

JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Utilizando Autoencoders Variacionais,  
Redes Generativas Adversariais e Aumento  
de Dados para melhorar a detecção da  
Esclerose Múltipla.**

**Gustavo Gonçalves Silva**

JUIZ DE FORA  
MARÇO, 2025

# **Utilizando Autoencoders Variacionais, Redes Generativas Adversariais e Aumento de Dados para melhorar a detecção da Esclerose Múltipla.**

GUSTAVO GONÇALVES SILVA

Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Ciência da Computação

Orientador: Barbara de Melo Quintela  
Coorientador: Marcelo Lobosco

JUIZ DE FORA  
MARÇO, 2025

**UTILIZANDO AUTOENCODERS VARIACIONAIS, REDES  
GENERATIVAS ADVERSARIAIS E AUMENTO DE DADOS  
PARA MELHORAR A DETECÇÃO DA ESCLEROSE MÚLTIPLA.**

Gustavo Gonçalves Silva

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Barbara de Melo Quintela  
Doutora em Modelagem Computacional

Marcelo Lobosco  
Doutor em Ciência da Computação

Rodrigo Weber dos Santos  
Doutor em Ciência da Computação

Leonardo Goliatt  
Doutor em Modelagem Computacional

JUIZ DE FORA  
DIA DE MARÇO, 2025

*Aos meus pais, por tudo. À minha namorada,  
por estar comigo. Aos meus amigos, por dar  
sentido a faculdade.*

## Resumo

A Esclerose Múltipla (EM) é uma doença neurodegenerativa cuja avaliação clínica depende fundamentalmente da identificação e quantificação de lesões em exames de Ressonância Magnética (RM). Embora a segmentação automática baseada em Aprendizado Profundo (*Deep Learning*) tenha alcançado o estado da arte, sua eficácia é frequentemente limitada pela escassez de dados médicos anotados e pela complexidade da obtenção de novos exemplos. Este trabalho propõe e avalia um *pipeline* de *Data Augmentation* generativo para mitigar esse problema, investigando o impacto de imagens sintéticas na qualidade de modelos de segmentação. O método desenvolvido combina um *Variational Autoencoder* (VAE) para a geração probabilística de máscaras de lesões e uma Rede Generativa Adversarial Condisional (CGAN) para a síntese de texturas de RM realistas a partir dessas máscaras. As imagens geradas foram integradas ao treinamento de uma rede SegResNet para a tarefa de segmentação supervisionada. Os resultados demonstraram que a inclusão de dados sintéticos promoveu um aumento significativo na métrica Dice em comparação ao uso exclusivo de dados reais. Observou-se que a adição de 750 imagens sintéticas proporcionou o melhor ganho de qualidade na segmentação, indicando saturação do aprendizado em volumes superiores devido a limitações na síntese de lesões de pequenas dimensões. Conclui-se que o uso de modelos generativos adversariais é uma estratégia promissora para enriquecer bases de dados médicas e aprimorar a generalização de diagnósticos automatizados.

**Palavras-chave:** Esclerose Múltipla; Segmentação Automática; Data Augmentation; GAN; Variational Autoencoder.

## Abstract

Multiple Sclerosis (MS) is a neurodegenerative disease where clinical assessment relies fundamentally on the identification and quantification of lesions in Magnetic Resonance Imaging (MRI). Although Deep Learning-based automated segmentation has reached state-of-the-art status, its effectiveness is often limited by the scarcity of annotated medical data and the complexity of acquiring new samples. This work proposes and evaluates a generative Data Augmentation pipeline to mitigate this issue, investigating the impact of synthetic images on segmentation model quality. The developed methodology combines a Variational Autoencoder (VAE) for the probabilistic generation of lesion masks and a Conditional Generative Adversarial Network (CGAN) for synthesizing realistic MRI textures from these masks. The generated images were integrated into the training of a SegResNet network for the supervised segmentation task. Results showed that the inclusion of synthetic data promoted a significant increase in the Dice metric compared to the exclusive use of real data. It was observed that adding 750 synthetic images provided the best performance gain, with learning saturation occurring at higher volumes due to limitations in synthesizing small-dimension lesions. It is concluded that the use of generative adversarial models is a promising strategy for enriching medical datasets and improving the generalization of automated diagnoses.

**Keywords:** Multiple Sclerosis; Automated Segmentation; Data Augmentation; GAN; Variational Autoencoder.

## **Agradecimentos**

À minha família, em especial minha mãe Eulália, meu pai Heleno e meu irmão Luis Felipe, por me apoiarem nessa etapa.

À minha namorada, Rafaela, por me ajudar em momentos difíceis durante a faculdade, e na confecção desse texto.

Aos meus orientadores, Bárbara e Lobosco, que foram luz durante todo meu processo de formação como Cientista da Computação.

