionais como colposcopia, biópsia e, possivelmente, tratamento para remover as células anormais. A colposcopia permite uma visualização detalhada do colo do útero e a biópsia pode confirmar o grau de anormalidade e a extensão das lesões. Tratamentos podem variar desde procedimentos locais, como a excisão das áreas afetadas, até terapias mais abrangentes dependendo da extensão da lesão.

Sobre o comportamento celular desta classe, as células em HSIL apresentam anormalidades mais significativas e estão mais desorganizadas. Elas têm núcleos muito grandes e irregulares, com uma proporção núcleo-citoplasma acentuadamente aumentada, e os núcleos são hipercromáticos (mais escuros devido ao aumento do conteúdo de DNA). Essas células indicam uma proliferação celular descontrolada, que pode ser precursora do câncer invasivo (VOIDĂZAN et al., 2024). O comportamento celular em HSIL é caracterizado por um crescimento desordenado e acelerado, aumentando o risco de progressão para carcinoma se não tratadas. A Figura 3 apresenta a lâmina contendo essa classificação celular:

Figura 3 – Comportamento Celular HSIL



Fonte: Retirado de (INCA) (2023)

2.3.4 Classe Carcinoma

O carcinoma é um tipo de câncer que se origina nas células epiteliais, que revestem as superfícies internas e externas do corpo, incluindo a pele, os revestimentos internos dos órgãos e glândulas (PORCHERI, 2019). É a forma mais comum de câncer e inclui subtipos como carcinoma de células escamosas e adenocarcinoma. Segundo Porcheri (2019), o carcinomas podem se desenvolver em vários órgãos, como pele, pulmões, mama, próstata e colo do útero, entre outros. Embora câncer seja um termo mais amplo que abrange todos os tipos de tumores malignos, carcinoma refere-se especificamente aos que se originam no tecido epitelial.

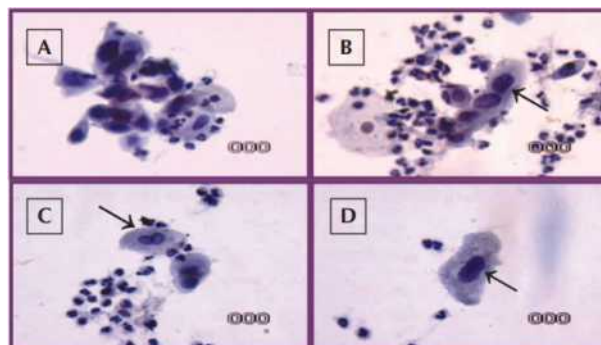
O carcinoma de colo de útero, também conhecido como câncer cervical, é um dos tipos mais comuns de câncer entre as mulheres no Brasil, especialmente na região Norte (Instituto Nacional de Câncer (INCA), 2022). Esse tipo de câncer se desenvolve nas células epiteliais que revestem o colo do útero. Segundo o Instituto Nacional de Câncer (INCA) (2022) em 2022, foram estimados 16.710 novos casos de câncer de colo de útero no Brasil, com uma taxa de incidência mais elevada nas regiões Norte e Nordeste, onde a cobertura de programas de rastreamento é mais limitada e o acesso ao tratamento é menor.

Na região Norte do Brasil, o cenário é preocupante devido à infraestrutura de saúde insuficiente e à falta de acesso a programas de rastreamento regular, como o exame de Papanicolau (SOUSA et al., 2021). O diagnóstico tardio é comum, o que reduz significativamente as chances de cura. Segundo dados recentes, a taxa de mortalidade por câncer de colo de útero na região Norte é uma das mais altas do país. A falta de conscientização, associada às barreiras geográficas e culturais, contribui para que muitas mulheres não realizem exames preventivos regularmente (RAMOS et al., 2020).

Sobre o comportamento celular desta classe, as células cancerígenas no carcinoma mostram severas anormalidades. Elas têm núcleos extremamente grandes e irregulares, nucléolos proeminentes, e um alto índice núcleo-citoplasma. Há significativa desorganização e agrupamento celular, indicando invasão tecidual e perda de aderência celular. O compor-

tamento dessas células é altamente agressivo, com capacidade de invasão e metastatização para outros tecidos e órgãos. A presença de necrose e inflamação no fundo do tecido é comum, refletindo a destruição tecidual causada pela proliferação descontrolada das células malignas (CANEPA; PATEL; GARCIA-MOLINER, 2019). Essas características indicam um estágio avançado de malignidade que requer intervenção médica imediata. A Figura 4 apresenta a lâmina contendo essa classificação celular:

Figura 4 – Comportamento Celular Carcinoma



Fonte: Retirado de (INCA) (2023)

Esse diagnóstico requer ação médica imediata para determinar a extensão do câncer e iniciar o tratamento adequado. O manejo clínico de pacientes com carcinoma cervical envolve uma série de abordagens terapêuticas, que podem incluir cirurgia para remover o tecido canceroso, radioterapia, quimioterapia ou uma combinação dessas modalidades (GENNIGENS et al., 2022). O plano de tratamento depende do estágio do câncer, da saúde geral da paciente e de outros fatores individuais. Além do tratamento, é fundamental o acompanhamento contínuo para monitorar possíveis recidivas e gerenciar efeitos colaterais.

2.4 Inteligência Artificial

A inteligência artificial é um campo da ciência da computação que se dedica à criação de sistemas capazes de realizar tarefas que normalmente exigiriam inteligência humana. Isso inclui processos como aprendizado, raciocínio, percepção, reconhecimento de fala e tomada de decisões (GAJJAR, 2023).

A IA está dividida em várias subáreas principais. A primeira é a aprendizagem de máquina (*machine learning*), que envolve o desenvolvimento de algoritmos que permitem aos computadores aprender e melhorar a partir de experiências passadas. Dentro dessa subárea, destaca-se o aprendizado profundo (*deep learning*), que utiliza redes neurais artificiais para modelar padrões complexos em grandes conjuntos de dados. Outra subárea importante é o processamento de linguagem natural (*natural language processing*), que

se concentra na interação entre computadores e linguagem humana, permitindo que as máquinas compreendam e gerem texto (COFINO; ESCORIAL; ENQUILINO, 2024).

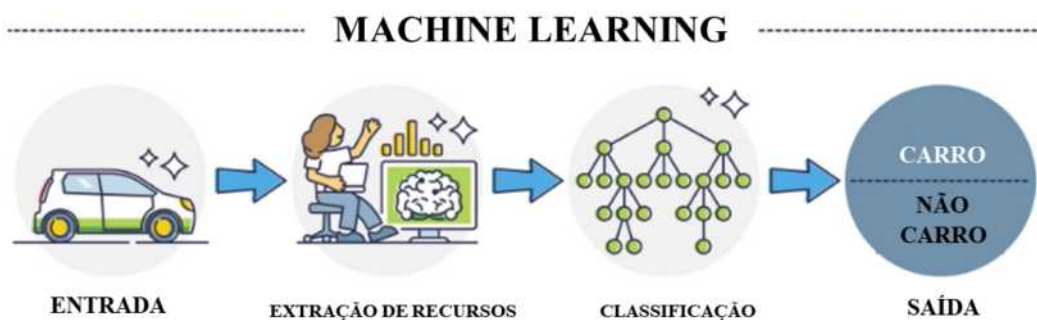
Além dessas subáreas, a visão computacional também desempenha um papel fundamental na inteligência artificial. Segundo Cofino, Escorial e Enquilino (2024), a visão computacional permite que os sistemas interpretem e processem informações visuais do mundo, funcionando de maneira semelhante ao olho humano. Outras subáreas da IA incluem a robótica, que aplica inteligência artificial para criar máquinas capazes de realizar tarefas físicas, e os sistemas especialistas, que são programas desenvolvidos para tomar decisões em domínios específicos utilizando conhecimento especializado.

2.4.1 Aprendizado de Máquina

A aprendizagem de máquina é a subárea que permite que sistemas computacionais aprendam e façam previsões ou decisões baseadas em dados. Em vez de serem programados com instruções específicas para cada tarefa, esses sistemas usam algoritmos que identificam padrões e relações nos dados, ajustando-se e melhorando continuamente com o tempo (GAJJAR, 2023).

O funcionamento da aprendizagem de máquina envolve várias etapas. Primeiro, um conjunto de dados é coletado e preparado, incluindo a limpeza e a formatação dos dados para que possam ser usados pelos algoritmos. Em seguida, um modelo é selecionado e treinado com esses dados. Durante o treinamento, o algoritmo ajusta seus parâmetros internos para minimizar erros e melhorar a precisão das previsões ou classificações. De acordo com Studer et al. (2021) esse processo pode ser iterativo, onde o modelo é constantemente ajustado e validado até atingir um desempenho satisfatório. Após o treinamento, o modelo pode ser testado em novos dados para avaliar sua capacidade de generalizar e fazer previsões precisas em situações não vistas anteriormente. A Figura 5 apresenta uma estrutura minimalista de sistemas de classificação baseados em aprendizado de máquina.

Figura 5 – Estrutura de Sistemas de Classificação baseados em ML



Fonte: Adaptado de Depositphotos (2024)

Os métodos de aprendizagem de máquina são variados e incluem técnicas supervisionadas, não supervisionadas e de reforço. Na aprendizagem supervisionada, os algoritmos são treinados com dados rotulados, onde a resposta correta é conhecida. Exemplos incluem regressão linear, que prevê valores contínuos, e árvores de decisão, que categorizam dados em diferentes classes. Na aprendizagem não supervisionada, os algoritmos trabalham com dados não rotulados e buscam padrões ou agrupamentos dentro dos dados. Métodos comuns incluem *clustering* (agrupamento), como o *k-means*, e PCA (Análise de Componentes Principais), que reduz a dimensionalidade dos dados. A aprendizagem por reforço envolve um agente que interage com um ambiente e aprende a tomar ações que maximizem uma recompensa cumulativa (MEMARIAN; DOLECK, 2024).

As funções de ML são amplamente diversificadas, elas incluem classificação, onde o objetivo é categorizar dados em classes específicas, como na detecção de *spam* em *e-mails*. Outra função é a regressão, que prevê valores numéricos contínuos, como a previsão de preços de ações (CHAUDHARY; KHURANA; AYALASOMAYAJULA, 2024). Além disso, a detecção de anomalias é usada para identificar padrões que desviam significativamente da norma, útil em segurança cibernética para detectar atividades fraudulentas. A recomendação é outra função importante, usada em sistemas de recomendação como os encontrados em serviços de *streaming* e *e-commerce*. Por fim, a segmentação de clientes agrupa consumidores com características semelhantes para personalizar estratégias de marketing. As técnicas e essas aplicações estão dispostas na Figura 6

Figura 6 – Técnicas e Aplicações Baseadas em ML



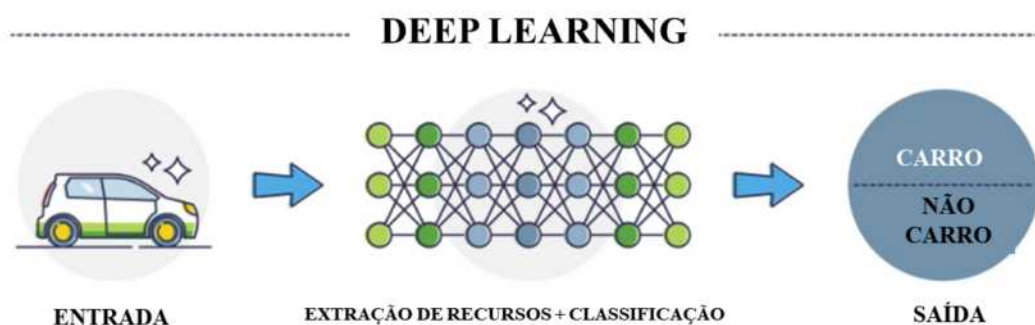
Fonte: Retirado de Zanette (2024)

2.4.2 Aprendizagem Profunda

Trata-se de uma subárea da aprendizagem de máquina que utiliza redes neurais artificiais compostas por várias camadas para modelar padrões complexos em dados. Inspirada no funcionamento do cérebro humano, essa técnica permite que computadores aprendam representações de dados em múltiplos níveis de abstração (KRAUSS et al., 2021). Isso é, em vez de depender de características criadas manualmente, como em métodos de aprendizagem de máquina tradicionais, as redes neurais profundas podem automaticamente descobrir as melhores características para representar os dados.

O funcionamento da aprendizagem profunda envolve várias camadas de neurônios artificiais, organizadas em uma estrutura chamada de rede neural profunda (GAJJAR, 2023). A Figura 7 apresenta uma estrutura minimalista de sistemas de classificação baseados em aprendizagem profunda.

Figura 7 – Estrutura de Sistemas de Classificação baseados em DL



Fonte: Adaptado de Depositphotos (2024)

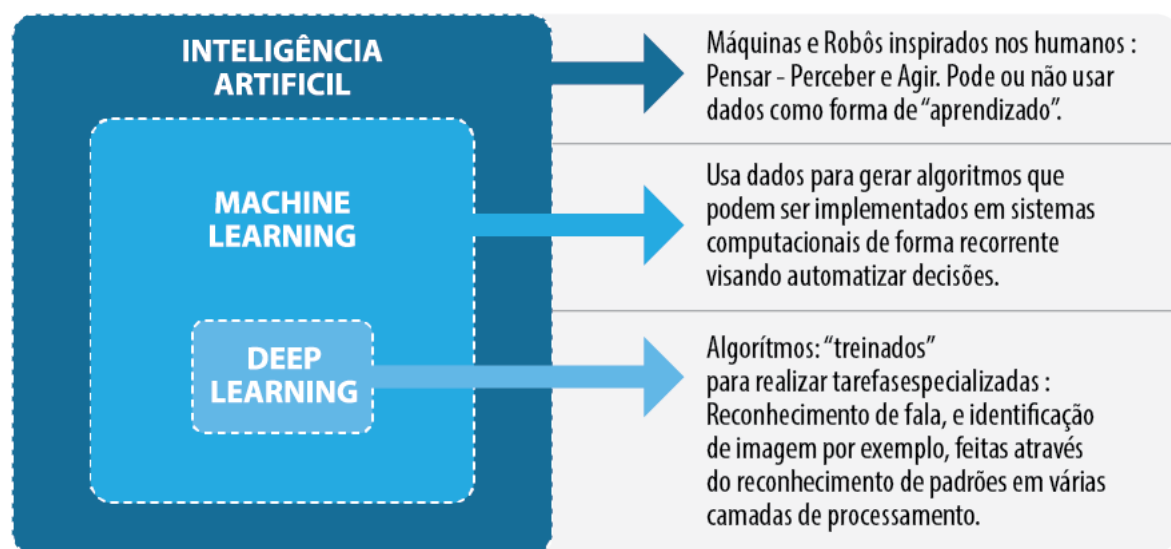
Os métodos de aprendizagem profunda incluem várias arquiteturas de redes neurais, cada uma projetada para diferentes tipos de tarefas. O autor Al-Fraihat et al. (2024) destaca que as redes neurais convolucionais são amplamente utilizadas para reconhecimento de imagens e processamento de vídeo, aproveitando operações de convolução para capturar padrões espaciais. Além disso, as redes neurais recorrentes (RNNs) são adequadas para dados sequenciais, como texto e séries temporais, pois possuem mecanismos de *feedback* que permitem a retenção de informações anteriores. As redes LSTM (*Long Short-Term Memory*) e GRU (*Gated Recurrent Unit*) são variantes das RNNs, desenvolvidas para resolver problemas de dependência de longo prazo. Além disso, as redes de *transforms*, se destacam em tarefas de processamento de linguagem natural, como tradução automática e análise de sentimentos.

Na visão computacional, os métodos de aprendizagem profunda são usados para reconhecimento facial, detecção de objetos e segmentação de imagens. No processamento de linguagem natural, é aplicado em traduções automáticas, *chatbots* e análise de sentimentos. Na área de saúde, utilizam *deep learning* para diagnóstico de doenças, análise de imagens

médicas e descoberta de medicamentos (LI, 2022). Em veículos autônomos, é essencial para a percepção do ambiente, reconhecimento de sinais de trânsito e tomada de decisões em tempo real. Além disso, a aprendizagem profunda é utilizada em recomendações de produtos, filtragem de *spam*, previsão de mercado financeiro e muitos outros domínios onde a modelagem de dados complexos é necessária.

A Figura 8 apresenta a relação entre os conceitos de IA, ML e DL.

Figura 8 – Relação entre os conceitos de Inteligência Artificial, Aprendizado de Máquina e Aprendizagem Profunda



Fonte: Retirado de Santos (2024)

2.4.3 Visão Computacional

A visão computacional é um campo interdisciplinar da ciência da computação que se concentra em desenvolver métodos para que computadores possam interpretar e compreender o mundo visual de maneira semelhante aos humanos. Utilizando algoritmos e modelos matemáticos, a visão computacional permite a análise e interpretação automática de imagens e vídeos, possibilitando a realização de tarefas como reconhecimento de objetos, detecção de padrões e análise de movimento (VOULODIMOS et al., 2018). Os principais objetivos da visão computacional incluem a automação de processos que requerem interpretação visual, a melhoria da precisão em tarefas de análise de imagem e a integração de sistemas visuais em aplicações práticas, como veículos autônomos, diagnósticos médicos e sistemas de vigilância.

O histórico e a evolução da visão computacional começaram nas décadas de 1960 e 1970, quando os pesquisadores começaram a explorar a possibilidade de desenvolver algoritmos para processar e interpretar imagens digitais. Os primeiros trabalhos focaram em problemas básicos de percepção visual, como a detecção de bordas e a segmentação de imagens (HE et al., 2022). Com o avanço da capacidade computacional e o desenvolvimento

de novos algoritmos, o campo evoluiu significativamente nas décadas seguintes. Na década de 1990, a introdução de técnicas baseadas em aprendizado de máquina trouxe melhorias consideráveis na capacidade de reconhecimento de padrões. Mais recentemente, o advento das redes neurais convolucionais (CNNs) e o aumento da disponibilidade de conjuntos de dados rotulados revolucionaram o campo, permitindo a criação de modelos mais precisos (CHEN et al., 2021).

