

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

GABRIEL MANTOVANI PINHAL

**Uso de Redes Generativas Adversariais na Criação de Imagens Médicas para
Classificação de COVID-19**

Botucatu

2025

GABRIEL MANTOVANI PINHAL

**Uso de Redes Generativas Adversariais na Criação de Imagens Médicas para
Classificação de COVID-19**

Trabalho de conclusão de curso apresentada
ao Instituto de Biociências de Botucatu -
UNESP, como parte dos requisitos necessários
para obtenção do título de Bacharel em Física
Médica.

Orientador: Prof. Dr. José Luiz Rybarczyk
Filho

Botucatu

2025

P654u

Pinhal, Gabriel Mantovani

Uso de redes generativas adversariais na criação de imagens médicas para classificação de COVID-19 / Gabriel Mantovani Pinhal. -- Botucatu, 2025

37 f. : il., tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Física Médica) -
Universidade Estadual Paulista (UNESP), Instituto de Biociências, Botucatu
Orientador: José Luiz Rybarczyk-Filho

1. Ciências Físicas. 2. Física médica. 3. Inteligência artificial. 4.
Inteligência artificial Aplicações médicas. 5. Redes neurais (Computação). I.
Título.

GABRIEL MANTOVANI PINHAL

USO DE REDES GENERATIVAS ADVERSARIAIS NA CRIAÇÃO DE IMAGENS MÉDICAS PARA CLASSIFICAÇÃO DE COVID-19

Trabalho de Conclusão de Curso,
apresentado a Universidade Estadual
Paulista, como parte das exigências para
a obtenção do título de Bacharel, do curso
de Graduação em Física Médica.

Local, 17 de novembro de 2025.
(data da defesa)

BANCA EXAMINADORA

Prof. Dr. José Luiz Rybarczyk Filho
Departamento de Biofísica e Farmacologia
Instituto de Biociências de Botucatu - UNESP

Prof. Dr. Roberto Morato Fernandez
Departamento de Biofísica e Farmacologia
Instituto de Biociências de Botucatu - UNESP

Prof. Dr. Joel Mesa Hormaza
Departamento de Biofísica e Farmacologia
Instituto de Biociências de Botucatu - UNESP

Agradecimentos

- À Universidade Estadual Paulista (UNESP) pela excelente formação acadêmica e pelo proporcionamento de um ambiente de ensino e pesquisa que foi decisivo para meu desenvolvimento profissional e pessoal.
- Esta pesquisa tornou-se possível graças aos recursos computacionais disponibilizados pelo Núcleo de Computação Científica (NCC/GridUNESP) da Universidade Estadual Paulista (UNESP).
- Ao meu Orientador, que me acolheu e guiou e foi além do necessário para me ajudar em qualquer dúvida que tinha.
- Aos meus pais, que me deram suporte, especialmente durante a pandemia, para eu não desistir do curso e continuar firme nos meus estudos.
- Aos meus irmãos, que me incentivaram à fazer este curso quando ainda estava em dúvida sobre qual curso fazer.

“Independentemente das circunstâncias, devemos ser sempre humildes, recatados e despidos de orgulho.”

(Dalai Lama)

Resumo

A pandemia de COVID-19 resultou em um aumento significativo na demanda por exames de imagem, incluindo radiografias de tórax, amplamente utilizadas na detecção de pneumonia, uma manifestação comum da doença. O desenvolvimento de técnicas de aprendizado profundo, especialmente redes neurais convolucionais, possibilitou avanços expressivos na análise automatizada dessas imagens. Entretanto, a escassez e a heterogeneidade de bancos de dados médicos dificultam o treinamento robusto de modelos de classificação.

Neste estudo, utilizamos a arquitetura UBNET treinada exclusivamente com imagens reais, e, em seguida, empregamos uma *Deep Convolutional Generative Adversarial Network* (DCGAN) para gerar imagens sintéticas de tórax. O objetivo foi avaliar o impacto da inclusão dessas imagens sintéticas no desempenho do modelo, comparando-o ao modelo treinado somente com dados reais. Para isso, utilizamos métricas de classificação (acurácia, precisão, recall e F1-Score), bem como métricas específicas para avaliação de GANs, como *Fréchet Inception Distance* (FID) e *Kernel Inception Distance* (KID), além de matrizes de confusão.

Os resultados evidenciam que, embora seja possível gerar imagens sintéticas visualmente plausíveis, sua utilização direta no treinamento pode degradar o desempenho da UBNET quando comparado ao uso exclusivo de dados reais. Entretanto, a análise também indica que a viabilidade do uso de dados sintéticos depende fortemente da qualidade das imagens geradas e da capacidade da GAN em preservar características diagnósticas relevantes. Dessa forma, conclui-se que modelos geradores mais sofisticados e estratégias de curadoria de dados sintéticos podem tornar essa abordagem promissora em cenários de escassez de dados médicos. **Palavras-chaves:** Rede Adversária Generativa, GAN, imagens médicas,

síntese de imagens médicas, Rede Adversária Generativa Convolutional Profunda, DCGAN, aprendizado profundo, aprendizado de máquina, Redes Neurais Convolucionais, CNN.

Abstract

The COVID-19 pandemic led to a significant increase in the demand for medical imaging, particularly chest X-rays, which are widely used for detecting pneumonia, a common manifestation of the disease. Advances in deep learning, especially convolutional neural networks, have enabled remarkable progress in the automated analysis of such images. However, the scarcity and heterogeneity of medical databases hinder the robust training of classification models.

In this study, we employed the UBNET architecture trained exclusively on real chest X-ray images and subsequently used a Deep Convolutional Generative Adversarial Network (DCGAN) to generate synthetic chest images. The objective was to evaluate the impact of including these synthetic images on model performance, comparing it with the model trained solely on real data. To this end, we used classification metrics (accuracy, precision, recall, and F1-score), as well as GAN-specific evaluation metrics such as Fréchet Inception Distance (FID) and Kernel Inception Distance (KID), in addition to confusion matrices.

The results show that, although it is possible to generate visually plausible synthetic images, their direct use in training may degrade UBNET performance compared to models trained exclusively with real data. Nevertheless, the analysis also indicates that the feasibility of using synthetic data strongly depends on the quality of the generated images and the GAN's ability to preserve diagnostically relevant features. Therefore, it is concluded that more sophisticated generative models and improved curation strategies for synthetic data may make this approach promising in scenarios of limited medical data availability.

Keywords: Generative Adversarial Network, GAN, medical images, medical image synthesis, Deep Convolutional Generative Adversarial Network, DCGAN, deep learning, machine learning, Convolutional Neural Networks, CNN.

Lista de figuras

Figura 1 – Os ossos da mão da esposa de Wilhelm Röntgen, com um anel em um dos dedos, vistos através de um raio-X. Retirado de (MARKEL, 2012).	11
Figura 2 – Esquema ilustrativo de um tubo de raio-x. Fonte: (SILVA, 2016).	12
Figura 3 – Distribuição de energia para um espectro de raio-X produzidos pelo freamento Bremsstrahlung, para um nível de energia de 90 kV.	13
Figura 4 – Raio-X de tórax mostrando uma área em formato de cunho do espaço aéreo pulmonar, característica de pneumonia bacteriana, no pulmão direito. <i>From the (medical) gallery of Mikael Häggström, M.D.. Last updated: 2025-07-28. Licensing (unless other license is given for image): Creative Commons CC0 1.0 Universal Public Domain Dedication, used with permission.</i>	14
Figura 5 – comparação entre raios-X de tórax de pulmões saudáveis, pulmões com pneumonia, e pulmões com pneumonia causada por Covid-18. Fonte: (ULLAH et al., 2023).	15
Figura 6 – Esquema ilustrando a arquitetura de uma GAN.	16
Figura 7 – Arquitetura do Gerador, criada utilizando PlotNeuralNet. Fonte: próprio autor.	21
Figura 8 – Arquitetura do Discriminador, criada utilizando PlotNeuralNet. Fonte: próprio autor.	22
Figura 9 – Arquitetura UBNNet. Fonte: Widodo, 2022.	23
Figura 10 – Imagens sintéticas no início do treinamento, época 3.	25
Figura 11 – Imagens sintéticas durante no início do treinamento, época 7.	26
Figura 12 – Imagens sintéticas na metade do treinamento, época 58.	27
Figura 13 – Imagens sintéticas na metade do treinamento, época 62.	28
Figura 14 – Imagens sintéticas no final do treinamento, época 100.	29
Figura 15 – Imagens da classe normal geradas pelo modelo.	29
Figura 16 – Imagens da classe pneumonia geradas pelo modelo.	30
Figura 17 – Histórico de acurácia e de perda ao longo do treinamento.	30
Figura 18 – Matriz de confusão para imagens reais.	31
Figura 19 – Matriz de confusão para imagens sintéticas.	31

Lista de tabelas

Tabela 1 – Relatório de Classificação (Imagens reais)	24
Tabela 2 – Relatório de Classificação (Imagens sintéticas)	30
Tabela 3 – Valores de FID e KID do modelo DDGAN	30

Lista de abreviaturas e siglas

GAN	Generative Adversarial Networks
DCGAN	Deep Convolutional Generative Adversarial Networks
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
FID	Fréchet Inception Distance
KID	Kernel Inception Distance
MMD	Maximum Mean Discrepancy

Sumário

1	INTRODUÇÃO	10
1.1	<i>Pandemia</i>	10
1.2	<i>Raio - X</i>	10
1.2.1	Produção de Raio - X	12
1.2.2	Imagens médicas	13
1.2.3	Raio-X de tórax	14
1.3	<i>Aprendizado profundo - Generativo</i>	15
1.3.1	Colapso de modelo	16
1.3.2	Imagens médicas sintéticas	17
1.4	<i>Justificativa</i>	18
2	OBJETIVOS	19
3	MATERIAIS E MÉTODOS	20
3.1	<i>Base de Dados</i>	20
3.2	<i>Redes Adversariais Generativas</i>	20
3.2.1	Modelo DCGAN	21
3.2.2	Modelo UBNet	22
4	RESULTADOS	24
4.1	<i>O modelo DCGAN</i>	24
4.1.1	Modelo UBNet	24
5	DISCUSSÃO	32
6	CONCLUSÃO	34
	Referências¹	35

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 INTRODUÇÃO

1.1 *Pandemia*

A pandemia colocou uma enorme pressão nos sistemas hospitalares no mundo inteiro, sobrecarregando muitos hospitais que não conseguiram lidar com o aumento da carga de pacientes afetados pela Covid-19, inicialmente detectada em um grupo de pessoas com pneumonia na região de Wuhan, China (KIRU et al., 2022). Como o SARS-CoV-2 é altamente contagioso e infeccioso, uma grande quantidade de pessoas ficaram doentes e necessitaram de atendimento hospitalar. Esse influxo de pacientes excedeu a capacidade dos hospitais para tratá-los, causando falta de atendimento médico e escassez de recursos, o que agravou ainda mais o problema de superlotação.

