

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Redes Neurais Convolucionais no diagnóstico de Câncer Cervical

Lucas Soares Amaral

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Lucas Soares Amaral

Redes Neurais Convolucionais no diagnóstico de Câncer Cervical

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Profa. Dra. Agma Juci Machado Traina

Versão original

São Carlos
2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

A485r Amaral, Lucas Soares
Redes Neurais Convolucionais no diagnóstico de
Câncer Cervical / Lucas Soares Amaral; orientadora
Agma Juci Machado Traina. -- São Carlos, 2024.
62 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024.

1. Papanicolaou. 2. Câncer cervical. 3. Diagnóstico por computador. 4. Inteligência Artificial. 5. Redes neurais convolucionais. I. Traina, Agma Juci Machado, orient. II. Título.

Lucas Soares Amaral

Convolutional Neural Networks for Cervical Cancer diagnosis

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Agma Juci Machado Traina

Original version

**São Carlos
2024**

AGRADECIMENTOS

À minha esposa e melhor amiga, Emilyane, pela parceria, compreensão e dedicação à nossa vida e aos nossos objetivos.

Aos meus pais e irmã, por toda dedicação e carinho ao longo de toda minha jornada.

À Deus, por ter preparado todo o caminho até aqui.

Aos professores do curso de MBA em Inteligência Artificial e Big Data do ICMS/USP pelo conhecimento compartilhado.

À minha orientadora, Prof. Dra. Agma Juci Machado Traina, por sua disponibilidade, ensinamentos e compreensão ao longo do desenvolvimento deste trabalho.

RESUMO

Amaral, L. S. **Redes Neurais Convolucionais no diagnóstico de Câncer Cervical.** 2024. 62p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

O câncer de colo do útero, embora amplamente prevenível e tratável, permanece um significativo problema de saúde pública, sendo uma das principais causas de morte por câncer entre mulheres em todo o mundo. Os exames citopatológicos, como o Papanicolau, desempenham um papel essencial na prevenção e no rastreamento precoce da doença, permitindo diagnósticos precoces e aumentando as chances de cura. No entanto, a eficácia desse método depende de uma análise minuciosa e precisa, destacando a necessidade de aprimorar técnicas de detecção e diagnóstico. Com o objetivo de tornar o processo de diagnóstico via Papanicolau mais eficiente e superar as limitações dos métodos convencionais de análise de lâminas, este trabalho avaliou o desempenho das redes neurais convolucionais VGG16, ResNet50, EfficientNetB0, EfficientNetV2B0, InceptionV3 e MobileNetV3Small na classificação de células cervicais em imagens de 75x75 e 100x100 pixels, com base no Sistema Bethesda. A análise dos modelos treinados revelou que a EfficientNetV2B0 e a ResNet50 foram os que apresentaram melhor desempenho geral, com destaque para a EfficientNetV2B0, que obteve um F1-score ponderado de 0.8036 e áreas sob a curva consistentemente elevadas em todas as classes (variando entre 0.90 e 0.99). Os resultados indicam que as redes neurais convolucionais conseguem alcançar altos níveis de acurácia e desempenho, oferecendo um método robusto e confiável para o diagnóstico automatizado de exames citológicos.

Palavras-chave: Papanicolau; Câncer cervical; Diagnóstico por computador; Inteligência Artificial; Redes neurais; Redes neurais convolucionais.

ABSTRACT

Amaral, L. S. **Convolutional Neural Networks for Cervical Cancer diagnosis.** 2024. 62p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Cervical cancer, although largely preventable and treatable, remains a significant public health issue and one of the leading causes of cancer-related deaths among women worldwide. Cytopathological exams, such as the Pap Smear test, play a crucial role in preventing and screening for the disease, enabling early diagnosis and increasing the chances of cure. However, the effectiveness of this method relies on careful and precise analysis, highlighting the need to enhance detection and diagnostic techniques. Aiming to make the diagnostic process through the Pap Smear test more efficient and to overcome the limitations of conventional slide analysis methods, this study evaluated the performance of the convolutional neural networks VGG16, ResNet50, EfficientNetB0, EfficientNetV2B0, InceptionV3, and MobileNetV3Small in classifying cervical cells in Pap smear exams using image resolutions of 75x75 and 100x100 pixels, based on the Bethesda System. The analysis of the trained models revealed that EfficientNetV2B0 and ResNet50 were the top-performing models overall, with EfficientNetV2B0 standing out by achieving a weighted F1-score of 0.8036 and consistently high area under the curve across all classes (values between 0.90 and 0.99). The results indicate that convolutional neural networks can achieve high levels of accuracy and performance, offering a robust and reliable method for the automated diagnosis of cytological exams.

Keywords: Papanicolaou; Cervical cancer; Computer-assisted diagnosis; Artificial Intelligence; Neural networks; Convolutional neural networks.

LISTA DE FIGURAS

Figura 1 – Curva ROC e área sob a curva (AUC). Fonte: Zewde <i>et al.</i> (2024)	30
Figura 2 – Exemplificação da arquitetura de uma CNN. Fonte: Clement <i>et al.</i> (2023)	31
Figura 3 – Arquitetura EfficientNet. Fonte: Chen <i>et al.</i> (2024)	31
Figura 4 – Arquitetura EfficientNetV2. Fonte: Gardner <i>et al.</i> (2023)	32
Figura 5 – Arquitetura MobileNetV3. Fonte: Elaziz <i>et al.</i> (2021)	33
Figura 6 – Arquitetura InceptionV3. Fonte: Abdalla, Mohammed and Saeed (2023)	33
Figura 7 – Arquitetura ResNet. Fonte: Clement <i>et al.</i> (2023)	34
Figura 8 – Arquitetura ResNet. Fonte: Bakasa and Viriri (2023)	35
Figura 9 – Diagrama representativo do <i>pipeline</i> proposto compreendendo as etapas de 1) obtenção e pré-processamento das imagens; 2) aumento de dados; e 3) classificação.	39
Figura 10 – Exemplificação de células e sua respectiva classificação conforme o Sistema Bethesda.	40
Figura 11 – Distribuição inicial das imagens entre as seis classificações possíveis.	41
Figura 12 – Operações de transformação realizadas para aumento dos dados.	42
Figura 13 – Distribuição das imagens entre as seis classificações possíveis no <i>dataset</i> de treinamento após realização do aumento de dados.	42
Figura 14 – Ilustração da rede gerada, consistindo 1) Das camadas congeladas do modelo base, pré-treinado; 2) Uma camada <i>Flatten</i> , para transformar as saídas convolucionais em uma única dimensão; 3) Uma camada <i>Dense</i> de 256 neurônios com ativação <i>ReLU</i> , para adicionar poder de decisão não-linear; 4) <i>Dropout</i> de 50% para evitar <i>overfitting</i> ; e 5) Camada de <i>output Softmax</i> para classificação das seis categorias de células.	44
Figura 15 – Curva ROC para ResNet50 e EfficientNetB0V2 (75x75).	51
Figura 16 – Curva ROC para ResNet50 e EfficientNetB0V2 (100x100).	52
Figura 17 – Evolução da acurácia e perda ao longo das épocas para ResNet50 e EfficientNetB0V2 (100x100).	53
Figura 18 – Matriz de confusão para ResNet50 e EfficientNetB0V2 (100x100).	54

LISTA DE TABELAS

Tabela 1 – Matriz de confusão para classificação binária.	28
Tabela 2 – Métricas por classe e modelo (75x75).	48
Tabela 3 – Métricas por classe e modelo (100x100).	50

LISTA DE ABREVIATURAS E SIGLAS

CCU	Câncer do colo do útero ou câncer cervical
OMS	Organização Mundial da Saúde
HPV	Papilomavírus Humano
CNNs	Redes neurais convolucionais
TBS	Sistema Bethesda
NILM	Negativo para Neoplasia Intraepitelial ou Malignidade
ASC-US	Célula Escamosa Atípica de Significado Indeterminado
ASC-H	Célula Escamosa Atípica que não pode excluir lesão intraepitelial escamosa de alto grau
LSIL	Lesão Intraepitelial Escamosa de Baixo Grau
HSIL	Lesão Intraepitelial Escamosa de Alto Grau
SCC	Carcinoma de Células Escamosas
RNAs	Redes Neurais Artificiais
RNPs	Redes neurais profundas
ROC	<i>Receiver Operating Characteristics</i>
AUC	<i>Area Under The Curve</i>
VP	Verdadeiros positivos
VN	Verdadeiros negativos
FP	Falsos positivos
FN	Falsos negativos
GANs	Redes adversariais generativas

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização, Motivação e Lacunas	21
1.2	Hipóteses e Objetivos	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Câncer Cervical	23
2.2	Papilomavírus Humano	23
2.3	Papanicolau	23
2.4	Sistema Bethesda	24
2.4.1	Negativo para Neoplasia Intraepitelial ou Malignidade (NILM)	24
2.4.2	Célula Escamosa Atípica de Significado Indeterminado (ASC-US)	24
2.4.3	Célula Escamosa Atípica que não pode excluir lesão intraepitelial escamosa de alto grau (ASC-H)	24
2.4.4	Lesão Intraepitelial Escamosa de Baixo Grau (LSIL)	24
2.4.5	Lesão Intraepitelial Escamosa de Alto Grau (HSIL)	24
2.4.6	Carcinoma de Células Escamosas (SCC)	25
2.5	Iniciativa para a Eliminação do Câncer Cervical	25
2.6	Redes Neurais Artificiais	25
2.6.1	Aumento de dados	25
2.6.2	Transferência de aprendizado	27
2.6.3	Métricas	27
2.6.3.1	Matriz de Confusão	28
2.6.3.2	Acurácia	28
2.6.3.3	Precisão	28
2.6.3.4	<i>Recall</i>	29
2.6.3.5	F1-Score	29
2.6.3.6	Taxa de erro	29
2.6.3.7	Curva ROC e AUC	29
2.7	Redes Neurais Convolucionais	30
2.7.1	EfficientNet	31
2.7.2	EfficientNetV2	31
2.7.3	MobileNetV3	32
2.7.4	InceptionV3	33
2.7.5	ResNet	34
2.7.6	VGG16	34
2.8	Estado da arte	35

3	METODOLOGIA E DESENVOLVIMENTO	39
3.1	Proposta	39
3.1.1	Conjunto de dados	39
3.2	Pré-processamento	40
3.3	Classificação	43
3.3.1	Modelos de classificação	43
4	ANÁLISE DOS RESULTADOS	47
5	CONCLUSÕES	55
	Referências	57

1 INTRODUÇÃO

1.1 Contextualização, Motivação e Lacunas

De acordo com dados da Agência Internacional de Pesquisa sobre o Câncer, o câncer do colo do útero ou câncer cervical (CCU) é o quarto tipo de câncer mais incidente em mulheres, atingindo 604.127 mulheres em todo o mundo no ano de 2020 (FERLAY *et al.*, 2020). Apesar de apresentar grandes chances de prevenção e cura, quando diagnosticado precocemente através de técnicas de rastreamento, o CCU foi identificado como a causa do óbito de 341.831 mulheres no ano de 2020, representando a maioria das mortes por câncer em 42 países, principalmente no continente Africano (FERLAY *et al.*, 2020).

Em 2020, a Organização Mundial de Saúde (OMS) lançou a "Iniciativa para a Eliminação do Câncer Cervical" para acelerar a redução e eliminação dos casos de CCU até 2030. A iniciativa se baseia na prevenção, rastreamento e tratamento do CCU e objetiva atingir 90% de cobertura na vacinação de garotas até 15 anos de idade contra o Papilomavírus Humano (HPV), 70% de cobertura no rastreamento e 90% de cobertura no tratamento de lesões pré-cancerosas (WHO, 2020). Dada a relação entre infecções por HPV e o desenvolvimento de CCU, a vacinação e a aplicação de métodos de rastreamento são importantes ferramentas para a prevenção e intervenção precoce ao desenvolvimento de CCU (BUSKWOFIE; DAVID-WEST; CLARE, 2020).

O exame Papanicolau é um dos métodos mais conhecidos e utilizados para a detecção de alterações nas células do útero, inclusive CCU. Por ser um método relativamente barato e eficaz, ele ainda é largamente utilizado, principalmente em países subdesenvolvidos, apesar do surgimento de novos métodos e tecnologias de diagnóstico (SACHAN *et al.*, 2018). Além disso, a implementação de programas eficientes para a utilização do Papanicolau se mostrou eficaz em países de alta renda, possibilitando uma redução de cerca de 60% na incidência e mortalidade por CCU nos últimos 40 anos (FERLAY *et al.*, 2018), o que reforça a importância dessas políticas, principalmente em países de baixa renda, onde a implementação de técnicas mais caras substitutas ou conjuntas ao Papanicolau é dificultada (MEGGIOLARO *et al.*, 2016).

O relativo baixo custo do Papanicolau é compensado pelo esforço necessário para a realização do diagnóstico, cuja interpretação pode ser demorada e subjetiva (REZENDE *et al.*, 2021b). Sendo assim, a criação e a aplicação de métodos computacionais para automatização de processos na área médica, bem como da utilização de modelos de aprendizado de máquina para a interpretação de imagens tem sido amplamente discutidas e impulsionadas pela evolução tecnológica e a crescente geração de dados relacionados ao tratamento de câncer (JHA *et al.*, 2023).

Apesar do crescente volume de dados, a criação de modelos de aprendizado de máquina capazes de identificar a existência de CCU requer a disponibilização de imagens de alta qualidade representativas de todas as classes de alterações pré-neoplásicas (JANTZEN *et al.*, 2005). Além disso, é necessária a realização de processos de validação do modelo com grandes volumes de dados, realização de ensaios clínicos e uma avaliação cuidadosa dos resultados obtidos (JHA *et al.*, 2023).

1.2 Hipóteses e Objetivos

Visando tornar o processo de diagnóstico através do Papanicolau mais eficiente e superar as limitações dos métodos convencionais de análise de lâminas, a proposição de diferentes modelos de aprendizado de máquina se faz importante para a geração de resultados de alta qualidade e para que a tecnologia desenvolvida seja, de fato, integrada aos procedimentos de laboratórios de patologia.

Este trabalho objetiva avaliar e comparar a performance de redes neurais convolucionais (CNNs) na identificação de células cancerígenas em lâminas provenientes do Papanicolau, através da classificação das células utilizando a nomenclatura sugerida pelo Sistema Bethesda (TBS).

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Câncer Cervical

O CCU é uma neoplasia maligna caracterizada pela proliferação anormal de células no colo do útero ocasionada principalmente pela infecção persistente pelos tipos de alto risco do HPV (JOHNSON *et al.*, 2019).

O CCU é a quarta causa mais comum de câncer (RAJARAM; GUPTA, 2021) e uma das principais causas de morte por câncer em mulheres em todo o mundo (OLUSOLA *et al.*, 2019), apesar de ser prevenível e tratável quando detectado precocemente (PIMPLE; MISHRA, 2019). A incidência e a mortalidade do CCU variam significativamente entre diferentes regiões geográficas, sendo mais prevalente em países de baixa e média renda devido à menor cobertura de programas de rastreamento e vacinação contra o HPV. (ARBYN *et al.*, 2020; VU *et al.*, 2018).

As estratégias de prevenção e tratamento incluem vacinação contra o HPV, rastreamento por meio de exames de Papanicolau e teste de DNA para HPV, bem como tratamentos como cirurgia, radioterapia e quimioterapia (JALLAH; ANJANKAR; NAN-KONG, 2023).