A visão computacional envolve várias técnicas para analisar e interpretar imagens e vídeos. Três dos principais processos utilizados neste campo são a classificação, a detecção e a segmentação (ALOMAR; AYSEL; CAI, 2023). Cada um desses processos serve a diferentes propósitos e níveis de detalhe na análise de dados visuais.

2.4.3.1 Classificação em Visão Computacional

A classificação em visão computacional refere-se ao processo de atribuir uma etiqueta ou categoria a uma imagem inteira. Este processo envolve o treinamento de um modelo de aprendizado de máquina em um conjunto de dados rotulados, onde cada imagem está associada a uma classe específica. Durante o treinamento, o modelo aprende a reconhecer padrões e características distintivas de cada classe. Após o treinamento, o modelo pode ser usado para prever a classe de novas imagens desconhecidas (PAUL et al., 2021). Por exemplo, em sistemas de diagnóstico médico, a classificação pode ser usada para determinar se uma imagem de lâmina mostra sinais de uma doença específica. Em sistemas de segurança, pode ajudar a identificar pessoas ou objetos em imagens de vigilância.

As redes neurais convolucionais são amplamente utilizadas para tarefas de classificação devido à sua capacidade de aprender automaticamente características hierárquicas das imagens. As CNNs consistem em várias camadas, incluindo convolução, *pooling* e camadas totalmente conectadas, que juntas permitem a extração de características relevantes e a tomada de decisões baseadas nesses dados (KARKUZHALI R MURUGESHWARI, 2021). A precisão da classificação pode ser melhorada com técnicas como aumento de dados, que cria variações das imagens de treinamento para aumentar a robustez do modelo.

2.5 Inteligência Artificial na Medicina

A saúde está em constante evolução, impulsionada por avanços tecnológicos significativos, especialmente na área de diagnósticos. Considerada uma prioridade fundamental para o ser humano, a saúde é vital para o desenvolvimento e bem-estar (MUN et al., 2021). Nesse contexto, a saúde é um recurso essencial para a humanidade, sempre sujeita a inovações e adaptações para garantir uma melhor qualidade de vida.

A aplicabilidade da Inteligência Artificial (IA) na medicina é vasta e diversificada. Um dos usos mais notáveis é no diagnóstico de doenças através de imagens médicas, como

raios-X, ressonâncias magnéticas e exames. A IA pode detectar anomalias com bastante precisão, muitas vezes em estágios iniciais, permitindo intervenções mais rápidas e eficazes. Além disso, a IA está sendo utilizada na análise de dados médicos para identificar padrões e prever surtos de doenças (WANG; KHALIL; FIRDI, 2022).

No entanto, a introdução da IA na medicina também levanta questões importantes sobre sua regulamentação e aplicação ética. No Brasil, por exemplo, o Marco Legal da Inteligência Artificial busca criar um quadro normativo para o uso seguro e responsável dessa tecnologia, assegurando a proteção dos dados dos pacientes e o respeito aos direitos fundamentais. Em nível global, as discussões se concentram na criação de normas que garantam a transparência, a segurança e a responsabilidade no uso da IA, especialmente em áreas sensíveis como a saúde, onde decisões automatizadas podem ter impactos significativos. Assim, é importante promover a colaboração entre desenvolvedores, profissionais de saúde e órgãos reguladores para garantir que a IA seja utilizada de maneira eficaz e ética.

A integração da IA no fluxo de trabalho clínico tem o potencial de melhorar significativamente a eficiência dos processos médicos. Algoritmos de IA treinados para triagem podem priorizar casos que necessitam de uma revisão detalhada, auxiliando médicos a tomarem decisões mais rápidas e informadas. Essa automação aumenta a precisão dos diagnósticos, agiliza o atendimento e permite que os profissionais de saúde se concentrem em casos mais complexos e no cuidado direto ao paciente.

Contudo, desafios importantes permanecem, incluindo a proteção da privacidade e a segurança dos dados dos pacientes, assim como a qualidade e a integridade dos dados utilizados para treinar os algoritmos. Também é necessário desenvolver soluções para lidar com dados incompletos ou enviesados (SHAHIN et al., 2024). Por isso, é fundamental estabelecer um equilíbrio entre o avanço tecnológico e a proteção dos direitos humanos, garantindo que a IA seja aplicada de forma justa, transparente e segura na medicina.

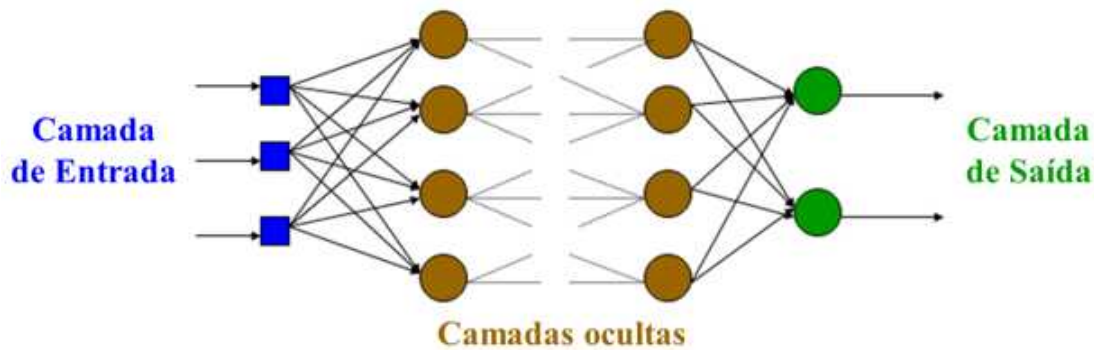
2.6 Redes Neurais

As redes neurais surgiram inspiradas na estrutura e funcionamento do cérebro humano, buscando criar sistemas computacionais capazes de aprender e tomar decisões de maneira semelhante a neurônios biológicos. A ideia inicial foi proposta por *Warren McCulloch* e *Walter Pitts* em 1943, que desenvolveram o primeiro modelo matemático de um neurônio artificial. No entanto, foi apenas em 1958, com o trabalho de *Frank Rosenblatt*, que o *perceptron*, uma versão mais prática de neurônio artificial, foi criado, marcando um avanço significativo no campo (VOULODIMOS et al., 2018).

Foram projetadas para reconhecer padrões e interpretar dados através de um processo de aprendizado. Esses sistemas são compostos por unidades básicas chamadas neurônios artificiais, que são organizados em camadas: uma camada de entrada, uma ou

mais camadas ocultas e uma camada de saída, dispostos na Figura 9 (HE et al., 2022). Cada neurônio recebe entradas, aplica uma função de ativação para processar esses dados e gera uma saída que é passada para os neurônios na próxima camada (VOULODIMOS et al., 2018). Durante o treinamento, os pesos das conexões entre os neurônios são ajustados iterativamente usando algoritmos de otimização, como o gradiente descendente, para minimizar o erro nas previsões do modelo em relação aos dados de treinamento.

Figura 9 – Arquitetura de uma Rede Neural Artificial Simples



Fonte: Adaptado de Nwankwo e Ihueze (2018)

Ao longo das décadas, as redes neurais evoluíram significativamente, levando ao desenvolvimento de várias arquiteturas, cada uma adequada para tarefas específicas (VOULODIMOS et al., 2018).

2.6.1 Redes Neurais Convolucionais

As CNNs são uma classe específica de redes neurais, especialmente eficazes para tarefas de visão computacional (ALOMAR; AYSEL; CAI, 2023). Elas utilizam camadas convolucionais que aplicam filtros ou *kernels* para extrair características espaciais importantes de uma imagem, como bordas, texturas e formas (VOULODIMOS et al., 2018). Essas camadas convolucionais são seguidas por camadas de *pooling*, que reduzem a dimensionalidade dos dados, mantendo as características mais relevantes e tornando o processamento mais eficiente.

Essas redes são capazes de capturar hierarquias de características, desde elementos básicos até representações mais complexas, o que as torna ideais para tarefas como reconhecimento de objetos, detecção de faces e segmentação de imagens (ALOMAR; AYSEL; CAI, 2023). A estrutura das CNNs permite que elas aprendam automaticamente as características importantes diretamente das imagens brutas, eliminando a necessidade de extração manual de características.

A principal diferença entre redes neurais tradicionais e redes neurais convolucionais reside na forma como elas processam os dados e extraem características. Redes neurais

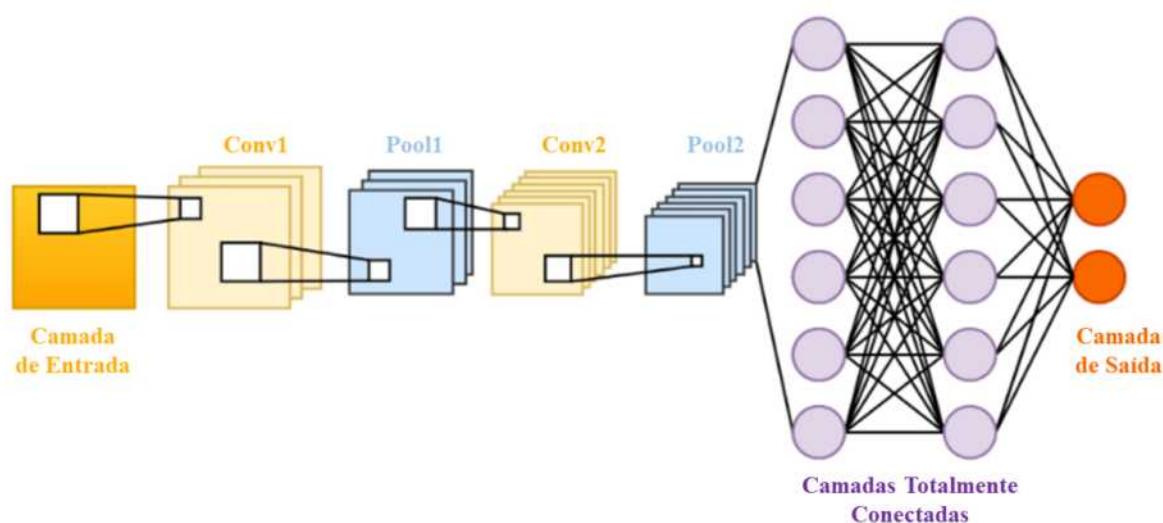
tradicionais, também conhecidas como redes neurais densamente conectadas ou *feedforward*, tratam os dados de entrada de forma global, conectando cada neurônio de uma camada a todos os neurônios da próxima camada (VOULODIMOS et al., 2018). Isso pode ser eficaz para dados estruturados, mas não captura eficientemente as relações espaciais presentes em imagens. Em contraste, as redes neurais convolucionais são projetadas especificamente para lidar com dados visuais, utilizando camadas convolucionais para extrair características locais através de filtros que se movem sobre a imagem. Essa abordagem permite que as CNNs identifiquem padrões espaciais e hierárquicos, tornando-as muito mais eficientes e precisas para tarefas de visão computacional (HE et al., 2022). Além disso, a estrutura das CNNs reduz significativamente o número de parâmetros, tornando o treinamento mais rápido e menos propenso ao sobreajuste em comparação com as redes neurais tradicionais.

2.6.1.1 Arquitetura de uma Rede Neural Convolucional

A arquitetura de uma Rede Neural Convolucional é projetada para simular de forma eficiente a percepção visual humana, representando, no entanto, uma evolução do *Perceptron* Multicamadas, diferenciando-se principalmente pelo arranjo especializado de suas conexões neuronais, inspiradas na organização do córtex visual dos animais (ALOMAR; AYSEL; CAI, 2023). Cada neurônio em uma CNN processa dados apenas de uma área restrita do campo visual, conhecida como campo receptivo, similar ao funcionamento dos neurônios no sistema visual biológico, que também respondem de maneira seletiva a estímulos em regiões específicas do campo visual (ALOMAR; AYSEL; CAI, 2023).

Essas redes são compostas primariamente por três tipos de camadas: convolucionais, de *pooling* e totalmente conectadas. Essa estruturação está indicada na Figura 10.

Figura 10 – Arquitetura de uma Rede Neural Convolucional



Fonte: Adaptado de Wang, Khalil e Firdi (2022)

2.6.1.2 Camada Convolutiva

As camadas convolucionais são fundamentais nas Redes Neurais Convolucionais por serem responsáveis na análise e interpretação de imagens. Estruturadas em torno de conjuntos de filtros ou *kernels* (HE et al., 2022), essas camadas aplicam-se diretamente às regiões locais da imagem de entrada, cada uma projetada para identificar características distintas (HE et al., 2022). Ao deslizar sobre a imagem durante as operações de convolução, os filtros ajustam os valores dos *pixels* de acordo com os pesos atribuídos a cada filtro em questão.

O processo de convolução resulta na criação de mapas de características, que são representações visuais onde as características detectadas pelos filtros são destacadas (VOULODIMOS et al., 2018). Conforme as informações passam pelas camadas subsequentes da CNN, os mapas de características são integrados e reprocessados para extrair e combinar informações cada vez mais abstratas (ALOMAR; AYSEL; CAI, 2023). Um exemplo desta operação está indicado na Figura 11.

Figura 11 – Exemplo de Aplicação de uma Camada de Convolução

$$\mathbf{F} \left[\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ \hline 7 & 8 & 9 \\ \hline \end{array} \otimes \begin{array}{|c|c|} \hline \text{I} & \text{II} \\ \hline \text{III} & \text{IV} \\ \hline \end{array} \right] = \begin{array}{|c|c|} \hline \text{A} & \text{B} \\ \hline \text{C} & \text{D} \\ \hline \end{array}$$

Entrada Kernel Saída

Fonte: Adaptado de Unzueta (2021)

Por outro lado, as funções de ativação em Redes Neurais Convolucionais são essenciais para introduzir não-linearidades no modelo, permitindo que as redes aprendam e modelam complexidades nos dados (ALOMAR; AYSEL; CAI, 2023). Cada função de ativação decide como um neurônio deve ativar sua saída com base na entrada recebida. A mais comum nas CNNs é a *Rectified Linear Unit* (VOULODIMOS et al., 2018). A ReLU destaca-se por sua simplicidade e eficácia em evitar o problema do desaparecimento do gradiente, facilitando o treinamento de redes profundas.

Outras funções de ativação incluem a *Sigmoid* e a Tangente Hiperbólica. A *Sigmoid*, que limita as saídas entre 0 e 1, é ideal para camadas de saída em problemas de classificação binária, enquanto a *Tanh*, que oferece saídas de -1 a 1 (ALOMAR; AYSEL; CAI, 2023), pode ser benéfica por centralizar os dados em zero. Ambas, no entanto, são menos utilizadas em camadas ocultas devido à possibilidade de causar extinção de gradientes em redes muito profundas.

Para a camada de saída, especialmente em tarefas de classificação multiclasse, a

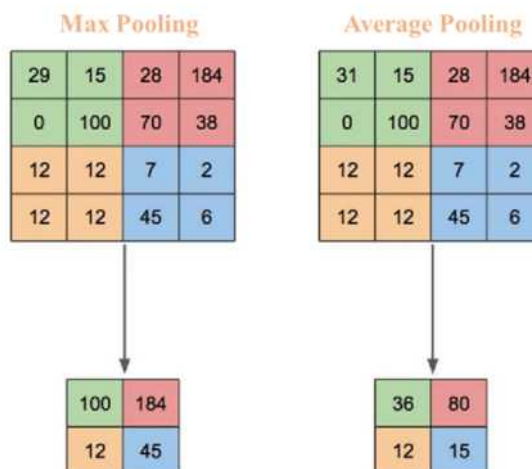
função *Softmax* é comum, pois converte os *logits* das camadas anteriores em probabilidades que somam 100% (ALOMAR; AYSEL; CAI, 2023). A escolha da função de ativação pode impactar significativamente o desempenho da rede e é influenciada pelo tipo específico de dados e do problema abordado (HE et al., 2022).

2.6.1.3 Camada de Pooling

A camada de *pooling*, também conhecida como camada de *subsampling* ou *down-sampling*, é o componente que segue as camadas convolucionais. Sua principal função é reduzir progressivamente o tamanho espacial da representação aprendida para diminuir a quantidade de parâmetros e computação na rede (HE et al., 2022). Isso é realizado através da operação de *pooling*, que agrega a saída das unidades em uma região definida, geralmente utilizando o máximo (*max pooling*) ou a média (*average pooling*) dos valores naquela região. O *max pooling* retorna o maior valor de um conjunto de dados de entrada, enquanto o *average pooling* calcula a média (HE et al., 2022). Essa redução de dimensionalidade resulta em uma representação mais compacta e robusta das características originais.

A importância dessa camada nas CNNs reside em sua capacidade de proporcionar invariância translacional às entradas da rede. Ao reduzir a resolução dos mapas de características, ela ajuda a tornar a rede menos sensível à localização exata das características dentro da imagem, o que é importante para tarefas como reconhecimento de imagens, onde o objeto de interesse pode aparecer em diferentes posições (HE et al., 2022). Além disso, ao condensar as informações, a camada em questão ajuda a prevenir o *overfitting*, melhorando a generalização do modelo ao aprender características hierárquicas (VOULODIMOS et al., 2018). Um exemplo desta operação está indicado na Figura 12.