Foram criados testes para detecção de Covid-19, mas muitos apresentavam uma alta taxa de falsos positivos ou levavam muito tempo para fornecer o resultado. Dessa forma, tornou-se necessária a adoção de um teste rápido e confiável, sendo esse o Raio-X do tórax. Porém, é necessário um radiologista para examinar o Raio-X e indicar se há ou não infecção por Covid-19.

Com os avanços em *Deep Learning* e Redes Neurais, tornou-se possível criar modelos capazes de realizar essa tarefa automaticamente, de forma rápida e com alta acurácia, com maior atenção sendo dada às GANs (*Generative Adversarial Networks*). GANs são conhecidas por sua habilidade de produzir imagens realistas e lidar com mudanças de domínio (HAN et al., 2018). O uso de GANs em imageamento médico está revolucionando acurácia diagnóstica e o aprimoramento de imagens, ao gerar imagens médicas de alta qualidade a partir de *datasets* limitados (ISLAM et al., 2024).

1.2 *Raio - X*

O Raios-X foram descobertos e estudados pelo professor de física alemão Wilhelm Röntgen em 8 de novembro de 1895. Röntgen denominou essa radiação de Raio-X por se tratar de um tipo de devido ser um tipo de radiação desconhecida até então. O potencial médico dessa descoberta foi evidenciado quando o pesquisador obteve a imagem da mão de sua esposa, visto na figura 1, em uma placa fotográfica exposta à raios-X, tornando-se o primeiro registro de imagem radiográfica ((MARKEL, 2012)).



Figura 1 – Os ossos da mão da esposa de Wilhelm Röntgen, com um anel em um dos dedos, vistos através de um raio-X. Retirado de (MARKEL, 2012).

Os raios-X possuem grande importância na área médica, sendo amplamente utilizados no diagnóstico por imagem e no tratamento de tumores. Isso ocorre porque os fótons de raio-X apresentam energia suficiente para atravessar diferentes materiais, interagindo com os tecidos e depositando parte de sua energia ao longo do trajeto. Na formação de imagens, os fótons são atenuados de maneira distinta pelos diversos tecidos do corpo humano. Por exemplo, os ossos contêm cálcio, elemento de número atômico elevado, que absorve mais fótons do que os tecidos moles, resultando em maior contraste na imagem radiográfica (OKUNO; YOSHIMURA, 2016).

No tratamento de tumores, essa mesma interação entre a radiação e os tecidos é explorada de forma controlada, concentrando a dose de radiação na região alvo para induzir morte celular seletiva.

A atenuação dos raios-X ocorre principalmente pelos processos de absorção e espalhamento dos fótons. O efeito fotoelétrico predomina em faixas de baixa energia (inferiores a 26 keV), sendo mais comum em tecidos com elevado número atômico. Já

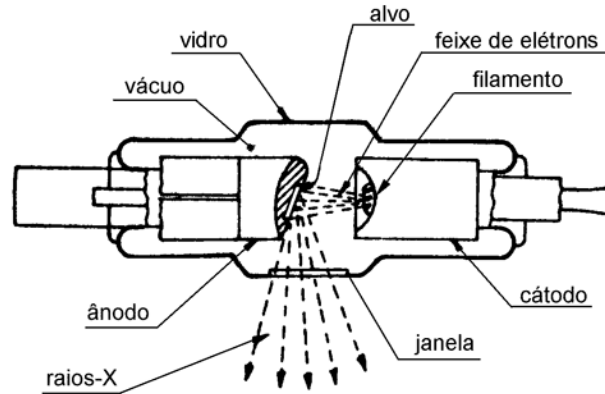


Figura 2 – Esquema ilustrativo de um tubo de raio-x. Fonte: (SILVA, 2016).

o efeito Compton torna-se mais significativo para fótons de maior energia e é típico em tecidos moles, que apresentam baixo número atômico Z . O espalhamento Rayleigh, por sua vez, contribui de forma reduzida à atenuação total, representando cerca de 10% das interações em mamografias e aproximadamente 5% das interações em radiografias de tórax (BUSHBERG; BOONE, 2011).

1.2.1 Produção de Raio - X

Um feixe de raio-X é produzido quando elétrons de energia suficientemente alta interagem com a matéria, convertendo parte de sua energia em radiação eletromagnética. Um exemplo disso é o tubo de raio-X comumente utilizado na área do Radiodiagnóstico. No tubo de raio-X, os elétrons, termicamente emitidos no cátodo, são acelerados em direção a um alvo metálico por meio de um diferencial de potencial, denominado de ânodo, conforme ilustrado na figura 2.

O feixe produzido contém dois tipos de raio-X: um chamado de espectro contínuo, criado pelo freamento (*Bremsstrahlung*) dos elétrons que interagem com o material do ânodo; e o outro, chamado de raio-X característico, criado quando um elétron colide com um elétron de uma camada interna do átomo, ejetando-o e gerando uma vacância, que é preenchida por um elétron de uma camada mais externa, liberando um fóton de raio-X com energia característica para cada elemento (EISBERG; RESNICK, 1985).

O espectro da radiação de *Bremsstrahlung* é bastante amplo, com a energia dos raios-X variando desde a ordem de elétron-volts (eV) até a energia máxima dos elétrons acelerados que atingem o ânodo, sendo a energia média do espectro aproximadamente

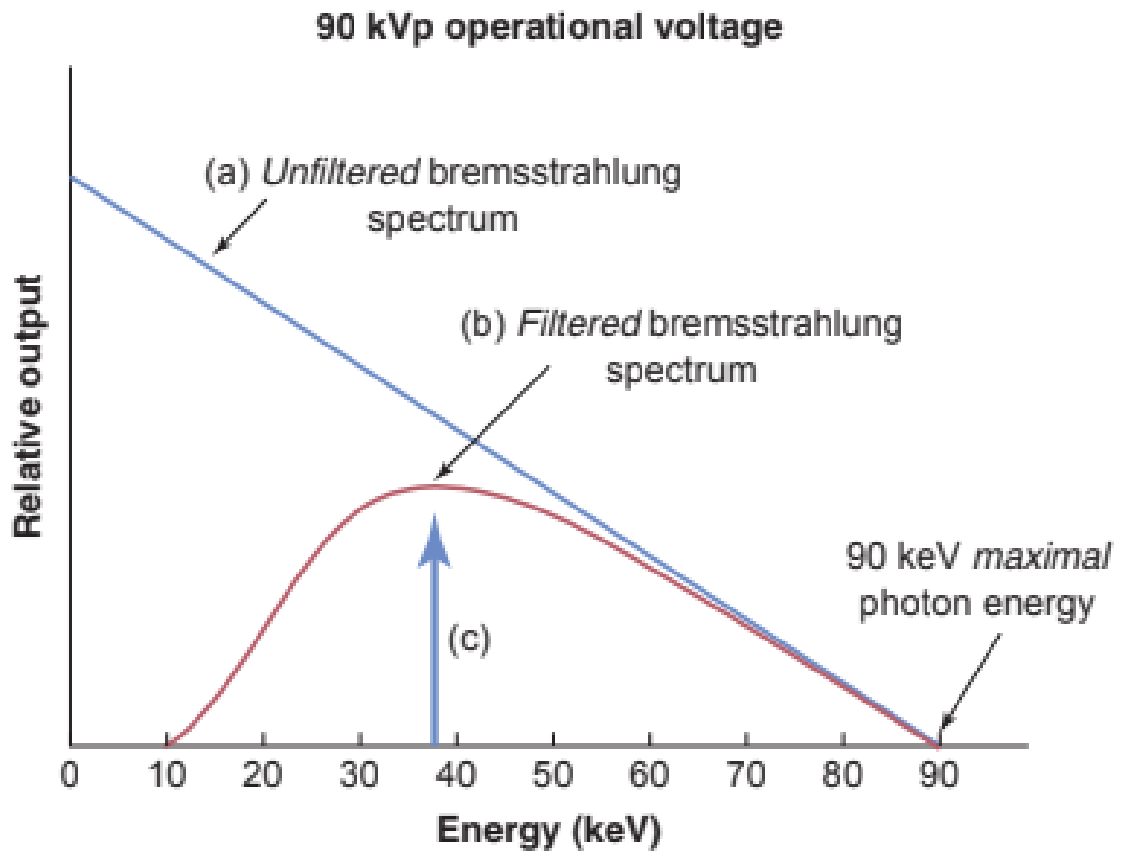


Figura 3 – Distribuição de energia para um espectro de raio-X produzidos pelo freamento Bremsstrahlung, para um nível de energia de 90 kV.

metade ou um terço da energia máxima. Isso ocorre devido à natureza probabilística da interação entre os elétrons e os núcleos dos átomos alvo.

No tubo de raio-X, são inseridos filtros que atenuam o feixe, removendo os fótons de baixa energia, para aumentar a energia média do feixe e torná-lo mais adequado ao imageamento com baixa dose.

Na figura 3 está ilustrado um exemplo da distribuição de energia *Bremsstrahlung* para um diferencial de potencial de 90 kV no tubo (BUSHBERG; BOONE, 2011).

1.2.2 Imagens médicas

Os exames radiológicos são obtidos a partir da detecção dos fótons que atravessam o paciente e formam uma imagem em um detector sensível à radiação. O contraste entre as

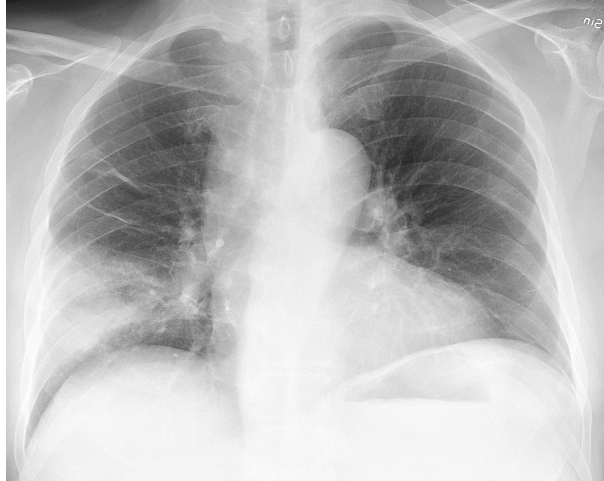


Figura 4 – Raio-X de tórax mostrando uma área em formato de cunho do espaço aéreo pulmonar, característica de pneumonia bacteriana, no pulmão direito. *From the (medical) gallery of Mikael Häggström, M.D.*. Last updated: 2025-07-28. Licensing (*unless other license is given for image*): Creative Commons CC0 1.0 Universal Public Domain Dedication, used with permission.

diferentes estruturas e tecidos é obtido pela absorção e/ou atenuação dos fótons, determinada pelas diferenças de densidade e número atômico dos tecidos(OKUNO; YOSHIMURA, 2016).