Por fim, gostaria de dedicar um agradecimento especial a todos aqueles que estiveram comigo durante todo o processo de graduação. Em especial, Guilherme, Nícolas, Abraão, Bonaldi, Lucas, Maria Luisa, Maycon e Gabriela. Se o processo fez sentido, foi por causa de vocês.

*“O que toda história tem em comum? É que elas acabam”.*

*Brandon Sanderson*

# Conteúdo

<b>Lista de Figuras</b>	<b>7</b>
<b>Lista de Tabelas</b>	<b>8</b>
<b>Lista de Abreviações</b>	<b>9</b>
<b>1 Introdução</b>	<b>10</b>
1.1 Objetivos . . . . .	11
<b>2 Fundamentos básicos</b>	<b>12</b>
2.1 Sistema Imune . . . . .	12
2.1.1 Resposta Imune Inata . . . . .	12
2.1.2 Resposta Imune Adquirida . . . . .	12
2.2 Esclerose Múltipla . . . . .	13
2.3 Redes Neurais . . . . .	14
2.4 Segmentação automática . . . . .	15
2.5 Formato <i>NIFTI</i> . . . . .	16
2.6 Formas de aquisição . . . . .	17
2.7 Geração de imagens e <i>Data augmentation</i> . . . . .	18
<b>3 Materiais e Métodos</b>	<b>21</b>
3.1 Base de dados . . . . .	21
3.2 Ambiente computacional . . . . .	22
3.3 Modelo de segmentação . . . . .	24
3.3.1 Aumento de dados clássico . . . . .	25
3.3.2 Formulação da Função de Perda . . . . .	26
3.4 Geração artificial de imagens . . . . .	27
3.4.1 <i>Variational Autoencoder</i> . . . . .	28
3.4.2 Otimização Variacional e Termos de Regularização . . . . .	29
3.4.3 Estratégia de Treinamento . . . . .	30
3.5 Modelo Generativo Adversarial . . . . .	31
3.5.1 Função Objetivo Composta . . . . .	32
3.5.2 Detalhes de Implementação . . . . .	34
<b>4 Resultados</b>	<b>35</b>
4.1 <i>Variational Autoencoder</i> . . . . .	35
4.2 GAN . . . . .	37
4.3 Modelo de segmentação . . . . .	40
<b>5 Conclusão</b>	<b>46</b>
5.1 Trabalhos futuros . . . . .	47
<b>Bibliografia</b>	<b>48</b>

## Lista de Figuras

2.1	Exemplo das diferentes formas de aquisição de um paciente presente no dataset de Esclerose Múltipla utilizado nesse trabalho. . . . .	18
2.2	Estrutura padrão de uma GAN. Adaptado de Aggarwal, Mittal e Battineni (2021) . . . . .	20
3.1	Exemplo de sobreposição de máscara e imagem de um corte no eixo $z$ de um paciente com lesões de Esclerose Múltipla identificadas. . . . .	22
3.2	Exemplo de imagem fortemente dependente da posição ao longo do eixo $y$ , na qual a ausência de informação espacial volumétrica pode levar o modelo a interpretações ambíguas, resultando em falhas de generalização ou colapso do aprendizado . . . . .	23
4.1	Valores da função de perda total no conjunto de treino. . . . .	35
4.2	Valores da função de perda de reconstrução no conjunto de treino. . . . .	36
4.3	Valores da função de perda kl de reconstrução no conjunto de treino. . . .	36
4.4	Valores da função de perda total no conjunto de validação. . . . .	37
4.5	Valores da função de perda de reconstrução no conjunto de validação. . . .	37
4.6	Valores da função de perda de reconstrução no conjunto de validação . . .	38
4.7	Exemplo de máscaras falsas geradas pelo Variational Autoencoder . . . .	38
4.8	Perda total do Gerador ao longo das épocas. . . . .	39
4.9	Perda total do discriminador ao longo das épocas. . . . .	39
4.10	Comparação de texturas reais do cérebro com texturas geradas pela Rede Generativa Adversarial, sendo, a primeira linha a máscara (vinda diretamente do dataset original), segunda linha a textura falsa e a terceira o corte real. . . . .	40
4.11	Comparação de textura gerada a partir de máscaras falsas produzidas através do Variational Autoencoder. . . . .	40
4.12	Imagens com artefatos visuais devido ao pequeno tamanho das lesões contidas na máscara de condicionamento . . . . .	41
4.13	Métrica de Dice associadas ao modelo treinado com conjunto de dados padrão	42
4.14	Comparação entre diferentes modelos, onde o default é o treinado no dataset padrão, um com 500 entradas extras e outro com 1000 entradas extras	42
4.15	Boxplot da métrica de dice médio dos quartis do conjunto de validação, avaliado no melhor modelo obtido . . . . .	43
4.16	Gráfico do dice médio de cada item no conjunto de validação separado por quartis relacionados ao tamanho das lesões, gráfico avaliado no melhor modelo obtido . . . . .	44
4.17	Gráfico do PICO de uso de VRAM durante cada etapa do treinamento, esse gráfico desconsidera o cache utilizado pelo Pytorch devido a alta volatilidade dessa informação durante diversos momentos no treinamento . . .	45