Figura 12 – Exemplo de Aplicação de uma Camada de *Pooling*



Fonte: Adaptado de Yani, Irawan S e Setiningsih ST (2019)

2.6.1.4 Camada Totalmente Conectada

As camadas totalmente conectadas, também conhecidas como camadas densas, estão posicionadas, geralmente, após as camadas convolucionais e de *pooling* (MUHAMMAD K KIANI, 2023). A principal função dessas camadas é combinar as características extraídas pelas camadas anteriores para formar as previsões finais. Isso pois em uma camada totalmente conectada, cada neurônio está conectado a todos os neurônios na camada anterior, o que significa que os padrões aprendidos pelas camadas convolucionais e de *pooling* são sintetizados e usados para realizar tarefas específicas, como classificação ou regressão (MUHAMMAD K KIANI, 2023). Essas camadas são capazes de aprender uma função global de não-linearidade, ajudando a mapear a distribuição dos dados de entrada para as saídas desejadas.

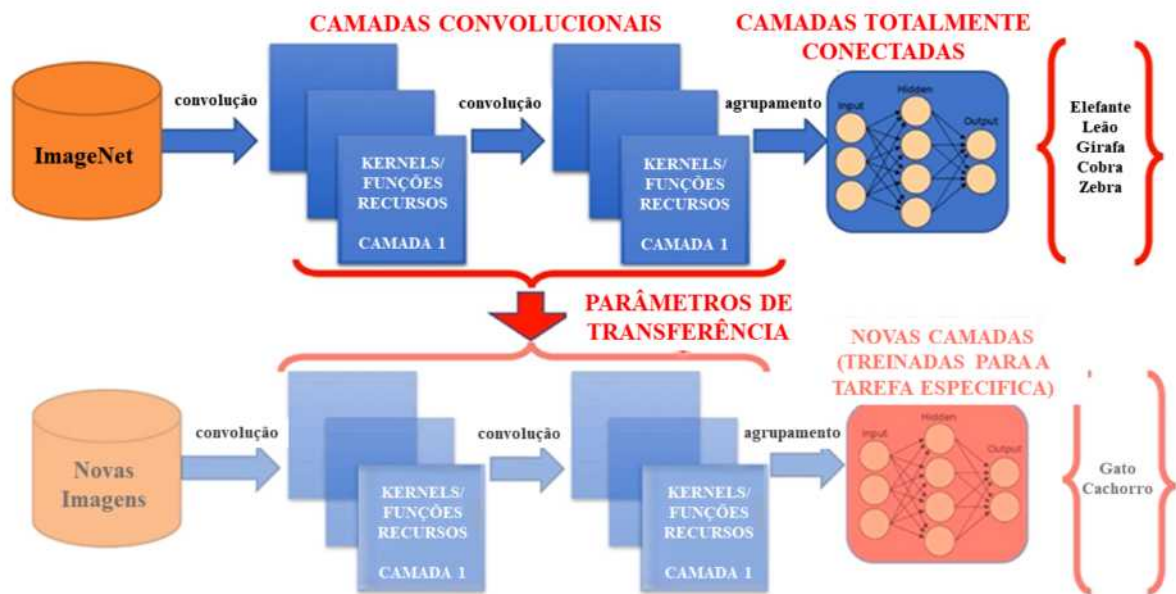
Assim, a importância das camadas totalmente conectadas em CNNs reside na sua capacidade de integrar informações locais coletadas pelas camadas convolucionais em conclusões globais que são essenciais para decisões finais.

2.7 Transferência de Aprendizado

Como forma de evitar ou minimizar o problema do *overfitting* existem várias técnicas que podem ser empregadas, como o *dropout*, *batch normalization*, transferência de aprendizado, *pretraining*, funções de regularização e *data augmentation* (CRUZ, 2021). O uso da técnica de transferência de aprendizado é uma das mais populares.

A transferência de aprendizado é uma técnica poderosa na área de aprendizado de máquina, particularmente útil em contextos onde dados de treinamento são escassos ou quando é necessário economizar recursos computacionais (YANI; IRAWAN S; SETININGSIH ST, 2019). O princípio subjacente é aproveitar o conhecimento aprendido por um modelo em uma tarefa específica e aplicá-lo a uma tarefa diferente, mas relacionada. Essa técnica é frequentemente empregada em Redes Neurais Convolucionais, onde modelos pré-treinados em grandes conjuntos de dados, como *ImageNet*, são adaptados para tarefas específicas (YANI; IRAWAN S; SETININGSIH ST, 2019). O processo geralmente envolve a reutilização das camadas iniciais de um modelo, que aprendem características gerais de imagens, enquanto as camadas superiores, mais especializadas nas tarefas originais, são ajustadas ou treinadas do zero para atender aos novos requisitos. Esse processo exemplificado está disposto na Figura 13.

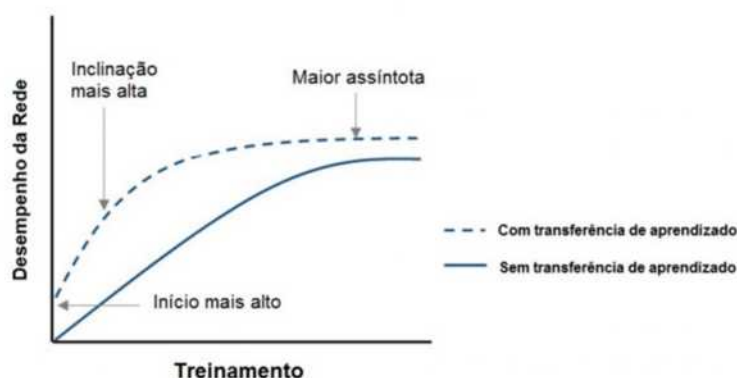
Figura 13 – Processo de Transferência de Aprendizagem



Fonte: Adaptado de Arcanjo (2023)

A função da transferência de aprendizado em CNNs visa reduzir significativamente o tempo e o esforço necessários para o desenvolvimento de modelos de aprendizado profundo (YANI; IRAWAN S; SETININGSIH ST, 2019). Ao iniciar o treinamento com pesos de um modelo que já capturou padrões complexos e úteis de grandes volumes de dados, os praticantes podem alcançar desempenho alto em suas tarefas específicas com muito menos dados de treinamento (YANI; IRAWAN S; SETININGSIH ST, 2019). Além disso, a transferência de aprendizado minimiza o risco de *overfitting*, já que o modelo não precisa aprender todas as características desde o início, permitindo que a rede generalize melhor para novos dados. Na Figura 14 está disposto um comparativo do desempenho de redes com transferência de aprendizado.

Figura 14 – Desempenho de Redes com Processo de Transferência de Aprendizagem

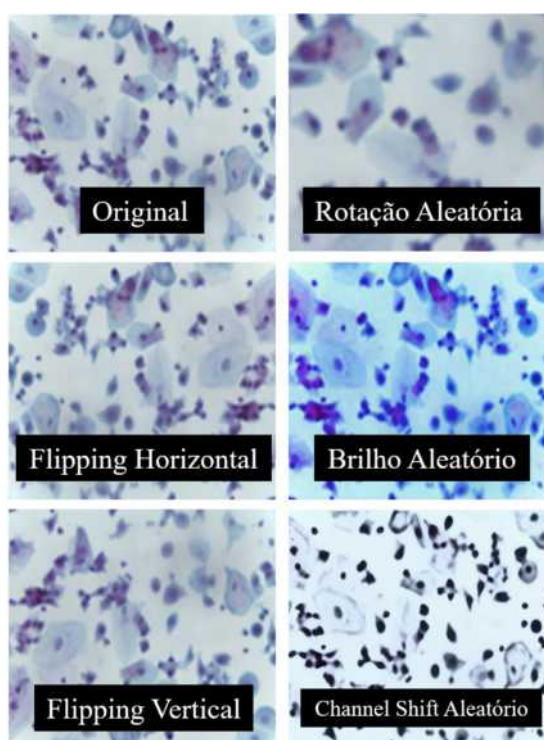


Fonte: Adaptado de Mathworks (2019)

2.8 Data Augmentation

Data augmentation é uma técnica utilizada para expandir artificialmente o tamanho do conjunto de dados, gerando diversas variantes realistas de cada instância (ALOMAR; AYSEL; CAI, 2023). Esta técnica de regularização visa reduzir o sobreajuste ao adicionar novas instâncias ao conjunto, resultantes de transformações que são aplicadas de maneira aleatória, enriquecendo o conjunto de dados de treinamento.

Figura 15 – Exemplos de Transformações Aplicadas em imagens Histopatológicas



Fonte: Gerada pela autora

A Figura 15 ilustra o processo de *data augmentation* aplicado a imagens histopatológicas para aumentar a diversidade do conjunto de dados. A imagem original, mostrada no canto superior esquerdo, serve como base para várias transformações (ALOMAR; AYSEL; CAI, 2023). Primeiramente, a imagem foi submetida a uma rotação aleatória, onde foi rotacionada em ângulos diversos para garantir que o modelo pudesse reconhecer as estruturas histopatológicas independentemente de sua orientação (HE et al., 2022). Em seguida, foram aplicadas as técnicas de *flipping* horizontal e vertical, espelhando a imagem lateralmente e verticalmente, respectivamente, o que adiciona variações de perspectiva e permite ao modelo identificar as mesmas estruturas sob diferentes orientações.

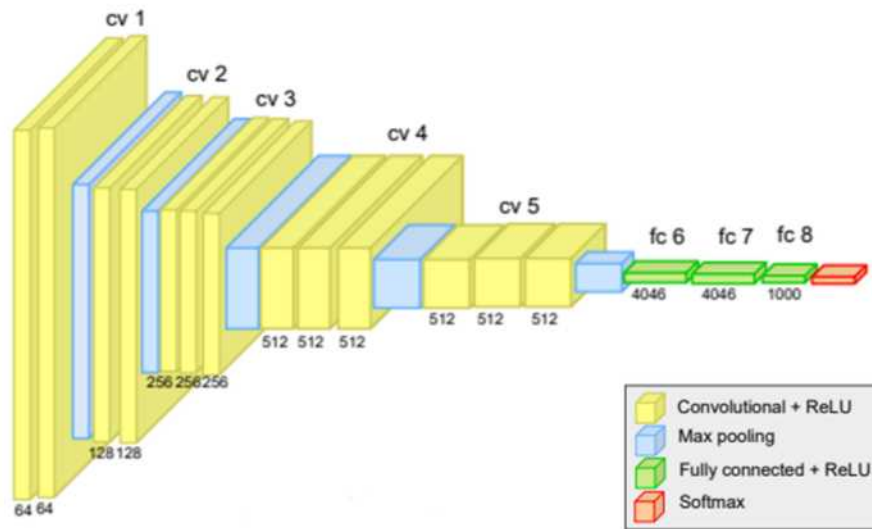
Além disso, ajustes de brilho aleatórios foram realizados, modificando o nível de brilho da imagem para ajudar o modelo a ser mais resistente a diferentes condições de iluminação que podem ocorrer durante a captura das imagens (ALOMAR; AYSEL; CAI, 2023). Por fim, a técnica de *channel shift* aleatório foi aplicada, onde as cores dos canais (vermelho, verde e azul) da imagem foram alteradas aleatoriamente.

2.9 Rede Neural Convolutacional VGG 16

Desenvolvida pelo grupo *Visual Geometry Group* da Universidade de *Oxford*, a VGG16 foi concebida com o objetivo de melhorar a precisão na classificação de imagens (BODDU AV KANUMURI, 2023). Composta por 16 camadas de aprendizado, sendo 13 camadas convolucionais e 3 camadas totalmente conectadas, a VGG16 segue uma estrutura sequencial que permite a extração de características visuais das imagens, utilizando *kernels* de convolução de tamanho 3x3, que são pequenos o suficiente para capturar detalhes finos, e ainda assim aprender padrões complexos (BODDU AV KANUMURI, 2023).

Proposta por Simonyan e Zisserman (2014) é uma rede profunda baseada na rede *AlexNet*, onde o número de *kernels* de convolução aumenta gradualmente de 64 na primeira camada para 128, 256 e, por fim, 512 na última camada. A camada de *pooling* utiliza um *kernel* de convolução de tamanho 2x2 e um passo de 2. Nas camadas totalmente conectadas as configurações usadas são as mesmas em todas as redes. A camada final, por sua vez, é a camada *softmax*. E, todas as camadas ocultas são equipadas com não linearidade de retificação. A estrutura mais detalhada desta rede está disposta na Figura 16.

Figura 16 – Estrutura da CNN VGG16



Fonte: Adaptado de Silva (2022)

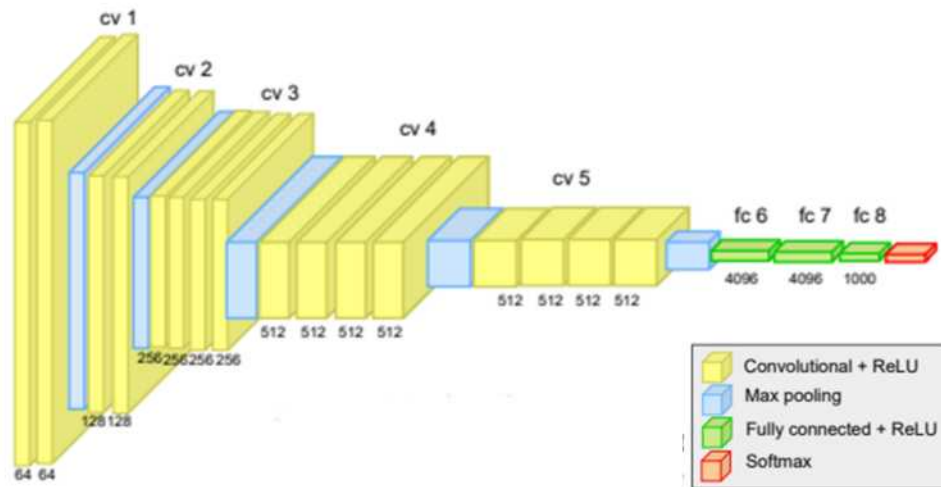
Apesar de ser superada por arquiteturas mais recentes, a VGG16 continua sendo uma escolha popular devido à sua simplicidade e boa adequação dos modelos. Sua estrutura facilita a implementação e a modificação para diversas aplicações de visão computacional (MURTHY HV CHAITRA, 2021). Além disso, a VGG16 oferece um bom ponto de partida para transferência de aprendizado, pois pode ser pré-treinada em grandes *datasets* como o *ImageNet*.

2.10 Rede Neural Convolucional VGG 19

Esta rede é uma extensão da VGG16, composta por 19 camadas de aprendizado, sendo 16 camadas convolucionais e 3 camadas totalmente conectadas. As camadas convolucionais utilizam *kernels* de tamanho 3x3 com um passo de 1. O aumento no número de camadas convolucionais na VGG19, em comparação com a VGG16, proporciona uma maior capacidade de captura de padrões complexos, o que pode resultar, geralmente, em um desempenho melhor em tarefas de visão computacional (DAWOOD, 2023).

A arquitetura da VGG19 segue a mesma filosofia da VGG16, com camadas de *pooling* intercaladas para reduzir a dimensionalidade espacial das características e camadas totalmente conectadas nas etapas finais, culminando em uma camada *softmax* para a classificação das imagens (DAWOOD, 2023). A utilização de funções de ativação ReLU em todas as camadas ocultas ajuda a introduzir não linearidade e acelerar a convergência durante o treinamento. A estrutura mais detalhada desta rede está disposta na Figura 17.

Figura 17 – Estrutura da CNN VGG19



Fonte: Adaptado de Silva (2022)

2.11 Métricas de Avaliação

As métricas de avaliação em visão computacional são fundamentais para medir a eficácia e precisão dos modelos propostos. Essas métricas fornecem uma base quantitativa para comparar diferentes modelos e técnicas, ajudando a determinar quais são mais adequados para uma aplicação específica (HE et al., 2022). Além disso, elas ajudam a identificar possíveis áreas de melhorias, orientando ajustes e refinamentos nos algoritmos.

Entre as métricas mais comuns estão a acurácia, precisão, *recall*, *F1-score*, AUC-ROC, entre outras, cada uma oferecendo uma perspectiva diferente sobre o desempenho do modelo. A seleção adequada de métricas de avaliação é importante para a interpretação correta dos resultados (HE et al., 2022). Além disso, as métricas de avaliação permitem a comparação direta entre diferentes abordagens e algoritmos, facilitando o progresso contínuo no desenvolvimento de soluções de visão computacional.

2.11.1 Matriz de Confusão

A matriz de confusão é uma ferramenta importante e amplamente utilizada para avaliação de sistemas de classificação de dados, pois permite analisar de forma rápida e intuitiva o desempenho de cada sistema (HE et al., 2022). Os valores que compõem a matriz são obtidos fornecendo os segmentos do conjunto de teste ao método de classificação e comparando sua predição com a classe correta de cada segmento.

Ela é uma tabela que resume os resultados da predição, comparando os valores previstos com os valores reais. A matriz é composta por quatro componentes principais:

verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN), exemplificados na Figura 18. Esses componentes ajudam a entender não apenas a acurácia geral do modelo, mas também como ele se comporta em termos de precisão e *recall*.

Figura 18 – Exemplo de matriz de confusão para classificadores binários



Fonte: Adaptado de Ruback, Avila e Cantero (2021)

A partir da matriz de confusão, é possível extrair várias métricas para medir a qualidade da classificação. Nesse contexto, as principais métricas utilizadas para avaliação e validação em sistemas de classificação de imagens são: acurácia, precisão, *recall*, *F1-score*, e a curva ROC-AUC (MOHAPATRA, 2024).