1.2.3 Raio-X de tórax

A radiografia do tórax é utilizada para diagnosticar doenças que afetam o tórax, os órgãos contidos nele, e estruturas próximas dele. As estruturas imageadas são os pulmões, o coração, as cavidades presentes na caixa torácica, e os vasos sanguíneos como a aorta e as veias cavas. No caso dos pulmões, pneumonia é comumente diagnosticada por esse método. Em um raio-X de tórax, pneumonia é detectada dependendo do seu tipo: em pneumonia bacteriana observa-se a consolidação pulmonar de um lobo do pulmão, enquanto na pneumonia viral há mais variedade em como se manifesta, podendo parecer regulares, parecer hiperinfladas, ter áreas bilaterais irregulares, ou apresentar semelhanças à pneumonia bacteriana com consolidação lobar. Consolidação se refere ao preenchimento do espaço alveolar por líquidos como pus ou sangue, criando uma região opaca. As consolidações são radiopacas e, portanto, aparecem mais esbranquiçadas na imagem. As figuras 4 e 5 mostram exemplos de pneumonia bacteriana e viral numa imagem de raio-X de tórax.

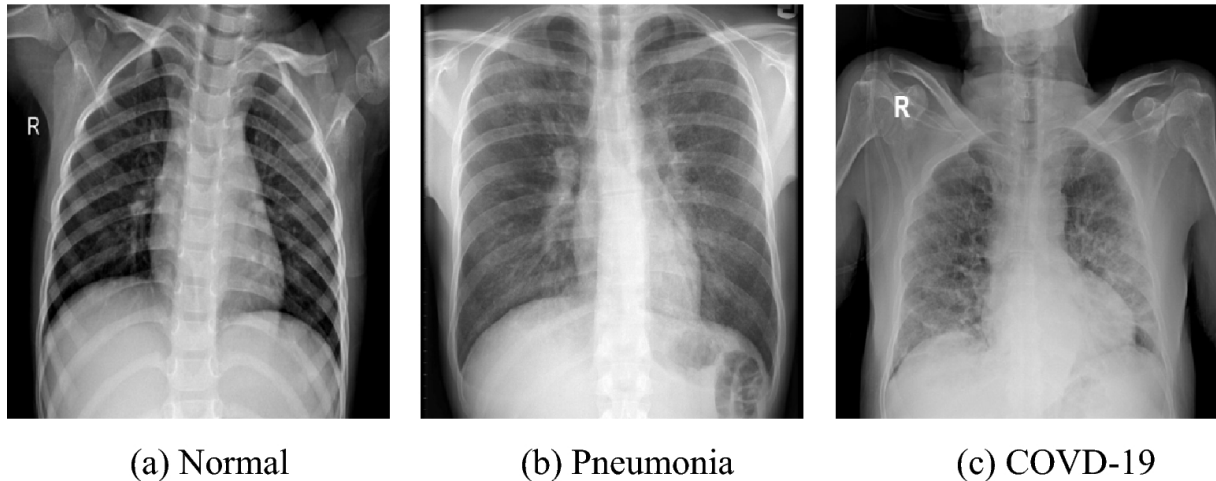


Figura 5 – comparação entre raios-X de tórax de pulmões saudáveis, pulmões com pneumonia, e pulmões com pneumonia causada por Covid-18. Fonte: (ULLAH et al., 2023).

1.3 Aprendizagem profunda - Generativo

O Aprendizado de Máquina Profundo, ou *Deep Learning*, utiliza várias camadas de redes neurais para realizar várias tarefas, tais como, classificação, regressão, ou extração de características. Alguns tipos de arquiteturas de *deep learning* são mais comuns, como a Rede Neural Convolutiva (CNN), a Rede Totalmente Conectada, a Rede Generativa Adversarial (GAN), e Redes Neurais Recorrentes (RNN). Essas arquiteturas são bastante utilizadas nas áreas de visão computacional, reconhecimento de fala, processamento de linguagem natural, *design* de fármacos, e análise de imagens médicas. As CNNs necessitam de grandes quantidades de dados rotulados para aprender a tarefa de detecção de anomalias em imagens (MOTAMED; ROGALLA; KHALVATI, 2021b). Nesse sentido, as GANs tornam-se uma alternativa interessante porque são ótimas em gerar imagens. GANs consistem de duas partes: o gerador e o discriminador. O gerador produz imagens a partir de uma entrada de ruído aleatório, enquanto o discriminador avalia se uma imagem é real ou sintética (MUNAWAR et al., 2020). O funcionamento de uma GAN está ilustrado na figura 6. As duas redes são treinadas de forma competitiva: o gerador busca criar imagens cada vez mais realistas, ao mesmo tempo em que o discriminador aprende a diferenciá-las das imagens reais. Esse processo iterativo prossegue até que o gerador produza imagens indistinguíveis para o discriminador (SINGH; RAZA, 2020). Essa capacidade de gerar

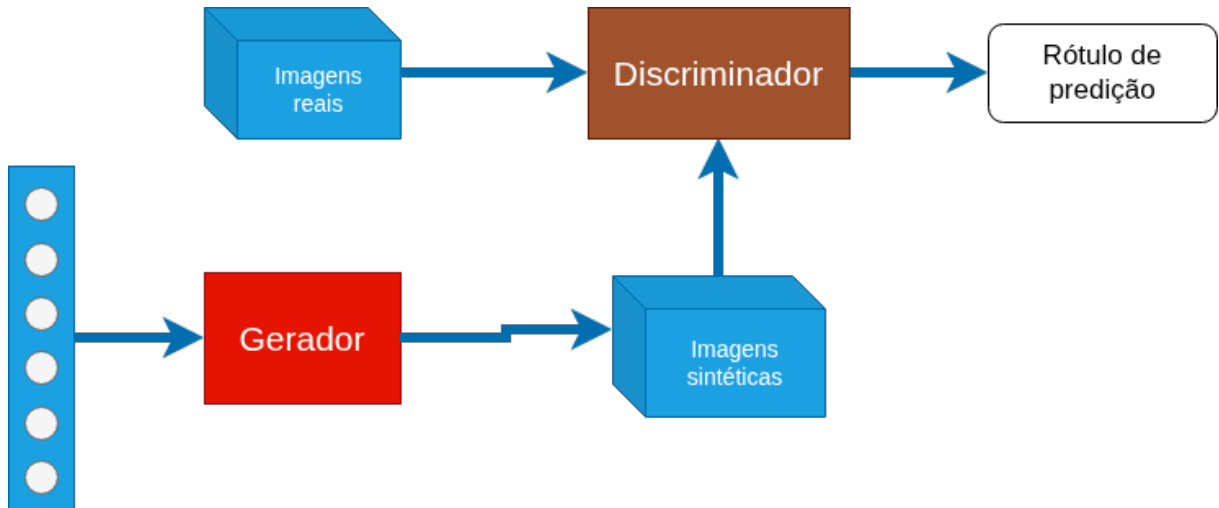


Figura 6 – Esquema ilustrando a arquitetura de uma GAN.

dados sintéticos úteis torna as GANs especialmente valiosas em contextos hospitalares, onde a disponibilidade de imagens rotuladas frequentemente é limitada.

1.3.1 Colapso de modelo

O colapso de modelo é um processo degenerativo que pode ocorrer em modelos generativos quando estes passam a ser treinados repetidamente com dados sintéticos produzidos por versões anteriores. Nesse cenário, pequenos erros introduzidos a cada geração são cumulativamente propagados, degradando progressivamente a qualidade do modelo (SHUMAILOV et al., 2024). No colapso inicial, observa-se a perda da capacidade de representar regiões extremas da distribuição real. Já no colapso tardio, o modelo converge para uma distribuição artificial, de baixa variabilidade e pouca semelhança com a distribuição original.

Esse fenômeno está relacionado às limitações inerentes aos modelos generativos. O erro de aproximação estatística surge devido ao número finito de amostras disponíveis; o erro de expressividade funcional decorre da capacidade limitada da rede de representar, de forma precisa, a distribuição real; e o erro de aproximação funcional é resultado de restrições no processo de otimização, viés dos dados de treinamento ou formulação do objetivo do modelo (SHUMAILOV et al., 2023).

Nas GANs, um caso particular desse problema é o colapso de modo (*mode collapse*), em que o gerador passa a produzir imagens muito semelhantes entre si, reduzindo significa-

tivamente a diversidade do conjunto gerado. Quando isso ocorre, o discriminador pode não perceber essa perda de variedade, resultando em pior desempenho e menor aplicabilidade prática do modelo (PETERSON, 2025).

1.3.2 Imagens médicas sintéticas

O uso de imagens médicas sintéticas é uma excelente solução para ampliar os conjuntos de dados para treinamento de redes neurais. Porém, seu uso apresenta problemas técnicos que afetam sua confiabilidade e segurança na prática clínica. Existem poucos bancos de dados de imagens médicas de larga escala publicamente disponíveis (WANG et al., 2018), agravado pela dificuldade de coletar imagens médicas, pois a coleta e rotulagem de dados médicos enfrentam problemas de privacidade de dados como a necessidade de explicações especializadas (YADAV; JADHAV, 2019). O compartilhamento de dados médicos é complicado devido à questões éticas, como privacidade do paciente, sendo necessário realizar desidentificação dos dados para possibilitar o acesso aos mesmos. Além disso, os centros de dados são descentralizados e a interoperabilidade limitada entre sistemas também contribui para esse problema (KOETZIER et al., 2024).

Modelos de redes neurais treinados exclusivamente em imagens sintéticas apresentam desempenho reduzido em cenários clínicos na vida real, devido a imagens reais possuírem ruídos, artefatos, e outras imperfeições causadas pelos instrumentos de imageamento. Dados sintéticos possuem maior dificuldade em replicar essas peculiaridades com precisão.

Além disso, dados sintéticos também carecem de autenticidade clínica, uma vez que, apesar de métricas computacionais indicarem que são altamente realistas e de alta qualidade, especialistas são capazes de identificar facilmente quais imagens são sintéticas e quais são reais.

Modelos generativos podem introduzir artefatos não presentes em exames reais ou que não são inerentes ao tipo de exame realizado, como consequência do viés adquirido durante o treinamento (KOÇAK et al., 2025). Esses modelos também podem falhar em representar casos de borda (edge cases), justamente aqueles em que a atuação da rede neural seria mais necessária.

1.4 *Justificativa*

A radiografia de tórax é um dos exames mais utilizados no diagnóstico de doenças pulmonares, como pneumonia e Covid-19, devido ao seu baixo custo, rapidez e ampla disponibilidade na prática clínica. Entretanto, o processo de interpretação dessas imagens exige profissionais especializados, podendo ser influenciado por variáveis como experiência do avaliador, carga de trabalho e condições de atendimento. Nesse contexto, modelos de aprendizado profundo têm sido explorados como ferramentas auxiliares para apoiar médicos na detecção de anomalias, aumentando a precisão diagnóstica e reduzindo o tempo de análise. Contudo, o treinamento eficaz desses modelos depende da disponibilidade de grandes conjuntos de dados devidamente rotulados, o que representa um desafio significativo, uma vez que imagens médicas são sensíveis, envolvem questões éticas e legais, e demandam curadoria especializada.