2.2 Papilomavírus Humano

O HPV é um grupo de vírus pertencentes à família *Papillomaviridae* que inclui mais de 200 variações genéticas distintas (de Sanjose *et al.*, 2019; de Villiers, 2013). Cerca de 40 destas variações podem infectar o trato genital, sendo o HPV dos tipos 16 e 18 responsáveis por aproximadamente 70% dos casos de CCU (STEBEN; DUARTE-FRANCO, 2007). A infecção por HPV é geralmente assintomática e transitória, mas em alguns casos, pode persistir e levar ao desenvolvimento de lesões pré-cancerosas ou cancerosas (BOSCH; SANJOSÉ, 2020). A transmissão do HPV ocorre principalmente por via sexual, e a infecção persistente por tipos de alto risco é um fator agravante significativo para o desenvolvimento do CCU (BURD, 2003).

2.3 Papanicolau

O teste de Papanicolau, ou exame citopatológico, é um método de rastreamento para o CCU que envolve a coleta de células do colo do útero para identificar possíveis alterações pré-cancerosas. Este teste reduziu significativamente a mortalidade por CCU desde a sua introdução na década de 1940 (STEWART; WILD, 2014), ao passo que auxilia no rastreamento e detecção precoce de lesões cervicais precursoras e CCU.

2.4 Sistema Bethesda

O TBS é uma classificação utilizada para relatar os resultados dos exames de Papanicolau (PANGARKAR, 2022). Ele fornece terminologia padronizada para descrever as anormalidades celulares encontradas, facilitando a comunicação entre profissionais de saúde e garantindo uma abordagem consistente no tratamento e acompanhamento de pacientes (NAYAR; WILBUR, 2015).

A característica mais importante do TBS é a sua relevância clínica proporcionada por uma categorização com implicações clínicas claras, baseadas em evidências sólidas e consenso mundial (PANGARKAR, 2022).

A partir da utilização do TBS, algumas categorias podem ser identificadas (PANGARKAR, 2022; REZENDE *et al.*, 2021a; WANG *et al.*, 2023):

2.4.1 Negativo para Neoplasia Intraepitelial ou Malignidade (NILM)

Este resultado indica que não foram encontradas alterações neoplásicas no exame Papanicolau. É considerado um resultado normal e sugere baixo risco de CCU (WANG *et al.*, 2023; PANGARKAR, 2022).

2.4.2 Célula Escamosa Atípica de Significado Indeterminado (ASC-US)

Este resultado indica presença de células escamosas atípicas de significado indeterminado (REZENDE *et al.*, 2021a). Ou seja, indica a presença de células escamosas com alterações que são levemente anormais, mas não conclusivas para uma lesão pré-cancerosa (WANG *et al.*, 2023).

2.4.3 Célula Escamosa Atípica que não pode excluir lesão intraepitelial escamosa de alto grau (ASC-H)

Indica a presença de células escamosas com alterações mais pronunciadas (PANGARKAR, 2022), sugerindo a possibilidade de alterações citológicas sugestivas de neoplasia intraepitelial (WANG *et al.*, 2023; PANGARKAR, 2022).

2.4.4 Lesão Intraepitelial Escamosa de Baixo Grau (LSIL)

Indica a presença de células escamosas com alterações sugestivas de lesão pré-cancerosa de baixo grau (PANGARKAR, 2022; WANG *et al.*, 2023).

2.4.5 Lesão Intraepitelial Escamosa de Alto Grau (HSIL)

Indica a presença de células escamosas com alterações significativas, sugestivas de lesão pré-cancerosa de alto grau. Essas alterações têm maior probabilidade de progredir para câncer invasivo se não forem tratadas (WANG *et al.*, 2023; REZENDE *et al.*, 2021a).

2.4.6 Carcinoma de Células Escamosas (SCC)

Esta categoria é reservada para casos em que as células observadas no exame são consistentes com SCC, indicando a presença de câncer invasivo (REZENDE *et al.*, 2021a).

2.5 Iniciativa para a Eliminação do Câncer Cervical

A "Iniciativa para a Eliminação do Câncer Cervical", liderada pela OMS, tem como objetivo reduzir significativamente a incidência e mortalidade por CCU por meio de estratégias abrangentes de prevenção, rastreamento e tratamento em todo o mundo (WHO, 2020).

Anunciada pelo Diretor Geral da OMS, a "Iniciativa para a Eliminação do Câncer Cervical" visa traçar uma estratégia até 2030 para a eliminação do CCU até o final do século (ORGANIZATION, 2020).

Para ser bem sucedida, é previsto que todos os países mantenham uma incidência abaixo de 4 a cada 100.000 mulheres através da implantação de políticas que garantam o atingimento da meta 90-70-90 (90% das adolescentes até 15 anos totalmente vacinadas contra o HPV; 70% das mulheres rastreadas por métodos de alto desempenho aos 35 e 45 anos; 90% das mulheres com câncer invasivo tratadas) (ORGANIZATION, 2020).

2.6 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso humano, capazes de realizar tarefas complexas, como classificação e previsão, através de grandes volumes de dados complexos, consistindo em camadas de neurônios conectados que processam entradas, realizam cálculos e produzem saídas. As RNAs possuem ampla aplicabilidade em tarefas que demandam reconhecimento de padrões em imagens e sons, além de previsões de séries temporais e diagnósticos automatizados em exames médicos. No contexto da saúde, são particularmente valiosas para análise de dados médicos, onde podem identificar padrões que não são facilmente visíveis para humanos (LECUN; BENGIO; HINTON, 2015).

Apesar de serem amplamente aplicadas em diferentes áreas e contextos, as RNAs demandam a disponibilidade de grandes volumes de dados para alcançarem a precisão necessária, podendo ser computacionalmente custosas. Além disso, sua estrutura complexa faz com que seja difícil interpretar como as decisões são tomadas internamente (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.6.1 Aumento de dados

O aumento de dados é uma técnica amplamente utilizada em problemas de aprendizado de máquina, particularmente em visão computacional, para expandir o conjunto

de dados disponível, aumentar a variabilidade intra-classe e melhorar a capacidade de generalização dos modelos. Em situações onde há escassez de dados rotulados, o aumento de dados se torna uma estratégia fundamental, permitindo que novas amostras sejam geradas a partir de dados existentes por meio de transformações que simulam variações naturais presentes no conjunto de dados (MUMUNI; MUMUNI, 2022).

Ao ser exposto a um conjunto de dados de tamanho adequado, o modelo é menos propenso a memorizar padrões específicos dos dados de treinamento e, em vez disso, aprende características mais robustas e generalizáveis (MUMUNI; MUMUNI, 2022). Isso é particularmente importante em cenários médicos, onde a coleta de novos dados pode ser limitada ou custosa (SHORTEN; KHOSHGOFTAAR, 2019), a variabilidade dos dados pode ser alta e é crucial que o modelo tenha a capacidade de lidar com diferentes tipos de variações nas imagens (PEREZ; WANG, 2017a).

Entre os principais métodos de aumento de dados aplicados à visão computacional estão as transformações geométricas, que incluem rotações, espelhamentos e cortes, e as transformações fotométricas, que ajustam as propriedades de cor, brilho e contraste das imagens (PEREZ; WANG, 2017b). Essas transformações permitem que o modelo se torne mais robusto a variações de posicionamento e iluminação que podem ocorrer durante a captura das imagens, garantindo que ele seja capaz de identificar as características relevantes, independentemente das condições externas e, consequentemente, expondo o modelo a uma maior diversidade de dados durante o treinamento (SHORTEN; KHOSHGOFTAAR, 2019).

O aumento de dados também desempenha um papel importante no balanceamento de classes (ISLAM *et al.*, 2024), um desafio comum em aplicações médicas, ajudando a equilibrar o número de amostras entre as classes e minimizando o risco de enviesamento do modelo para a classe majoritária (BUDA; MAKI; MAZUROWSKI, 2018).

O aumento de dados também ajuda a reduzir as chances de ocorrência de *overfitting* (PEREZ; WANG, 2017c; SHORTEN; KHOSHGOFTAAR, 2019), um problema comum em modelos de RNNs, especialmente quando o conjunto de dados é limitado. Ao ampliar o número de amostras de treinamento, os dados aumentados passam a abranger uma variedade maior de possíveis cenários, reduzindo a diferença entre os conjuntos de treinamento, teste e futuros conjuntos de validação. Isso ajuda o modelo a generalizar melhor, resultando em um desempenho aprimorado em novos dados (SHORTEN; KHOSHGOFTAAR, 2019).

Apesar de suas vantagens, o aumento de dados apresenta algumas limitações como necessidade de memória adicional, aumento no tempo de treinamento do modelo e cuidado adicional para evitar alterações no rótulo da imagem (SHORTEN; KHOSHGOFTAAR, 2019). Além disso, transformações excessivas ou irreais podem introduzir ruído nos dados, o que pode prejudicar o desempenho do modelo, em vez de aprimorá-lo. Por isso, é crucial que as transformações aplicadas sejam representativas das variações reais dos dados e não

alterem de forma drástica a semântica das imagens originais.

2.6.2 Transferência de aprendizado

A transferência de aprendizado é uma técnica que tem ganhado grande relevância no campo de aprendizado de máquina, especialmente em problemas envolvendo RNP (GU; LEE, 2024). A transferência de aprendizado consiste na reutilização de um modelo previamente treinado em uma tarefa, para aplicá-lo a uma nova tarefa que compartilhe algumas características semelhantes (PAN; YANG, 2010). Essa abordagem tem mostrado resultados notáveis em diversas áreas, particularmente em visão computacional e processamento de imagens médicas, onde os conjuntos de dados são muitas vezes limitados (GU; LEE, 2024).

Para classificação de imagens, o treinamento de RNP pode ser desafiador devido à escassez de dados rotulados e à necessidade de recursos computacionais consideráveis. A transferência de aprendizado permite contornar essas limitações, aproveitando o conhecimento adquirido por modelos pré-treinados em grandes conjuntos de dados, como o *ImageNet*, para acelerar o treinamento em tarefas específicas. Ao utilizar uma rede previamente treinada em um conjunto massivo de imagens, as camadas iniciais da rede já estão capacitadas para detectar padrões genéricos, como bordas, texturas e formas simples, que podem ser reutilizados em novas tarefas, economizando tempo e melhorando a performance (ZOPH *et al.*, 2017). Com isso, apenas as camadas finais precisam ser ajustadas à tarefa específica, reduzindo significativamente a complexidade inerente ao processo de treinamento e a necessidade de grandes volumes de dados (SHIN *et al.*, 2016a).

Embora a transferência de aprendizado ofereça benefícios significativos, há algumas limitações. A principal desvantagem está na dependência da similaridade entre o conjunto de dados original e a nova tarefa. Se as características das imagens da nova tarefa forem muito diferentes das que o modelo foi originalmente treinado, a eficácia da transferência de aprendizado pode ser limitada (HUH; AGRAWAL; EFROS, 2016). Além disso, o ajuste fino (*fine-tuning*) de camadas pode exigir um cuidado especial para evitar *overfitting*, especialmente quando o novo conjunto de dados é muito pequeno.

2.6.3 Métricas

Em problemas de aprendizado supervisionado, o uso de métricas adequadas permite uma análise detalhada sobre como o modelo está se comportando em relação às diversas classes de saída (CASTRO; BRAGA, 2011). Algumas das métricas comumente utilizadas para avaliações de RNP incluem a acurácia, precisão, *recall* (sensibilidade), F1-score, F1-score ponderado, taxa de erro e o gráfico da curva *Receiver Operating Characteristic* (ROC) e a área sob a curva (*Area Under The Curve/AUC*) (NAIDU; ZUVA; SIBANDA, 2023).

2.6.3.1 Matriz de Confusão

As previsões de um modelo treinado resultam em quatro possíveis desfechos representados por uma matriz de confusão: previsões corretas – verdadeiros positivos (VP) e verdadeiros negativos (VN) – e previsões incorretas – falsos positivos (FP) e falsos negativos (FN). Quando uma amostra positiva é corretamente classificada como positiva, ela é chamada de VP. Se for incorretamente classificada como negativa, é um FN. Da mesma forma, uma amostra negativa corretamente classificada como negativa é um VN, enquanto, se for incorretamente classificada como positiva, é considerada um FP (THARWAT, 2018).

O modelo de uma matriz de confusão pode ser observada na Tabela 1.

Tabela 1 – Matriz de confusão para classificação binária.

	Previsão Positiva	Previsão Negativa
Classe Positiva	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Classe Negativa	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Com base nos dados que podem ser extraídos desta matriz de confusão, muitas métricas de classificação podem ser calculadas (THARWAT, 2018; NAIDU; ZUVA; SIBANDA, 2023).

2.6.3.2 Acurácia

A acurácia é uma das métricas mais comuns e simples em problemas de classificação, e visa calcular a proporção de previsões corretas em relação ao total de previsões realizadas, sendo útil para fornecer uma visão geral do desempenho do modelo em todo o conjunto de dados (NAIDU; ZUVA; SIBANDA, 2023).

O cálculo da acurácia pode ser representado pela fórmula abaixo:

$$\text{Acurácia} = \frac{\text{Número de previsões corretas}}{\text{Número total de amostras}}$$

2.6.3.3 Precisão

A precisão é uma métrica que reflete a proporção de previsões corretas para uma determinada classe em relação a todas as previsões feitas para essa mesma classe (NAIDU; ZUVA; SIBANDA, 2023).

A precisão é calculada da seguinte forma:

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

2.6.3.4 *Recall*

O *recall*, também conhecido como sensibilidade, mede a capacidade do modelo de identificar corretamente todas os resultados VP de cada categoria disponível e quão eficaz o modelo foi capaz de prevê-los (NAIDU; ZUVA; SIBANDA, 2023).

O *recall* é calculado da seguinte maneira:

$$\text{Recall} = \frac{VP}{(VP + FN)}$$

2.6.3.5 F1-Score

O F1-score é a métrica que harmoniza a precisão e o *recall* e é útil ao medir a eficácia de conjuntos de dados desbalanceados (NAIDU; ZUVA; SIBANDA, 2023).

O F1-score é calculado pela fórmula:

$$F1\text{-score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

Um alto F1-score indica que o modelo consegue tanto identificar a maioria dos VP (alto *recall*) quanto evitar muitos FP (alta precisão), sendo uma métrica central para avaliar o desempenho em classes minoritárias.

2.6.3.6 Taxa de erro

A taxa de erro representa a porcentagem de previsões incorretas em relação ao número total de instâncias analisadas (NAIDU; ZUVA; SIBANDA, 2023).