2.11.2 Acurácia

A acurácia mede a proporção de predições corretas feitas por um modelo em relação ao total de predições realizadas (HE et al., 2022). É calculada como o número de predições corretas dividido pelo número total de predições, obtida através da Equação (2.1).

$$Acc. = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.1)$$

No contexto de diagnósticos médicos, a acurácia é importante porque oferece um retorno sobre a eficácia de um modelo em identificar corretamente os casos de uma condição específica (HE et al., 2022). No entanto, enquanto uma alta acurácia é desejável, ela pode ser insuficiente por si só, especialmente em cenários de saúde onde a proporção de casos positivos, geralmente, é menor do que a de negativos.

Por isso, além da acurácia, outras métricas também devem ser consideradas para garantir que o modelo seja efetivo na identificação de casos positivos e negativos, proporcionando um diagnóstico médico mais confiável e equilibrado.

2.11.3 Precisão

A precisão indica a proporção de predições positivas corretas em relação ao total de predições positivas feitas pelo modelo (HE et al., 2022). Em suma, a precisão mede

quantos dos casos que o modelo classificou como positivos são realmente positivos. A fórmula para calcular a precisão é dada pela Equação (2.2)

$$Prec. = \frac{VP}{VP + FP} \quad (2.2)$$

No contexto de diagnósticos médicos, a precisão é importante porque uma alta precisão significa que o modelo tem uma baixa taxa de falsos positivos, o que é necessário para evitar diagnósticos incorretos que podem levar a tratamentos desnecessários e estresse para os pacientes.

2.11.4 Sensibilidade ou *Recall*

Mede a proporção de verdadeiros positivos em relação ao número total de casos reais positivos (TING et al.,). Em suma, a sensibilidade indica a capacidade do modelo de identificar corretamente todos os casos positivos, é calculada segundo a Equação (2.3).

$$Sens. = \frac{VP}{VP + FN} \quad (2.3)$$

No contexto de diagnósticos médicos, a sensibilidade é extremamente importante porque uma alta sensibilidade garante que a maioria dos pacientes com a condição sejam identificados corretamente, minimizando o risco de casos não detectados (TING et al.,).

2.11.5 *F1-score*

É uma métrica de avaliação que combina a precisão e a sensibilidade em uma única medida harmônica, oferecendo um balanço entre as duas (IBANGA, 2024). Dada pela Equação (2.4), ele é especialmente útil em situações onde existe um desequilíbrio entre as classes ou quando tanto os falsos positivos quanto os falsos negativos são importantes (IBANGA, 2024).

$$F1 = 2 * \frac{Prec. * Sens.}{Prec. + Sens.} \quad (2.4)$$

No contexto de diagnósticos médicos, o *F1-score* é importante porque equilibra a necessidade de minimizar falsos positivos e falsos negativos. Isso pois um falso positivo pode levar a tratamentos desnecessários, enquanto um falso negativo pode resultar na falta de tratamento para uma condição grave (CONCIATORI A VALLETTA, 2024).

2.12 Considerações Finais

Neste capítulo, procedeu-se à fundamentação teórica de todos os conceitos e técnicas empregados ao longo deste estudo, visando aprimorar a compreensão dos próximos capítulos. Na sequência, Capítulo 3, são expostas as pesquisas que abordam temáticas correlatas à presente investigação.

3 Trabalhos Correlatos

3.1 Considerações Iniciais

Este capítulo visa apresentar os principais trabalhos relacionados ao uso de classificação de imagens em visão computacional. Também aborda arquiteturas focadas na detecção de câncer cervical e lesões cancerosas, com ênfase na identificação de padrões citológicos anormais, com o intuito de auxiliar profissionais da saúde.

3.2 Levantamento do Estado da Arte

Para dar-se início ao levantamento bibliográfico deste trabalho, foi realizada uma pesquisa no portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). O portal CAPES, possui uma biblioteca virtual criada para proporcionar acesso global, atualizado e de alta qualidade à produção científica. A plataforma cobre todas as áreas do conhecimento e disponibiliza bases de dados textuais e de referências, incluindo mais de 50.000 títulos de periódicos de âmbito nacional e internacional. Ademais, oferece à comunidade acadêmica – composta por estudantes, professores, pesquisadores e técnicos – um dos maiores acervos bibliográficos do mundo, contendo artigos, teses, patentes, trabalhos de conferências, livros eletrônicos e outros recursos fundamentais (CAPES et al., 2020).

Para proporcionar uma visualização mais eficiente do levantamento do estado da arte, utilizou-se a ferramenta VOSviewer (2024) para a análise de redes bibliométricas. De acordo com Eck e Waltman (2010), esta ferramenta é amplamente utilizada para a análise de dados bibliométricos, incluindo artigos científicos, referências, palavras-chave e citações.

A utilização do VOSViewer permite a criação de representações gráficas das interligações de citação, facilitando a identificação dos principais autores e a correlação das palavras-chave. Este processo promove a detecção de tendências em diversas áreas do conhecimento, contribuindo significativamente para a compreensão da dinâmica e da evolução da produção científica.

Além disso, a análise de redes bibliométricas com o VOSViewer possibilita uma visão detalhada das colaborações entre pesquisadores e das influências exercidas por trabalhos relevantes. A identificação dessas relações é importante para a elaboração de estratégias de pesquisa e desenvolvimento, bem como para a orientação de investimentos em áreas emergentes e promissoras.

Para a construção da rede bibliométrica, utilizaram-se as palavras-chave "*Pap*

desenvolvida.

No trabalho proposto por Alsalatie et al. (2022), é apresentada uma abordagem baseada em deep learning para a análise de imagens de lâminas inteiras (Whole Slide Images - WSI) de esfregaços cervicais. O estudo utiliza uma estrutura de aprendizado profundo em ensemble, que foca na classificação automática de imagens completas, em vez de células individuais, oferecendo uma análise abrangente e detalhada. A abordagem classifica as células cervicais em quatro categorias: normal, carcinoma de células escamosas (SCC), lesão intraepitelial escamosa de baixo grau (LSIL) e lesão intraepitelial escamosa de alto grau (HSIL). A proposta visa melhorar a precisão e a eficiência no diagnóstico de anormalidades cervicais, proporcionando uma ferramenta valiosa para a detecção precoce e tratamento do câncer cervical.

O autor Lima et al. (2023) faz a utilização de uma rede neural convolucional baseada em máscara (Mask RCNN) para segmentação e classificação automática de células cervicais em exames de Papanicolau. O estudo utiliza três conjuntos de dados, combinando um conjunto privado e bancos de dados abertos (SIPaKMeD e Mendeley), permitindo a análise precisa de células em diferentes estágios de desenvolvimento. Essa abordagem classifica as células cervicais em até cinco categorias, incluindo células normais, pré-cancerosas e cancerosas, com uma precisão média (mAP) superior a 60 % e F1-score acima de 70%. O objetivo é automatizar o processo de diagnóstico, fornecendo relatórios completos aos consultores médicos e melhorando a precisão e a eficiência na detecção precoce de anormalidades cervicais, contribuindo para um tratamento mais rápido e eficaz das pacientes.

No estudo proposto por Alsalatie et al. (2022), foi realizada uma comparação entre a citologia em base líquida (ThinPrep) e o teste de Papanicolaou convencional para a triagem de câncer cervical em uma amostra de 5.428 pacientes. A pesquisa, conduzida em seis centros na França, mostrou que o método ThinPrep detectou 29% mais casos de ASCUS (células escamosas atípicas de significado indeterminado) e 39% mais lesões intraepiteliais escamosas de baixo e alto grau (LSIL e HSIL) em comparação ao método convencional. Além disso, o ThinPrep apresentou uma maior precisão e qualidade de amostras, reduzindo artefatos como sangue e inflamação. O estudo concluiu que o ThinPrep é uma alternativa mais eficaz para a detecção precoce de lesões pré-cancerosas, otimizando a triagem do câncer cervical.

O estudo realizado por Chen et al. (2020) tem como objetivo melhorar a classificação de células cervicais assistida por computador, essencial na detecção de câncer cervical. Para isso, os autores propõem um método baseado em transferência de aprendizado e *ensemble de snapshots* (TLSE - Transfer Learning based Snapshot Ensemble), que combina as vantagens do aprendizado por transferência e do *ensemble de snapshots* em um único processo de treinamento, aplicando-o à classificação fina de células cervicais em sete

categorias, usando a base de dados Herlev. A rede utilizada é um modelo de rede neural convolucional (CNN) pré-treinado (os modelos VGG, ResNet-18 e Inception-ResNet) no conjunto de dados *ImageNet*, ajustado para o conjunto de dados de células cervicais. O modelo base é modificado com a adição de camadas totalmente conectadas e técnicas de regularização, como *dropout* e normalização de lotes, para melhorar a generalização e a robustez. O *ensemble de snapshots* é implementado para obter múltiplos modelos em um único ciclo de treinamento, ajustando a taxa de aprendizado de forma cíclica para permitir que o modelo atinja vários mínimos locais durante o processo de otimização. O método TLSE mostrou-se superior a outros métodos existentes, alcançando uma acurácia de 65,56% na classificação das imagens do conjunto de dados Herlev, demonstrando sua eficácia para tarefas de classificação fina de células cervicais.

Chen et al. (2022), propõe um modelo leve e eficiente para a classificação de células cervicais, ideal para sistemas de diagnóstico com recursos computacionais limitados, comuns em países em desenvolvimento. Utilizando a técnica de *knowledge distillation*, o poder de um modelo robusto *teacher* é transferido para um modelo leve *student*. Das imagens do *dataset* Herlev são extraídos *patches* de modo a aumentar os dados para balancear o treinamento. Modelos CNN pré-treinados são refinados, escolhendo o melhor como modelo *teacher*, cujas saídas são suavizadas em *soft labels* para treinar os modelos *student* com uma função de perda híbrida que combina *soft labels* e *hard labels*. Os resultados mostram que modelos leves treinados com *soft labels* superam aqueles treinados apenas com *hard labels*, com o modelo *Xception* atingindo precisão comparável ao Inception-ResnetV2, mas com 40% do tamanho e dos parâmetros. A combinação de CNNs leves e destilação de conhecimento melhora significativamente a precisão em ambientes com recursos limitados, aplicando-se também a outros sistemas de análise de imagens médicas.

O trabalho desenvolvido por Akbar et al. (2024), teve como objetivo desenvolver modelos de rede neural convolucional (CNN) e máquina de vetor de suporte (SVM) para identificar câncer cervical através de imagens de Papanicolau, utilizando o conjunto de dados SIPaKMeD com 4049 imagens. A metodologia incluiu o pré-processamento dos dados e a implementação de quatro cenários de modelagem CNN, utilizando ResNet50 e ResNet50V2. Foi feito o teste para alguns cenários e o cenário 4, que utilizou ResNet50V2, apresentou o melhor desempenho com uma precisão de 97.09%. Este trabalho evidenciou que a combinação de CNN com SVM, especialmente com a arquitetura ResNet50V2, foi a mais eficaz, oferecendo uma abordagem eficiente para a detecção automática de câncer cervical.

A pesquisa feita por Joynab et al. (2024), aborda um sistema de aprendizado federado para classificar imagens de Papanicolau, preservando a privacidade dos dados dos pacientes. Utilizando diferentes arquiteturas de rede neural convolucional (CNN) no *dataset* SIPaKMeD, o estudo testou três cenários experimentais e alcançou uma

precisão de teste de 94.36% em dados IID e 78.4% em dados não-IID. A arquitetura baseada em aprendizado federado demonstrou superioridade em comparação com modelos tradicionais de aprendizado de máquina, destacando-se pela precisão e capacidade de manter a privacidade. O estudo sugere futuras pesquisas para incorporar técnicas de privacidade diferencial e utilizar *datasets* mais diversos para aprimorar o sistema.

Os estudos previamente mencionados apresentam abordagens variadas de análise de imagens do exame papanicolau na mesma área deste projeto e introduzem conceitos relevantes relacionados à pesquisa proposta. A Tabela 1 resume de forma concisa os trabalhos discutidos, juntamente com suas respectivas limitações.

Tabela 1 – Resumo dos Trabalhos Correlatos

Referência	Objetivos	Lacunas	Data-set	Classificação
(ALSALATIE et al., 2022)	Desenvolver uma abordagem automatizada para analisar esfregaços cervicais com aprendizado profundo, melhorando a precisão.	Ausência de gráficos de precisão e perda, o que impede um melhor direcionamento e compreensão por parte do leitor, além de não avaliar dados desbalanceados.	<i>Liquid Based Cytology Pap Smear Images for Multi-Class Diagnosis of Cervical Cancer</i> - 963 imagens	<i>Bethesda</i> : Normal, LSIL, HSIL e SCC
(LIMA et al., 2023)	Desenvolver e testar uma arquitetura <i>Mask RCNN</i> para automatizar a detecção, classificação e segmentação de células cervicais em exames de Papanicolaú, agilizando o diagnóstico e tratamento do câncer cervical.	O desempenho não foi satisfatório nas classes de pré-câncer, como Metaplasia e Koilocytótica.	Dados privado (1.935 imagens), SIPaKMeD - 4.049 imagens, e <i>Liquid Based Cytology Pap Smear Images for Multi-Class Diagnosis of Cervical Cancer</i> - 963 imagens	Células Normais: SSE, ISE, CE, Endocervical. Células Pré-Cancerosas: SNKD, SNKD. Células Cancerosas: SNKD, SC-CIS
(CHEN et al., 2020)	Aprimorar a classificação de células cervicais com transferência de aprendizado e <i>ensemble snapshot</i> .	Número limitado de amostras, dificultando o treinamento de redes neurais profundas.	<i>Herlev dataset</i> - 917 imagens.	<i>Bethesda</i> : SSE, ISE, CE, LSIL, MSIL, SSIL, SCC.
(CHEN et al., 2022)	Criar um modelo leve para classificação de células cervicais usando destilação de conhecimento.	Restrito ao conjunto de dados <i>Herlev Pap smear</i> ; falta análise do impacto do balanceamento de dados.	<i>Herlev dataset</i> - 917 imagens.	3 classes normais e 4 anormais.
(AKBAR et al., 2024)	Desenvolver modelos de CNN e SVM para identificar câncer cervical em imagens de Papanicolaú.	Uso exclusivo do conjunto de dados SIPaKMeD, sem abrangência da diversidade de casos clínicos.	SIPaKMeD - 4.049 imagens	5 classes: Superficial Intermediate, Parabasal, Koilocytotic, Dyskeratotic, Metaplastic.
(JOYNAB et al., 2024)	Criar um sistema de aprendizado federado para classificar imagens de Papanicolaú, preservando a privacidade.	<i>Dataset</i> com diversidade limitada, o que pode reduzir a eficácia do modelo em cenários reais.	SIPaKMeD - 4.049 imagens	5 classes: Superficial Intermediate, Parabasal, Koilocytotic, Dyskeratotic, Metaplastic.

Foram selecionados dois trabalhos de cada um dos bancos de dados: *Liquid Based Cytology Pap Smear Images for Multi-Class Diagnosis of Cervical Cancer*, Herlev e *SIPaK-MeD*, considerando que esses são os principais bancos de dados utilizados na literatura para pesquisas sobre o diagnóstico multiclases de câncer cervical. A inclusão dessas diferentes bases de dados enriqueceu o embasamento teórico da pesquisa, permitindo uma análise abrangente das abordagens e características investigadas em diferentes contextos. Dentre essas bases, uma foi eleita como principal para o desenvolvimento deste trabalho, escolha que será detalhada no Capítulo 4.

3.3 Considerações Finais

De acordo com as limitações dos trabalhos relacionados estudados, observou-se que a análise e classificação de imagens de citologia cervical enfrentam desafios significativos, o que compromete a precisão dos resultados, tais como a limitação do número de amostras de treinamento e a dificuldade em otimizar hiperparâmetros. O uso de conjuntos de dados específicos e limitados, também levanta preocupações sobre a representatividade e a capacidade dos modelos de generalizar para cenários clínicos reais.

Para superar essas limitações neste trabalho é proposto o aumento do conjunto de dados de treinamento por meio de técnicas de aumento de dados, garantindo uma amostra mais diversificada e representativa bem como o teste de diversas topologias. Além de propor uma análise comparativa mais abrangente com relação ao impacto do uso de dados balanceados e de como o modelo reage com a variação de parâmetros.

Assim, este capítulo delineou os trabalhos correlatos que fundamentaram a realização desta pesquisa e forneceram a base teórica para o desenvolvimento dos modelos propostos. A seção subsequente expõe detalhadamente a metodologia empregada neste estudo.

4 Materiais e Métodos

Neste trabalho foi proposto o desenvolvimento de dois modelos de redes da família VGG para a classificação automática de resultados a partir de exame preventivo Papanicolau. Este capítulo oferece uma visão geral das etapas metodológicas empregadas no desenvolvimento deste estudo.