Para contornar a escassez de dados, o uso de imagens médicas sintéticas geradas por Redes Generativas Adversariais (GANs) surge como uma alternativa promissora, permitindo expandir bancos de imagens e aumentar a variabilidade dos dados de treinamento. No entanto, modelos treinados exclusivamente com imagens sintéticas tendem a apresentar baixo desempenho em ambientes clínicos reais, uma vez que essas imagens geralmente não reproduzem com fidelidade os ruídos, artefatos e características sutis presentes em exames adquiridos por equipamentos médicos. Além disso, modelos generativos podem introduzir padrões artificiais ou distorções que podem levar a interpretações equivocadas, especialmente em casos de borda (edge cases), onde o suporte da inteligência artificial é mais necessário. Assim, torna-se fundamental investigar o impacto da inclusão de imagens sintéticas no treinamento de classificadores, avaliando se essas imagens contribuem efetivamente para melhorar o desempenho do modelo ou se podem comprometer sua confiabilidade clínica.

2 OBJETIVOS

Objetivo geral

O objetivo geral deste trabalho é verificar se imagens sintéticas podem ser usadas no treinamento de CNNs para mitigar problemas oriundos de baixo número de dados disponíveis.

Objetivos específicos

- Treinar um modelo DCGAN para gerar imagens sintéticas de alta qualidade.
- Treinar o modelo da UBNNet utilizando imagens reais e avaliar o desempenho do modelo com imagens sintéticas, e comparar as métricas de avaliação entre os dois modelos.

3 MATERIAIS E MÉTODOS

3.1 Base de Dados

A base de dados utilizada foi disponibilizada pelo artigo "UBNet: Deep learning-based approach for automatic X-ray image detection of pneumonia and COVID-19 patients". A base contém dois conjuntos de imagens, um de teste e outro de treino, com o primeiro separado em 234 imagens normais e 390 com pneumonia, e o último separado em 1349 normais e 3884 com pneumonia, totalizando 5860 imagens de raio-X torácico.

3.2 Redes Adversariais Generativas

Redes Adversariais Generativas (GAN) são compostas por duas redes neurais, uma chamada de Gerador (G) e a outra chamada de Discriminador (D). O Gerador produz imagens sintéticas a partir de ruído aleatório, visando replicar a autenticidade de imagens reais, enquanto que o Discriminador avalia imagens e visa diferenciá-las em reais ou sintéticas (LIU et al., 2022). As duas redes competem entre si, o Gerador com o objetivo de gerar imagens sintéticas que o Discriminador é incapaz de dizer que são sintéticas, e o Discriminador tem o objetivo de ser capaz de distinguir entre imagens reais e imagens geradas pelo Gerador. O Discriminador visa maximizar a probabilidade de classificar corretamente os exemplos fornecidos ambos pelos dados de treinamento e os exemplos fornecidos pelo Gerador, enquanto que o último visa minimizar essa probabilidade (MOTAMED; ROGALLA; KHALVATI, 2021a). Deep Convolutional GANs (DCGAN) é uma variante da GAN que mantém a arquitetura geral do modelo porém fazem o uso de *convolutional neural networks* (CNN) para o gerador e o discriminador, integrando métodos como "*batch normalization*", que normaliza os outputs para estabilizar o treinamento, e uso de *leakyReLU activation* ao invés de *ReLU activation*, para evitar a ocorrência de "neurônios mortos", para melhorar a estabilidade do treinamento e a qualidade das imagens de output (AKHIL et al., 2024) (RADFORD; METZ; CHINTALA, 2015), além de melhorar significativamente a otimização de redes neurais e ser altamente eficiente em DCGANs (SRIVASTAV; BAJPAI; SRIVASTAVA, 2021). Os discriminadores e os geradores de DCGAN são projetados para alcançar uma performance maior no quesito de geração de imagens (VJ, 2021).

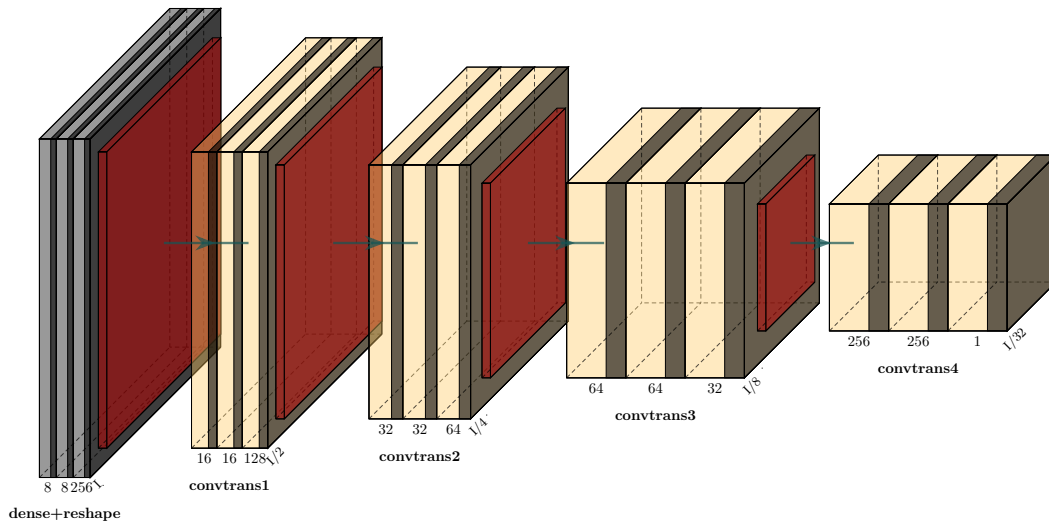


Figura 7 – Arquitetura do Gerador, criada utilizando PlotNeuralNet. Fonte: próprio autor.

3.2.1 Modelo DCGAN

O modelo DCGAN consiste de dois modelos, um modelo gerador e um modelo discriminador. O modelo gerador faz uso de convoluções de passo fracionário que o possibilita aprender a partir de um vetor de ruído de entrada (*input*) aleatório através do seu próprio *upsampling* para gerar uma imagem a partir dele (SINGH; RAZA, 2020), e consiste em quatro camadas de convolução tranposta, com a três primeiras dobrando o tamanho da imagem gerada a partir do input do vetor de ruído desde 8×8 pixels até 64×64 pixels, que é quadruplicado pela última camada de convolução para atingir o tamanho desejado de 256×256 pixels. Para todas as camadas é utilizado *batchnormalization* e função de ativação *LeakyReLU*. O modelo discriminador possui uma arquitetura mais tradicional para CNNs, utilizando três camadas de convolução, cada uma fazendo *downsampling*, reduzindo o tamanho da imagem pela metade, e uma camada totalmente conexa (Densa) que produz um número escalar que representa a decisão do modelo sobre a veracidade da imagem, com valores positivos sendo imagens reais e valores negativos sendo imagens sintéticas. As arquiteturas do gerador e do discriminador estão ilustradas pelas figuras 7 e 8.

Métricas comumente utilizadas para avaliar GANs são *Fréchet Inception Distance* (FID) e *Kernel Inception Distance* (KID), o primeiro avaliando a semelhança estatística

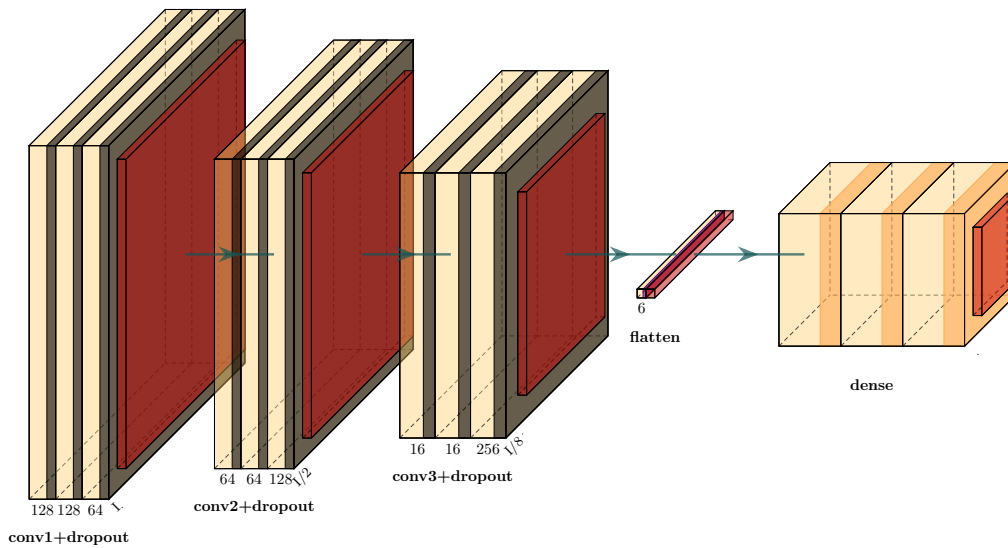


Figura 8 – Arquitetura do Discriminador, criada utilizando PlotNeuralNet. Fonte: próprio autor.

entre a distribuição de características das imagens reais e a distribuição de características das imagens sintéticas, e o último utilizando distância de kernel (*Maximum Mean Discrepancy*, MMD) ao invés da distância de Fréchet para comparar as distribuições de características das imagens. Quanto menor o valor de FID e o valor de KID maior é a variedade de amostras e, portanto, mais visualmente semelhantes são as imagens sintéticas em relação as imagens reais (HO; SALIMANS, 2022).

3.2.2 Modelo UBNNet

O modelo UBNNet, como descrito no artigo (WIDODO et al., 2022), tem a arquitetura típica de CNNs, com sete camadas de convolução, utilizando operações de *max pooling* para fazer o *downsampling* para extrair as informações mais importantes da imagem, e três camadas totalmente conexas, com todas as camadas utilizando função de ativação ReLU com exceção da última camada Densa, que utiliza a função de ativação *softmax*. Essa função é utilizada na última camada pois é ela que determina a tendência de *output* de cada classe, com o valor sendo entre zero e um.

Este modelo serve como base para verificar a viabilidade de imagens sintéticas no treinamento de modelos de classificação. Sua arquitetura está ilustrada na figura 9.