A fórmula para a taxa de erro é:

$$\text{Taxa de erro} = \frac{FP + FN}{(VP + FP + VN + FN)}$$

2.6.3.7 Curva ROC e AUC

A curva ROC é um método de avaliação muito popular (SAITO; REHMSMEIER, 2015; FAWCETT, 2006a; THARWAT, 2018) e é visualizado como um gráfico que representa o desempenho de um modelo de classificação em diferentes limiares de decisão (NAIDU; ZUVA; SIBANDA, 2023). A curva ROC mostra a relação entre a taxa de VP e a taxa de FP - duas variáveis plotadas ao longo da curva ROC (NAIDU; ZUVA; SIBANDA, 2023; THARWAT, 2018):

$$\text{Taxa de verdadeiros positivos} = \frac{TP}{(TP + FN)}$$

$$\text{Taxa de falsos positivos} = \frac{FP}{(FP + VN)}$$

Sendo assim, o gráfico ROC mostra o *tradeoff* entre especificidade e sensibilidade (SAITO; REHMSMEIER, 2015; FAWCETT, 2006a)

A AUC, por sua vez, condensa essa informação em um único valor (entre 0 e 1), que reflete a capacidade do classificador em separar corretamente as diferentes classes. A AUC é 0,5 para classificadores aleatórios e 1 para classificadores perfeitos (SAITO; REHMSMEIER, 2015; FAWCETT, 2006b; NAIDU; ZUVA; SIBANDA, 2023; ZEWDE *et al.*, 2024) (Figura 1).

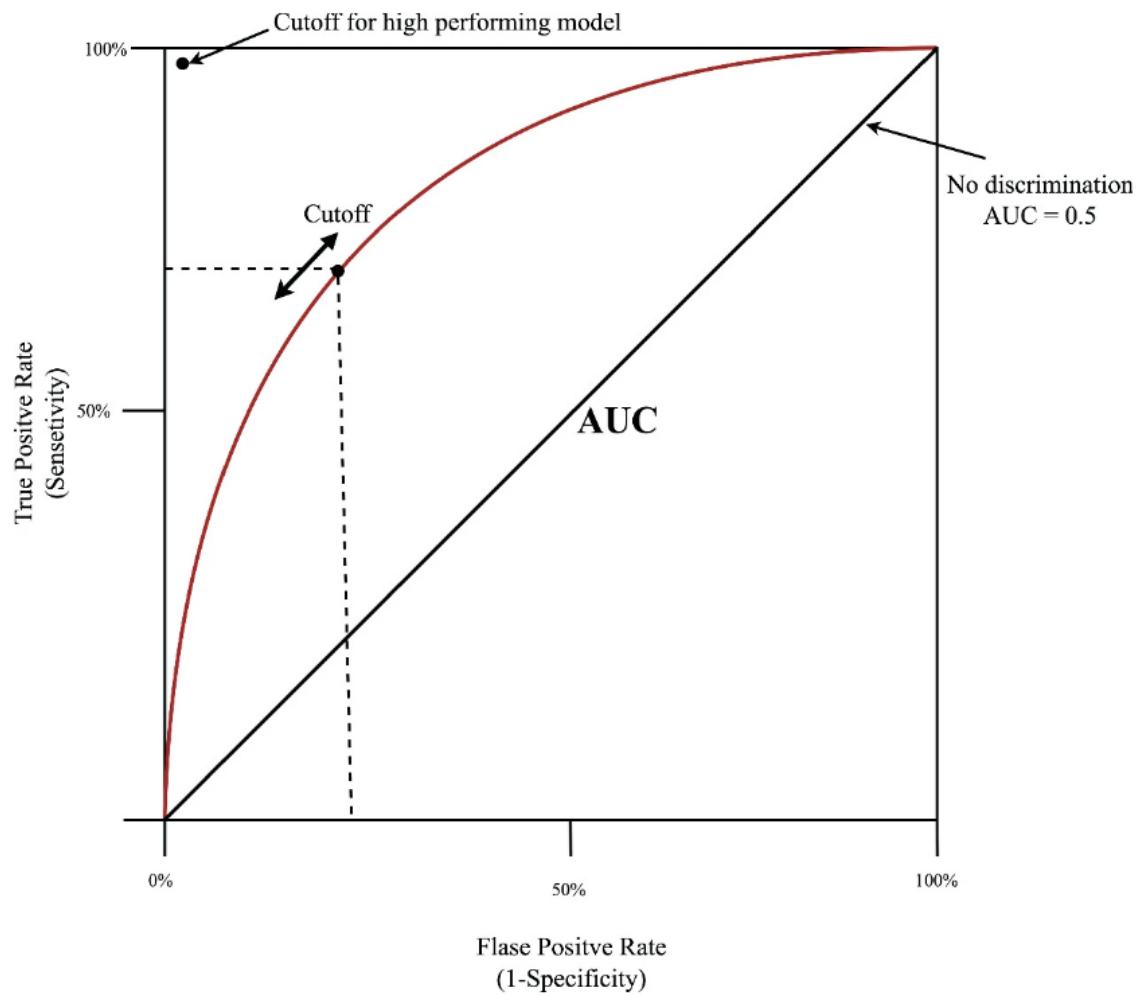


Figura 1 – Curva ROC e área sob a curva (AUC). Fonte: Zewde *et al.* (2024)

2.7 Redes Neurais Convolucionais

As redes neurais convolucionais (CNNs) são um tipo específico de rede neural projetadas para processar dados com estrutura espacial, como imagens. Elas são compostas por camadas convolucionais que extraem características das imagens de forma hierárquica, sendo extremamente eficazes na análise visual (YAMASHITA *et al.*, 2018), como mostrado na Figura 2.

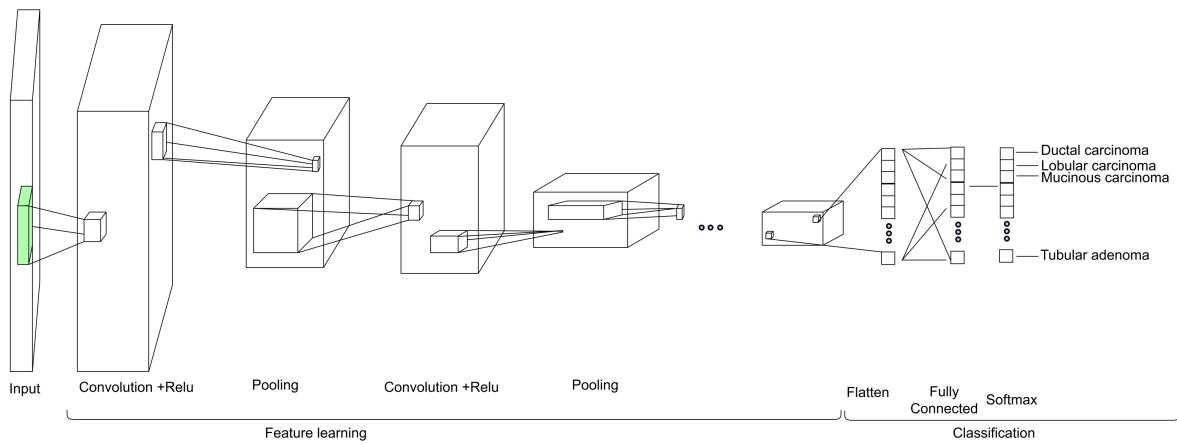


Figura 2 – Exemplificação da arquitetura de uma CNN. Fonte: Clement *et al.* (2023)

2.7.1 EfficientNet

Introduzida por Tan and Le (2019), a EfficientNet é uma CNN proposta como uma arquitetura otimizada que introduz o conceito de escalabilidade uniforme de todas as dimensões da rede com um conjunto fixo de fatores de escala. A proposta central da EfficientNet é escalar a profundidade, a largura e a resolução da rede de maneira coordenada, através de uma técnica chamada *Compound Scaling*, o que resulta em um modelo mais eficiente em termos computacionais, sem sacrificar a precisão (MARQUES; AGARWAL; DÍEZ, 2020; TAN; LE, 2019). Isto permite que ela alcance uma eficiência e eficácia melhoradas em tarefas computacionais intensivas, como processamento de imagens médicas (TAN; LE, 2019). A Figura 3 possui um diagrama representativo da arquitetura da rede EfficientNet.

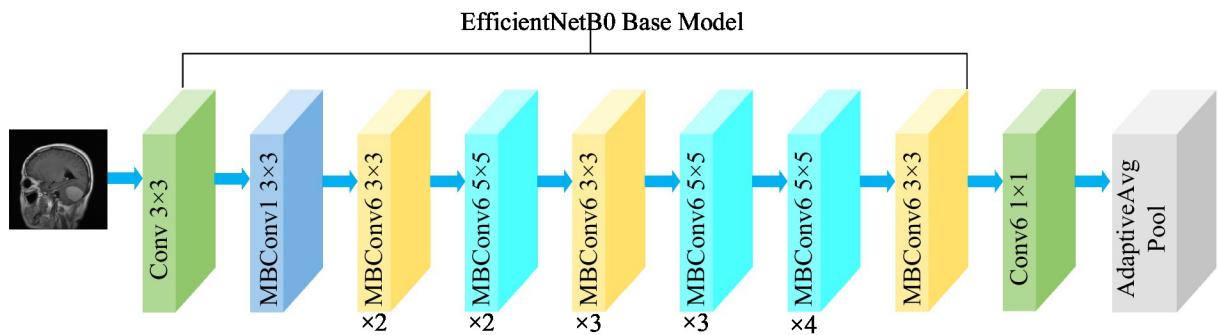


Figura 3 – Arquitetura EfficientNet. Fonte: Chen *et al.* (2024)

2.7.2 EfficientNetV2

O EfficientNetV2 é uma versão aprimorada do modelo EfficientNet, que se baseia nos mesmos princípios de escalabilidade, mas com melhorias no treinamento e eficiência.

Ele introduz um novo bloco de construção mais eficiente, além de técnicas que aceleram o treinamento da rede, como Fused-MBConv (Figura 4) e método aprimorado de *progressive learning* (TAN; LE, 2021).

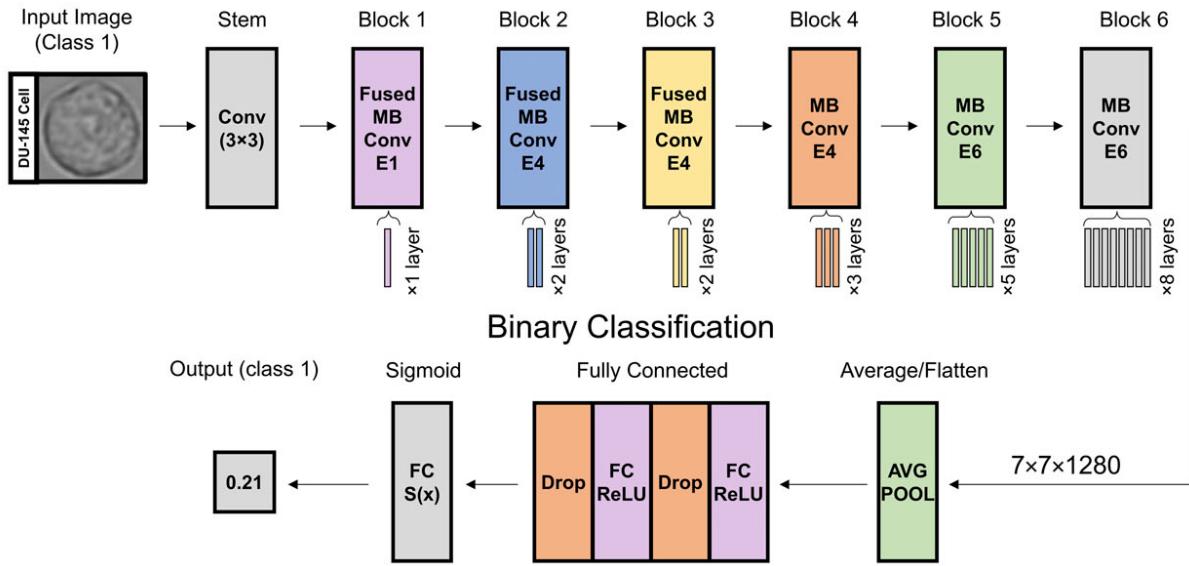


Figura 4 – Arquitetura EfficientNetV2. Fonte: Gardner *et al.* (2023)

Essa versão do modelo permite tanto maior precisão quanto menor tempo de treinamento em comparação ao EfficientNet original, sem comprometer a eficiência no uso de recursos computacionais (TAN; LE, 2021; TUMMALA *et al.*, 2023).

2.7.3 MobileNetV3

O MobileNetV3 é parte da família MobileNet, desenvolvida para ser eficiente em termos de recursos computacionais, com foco na execução de redes neurais em dispositivos móveis e ambientes com pouca capacidade de processamento. O MobileNetV3 combina camadas convolucionais com operações otimizadas (Figura 5), como *inverted residuals* e *linear bottlenecks*, o que permite alta eficiência em termos de uso de memória e processamento (HOWARD *et al.*, 2019; YULDASHEV *et al.*, 2023).

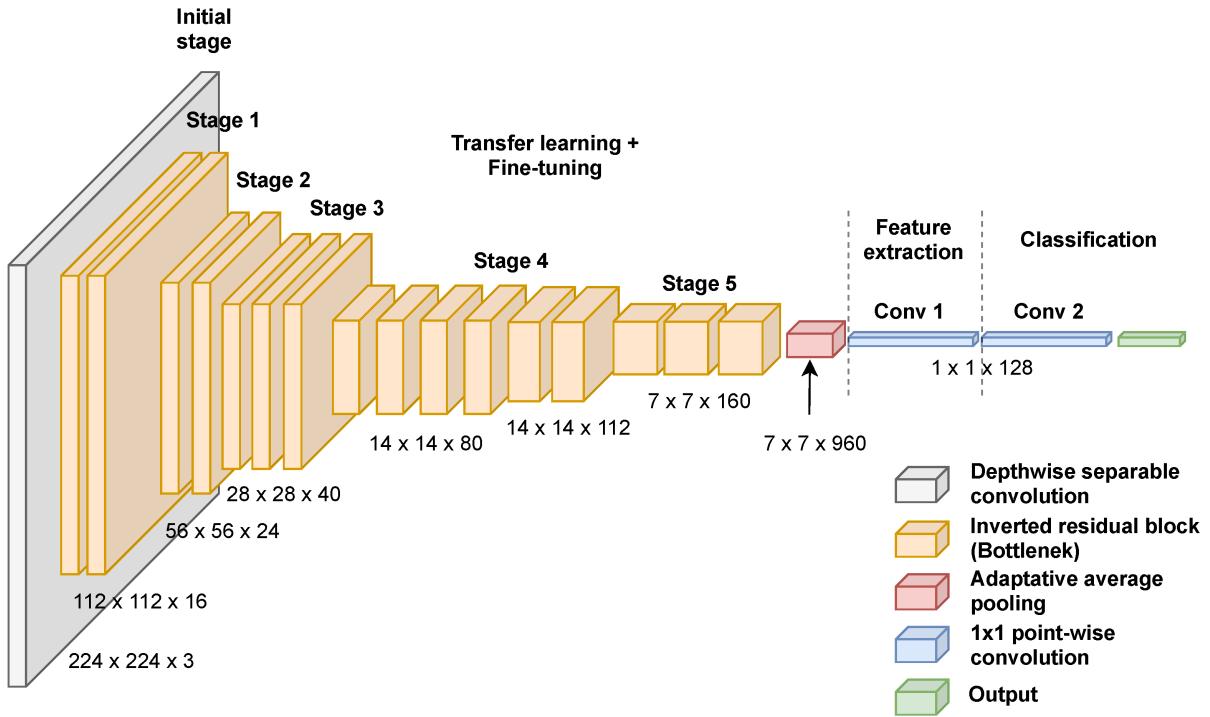


Figura 5 – Arquitetura MobileNetV3. Fonte: Elaziz *et al.* (2021)

2.7.4 InceptionV3

Desenvolvida por Szegedy *et al.* (2016), a arquitetura InceptionV3 faz parte da família de modelos Inception (GoogleNet) e é conhecida por sua eficiência em termos de computação e alta precisão em tarefas de classificação de imagens. A InceptionV3 introduz várias inovações, incluindo fatores de normalização em lote e módulos de fatorização, que reduzem o custo computacional e melhoram o desempenho em tarefas de visão computacional (SZEGEDY *et al.*, 2016; AGARWAL *et al.*, J Med Imaging (Bellingham); GUAN *et al.*, 2019). A Figura 6 possui uma representação da arquitetura InceptionV3.