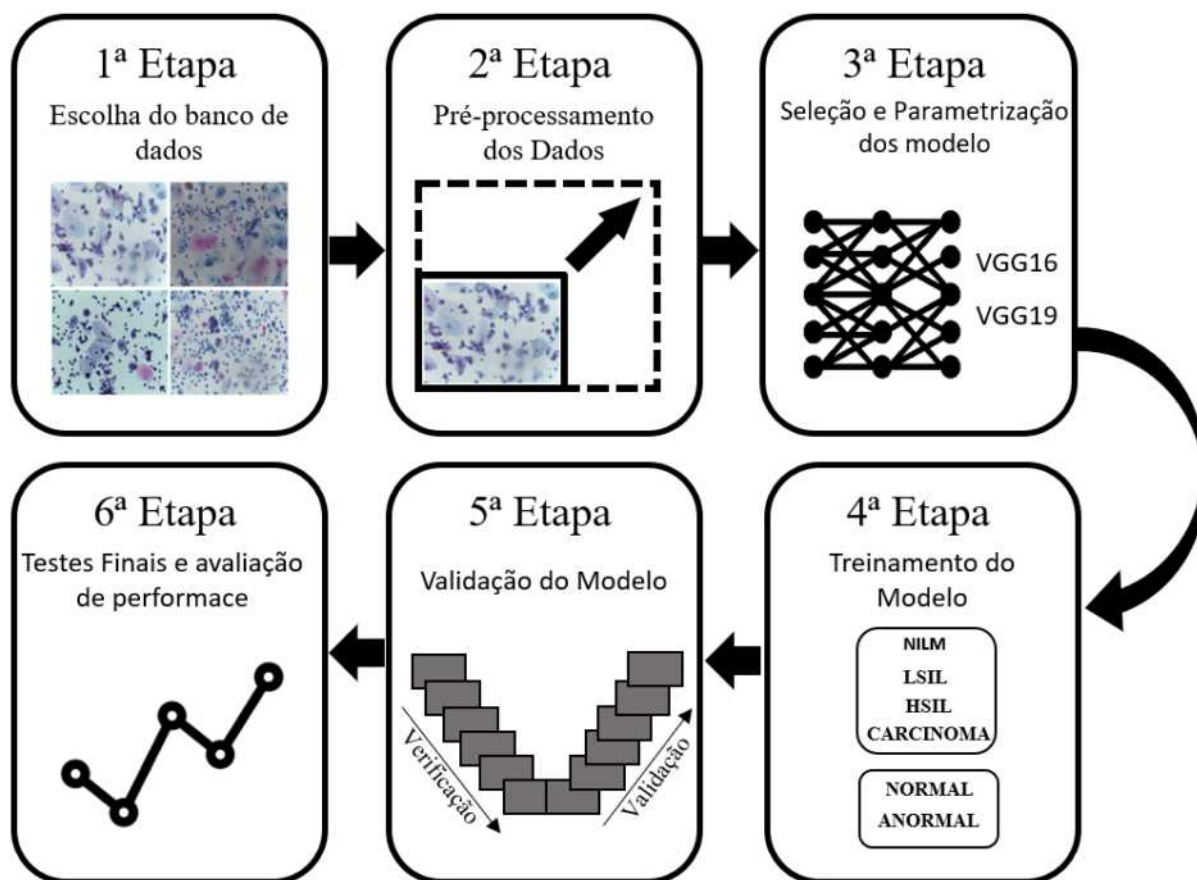
4.1 Considerações Iniciais

O trabalho se inicia a partir da seleção de um banco de dados contendo imagens de lâminas de Papanicolau, classificadas em quatro categorias: NILM, LSIL, HSIL e Carcinoma. Em seguida, realiza-se o pré-processamento dos dados, que inclui a normalização das imagens para adequá-las aos modelos propostos.

Após isso, é realizada a importação dos modelos VGG16 e VGG19 pré-treinados, adicionando camadas de classificação específicas e definindo os hiperparâmetros necessários. O treinamento é conduzido utilizando o conjunto de imagens de treinamento. Após essa etapa, os modelos são validados com o conjunto de imagens de validação para avaliar as métricas de desempenho.

A etapa final consiste nos testes e verificação de performance. Os modelos são avaliados no conjunto de teste para medir sua eficácia e, posteriormente, passa por um pós-processamento para compressão e otimização. A Figura 20 ilustra a metodologia proposta neste estudo por meio de um fluxograma, detalhando cada uma das etapas descritas.

Figura 20 – Fluxograma da Metodologia de Análise de Resultados do Exame Papanicolau Utilizando CNNs



Fonte: Elaborado pela autora

4.2 Conjunto de dados

A escolha do conjunto de dados a ser utilizado para o treinamento e validação dos modelos foi realizada sobre alguns seguintes critérios que constam abaixo juntamente com suas devidas justificativas de eleição:

1. Disponibilidade em repositórios de domínio público: já que facilita o acesso e a transparência da pesquisa científica, permitindo que outros pesquisadores repliquem estudos, validem resultados e contribuam para a evolução do conhecimento. Além disso, assegura a conformidade com práticas éticas e legais, especialmente em áreas como a saúde, onde a disponibilidade de dados precisa seguir padrões de seguranças de dados sensíveis.
2. Adequação a proposta de classificação, isto é, previamente segmentadas e/ou rotuladas por especialistas: porque minimizam erros e ambiguidades que poderiam

comprometer a performance dos modelos, proporcionando uma base confiável para o treinamento e validação.

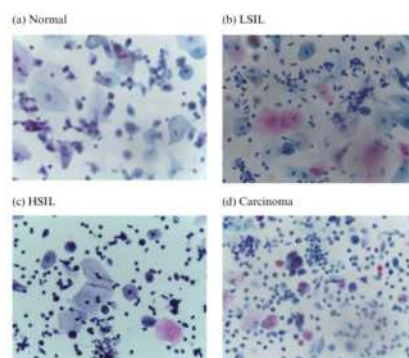
3. Coloração de Hematoxilina e Eosina (H&E): uma vez que se pretende aplicar os modelos futuramente em ambientes práticos, a coloração torna-se fundamental para o bom desempenho de trabalhos futuros, pois assegura que os modelos propostos possam ser diretamente integrados aos fluxos de trabalho clínicos, onde a coloração em questão é um padrão estabelecido para diagnósticos histopatológicos.
4. Maior número de amostras possíveis: um maior volume de amostras proporciona uma diversidade mais ampla de casos e variações, permitindo que o modelo aprenda a generalizar melhor e a lidar com uma gama mais ampla de situações. Isso reduz a probabilidade de *overfitting* e melhora a capacidade do modelo de realizar previsões precisas em novos dados não vistos anteriormente. Além disso, conjuntos de dados extensivos aumentam a representatividade das classes, garantindo que o modelo seja treinado em exemplos suficientemente variados

O conjunto de dados que melhor satisfaz os critérios acima, segundo a pesquisa realizada, foi o intitulado *Liquid Based Cytology Pap Smear Images for Multi-Class Diagnosis of Cervical Cancer* proposto por Hussain (2019). Este conjunto de dados é composto por 958 imagens, distribuídas em quatro classes distintas: NILM, LSIL, HSIL e Carcinoma.

A escolha desse banco de dados foi estratégica e importante para a pesquisa, pois oferece um conjunto de imagens com ampliação de 400x, obtidas com o microscópio Leica ICC50 HD. Esta ampliação fornece uma visão superior das características celulares em comparação com ampliações de 100x e 200x, permitindo uma análise mais detalhada das categorias celulares em questão. Além disso, a quantidade de imagens disponíveis no banco de dados facilita a validação e o treinamento dos modelos de redes empregados no estudo.

A Figura 21 representa visualmente as classes de imagens contidas no *dataset*.

Figura 21 – Representação Visual das Classes de Imagens no Dataset



Fonte: Retirado de Hussain (2019)

O banco de dados foi utilizado de duas maneiras neste trabalho. Inicialmente, foi empregado sem o balanceamento das classes para avaliar o desempenho dos modelos com a distribuição original dos dados. Em um segundo momento, foi realizado o pré-processamento das imagens e o balanceamento das classes. Essa abordagem permitiu analisar e comparar o comportamento dos modelos em ambas as situações, fornecendo observações sobre a influência do balanceamento de classes na eficácia dos modelos de classificação propostos.

Nas duas situações o banco de imagens foi dividido em três partes, seguindo a proporção de 70% para treino e 15% para teste e 15% para validação, conforme as práticas recomendadas para o desenvolvimento de modelos de *Machine Learning*.

4.2.1 Primeiro Momento- Classes Desbalanceadas

A utilização inicial do banco de dados sem o balanceamento das classes é fundamental para entender como os modelos de classificação se comportam com a distribuição original dos dados, que frequentemente reflete cenários reais onde certas classes podem estar sub-representadas. Esse passo é essencial para identificar possíveis vieses e limitações dos modelos quando confrontados com dados desequilibrados, fornecendo uma linha de base para o desempenho. Além disso, permite avaliar a necessidade e o impacto de técnicas de balanceamento, ajudando a desenvolver estratégias para melhorar a precisão dos modelos de redes em questão.

Nesse sentido, no primeiro momento, para o conjunto de treino, foram utilizadas 672 imagens, e do restante ficaram divididas 143 para teste e 143 para validação, distribuídas de acordo com a Tabela 2 abaixo:

Tabela 2 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 4 Classes Desbalanceadas

Classe	Treino	Teste	Validação	Total
NILM	428	92	92	612
LSIL	79	17	17	113
HSIL	115	24	24	163
Carcinoma	50	10	10	70
Total	672	143	143	958

Neste primeiro momento, também foi analisado o comportamento das classes desbalanceadas em uma classificação binária. Para isso, as classes LSIL, HSIL e Carcinoma foram agrupadas na classe anormal, enquanto a classe normal permaneceu separada. Essa abordagem permitiu observar o desempenho dos modelos em distinguir entre amostras nor-

mais e anormais. A distribuição das classes para esta análise seguiu os dados apresentados na Tabela 3.

Tabela 3 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 2 Classes Desbalanceadas

Classe	Treino	Teste	Validação	Total
Normal	428	92	92	612
Anormal	244	51	51	346
Total	672	143	143	958

A escolha de realizar uma análise binária se deu pela necessidade de simplificar o problema inicial de classificação, isso facilita a detecção de qualquer tipo de anomalia, o que é importante em cenários onde a identificação precoce de problemas é essencial, como na problemática em questão. Por outro lado, a análise geral de quatro classes foi escolhida para proporcionar uma classificação mais detalhada e específica. Essa abordagem é fundamental em cenários onde é necessário distinguir entre diferentes tipos de anormalidades a fim de definir tratamentos específicos e prognósticos. Dessa forma, a combinação das duas análises oferece a este trabalho uma visão completa da eficácia dos modelos, permitindo tanto uma detecção inicial abrangente quanto uma classificação detalhada e específica das anormalidades.

4.2.2 Segundo Momento- Classes Balanceadas

Treinar modelos de redes neurais com classes balanceadas é fundamental porque assegura que o modelo aprenda de forma equitativa sobre todas as categorias presentes no conjunto de dados. Classes desbalanceadas podem levar o modelo a desenvolver um viés em favor das classes majoritárias, ignorando ou mal interpretando as classes minoritárias. Isso, geralmente, pode resultar em um desempenho ruim, especialmente em termos de precisão e *recall* para as classes menos representadas. Assim, o balanceamento das classes garante uma representação justa e imparcial dos dados, promovendo um aprendizado mais eficiente.

Para balancear as classes no conjunto de dados foi usado o processo de *data augmentation*, onde o objetivo foi igualar o número de instâncias em cada classe ao número da classe majoritária, que é a classe NILM com 612 imagens. Assim, para a classe NILM, que já possui 612 instâncias, permanece inalterada. Para as outras classes, foi aplicado o *data augmentation* para aumentar suas instâncias até atingir o mesmo número da classe NILM.

O processo resultou em uma distribuição igualitária de 612 instâncias para cada classe, tornando o conjunto de dados balanceado. Nesse sentido, no segundo momento, o

banco de dados ficou composto por imagens 2448 imagens. Sendo utilizadas 1716 imagens para treino, e do restante ficaram divididas 368 para teste e 368 para validação, distribuídas de acordo com a Tabela 4.

Tabela 4 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 4 Classes após Data Augmentation

Classe	Treino	Teste	Validação	Total
NILM	428	92	92	612
LSIL	428	92	92	612
HSIL	428	92	92	612
Carcinoma	428	92	92	612
Total	1712	368	368	2448

Neste segundo momento, foi analisado o comportamento das classes balanceadas em uma classificação binária. Para isso, realizou-se uma seleção aleatória nas classes LSIL, HSIL e Carcinoma, deixando cada uma com 153 imagens. Essas imagens foram agrupadas na classe anormal, totalizando 612 imagens. Enquanto isso, a classe normal permaneceu com suas 612 imagens originais. A distribuição das classes para esta análise seguiu os dados apresentados na Tabela 5.

Tabela 5 – Distribuição das Imagens nos Conjuntos de Treino, Teste e Validação com 2 Classes após Data Augmentation

Classe	Treino	Teste	Validação	Total
Normal	428	92	92	612
Anormal	428	92	92	612
Total	856	184	184	1224

4.3 Treinamento dos modelos

O treinamento de CNNs possui um custo computacional significativo, exigindo grande capacidade de memória e unidades de processamento gráfico (GPUs) de alto desempenho. Para contornar essas limitações, aplicou-se a técnica de transferência de aprendizado e utilizou-se as plataformas *Google Colaboratory* e *Kaggle*, que oferecem recursos computacionais acessíveis.

Para o treinamento dos modelos, foram utilizadas redes pré-treinadas nos *datasets*, especificamente no *ImageNet*, para as CNNs VGG16 e VGG19. Esse método permite

aproveitar os pesos já ajustados a partir de grandes bases de dados, reduzindo o tempo de treinamento e melhorando a performance dos modelos em tarefas específicas com conjuntos de dados menores.

4.3.1 Classificadores com CNNs

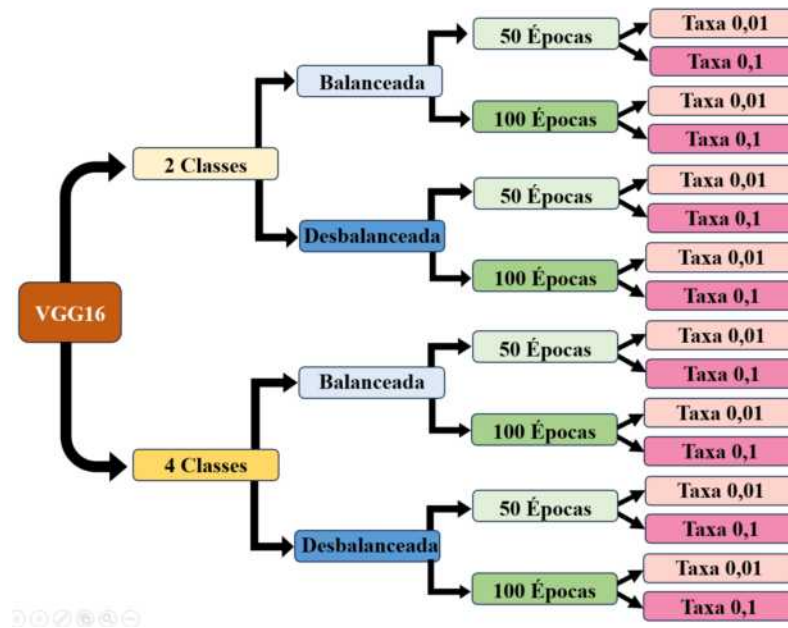
Os modelos propostos foram desenvolvidos após os resultados obtidos através de um estudo prévio entre as CNNs, as quais foram direcionadas para classificação de imagens médicas. Dessa forma, foi realizada a seleção e confecção dessas arquiteturas com adequação dos parâmetros ao conjunto de dados eleito. Para a classificação, utilizou-se das arquiteturas VGG16 e VGG19, descritas seção 2.9 e seção 2.10.

Optou-se por testar diversas topologias variando o número de épocas e a taxa de aprendizagem para otimizar o desempenho dos modelos de CNN. As variações no número de épocas entre 50 e 100 permitem observar como os modelos se comportam com diferentes níveis de treinamento, enquanto a taxa de aprendizagem ajustada entre 0,01 e 0,1 ajuda a identificar a melhor velocidade de ajuste dos pesos da rede neural. Esses parâmetros foram adotados por serem comumente utilizados na literatura para análises histológicas e por se mostrarem eficazes no contexto do modelo proposto.

A avaliação dessas topologias foi realizada nos dois modelos de CNN, levando em consideração tanto bancos de dados balanceados quanto desbalanceados, além de cenários de classificação em 4 classes (NILM, LSIL, HSIL e Carcinoma) e 2 Classes (Normal e Anormal). Essa abordagem abrangente permite entender como as variações nos parâmetros afetam o desempenho dos modelos em diferentes condições. Ao explorar esses diferentes cenários, busca-se identificar as configurações que proporcionam os melhores resultados.

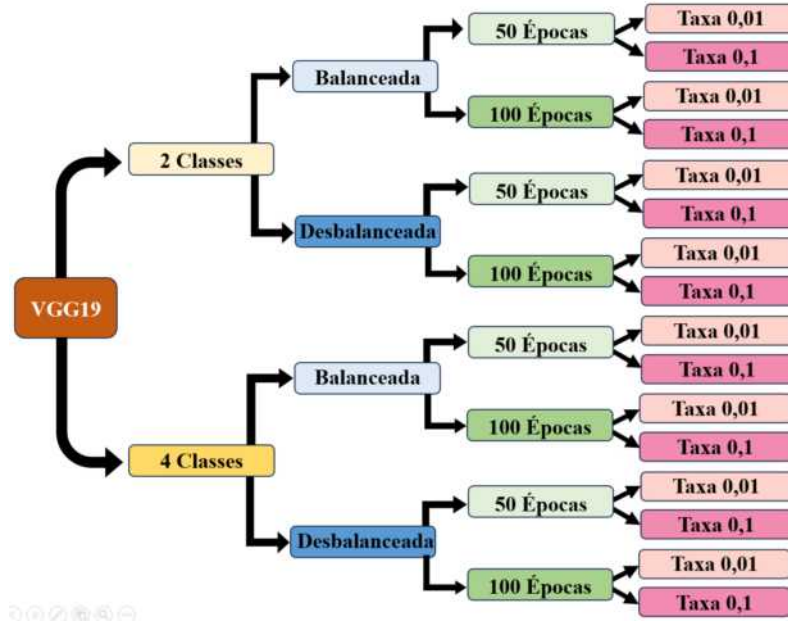
A relação estruturada das topologias em questão estão apresentadas na Figura 22 e Figura 23.