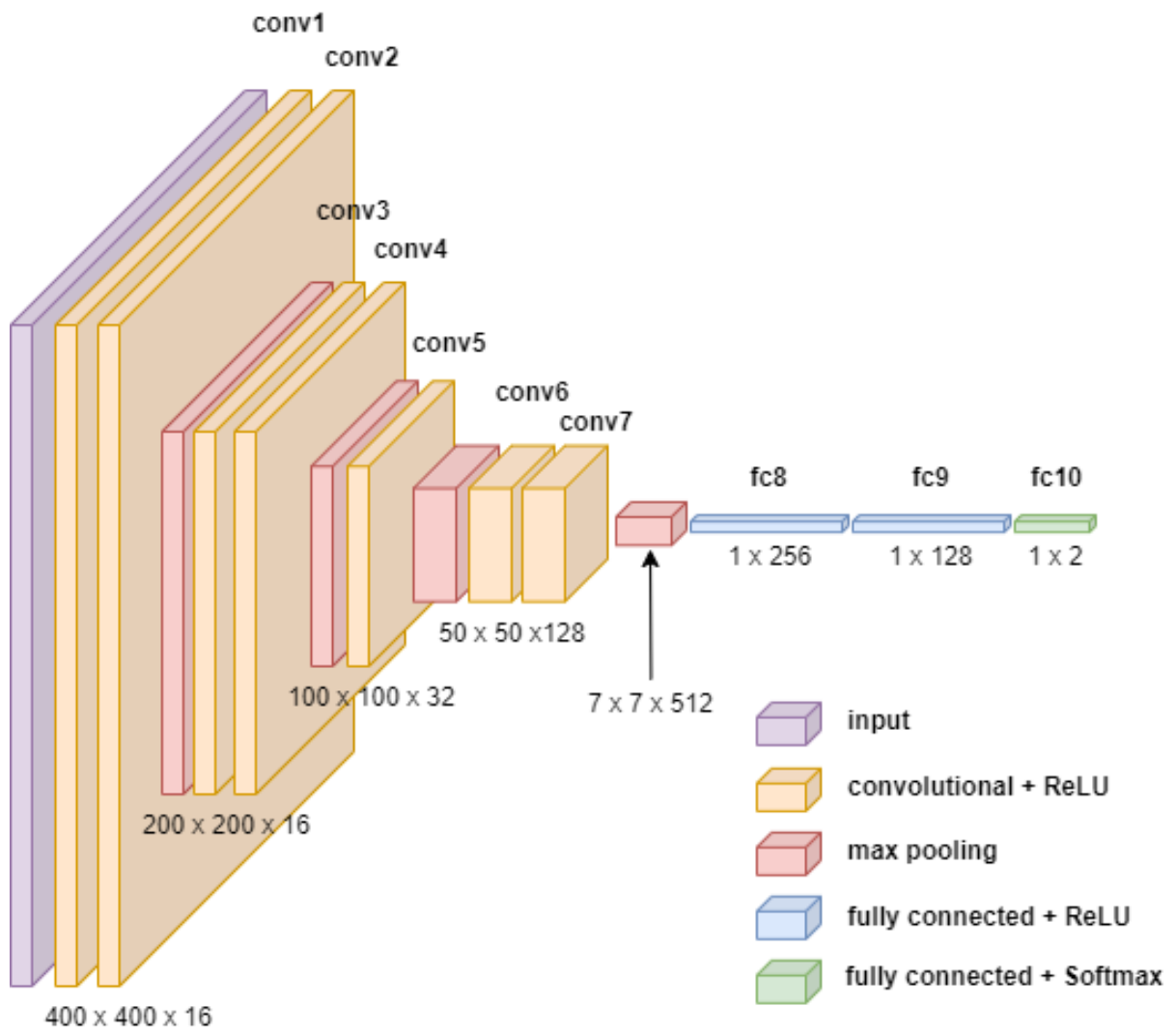


Figura 9 – Arquitetura UBNet. Fonte: Widodo, 2022.

4 RESULTADOS

4.1 O modelo DCGAN

Depois de cada época de treinamento, o modelo gerava 32 imagens para servir como monitoramento do processo de treinamento e visualizar o aprimoramento do gerador em criar imagens sintéticas de um raio-X de tórax. Alguns destes grupos de imagens criadas são mostradas nas figuras abaixo, revelando o progresso do modelo. No início, as imagens continham apenas a estrutura geral de um raio-X de tórax, como a área mais clara no centro sendo a coluna vertebral e a leve presença de listras sendo as costelas, e, conforme o treinamento progredia, o gerador aprendeu a criar imagens que se assemelham a um raio-X torácico, como mostrado nas figuras 10, 11, 12, 13, e 14. Após o término do treinamento, o modelo gerou mil imagens da classe Normal e mil imagens da classe Pneumonia, para uso no treinamento do outro modelo. As figuras 15 e 16 mostram alguns destes exemplos.

4.1.1 Modelo UBNNet

Foram realizados dois testes de avaliação para o modelo, um teste quando o modelo foi treinado apenas por imagens reais, e outro quando foram incorporadas as imagens sintéticas no treinamento do modelo. Os testes de avaliação são mostrados a seguir, organizados nas tabelas 1, 2, e 3, e nas figuras 17, 18, e 19.

Tabela 1 – Relatório de Classificação (Imagens reais)

Classe	Precisão (precision)	Revocação (recall)	F1-Score	Suporte (n)
NORMAL	0.9813	0.4487	0.6158	234
PNEUMONIA	0.7505	0.9949	0.8556	390
Acurácia global				0.7901
Médias macro:	Precisão: 0.8659	Revocação: 0.7218	F1: 0.7357	
Médias ponderadas:	Precisão: 0.8370	Revocação: 0.7901	F1: 0.7657	

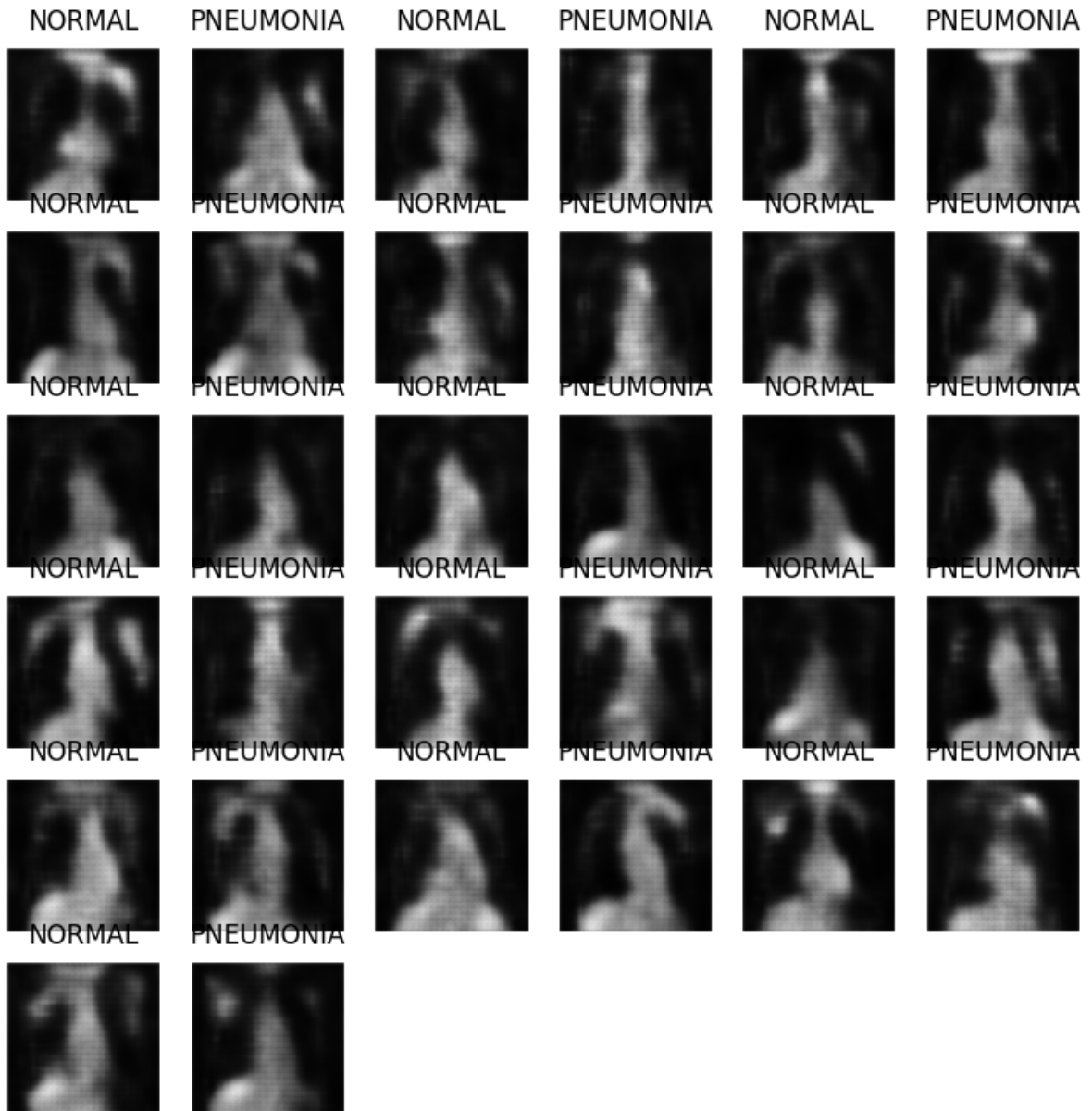


Figura 10 – Imagens sintéticas no início do treinamento, época 3.