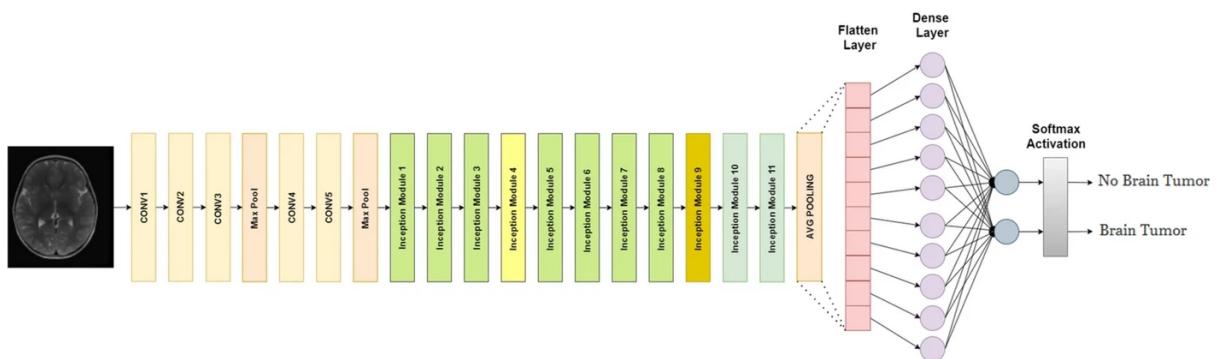


Figura 6 – Arquitetura InceptionV3. Fonte: Abdalla, Mohammed and Saeed (2023)

2.7.5 ResNet

A arquitetura ResNet (Residual Networks), introduzida por He *et al.* (2016), introduz o conceito de conexões residuais para permitir o treinamento de redes muito profundas sem o problema do desaparecimento do gradiente (Figura 7). Esse avanço permitiu que redes com centenas de camadas fossem treinadas eficientemente, o que resultou em melhorias significativas nas tarefas de classificação de imagens (AGARWAL *et al.*, J Med Imaging (Bellingham); HE *et al.*, 2016). É amplamente usada na classificação de imagens médicas devido à sua capacidade de aprender recursos hierárquicos (AGARWAL *et al.*, J Med Imaging (Bellingham)).

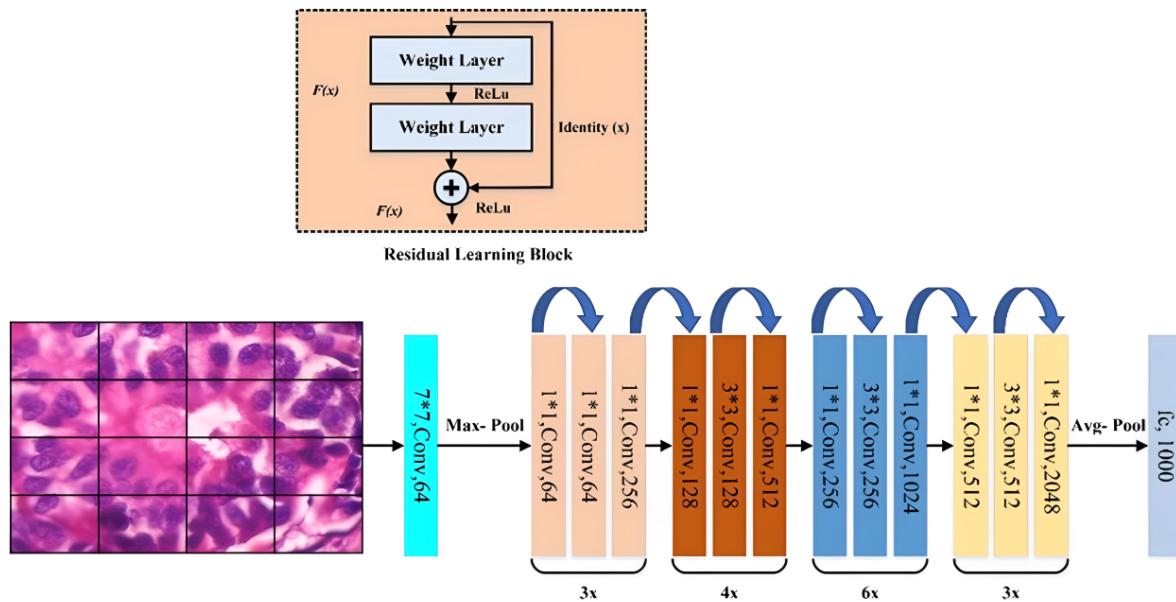


Figura 7 – Arquitetura ResNet. Fonte: Clement *et al.* (2023)

2.7.6 VGG16

Desenvolvida por Simonyan and Zisserman (2014), a VGG16 é uma rede neural profunda que usa uma abordagem simples e uniforme de aplicar convoluções 3x3 e camadas de *pooling* ao longo da rede, sendo composta por 16 camadas convolucionais (Figura 8). Devido à sua capacidade de extrair características discriminativas, ela tem sido aplicada em várias tarefas de visão computacional, incluindo diagnóstico médico, onde demonstrou alta precisão na detecção de características patológicas.

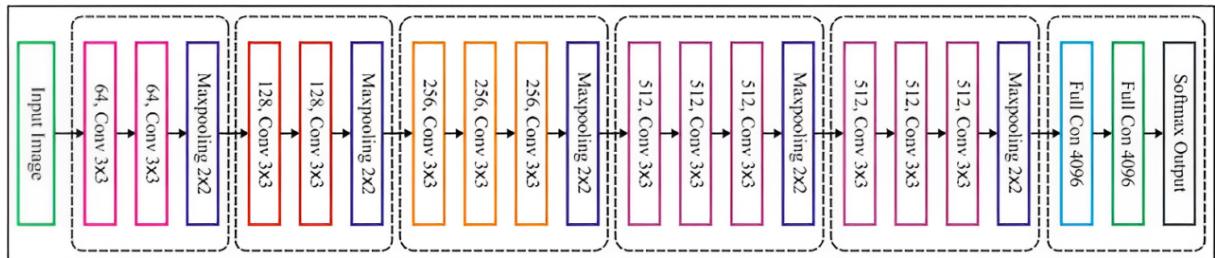


Figura 8 – Arquitetura ResNet. Fonte: Bakasa and Viriri (2023)

2.8 Estado da arte

Nos últimos anos, o uso de CNNs para análise de imagens médicas, particularmente no diagnóstico de câncer, tem sido amplamente explorado e apresenta resultados promissores (LI *et al.*, 2023). Em especial, a aplicação de CNNs para a classificação automática de células cervicais em exames de Papanicolau tem demonstrado grande potencial para aumentar a precisão, velocidade e eficiência desse processo, muitas vezes alcançando desempenho comparável ao de citopatologistas experientes (ZHANG *et al.*, 2017).

Em uma revisão sistemática e meta-análise realizada por Jha *et al.* (2023), que analisou modelos de predição utilizados no diagnóstico de CCU, as CNNs emergiram como uma das ferramentas mais promissoras, demonstrando alta sensibilidade e especificidade na detecção de lesões cervicais, destacando a eficácia dos modelos de predição em vários desfechos clínicos. A revisão ressalta que, embora o uso de CNNs seja uma abordagem robusta, existem desafios a serem superados, como a necessidade de maior interpretabilidade dos modelos para assegurar a confiabilidade clínica.

O diagnóstico tradicional de lesões cervicais e câncer é feito por meio de análises visuais manuais, que dependem da habilidade e experiência do citopatologista. Esse método, apesar de eficaz, pode ser demorado, suscetível a erros humanos e limitado pela variabilidade interobservador (HOU *et al.*, 2022). A introdução de CNNs nesse processo busca automatizar a classificação de células, ajudando a reduzir a subjetividade e fornecendo uma ferramenta confiável de apoio ao diagnóstico. Nesse contexto, CNNs são usadas para identificar características morfológicas sutis nas células, como variações na relação núcleo-citoplasma, hiperchromasia e alterações no formato nuclear, que são indicativos de lesões pré-cancerígenas ou cancerígenas. Essas redes profundas são capazes de aprender automaticamente essas características visuais a partir de grandes quantidades de imagens, sem necessidade de intervenção manual para a extração de atributos, o que aumenta significativamente a eficiência do processo de análise (SHIN *et al.*, 2016b).

Além da citologia cervical, as CNNs têm sido aplicadas com sucesso em várias áreas da oncologia, como na análise de mamografias, tomografias computadorizadas e ressonâncias magnéticas, para a detecção de diferentes tipos de câncer, incluindo o de

pulmão, mama e próstata (JIANG *et al.*, 2023). De acordo com Jha *et al.* (2023), o uso de CNNs em imagens oncológicas tem mostrado precisão e sensibilidade comparáveis às de especialistas humanos, especialmente na detecção de nódulos malignos e lesões suspeitas. Isso sugere que essas redes podem ser ferramentas valiosas na detecção precoce de câncer e no suporte à tomada de decisões clínicas.

Duas abordagens essenciais no uso de CNNs em diagnósticos médicos são a transferência de aprendizado e o aumento de dados (ZHAO *et al.*, 2024). A transferência de aprendizado permite que redes pré-treinadas em grandes bases de dados, como o ImageNet, sejam adaptadas para tarefas específicas, como a classificação de células cervicais, utilizando um número limitado de imagens rotuladas. Essa técnica tem se mostrado crucial em áreas médicas, onde a obtenção de grandes conjuntos de dados rotulados é desafiadora (KIM *et al.*, 2022). Shin *et al.* (2016b) demonstraram que CNNs treinadas com transferência de aprendizado podem atingir níveis de precisão semelhantes aos de especialistas humanos em tarefas de classificação de células cervicais, mesmo com conjuntos de dados relativamente pequenos.

O aumento de dados é outra técnica amplamente utilizada para melhorar o desempenho de CNNs, especialmente em cenários com desequilíbrio de classes, como em exames de Papanicolau, onde a maioria das amostras são de células normais e apenas uma fração das amostras são anormais. Ao gerar variações artificiais das imagens de treinamento, como rotações, ajustes de contraste e translações, é possível aumentar a diversidade do conjunto de dados, ajudando o modelo a generalizar melhor e a evitar *overfitting*. De acordo com Shorten and Khoshgoftaar (2019), o aumento de dados é essencial para melhorar o desempenho das CNNs, especialmente em classes minoritárias, garantindo que o modelo seja capaz de detectar células anormais com maior precisão.

Apesar dos avanços, ainda existem desafios a serem enfrentados para a plena adoção das CNNs na prática clínica. Um dos principais obstáculos é a interpretação das decisões tomadas pelas redes, uma vez que essas redes são frequentemente descritas como "caixas-pretas". Tornar essas decisões mais transparentes é crucial para aumentar a confiança de médicos e reguladores em sistemas automatizados de diagnóstico. Técnicas como Grad-CAM têm sido exploradas para tornar os resultados das CNNs mais compreensíveis, visualizando quais áreas da imagem influenciaram as decisões do modelo (SELVARAJU *et al.*, 2016).

Outra área em crescimento é a combinação de CNNs com outras abordagens de inteligência artificial, como redes neurais adversariais generativas (GANs), que podem ser usadas para gerar amostras sintéticas de células em casos de dados escassos, ampliando assim os conjuntos de dados de treinamento e melhorando a performance dos modelos (LAN *et al.*, 2020). Frid-Adar *et al.* (2018) demonstraram que GANs são eficazes na criação de imagens médicas sintéticas de alta qualidade, ajudando a aumentar a variabilidade do

conjunto de dados e melhorando a generalização dos modelos de CNN.

Esses avanços representam um passo significativo para a integração da inteligência artificial na prática clínica, oferecendo ferramentas poderosas para a personalização do tratamento do CCU, potencializando a precisão diagnóstica e a eficácia terapêutica. A validação desses modelos em estudos clínicos prospectivos e a avaliação em dados externos serão essenciais para confirmar sua utilidade e segurança em cenários clínicos reais (JHA *et al.*, 2023).

3 METODOLOGIA E DESENVOLVIMENTO

3.1 Proposta

O objetivo principal deste projeto é avaliar e comparar a performance de CNNs para a predição precoce de CCU a partir de imagens de Papanicolau. Para tanto, a metodologia proposta compreende duas etapas principais: pré-processamento das imagens provenientes do Papanicolau para identificação de células e classificação das células obtidas na etapa anterior conforme o TBS. A Figura 9 fornece uma representação visual detalhada deste processo.

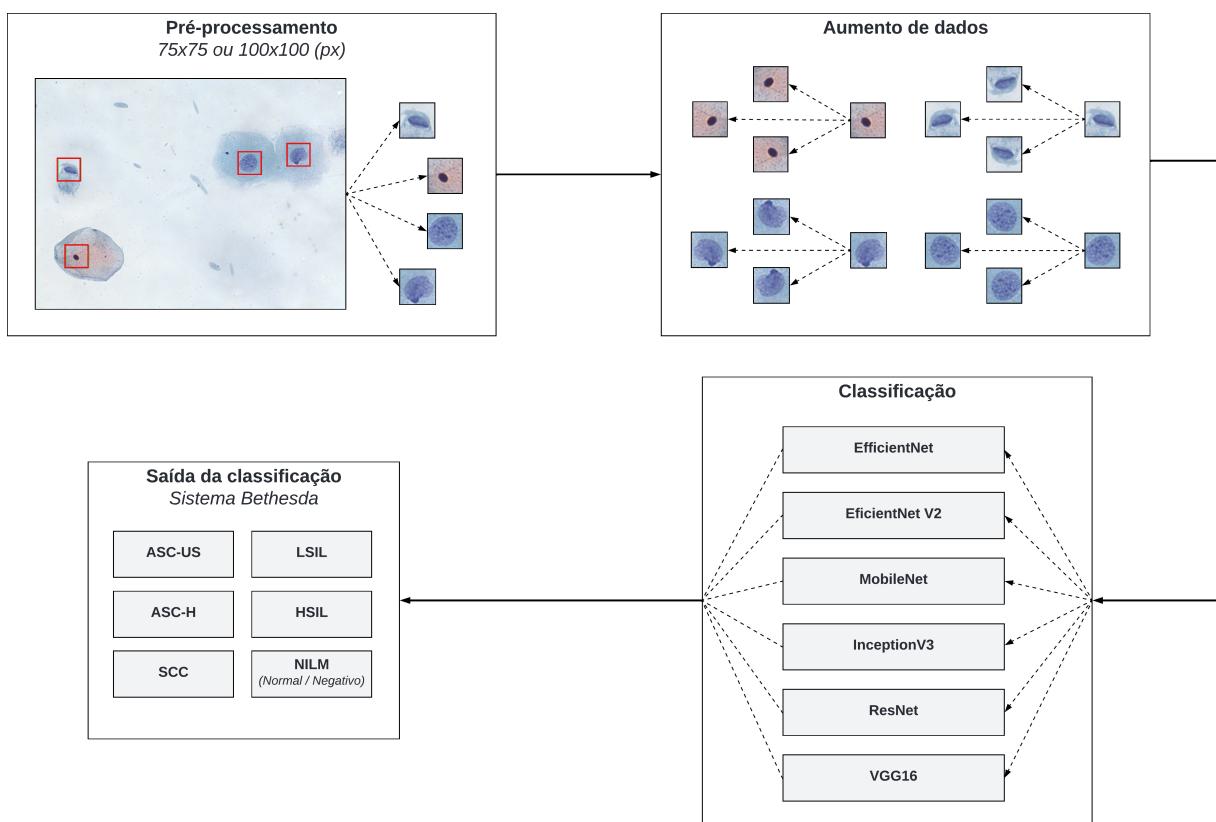


Figura 9 – Diagrama representativo do *pipeline* proposto compreendendo as etapas de 1) obtenção e pré-processamento das imagens; 2) aumento de dados; e 3) classificação.

