

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

VITÓRIA MARIA SILVA BISPO

Orientadora: Prof^ª. Dra. Andrea Gomes Campos Bianchi

Coorientadora: Natália Fernanda de Castro Meira

**GERAÇÃO DE IMAGENS CITOLÓGICAS SINTÉTICAS ATRAVÉS DE
REDES GENERATIVAS ADVERSÁRIAS**

Ouro Preto, MG
2024

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

VITÓRIA MARIA SILVA BISPO

**GERAÇÃO DE IMAGENS CITOLÓGICAS SINTÉTICAS ATRAVÉS DE REDES
GENERATIVAS ADVERSÁRIAS**

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

Orientadora: Prof^a. Dra. Andrea Gomes Campos Bianchi

Coorientadora: Natália Fernanda de Castro Meira

Ouro Preto, MG
2024



FOLHA DE APROVAÇÃO

Vitória Maria Silva Bispo

Geração de Imagens Citológicas Sintéticas Através de Redes Generativas Adversárias

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

Aprovada em 15 de Janeiro de 2024.

Membros da banca

Andrea Gomes Campos Bianchi (Orientadora) - Doutora - Universidade Federal de Ouro Preto
Natália Fernanda de Castro Meira (Coorientadora) - Mestre - Universidade Federal de Ouro Preto
Pedro Henrique Lopes Silva (Examinador) - Doutor - Universidade Federal de Ouro Preto
Mateus Coelho Silva (Examinador) - Doutor - Universidade Federal de Ouro Preto

Andrea Gomes Campos Bianchi, Orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 15/01/2024.



Documento assinado eletronicamente por **Andrea Gomes Campos Bianchi, PROFESSOR DE MAGISTERIO SUPERIOR**, em 17/01/2024, às 18:38, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0654018** e o código CRC **BEBEC039**.

Dedico este trabalho à minha mãe, que fez inúmeros sacrifícios para me proporcionar a oportunidade de alcançar tudo que almejei.

Agradecimentos

Primeiramente, gostaria de expressar minha profunda gratidão à minha família, por serem meu alicerce ao longo desta jornada. Nada disso seria possível sem o apoio de vocês.

Além disso, sou repleta de gratidão pelos meus amigos, pela constante troca de experiência e companheirismo ao longo de todos esses anos.

Agradeço também à professora Andrea e à Natália pelo suporte e pela orientação proporcionados durante a realização do presente trabalho. O apoio e a confiança que recebi de vocês foram essenciais para a concretização desta pesquisa.

Finalmente, agradeço à UFOP e aos professores por concederem esta valiosa oportunidade de aprendizado. Além disso, sou imensamente grata pela oportunidade de envolvimento em projetos, os quais desempenharam um papel fundamental no meu crescimento tanto pessoal, quanto profissional.

Agradecimentos ao CNPq, Capes e à Fapemig.

Resumo

O câncer de colo de útero é o quarto mais frequente nas mulheres brasileiras. Por ser uma condição sem sintomas aparentes e de evolução gradual, muitas vezes passa despercebida. No entanto, apresenta um significativo potencial de recuperação quando tratada precocemente. Um poderoso aliado em prol desse objetivo é o diagnóstico assistido por computador, que utiliza modelos computacionais para detecção e classificação das informações das células por meio da abstração de padrões. Para assegurar a exatidão da classificação realizada, é necessário treinar o modelo com uma grande quantidade de dados. Nesse contexto, este trabalho tem como objetivo ampliar tanto a quantidade quanto a diversidade das amostras de *datasets* citológicos, incorporando imagens artificiais geradas por Redes Generativas Adversárias Convolucionais Profundas (*Deep Convolutional Generative Adversarial Networks* ou DCGANs). Para isso, dois modelos foram construídos: uma DCGAN padrão e uma pré-treinada a partir do ponto de partida da primeira. Por fim, o gerador foi exportado para que fosse possível gerar a quantidade de amostras sintéticas desejada. Adicionalmente, foi proposta a utilização de uma Rede Neural Convolutiva (CNN), para investigar os impactos que o aumento de dados possui em uma rede classificadora. Os resultados apresentaram acurácia acima de 70% para o classificador, indicando que as imagens geradas possuem boa qualidade e podem representar células cervicais. No entanto, uma análise mais aprofundada com diversas redes e estratégias é essencial para obter uma compreensão abrangente e confiável das imagens geradas.

Palavras-chave: Aprendizado de Máquina. Modelos Generativos. Redes Generativas Adversárias Convolucionais Profundas. Aumento de Conjuntos de Dados. Câncer de Colo de Útero.

Abstract

Cervical cancer is the fourth most common cancer in Brazilian women. It often goes unnoticed because it is a condition without apparent symptoms and of gradual evolution. However, it has a significant recovery potential when treated early. Computer-aided diagnosis is a powerful ally in favor of this objective, which uses computational models for the detection and classification of cell information through the abstraction of patterns. To ensure classification accuracy, training the model with a large amount of data is necessary. This work aims to increase the quantity and diversity of samples from cytological datasets, incorporating artificial images generated by Deep Convolutional Generative Adversarial Networks (DCGANs). Two models were developed: a standard DCGAN and one pre-trained from the starting point of the first. The second model generator was exported so that it was possible to generate the desired number of synthetic samples. Additionally, the use of a Convolutional Neural Network (CNN) was proposed to investigate the impacts that increased data has on a classifier network. The results showed an accuracy above 70% for the classifier, indicating that the generated images have good quality and can represent cervical cells. However, further analysis with various networks and strategies is essential to obtain a comprehensive and reliable understanding of the generated images.

Keywords: Machine Learning. Generative Models. Deep Convolutional Generative Adversarial Networks. Data Augmentation. Cervical Cancer.

Lista de Ilustrações

Figura 2.1 – Modelo de um neurônio artificial.	5
Figura 2.2 – Diagrama do funcionamento das GANs.	8
Figura 2.3 – Categorias dos principais métodos de avaliação das GANs.	9
Figura 2.4 – Estrutura de um <i>Variational Autoencoder</i>	13
Figura 2.5 – Diagrama do funcionamento e organização da GAN utilizada no trabalho. . .	14
Figura 2.6 – Imagens geradas pela GAN.	15
Figura 2.7 – Diagrama do funcionamento das RCGANs.	16
Figura 3.1 – Fluxograma que descreve a metodologia aplicada na pesquisa.	21
Figura 3.2 – Exemplos de amostras do <i>dataset</i> SIPaKMeD.	22
Figura 3.3 – Exemplos de amostras do <i>dataset</i> CRIC Cervix.	23
Figura 3.4 – Células recortadas das imagens do <i>CRIC Database</i>	25
Figura 3.5 – Arquitetura das DCGANs implementadas.	26
Figura 4.1 – Células sintéticas geradas pela DCGAN padrão ao final das 250 épocas. . .	29
Figura 4.2 – Amostras de células sintéticas com lesão de alto grau.	30
Figura 4.3 – Amostras de células sintéticas com lesão de baixo grau.	31
Figura 4.4 – Amostras de células sintéticas não lesionadas.	31

Lista de Tabelas

Tabela 2.1 – Métricas da performance do classificador utilizando dados gerados pela GAN.	15
Tabela 2.2 – Métricas da performance do classificador sem e com os dados gerados pela RCGAN.	16
Tabela 2.3 – Métricas da performance do classificador VGG16 para cada estratégia.	17
Tabela 2.4 – Métricas da performance ELM para os dados reais e dados gerados pela DCGAN.	18
Tabela 2.5 – Métricas da performance do classificador para os dados gerados pela GAN.	19
Tabela 2.6 – Relação das estratégias abordadas e <i>datasets</i> utilizados nos experimentos.	20
Tabela 2.7 – Arquiteturas de GANs e CNNs utilizadas nos Trabalhos Relacionados.	20
Tabela 3.1 – Contagem de tipo de células contidas no SIPaKMeD.	22
Tabela 3.2 – Contagem de tipo de células contidas no CRIC <i>database</i> .	24
Tabela 3.3 – Parâmetros utilizados na implementação da DCGAN.	26
Tabela 4.1 – Resultados da CNN treinada com diferentes quantidades de imagens sintéticas.	32

Lista de Abreviaturas e Siglas

2D	<i>Two Dimensional</i>
3D	<i>Three dimensional</i>
ABNT	Associação Brasileira de Normas Técnicas
ANN	<i>Artificial Neural Network</i>
ASC-H	<i>Atypical Squamous Cells, Cannot Exclude a High-Grade Lesion</i>
ASC-US	<i>Atypical Squamous Cells of Undetermined Significance, Possibly Non-Neoplastic</i>
AUC	<i>Area Under the Curve</i>
BCE	<i>Binary Cross Entropy</i>
CCD	<i>Charge-Coupled Device</i>
CGAN	<i>Conditional Generative Adversarial Network</i>
CNN	<i>Convolutional Neural Network</i>
CRIC	<i>Center for Recognition and Inspection of Cells</i>
DCGAN	<i>Deep Convolutional Generative Adversarial Network</i>
DECOM	Departamento de Computação
DOI	<i>Digital Object Identifier</i>
ELM	<i>Extreme Learning Machine</i>
FC	<i>Fully Connected</i>
FID	<i>Fréchet Inception Distance</i>
GAN	<i>Generative Adversarial Network</i>
HSIL	<i>High-Grade Squamous Intraepithelial Lesion</i>
IS	<i>Inception Score</i>
KID	<i>Kernel Inception Distance</i>
LSIL	<i>Low-Grade Squamous Intraepithelial Lesion</i>

MAGAN	<i>Manifold-Aligning Generative Adversarial Network</i>
MNIST	<i>Modified National Institute of Standards and Technology Database</i>
MS	<i>Mode Score</i>
MS-SSIM	<i>Multi-Scale Structural Similarity for Image Quality</i>
NILM	<i>Negative for Intraepithelial Lesion or Malignancy</i>
RCGAN	<i>Residual Condition Generative Adversarial Network</i>
SCC	<i>Squamous Cell Carcinoma</i>
UFOP	Universidade Federal de Ouro Preto

Lista de Símbolos

μ Micro

φ Phi

Sumário

1	Introdução	1
1.1	Justificativa	2
1.2	Objetivos	3
1.2.1	Objetivo Geral	3
1.2.2	Objetivos Específicos	3
1.3	Organização do Trabalho	3
2	Revisão Bibliográfica	4
2.1	Fundamentação Teórica	4
2.1.1	<i>Machine Learning</i>	4
2.1.2	Redes Neurais Artificiais	4
2.1.2.1	Neurônios Artificiais	5
2.1.2.2	Aprendizado das Redes Neurais Artificiais	5
2.1.3	Redes Neurais Convolucionais	6
2.1.3.1	Arquitetura das CNNs	6
2.1.3.2	Métricas de Avaliação	6
2.1.4	Redes Generativas Adversárias	7
2.1.4.1	Métricas de Avaliação	9
2.1.5	<i>Data Augmentation</i>	11
2.1.5.1	Transformações Simples	12
2.1.5.2	Modelos Generativos	13
2.2	Trabalhos Relacionados	13
2.2.1	<i>Generative adversarial network-based data augmentation to improve cervical cell classification model</i>	14
2.2.2	<i>Cervical Cancer Single Cell Image Data Augmentation Using Residual Condition Generative Adversarial Networks</i>	15
2.2.3	<i>Cell image augmentation for classification task using GANs on Pap smear dataset</i>	17
2.2.4	<i>DCGAN-based Medical Image Augmentation to Improve ELM Classification Performance</i>	18
2.2.5	<i>Generative Adversarial Networks for Augmenting Training Data of Microscopic Cell Images</i>	18
2.2.6	Análise das Técnicas Utilizadas	19
3	Desenvolvimento	21
3.1	Caracterização dos <i>datasets</i>	21
3.1.1	<i>SIPaKMeD</i>	21
3.1.2	<i>CRIC Cervix</i>	22

3.2	Preparação da base de dados	24
3.3	Geração de imagens	25
3.3.1	Arquitetura do modelo DCGAN	25
3.3.2	Treinamento dos Modelos	27
3.4	Classificação de Imagens	27
3.4.1	Arquitetura CNN	27
4	Resultados	28
4.1	Ambiente de Desenvolvimento	28
4.2	Geração de imagens	28
4.2.1	DCGAN Padrão	28
4.2.2	DCGAN Pré-Treinada	29
4.3	Classificação de Imagens	32
4.3.1	Estratégias de Treinamento	32
4.4	Disponibilização do código implementado	33
5	Considerações Finais	34
5.1	Conclusão	34
5.2	Trabalhos Futuros	34
	Referências	36

1 Introdução

O Instituto Nacional do Câncer (2022) estima que no triênio de 2023 a 2025, o câncer cervical será o quarto mais frequente em mulheres no Brasil. Apesar disso, ele possui potencial de cura quando detectado previamente e tratado de forma adequada. A incidência se dá pelo fato da doença possuir evolução lenta e não possuir sintomas perceptíveis em seu estado inicial (Instituto Nacional do Câncer, 2002).

Entre os métodos para diagnóstico precoce do câncer cervical, o mais utilizado é o exame citopatológico ou Papanicolau, que além de ser uma abordagem de baixo custo, segura e indolor, é também uma técnica eficaz para o rastreamento da doença. O exame de Papanicolau possibilita a detecção de lesões citológicas causadoras do câncer cervical através da análise das imagens obtidas das células coletadas durante o exame (Instituto Nacional do Câncer, 2002).

O diagnóstico é realizado pelos citopatologistas ou citotécnicos e, mesmo usando procedimentos rigorosos, é suscetível a erros devido às limitações humanas e a subjetividade da interpretação das imagens. Esses erros podem ser causados devido às variações das características citomorfológicas, que torna difícil classificá-las com precisão, assim como a extensa carga de trabalho desses profissionais (REZENDE; GOMES; CARNEIRO, 2021).

O diagnóstico assistido por computador tem se tornado uma técnica importante para o apoio ao profissional da saúde (LANGE; FERRIS, 2005). No caso de imagens celulares, são propostos modelos computacionais para detecção e classificação das informações das células através da extração de padrões de dados, com o objetivo de prever se a imagem pertence à classe afetada pela doença ou não.

Modelos computacionais baseados em aprendizado profundo muitas vezes necessitam de grande quantidade de dados para essa tarefa (YU et al., 2021). Por isso, um problema frequentemente enfrentado por métodos de aprendizado de máquina é o desbalanceamento de dados, que ocorre quando uma ou mais classes possuem quantidade de exemplos significativamente menor do que outras classes. Como consequência, o modelo fica enviesado a classificar os dados sempre como a classe majoritária, visto que no processo de extração de padrões o algoritmo pode dar maior peso as características dos grupos mais frequentes (YU et al., 2021).