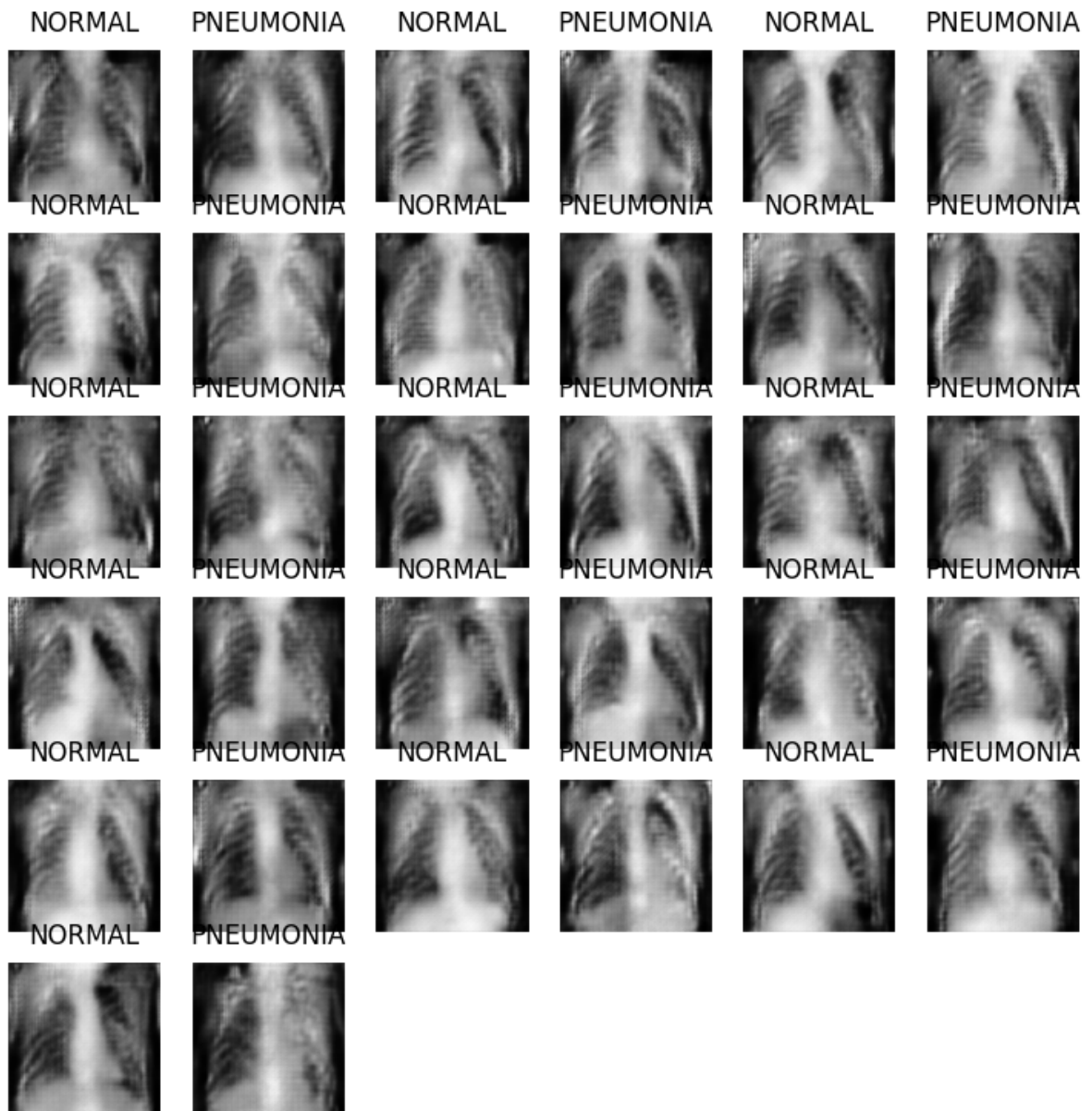


Figura 11 – Imagens sintéticas durante no início do treinamento, época 7.

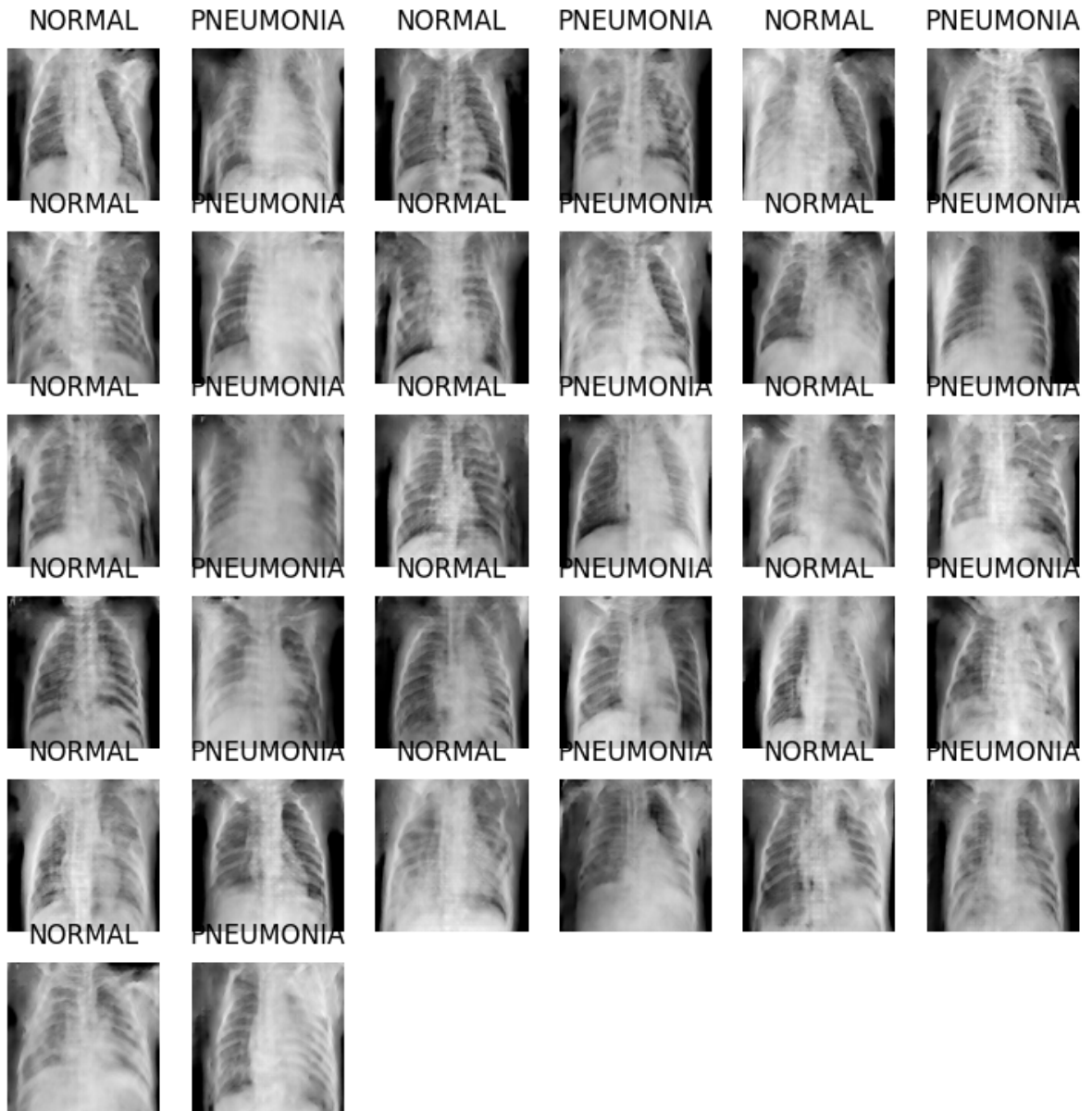


Figura 12 – Imagens sintéticas na metade do treinamento, época 58.

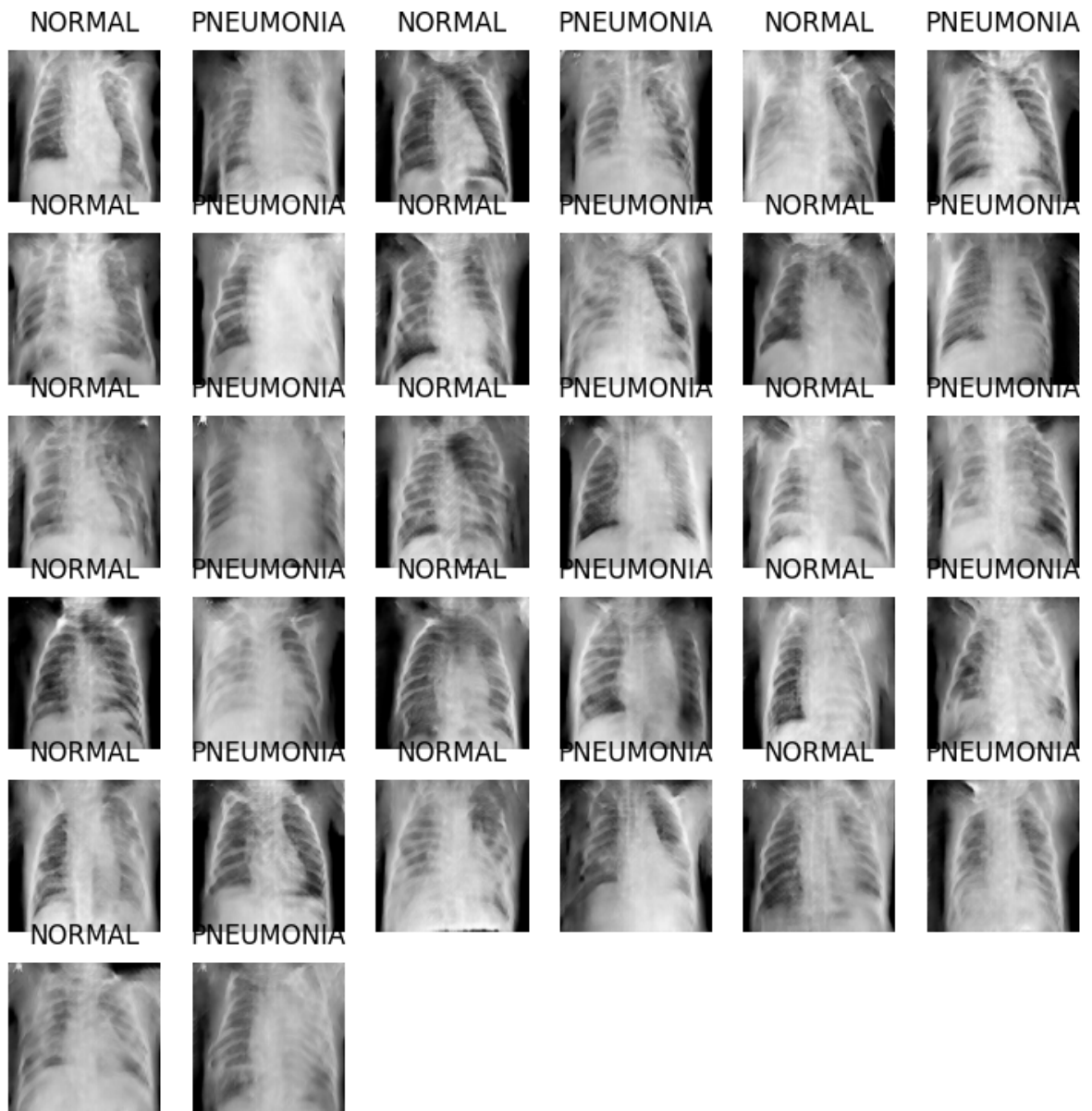


Figura 13 – Imagens sintéticas na metade do treinamento, época 62.