3.1.1 Conjunto de dados

O conjunto de dados utilizado foi obtido a partir da plataforma CRIC (REZENDE *et al.*, 2021b), e é composto por 400 imagens de lâminas provenientes do Papanicolau e metadados contendo a identificação dos pontos centrais de cada uma das células e sua

respectiva classificação (Figura 10). Ao todo, o dataset possui 11.533 células classificadas manualmente por especialistas.

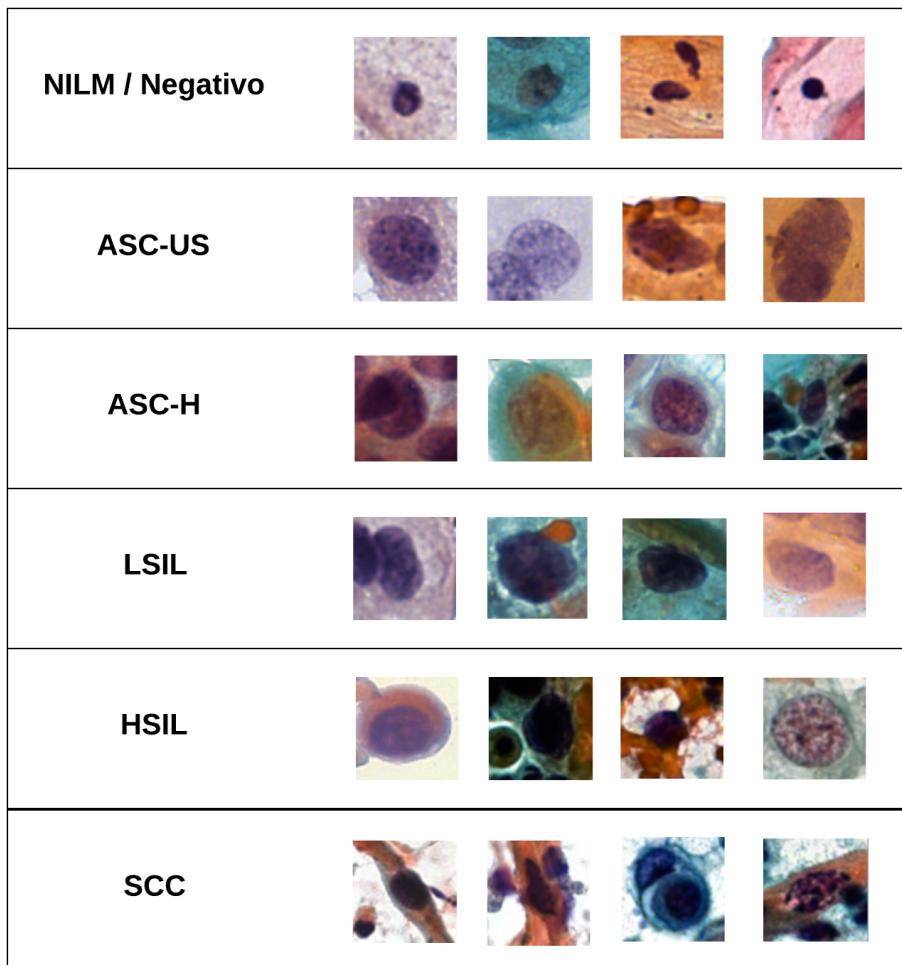


Figura 10 – Exemplificação de células e sua respectiva classificação conforme o Sistema Bethesda.

Além das imagens provenientes do Papanicolau e da classificação das células, o *dataset* utilizado fornece metadados contendo as coordenadas centrais de cada uma das células classificadas, simplificando o processo de pré-processamento e extração das células.

3.2 Pré-processamento

Para a realização do treinamento dos modelos para a classificação, as imagens passam por um pré-processamento que consiste na identificação e recorte das células a partir do *dataset* inicial. Partindo-se das coordenadas centrais de cada célula identificada no CRIC, são criados recortes de 75x75 e 100x100 pixels que serão utilizados para extrair cada uma das células presentes na imagem. Os recortes obtidos serão utilizados para a construção do *dataset* de células que será usado para treinamento e teste dos modelos de classificação.

As imagens obtidas foram organizadas em uma estrutura de diretórios que corresponde à classificação de cada uma das células e o tamanho do recorte. Ou seja, foram distribuídas em 6 diretórios (um para cada classificação possível) incluídos em dois diretórios, "75x75" e "100x100". Essa estrutura foi utilizada para possibilitar a utilização do método "*tf.keras.preprocessing.image_dataset_from_directory*" da biblioteca *Tensorflow* (ABADI *et al.*, 2015), que simplifica o carregamento dos dados e sua distribuição em *datasets* de treinamento e teste.

Para as etapas de treinamento e teste, os dados serão balanceados e divididos utilizando a escala inicial de 8x2 (80% para treinamento e 20% para teste). A Figura 11 mostra a distribuição das classes geradas a partir do *dataset* inicial.

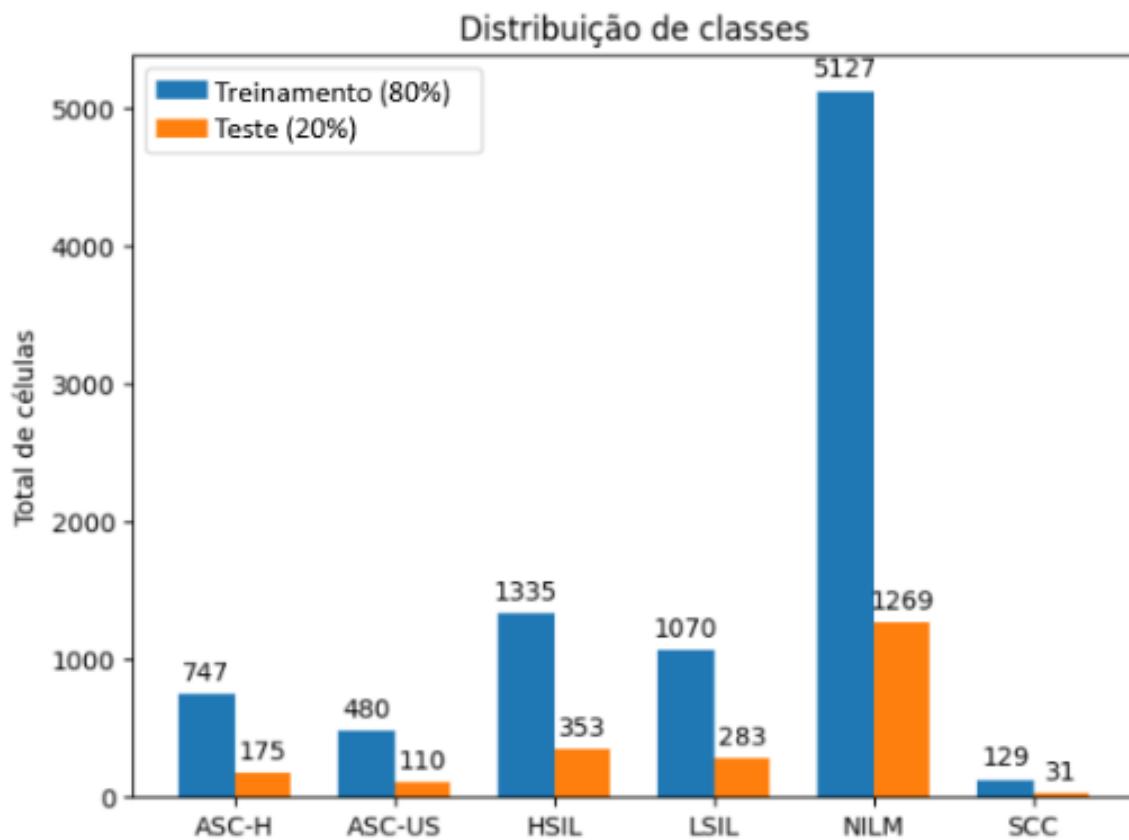


Figura 11 – Distribuição inicial das imagens entre as seis classificações possíveis.

Para reduzir a diferença de distribuição entre classes e aumentar a diversidade e a robustez do conjunto de dados para a classificação, são aplicadas técnicas de aumento de dados, como rotação, espelhamento e redimensionamento (Figura 12). Essas técnicas são cruciais para melhorar a eficiência do treinamento das CNNs, prevenindo *overfitting*, ou seja, garantindo que a classificação possa ser generalizada para dados não vistos anteriormente (PEREZ; WANG, 2017a; SHORTEN; KHOSHGOFTAAR, 2019). Além disso, o modelo é treinado em um conjunto de dados diversificado para evitar viéses (MEHRABI *et al.*,

2022). Como resultado do processo de aumento dados obteve-se uma distribuição mais equilibrada entre as classes possíveis, conforme ilustrado na Figura 13.

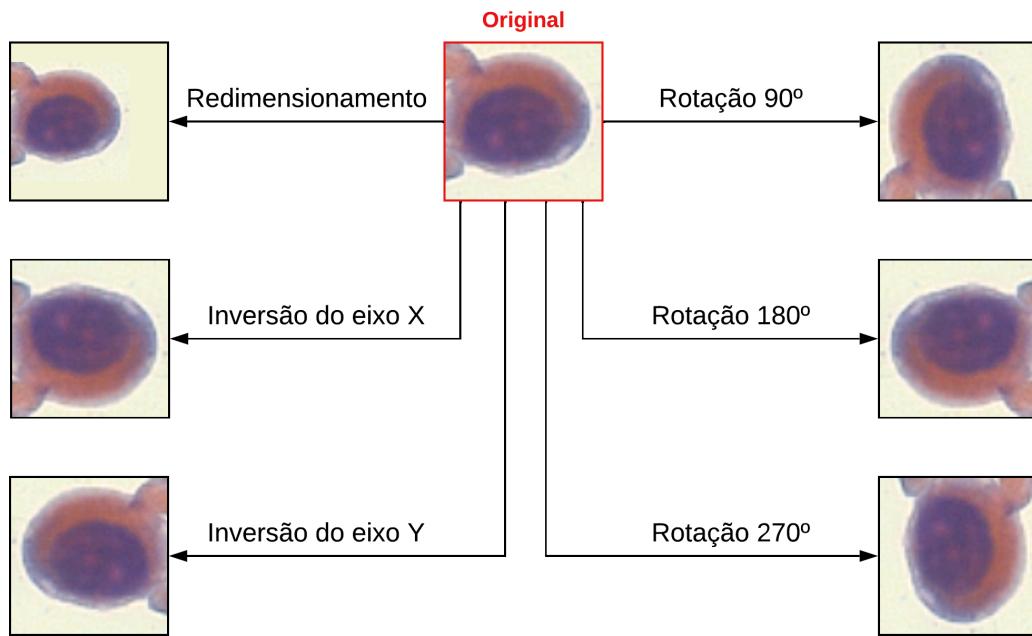


Figura 12 – Operações de transformação realizadas para aumento dos dados.

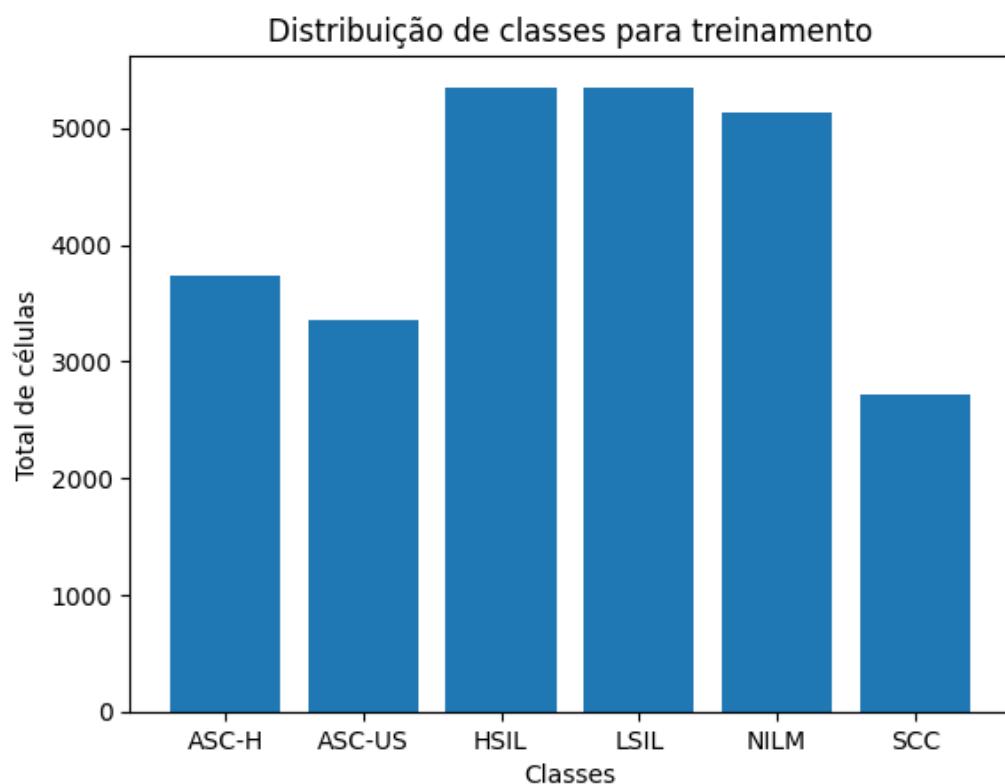


Figura 13 – Distribuição das imagens entre as seis classificações possíveis no *dataset* de treinamento após realização do aumento de dados.

Mesmo após o aumento dos dados, algumas classes ainda apresentaram uma quantidade consideravelmente menor de amostras quando comparadas às demais, como observado em ASC-H, ASC-US e, principalmente, SCC. Para lidar com esse desequilíbrio persistente, a função "*sklearn.utils.class_weight.compute_class_weight*" da biblioteca "*scikit-learn*" foi utilizada (PEDREGOSA *et al.*, 2011). Ao aplicar essa função, os pesos das classes são determinados com base em suas frequências, garantindo que o modelo leve em consideração as classes menos representadas de forma adequada, evitando que as classes majoritárias dominem o processo de aprendizagem. Essa abordagem é particularmente útil em tarefas de classificação onde o desequilíbrio de classes pode levar o modelo a apresentar um desempenho inferior nas classes minoritárias (PEDREGOSA *et al.*, 2011).