Figura 22 – Relação de Topologias testadas para a CNN VGG16



Fonte: Elaborado pela autora

Figura 23 – Relação de Topologias testadas para a CNN VGG19



Fonte: Elaborado pela autora

4.4 Avaliação de Performance

Após o treinamento e testes dos modelos, foi realizada a avaliação do desempenho através da Acurácia, Precisão, Sensibilidade e *F1-score*, que estão descritas na seção 2.11.

Analisar essas métricas como um todo permite considerar o contexto clínico e as prioridades específicas do problema médico para apontar a topologia adequada.

No entanto, para afunilar resultados, foram comparadas separadamente os resultados para as redes VGG16 e após isso para a VGG19. Para os seguintes moldes:

- Melhor Resultado para Classificação de 4 Classes para Dados Desbalanceados
- Melhor Resultado para Classificação de 2 Classes para Dados Desbalanceados
- Melhor Resultado para Classificação de 4 Classes para Dados Balanceados
- Melhor Resultado para Classificação de 2 Classes para Dados Balanceados

Para eleger esse melhor por sua vez usou-se a seguinte estratégia: escolher uma topologia com melhor acurácia e *F1-score*. Isso, porque a acurácia garante que a maioria das previsões está correta, refletindo um bom desempenho geral do modelo, enquanto o *F1-score*, sendo a média harmônica entre precisão e sensibilidade, assegura um equilíbrio entre a identificação correta dos casos positivos (importante para não perder diagnósticos críticos) e a minimização dos falsos positivos (importante para evitar tratamentos desnecessários).

Optou-se por apresentar nas tabelas de resultados todas as métricas, ainda que o foco principal tenha sido na acurácia e no *F1-score*, de forma a oferecer uma referência completa para futuros pesquisadores e garantir a transparência do estudo, facilitando sua replicação e verificação.

Após isso, foram comparados entre os resultados obtidos para os dois modelos de CNNs qual apresentavam melhor desempenho de classificação nesses moldes. Foi gerado então os Gráficos de precisão nos conjuntos de treinamento e validação para os melhores resultados para esses cenários.

4.5 Considerações Finais

Este capítulo teve como objetivo analisar os elementos das arquiteturas propostas no contexto da tarefa de classificação de resultados do exame papanicolau. Foi detalhado o fluxo e cada etapa integrante da solução: dataset, pré-processamento, confecção, treinamento e análise de desempenho dos modelos desenvolvidos. No Capítulo 5 serão apresentados os resultados obtidos através dos treinamentos para cada modelo e um comparativo do desempenho entre estes.

5 Resultados e Discussão

O presente capítulo apresentará os resultados obtidos a partir da aplicação da metodologia proposta neste trabalho no que tange a classificação de resultados do Exame Papanicolau. Os resultados obtidos através do treinamento dos modelos propostos, aplicados aos conjuntos de dados determinado, foram avaliados mediante as métricas de avaliação dispostas na seção 4.4.

Para facilitar as análises dos resultados foram separados quatro subseções contendo os resultados e análise das comparações das rede VGG16 e VGG19, para os 4 cenários mencionados na metodologia do trabalho.

5.1 Resultados para a Classificação com 4 Classes Desbalanceadas

Primeiramente, foram gerados os resultados para a classificação com 4 classes (NILM, LSIL, HSIL e Carcinoma) utilizando o banco de dados original desbalanceado, para 50 e 100 épocas, e com taxas de aprendizagem de 0.1 e 0.01 para a rede VGG16. Os resultados obtidos estão na Tabela 6.

Tabela 6 – Resultados de desempenho para a VGG16 com 4 Classes Desbalanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.75	0.70	0.67	0.68
50	0.01	0.72	0.67	0.64	0.65
100	0.1	0.74	0.69	0.66	0.67
100	0.01	0.73	0.68	0.65	0.66

Observa-se que, com 50 épocas e uma taxa de aprendizado de 0.1, a acurácia foi a maior (0.75), acompanhada de uma precisão de 0.70, sensibilidade de 0.67 e *F1-Score* de 0.68. Reduzindo a taxa para 0.01 nas mesmas 50 épocas, todos os valores das métricas caíram. Para 100 épocas, a acurácia e outras métricas melhoram um pouco em ambas as taxas, mas a taxa de 0.1 ainda apresenta melhores resultados gerais. Logo, uma taxa de aprendizado maior com menos épocas é mais eficaz neste cenário, possivelmente devido a uma convergência mais rápida do modelo.

Comparando os resultados para a mesma análise com a Rede VGG19, o padrão de resultados estão dispostos na Tabela 7.

Tabela 7 – Resultados de desempenho para a VGG19 com 4 Classes Desbalanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.77	0.72	0.68	0.70
50	0.01	0.78	0.73	0.69	0.71
100	0.1	0.79	0.74	0.70	0.72
100	0.01	0.80	0.75	0.71	0.73

Nesse caso a topologia mais eficaz é com 100 épocas e taxa de aprendizado de 0.01, proporcionando os melhores valores em todas as métricas, o que sugere que a VGG19, assim como a VGG16, se beneficia de um treinamento prolongado e uma taxa de aprendizado menor para alcançar um desempenho otimizado. Comparando os resultados da Tabela 7 com a Tabela 6, observa-se que a VGG19 consistentemente supera a VGG16 em termos de todas as métricas de desempenho em todas as topologias para a análise com 4 classes desbalanceadas. Essas diferenças podem ser atribuídas à maior profundidade e capacidade da VGG19.

5.2 Resultados para a Classificação com 2 Classes Desbalanceadas

No próximo cenário de análise, são apresentados os resultados para a classificação com 2 classes (Normal e Anormal), também utilizando o banco de dados original desbalanceado, para 50 e 100 épocas, e com taxas de aprendizagem de 0.1 e 0.01.

Os resultados obtidos para a rede VGG16 estão na Tabela 8 e as taxas de aprendizado 0.01 mostram aumento em todas as métricas com relação as taxas 0.1. A topologia com os melhores resultados é a que utiliza 100 épocas com taxa de 0.01, já que apresenta os valores mais altos em todas as métricas, indicando que uma taxa de aprendizado menor com mais épocas permite uma convergência mais estável e eficaz do modelo neste cenário.

Tabela 8 – Resultados de desempenho para a VGG16 com 2 Classes Desbalanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.76	0.75	0.74	0.74
50	0.01	0.78	0.77	0.76	0.76
100	0.1	0.80	0.79	0.78	0.78
100	0.01	0.82	0.81	0.80	0.80

Os resultados obtidos para este cenário pela Rede VGG19 estão na Tabela 9. Para

uma classificação binária com dados desbalanceados, a rede atinge uma acurácia de 0.88 e *F1-Score* de 0.83, superando os 0.82 e 0.80 da VGG16.

Tabela 9 – Resultados de desempenho para a VGG19 com 2 Classes Desbalanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.75	0.76	0.74	0.74
50	0.01	0.78	0.79	0.76	0.77
100	0.1	0.81	0.80	0.78	0.79
100	0.01	0.88	0.83	0.82	0.83

Isso indica que, embora ambas as arquiteturas se beneficiem de um treinamento prolongado e de uma taxa de aprendizado menor, a VGG19, com sua maior profundidade, consegue um ajuste mais preciso e um desempenho superior em cenários desbalanceados. Portanto, a VGG19 é mais vantajosa para esse cenário, proporcionando melhor precisão e estabilidade no desempenho.

5.3 Resultados para a Classificação com 4 Classes Balanceadas

As classificações seguintes evidenciam os resultados para a classificação usando o banco de dados balanceado, a Tabela 10 mostra os resultados da rede VGG16 testada com 4 classes (NILM, LSIL, HSIL e Carcinoma) enquanto a Tabela 11 apresenta os resultados da Rede VGG19

Tabela 10 – Resultados de desempenho para a VGG16 com 4 Classes Balanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.76	0.74	0.75	0.75
50	0.01	0.82	0.80	0.81	0.81
100	0.1	0.87	0.86	0.85	0.85
100	0.01	0.90	0.88	0.89	0.89

A partir dos resultados apresentados nas tabelas Tabela 10 e Tabela 11 para as redes VGG16 e VGG19, pode-se observar que ambas as redes atingem valores satisfatórios de acurácia e métricas relacionadas, com VGG19 ligeiramente superior em vários aspectos. Na Tabela 10, para a rede VGG16, com 100 épocas e uma taxa de aprendizado de 0.01, a acurácia chega a 0.90 e o *F1-Score* a 0.88, indicando um bom desempenho na classificação das quatro classes. Entretanto, na Tabela 11, os resultados para a rede VGG19 mostram

Tabela 11 – Resultados de desempenho para a VGG19 com 4 Classes Balanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.91	0.90	0.89	0.90
50	0.01	0.92	0.91	0.90	0.91
100	0.1	0.94	0.93	0.92	0.93
100	0.01	0.96	0.95	0.94	0.95

uma acurácia máxima de 0.96 e *F1-Score* de 0.95 , quando utilizada a mesma configuração de 100 épocas e taxa de aprendizado de 0.01. Isso sugere que a VGG19 tem uma capacidade ligeiramente melhor de identificar as classes corretamente em comparação com a VGG16, indicando uma maior capacidade de detecção de verdadeiros positivos nos cenários de classificações mais abrangentes com dados balanceados.

5.4 Resultados para a Classificação com 2 Classes Balanceadas

Por fim, as tabelas Tabela 12 e Tabela 13 apresentam os resultados das redes VGG16 e VGG19, respectivamente para o cenário de classificação binária com as classes balanceadas.

Tabela 12 – Resultados de desempenho para a VGG16 com 2 Classes Balanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.96	0.94	0.92	0.93
50	0.01	0.97	0.95	0.93	0.94
100	0.1	0.98	0.96	0.94	0.95
100	0.01	0.98	0.97	0.95	0.96

Tanto na classificação binária quanto na classificação de 4 classes, no cenário balanceado, a rede VGG 16 apresentou uma melhora progressiva à medida que o número de épocas aumentou e a taxa de aprendizado foi ajustada com maior precisão. No entanto, observou-se que o desempenho na classificação binária superou o da classificação com 4 classes. Esse resultado é esperado, uma vez que a tarefa de classificar um número menor de classes envolve menor complexidade, permitindo que o modelo alcance maior precisão e sensibilidade sob os mesmos parâmetros de treinamento. Em ambos os casos, a configuração com 100 épocas e uma taxa de aprendizado de 0,01 gerou os melhores resultados.

Esses resultados indicam que, para um cenário balanceado com menos classes, o modelo VGG16 se beneficia de um treinamento prolongado e uma taxa de aprendizado menor, permitindo uma convergência mais refinada e precisa.

Tabela 13 – Resultados de desempenho para a VGG19 com 2 Classes Balanceadas

Época	Taxa	Acurácia	Precisão	Sensibilidade	F1-Score
50	0.1	0.90	0.89	0.88	0.88
50	0.01	0.91	0.90	0.89	0.90
100	0.1	0.93	0.92	0.91	0.92
100	0.01	0.95	0.94	0.93	0.94

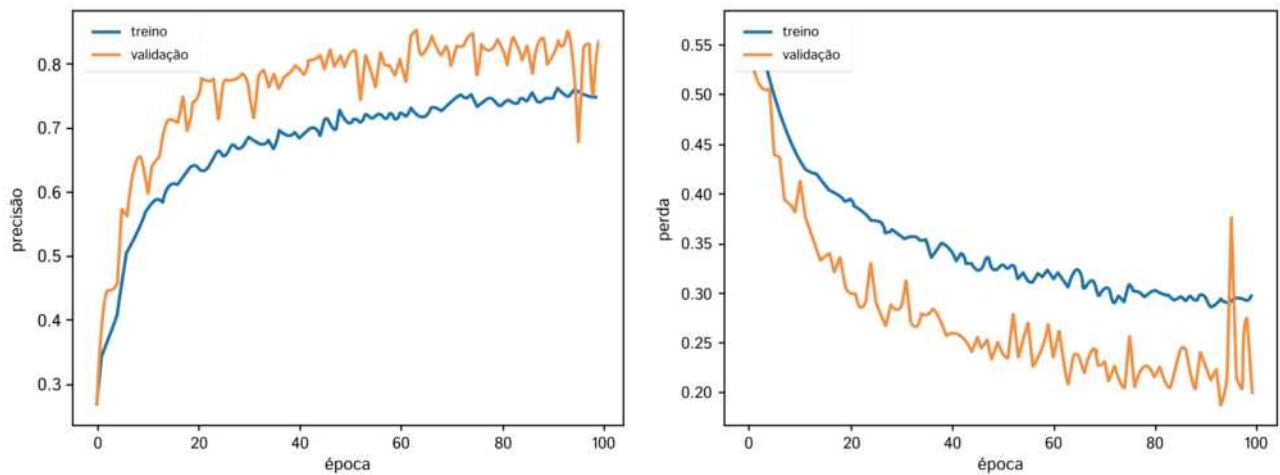
Comparando os resultados para este cenário das redes VGG16 e VGG19, a acurácia da VGG16 varia de 0.96 a 0.98 e o *F1-Score* de 0.93 a 0.96. Na VGG19 a acurácia varia de 0.90 a 0.95 e o *F1-Score* de 0.88 a 0.94. Observa-se que, enquanto os valores são relativamente próximos, a VGG16 apresenta resultados ligeiramente superiores em todas as métricas, especialmente com 100 épocas e uma taxa de aprendizado de 0.01. Isso sugere que a VGG16 é ligeiramente mais eficaz em cenários com classificação binária com classes balanceadas, possivelmente devido à menor complexidade do modelo que pode evitar *overfitting*.

5.5 Resultados Finais

Diante do exposto nas seções anteriores as topologias testadas foram eleitas com os melhores resultados para os cenários analisados. Para cada uma dessas foram desenvolvidas a representação na forma gráfica tanto nos estágios de treinamento quanto nos de teste. Os valores correspondentes de acurácia e perda para cada uma das CNNs eleitas estão representados nas Figuras correspondentes e indexadas em cada item que se segue como norteamto dos resultados.

- Melhor Resultado para classificação de 4 Classes para Dados Desbalanceados: VGG19, 100 épocas, taxa 0.01 (Acurácia 0.80, F1-Score 0.73)

Figura 24 – Resultado para classificação de 4 Classes para Dados Desbalanceados



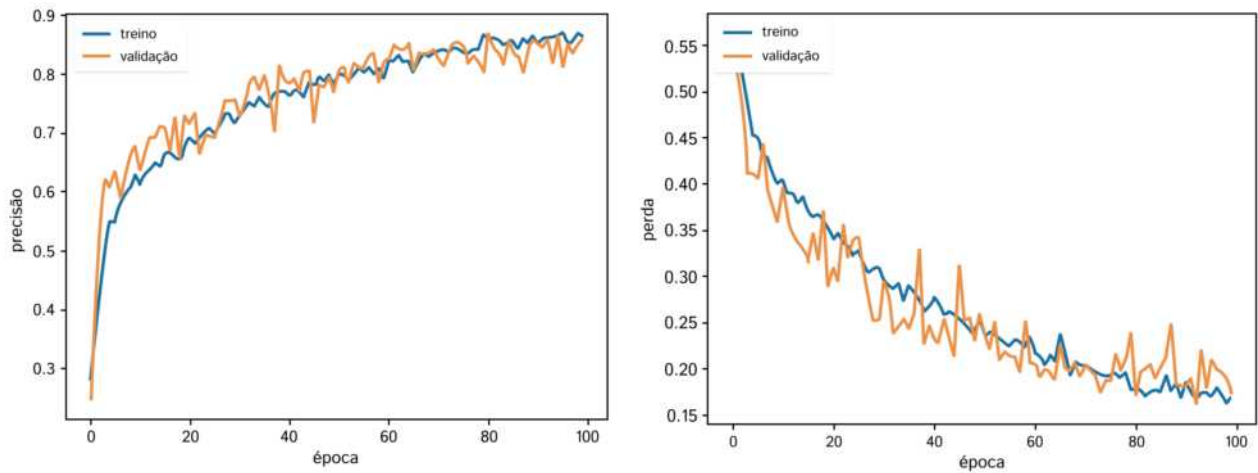
Fonte: Elaborado pela autora

O gráfico à esquerda mostra que, ao longo das épocas, tanto o conjunto de treino quanto o de validação melhoraram em termos de precisão. A curva de precisão para o conjunto de validação aumenta rapidamente nas primeiras épocas, mas a partir de certo ponto, ela se estabiliza e começa a oscilar em torno de 0,8. Isso sugere que o modelo teve uma fase inicial de aprendizado rápido, mas enfrentou dificuldades em continuar melhorando a generalização para os dados de validação.

O gráfico à direita mostra um padrão semelhante, com a perda de validação diminuindo inicialmente, mas caindo de forma mais rápida que a perda do treino, indicando que o modelo se ajusta mais rapidamente aos dados de validação. No entanto, há uma queda constante na qualidade da generalização, com a perda de validação sendo consistentemente maior que a de treino. Essa diferença pode indicar *overfitting*, onde o modelo se ajusta excessivamente aos dados de treino, sacrificando o desempenho em dados que ele não viu.