Uma abordagem eficiente para lidar com o desbalanceamento de um conjunto de dados é a aplicação de modelos generativos para a geração de dados realistas. Dentre as abordagens existentes, a Rede Generativa Adversária (*Generative Adversarial Networks* ou GAN) se trata de uma técnica de aprendizado de máquina com duas redes neurais artificiais: o gerador e o discriminador, que estabelecem o jogo característico das GANs (GOODFELLOW et al., 2014). O gerador possui a função de criar novas amostras realistas e o discriminador, verifica se elas pertencem ou não ao conjunto de dados de imagens reais. Essa configuração é importante para

o aprendizado do modelo, visto que ambos os componentes são estimulados a reduzir os erros (GOODFELLOW et al., 2014).

Finalmente, a geração de novas amostras aumentará a diversidade de bancos de dados citológicos a fim de balancear os tipos celulares. Dessa forma, os modelos de classificação tendem a abstrair informações de todas as classes igualmente, de forma que, ao final, eles aprenderão a classificar todos os grupos corretamente.

1.1 Justificativa

O câncer cervical possui uma alta taxa de mortalidade no Brasil, visto que nem sempre é detectado a tempo de ser tratado. Para esse propósito, é essencial que os exames de Papanicolau sejam realizados de forma regular (Instituto Nacional do Câncer, 2002).

Embora existam exames para detecção do câncer de colo de útero, modelos classificadores baseados em aprendizado de máquina podem ser aliados dos profissionais na etapa de análise das amostras (REZENDE; GOMES; CARNEIRO, 2021). Apesar da eficiência dessa abordagem, é necessário fornecer uma grande quantidade de dados de alta qualidade para que o modelo abstraia informações de todas as classes (REZENDE; GOMES; CARNEIRO, 2021). Essa demanda se dá pela variabilidade das células, de forma que seja necessário fornecer um conjunto de dados diverso para que as particularidades sejam aprendidas e bem representadas.

Segundo Rezende, Gomes e Carneiro (2021), a variedade desses conjuntos de dados é limitada, de forma que existem mais amostras de células saudáveis do que anormais. Isso ocorre devido à maioria dos casos examinados resultarem em células saudáveis. Assim, um *dataset* com poucas amostras de células afetadas pelo câncer de colo do útero pode resultar em um modelo de classificação enviesado.

Por isso, surge a necessidade de aplicações de técnicas que aumentem a variedade de amostras da classe minoritária em um conjunto de dados. A abordagem proposta no presente trabalho visa solucionar esse problema por meio da geração de imagens artificiais realistas através das GANs, que são frequentemente utilizadas para balancear conjuntos de dados a fim de prevenir *overfitting* e tratar a sub-representação de dados.

O trabalho possui relevância para o aumento da eficiência do diagnóstico de doenças citológicas. Isso porque ele visa aumentar a diversidade e a quantidade de amostras nos *datasets* que são frequentemente utilizados em modelos classificadores que oferecem suporte nesse processo.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral deste trabalho consiste em implementar uma Rede Generativa Adversária capaz de gerar imagens sintéticas de citologia cervical, que possam ser utilizadas para aumentar conjuntos de dados, frequentemente utilizados para treinar métodos de diagnóstico assistido por computador.

1.2.2 Objetivos Específicos

- Explorar o estado atual das técnicas de geração de imagens citológicas utilizando Redes Generativas Adversárias (GANs);
- Elaborar conjuntos de dados para treinamento de GANs a partir de amostras das bases de dados SiPaKMeD e *Center for Recognition and Inspection of Cells* (CRIC);
- Implementar uma DCGAN para gerar imagens citológicas realistas sintéticas com e sem lesões;
- Aumentar o conjunto de dados de treinamento de um classificador utilizando os dados sintéticos gerados pela DCGAN;
- Comparar os resultados dos algoritmos classificadores antes e depois do aumento da base de dados;
- Aprimorar as métricas do algoritmo classificador escolhido através da ampliação do *dataset*.

1.3 Organização do Trabalho

O presente trabalho está organizado em cinco capítulos. No Capítulo 2 é realizada a Revisão Bibliográfica, que é dividida em duas seções. A primeira delas é a Fundamentação Teórica (Seção 2.1), onde são apresentados os conceitos básicos para o entendimento do trabalho e Trabalhos Relacionados (Seção 2.2), que descreve o estado da arte no que diz respeito à geração de imagens citológicas através de Redes Generativas Adversárias. Em seguida, o Capítulo 3 apresenta a metodologia aplicada durante o desenvolvimento da pesquisa.

O Capítulo 4 apresenta os resultados obtidos a partir da aplicação dos classificadores em um *dataset* de células e reais e em um conjunto de dados aumentado. Por fim, o Capítulo 5 apresenta as considerações finais do trabalho, incluindo os desafios enfrentados, limitações e possíveis trabalhos futuros.

2 Revisão Bibliográfica

Este capítulo tem a finalidade de oferecer uma visão contextualizada dos principais temas abordados neste Trabalho de Conclusão de Curso. Portanto, na Seção 2.1 são apresentados os conceitos básicos para a compreensão do trabalho e na Seção 2.2, as principais pesquisas (estado da arte) no que diz respeito às Redes Generativas Adversárias.

2.1 Fundamentação Teórica

O propósito desta seção é oferecer embasamento teórico para compreensão deste trabalho. Para alcançar esse objetivo, serão abordados os conceitos e técnicas fundamentais que são relevantes para a compreensão e produção de imagens por meio das GANs.

2.1.1 *Machine Learning*

O Aprendizado de Máquina (*Machine Learning* ou ML) trata da aplicação de técnicas capazes de aprender padrões, cujos modelos são construídos para realizar previsões. O aprendizado é realizado através do fornecimento de dados ou amostras (ZHOU, 2021).

Durante a construção do modelo de representação dos dados, as bases de dados são divididas em treino (*training*) e teste (*testing*). O primeiro grupo de amostras é empregado na etapa de extração de informações, aprendizado de padrões e a construção do modelo, enquanto o segundo é utilizado para simular a aplicação do algoritmo, o que torna possível avaliar o desempenho da técnica desenvolvida em dados que não foram usados para o treinamento (ZHOU, 2021).

O aprendizado pode se dar de diferentes formas, a primeira delas é a supervisionada, onde cada exemplo de dado de entrada é combinado com seu rótulo (exemplo: tarefas de classificação e regressão). Outra abordagem é o aprendizado não-supervisionado, onde o treinamento não associa a entrada com um resultado específico (exemplo: clusterização) e por último, o semi-supervisionado, onde parte das amostras possuem rótulos e outra parte não (exemplo: sistemas de recuperação de texto/imagens) (NAQA; MURPHY, 2015).

2.1.2 **Redes Neurais Artificiais**

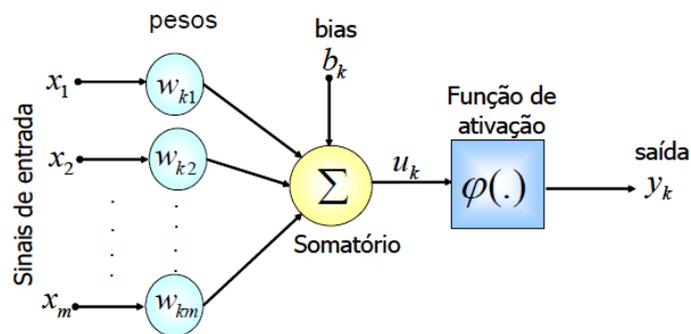
A Rede Neural Artificial, ou *Artificial Neural Network* (ANN) é um método de ML inspirado na estrutura do cérebro humano. A ANN é composta por diversas unidades de processamento simples, denominados neurônios interligados em uma ou mais camadas a fim de resolver problemas através do cálculo de funções matemáticas lineares e não lineares. A abstração do

conhecimento se dá através da associação à pesos, que armazenam o que foi aprendido pelo modelo (BRAGA; CARVALHO; LUDERMIR, 2000).

2.1.2.1 Neurônios Artificiais

As ANNs são compostas por uma rede de neurônios conectados entre si, que recebem sinais uns dos outros (HAYKIN, 2000). A Figura 2.1 apresenta um modelo de um neurônio artificial. Ao analisá-la, é possível identificar três elementos: um conjunto de pesos (x_i), que é conectada com um outro neurônio e cada uma possui um peso (w_{ki}). Um somatório, com a finalidade de somar os sinais de entrada e uma função de ativação ($\varphi(x)$), que tem a finalidade de limitar a amplitude da saída (y_k) de um neurônio (HAYKIN, 2000).

Figura 2.1 – Modelo de um neurônio artificial.



Fonte: Soares e Silva (2011)

O funcionamento do neurônio se dá da seguinte forma: o conjunto de sinapses recebe os estímulos, ou seja, os sinais de entrada, que são multiplicados pelos pesos sinápticos. Para produzir um nível de atividade, é realizada a soma ponderada dos sinais, que deve exceder determinado limite definido pela função de ativação para que a unidade processadora produza uma saída (FURTADO, 2019).

2.1.2.2 Aprendizado das Redes Neurais Artificiais

Por se tratar de um modelo de ML, o processo de aprendizado das ANNs também pode ser supervisionado ou não (RAUBER, 2005). O primeiro deles acontece quando o conjunto de treino contém as amostras e seus rótulos, ou seja, os resultados esperados. Portanto, a tarefa do modelo supervisionado é tentar relacionar os dados com os rótulos corretos. Em relação ao segundo tipo, apenas as amostras estão disponíveis e o modelo tem como objetivo perceber os padrões entre os exemplos do conjunto de treino através da descoberta dos atributos relevantes para o aprendizado do modelo (RAUBER, 2005).

Algoritmos de ML aprendem através da extração de padrões dos dados fornecidos e da atualização nos pesos da rede, de forma que os valores da saída da ANN se aproximam do valor desejado (RAUBER, 2005). Uma das principais formas de garantir que o modelo aprenda de maneira satisfatória é o treinamento em uma base de dados diversa e com uma grande quantidade

de amostras. Portanto, existem diversos problemas relacionados ao *dataset* que devem ser levados em consideração ao implementar um modelo de ML (RAUBER, 2005).

2.1.3 Redes Neurais Convolucionais

A Rede Neural Convolucional (*Convolutional Neural Network*, ou CNN) se trata de um modelo de redes neurais no campo de aprendizado profundo (LI et al., 2022). Semelhante às ANNs, as CNNs são compostas por um conjunto de neurônios conectados entre si, capazes de aprender por meio da abstração de informações e auto-otimização (O'SHEA; NASH, 2015).

Amplamente utilizadas em aplicações que envolvem reconhecimento de padrões, processamento de imagens e até mesmo reconhecimento de voz, as CNNs demonstram excelentes resultados na resolução de tarefas complexas (O'SHEA; NASH, 2015; ALBAWI; MOHAMMED; AL-ZAWI, 2017).

2.1.3.1 Arquitetura das CNNs

Além da aplicação em problemas complexos, uma das principais características de uma CNN é a sua composição, uma vez que possui três tipos de camadas: *convolutional layers*, *pooling layers* e *fully-connected layers* (O'SHEA; NASH, 2015).

As camadas convolucionais (*convolutional layers*) estão diretamente relacionadas com a forma que a CNN abstrai as informações e, conseqüentemente, aprende (O'SHEA; NASH, 2015). A camada convolucional é composta por diversos *kernels* de convolução, que têm a função de calcular diferentes mapas de características, de modo que seja possível aprender as representações das características de entrada (LI et al., 2022).

Por sua vez, as camadas de *pooling* diminuem a complexidade computacional do modelo, bem como a quantidade de parâmetros através da redução do número de conexões entre as camadas convolucionais (GU et al., 2018). Por fim, as camadas totalmente conectadas (*fully-connected*) desempenham as mesmas funções que realizariam em uma ANN. Dessa forma, cada nó está conectado a todos os nós das camadas imediatamente anteriores e subseqüentes (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

2.1.3.2 Métricas de Avaliação

As métricas de avaliação representam ferramentas fundamentais em aplicações de aprendizado de máquina, proporcionando uma análise do desempenho do modelo implementado de forma quantitativa (DALIANIS, 2018). Essas métricas utilizam cálculos para mensurar a eficácia do modelo durante o processo de execução, desempenhando um papel crucial na compreensão do comportamento do modelo e na identificação de potenciais melhorias.

Acurácia

A acurácia se trata da porcentagem de classificações corretas, tanto positivas quanto negativas em relação ao total de previsões que foram realizadas pelo modelo. Um alto valor de acurácia indica que o sistema está classificando a maioria das instâncias de maneira correta (DALIANIS, 2018).

Precisão

A precisão representa a porcentagem de previsões corretas feitas para a classe de interesse em relação ao número total de previsões dessa classe. Em outras palavras, essa métrica indica a proporção de instâncias identificadas como pertencentes a uma classe que realmente fazem parte dessa classe (DALIANIS, 2018).

Recall

A métrica de avaliação *recall* representa a porcentagem de instâncias da classe de interesse que o modelo identificou corretamente. Essa métrica quantifica a proporção de instâncias pertencentes a essa classe que foram corretamente classificadas em relação ao total de instâncias reais pertencentes à classe de interesse (DALIANIS, 2018).

F1-Score

O F1-Score é uma média ponderada entre a precisão e o *recall*, representando a média harmônica dessas duas métricas (DALIANIS, 2018).