3.3 Classificação

Após o pré-processamento das imagens, a próxima etapa envolve o uso dos dados extraídos das células para realizar a classificação com base no TBS. O conjunto de dados obtido na etapa anterior é composto por imagens individuais de células, cada uma associada a um rótulo de classificação. Este rótulo indica o grau de saúde celular, variando entre normalidade e presença de características compatíveis com CCU (REZENDE *et al.*, 2021b).

3.3.1 Modelos de classificação

Conforme mostrado na Figura 9, as imagens de células pré-processadas são utilizadas para treinamento dos modelos de classificação para indicar a presença de células normais (NILM / negativa) ou células com sinais de alterações (ASC-US, ASC-H, LSIL, HSIL ou SCC).

Neste estudo, seis arquiteturas populares de CNNs foram comparadas no processo de classificação de células cervicais extraídas de exames de Papanicolau. As redes incluíram VGG16, ResNet (ResNet50), EfficientNet (EfficientNetB0), EfficientNetV2 (EfficientNetV2B0), InceptionV3 e MobileNetV3 (MobileNetV3Small). A escolha dessas arquiteturas se baseou na sua eficácia comprovada em tarefas de visão computacional e classificação de imagens médicas. O objetivo foi analisar o desempenho de cada modelo em termos de acurácia, sensibilidade e precisão.

Todos os modelos utilizaram suas arquiteturas base pré-treinadas no conjunto de dados ImageNet para facilitar o processo de transferência de aprendizado. Essa técnica permite reutilizar os pesos das redes previamente treinadas em grandes conjuntos de dados, adaptando-os para uma tarefa específica com um conjunto de dados menor, reduzindo o tempo de treinamento e melhorando a precisão (SHIN *et al.*, 2016a). As camadas da arquitetura base foram congeladas, permitindo que apenas as camadas adicionadas fossem ajustadas durante o treinamento (Figura 14).

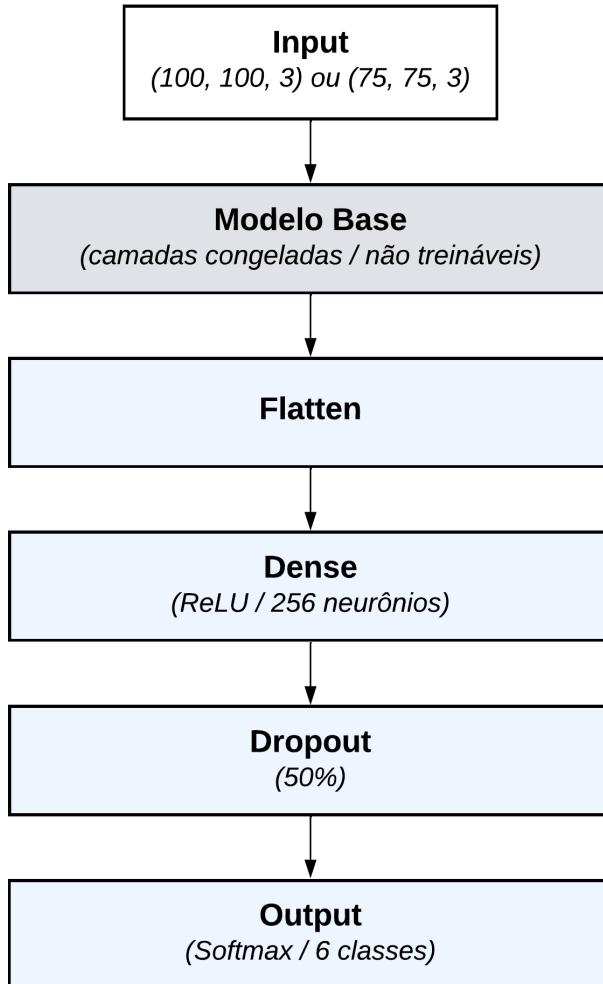


Figura 14 – Ilustração da rede gerada, consistindo 1) Das camadas congeladas do modelo base, pré-treinado; 2) Uma camada *Flatten*, para transformar as saídas convolucionais em uma única dimensão; 3) Uma camada *Dense* de 256 neurônios com ativação *ReLU*, para adicionar poder de decisão não-linear; 4) *Dropout* de 50% para evitar *overfitting*; e 5) Camada de *output Softmax* para classificação das seis categorias de células.

A maioria dos modelos foi compilada utilizando o otimizador *Adam* (KINGMA; BA, 2017), que combina as vantagens de adaptação de taxa de aprendizado com *momentums* e mostrou-se eficaz em uma ampla gama de tarefas de aprendizado profundo. O *Adam* foi escolhido pela sua robustez e velocidade de convergência. A função de perda usada foi a *categorical cross-entropy*, ideal para problemas de classificação multiclasse, como o deste estudo.

Para o modelo InceptionV3, o otimizador *Adam* foi substituído pelo *RMSprop*, pelo fato de que esse otimizador tende a funcionar melhor para arquiteturas mais profundas e complexas, onde o ajuste da taxa de aprendizado em diferentes direções pode ser mais eficaz na redução da função de perda (SZEGEDY *et al.*, 2016). O *RMSprop*, com uma taxa de aprendizado inicial de 1e-4, garantiu uma convergência mais estável para esse modelo

específico.

O treinamento foi realizado utilizando 100 épocas para todos os modelos, com *callbacks* para *EarlyStopping* e *ModelCheckpoint*. O *EarlyStopping* monitorou a função de perda do teste e interrompeu o treinamento caso a função de perda não melhorasse após 10 épocas, evitando *overfitting* e garantindo que o melhor modelo fosse salvo com base no menor valor de perda.

Cada modelo foi treinado e testado em um conjunto de dados balanceado após o aumento, com as métricas sendo monitoradas em cada época: acurácia, *recall* e precisão. Essas métricas foram escolhidas por fornecerem uma visão equilibrada do desempenho do modelo, especialmente em problemas de classificação com classes desbalanceadas, onde a acurácia isolada pode não refletir adequadamente a performance nas classes minoritárias.

Após o treinamento de cada um dos modelos base, os resultados foram avaliados considerando as métricas de perda e acurácia, precisão, *recall* e F1-score. Adicionalmente, para a análise mais aprofundada dos modelos, foram geradas matrizes de confusão e curvas ROC, utilizando as previsões realizadas no conjunto de teste. Essas ferramentas permitiram avaliar o desempenho dos modelos na diferenciação das classes, identificando padrões de erro e possíveis áreas de melhoria.

4 ANÁLISE DOS RESULTADOS

Neste estudo, avaliamos seis arquiteturas de CNNs: VGG16, ResNet50, EfficientNetB0, EfficientNetV2B0, InceptionV3 e MobileNetV3Small. O objetivo foi analisar o desempenho dessas redes na classificação de células cervicais em exames de Papanicolau, utilizando blocos de imagem de 75x75 e 100x100 pixels, centrados na célula de interesse. As redes foram treinadas e testadas em um conjunto de dados desbalanceado, composto por células normais (NILM) e classes minoritárias (ASC-US, ASC-H, LSIL, HSIL e SCC). As métricas utilizadas para a avaliação incluem acurácia, *recall*, precisão, F1-score e AUC. Também foram incluídas análises visuais como curvas ROC, gráficos de acurácia e perda, e a matriz de confusão. O código fonte e demais dados gerados podem ser acessados através do repositório localizado em <<https://github.com/lucassoamaral/cervical-cnn-comparison>>.

Na execução realizada para *inputs* com resolução 100x100 pixels, todos os modelos foram testados utilizando tanto *Adam*, quanto *RMSprop* como otimizadores, com base nos resultados iniciais que sugeriram melhorias com este otimizador para modelos mais complexos, como o InceptionV3.

Ao treinar os modelos com blocos de 75x75 pixels, o ResNet50 apresentou o melhor desempenho global, com uma acurácia de 76% e um F1-score ponderado de 0.7595. O modelo demonstrou boa capacidade de classificação em todas as classes, inclusive nas minoritárias, como a SCC, para a qual obteve um F1-score de 0.58. Além disso, suas curvas ROC indicaram capacidade discriminatória elevada, com AUC de 0.93, 0.93 e 0.98 para as classes ASC-H, NILM e SCC, respectivamente. Esses resultados mostram que o modelo foi capaz de classificar de maneira confiável as classes mais desafiadoras e as mais frequentes, mantendo uma taxa alta de VP e baixa de FP.

A EfficientNetB0 também apresentou um desempenho sólido, com o F1-score ponderado de 0.7247, além de bons valores para ROC-AUC, em especial para NILM (0.94) e SCC (0.96). Por outro lado, a VGG16, apesar de apresentar resultados estáveis, com um F1-score ponderado de 0.6858, demonstrou bons resultados para classes mais comuns, como NILM e HSIL, com ROC-AUC de 0.92 para ambas. O modelo, no entanto, apresentou precisão mais baixa para classes minoritárias, como ASC-US e SCC, com valores de precisão de 0.27 e 0.29, respectivamente.

No caso da InceptionV3, observou-se uma melhora significativa ao substituir o otimizador *Adam* pelo *RMSprop*, apresentando uma variação de 0.0115 para 0.5017 para F1-score ponderado. Apesar de ter apresentado uma melhora na classificação de classes majoritárias com *RMSprop*, o modelo apresentou resultados abaixo do esperado, com uma acurácia de 46% e um F1-score ponderado de 0.5017. Esse desempenho pode estar

relacionado à complexidade da arquitetura do InceptionV3, que pode não ter se beneficiado do uso de blocos menores.

A Tabela 2 apresenta os resultados completos de todas as métricas coletadas para todos os modelos com blocos de 75x75 pixels.

Tabela 2 – Métricas por classe e modelo (75x75).

Classe	Modelo	Precisão	Recall	F1-Score	ROC-AUC
ASC-H	EfficientNetB0	0.43	0.55	0.48	0.91
	EfficientNetB0V2	0.40	0.50	0.44	0.89
	MobileNet	0.31	0.45	0.37	0.85
	InceptionV3 (<i>RMSprop</i>)	0.14	0.39	0.20	0.77
	ResNet50	0.48	0.66	0.55	0.93
	VGG16	0.31	0.55	0.40	0.88
ASC-US	EfficientNetB0	0.34	0.52	0.41	0.87
	EfficientNetB0V2	0.28	0.54	0.3	0.87
	MobileNet	0.31	0.54	0.39	0.87
	InceptionV3 (<i>RMSprop</i>)	0.13	0.65	0.22	0.80
	ResNet50	0.62	0.37	0.47	0.87
	VGG16	0.27	0.57	0.37	0.87
HSIL	EfficientNetB0	0.74	0.68	0.71	0.95
	EfficientNetB0V2	0.73	0.64	0.68	0.95
	MobileNet	0.70	0.54	0.61	0.94
	InceptionV3 (<i>RMSprop</i>)	0.69	0.19	0.29	0.87
	ResNet50	0.82	0.72	0.77	0.96
	VGG16	0.67	0.54	0.60	0.92
LSIL	EfficientNetB0	0.63	0.24	0.35	0.87
	EfficientNetB0V2	0.59	0.26	0.36	0.88
	MobileNet	0.53	0.20	0.29	0.87
	InceptionV3 (<i>RMSprop</i>)	0.25	0.01	0.01	0.71
	ResNet50	0.49	0.54	0.52	0.88
	VGG16	0.53	0.23	0.32	0.83
NILM	EfficientNetB0	0.85	0.91	0.88	0.94
	EfficientNetB0V2	0.88	0.89	0.89	0.94
	MobileNet	0.85	0.88	0.86	0.93
	InceptionV3 (<i>RMSprop</i>)	0.91	0.63	0.74	0.86
	ResNet50	0.88	0.86	0.87	0.93
	VGG16	0.88	0.85	0.8	0.92
SCC	EfficientNetB0	0.46	0.61	0.53	0.96
	EfficientNetB0V2	0.30	0.68	0.41	0.96
	MobileNet	0.23	0.58	0.33	0.96
	InceptionV3 (<i>RMSprop</i>)	0.06	0.35	0.10	0.85
	ResNet50	0.48	0.74	0.58	0.98
	VGG16	0.29	0.48	0.37	0.92

Dada a melhora no desempenho apresentada pela InceptionV3 quando utilizando o otimizador *RMSprop*, para os blocos de 100x100 a comparação dos otimizadores foi

realizada para todos os modelos avaliados, tendo sido observada uma ligeira melhora nos resultados obtidos com o otimizador *RMSprop*.

Com blocos de 100x100 e otimizador *RMSprop*, os modelos apresentaram uma melhora significativa. Os modelos ResNet50 e EfficientNetB0V2 se destacaram novamente, com 80% de acurácia e F1-score ponderado de 0.7987 e 0.8036, respectivamente. Esses resultados demonstram a capacidade desses modelos de capturar informações contextuais adicionais proporcionadas pelos blocos maiores. Notavelmente, o EfficientNetB0V2 obteve um F1-score de 0.60 para SCC, o que indica uma melhor generalização para as classes minoritárias em comparação com *inputs* de menor resolução.

O VGG16 também apresentou uma melhora em seu desempenho geral, com uma acurácia de 69% e um F1-score ponderado de 0.6936. Porém, continuou apresentando dificuldades em classes minoritárias, como ASC-US e LSIL.

O InceptionV3, mesmo com o otimizador *RMSprop*, continuou a apresentar desempenho inferior em comparação aos outros modelos, obtendo uma acurácia de apenas 38% e um F1-score ponderado de 0.4279.

A Tabela 3 apresenta os resultados detalhados para blocos de 100x100 pixels.

Tabela 3 – Métricas por classe e modelo (100x100).

Classe	Modelo	Precisão	Recall	F1-Score	ROC-AUC
ASC-H	EfficientNetB0	0.49	0.62	0.55	0.93
	EfficientNetB0V2	0.56	0.77	0.64	0.95
	MobileNet	0.51	0.51	0.51	0.91
	InceptionV3	0.13	0.58	0.21	0.67
	ResNet50	0.61	0.72	0.66	0.95
	VGG16	0.36	0.60	0.45	0.88
ASC-US	EfficientNetB0	0.42	0.35	0.39	0.87
	EfficientNetB0V2	0.41	0.51	0.45	0.90
	MobileNet	0.42	0.55	0.47	0.88
	InceptionV3	0.13	0.65	0.21	0.76
	ResNet50	0.47	0.42	0.44	0.90
	VGG16	0.27	0.44	0.34	0.82
HSIL	EfficientNetB0	0.83	0.68	0.74	0.97
	EfficientNetB0V2	0.81	0.81	0.81	0.97
	MobileNet	0.78	0.69	0.73	0.96
	InceptionV3	0.64	0.10	0.18	0.87
	ResNet50	0.84	0.84	0.84	0.98
	VGG16	0.76	0.55	0.63	0.94
LSIL	EfficientNetB0	0.61	0.44	0.51	0.89
	EfficientNetB0V2	0.71	0.49	0.58	0.91
	MobileNet	0.56	0.50	0.53	0.89
	InceptionV3	0.11	0.01	0.01	0.68
	ResNet50	0.66	0.46	0.55	0.91
	VGG16	0.46	0.29	0.36	0.83
NILM	EfficientNetB0	0.85	0.92	0.89	0.94
	EfficientNetB0V2	0.92	0.90	0.91	0.96
	MobileNet	0.88	0.89	0.89	0.94
	InceptionV3	0.91	0.50	0.65	0.81
	ResNet50	0.88	0.92	0.90	0.95
	VGG16	0.86	0.85	0.86	0.91
SCC	EfficientNetB0	0.49	0.58	0.53	0.96
	EfficientNetB0V2	0.49	0.77	0.60	0.99
	MobileNet	0.35	0.65	0.45	0.96
	InceptionV3	0.08	0.26	0.12	0.82
	ResNet50	0.53	0.65	0.58	0.97
	VGG16	0.26	0.58	0.36	0.91

A análise das curvas ROC para cada modelo confirma os resultados observados. As Figuras 15 e 16 ilustram a performance discriminatória dos modelos que apresentaram a melhor performance (ResNet50 e EfficientNetV2B0). Os gráficos mostram que esses modelos são eficazes na classificação de classes mais desafiadoras, como SCC e HSIL, mantendo AUCs superiores a 0.95 e sua alta capacidade discriminatória, com baixa taxa de FP.