- Melhor Resultado para classificação de 2 Classes para Dados Desbalanceados: VGG19, 100 épocas, taxa 0.01 (Acurácia 0.88, F1-Score 0.83)

Figura 25 – Resultado para classificação de 2 Classes para Dados Desbalanceados



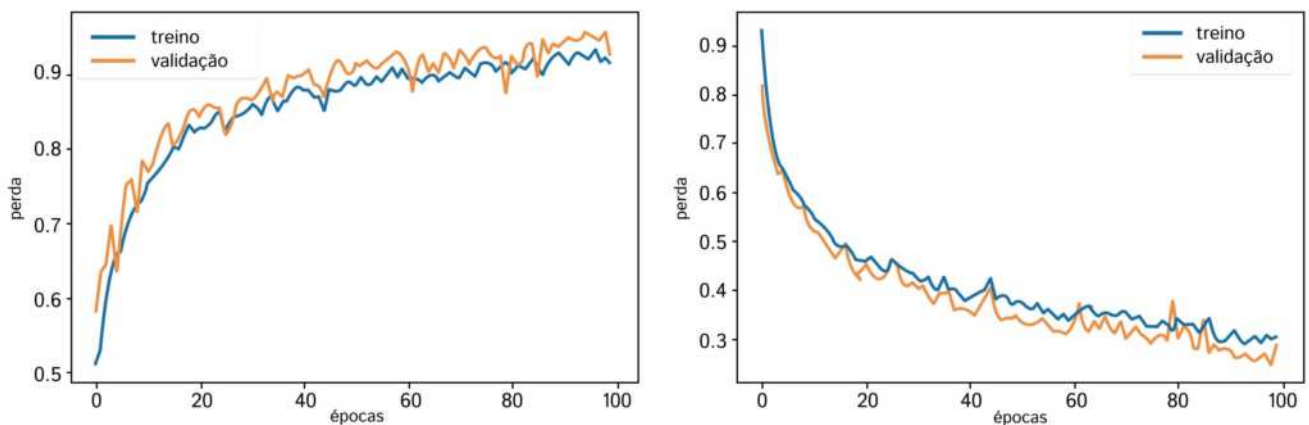
Fonte: Elaborado pela autora

Comparando este gráfico com o de 4 classes, percebe-se uma diferença mais significativa na convergência da precisão de treino e validação. Tanto a curva de treino quanto a de validação sobem juntas e de maneira mais uniforme, estabilizando em torno de 0.85. Isso mostra que a classificação binária, devido à sua menor complexidade, permitiu que o modelo aprendesse de maneira mais eficaz e generalizasse melhor.

A perda para a classificação binária também diminui de forma mais consistente ao longo do tempo, com uma diferença relativamente pequena entre o treino e a validação, o que reflete um bom ajuste do modelo. Não há sinais tão claros de *overfitting* como no caso de 4 classes.

- Melhor Resultado para Classificação de 4 Classes para Dados Balanceados: VGG19, 100 épocas, taxa 0.01 (Acurácia 0.96, F1-Score 0.95)

Figura 26 – Resultado para classificação de 4 Classes para Dados Balanceados



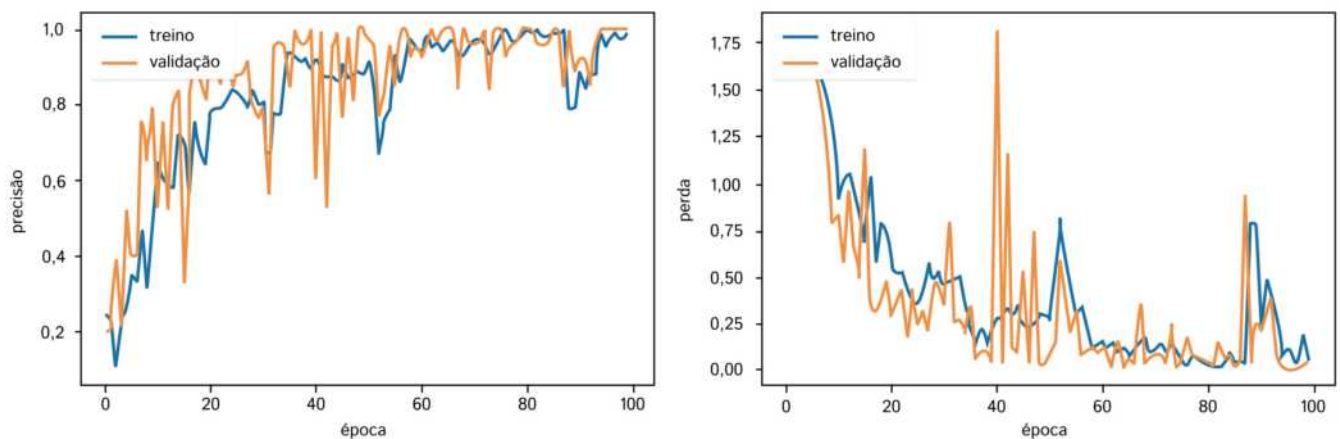
Fonte: Elaborado pela autora

No cenário balanceado com 4 classes, observa-se uma melhoria significativa na precisão, com as curvas de treino e validação subindo juntas de maneira mais consistente em comparação ao cenário desbalanceado. A precisão final de validação fica em torno de 0.9, indicando que o modelo conseguiu aprender melhor com dados balanceados, o que é esperado, dado que a ausência de classes majoritárias ou minoritárias tende a melhorar a capacidade do modelo de generalizar bem para todas as classes.

A perda tanto no treino quanto na validação é muito mais próxima do que no cenário desbalanceado. Aqui, o modelo parece estar menos sujeito ao *overfitting*, pois as curvas seguem trajetórias semelhantes, sugerindo um melhor ajuste e generalização.

- Melhor Resultado para Classificação de 2 Classes para Dados Balanceados: VGG16, 100 épocas, taxa 0.01 (Acurácia 0.98, F1-Score 0.96)

Figura 27 – Resultado para classificação de 2 Classes para Dados Balanceados



Fonte: Elaborado pela autora

Para o cenário balanceado com 2 classes, as curvas de precisão tanto no treino quanto na validação oscilam bastante ao longo do tempo, especialmente no caso da validação. Isso pode indicar instabilidade durante o treinamento, possivelmente devido à dificuldade de generalizar perfeitamente entre as classes no caso balanceado. Apesar das oscilações, a precisão permanece alta, sugerindo que o modelo ainda foi capaz de manter um bom desempenho médio, mesmo com essas flutuações.

O gráfico de perda reflete bem a instabilidade vista na precisão, com a perda de validação subindo e descendo ao longo das épocas. Essa variabilidade pode ser um sinal de que o modelo está sendo sensível a pequenas variações nos dados de validação ou pode estar enfrentando dificuldades em convergir de forma estável neste cenário balanceado.

De modo geral, os resultados mostrados na Figura 24, Figura 25, Figura 26 e Figura 27 reforçam que o desempenho das arquiteturas VGG16 e VGG19 varia conforme

o número de classes e o balanceamento dos dados. Em cenários com menos classes, ambas as arquiteturas tendem a alcançar uma performance superior, evidenciando que a complexidade reduzida facilita a tarefa de classificação. Além disso, os dados balanceados favorecem melhores resultados, uma vez que o modelo pode aprender de maneira mais uniforme as características das diferentes classes. Bem como, um treinamento prolongado permite ao modelo ajustar-se mais precisamente aos padrões dos dados.

Em termos de comparação entre VGG16 e VGG19, a VGG19, geralmente, apresenta desempenho superior em cenários mais complexos e com dados desbalanceados. Entretanto, a VGG16 ainda demonstra robustez e pode ser preferível em tarefas menos complexas ou com dados balanceados, devido à sua eficiência computacional e menor propensão ao *overfitting*. Assim, a escolha da arquitetura ideal deve considerar tanto a complexidade da tarefa quanto a natureza dos dados disponíveis. Para facilitar essa análise foi sumarizado os principais resultados na Tabela 14.

Tabela 14 – Resultados de Desempenho Sintetizados

Classificação Final	Arquitetura	Épocas	Taxa
Melhor Resultado para Classificação de 4 Classes para Dados Desbalanceados	VGG 19	100	0.01
Melhor Resultado para Classificação de 2 Classes para Dados Desbalanceados	VGG 19	100	0.01
Melhor Resultado para Classificação de 4 Classes para Dados Balanceados	VGG 19	100	0.01
Melhor Resultado para Classificação de 2 Classes para Dados Balanceados	VGG 16	100	0.01

Por fim, para a implementação da ferramenta, a topologia mais adequada seria a VGG19. Essa escolha é fundamentada nos resultados obtidos, que indicam a superioridade da VGG19 em diferentes cenários de classificação, especialmente ao lidar com dados desbalanceados e conjuntos com 4 classes. A arquitetura VGG19 demonstrou uma maior capacidade de processamento e melhor acurácia, conseguindo manter um equilíbrio entre precisão, sensibilidade e F1-Score. Isso é essencial para diagnósticos precisos em ambientes com grande variação nos dados. Vale destacar que, antes da implementação, será necessário realizar ajustes finos nos modelos e testes rigorosos para verificar e mitigar possíveis casos de *overfitting*, garantindo, assim, uma análise mais confiável.

6 Conclusões

O câncer cervical é uma das patologias que mais matam mulheres no mundo, e como a única maneira de evitar essa morbidade é fazendo a detecção precoce da doença, torna-se vital o estudo de métodos voltados para essa detecção de forma eficiente, permitindo, portanto, um tratamento mais oportuno e eficaz. Nesse sentido, o uso de tecnologias que venham a complementar o diagnóstico humano se torna imprescindível.

Neste trabalho mostrou-se uma solução potencial baseada em inteligência artificial para complementar o diagnóstico tradicional. A problemática proposta visou a classificação de resultados do exame Papanicolau utilizando redes neurais convolucionais, especificamente as arquiteturas VGG16 e VGG19. O estudo se concentrou em avaliar a eficiência dessas redes na distinção de diferentes classes de anomalias cervicais, tanto em cenários de dados balanceados, quanto desbalanceados. A análise se baseou em métricas de desempenho como acurácia, precisão, sensibilidade e *F1-Score*, para determinar a eficácia dos modelos em contextos clínicos típicos.

Os resultados obtidos indicam que ambas as arquiteturas apresentam um desempenho bom, com a VGG19 demonstrando uma ligeira superioridade em cenários mais complexos, especialmente com dados desbalanceados. A maior profundidade da VGG19 parece contribuir para uma melhor capacidade de representação, permitindo ao modelo ajustar-se mais precisamente aos padrões presentes nos dados. No entanto, a VGG16, apesar de ser menos complexa, mostrou-se eficaz em cenários com dados balanceados e menos classes.

Constatou-se que independente da arquitetura, configurações que envolvem um maior número de épocas e uma taxa de aprendizado menor consistentemente resultaram em melhores métricas de desempenho. Além disso, os resultados destacam a importância do balanceamento dos dados para a eficácia da classificação. Os cenários com dados balanceados apresentaram métricas de desempenho significativamente superiores, reforçando a necessidade de um pré-processamento dos dados para assegurar que o modelo possa aprender de maneira uniforme as características das diferentes classes. A classificação binária (Normal e Anormal) também se mostrou mais eficaz, devido à menor complexidade da tarefa.

Em conclusão, é evidente que a escolha da arquitetura ideal para a classificação de resultados do exame Papanicolau deve considerar a complexidade da tarefa e a natureza dos dados disponíveis. A VGG19 é mais adequada para cenários complexos e desbalanceados, enquanto a VGG16 é preferível em tarefas menos complexas ou com dados balanceados.

Logo, os resultados mostraram que o uso de CNNs pode levar a um rastreamento au-

tomático e eficaz da doença. Salienta-se, por sua vez, que independente do resultado obtido com o preventivo, o diagnóstico médico só será fechado a partir de exames complementares. Destacando, assim, a importância do acompanhamento de profissionais especializados para a conclusão diagnóstica e orientação para o melhor plano de tratamento.

6.1 Limitações e Dificuldades Encontradas

Durante a realização deste estudo, algumas limitações e dificuldades foram enfrentadas e estão elencadas abaixo:

- A quantidade reduzida de banco de dados existentes de citologia cervical dificulta análise comparativa de desempenho com outros bancos.
- Em comparação com outras áreas da medicina, como a detecção de diversas patologias, a quantidade de imagens disponíveis publicamente para estudos histopatológicos é extremamente limitada.
- A classificação correta das imagens citológicas exige um conhecimento especializado profundo, o que dificulta a avaliação da eficácia e da confiabilidade dos resultados gerados pela ampliação do conjunto de dados.

6.2 Trabalhos Futuros

Para dar continuidade à pesquisa desenvolvida, a seguir são apresentados possíveis direcionamentos para investigações futuras:

- Criar um conjunto de dados próprio com uma quantidade mais significativa de imagens para melhorar a representatividade e a robustez do modelo;
- Realizar uma investigação detalhada sobre a otimização dos hiperparâmetros dos modelos propostos, utilizando a biblioteca *Keras*, visando aprimorar a performance;
- Desenvolver novos modelos baseados em diferentes CNNs, com ênfase em redes *ResNets* com profundidades superiores a 50 camadas, para aumentar a precisão das classificações;
- Desenvolver modelos que incorporem algoritmos de detecção e segmentação de objetos, adaptados às necessidades específicas da problemática médica;
- Desenvolver uma solução *web* ou *mobile* para permitir testes práticos em ambiente clínico, facilitando a validação e a implementação da tecnologia em uso real.

- Explorar métodos de verificação e mitigação de *overfitting* e realizar testes futuros aplicando técnicas avançadas de regularização, como *Dropout*, *Early Stopping*, e *L2 Regularization*, para minimizar os efeitos do *overfitting* observados em alguns modelos, como forma de garantir que o modelo generalize melhor para novos dados, aumentando a confiabilidade das classificações em cenários clínicos e com variação significativa nos dados.

6.3 Trabalhos Publicados

Publicações realizadas relativas à pesquisa durante o mestrado:

Costa, E. S. S., Ferreira, S. C. B., Assumpção, P. P., Araújo, J. P. L. de, Cardoso, E. H. S., Francês, C. R. L. (2023). Utilização da Rede Neural VGG16 para classificação citológica. Anais do Congresso Interdisciplinar de Ciências, Tecnologia e Informação em Saúde (COINTES), 89-91. ISBN: 978-65-85179-29-4. DOI: 10.29327/1312822.

Costa, E. S. S., Ferreira, S. C. B., Melo, I. J. da S., Souza, J. D. de, Araújo, J. P. L. de, Cardoso, E. H. S., Guerreiro, J. F., Assumpção, P. P., Francês, C. R. L. (2023). Classificação de Imagens de Exames de Preventivo Usando Rede Neural Convolucional. Anais do XIII Congresso Nacional em Comunicações, Redes e Segurança da Informação (ENCOM), 13, 31-32.

Referências

- AKBAR, A. H. et al. Layer selection on residual network for feature extraction of pap smear images. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, v. 36, n. 2, p. 56 – 66, 2024. Cited by: 0; All Open Access, Hybrid Gold Open Access. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85181455469&doi=10.37934%2ffaraset.36.2.5666&partnerID=40&md5=c706966e3beef8b4bc69bdf0f5bb0490>>. Citado 2 vezes nas páginas 34 e 36.
- AL-FRAIHAT, D. et al. Speech recognition utilizing deep learning: A systematic review of the latest developments. *HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES*, KOREA INFORMATION PROCESSING SOC 1002HO YONGSUNGBIZTEL 314-1 2GA HANKANGRO ..., v. 14, 2024. Citado na página 14.
- ALLAHYARI, E.; MOODI, M.; TAHERGORABI, Z. Artificial neural networks (anns) for modeling efficient factors in predicting pap smear screening behavior change stage. *BioMedicine*, China Medical University, v. 12, n. 2, p. 10, 2022. Citado na página 2.
- ALOMAR, K.; AYSEL, H. I.; CAI, X. Data augmentation in classification and segmentation: A survey and new strategies. *Journal of Imaging*, MDPI, v. 9, n. 2, p. 46, 2023. Citado 7 vezes nas páginas 16, 18, 19, 20, 21, 24 e 25.
- ALRAJJAL, A. et al. Squamous intraepithelial lesions (sil: Lsil, hsil, ascus, asc-h, lsil-h) of uterine cervix and Bethesda system. *Cytojournal*, Cytopathology Foundation, Inc., v. 18, 2021. Citado na página 8.
- ALREFAI, E. A.; ALHEJAILI, R. T.; HADDAD, S. A. Human papillomavirus and its association with cervical cancer: A review. *Cureus*, Cureus Inc., v. 16, n. 4, 2024. Citado na página 8.
- ALSALATIE, M. et al. Analysis of cytology pap smear images based on ensemble deep learning approach. *Diagnostics*, v. 12, n. 11, 2022. ISSN 2075-4418. Disponível em: <<https://www.mdpi.com/2075-4418/12/11/2756>>. Citado 2 vezes nas páginas 33 e 36.
- AMIN, R. et al. Healthcare techniques through deep learning: issues, challenges and opportunities. *IEEE Access*, IEEE, v. 9, p. 98523–98541, 2021. Citado na página 3.
- ARCANJO, J. *Deep Learning: How to apply Transfer Learning to your projects*. 2023. Acessado em: [data de acesso]. Disponível em: <<https://medium.com/jonys-arcujo/deep-learning-como-aplicar-transfer%C3%A2ncia-de-aprendizagem-em-seus-projetos-1b3fcc6c759>>. Citado na página 23.
- BANERJEE, D. et al. Screening technologies for cervical cancer: Overview. *Cytojournal*, Cytopathology Foundation, Inc., v. 19, 2022. Citado 2 vezes nas páginas 1 e 8.
- BODDU AV KANUMURI, D. T. C. R. S. Fake images detection: A comparative study using cnn and vgg-16 models. *International Journal of Advanced Engineering and Management*, 2023. Disponível em: <https://ijaem.net/issue_dcp/Fake%20Images%20Detection%20%20A%20Comparative%20Study%20Using%20CNN%20and%20VGG%2016%20Models.pdf>. Citado na página 25.