2.1.4 Redes Generativas Adversárias

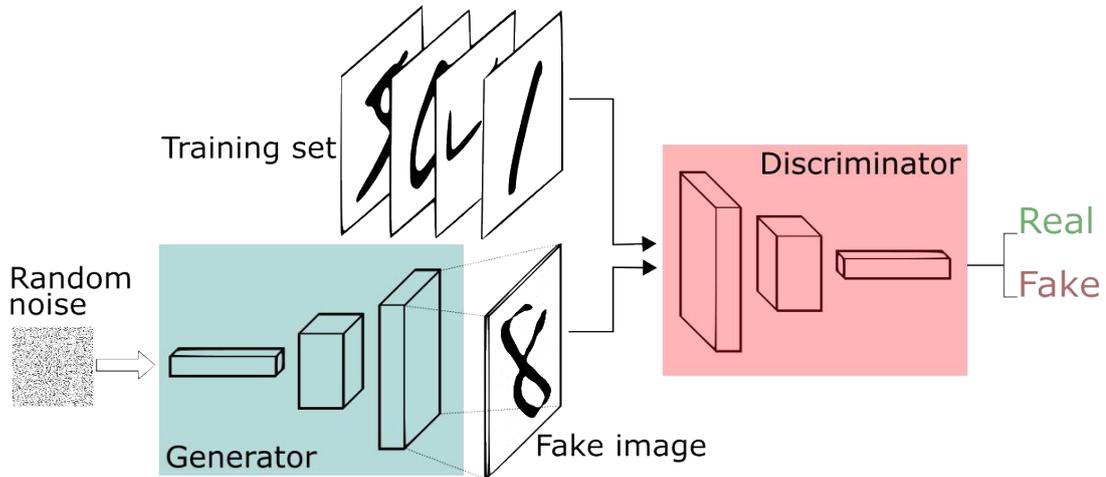
Em 2014, Goodfellow et al. (2014) propuseram a GAN, que consiste em um modelo não-supervisionado que se baseia em *game theory*, ou seja, dois jogadores adversários são criados e ambos possuem o mesmo objetivo: derrotar o outro participante.

Esse jogo se dá entre o *gerador* e o *discriminador*, dois modelos de aprendizado de máquina que utilizam redes neurais. O primeiro deles, o gerador, é responsável por aprender a gerar amostras falsas através de ruídos; e o segundo, o discriminador, examina as amostras produzidas e retorna uma estimativa de quão real ou falsa ela é (GOODFELLOW et al., 2020).

Conforme apresentado pelo diagrama da Figura 2.2, a produção das amostras tem início no gerador (*generator*), que recebe como entrada um vetor de ruído aleatório (*random noise*), que será processado pela rede neural do gerador e como saída, obtêm-se uma imagem falsa (*fake image*). A amostra resultante é utilizada como entrada no discriminador (*discriminator*), juntamente com amostras verdadeiras do conjunto de treinamento (*training set*). Como o discriminador se trata

de um modelo classificador, ele retorna a probabilidade da amostra produzida pelo gerador ser verdadeira (*real*) ou falsa (*fake*) (GOODFELLOW et al., 2020).

Figura 2.2 – Diagrama do funcionamento das GANs.



Fonte: Silva (2017)

A perda é computada utilizando a probabilidade estimada pela rede, e esse cálculo é realizado por meio de uma função de perda, como a *Binary Cross Entropy* (BCE), conforme definida na Equação (2.1). Essa função leva em conta o rótulo original da amostra (Y), que é 0 para amostras falsas e 1 para amostras reais, enquanto \hat{Y} representa o rótulo atribuído pelo discriminador (NIE et al., 2018).

$$BCE(\hat{Y}, Y) = - \sum_i Y^i \log(\hat{Y}^i) + (1 - Y^i) \log(1 - \hat{Y}^i) \quad (2.1)$$

O resultado da equação é utilizado como parâmetro para melhorar o desempenho do discriminador e do gerador durante a etapa de treinamento, visto que o sinal negativo no início da equação faz com que o valor resultante seja uma perda (LAN et al., 2020). Essa definição evidencia o problema de otimização característico de redes neurais e mais especificamente, o objetivo estabelecido pelas GANs, o *min-max*. Uma vez que o objetivo do gerador é fazer com que o resultado da Equação (2.1) se aproxime de 1 maximizando assim o resultado e o discriminador seja minimizado aproximando o valor de 0 (LAN et al., 2020).

Os treinamentos do gerador e do discriminador acontecem de forma que ambos evoluem gradativamente e para isso é utilizado um otimizador gradiente (GOODFELLOW et al., 2020). O treinamento simultâneo possui grande importância para garantir o bom funcionamento da GAN, visto que treinar mais o discriminador do que o gerador sempre vai acusar a amostra como falsa e um gerador que foi mais treinado que o discriminador sempre produzirá amostras que serão classificadas como verdadeiras.

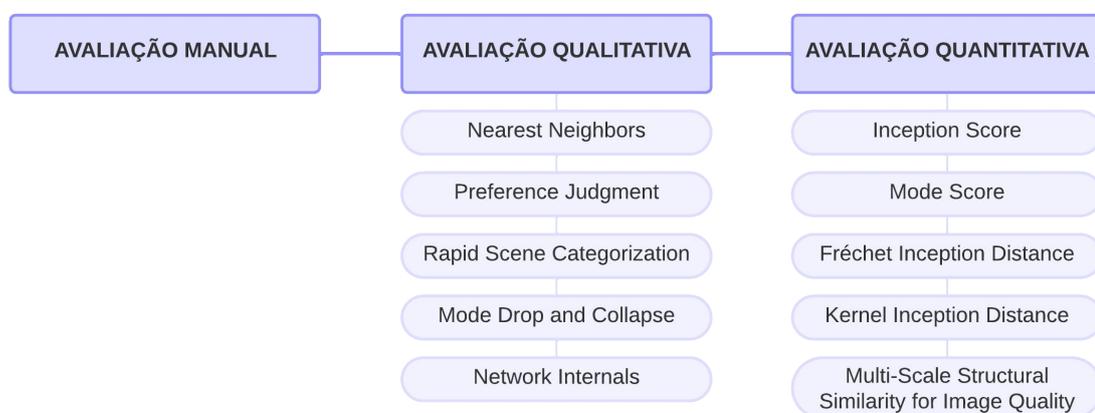
Ao final do processo de treinamento da GAN, elas são capazes de produzir amostras realistas mesmo para os mais complexos *datasets* contendo imagens de alta resolução (GOODFELLOW

et al., 2020).

2.1.4.1 Métricas de Avaliação

Dada a popularidade dos modelos generativos, diversas métricas de avaliação foram propostas para que seja possível mensurar seu desempenho (BETZALEL et al., 2022). Entretanto, esse ainda é um problema aberto, uma vez que avaliações, subjetivas ou não, são difíceis de serem mensuradas. A Figura 2.3 explicita os principais métodos de avaliação das GANs de acordo com Salehi, Chalechale e Taghizadeh (2020), que serão discutidos detalhadamente abaixo.

Figura 2.3 – Categorias dos principais métodos de avaliação das GANs de acordo com Salehi, Chalechale e Taghizadeh (2020).



Fonte: Produzida pela autora.

Avaliação Manual

A avaliação manual se trata de uma técnica intuitiva frequentemente utilizada para avaliar a qualidade e a diversidade das imagens geradas por uma GAN (SALEHI; CHALECHALE; TAGHIZADEH, 2020). A abordagem consiste em realizar comparações entre a imagem real e a artificial através da interpretação de pessoas avaliadoras. Uma vez que a avaliação manual depende da interpretação, ela se trata de uma técnica subjetiva e além disso, consome tempo e esforço humano. Outro ponto negativo dessa abordagem é a dificuldade de saber o momento de interromper o processo de treinamento da GAN, visto que não são utilizadas métricas estatísticas para avaliar a performance (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

Avaliação Qualitativa

Outra abordagem frequentemente utilizada é a avaliação qualitativa, que se trata da aplicação de métricas não numéricas com a finalidade de mensurar a performance do modelo através de avaliações comparativas ou subjetivas (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

A abordagem *Nearest Neighbors* consiste em utilizar a distância Euclidiana em cada pixel das imagens artificiais a fim de selecionar as imagens mais similares às imagens reais, medida de similaridade entre imagens. A partir dessa distância, é realizada uma comparação para compreender o quão real a amostra gerada é (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

Os métodos de avaliação qualitativa *Preference Judgment* e *Rapid Scene Categorization* são extensões da avaliação manual, visto que utilizam pessoas para determinar a qualidade das imagens geradas pelo modelo. A primeira abordagem consiste em utilizar avaliadores humanos, que recebem as imagens produzidas e as classificam em valores de precisão (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Por sua vez, a técnica *Rapid Scene Categorization* se baseia em mostrar diversas amostras para os avaliadores por um curto período de tempo para que eles possam classificá-las em reais e falsas (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Partindo da ideia que pode conter disparidades nas imagens contidas em cada um dos dois grupos, são chamados diversos avaliadores, a fim de diminuir a possível variância nos resultados (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

A abordagem *Mode Drop and Collapse* se dá pela avaliação de dois possíveis problemas de GANs. O primeiro deles é o *Mode Drop*, que ocorre quando o modelo gera saídas diferentes para entradas similares e o *Mode Collapse* ocorre quando a diversidade das amostras geradas é limitada (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

Por fim, Salehi, Chalechale e Taghizadeh (2020) apresenta a utilização de *Network Internals* para descobrir quantas características são consideradas na camada latente das redes. Essa abordagem é aplicada a fim de compreender a qualidade das representações internas.

Avaliação Quantitativa

A avaliação quantitativa se baseia no emprego de cálculos com a finalidade de comparar a qualidade das imagens geradas pelo modelo (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

O *Inception Score* (IS) se trata de uma métrica que avalia a qualidade e a diversidade das imagens geradas (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Para isso, o IS utiliza a rede *Inception V3* pré-treinada no conjunto de dados *ImageNet* (XU et al., 2018). A rede é utilizada para obter a distribuição condicional dos rótulos através da classificação das imagens, de forma que seja possível calcular a probabilidade do modelo atribuir diferentes rótulos para a imagem gerada (SALIMANS et al., 2016). Valores elevados de IS indicam que as amostras geradas pelo modelo possuem alta qualidade (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

Baseada no IS, a métrica *Mode Score* (MS) avalia a qualidade e diversidade das imagens geradas (SALEHI; CHALECHALE; TAGHIZADEH, 2020). O adicional do MS é a sensibilidade à distribuição dos rótulos dos dados reais, de forma que seja possível avaliar como as novas amostras se comparam às reais através da distribuição dos rótulos (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

A métrica *Fréchet Inception Distance* (FID) visa comparar o quão semelhante as imagens artificiais e reais são (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Para isso, o FID analisa como as duas amostras se comportam em um espaço de características da rede *Inception* e calcula uma média. Por fim, a métrica calcula a distância entre a média e a covariância entre as amostras a fim de avaliar a qualidade das imagens geradas (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Salehi, Chalechale e Taghizadeh (2020) destacam que a métrica *Kernel Inception Distance* (KID) foi proposta para resolver problemas de *overfitting*. De maneira similar à FID, essa métrica conduz comparações entre as distribuições de amostras sintéticas e reais para avaliar a distância entre elas.

A métrica *Multi-Scale Structural Similarity for Image Quality* (MS-SSIM) visa avaliar a similaridade entre duas imagens através de dados de luminância e contraste (WANG; SIMON-CELLI; BOVIK, 2003). Dessa forma, a MS-SSIM pode ser utilizada para determinar a diversidade dos dados gerados pela GAN (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

2.1.5 *Data Augmentation*

Durante o processo de treinamento de um modelo supervisionado, ambos os recursos são oferecidos para que ocorra o processo de aprendizado, mas a fim de avaliar a performance, a etapa de teste fornece apenas os atributos e ao final da execução, é realizada uma comparação dos rótulos resultantes com o resultado esperado (KOEHRSEN, 2018).

O desbalanceamento de dados é um problema regularmente enfrentado por modelos de ML, uma vez que determinadas classes são menos usuais do que outras (YU et al., 2021). Embora existam dados suficientes para o treinamento de algumas classes, o mesmo não ocorre para outras que são igualmente importantes (CLARO et al., 2020). Esse desequilíbrio dos dados faz com que o modelo abstraia mais padrões de uma classe do que de outra, que resulta em um sistema de ML de baixa precisão para o grupo minoritário (CLARO et al., 2020).

Underfitting e *overfitting* são outros dois problemas frequentemente enfrentados por modelos de ML. Ambos dizem respeito ao grau de flexibilidade do modelo ao representar um conjunto de dados (KOEHRSEN, 2018). O primeiro deles ocorre quando o modelo é muito simples e não é capaz de capturar a variabilidade dos dados (JABBAR; KHAN, 2015). Por sua vez, o *overfitting* ocorre quando o modelo perde a capacidade de generalização e apenas se ajusta aos dados de treinamento. Dessa forma, o modelo resultante não se adaptará para novos dados, uma vez que apenas aprendeu a regularidade existente nos dados utilizados no treinamento (JABBAR; KHAN, 2015).

Claro et al. (2020) mencionam que abordagens de *data augmentation*, ou aumento de dados, são frequentemente empregadas para amenizar ou evitar os problemas decorrentes do desbalanceamento de dados. Diversas técnicas que visam aumentar a quantidade de dados em um *dataset* foram desenvolvidas a fim de resolvê-los e algumas delas são detalhadas nas seções

subsequentes (CLARO et al., 2020).

2.1.5.1 Transformações Simples

Transformações simples são procedimentos de manipulação de imagens frequentemente empregados para aumentar o conjunto de dados de modelos de ML (CLARO et al., 2020). Dentre as possíveis transformações, as mais utilizadas são geométricas e fotométricas.

Abordagens de transformações geométricas envolvem a manipulação de atributos da imagem original e seu rótulo, por esse motivo, são caracterizados por serem de fácil implementação. Por sua vez, a transformação fotométrica se baseia em manipular os valores armazenados nos *pixels* das imagens (CLARO et al., 2020).

A seguir estão alguns exemplos de transformações que podem ser realizadas para proporcionar o aumento da diversidade de imagens na base de dados apresentadas no trabalho de Claro et al. (2020);

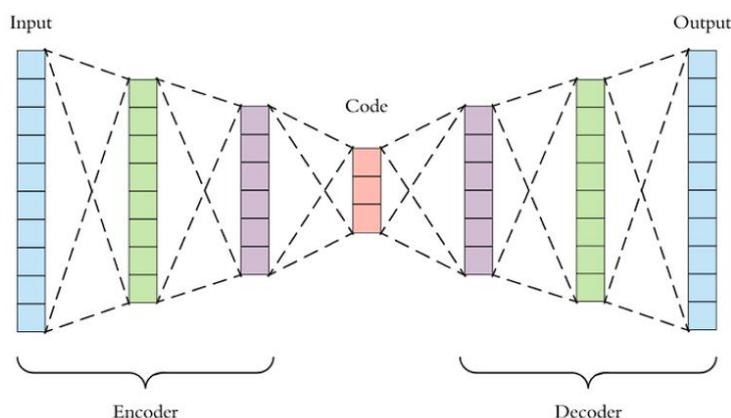
- **Inversão:** a técnica envolve rotacionar a imagem em 180° horizontal ou verticalmente;
- **Rotação:** a abordagem de rotação se baseia na rotação da imagem para a direita ou esquerda em um eixo entre 1° e 359°;
- **Translação:** a aplicação da translação envolve movimentar a imagem nos eixos X ou Y;
- **Recorte:** a abordagem se baseia em selecionar uma porção da imagem por meio do recorte;
- **Zoom:** a técnica se baseia em ampliar ou reduzir a escala da imagem;
- **Cisalhamento:** busca alongar a imagem através da fixação de um eixo e alongamento da imagem em determinado ângulo;
- **Adição de ruído:** caracterizada pela adição de valores aleatórios em alguns *pixels* da imagem original;
- **Transformação no espaço de cores:** a técnica busca realizar alterações nos valores do histograma de cores da imagem, de forma que ocorram alterações na iluminação e contraste da mesma;
- **Filtragem com *kernel*:** envolve deslizar uma janela sobre a imagem e em cada posição, aplicar operações como borramento e aguçamento;
- **Combinação de imagens:** abordagem que gera novas imagens através da combinação de outras imagens do *dataset*, e;
- **Random erasing:** consiste em ocultar determinadas áreas da imagem através da inserção de ruídos.