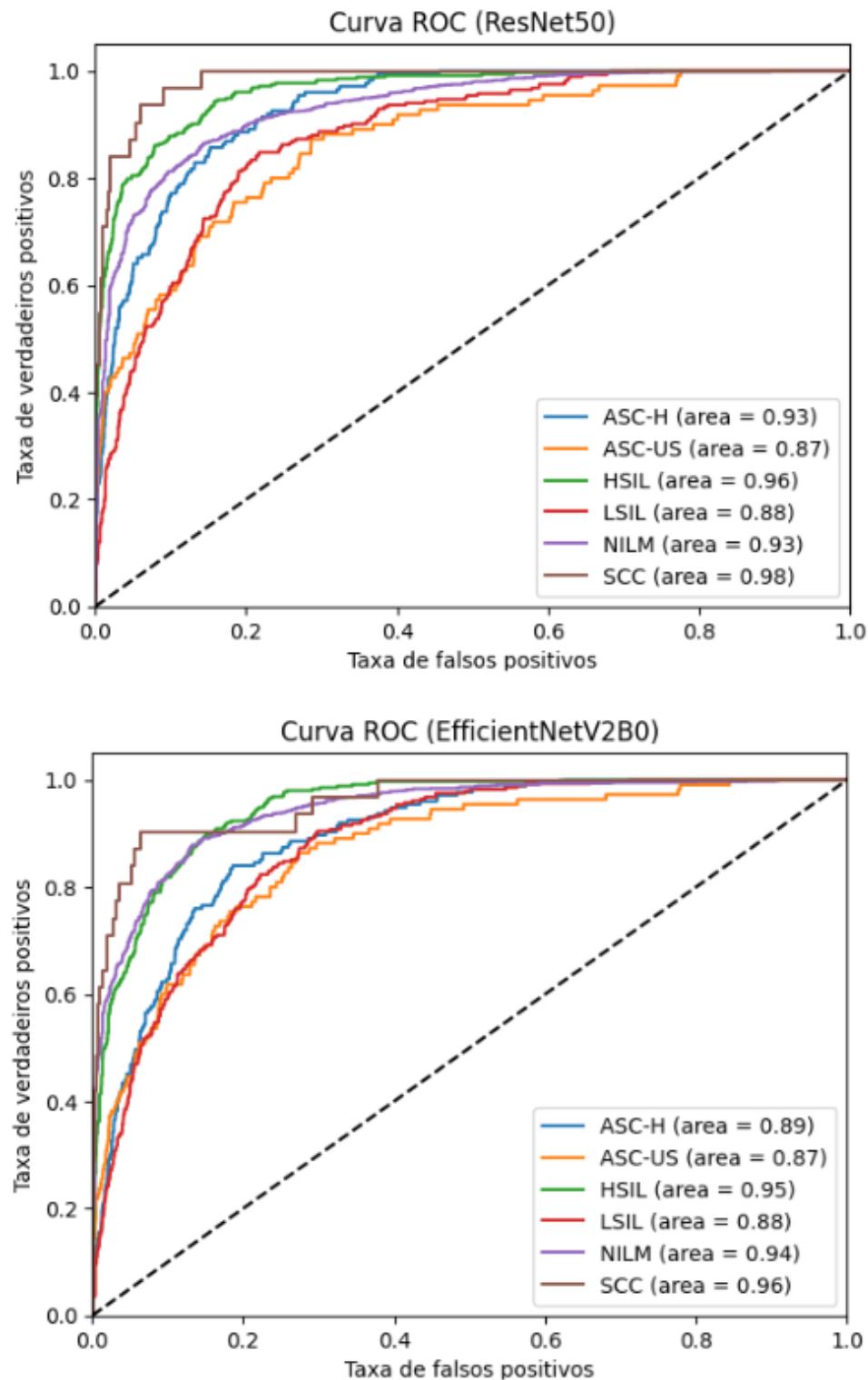


Figura 15 – Curva ROC para ResNet50 e EfficientNetB0V2 (75x75).

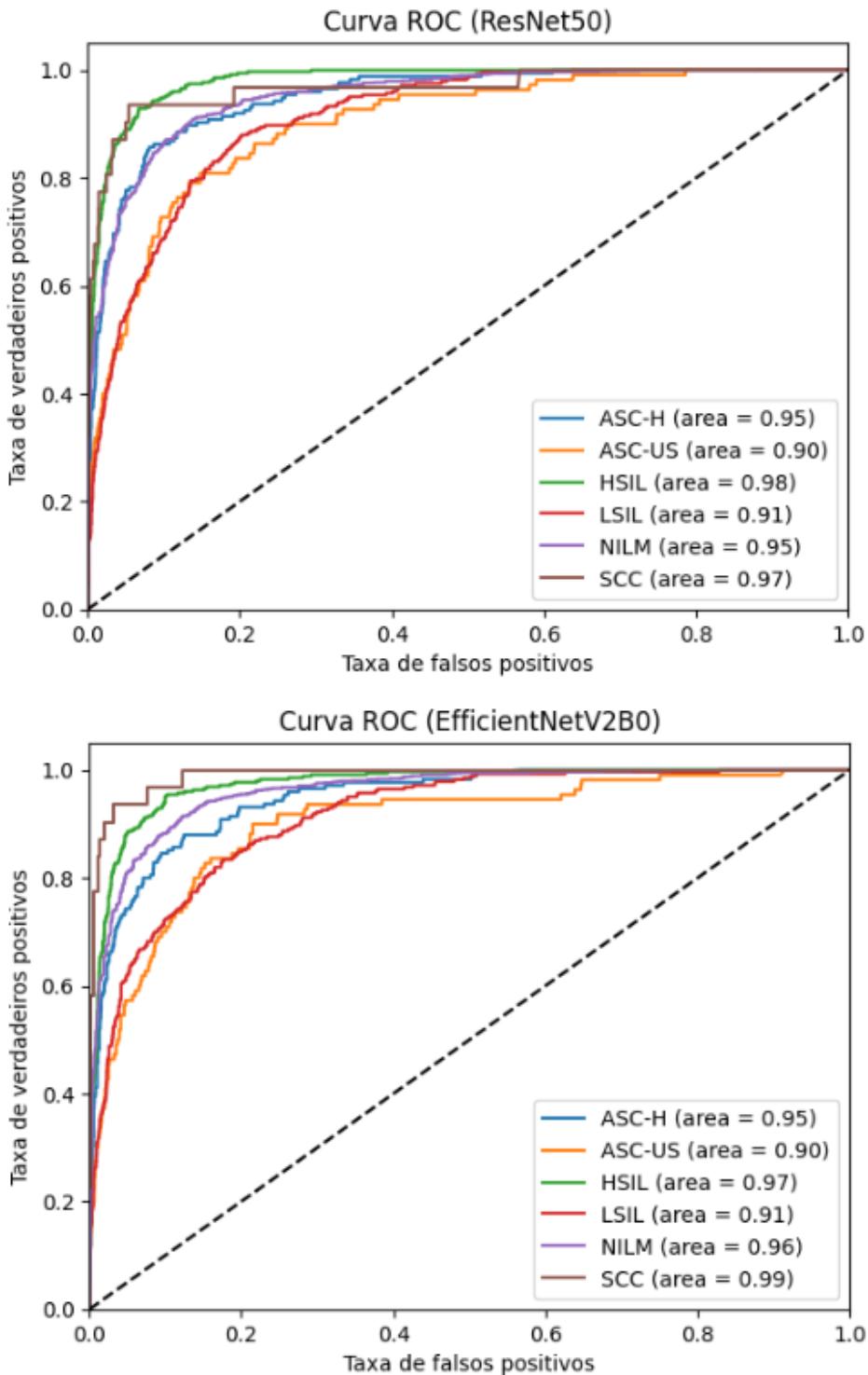


Figura 16 – Curva ROC para ResNet50 e EfficientNetB0V2 (100x100).

No que se refere à acurácia e perda, observamos na Figura 17 que tanto o ResNet50 quanto o EfficientNetV2B0 exibiram um padrão de convergência rápido, com acurácia em crescimento contínuo e perda mínima após as primeiras épocas. Esses gráficos confirmam que os modelos conseguiram aprender as características das imagens de maneira eficiente e estável ao longo do treinamento.

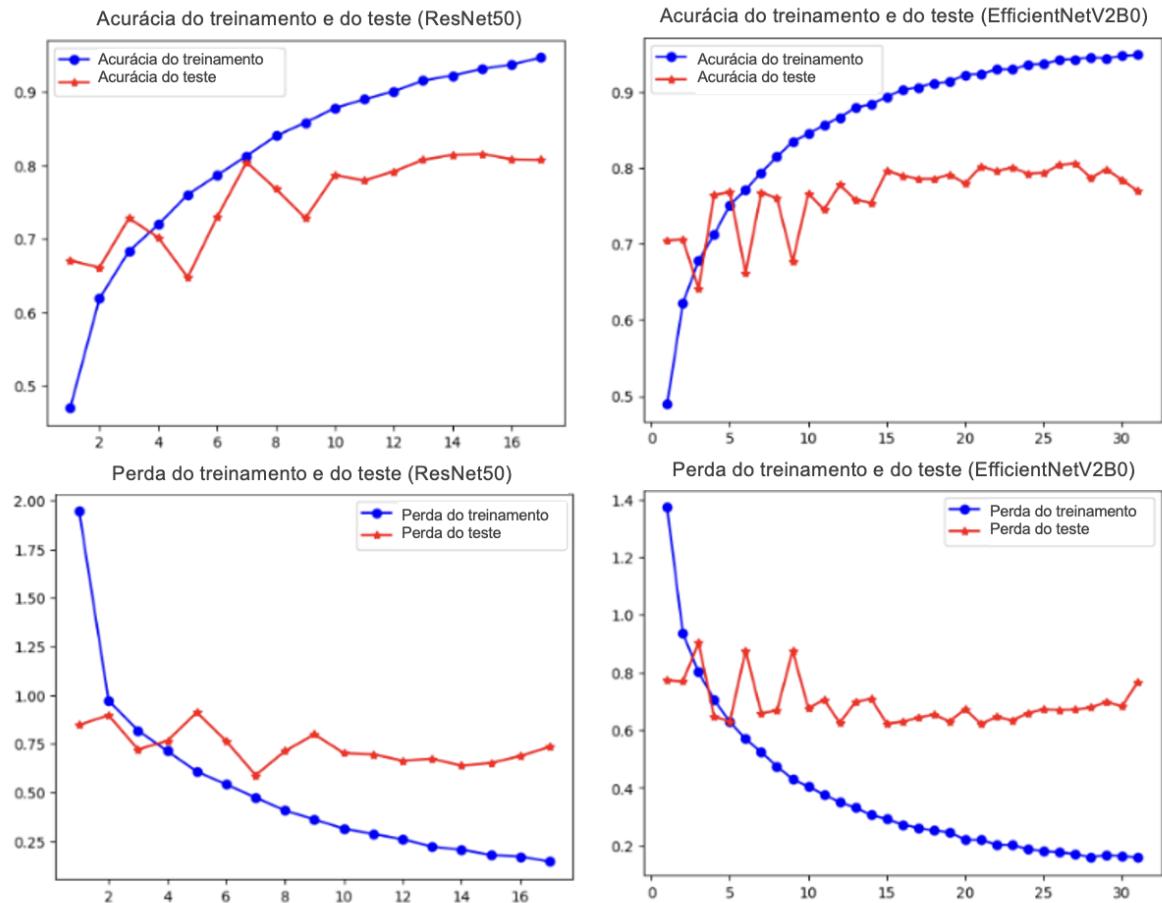


Figura 17 – Evolução da acurácia e perda ao longo das épocas para ResNet50 e EfficientNetB0V2 (100x100).

A matriz de confusão (Figura 18) para o EfficientNetV2B0 revela que o modelo consegue classificar corretamente a maioria das instâncias das classes NILM e HSIL, com poucas confusões. No entanto, há uma ligeira confusão entre as classes LSIL e HSIL, o que impacta a precisão geral em classes menos representadas como ASCUS.

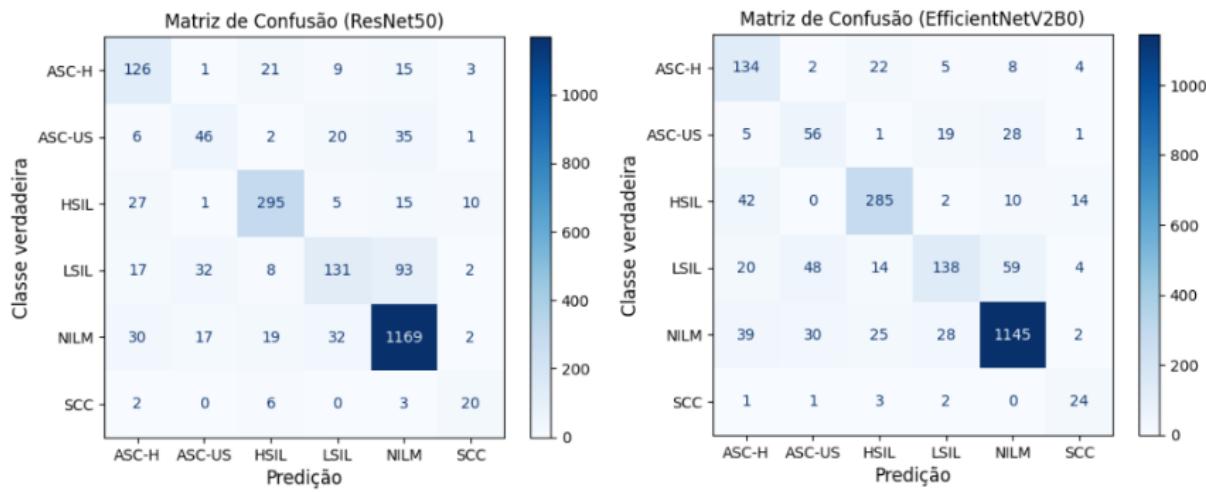


Figura 18 – Matriz de confusão para ResNet50 e EfficientNetB0V2 (100x100).

5 CONCLUSÕES

Este trabalho demonstrou a eficácia das CNNs na classificação de células cervicais em exames de Papanicolau, utilizando técnicas avançadas como transferência de aprendizado, aumento de dados e ajustes de pesos de classe para lidar com o desbalanceamento dos dados. Comparando diferentes arquiteturas, como VGG16, ResNet50, EfficientNetB0, EfficientNetV2B0, InceptionV3 e MobileNetV3Small, foi possível observar que cada modelo apresentou características específicas de desempenho.