- BONFIGLIO, T. A. Gynecologic cytopathology: Historical perspective, current status, and future outlook. *AJSP: Reviews & Reports*, LWW, v. 10, n. 3, p. 98–105, 2005. Citado na página 6.
- CANEPA, M.; PATEL, N. R.; GARCIA-MOLINER, M. L. Pathology and molecular diagnosis of cervical cancer and precursor lesions. *Uterine Cervical Cancer: Clinical and Therapeutic Perspectives*, Springer, p. 61–87, 2019. Citado na página 11.
- CAPES, P. P. et al. Portal periódicos capes. Disponível em: <http://www.periodicos.capes.gov.br/ez88.periodicos.capes.gov.br/index.php>, 2020. Citado 2 vezes nas páginas 31 e 32.
- CHAUDHARY, P. S.; KHURANA, M. R.; AYALASOMAYAJULA, M. Real-world applications of data analytics, big data, and machine learning. In: *Data Analytics and Machine Learning: Navigating the Big Data Landscape*. [S.l.]: Springer, 2024. p. 237–263. Citado na página 13.
- CHEN, W. et al. Lightweight convolutional neural network with knowledge distillation for cervical cells classification. *Biomedical Signal Processing and Control*, v. 71, 2022. Cited by: 26. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85115121442&doi=10.1016%2fj.bspc.2021.103177&partnerID=40&md5=a6dda92466876947337e5efd1ad45ce9>. Citado 2 vezes nas páginas 34 e 36.
- CHEN, W. et al. Improving computer-aided cervical cells classification using transfer learning based snapshot ensemble. *Applied Sciences (Switzerland)*, v. 10, n. 20, p. 1 – 14, 2020. Cited by: 22; All Open Access, Gold Open Access. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092719947&doi=10.3390%2fapp10207292&partnerID=40&md5=d966e19bb983c7bb515fde9172123661>. Citado 2 vezes nas páginas 33 e 36.
- CHEN, Y. et al. Combining deep learning and mathematical morphology for historical map segmentation. In: SPRINGER. *International Conference on Discrete Geometry and Mathematical Morphology*. [S.l.], 2021. p. 79–92. Citado na página 16.
- CHIKHAOUI, E.; ALAJMI, A.; LARABI-MARIE-SAINTTE, S. Artificial intelligence applications in healthcare sector: ethical and legal challenges. *Emerging Science Journal*, v. 6, n. 4, p. 717–738, 2022. Citado na página 3.
- CHINTALA, S. Data privacy and security challenges in ai-driven healthcare systems in india. *Journal of Data Acquisition and Processing*, v. 37, n. 5, p. 2769–2778, 2022. Citado na página 3.
- COFINO, C. L.; ESCORIAL, R. B.; ENQUILINO, D. L. B. A literature review on natural language processing (nlp) in aiding industry to progress. *Journal Name*, 2024. Citado na página 12.
- CONCIATORI A VALLETTA, A. S. M. Improving the quality evaluation process of machine learning algorithms applied to landslide time series analysis. *Computers & Geosciences*, Elsevier, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0098300424000141>. Citado na página 29.

CRUZ, J. S. An ensemble approach for multi-stage transfer learning models for covid-19 detection from chest ct scans. *Intelligence-Based Medicine*, Elsevier, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S266652122100003X>>. Citado na página 22.

DAWOOD, A. A comparative study using deep learning models and transfer learning for detection and classification of alzheimer's disease. *International Journal of Computer Applications*, 2023. Citado na página 26.

DEPOSITPHOTOS. *Machine vs Deep Learning as Artificial Intelligence Principle Outline Diagram*. 2024. Accessed: April 23, 2024. Disponível em: <<https://depositphotos.com/br/vector/machine-vs-deep-learning-as-artificial-intelligence-principle-outline-diagram-535502116.html>>. Citado 2 vezes nas páginas 12 e 14.

DESIANI, A. et al. Bi-path architecture of cnn segmentation and classification method for cervical cancer disorders based on pap-smear images. *IAENG International Journal of Computer Science*, v. 48, n. 3, 2021. Citado na página 2.

ECK, N. van; WALTMAN, L. Software survey: Vosviewer, um programa de computador para mapeamento bibliométrico. *Cienciometria*, v. 84, n. 2, p. 523–538, August 2010. EPub 2009 31 de dezembro. Citado na página 31.

FARUQUI, N. et al. Healthcare as a service (haas): Cnn-based cloud computing model for ubiquitous access to lung cancer diagnosis. *Heliyon*, Elsevier, v. 9, n. 11, 2023. Citado na página 2.

GAJJAR, D. Artificial intelligence: An explainer. *Artificial intelligence*, 2023. Citado 3 vezes nas páginas 11, 12 e 14.

GENNIGENS, C. et al. Recurrent or primary metastatic cervical cancer: current and future treatments. *ESMO open*, Elsevier, v. 7, n. 5, p. 100579, 2022. Citado na página 11.

HE, M. et al. Image segmentation of a sewer based on deep learning. *Sustainability*, MDPI, v. 14, n. 11, p. 6634, 2022. Citado 8 vezes nas páginas 15, 18, 19, 20, 21, 25, 27 e 28.

HULL, R. et al. Cervical cancer in low and middle-income countries. *Oncology letters*, Spandidos Publications, v. 20, n. 3, p. 2058–2074, 2020. Citado na página 1.

HUSSAIN, E. *Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer*. Mendeley Data, 2019. Disponível em: <<https://data.mendeley.com/datasets/zddtpgzv63/3>>. Citado na página 40.

IBANGA, J. Resolving data imbalance in financial fraud detection by combining machine learning models and ensemble learning strategies. *ResearchGate*, 2024. Disponível em: <https://www.researchgate.net/profile/Joseph-Ibanga-4/publication/380877850_Resolving_data_imbalance_in_financial_fraud_detection_by_combining_machine_learning_links/66527133bc86444c72018ad5/Resolving-data-imbalance-in-financial-fraud-detection-by-combining-machine-learning-models-and-ensemble-learning-strategies.pdf>. Citado na página 29.

(INCA), N. C. I. Adenocarcinomas of the uterine cervix, endometrium, and their main benign differential diagnoses. 2023. Accessed: April 23, 2024. Disponível em: <<https://ninho.inca.gov.br/jspui/bitstream/123456789/2232/3/Adenocarcinomas%20do%20cervix%20e%20endometrio.pdf>>.

20do%20colo%20uterino%2C%20de%20endom%20C3%A9trio%20e%20seus%20principais%20diagn%C3%B3sticos%20diferenciais%20ben%C3%ADgnos.pdf>. Citado 5 vezes nas páginas 1, 8, 9, 10 e 11.

Instituto Nacional de Câncer (INCA). *Dados e Números do Câncer de Colo do Útero*. 2022. Acesso em: 18 de junho de 2024. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/dados_e_numeros_colo_22setembro2022.pdf>. Citado na página 10.

JOYNAB, N. S. et al. A federated learning aided system for classifying cervical cancer using pap-smear images. *Informatics in Medicine Unlocked*, v. 47, 2024. Cited by: 0; All Open Access, Gold Open Access. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85190724833&doi=10.1016%2fj.imu.2024.101496&partnerID=40&md5=2df7c12021ded0b8dd7d984883163c87>>. Citado 2 vezes nas páginas 34 e 36.

JUG RACHEL M.B.B.CH., B.; BEAN, S. M. M. *Cervix Cytology and the Bethesda System*. 2023. Last author update: 3 May 2023. Last staff update: 3 May 2023. Page views in 2023: 46,004. Citado na página 9.

KALBHOR, M. et al. Pap smear-based cervical cancer detection using hybrid deep learning and performance evaluation. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, v. 11, n. 5, p. 1615–1624, 2023. Citado 2 vezes nas páginas 1 e 2.

KARKUZHALI R MURUGESHWARI, V. U. S. Facial emotion recognition and synthesis with convolutional neural networks. *International Journal of ...*, 2021. Disponível em: <https://www.researchgate.net/profile/Karkuzhali-S/publication/379188944_Facial_Emotion_Recognition_and_Synthesis_with_Convolutional_Neural_Networks/links/65fe6424a4857c79626d226a/Facial-Emotion-Recognition-and-Synthesis-with-Convolutional-Neural-Networks.pdf>. Citado na página 16.

KIPTOO, S.; NDERU, L.; MUTANU, L. Automated detection of cervical pre-cancerous lesions using regional-based convolutional neural network. In: *ACSE*. [S.l.: s.n.], 2020. Citado na página 2.

KOSS, L. G. The papanicolaou test for cervical cancer detection: a triumph and a tragedy. *Jama*, American Medical Association, v. 261, n. 5, p. 737–743, 1989. Citado na página 6.

KRAUSS, P. et al. Analysis and visualization of sleep stages based on deep neural networks. *Neurobiology of sleep and circadian rhythms*, Elsevier, v. 10, p. 100064, 2021. Citado na página 14.

LI, K. Applications of deep learning in object detection. In: IEEE. *2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE)*. [S.l.], 2022. p. 436–442. Citado na página 15.

LIMA, C. R. de et al. Mask region-based cnns for cervical cancer progression diagnosis on pap smear examinations. *Heliyon*, Elsevier, v. 9, n. 11, 2023. Citado 2 vezes nas páginas 33 e 36.

MACIOS, A.; NOWAKOWSKI, A. False negative results in cervical cancer screening—risks, reasons and implications for clinical practice and public health. *Diagnostics*, MDPI, v. 12, n. 6, p. 1508, 2022. Citado na página 1.

- MAHMOOD, F. et al. Deep learning with cinematic rendering: fine-tuning deep neural networks using photorealistic medical images. *Physics in Medicine & Biology*, IOP Publishing, v. 63, n. 18, p. 185012, 2018. Citado na página 3.
- MATHWORKS. *Transfer Learning*. 2019. Acessado em: [data de acesso]. Disponível em: <<https://www.mathworks.com/discovery/transfer-learning.html>>. Citado na página 24.
- MBUNGE, E.; BATANI, J. Application of deep learning and machine learning models to improve healthcare in sub-saharan africa: Emerging opportunities, trends and implications. *Telematics and Informatics Reports*, Elsevier, p. 100097, 2023. Citado na página 3.
- MEMARIAN, B.; DOLECK, T. A scoping review of reinforcement learning in education. *Computers and Education Open*, Elsevier, p. 100175, 2024. Citado na página 13.
- MERRICK, M. et al. Screening for cervical dysplasia and cancer in adults with hiv. *Medical Care*, v. 2, 2022. Citado na página 7.
- MOHAPATRA, S. T. U. An efficient convolutional neural network-based classifier for an imbalanced oral squamous carcinoma cell dataset. *International Journal of Artificial Intelligence*, 2024. Disponível em: <<https://pdfs.semanticscholar.org/fa1c/291c12b525b01aa73c061ac908a5af03c0f0.pdf>>. Citado na página 28.
- MUHAMMAD K KIANI, T. M. N. T. A. Skinlesnet: Classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional neural network. *Cancers*, 2023. Disponível em: <<https://salford-repository.worktribe.com/OutputFile/2850276>>. Citado na página 22.
- MUN, S. K. et al. Artificial intelligence for the future radiology diagnostic service. *Frontiers in molecular biosciences*, Frontiers Media SA, v. 7, p. 614258, 2021. Citado 2 vezes nas páginas 3 e 16.
- MUÑOZ, N.; BOSCH, F. X. The causal link between hpv and cervical cancer and its implications for prevention of cervical cancer. *Bulletin of the Pan American Health Organization (PAHO)*; 30 (4), dec. 1996, 1996. Citado na página 6.
- MURTHY HV CHAITRA, E. V. B. M. G. A novel transfer learning approach to improve breast cancer diagnosing on screening mammography. *International Journal of Engineering and Emerging Research*, 2021. Disponível em: <<https://ijeer.forexjournal.co.in/papers-pdf/ijeer-120141.pdf>>. Citado na página 26.
- NJUE, J. K. et al. Primary and triage cervical screening diagnostic value of methods for the detection of cervical dysplasia. *BioMed Research International*, Wiley Online Library, v. 2022, n. 1, p. 1930102, 2022. Citado na página 7.
- NWANKWO, C. O.; IHUEZE, C. C. Corrosion rate models for oil and gas pipeline systems a numerical approach. *International Journal of Engineering Research and Technology*, 2018. Citado na página 18.
- ORGANIZATION, W. H. et al. *Accelerating the elimination of cervical cancer as a public health problem: Towards achieving 90–70–90 targets by 2030*. [S.l.], 2022. Citado na página 6.
- PANGARKAR, M. A. The bethesda system for reporting cervical cytology. *Cytojournal*, Cytopathology Foundation, Inc., v. 19, 2022. Citado na página 7.

- PAUL, W. et al. Defending medical image diagnostics against privacy attacks using generative methods: Application to retinal diagnostics. In: *Privacy-Preserving Machine Learning*. Springer, 2021. p. 389–412. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-030-90874-4_17>. Citado na página 16.
- PERKINS, R. B. et al. 2019 asccp risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors. *Journal of lower genital tract disease*, LWW, v. 24, n. 2, p. 102–131, 2020. Citado na página 9.
- PORCHERI, C. e. a. Multifactorial contribution of notch signaling in head and neck squamous cell carcinoma. *International journal of molecular sciences*, MDPI, v. 20, n. 6, p. 1520, 2019. Citado na página 10.
- PUNHANI, P.; AHLUWALIA, C. Biomarkers in the screening and management of cervical cancer. *Journal of Colposcopy and Lower Genital Tract Pathology*, LWW, v. 2, n. 1, p. 6–14, 2024. Citado na página 7.
- RAMOS, J. L. S. et al. Health services, socioeconomic indicators, and primary care coverage in mortality by lower genital tract and breast neoplasias in brazilian women during reproductive and non-reproductive periods. *International journal of environmental research and public health*, MDPI, v. 17, n. 16, p. 5804, 2020. Citado na página 10.
- RAPARTHI, M.; DODDA, S. B.; MARUTHI, S. Ai-enhanced imaging analytics for precision diagnostics in cardiovascular health. *European Economic Letters (EEL)*, v. 11, n. 1, 2021. Citado na página 3.
- RODRÍGUEZ, G. et al. Cervical cancer prevention by vaccination. *Frontiers in Oncology*, Frontiers Media SA, v. 14, 2024. Citado na página 6.
- RUBACK, L.; AVILA, S.; CANTERO, L. Machine learning biases and their social implications: A case study in facial recognition. In: . [S.l.: s.n.], 2021. p. 90–101. Citado na página 28.
- SANTOS, F. dos. *Machine Learning, Inovação Angolana, AI Explain*. 2024. Accessed: May 17, 2024. Disponível em: <https://pt.linkedin.com/posts/fidel-dos-santos-6b197928b_machinelearning-inova%C3%A7%C3%A3oangolana-aiexplain-activity-7151410049804894209-XPsa>. Citado na página 15.
- SHAHIN, M. et al. Deploying deep convolutional neural network to the battle against cancer: Towards flexible healthcare systems. *Informatics in Medicine Unlocked*, Elsevier, v. 47, p. 101494, 2024. Citado 2 vezes nas páginas 2 e 17.
- SHAW, P. A. The history of cervical screening i: the pap. test. *Journal SOGC*, Elsevier, v. 22, n. 2, p. 110–114, 2000. Citado na página 6.
- SILVA, F. G. d. Evaluation of machine learning techniques in the classification of maritime targets in c-band sar images. 2022. Citado 2 vezes nas páginas 26 e 27.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 25.
- SOUSA, L. V. d. A. et al. Inequalities in mortality and access to hospital care for cervical cancer—an ecological study. *International Journal of Environmental Research and Public Health*, MDPI, v. 18, n. 20, p. 10966, 2021. Citado na página 10.

- SOUSA, L. V. d. A. et al. Trends in hospitalization and mortality from cervical cancer in brazil are linked to socioeconomic and care indicators. *Women*, MDPI, v. 2, n. 3, p. 274–284, 2022. Citado na página 1.
- STUDER, S. et al. Towards crisp-ml (q): a machine learning process model with quality assurance methodology. *Machine learning and knowledge extraction*, MDPI, v. 3, n. 2, p. 392–413, 2021. Citado na página 12.
- TING, D. et al. Deep learning models for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *The Lancet Digital Health*, v. 2, n. 1, p. e20–e30. Citado na página 29.
- UNZUETA, D. Convolutional layers vs. fully connected layers. *Towards Data Science*, 2021. Citado na página 20.
- VOIDĂZAN, T. S. et al. The importance of combined hpv and cintec® plus genotyping testing for p16 in women with cervical squamous cell carcinoma. *Romanian Journal of Morphology and Embryology*, Romanian Academy Publishing House, v. 65, n. 1, p. 99, 2024. Citado na página 9.
- VOSviewer. *VOSviewer: Visualizing scientific landscapes*. 2024. <<https://www.vosviewer.com/>>. Accessed: 2024-01-03. Citado na página 31.
- VOULODIMOS, A. et al. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, Wiley Online Library, v. 2018, n. 1, p. 7068349, 2018. Citado 6 vezes nas páginas 15, 17, 18, 19, 20 e 21.
- WANG, C.-W.; KHALIL, M.-A.; FIRDI, N. P. A survey on deep learning for precision oncology. *Diagnostics*, MDPI, v. 12, n. 6, p. 1489, 2022. Citado 2 vezes nas páginas 17 e 19.
- YANI, M.; IRAWAN S, S. M. B.; SETININGSIH ST, M. C. Application of transfer learning using convolutional neural network method for early detection of terry’s nail. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2019. v. 1201, n. 1, p. 012052. Citado 3 vezes nas páginas 21, 22 e 23.
- YASMIN, F. et al. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. *Reviews in cardiovascular medicine*, IMR Press, v. 22, n. 4, p. 1095–1113, 2021. Citado na página 3.
- ZANETTE, G. *Types of Learning in Machine Learning*. 2024. Accessed: May 17, 2024. Disponível em: <<https://medium.com/@gabezanette.dev/tipos-de-aprendizagem-em-machine-learning-45bc6d55a25a>>. Citado na página 13.