2.1.5.2 Modelos Generativos

Outro tipo de abordagem frequentemente utilizada para aumentar a quantidade e diversidade de dados em um *dataset* são os modelos generativos. Essa técnica consiste em gerar amostras artificiais similares ao conjunto de dados de imagens reais (CLARO et al., 2020).

As Redes Autocodificadoras Variacionais (*Variational Autoencoder*, ou VAE) são um tipo de modelo generativo. As VAEs são compostas por um *encoder* e um *decoder* (CLARO et al., 2020). A Figura 2.4 representa a estrutura dessa rede, que é composta por três camadas: entrada (*input*), codificação (*code*) e saída (*output*) (HARSHVARDHAN et al., 2020).

Figura 2.4 – Estrutura de um *Variational Autoencoder*.



Fonte: Claro et al. (2020)

Basicamente, o funcionamento da VAE se dá através da inserção de uma imagem na entrada do *encoder*, que a codificará a fim de produzir um vetor de baixa dimensionalidade. Esse vetor é armazenado na camada *code* e a função do *decoder* é reconstruir essa imagem através do vetor. Por se tratar de uma reconstrução, a nova amostra gerada possuirá novas características, ou seja, uma nova imagem será criada (CLARO et al., 2020).

As Redes Generativas Adversárias (*Generative Adversarial Networks*, ou GANs) referem-se a outro tipo de modelos generativos. A Seção 2.1.4 explorou com maior profundidade o funcionamento desse gênero de modelo generativo.

2.2 Trabalhos Relacionados

Esta seção visa fornecer o contexto atual em relação à geração de imagens citológicas através de GANs. Trabalhos relacionados serão apresentados e analisados a fim de demonstrar o estado da arte do tema abordado.

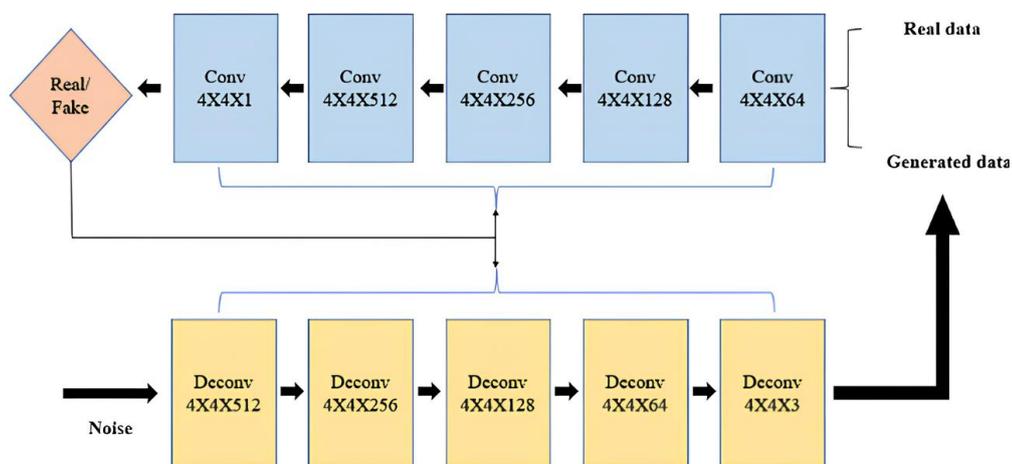
Os estudos apresentados foram selecionados de maneira a garantir que todos utilizassem métodos semelhantes para avaliar o resultados das imagens sintéticas geradas. Isso permite uma comparação direta entre os resultados deste trabalho e os dos autores mencionados.

2.2.1 Generative adversarial network-based data augmentation to improve cervical cell classification model

O artigo de pesquisa desenvolvido por Yu et al. (2021), visa resolver um problema frequentemente enfrentado em modelos de aprendizado de máquina: o desbalanceamento de *datasets*. O trabalho é realizado com o foco em proporcionar aumento de dados para melhorar modelos de classificação aplicados à células cervicais, visto que se trata de um processo demorado e que pode conter erros, faz-se necessário o desenvolvimentos de abordagens que tenham a finalidade de automatizá-lo.

Em relação à arquitetura da GAN desenvolvida, o trabalho mostra que a rede utilizada possui cinco camadas de convolução tanto no gerador, quanto no discriminador (Figura 2.5). Pela baixa quantidade de imagens disponíveis, os autores pré-treinaram a GAN com outro conjunto de dados para que ao iniciar o treinamento com as células, o modelo já fosse capaz de gerar estruturas básicas. O *dataset* utilizado no trabalho foi coletado no hospital da cidade de Baoding, na China e contém 22.124 imagens celulares, sendo elas 1.202 amostras com anomalias e 20.922 amostras normais.

Figura 2.5 – Diagrama do funcionamento e organização da GAN utilizada no trabalho.



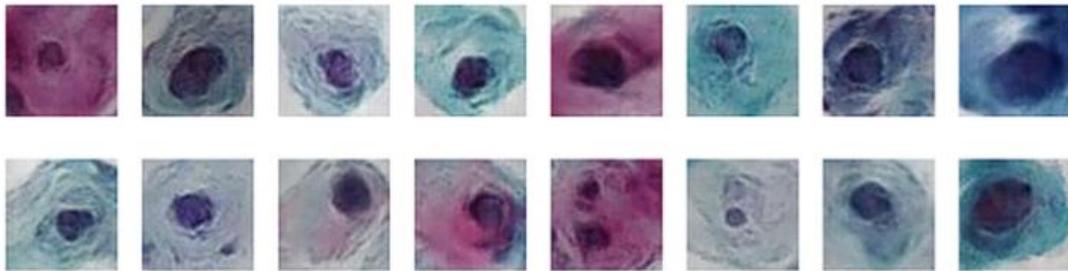
Fonte: Yu et al. (2021)

Os autores desenvolveram um modelo de classificação baseado na arquitetura Alexnet a fim de verificar a qualidade das amostras produzidas pela GAN, com a finalidade de fornecer rótulos para o treinamento do modelo durante a tarefa de classificação. Um conjunto de patologistas recortaram as células das imagens do *dataset* e as categorizaram em duas classes: amostras normais e com anomalias. O processo de treinamento realizado por Yu et al. (2021) envolveu quatro diferentes estratégias:

1. Treinamento contendo apenas dados reais;
2. Pré-treinamento com dados do ImageNet;

3. Treinamento com dados reais e imagens geradas pela GAN;
4. Pré-treinamento com os dados gerados pela GAN;

Figura 2.6 – Imagens geradas pela GAN.



Fonte: Yu et al. (2021)

Cada uma dessas estratégias se destacou em diferentes pontos, mas a terceira e a quarta obtiveram maior êxito, uma vez que utilizaram amostras artificiais no treinamento e pré-treinamento. A partir dessa constatação, Yu et al. (2021) observaram que as métricas de um problema com poucos dados podem ser melhoradas ao combinar os dados reais e artificiais durante o processo de aprendizado do modelo classificador.

Tabela 2.1 – Métricas da performance do classificador utilizando dados gerados pela GAN.

Task	Precision	Sensitivity	Specificity	Accuracy	F1-score	H-mean	AUC
1	28.9%	73.4%	89.0%	88.1%	41.5%	80.5%	0.859
2	38.3%	93.8%	90.8%	91.0%	54.4%	92.3%	0.975
3	54.5%	92.5%	95.3%	95.1%	68.6%	93.9%	0.982
4	47.8%	95.9%	93.6%	93.8%	63.8%	94.7%	0.984

Fonte: Yu et al. (2021)

Ao final do trabalho, Yu et al. (2021) chegaram à conclusão que treinar o modelo com os dados produzidos pela GAN aumentou as métricas estatísticas e além disso, uma possível melhoria seria pré-treinar o modelo classificador com os dados gerados pela GAN e realizar os ajustes do modelo com os dados reais. Como mostrado na Tabela 2.1, os valores de acurácia (*accuracy*) do classificador da terceira e da quarta estratégia foram respectivamente 95,3% e 93,8%. O que demonstra que o aumento de *datasets* através de GANs é uma forma eficiente para tratar o problema de desbalanceamento de dados.

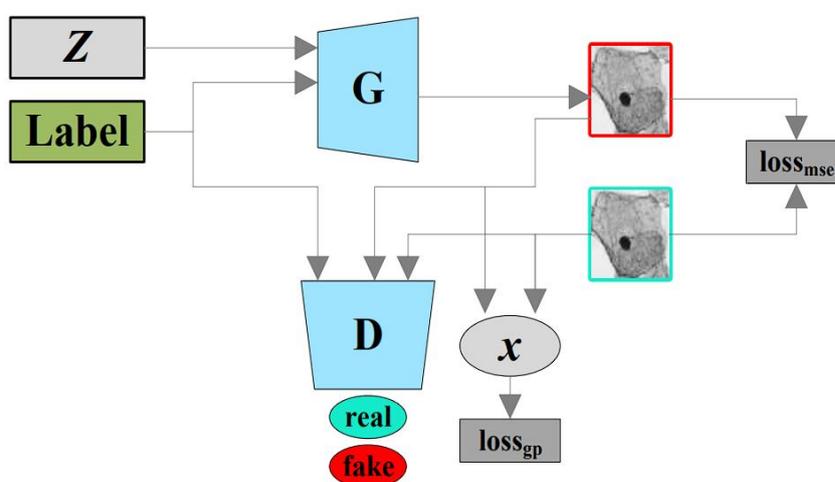
2.2.2 Cervical Cancer Single Cell Image Data Augmentation Using Residual Condition Generative Adversarial Networks

O trabalho desenvolvido por Chen et al. (2020) possui o objetivo de aumentar *datasets* que contém imagens de células de câncer cervical para melhorar modelos de classificação. Diante disso, os autores propõem um modelo generativo baseado em condição residual, a *Residual*

Condition Generative Adversarial Network (RCGAN), que se trata de uma arquitetura de Redes Neurais Convolucionais em que a entrada da camada anterior é somada à saída da atual, ou seja, utiliza-se os resíduos de uma camada na camada seguinte (SHAFIQ; GU, 2022).

No que diz respeito à arquitetura do modelo (Figura 2.7), no gerador (G) foram utilizados blocos residuais e 21 camadas convolucionais, que receberam como entrada ruídos aleatórios Gaussianos (Z) e rótulos da classe a ser gerada ($Label$), que resulta em uma nova imagem da classe desejada. Por sua vez, o discriminador (D) recebe como entrada a imagem gerada, a imagem real e o rótulo da amostra. Ao final, obtém-se a probabilidade da amostra gerada ser originada do conjunto de imagens reais.

Figura 2.7 – Diagrama do funcionamento das RCGANs.



Fonte: Chen et al. (2020)

O *dataset* utilizado no trabalho foi coletado pela *Danish Technical University* e pela *Herlev University* (JANTZEN et al., 2005a), os dados estavam divididos em duas classes: células normais, com 205 imagens e células com anomalias, com 585 imagens, totalizando assim 790 amostras. Os autores mencionam que para eliminar os erros que manchas poderiam causar, a imagem foi convertida para escala de cinza e recortadas de diferentes tamanhos para 72×72 pixels. Além disso, os valores de intensidade dos *pixels* foram normalizados para $(-1, 1)$.

Tabela 2.2 – Métricas da performance do classificador sem e com os dados gerados pela RCGAN.

Treinamento com amostras reais				Treinamento com amostras geradas			
Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
84.25%	71.57%	99.65%	83.31%	95.18%	96.10%	98.50%	97.28%

Fonte: Chen et al. (2020)

Em relação os experimentos realizados, Chen et al. (2020) utilizaram um classificador para medir a influência causada pelas amostras geradas via RCGAN. Os resultados das métricas estatísticas para o modelo treinado apenas com imagens reais e treinado com imagens do modelo proposto no trabalho é exibido na Tabela 2.2. Através da análise dos resultados, é possível perceber

que houve um aumento no desempenho do classificador ao utilizar o *dataset* aumentado, o que mostra que a abordagem proposta contribuiu para classificação correta das imagens de células.

2.2.3 *Cell image augmentation for classification task using GANs on Pap smear dataset*

O artigo desenvolvido por Zak et al. (2022) tem como objetivo realizar o aumento de dados através de GANs para que modelos de Redes Neurais Convolucionais possuam melhores métricas ao serem executados. O proposto é utilizado para aumentar a quantidade de imagens do *dataset* público, *HErlev Pap Smear* (JANTZEN et al., 2005b), que é frequentemente utilizado para treinar CNNs na classificação de sete classes. Além desse *dataset*, os autores também mencionam outros conjuntos de dados com imagens de citologia ou fragmentos de imagens: CRIC (REZENDE et al., 2021) e SIPakMeD (PLISSITI et al., 2018).

O método proposto no trabalho é dividido em duas etapas: geração da imagem artificial e mapa baseados na original e melhoria na qualidade da imagem. Essa arquitetura se baseia na *pix2pix network*, que foi proposta por Isola et al. (2016).

Em relação aos experimentos realizados, Zak et al. (2022) utilizaram *VGG16* (SIMONYAN; ZISSERMAN, 2015), *InceptionV3* (SZEGEDY et al., 2015), *ResNet50* (HE et al., 2015), *DenseNet121* (HUANG et al., 2018) para avaliar o desempenho das imagens geradas pelo método desenvolvido por eles através da classificação das amostras. Os autores desenvolveram três diferentes estratégias de treinamento:

1. Treinamento das CNNs apenas com imagens do dataset original;
2. Treinamento das CNNs com *dataset* aumentado;
3. Treinamento das CNNs com *dataset* aumentado melhorado.