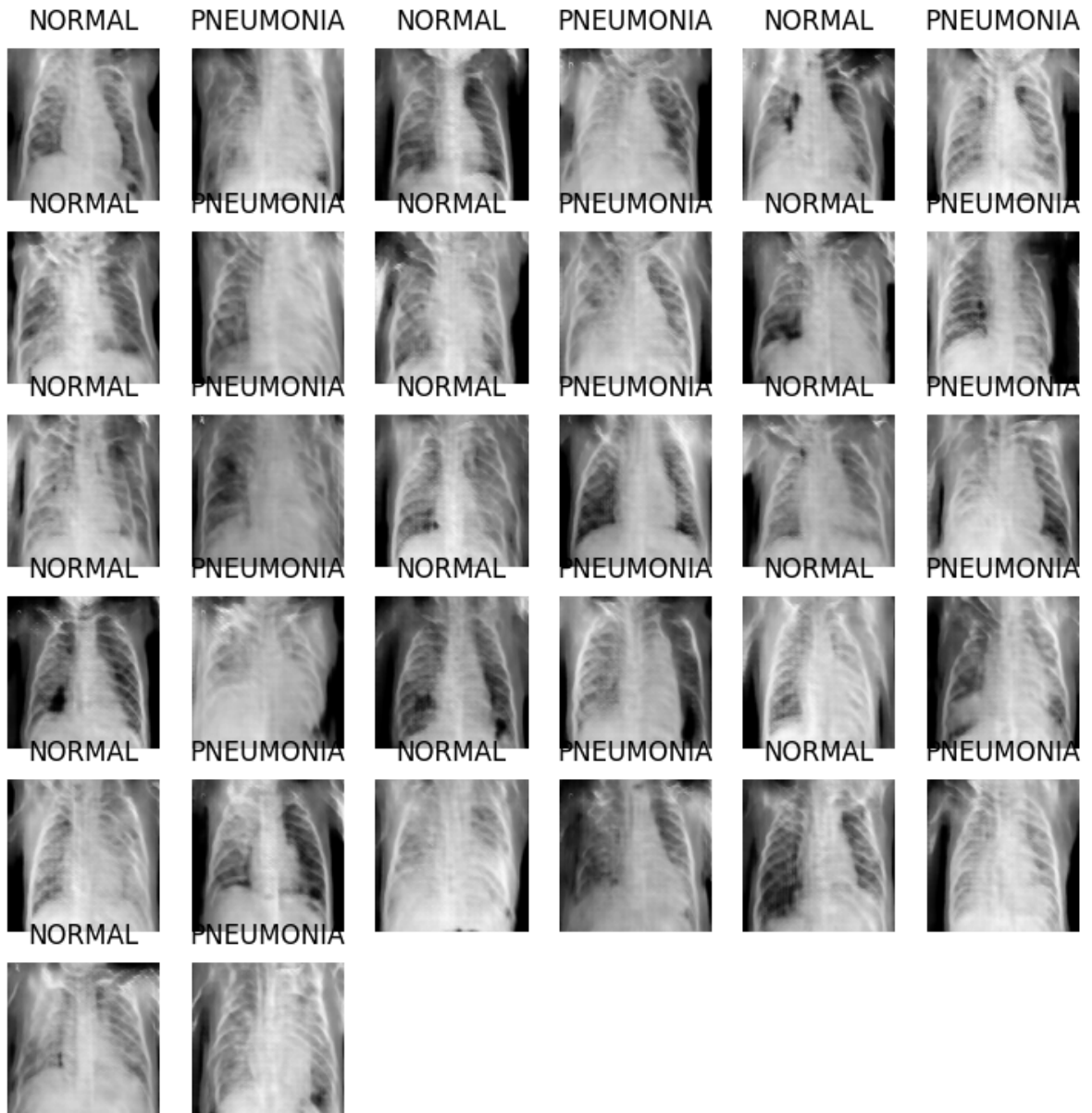


Figura 14 – Imagens sintéticas no final do treinamento, época 100.

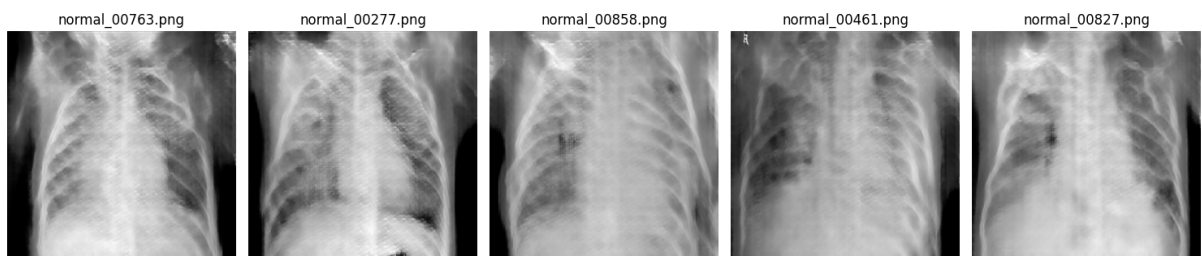


Figura 15 – Imagens da classe normal geradas pelo modelo.

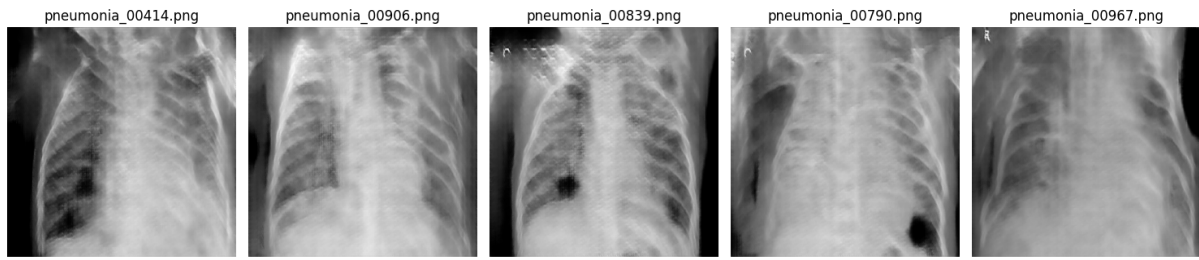


Figura 16 – Imagens da classe pneumonia geradas pelo modelo.

Tabela 2 – Relatório de Classificação (Imagens sintéticas)

Classe	Precisão (Precision)	Revocação (Recall)	F1-Score	Suporte (n)
NORMAL	0.4932	0.0360	0.0671	1000
PNEUMONIA	0.4997	0.9630	0.6580	1000
Acurácia global: 0.4995				
Médias macro: Precisão: 0.4964 — Revocação: 0.4995 — F1: 0.3626				
Médias ponderadas: Precisão: 0.4964 — Revocação: 0.4995 — F1: 0.3626				

Tabela 3 – Valores de FID e KID do modelo DDGAN

grupo	n_real	n_sint	FID	KID	KID se
GLOBAL	507	546	249.271102	6.190507	1.586688e-03
NORMAL	234	273	313.809291	8.758761	4.843418e-03
PNEUMONIA	273	273	225.792652	4.930082	1.873673e-07

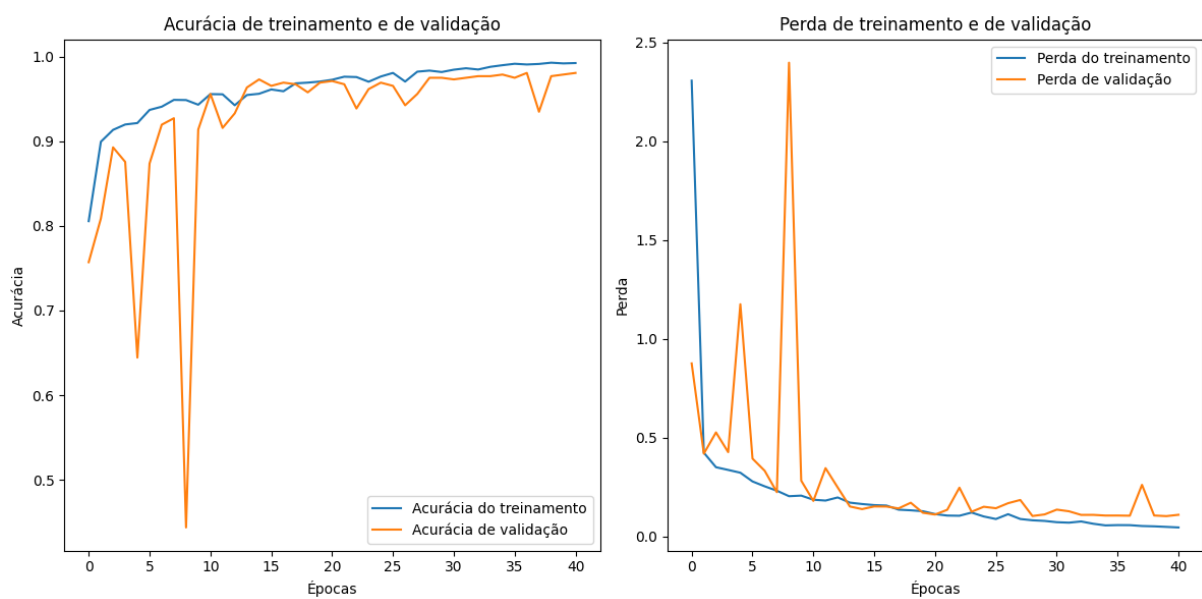


Figura 17 – Histórico de acurácia e de perda ao longo do treinamento.

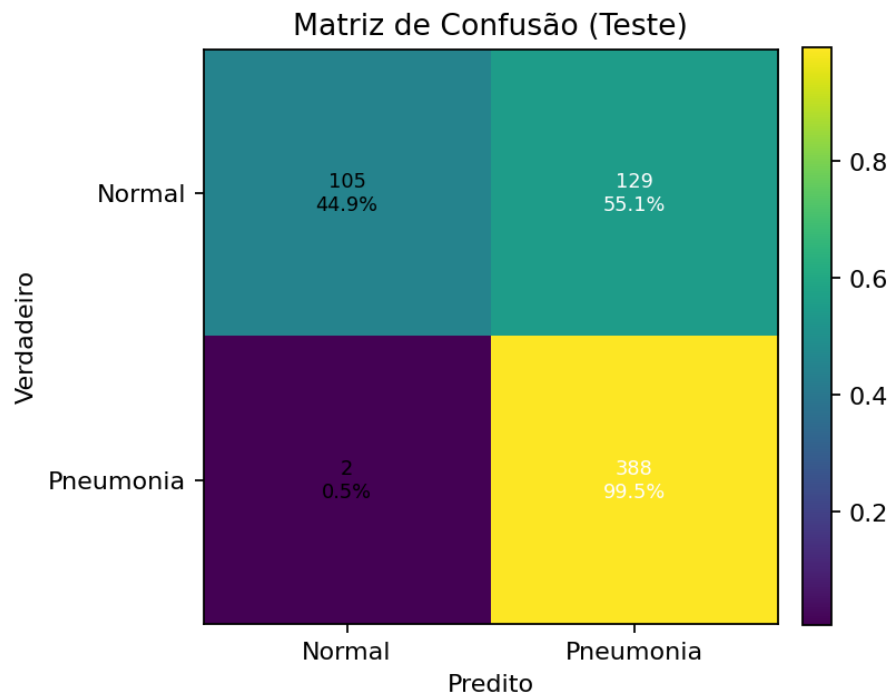


Figura 18 – Matriz de confusão para imagens reais.