A análise dos modelos treinados nas resoluções 75x75 e 100x100 revela que o EfficientNetV2B0 e o ResNet50 foram os modelos com o melhor desempenho geral, com destaque para o EfficientNetV2B0, que apresentou F1-score ponderado de 0.8036 e AUCs consistentemente elevadas em todas as classes. O MobileNetV3Small, apesar de ser uma arquitetura mais leve, também obteve bons resultados, especialmente nas classes mais representadas, tornando-se uma opção eficiente para cenários com recursos computacionais limitados. A adoção do *RMSprop* como otimizador na resolução 100x100 trouxe melhorias significativas para a maioria dos modelos, com destaque para o VGG16 e o InceptionV3, que se beneficiaram da mudança, embora o último ainda enfrente desafios com classes de menor representatividade.

Os resultados indicam que, mesmo em cenários de classes desbalanceadas, as CNNs, quando corretamente ajustadas, conseguem atingir altos níveis de acurácia e desempenho, oferecendo um método robusto e confiável para o diagnóstico automatizado de exames citológicos. A aplicação de pesos de classe foi fundamental para assegurar que o desempenho nas classes minoritárias não fosse comprometido, especialmente na detecção de células anormais.

Além disso, o uso de técnicas de aumento de dados permitiu a criação de um conjunto de treinamento mais robusto, essencial para melhorar a generalização dos modelos e evitar o *overfitting*. A análise do F1-score, em conjunto com métricas como precisão e *recall*, forneceu uma visão abrangente do desempenho dos modelos em um cenário de dados desbalanceados, demonstrando que essas métricas são essenciais para uma avaliação justa e precisa.

Em suma, a utilização de CNNs combinadas com técnicas de aprendizado profundo e ajustes apropriados para desequilíbrio de classes oferece uma abordagem promissora para o diagnóstico assistido por computador em exames de Papanicolau. No entanto, futuros trabalhos podem focar em aumentar a interpretabilidade dos modelos e explorar arquiteturas híbridas, como GANs, para expandir ainda mais a eficácia dessas abordagens na prática clínica.

REFERÊNCIAS

- ABADI, M. *et al.* **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Available at: <<https://www.tensorflow.org/>>.
- ABDALLA, P.; MOHAMMED, B.; SAEED, A. The impact of image augmentation techniques of mri patients in deep transfer learning networks for brain tumor detection. **Journal of Electrical Systems and Information Technology**, v. 10, p. 51, 2023.
- AGARWAL, R. *et al.* Automatic mass detection in mammograms using deep convolutional neural networks. **J Med Imaging (Bellingham)**, J Med Imaging (Bellingham).
- ARBYN, M. *et al.* Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. **The Lancet Global Health**, v. 8, n. 2, p. e191–e203, 2020.
- BAKASA, W.; VIRIRI, S. Vgg16 feature extractor with extreme gradient boost classifier for pancreas cancer prediction. **Journal of Imaging**, v. 9, n. 7, p. 138, 2023.
- BOSCH, F. X.; SANJOSÉ, S. de. Human papillomavirus and cervical cancer – burden and assessment of causality. **Journal of Clinical Pathology**, v. 73, n. 6, p. 321–329, 2020.
- BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. **Neural Networks**, v. 106, p. 249–259, 2018. ISSN 0893-6080. Available at: <<https://www.sciencedirect.com/science/article/pii/S0893608018302107>>.
- BURD, E. M. Human papillomavirus and cervical cancer. **CLINICAL MICROBIOLOGY REVIEWS**, 2003.
- BUSKWOFIE, A.; DAVID-WEST, G.; CLARE, C. A. A review of cervical cancer: incidence and disparities. **Journal of the National Medical Association**, Elsevier, v. 112, n. 2, p. 229–232, 2020.
- CASTRO, C. L. d.; BRAGA, A. P. Aprendizado supervisionado com conjuntos de dados desbalanceados. **Sba: Controle and Automação Sociedade Brasileira de Automática**, Sociedade Brasileira de Automática, v. 22, n. 5, p. 441–466, Sep 2011. ISSN 0103-1759. Available at: <<https://doi.org/10.1590/S0103-17592011000500002>>.
- CHEN, W. *et al.* A robust approach for multi-type classification of brain tumor using deep feature fusion. **Frontiers in Neuroscience**, v. 18, p. 1288274, 2024.
- CLEMENT, D. *et al.* Multi-class breast cancer histopathological image classification using multi-scale pooled image feature representation (mpifr) and one-versus-one support vector machines. **Applied Sciences**, v. 13, n. 1, 2023. ISSN 2076-3417. Available at: <<https://www.mdpi.com/2076-3417/13/1/156>>.
- de Sanjose, S. *et al.* Human papillomavirus vaccine disease impact beyond expectations. **Current Opinion in Virology**, v. 39, p. 16–22, 2019. ISSN 1879-6257. Viruses and cancer. Available at: <<https://www.sciencedirect.com/science/article/pii/S1879625719300434>>.

de Villiers, E.-M. Cross-roads in the classification of papillomaviruses. **Virology**, v. 445, n. 1, p. 2–10, 2013. ISSN 0042-6822. Special Issue: The Papillomavirus Episteme. Available at: <<https://www.sciencedirect.com/science/article/pii/S0042682213002456>>.

ELAZIZ, M. A. *et al.* Boosting covid-19 image classification using mobilenetv3 and aquila optimizer algorithm. **Entropy**, v. 23, n. 11, 2021. ISSN 1099-4300. Available at: <<https://www.mdpi.com/1099-4300/23/11/1383>>.

FAWCETT, T. An introduction to roc analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Available at: <<https://www.sciencedirect.com/science/article/pii/S016786550500303X>>.

FAWCETT, T. An introduction to roc analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Available at: <<https://www.sciencedirect.com/science/article/pii/S016786550500303X>>.

FERLAY, J. *et al.* Cancer incidence and mortality patterns in europe: Estimates for 40 countries and 25 major cancers in 2018. **European journal of cancer**, Elsevier, v. 103, p. 356–387, 2018.

FERLAY, J. *et al.* **Global Cancer Observatory: Cancer Today**. 2020. Lyon, France: International Agency for Research on Cancer. Accessed: 20 Janeiro 2024. Available at: <<https://gco.iarc.fr/today>>.

FRID-ADAR, M. *et al.* Synthetic data augmentation using gan for improved liver lesion classification. In: **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.: s.n.], 2018. p. 289–293.

GARDNER, K. *et al.* Label free identification of different cancer cells using deep learning-based image analysis. **APL Machine Learning**, v. 1, n. 2, p. 026110, 05 2023. ISSN 2770-9019. Available at: <<https://doi.org/10.1063/5.0141730>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.: s.n.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

GU, C.; LEE, M. Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia x-ray images. **Bioengineering**, 2024.

GUAN, Q. *et al.* Inception-v3 model for differential diagnosing of lymph node in cytological images: a pilot study. **Ann Transl Med**, 2019.

HE, K. *et al.* Deep residual learning for image recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HOU, X. *et al.* Artificial intelligence in cervical cancer screening and diagnosis. **Front Oncol**, 2022.

HOWARD, A. *et al.* Searching for mobilenetv3. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2019.

HUH, M.; AGRAWAL, P.; EFROS, A. A. What makes imagenet good for transfer learning? **CoRR**, abs/1608.08614, 2016. Available at: <<http://arxiv.org/abs/1608.08614>>.

- ISLAM, T. *et al.* A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. **Healthcare Analytics**, v. 5, p. 100340, 2024. ISSN 2772-4425. Available at: <<https://www.sciencedirect.com/science/article/pii/S277244252400042X>>.
- JALLAH, J. K.; ANJANKAR, A.; NANKONG, F. A. Public health approach in the elimination and control of cervical cancer: A review. **Cureus**, 2023.
- JANTZEN, J. *et al.* Pap-smear benchmark data for pattern classification. **Nature inspired smart information systems (NiSIS 2005)**, p. 1–9, 2005.
- JHA, A. K. *et al.* Systematic review and meta-analysis of prediction models used in cervical cancer. **Artificial Intelligence in Medicine**, Elsevier, p. 102549, 2023.
- JIANG, X. *et al.* Deep learning for medical image-based cancer diagnosis. **Cancers (Basel)**, 2023.
- JOHNSON, C. A. *et al.* Cervical cancer: An overview of pathophysiology and management. **Seminars in Oncology Nursing**, v. 35, n. 2, p. 166–174, 2019. ISSN 0749-2081. Gynecology Oncology. Available at: <<https://www.sciencedirect.com/science/article/pii/S0749208119300142>>.
- KIM, H. E. *et al.* Transfer learning for medical image classification: a literature review. **BMC Med Imaging**, 2022.
- KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2017. Available at: <<https://arxiv.org/abs/1412.6980>>.
- LAN, L. *et al.* Generative adversarial networks and its applications in biomedical informatics. **Front Public Health**, 2020.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.
- LI, M. *et al.* Medical image analysis using deep learning algorithms. **Front Public Health**, 2023.
- MARQUES, G.; AGARWAL, D.; DÍEZ, I. de la T. Automated medical diagnosis of covid-19 through efficientnet convolutional neural network. **Appl Soft Comput**, 2020.
- MEGGIOLARO, A. *et al.* The role of pap test screening against cervical cancer: a systematic review and meta-analysis. **Clin Ter**, v. 167, n. 4, p. 124–39, 2016.
- MEHRABI, N. *et al.* **A Survey on Bias and Fairness in Machine Learning**. 2022.
- MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array**, v. 16, p. 100258, 2022. ISSN 2590-0056. Available at: <<https://www.sciencedirect.com/science/article/pii/S2590005622000911>>.
- NAIDU, G.; ZUVA, T.; SIBANDA, E. M. A review of evaluation metrics in machine learning algorithms. In: SILHAVY, R.; SILHAVY, P. (ed.). **Artificial Intelligence Application in Networks and Systems**. Cham: Springer International Publishing, 2023. p. 15–25. ISBN 978-3-031-35314-7.

NAYAR, R.; WILBUR, D. C. **The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes.** [S.l.: s.n.]: Springer, 2015.

OLUSOLA, P. *et al.* Human papilloma virus-associated cervical cancer and health disparities. **Cells**, p. 1–12, 2019.

ORGANIZATION, W. H. **Global strategy to accelerate the elimination of cervical cancer as a public health problem.** Geneva: World Health Organization, 2020.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, 2010.

PANGARKAR, M. A. The bethesda system for reporting cervical cytology. **Cytojournal**, 2022.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PEREZ, L.; WANG, J. **The Effectiveness of Data Augmentation in Image Classification using Deep Learning.** 2017.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. **arXiv**, 2017.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. **CoRR**, abs/1712.04621, 2017. Available at: <<http://arxiv.org/abs/1712.04621>>.

PIMPLE, S. A.; MISHRA, G. A. Global strategies for cervical cancer prevention and screening. **Minerva Ginecol**, 2019.

RAJARAM, S.; GUPTA, B. Screening for cervical cancer: Choices and dilemmas. **Indian J Med Res**, p. 210–220, 2021.

REZENDE, M. T. *et al.* Cric searchable image database as a public platform for conventional pap smear cytology data. **Scientific Data**, 2021.

REZENDE, M. T. *et al.* Cric searchable image database as a public platform for conventional pap smear cytology data. **Scientific data**, Nature Publishing Group UK London, v. 8, n. 1, p. 151, 2021.

SACHAN, P. L. *et al.* A study on cervical cancer screening using pap smear test and clinical correlation. **Asia-Pacific journal of oncology nursing**, Elsevier, v. 5, n. 3, p. 337–341, 2018.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. **PLOS ONE**, Public Library of Science, v. 10, n. 3, p. 1–21, 03 2015. Available at: <<https://doi.org/10.1371/journal.pone.0118432>>.

SELVARAJU, R. R. *et al.* Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. **CoRR**, abs/1610.02391, 2016. Available at: <<http://arxiv.org/abs/1610.02391>>.

- SHIN, H.-C. *et al.* Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. **IEEE Transactions on Medical Imaging**, v. 35, n. 5, p. 1285–1298, 2016.
- SHIN, H.-C. *et al.* Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. **IEEE Transactions on Medical Imaging**, v. 35, n. 5, p. 1285–1298, 2016.
- SHORTEN, C.; KHOSHGOFTAAR, T. A survey on image data augmentation for deep learning. **Journal of Big Data**, p. 1–48, 2019.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
- STEBEN, M.; DUARTE-FRANCO, E. Human papillomavirus infection: Epidemiology and pathophysiology. **Gynecologic Oncology**, v. 107, n. 2, Supplement, p. S2–S5, 2007. ISSN 0090-8258. Reducing the Burden of HPV and HPV Related Diseases. Available at: <<https://www.sciencedirect.com/science/article/pii/S0090825807005446>>.
- STEWART, B. W.; WILD, C. P. (ed.). **World cancer report 2014**. Lyon, France: International Agency for Research on Cancer, 2014.
- SZEGEDY, C. *et al.* Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **International Conference on Machine Learning (ICML)**. [S.l.: s.n.], 2019.
- TAN, M.; LE, Q. V. Efficientnetv2: Smaller models and faster training. **CoRR**, abs/2104.00298, 2021. Available at: <<https://arxiv.org/abs/2104.00298>>.
- THARWAT, A. Classification assessment methods. **Applied Computing and Informatics**, 2018.
- TUMMALA, S. *et al.* Efficientnetv2 based ensemble model for quality estimation of diabetic retinopathy images from deepdrid. **Diagnostics (Basel)**, 2023.
- VU, M. *et al.* Cervical cancer worldwide. **Current Problems in Cancer**, 2018.
- WANG, T. *et al.* Updates in cervical cancer screening guidelines, the bethesda system for reporting cervical cytology, and clinical management recommendations. **J Clin Transl Pathol**, 2023.
- WHO. A study on cervical cancer screening using pap smear test and clinical correlation. In: GENEVA: WORLD HEALTH ORGANIZATION. [S.l.: s.n.], 2020.
- YAMASHITA, R. *et al.* Convolutional neural networks: an overview and application in radiology. **Insights Imaging**, 2018.
- YULDASHEV, Y. *et al.* Parking lot occupancy detection with improved mobilenetv3. **Sensors (Basel)**, 2023.

ZEWDE, E. T. *et al.* Machine learning based stroke segmentation and classification from ct-scan: A survey. In: DEBELEE, T. G. *et al.* (ed.). **Pan-African Conference on Artificial Intelligence**. Cham: Springer Nature Switzerland, 2024. p. 3–45. ISBN 978-3-031-57624-9.

ZHANG, L. *et al.* Deeppap: Deep convolutional networks for cervical cell classification. **IEEE Journal of Biomedical and Health Informatics**, v. 21, n. 6, p. 1633–1643, 2017.

ZHAO, Z. *et al.* A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. **Expert Systems with Applications**, v. 242, p. 122807, 2024. ISSN 0957-4174. Available at: <<https://www.sciencedirect.com/science/article/pii/S0957417423033092>>.

ZOPH, B. *et al.* Learning transferable architectures for scalable image recognition. **CoRR**, abs/1707.07012, 2017. Available at: <<http://arxiv.org/abs/1707.07012>>.