Tabela 2.3 – Métricas estatísticas da performance do classificador VGG16 para cada estratégia.

Strategies	F-score (%)	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)
Strategy 1	51.7 ± 20.3	82.8 ± 7.0	52.4 ± 26.3	59.5 ± 24.6	90.5 ± 7.1
Strategy 2	86.6 ± 11.8	90.9 ± 11.8	86.7 ± 15.5	89.1 ± 13.0	94.6 ± 5.8
Strategy 3	88.9 ± 5.3	93.5 ± 6.0	88.6 ± 7.7	90.6 ± 10.8	94.5 ± 7.5

Fonte: Zak et al. (2022)

O autor organizou os dados estatísticos da execução das CNNs para as três estratégias e ao analisá-los, foi possível perceber que a estratégia 3 obteve melhor resultado. A Tabela 2.3 exibe os dados da execução da estratégia com melhor desempenho. A partir da análise dos resultados, é possível perceber que a utilização de *pix2pix* para aumento de *datasets* é uma abordagem eficiente, visto que houve ganho significativo nas métricas do classificador.

2.2.4 DCGAN-based Medical Image Augmentation to Improve ELM Classification Performance

O artigo desenvolvido por [Rando et al. \(2022\)](#) propõe um método de aumento de *datasets* de imagens celulares baseado na arquitetura DCGAN. Tendo em vista que a tecnologia possui o potencial de ser grande aliada dos profissionais para promoção do diagnóstico prévio de câncer, se faz necessário aumentar a eficiência desses métodos, garantindo assim, a qualidade e rapidez da detecção. Um grande problema enfrentado é a limitação de imagens médicas para apoiar esses métodos e esse trabalho visa amenizá-lo.

Inicialmente, os autores apresentam a DCGAN, que foi utilizada para desenvolver a proposta. Ela foi escolhida por resolver um problema comum da GAN original: a instabilidade no treinamento. Para fins de avaliação das amostras geradas, foi utilizado o *Extreme Learning Machine* (ELM), método frequentemente utilizado para problemas de classificação. O *dataset* utilizado foi coletado pelo dr. Soetomo, do Hospital Surabaya, da Indonésia e é composto por 538 imagens de três diferentes classes: *Adenocarcinoma*, *HSIL* e *Squamous Cell Carcinoma*.

Tabela 2.4 – Métricas da performance ELM para os dados reais e dados gerados pela DCGAN.

Classe	Real Data			Real data with Synthetic Images		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Adenocarcinoma	67%	27%	39%	94%	90%	92%
HSIL	93%	93%	93%	95%	88%	91%
Squamous Cell Carcinoma	46%	74%	57%	81%	95%	87%

Fonte: [Rando et al. \(2022\)](#)

No que diz respeito ao treinamento da DCGAN, as imagens foram redimensionadas para 64x64 *pixels* e centralizadas. Ao final do trabalho, os autores utilizam o ELM para classificar as células dentro das três classes utilizando apenas dados reais e dados reais com a adição de imagens sintéticas, como ilustrado na Tabela 2.4. Além disso, [Rando et al. \(2022\)](#) realizam a comparação da acurácia entre essas duas medições.

O trabalho apresenta grande colaboração para o aumento de dados, visto que as imagens que foram geradas pelo modelo proposto aumentaram a eficiência do classificador ELM.

2.2.5 Generative Adversarial Networks for Augmenting Training Data of Microscopic Cell Images

O trabalho realizado por [Baniukiewicz et al. \(2019\)](#) tem como objetivo criar imagens de células microscópicas sintéticas em 2D e 3D através de GANs. Os autores destacam a relevância do processo de aumento de dados de um *dataset*, visto que, principalmente quando se trata de imagens 3D, existe uma limitação significativa na obtenção de dados de treinamento. Fato que resulta na necessidade de abordagens que utilizam imagens 2D para geração de imagens pseudo-3D, como a apresentada neste trabalho.

Diversos conceitos propostos por Isola et al. (2017) foram utilizados pelos autores, um deles foi a tarefa de traduzir uma possível representação de cena em outra, a *image-to-image translation* através de GANs condicionais. Baniukiewicz et al. (2019) utilizaram também a implementação do autor no gerador com arquitetura *U-Net*, que foi proposta por (ÇIÇEK et al., 2016).

Em relação ao discriminador, foi utilizado um classificador convolucional *PatchGAN*, desenvolvido por Li e Wand (2016). Além disso, para estabelecer uma relação entre as células segmentadas e suas imagens microscópicas, Baniukiewicz et al. (2019) utilizaram as *Perdas LI+cGAN*, por Isola et al. (2017). Dessa forma, máscaras binárias 2D e 3D foram utilizadas para gerar imagens realistas de células.

O autor menciona ao decorrer do trabalho que existe um alto custo computacional e grande dificuldade na obtenção de conjunto de dados suficientemente grandes para treinar modelos que gerem as imagens 3D. Para contornar essa limitação, foi proposta uma abordagem que redes 2D sejam utilizadas para sintetizar conjunto de dados pseudo-3D e, ao comparar os resultados com dados reais, foi evidenciado que as características relevantes das imagens foram preservadas.

Tabela 2.5 – Métricas da performance do classificador para os dados gerados pela GAN.

Network	Recall	F1-score	Precision
net_ABD	96%	98%	100%
net_cAR1	99%	97%	95%
net_TalA	100%	100%	100%

Fonte: Baniukiewicz et al. (2019)

O modelo foi treinado com imagens de células *Dictyostelium* rotuladas por três marcadores fluorescentes diferentes: ABD, CAR1 e TalA. Para fins de análise dos resultados do trabalho, experimentos foram realizados através de um classificador que terá a função de avaliar a performance da rede geradora para as três classes. O classificador utiliza Rede Neural Convolucional (CNN) e foi treinado com 2.592 imagens reais. Ao final, ele foi utilizado para prever qual tipo de marcador fluorescente foi utilizado e a Tabela 2.5 mostra as métricas estatísticas do classificador para cada uma das classes. Através da análise dos resultados, é possível perceber que o artigo de Baniukiewicz et al. (2019) possui grande contribuição para a geração de imagens de células, uma vez que apresenta uma forma eficiente de enriquecer a diversidades de amostras presentes nos *datasets* de imagens celulares.

2.2.6 Análise das Técnicas Utilizadas

Através da análise dos resultados apresentados nos Trabalhos Relacionados acima, nota-se que o emprego de GANs para aumento de *datasets* garante maior qualidade no processo de treinamento de redes classificadoras, visto que em todos os trabalhos, houve um ganho nas métricas de avaliação das redes de classificação utilizadas pelos autores,

Diversas arquiteturas de GANs foram utilizadas nos trabalhos e cada uma delas obteve resultados satisfatórios, cumprindo assim, os objetivos propostos pelos autores. Além disso, diferentes *datasets* foram utilizados, o que demonstra que existem diversos caminhos a serem seguidos ao implementar uma Rede Generativa Adversária para aumento de *datasets*. A Tabela 2.6 exibe as principais informações relacionadas ao treinamento e bases de dados utilizadas.

Tabela 2.6 – Relação das estratégias abordadas e *datasets* utilizados nos experimentos.

Trabalho	Experimentos	Base de dados
Yu et al. (2021)	4 estratégias	Hospital de Baoding, China
Chen et al. (2020)	<i>Real x Fake</i>	<i>Danish Technical University, Herlev University</i>
Zak et al. (2022)	3 estratégias	<i>HERlev Pap Smear</i>
Rando et al. (2022)	<i>Real x Real + Fake</i>	dr. Soetomo, Hospital Surabaya, Indonésia
Baniukiewicz et al. (2019)	<i>Train: real, Test: Fake</i>	<i>Células Dictyostelium</i>

Fonte: Produzida pela autora

Em todos os artigos abordados, a análise dos resultados foi realizada através da utilização de um classificador, visto que ao final, é possível obter diversas métricas, que facilitam a comparação de resultados. A Tabela 2.7 mostra as diferentes abordagens de GANs e CNNs para análise dos resultados.

Tabela 2.7 – Arquiteturas de GANs e CNNs utilizadas nos Trabalhos Relacionados.

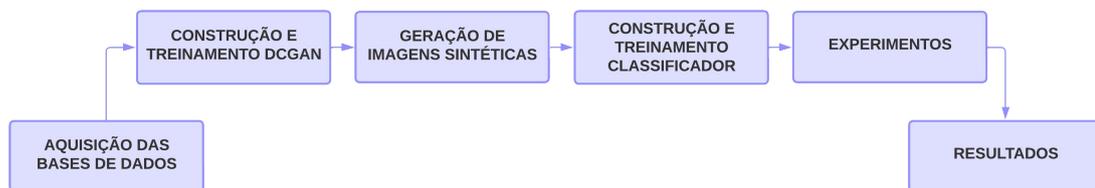
Trabalho	Arquitetura GAN	Arquitetura CNN
Yu et al. (2021)	<i>5-layer CNN</i>	Alex-net
Chen et al. (2020)	RCGAN	-
Zak et al. (2022)	CGAN (pix2pix)	VGG16, InceptionV3, ResNet50, DenseNet121
Rando et al. (2022)	DCGAN	ELM
Baniukiewicz et al. (2019)	CGAN	<i>4 Conv Layers + FC on output</i>

Fonte: Produzida pela autora

3 Desenvolvimento

Este capítulo tem como objetivo apresentar a metodologia aplicada no decorrer da pesquisa. A Figura 3.1 apresenta um fluxograma das etapas. Posteriormente, os conjuntos de dados empregados são detalhados na Seção 3.1. A Seção 3.2 apresenta o pré-processamento aplicado em um dos bancos de dados utilizados. Posteriormente, o estágio de implementação dos modelos generativos é introduzido na Seção 3.3, na qual é detalhada a arquitetura da DCGAN utilizada, o treinamento realizado e especificidades sobre as imagens geradas. Por fim, a Seção 3.4 apresenta o modelo classificador utilizado para os experimentos e avaliação das imagens geradas pela DCGAN.

Figura 3.1 – Fluxograma que descreve a metodologia aplicada na pesquisa.



Fonte: Produzido pela autora.

3.1 Caracterização dos *datasets*

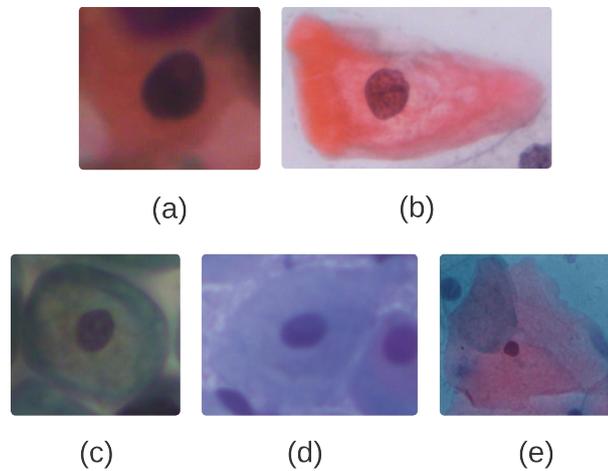
Os modelos de aprendizado de máquina implementados neste trabalho utilizam bancos de dados de imagens de células cervicais para o treinamento. Por isso, utilizou-se o *SIPaKMeD* (PLISSITI et al., 2018) e o *CRIC* (REZENDE et al., 2021) como fontes de imagens reais. Ambos conjuntos de dados possuem uma coleção de imagens microscópicas obtidas através do exame de Papanicolau.

3.1.1 *SIPaKMeD*

A criação da base de dados *SIPaKMeD* originou-se da demanda por conjuntos de dados públicos que possibilitassem aos autores a comparação do desempenho de seus modelos de classificação de tipos celulares. Isso se tornou essencial, pois grande parte dos modelos contavam com conjuntos de dados criados pelos próprios autores (PLISSITI et al., 2018).

O *dataset* *SIPaKMeD* é composto por 4.031 células manualmente recortadas e classificadas, originárias de 966 imagens de lâminas do exame de Papanicolau. A captura das células foi realizada através de uma câmera CCD (*Charge-Coupled Device*) *Infinity 1 Lumenera* adaptada a um microscópio óptico *OLYMPUS BX53F* (PLISSITI et al., 2018). A Figura 3.2 apresenta exemplos de células recortadas do *dataset*.

Figura 3.2 – Exemplos de amostras do *dataset* SIPaKMeD. (a) Amostra da classe *Dyskeratotic*; (b) Amostra da classe *Koilocytotic*; (c) Amostra da classe *Metaplastic*; (d) Amostra da classe *Parabasal*; (e) Amostra da classe *Superficial/Intermediate*.



Fonte: Plissiti et al. (2018). Imagens selecionadas pela autora.

As imagens presentes no SIPaKMeD estão classificadas em três grupos: células normais (Superficial-Intermediate e Parabasal), células anormais (Koilocytotic e Dyskeratotic) e células benignas (Metaplastic). Essa classificação foi realizada por citopatologistas profissionais de acordo com a aparência e morfologia celular (PLISSITI et al., 2018). A Tabela 3.1 demonstra a distribuição de imagens contidas no SIPaKMeD.

Tabela 3.1 – Contagem de tipo de células contidas no SIPaKMeD.

Classificação	Quantidade de Imagens	Quantidade de Células
Superficial/Intermediate	126	813
Parabasal	108	787
Koilocytotic	238	825
Metaplastic	271	793
Dyskeratotic	223	813
Total	966	4.031

Fonte: Plissiti et al. (2018)

O conteúdo da base de dados está disponível para *download* no *website* *SIPaKMeD Database*⁸, proposta por Plissiti et al. (2018).