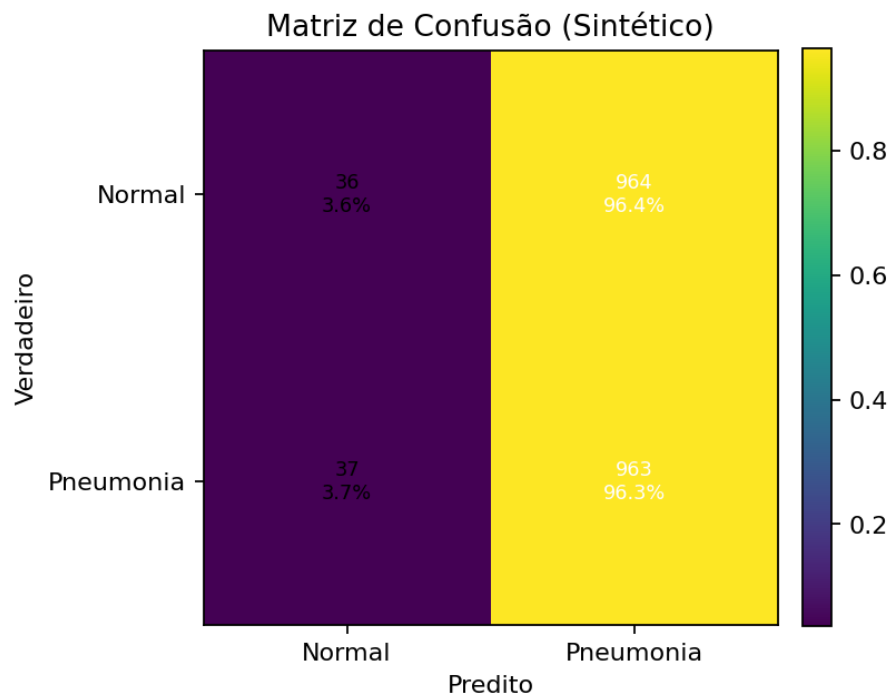


Figura 19 – Matriz de confusão para imagens sintéticas.

5 DISCUSSÃO

Os relatórios de classificação deixam em evidência uma queda significativa em todas as métricas do modelo, com o modelo UBNNet treinado apenas com imagens reais tendo uma acurácia média de 79,01%, precisão média de 86,59%, recall de 72,18%, e F1-Score de 73,57%, conforme a tabela 1, métricas boas mas não suficientes para um modelo classificador de imagens médicas, e o modelo UBNNet treinado com imagens sintéticas apresentando resultados abaixo de 50%, com acurácia média de 49,95%, precisão de 49,64%, recall de 49,95%, e F1-Score de 36,26% evidenciado na tabela 2, indicando que as imagens sintéticas tiveram um efeito negativo no treinamento do modelo. Isso é exemplificado pelas matrizes de confusão (figuras 18 e 19), com o modelo inicialmente classificando corretamente, com um leve viés para pneumonia, mas depois de treinado com imagens sintéticas esse viés intensificou, com a maioria das classificações sendo pneumonia e poucas sendo normal.

Essa queda de performance do modelo pode estar atrelada ao processo de colapso de modelo, visto que ocorreu devido a inclusão de imagens sintéticas no treinamento, porém também pode ser devido ao banco de dados estar desbalanceado, pois um modelo treinado com um banco de dados que contém mais registros de pneumonia do que registros de normal tenderá a classificar imagens como pneumonia com mais frequência do que normal. Porém, o banco de dados utilizado não possui um desbalanço significativo para resultar em um viés tão forte como o obtido, e se esse for o caso, o modelo apresentaria esse viés forte mesmo antes de ser treinado com as imagens sintéticas, o que não foi o caso. Analisando os valores de FID e de KID (tabela 3) confirma que as imagens sintéticas são de baixa qualidade, com um valor global de FID de 249. Apesar do valor global de KID ser ideal (próximo de zero), a discrepância com a FID indica que as imagens sintéticas possuem pouca semelhança com as imagens reais.

As imagens sintéticas foram de baixa qualidade, com vários possíveis fatores sendo responsáveis por isso. Estes sendo pré-processamento insuficiente dos dados, por exemplo, pequenos desvios no intervalo dos dados afetam significativamente as métricas (DOHMEN et al., 2024), modelos muito simples possuem dificuldade em aprender as informações importantes para recriar autenticamente os dados em que foram treinados, viés inerente ao banco de dados, e a complexidade das imagens médicas porque possuem pequenos detalhes

e características visuais, como ruído e imperfeições, demasiadamente complexas para um modelo conseguir replicar.

6 CONCLUSÃO

Este trabalho teve como objetivo verificar a viabilidade do uso de imagens sintéticas no treinamento de modelos de redes neurais convolucionais (CNNs) como forma de reduzir os problemas comuns aos bancos de dados de imagens médicas, por exemplo a baixa disponibilidade de dados. Infelizmente, o resultado obtido foi o oposto, a inclusão de imagens sintéticas causa uma queda na performance de modelos CNNs, inviabilizando este meio como uma solução para o problema de escassez de dados.

Estudos anteriores que compartilham este objetivo indicam que a utilização de imagens sintéticas constitui uma alternativa viável (AKHIL et al., 2024). Porém, a complexidade necessária para que modelos consigam atingir tal fim requer poder computacional que não estava disponível para este trabalho. Logo, para aperfeiçoamento desse trabalho e futuras perspectivas, podem ser construídos um modelo DCGAN ou talvez variações de GANs, tais como f-GAN ou WGAN, que resolvem ou, ao menos, mitigam os problemas do DCGAN como colapso de modelo, mais complexos e capazes de reproduzir as características mais delicadas de imagens médicas. Também podem ser aplicadas modelos de deep learning mais recentes e refinados, como modelos de difusão ou modelos pré-treinados, que frequentemente apresentam métricas de desempenho superiores.

Referências¹

- AKHIL, M. S. et al. Medical image synthesis using dcgan for chest x-ray images. In: IEEE. *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. [S.l.], 2024. v. 1, p. 1–8. Citado 2 vezes nas páginas 20 e 34.
- BUSHBERG, J. T.; BOONE, J. M. *The essential physics of medical imaging*. [S.l.]: Lippincott Williams & Wilkins, 2011. Citado 2 vezes nas páginas 12 e 13.
- DOHMEN, M. et al. Five pitfalls when assessing synthetic medical images with reference metrics. In: SPRINGER. *MICCAI Workshop on Deep Generative Models*. [S.l.], 2024. p. 150–159. Citado na página 32.
- EISBERG, R.; RESNICK, R. *Quantum physics of atoms, molecules, solids, nuclei, and particles*. [S.l.: s.n.], 1985. Citado na página 12.
- HAN, C. et al. Gan-based synthetic brain mr image generation. In: IEEE. *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. [S.l.], 2018. p. 734–738. Citado na página 10.
- HO, J.; SALIMANS, T. *Classifier-Free Diffusion Guidance*. 2022. Disponível em: <<https://arxiv.org/abs/2207.12598>>. Citado na página 22.
- ISLAM, S. et al. Generative adversarial networks (gans) in medical imaging: Advancements, applications, and challenges. *IEEE Access*, IEEE, v. 12, p. 35728–35753, 2024. Citado na página 10.
- KIRU, M. U. et al. Comparative analysis of some selected generative adversarial network models for image augmentation: a case study of covid-19 x-ray and ct images. *Journal of Intelligent & Fuzzy Systems*, SAGE Publications Sage UK: London, England, v. 43, n. 6, p. 7153–7172, 2022. Citado na página 10.
- KOÇAK, B. et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagnostic and interventional radiology*, v. 31, n. 2, p. 75, 2025. Citado na página 17.
- KOETZIER, L. R. et al. Generating synthetic data for medical imaging. *Radiology*, Radiological Society of North America, v. 312, n. 3, p. e232471, 2024. Citado na página 17.
- LIU, B. et al. Application of an improved dcgan for image generation. *Mobile information systems*, Wiley Online Library, v. 2022, n. 1, p. 9005552, 2022. Citado na página 20.
- MARKEL, H. 'i have seen my death': How the world discovered the x-ray. PBS, 2012. Citado 3 vezes nas páginas 6, 10 e 11.
- MOTAMED, S.; ROGALLA, P.; KHALVATI, F. Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images. *Informatics in medicine unlocked*, Elsevier, v. 27, p. 100779, 2021. Citado na página 20.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

- MOTAMED, S.; ROGALLA, P.; KHALVATI, F. Randgan: randomized generative adversarial network for detection of covid-19 in chest x-ray. *Scientific Reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 8602, 2021. Citado na página 15.
- MUNAWAR, F. et al. Segmentation of lungs in chest x-ray image using generative adversarial networks. *Ieee Access*, IEEE, v. 8, p. 153535–153545, 2020. Citado na página 15.
- OKUNO, E.; YOSHIMURA, E. M. *Física das radiações*. [S.l.]: Oficina de Textos, 2016. Citado 2 vezes nas páginas 11 e 14.
- PETERSON, A. J. Ai and the problem of knowledge collapse. *AI & SOCIETY*, Springer, p. 1–21, 2025. Citado na página 17.
- RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. Citado na página 20.
- SHUMAILOV, I. et al. The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493*, 2023. Citado na página 16.
- SHUMAILOV, I. et al. Ai models collapse when trained on recursively generated data. *Nature*, Nature Publishing Group UK London, v. 631, n. 8022, p. 755–759, 2024. Citado na página 16.
- SILVA, O. E. da. Estudo do exchange bias em filmes finos de nife/femn (bicamadas) e nife/irmn (multicamadas). 2016. Citado 2 vezes nas páginas 6 e 12.
- SINGH, N. K.; RAZA, K. Medical image generation using generative adversarial networks. *arXiv preprint arXiv:2005.10687*, 2020. Citado 2 vezes nas páginas 15 e 21.
- SRIVASTAV, D.; BAJPAI, A.; SRIVASTAVA, P. Improved classification for pneumonia detection using transfer learning with gan based synthetic image augmentation. In: *IEEE. 2021 11th international conference on cloud computing, data science & engineering (confluence)*. [S.l.], 2021. p. 433–437. Citado na página 20.
- ULLAH, Z. et al. Densely attention mechanism based network for covid-19 detection in chest x-rays. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 261, 2023. Citado 2 vezes nas páginas 6 e 15.
- VJ, S. Deep learning algorithm for covid-19 classification using chest x-ray images. *Computational and Mathematical Methods in Medicine*, Wiley Online Library, v. 2021, n. 1, p. 9269173, 2021. Citado na página 20.
- WANG, X. et al. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 9049–9058. Citado na página 17.
- WIDODO, C. S. et al. Ubnet: Deep learning-based approach for automatic x-ray image detection of pneumonia and covid-19 patients. *Journal of X-ray science and technology*, SAGE Publications Sage UK: London, England, v. 30, n. 1, p. 57–71, 2022. Citado na página 22.

YADAV, S. S.; JADHAV, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big data*, Springer, v. 6, n. 1, p. 1–18, 2019. Citado na página 17.