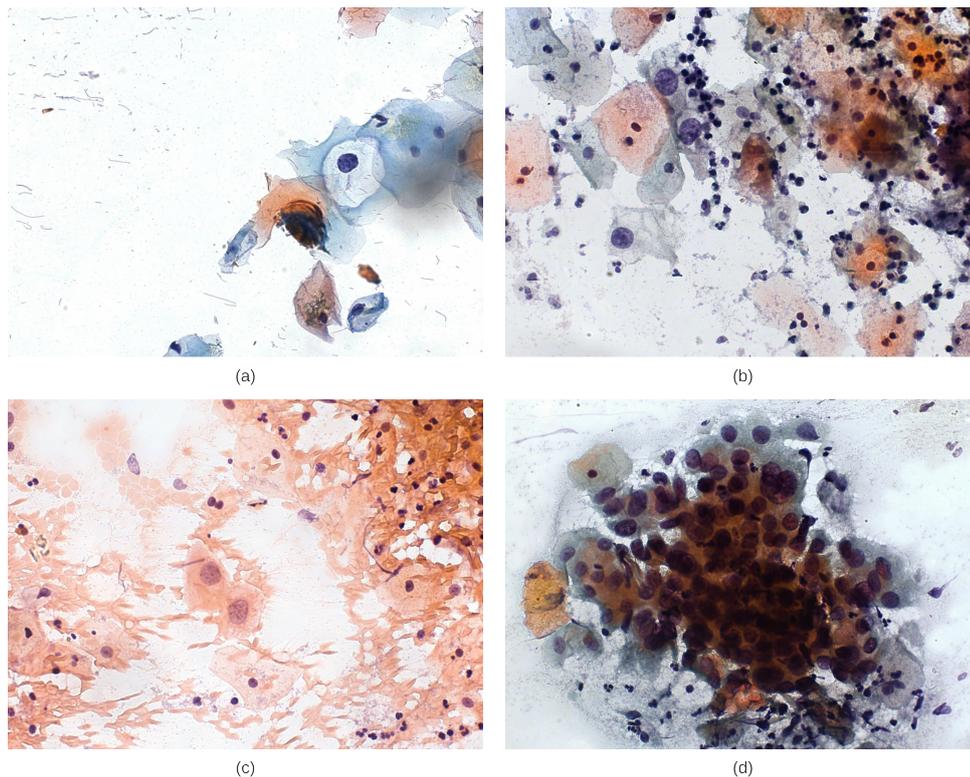
3.1.2 CRIC Cervix

O banco de dados *CRIC Cervix* do *Center for Recognition and Inspection of Cells*, apresentado no trabalho (REZENDE et al., 2021), possui uma coleção de imagens obtidas através do esfregaço convencional de Papanicolau.

⁸ <https://www.cs.uoi.gr/marina/sipakmed.html>

A base de dados contém 400 imagens coletadas de exames de papanicolau. Essas amostras contêm 11.534 células que foram identificadas e classificadas manualmente com base na possível lesão que apresentam. Esse processo foi realizado com a validação de três citopatologistas, os quais seguiram o protocolo de classificação de lesões celulares definidos pelo Sistema *Bethesda*. Todas as imagens estão com 1.376x1.020 *pixels* de dimensão e resolução de 0.228 $\mu\text{m}/\text{pixel}$ (REZENDE et al., 2021). A Figura 3.3, apresenta exemplos de amostras do *dataset*.

Figura 3.3 – Exemplos de amostras do *dataset* CRIC. (a) Imagem 164 da base de dados; (b) Imagem 349 da base de dados; (c) Imagem 213 da base de dados; (d) Imagem 395 da base de dados.



Fonte: Rezende, Gomes e Carneiro (2021). Imagens selecionadas pela autora.

O conteúdo da base de dados está disponível para *download* na plataforma *CRIC Searchable Image Database*⁹, proposta em (REZENDE et al., 2021). O *dataset* possui os seguintes atributos:

- *[image_id]*: Identificação da imagem, varia de 1 a 400;
- *[image_filename]*: Nome e extensão da imagem;
- *[image_doi]*: Identificador de Objeto Digital (DOI) da imagem na plataforma Figshare;
- *[cell_id]*: Identificação da célula encontrada na imagem, varia de 1 a 11.534;
- *[bethesda_system]*: Classificação segundo o Sistema *Bethesda*;

⁹ <https://database.cric.com.br>

- *[nucleus_x]*: Eixo X da coordenada do núcleo da célula na imagem;
- *[nucleus_y]*: Eixo Y da coordenada do núcleo da célula na imagem.

A Tabela 3.2 apresenta as seis classes possíveis de lesões celulares e suas quantidades. Abaixo são apresentados os nomes de cada classe de acordo com o Sistema *Bethesda*:

1. Negativo para lesão intraepitelial ou malignidade (NILM)
2. Células escamosas atípicas de significado indeterminado, possivelmente não neoplásticas (ASC-US)
3. Células escamosas atípicas, não é possível excluir lesão de alto grau (ASC-H)
4. Lesão intraepitelial escamosa de baixo grau (LSIL)
5. Lesão intraepitelial escamosa de alto grau (HSIL)
6. Carcinoma de células escamosas (SCC)

Tabela 3.2 – Contagem de tipo de células contidas no CRIC *database*.

Classificação	Quantidade de Imagens
NILM	6.779
ASC-US	606
ASC-H	925
LSIL	1.360
HSIL	1.703
SCC	161
Total	11.534

Fonte: Rezende et al. (2021)

Apesar do conjunto de dados ser dividido em seis classes, existem outras possibilidades de divisão. Pode-se dividir em células lesionadas e não lesionadas, ou por grau de lesão, conforme demonstrado em (DINIZ et al., 2021). A metodologia adotada neste trabalho foi a utilização dos três graus de lesão, uma vez que treinar modelos para as seis classes seria custoso, e treinar apenas para duas poderia generalizar demais as diferenças entre lesões de alto e baixo grau. Portanto, os dados foram divididos em células não lesionadas (NILM), lesão de baixo grau (ASC-US e LSIL) e lesão de alto grau (ASC-H, HSIL e SCC).

3.2 Preparação da base de dados

Um dos modelos implementados utilizou o conjunto de dados CRIC como fonte de imagens reais. Porém, esse *dataset* é composto por imagens que possuem diversas células com

diferentes rótulos do Sistema *Bethesda*. Portanto, fez-se necessário incluir uma etapa de pré-processamento nas imagens, que consiste na extração de regiões que contenham os núcleos celulares. Esta necessidade surge da hipótese de que o núcleo contém informações adequadas para representar as lesões celulares. Portanto, optou-se por realizar o recorte dos núcleos.

Para a extração dos núcleos, foi desenvolvido um algoritmo na linguagem de programação *Python* que, com base nas coordenadas de cada núcleo, realizou o recorte dessas estruturas, seguindo as dimensões de 100×100 *pixels*, valor obtido através de experimentos empíricos. Após a extração, realizou também a associação da imagem com uma das seis classes. A Figura 3.4 apresenta exemplos de regiões de células que foram recortadas das imagens do *CRIC Database*.

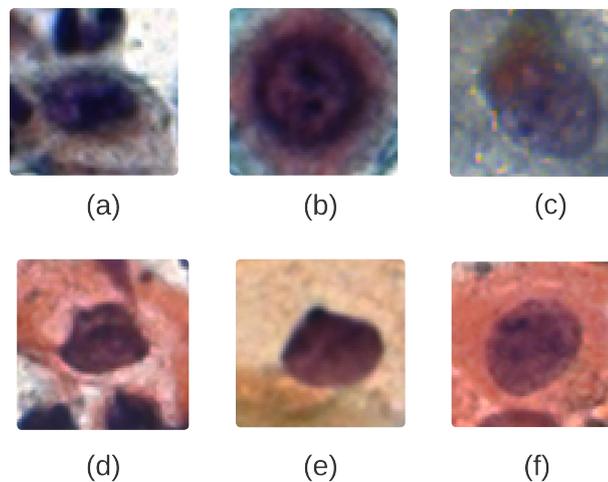


Figura 3.4 – Células recortadas das imagens do *CRIC Database*. (a) e (b) Lesões de alto grau. (c) e (d) Lesões de baixo grau. (e) e (f) Células não lesionadas. Fonte: Produzido pela autora.

3.3 Geração de imagens

Assim como o trabalho realizado por [Rando et al. \(2022\)](#), este estudo visa gerar imagens sintéticas de células cervicais através da DCGAN, uma arquitetura de GAN proposta no artigo ([RADFORD; METZ; CHINTALA, 2016](#)). Adicionalmente, implementou-se a estratégia de pré-treinamento. Por fim, os pesos do gerador foram salvos, possibilitando sua utilização independente para gerar a quantidade de desejada de imagens. As seções a seguir descrevem a metodologia adotada no processo de implementação da DCGAN e na geração de imagens.

3.3.1 Arquitetura do modelo DCGAN

O trabalho proposto por [RADFORD; METZ; CHINTALA](#) introduziu uma estrutura de GANs convolucionais denominada *Deep Convolutional Generative Adversarial Networks*. Segundo os autores, as DCGANs demonstraram ser uma estratégia eficaz ao serem treinadas com conjuntos de dados provenientes de diversas áreas, apresentando resultados satisfatórios

em diversas aplicações. Além disso, a escolha desta arquitetura foi motivada pelos resultados apresentados no trabalho de [Rando et al. \(2022\)](#) e pela disponibilidade de conteúdos explicativos para serem utilizados como referência na construção do modelo.

Em relação a arquitetura, os autores [RADFORD; METZ; CHINTALA](#) implementaram modificações na estrutura básica da GAN, a fim de preencher as lacunas identificadas no desempenho do modelo apresentado em ([GOODFELLOW et al., 2014](#)). Essas adaptações incluíram a eliminação das camadas totalmente conectadas. A normalização em lote foi incorporada tanto no gerador, quanto no discriminador. No gerador, a ativação *ReLU* foi adotada em todas as camadas, exceto na camada de saída, que utiliza a função de tangente hiperbólica. Quanto ao discriminador, a ativação *LeakyReLU* foi aplicada em todas as camadas ([RADFORD; METZ; CHINTALA, 2016](#)). A Figura 3.5 ilustra a arquitetura das DCGANs implementadas neste trabalho.

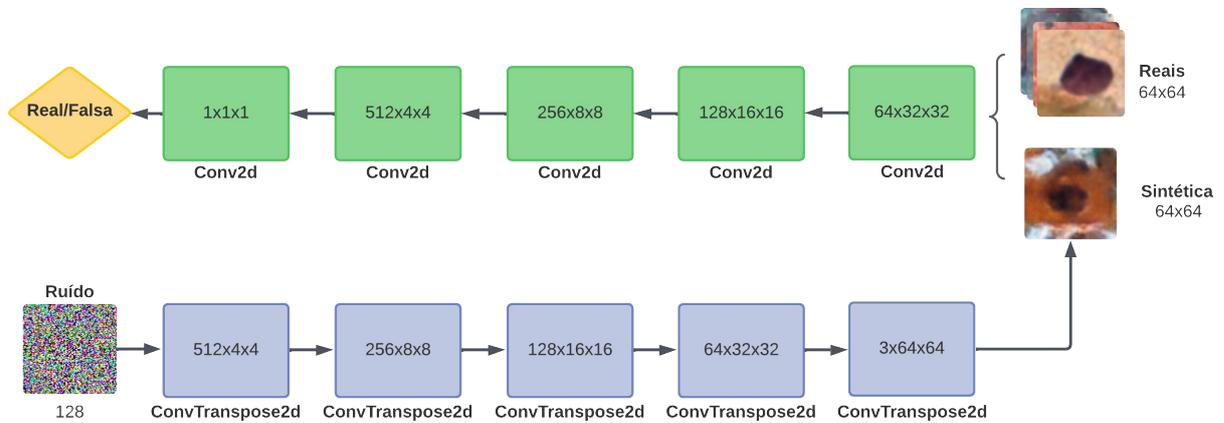


Figura 3.5 – Arquitetura das DCGANs implementadas.

Além disso, [RADFORD; METZ; CHINTALA](#) sugeriram alguns parâmetros em seu artigo, valores que foram utilizados na implementação realizada neste estudo (Tabela 3.3). Os modelos construídos neste trabalho utilizaram as especificações do artigo dos autores.

Tabela 3.3 – Parâmetros utilizados na implementação da DCGAN.

Parâmetro	Valor
Tamanho mini-batch	128
Função de ativação Tahn	[-1. 1]
Desvio padrão	0,02
LeakyReLU	0,2
Learning Rate	0,0002
B1	0,5

Fonte: Radford, Metz e Chintala (2016)

3.3.2 Treinamento dos Modelos

Para este trabalho, foram empregadas duas DCGANs, a primeira foi treinada do zero usando os dados do SIPaKMeD até a 250^a época por apresentar menor quantidade de ruídos nas imagens produzidas. A segunda DCGAN foi treinada a partir desse ponto de partida, mas utilizando os dados do CRIC. Em ambos os modelos, as imagens de treinamento foram redimensionadas para 64x64 *pixels*, e o vetor Z , utilizado como entrada no gerador, foi inicializado com o valor 128, para que fossem geradas amostras diversas. A quantidade de épocas e o valor utilizado no vetor Z foram selecionados através da realização de experimentos empíricos.

3.4 Classificação de Imagens

Após a geração das imagens artificiais, que é o foco principal deste estudo, um algoritmo classificador foi desenvolvido para avaliar se os dados sintéticos têm a capacidade de melhorar o desempenho de modelos de diagnóstico assistido por computador.

3.4.1 Arquitetura CNN

Com a finalidade de construir um modelo simples, a CNN foi configurada com três camadas convolucionais, utilizando um tamanho de *kernel* igual a 3, *stride* e *padding* ambos iguais a 1, e incorporando batch normalization. O otimizador escolhido foi Adam, e, por meio de experimentos, definiu-se o valor de *learning rate* em 0,0001 e *weight decay* em 1e-6. O treinamento foi conduzido por 15 épocas, com um *batch size* de 128.

Foram empregadas duas abordagens distintas para treinamento e teste. A primeira delas envolve o uso do conjunto de dados CRIC balanceado, sem a inclusão de imagens sintéticas para treinamento e teste do modelo. A segunda delas consiste no treinamento com o conjunto de dados aumentado, seguido pelo teste exclusivamente com imagens reais, ambos balanceados.

A avaliação dos resultados obtidos foi realizada através da média das métricas obtidas na execução das cinco iterações para cada técnica. A apresentação e discussão desses resultados são feitas no Capítulo 4.

4 Resultados

Este capítulo visa apresentar os experimentos (características) e destacar os resultados alcançados ao longo deste estudo. Portanto, as etapas realizadas antes e durante a execução dos modelos são relatadas neste capítulo.

4.1 Ambiente de Desenvolvimento

Os modelos de aprendizado de máquina foram desenvolvidos no Google Colaboratory¹ em sua versão Pro, com a linguagem Python² para a implementação do código. As principais bibliotecas empregadas foram Pytorch³, Numpy⁴, Pandas⁵, scikit-learn⁶ e Matplotlib⁷. Em termos de configurações do ambiente de execução, foram aplicadas as seguintes especificações: 51 GB de memória RAM, 15 GB de memória RAM da GPU, e 166,8 GB de espaço em disco, com a utilização de uma GPU T4.

4.2 Geração de imagens

Assim como o estudo desenvolvido por YU et al., este trabalho empregou técnicas de transferência de aprendizado. Para essa finalidade, foram desenvolvidas duas DCGANs, modelos generativos capazes de gerar imagens sintéticas a partir de dados reais. Ambos os modelos compartilham implementações semelhantes, no entanto, cada um tem uma finalidade específica para a implementação. As especificações dos modelos desenvolvidos em cada experimento são descritas a seguir.

4.2.1 DCGAN Padrão

A primeira DCGAN foi implementada seguindo as recomendações de parâmetros do trabalho de Radford, Metz e Chintala (2016). Após a implementação do modelo, deu-se início ao processo de treinamento. Nesse sentido, optou-se pela utilização da base de dados SIPaKMeD para fornecer ao modelo células cervicais reais, uma vez que ela abrange uma variedade de imagens citológicas provenientes de exames de Papanicolau. O objetivo desta etapa era permitir

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

² <https://www.python.org/>

³ <https://pytorch.org/>

⁴ <https://numpy.org/>

⁵ <https://pandas.pydata.org/>

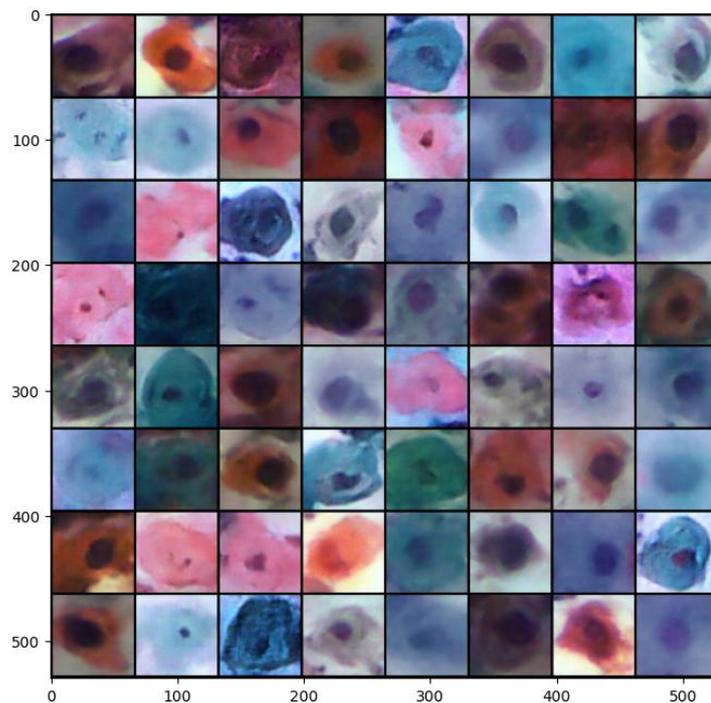
⁶ <https://scikit-learn.org/>

⁷ <https://matplotlib.org/>

que o modelo aprendesse a representar estruturas celulares básicas, não ocorreu a divisão entre as classes do conjunto, ou seja, todas as células foram agrupadas em um único conjunto.

O modelo foi executado por 250 épocas e ao término desse processo, o peso mais recente foi armazenado para servir como ponto de partida na construção da DCGAN pré-treinada. Na Figura 4.1, são exibidas as células sintéticas geradas pelo modelo ao longo do treinamento, seguindo as configurações mencionadas.

Figura 4.1 – Células sintéticas geradas pela DCGAN padrão ao final das 250 épocas.



Fonte: Produzida pela autora.

4.2.2 DCGAN Pré-Treinada

O modelo padrão foi construído para ser utilizado como base na construção da DCGAN pré-treinada, ou seja, o treinamento do primeiro modelo é aproveitado como ponto de partida no segundo. Em relação à arquitetura, os dois modelos não se diferem, uma vez que tanto o gerador, quanto o discriminador possuem as mesmas configurações.

Com a conclusão da implementação do modelo e do pré-processamento dos dados (Seção 3.2), deu-se início ao treinamento do modelo através do *dataset* CRIC. A partir da configuração mencionada, treinamos três modelos de DCGAN por 350 épocas cada, de tal forma que cada modelo correspondesse a cada classe: sem lesão, lesão de baixo grau e lesão de alto grau. Após cada treinamento, os pesos de cada classe foram salvos separadamente.

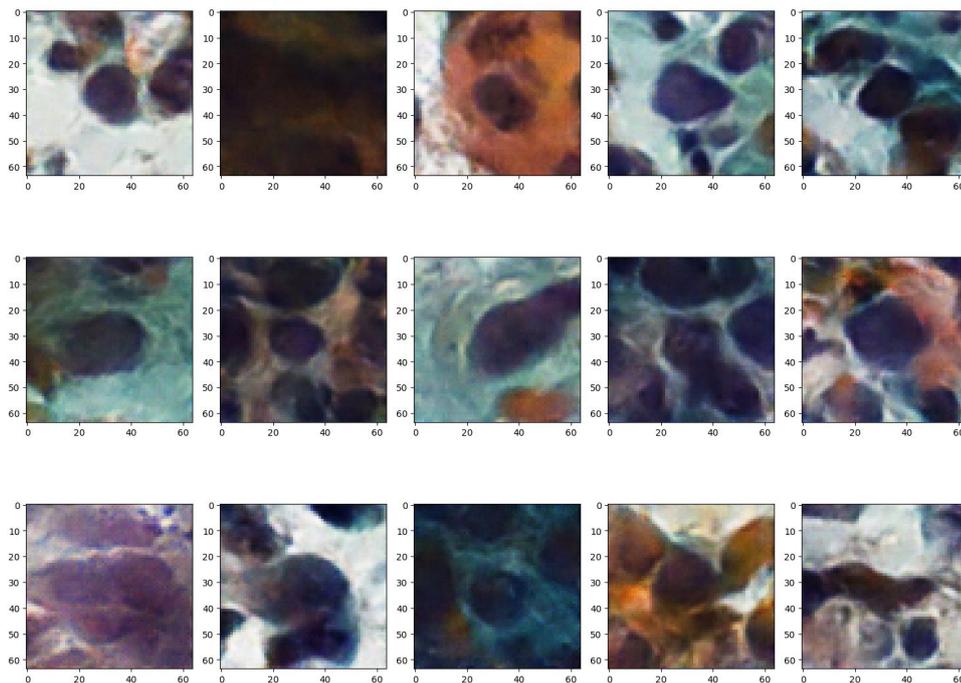
Ao final, utilizou-se o peso de cada gerador da DCGAN treinada para gerar a quantidade desejada de imagens.

As Figuras 4.2, 4.3 e 4.4 exibem 15 amostras de células geradas pela DCGAN pré-treinada. Cada figura representa uma das classes, respectivamente, células com lesão de alto grau, baixo grau e não lesionadas, respectivamente.

Observa-se que, mesmo utilizando os pesos da DCGAN treinada a partir da base de dados SIPaKMeD, o modelo foi capaz de gerar imagens sintéticas que se assemelham às do conjunto de dados CRIC, utilizado no treinamento deste segundo modelo.

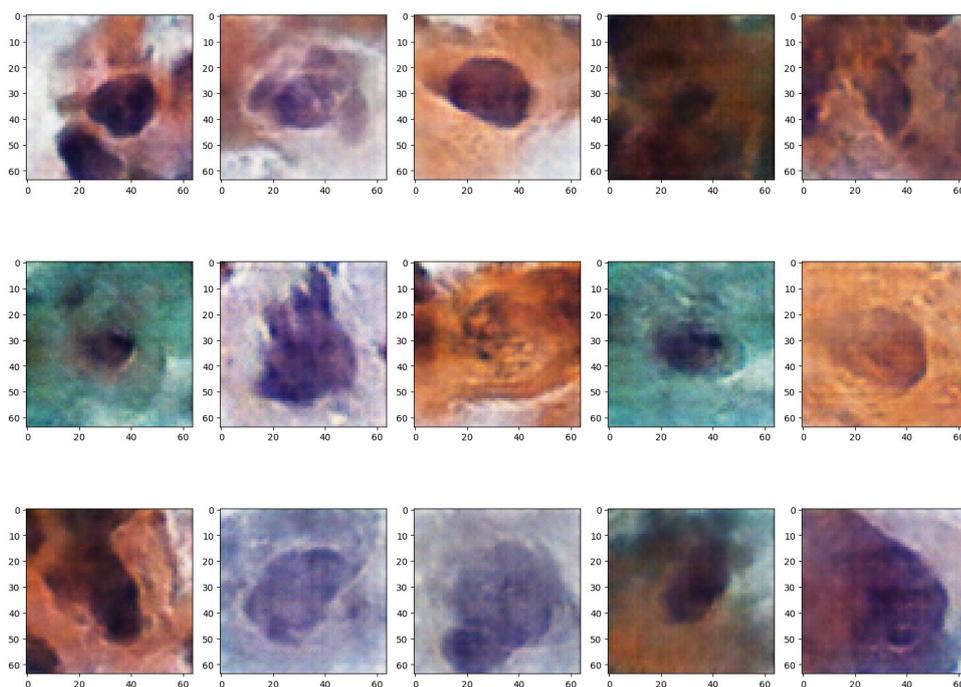
A utilização do aprendizado por transferência a partir da base de dados SIPaKMeD, a GAN treinada com os dados da base CRIC foi capaz de gerar imagens sintéticas para a base CRIC. As imagens de cada classe contém a representação das células. Além disso, é possível notar que algumas imagens possuem ruídos que podem afetar a qualidade da imagem dependendo da aplicação, como são mostradas nas Figuras 4.3 e 4.4. Atribuímos essa limitação a quantidade de épocas de treinamento, ou seja, pode ser superado ao empregar um maior número de épocas durante o treinamento, possibilitando que o modelo adquira uma compreensão mais profunda das características das classes e, assim, melhore a sua representação visual.

Figura 4.2 – Amostras de células sintéticas com lesão de alto grau.



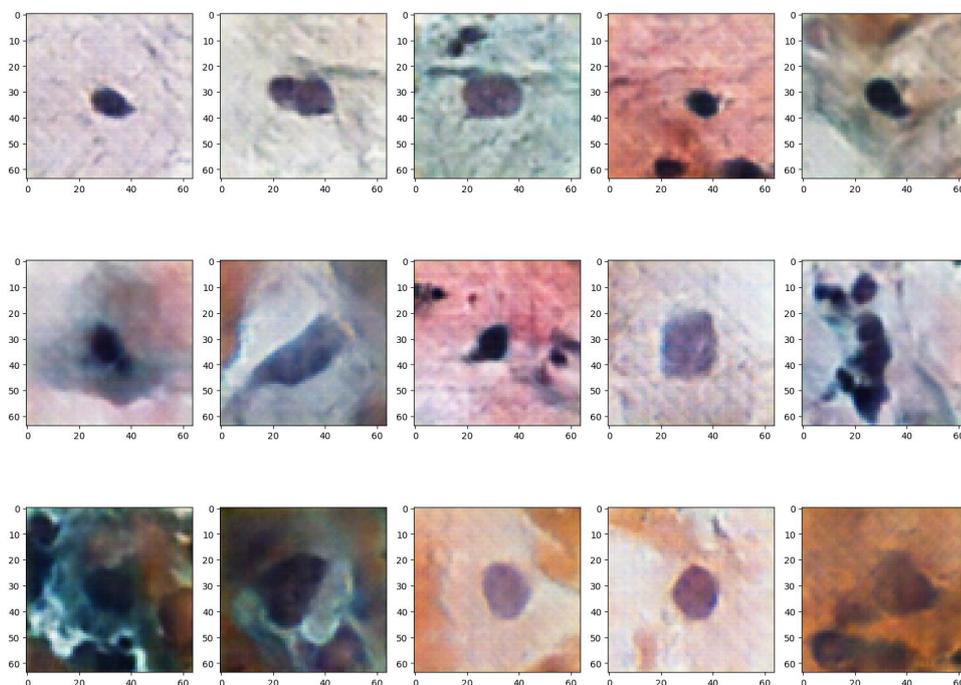
Fonte: Produzida pela autora.

Figura 4.3 – Amostras de células sintéticas com lesão de baixo grau.



Fonte: Produzida pela autora.

Figura 4.4 – Amostras de células sintéticas não lesionadas.



Fonte: Produzida pela autora.

4.3 Classificação de Imagens

Assim como os trabalhos realizados por YU et al., CHEN et al., ZAK et al., RANDO et al. e BANIUKIEWICZ et al., este trabalho optou desenvolver um modelo de classificação para avaliar o impacto da geração de imagens sintéticas. Portanto, a CNN especificada na Seção 3.4 foi implementada para tornar viável a avaliação de como a inclusão delas afeta o treinamento de uma rede neural convolucional.

Dado que deseja-se avaliar o comportamento do modelo ao receber um volume de imagens geradas pela DCGAN, torna-se necessário comparar as métricas do classificador antes e depois desse aumento de dados. Dessa forma, empregaram-se duas estratégias de treinamento: a utilização exclusiva de dados reais e a utilização de imagens sintéticas e reais. Em ambas as estratégias, o conjunto de dados de células recortadas provenientes do CRIC *database* foi balanceado, utilizando como referência a classe com menor quantidade de imagens, ou seja, a classe de lesões de baixo grau foi estabelecida como o limite para o número de imagens em todas as classes.

Adicionalmente, na segunda estratégia, foram incluídas imagens falsas, criando assim, um conjunto de dados aumentado. Por fim, o teste foi conduzido apenas com imagens reais para ambas as estratégias.

O desempenho do modelo classificador foi avaliado por meio das métricas mais comumente utilizadas nas pesquisas da seção Trabalhos Relacionados (Seção 2.2). Assim, a análise engloba as seguintes métricas: acurácia, precisão, *recall* e *F1-score*. As seções a seguir apresentam os resultados do classificador com o *dataset* de células recortadas após a adição de 600, 750 e 1.500 imagens sintéticas, que foram distribuídas igualmente para cada classe. Essa metodologia foi aplicada com a finalidade de entender o desempenho do classificador para diferentes tamanhos de bancos de dados.

4.3.1 Estratégias de Treinamento

A Tabela 4.1 exibe os resultados do teste para a execução da CNN com 0, 600, 750 e 1.500 imagens sintéticas. Desse modo, a base de dados sem imagens sintéticas é constituída exclusivamente por células recortadas e redimensionadas da base de dados CRIC, enquanto as demais incluem essas adições à base de dados original.

Tabela 4.1 – Resultados da CNN treinada com diferentes quantidades de imagens sintéticas.

Imagens Sintéticas	Acurácia	Precisão	Recall	F1-Score
0	84,33%	84,35%	84,33%	84,33%
600	78,40%	78,64%	78,40%	78,11%
750	79,55%	79,44%	79,55%	79,43%
1.500	76,38%	76,33%	76,38%	76,35%

Fonte: Produzida pela autora

Ao analisar os dados da Tabela 4.1, percebe-se que o modelo demonstrou um desempenho positivo nos quatro experimentos. Contudo, as estratégias que empregaram o banco de dados aumentado apresentaram resultados inferiores à abordagem que utilizou exclusivamente o CRIC, alcançando uma acurácia de 84,33%. Assim, diferentemente dos estudos apresentados na Seção 2.2, não foi possível validar os benefícios do aumento de dados através da geração de imagens artificiais. Portanto, os resultados indicam a necessidade de aprimoramentos na arquitetura e nos hiperparâmetros da CNN, bem como a exploração de outras estratégias de treinamento.

4.4 Disponibilização do código implementado

O código elaborado neste trabalho está disponível por meio da URL <<https://github.com/vitoriabispo/Monografia.git>> no GitHub⁸.

⁸ <https://github.com>

5 Considerações Finais

Este capítulo apresenta as considerações finais deste trabalho. Dessa forma, a Seção 5.1 apresenta as conclusões obtidas através dos experimentos. Por fim, a Seção 5.2 apresenta os trabalhos a serem realizados futuramente.

5.1 Conclusão

Esta monografia empregou DCGANs para gerar imagens sintéticas de citologia cervical a fim de aumentar a quantidade e a diversidade de dados de conjuntos de citologia cervical. A investigação visava determinar a viabilidade do emprego de GANs na geração de imagens sintéticas como uma alternativa eficaz para o aumento de conjuntos de dados biológicos.

A geração de imagens envolveu o uso de duas DCGANs: uma foi treinada com a base de dados SIPaKMeD, que envolve vários tipos de células cervicais, que foi desenvolvida por [PLISSITI et al.](#), e a segunda DCGAN foi pré-treinada utilizando os pesos resultantes da primeira, estratégia utilizada no trabalho de [YU et al.](#), e as células recortadas do conjunto de dados CRIC, conforme apresentado no trabalho de [REZENDE et al.](#) Por fim, o gerador do segundo modelo foi utilizado para gerar uma determinada quantidade de imagens de cada classe.

Quanto à avaliação dos resultados, inicialmente propusemos a implementação de uma CNN para classificar as imagens em três níveis de lesões: alto grau, baixo grau e não lesionadas. Os resultados obtidos para os classificadores superaram 70%, indicando preliminarmente que as imagens geradas possuem boa qualidade e podem representar células cervicais. No entanto, uma análise mais aprofundada é necessária, explorando diversas redes de classificação e estratégias de execução, a fim de obter uma compreensão mais abrangente e confiável sobre as imagens geradas.

Vários aspectos da implementação podem ter contribuído para este resultado. Assim, algumas abordagens para melhorar o desempenho do classificador ao lidar com dados sintéticos são discutidas na Seção 5.2.

5.2 Trabalhos Futuros

Este trabalho evidenciou que a utilização de GANs na geração de imagens sintéticas trata-se de uma alternativa viável para o aumento de conjuntos de dados biológicos. No entanto, diversas oportunidades de melhorias e novas abordagens surgem como perspectivas para futuras pesquisas.

Inicialmente destacamos a importância de empregarmos métricas de avaliação específicas

para GANs, com o objetivo de analisar a qualidade, a similaridade e a diversidade das imagens geradas pela DCGAN. A implementação dessas métricas proporcionaria uma compreensão aprofundada dos resultados obtidos neste estudo.

Outro fator a ser considerado é uma melhor investigação e um aprofundamento nos estudos acerca de algoritmos de classificação das lesões. Devido a limitações encontradas na implementação, não foi possível tirar conclusões mais qualitativas a respeito da melhoria da precisão de classificadores usando as imagens sintéticas. A estratégia consistiria na simulação da aplicação para diagnóstico assistido por computador. Além disso, é crucial desenvolver métodos que visem melhorar o desempenho do modelo classificador, uma vez que a etapa de pré-processamento envolveu apenas a extração da região dos núcleos e redimensionamento das imagens.

Outro aspecto relevante a ser explorado refere-se ao peso salvo no primeiro modelo da DCGAN. Este trabalho adotou o último peso, correspondente à 250 épocas. No entanto, queremos investigar a possibilidade de salvar o melhor peso, o que pode resultar em uma melhoria na qualidade das imagens geradas pela DCGAN pré-treinada.

Referências

- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)*. [S.l.: s.n.], 2017. p. 1–6.
- BANIUKIEWICZ, P.; LUTTON, E. J.; COLLIER, S.; BRETSCHNEIDER, T. Generative adversarial networks for augmenting training data of microscopic cell images. *Frontiers in Computer Science*, v. 1, 2019. ISSN 2624-9898. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fcomp.2019.00010>>.
- BETZALEL, E.; PENSO, C.; NAVON, A.; FETAYA, E. *A Study on the Evaluation of Generative Models*. 2022.
- BRAGA, A. de P.; CARVALHO, A. de L. F.; LUDERMIR, T. *Redes neurais artificiais: teoria e aplicações*. Livros Técnicos e Científicos, 2000. ISBN 9788521612186. Disponível em: <<https://books.google.com.br/books?id=cUgEaAEACAAJ>>.
- CHEN, S.; GAO, D.; WANG, L.; ZHANG, Y. Cervical cancer single cell image data augmentation using residual condition generative adversarial networks. In: *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. [S.l.: s.n.], 2020. p. 237–241.
- ÇIÇEK, Ö.; ABDULKADIR, A.; LIENKAMP, S. S.; BROX, T.; RONNEBERGER, O. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: OURSELIN, S.; JOSKOWICZ, L.; SABUNCU, M. R.; UNAL, G.; WELLS, W. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing, 2016. p. 424–432. ISBN 978-3-319-46723-8.
- CLARO, M.; VOGADO, L.; SANTOS, J.; VERAS, R. Utilização de técnicas de data augmentation em imagens: Teoria e prática. *Sociedade Brasileira de Computação*, 2020.
- CÂNCER, I. N. de. *Falando sobre Câncer do Colo do Útero*. [s.n.], 2002. Disponível em: <https://bvsmms.saude.gov.br/bvs/publicacoes/inca/falando_cancer_colo_uterio.pdf>.
- CÂNCER, I. N. do. *Estimativa | 2023 Incidência de Câncer no Brasil*. [s.n.], 2022. Disponível em: <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf>>.
- DALIANIS, H. Evaluation metrics and evaluation. In: _____. *Clinical Text Mining: Secondary Use of Electronic Patient Records*. Cham: Springer International Publishing, 2018. p. 45–53. ISBN 978-3-319-78503-5. Disponível em: <https://doi.org/10.1007/978-3-319-78503-5_6>.
- DINIZ, D. N.; REZENDE, M. T.; BIANCHI, A. G.; CARNEIRO, C. M.; USHIZIMA, D. M.; MEDEIROS, F. N. de; SOUZA, M. J. A hierarchical feature-based methodology to perform cervical cancer classification. *Applied Sciences*, MDPI, v. 11, n. 9, p. 4091, 2021.
- FURTADO, M. I. V. *Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula*. [s.n.], 2019. Disponível em: <<http://educapes.capes.gov.br/handle/capes/432794>>.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 63, n. 11, p. 139–144, oct 2020. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3422622>>.

GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. *Generative Adversarial Networks*. 2014.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, G.; CAI, J.; CHEN, T. Recent advances in convolutional neural networks. *Pattern Recognition*, v. 77, p. 354–377, 2018. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320317304120>>.

HARSHVARDHAN, G.; GOURISARIA, M. K.; PANDEY, M.; RAUTARAY, S. S. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, Elsevier, v. 38, p. 100285, 2020.

HAYKIN, S. *Redes Neurais - 2ed*. Bookman, 2000. ISBN 9788573077186. Disponível em: <<https://books.google.com.br/books?id=IBp0X5qfyjUC>>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. *Deep Residual Learning for Image Recognition*. 2015.

HUANG, G.; LIU, Z.; MAATEN, L. van der; WEINBERGER, K. Q. *Densely Connected Convolutional Networks*. 2018.

ISOLA, P.; ZHU, J.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. Disponível em: <<http://arxiv.org/abs/1611.07004>>.

ISOLA, P.; ZHU, J.-Y.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 5967–5976.

JABBAR, H.; KHAN, R. Z. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, Res. Publ Singapore, v. 70, n. 10.3850, p. 978–981, 2015.

JANTZEN, J.; NORUP, J.; DOUNIAS, G.; BJERREGAARD, B. Pap-smear benchmark data for pattern classification. *Nature Inspired Smart Information Systems (NiSIS)*, 01 2005.

JANTZEN, J.; NORUP, J.; DOUNIAS, G.; BJERREGAARD, B. Pap-smear benchmark data for pattern classification. In: *Proc. NiSIS 2005*. [S.l.]: NiSIS, 2005. p. 1–9. Nature inspired Smart Information Systems : EU co-ordination action, Nisis 2005 ; Conference date: 01-01-2005.

KOEHRSEN, W. Overfitting vs. underfitting: A complete example. *Towards Data Science*, v. 405, 2018.

LAN, L.; YOU, L.; ZHANG, Z.; FAN, Z.; ZHAO, W.; ZENG, N.; CHEN, Y.; ZHOU, X. Generative adversarial networks and its applications in biomedical informatics. *Frontiers in Public Health*, v. 8, 2020. ISSN 2296-2565. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpubh.2020.00164>>.

LANGE, H.; FERRIS, D. G. Computer-aided-diagnosis (cad) for colposcopy. In: SPIE. *Medical Imaging 2005: Image Processing*. [S.l.], 2005. v. 5747, p. 71–84.

- LI, C.; WAND, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 702–716. ISBN 978-3-319-46487-9.
- LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, v. 33, n. 12, p. 6999–7019, 2022.
- NAQA, I. E.; MURPHY, M. J. *What is machine learning?* [S.l.]: Springer, 2015.
- NIE, D.; TRULLO, R.; LIAN, J.; WANG, L.; PETITJEAN, C.; RUAN, S.; WANG, Q.; SHEN, D. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, v. 65, n. 12, p. 2720–2730, 2018.
- O’SHEA, K.; NASH, R. *An Introduction to Convolutional Neural Networks*. 2015.
- PLISSITI, M. E.; DIMITRAKOPOULOS, P.; SFIKAS, G.; NIKOU, C.; KRIKONI, O.; CHARCHANTI, A. V. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. *2018 25th IEEE International Conference on Image Processing (ICIP)*, p. 3144–3148, 2018.
- RADFORD, A.; METZ, L.; CHINTALA, S. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2016.
- RANDO; SETIAWAN, N. A.; PERMANASARI, A. E.; RULANINGTYAS, R.; SUKSMONO, A. B.; SITANGGANG, I. S. Dcgan-based medical image augmentation to improve elm classification performance. In: *2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*. [S.l.: s.n.], 2022. p. 206–211.
- RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, v. 29, 2005.
- REZENDE, M. T.; GOMES, A.; CARNEIRO, C. M. Cervical cancer: Automation of pap test screening. v. 49, n. 4, p. 559–574, Feb 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/dc.24708#pane-pcw-related>>.
- REZENDE, M. T.; SILVA, R.; BERNARDO, F. d. O.; TOBIAS, A. H.; OLIVEIRA, P. H.; MACHADO, T. M.; COSTA, C. S.; MEDEIROS, F. N.; USHIZIMA, D. M.; CARNEIRO, C. M. et al. Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific Data*, Nature Publishing Group, v. 8, n. 1, p. 1–8, 2021.
- SALEHI, P.; CHALECHALE, A.; TAGHIZADEH, M. Generative adversarial networks (gans): An overview of theoretical model, evaluation metrics, and recent developments. *CoRR*, abs/2005.13178, 2020. Disponível em: <<https://arxiv.org/abs/2005.13178>>.
- SALIMANS, T.; GOODFELLOW, I.; ZAREMBA, W.; CHEUNG, V.; RADFORD, A.; CHEN, X. *Improved Techniques for Training GANs*. 2016.
- SHAFIQ, M.; GU, Z. Deep residual learning for image recognition: A survey. *Applied Sciences*, MDPI, v. 12, n. 18, p. 8972, 2022.
- SILVA, T. S. A short introduction to generative adversarial networks. <https://sthalles.github.io>, 2017. Disponível em: <<https://sthalles.github.io/intro-to-gans/>>.

- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015.
- SOARES, P.; SILVA, J. da. Aplicação de redes neurais artificiais em conjunto com o método vetorial da propagação de feixes na análise de um acoplador direcional baseado em fibra Ótica. *Revista Brasileira de Computação Aplicada*, v. 3, 12 2011.
- SZEGEDY, C.; VANHOUCHE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. *Rethinking the Inception Architecture for Computer Vision*. 2015.
- WANG, Z.; SIMONCELLI, E.; BOVIK, A. Multiscale structural similarity for image quality assessment. In: *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. [S.l.: s.n.], 2003. v. 2, p. 1398–1402 Vol.2.
- XU, Q.; HUANG, G.; YUAN, Y.; GUO, C.; SUN, Y.; WU, F.; WEINBERGER, K. *An empirical study on evaluation metrics of generative adversarial networks*. 2018.
- YU, S.; ZHANG, S.; WANG, B.; DUN, H.; XU, L.; HUANG, X.; SHI, E.; FENG, X. Generative adversarial network based data augmentation to improve cervical cell classification model. *Mathematical Biosciences and Engineering*, v. 18, p. 1740–1753, 02 2021.
- ZAK, J.; GRZESZCZYK, M. K.; PATER, A.; ROSZKOWIAK, L.; SIEMION, K.; KORZYNSKA, A. Cell image augmentation for classification task using gans on pap smear dataset. *Biocybernetics and Biomedical Engineering*, v. 42, n. 3, p. 995–1011, 2022. ISSN 0208-5216. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0208521622000675>>.
- ZHOU, Z.-H. *Machine learning*. [S.l.]: Springer Nature, 2021.