## **Lista de Tabelas**

2.1	Comparação entre as sequências de RM T1, T2 e FLAIR no contexto da Esclerose Múltipla (EM) . . . . .	17
3.1	Configuração do modelo de segmentação através da biblioteca MONAI. . . . .	25
3.2	Pesos utilizados para as funções de perda do modelo Generativo Adversarial	33
4.1	Precisão em relação as lesões demarcadas em porcentagem(Ruído), sendo a taxa de lesões segmentadas que não existe nenhuma correspondencia na máscara real. . . . .	43
4.2	Valores obtidos para o as principais métricas nas diferentes configurações de treino . . . . .	43
4.3	Recursos computacionais necessários para execução de cada modelo . . . . .	44

## **Lista de Abreviações**

DCC Departamento de Ciência da Computação

UFJF Universidade Federal de Juiz de Fora

# 1 Introdução

De acordo com a Organização Mundial da Saúde, a Esclerose Múltipla (EM) é uma doença autoimune neurodegenerativa que afeta cerca de 1,8 milhões de indivíduos globalmente (WHO, 2025). A doença é caracterizada por lesões causadas pelo sistema imunológico à bainha de mielina, responsável por auxiliar na propagação de impulsos nervosos pelo corpo.

O diagnóstico da doença é comumente estabelecido por meio da análise de imagens de Ressonância Magnética (*Magnetic Resonance Imaging*, MRI), especialmente quando associadas aos sintomas clínicos característicos (BROWNLEE et al., 2017). A identificação e delimitação das áreas lesionadas no cérebro e na medula espinhal são tradicionalmente realizadas de forma manual por especialistas, utilizando ferramentas de segmentação, o que contribui não apenas para o diagnóstico, mas também para a avaliação da progressão da doença e da eficácia dos tratamentos instituídos.

A segmentação manual de imagens de Ressonância Magnética é um processo de notória complexidade, visto o elevado tempo demandado para sua realização e por estar sujeita a erros humanos, mesmo quando realizado por especialistas na tarefa. Nesse cenário, com o avanço de técnicas computacionais associadas ao processamento de imagens e visão computacional, a segmentação automática de lesões associadas à Esclerose Múltipla se mostra uma ferramenta promissora no que tange o diagnóstico e avaliação da progressão da doença e resposta do indivíduo às drogas que diminuem a degeneração (LLADÓ et al., 2012). Como demonstrado por Rosa et al. (2024), o uso de técnicas de aprendizado profundo, especialmente aproveitando-se dos resultados de alta qualidade providos por Redes Neurais Convolucionais (CNNs), permitiu avanços significativos nos algoritmos para segmentação automática.

Contudo, para aumento da eficácia dos métodos de segmentação automática, notoriamente aqueles que se baseiam em Aprendizado Profundo para seu funcionamento, deve-se prover uma quantidade significativa de dados aos modelos, para obtenção de métricas de qualidade satisfatórias. Entretanto, a obtenção desses dados é um processo

árduo, devido à sensibilidade inerente às informações clínicas de pacientes acometidos pela EM. Para contextos semelhantes, técnicas de *Data Augmentation* vêm sendo aplicadas, com o objetivo de utilizar métodos generativos para aumentar a disponibilidade dos dados já disponíveis.

O presente trabalho busca investigar o impacto de imagens geradas artificialmente através de Redes Generativas Adversárias na qualidade de modelos de segmentação de imagens de Ressonância Magnética de pacientes com Esclerose Múltipla. Para isso, foi proposto um pipeline de geração artificial composto por dois estágios. O primeiro utiliza um *VAE*(*Variational Autoencoder*) para geração de máscaras binárias de lesões de Esclerose Múltipla que se assemelham as lesões presentes em datasets reais. No segundo estágio, foi empregado uma *cGAN*(*Conditional Generative Adversarial Network*) para síntese de texturas cerebrais sob o condicionamento de máscaras de lesões.

## 1.1 Objetivos

Dada a relevância dos estudos voltados ao entendimento da Esclerose Múltipla, bem como a elevada complexidade do processo de segmentação manual das lesões e a escassez de bases de dados públicas de fácil acesso, torna-se necessário investigar abordagens alternativas. Nesse contexto, este trabalho tem como objetivo avaliar o impacto da inclusão de imagens geradas artificialmente na qualidade de um modelo baseado na arquitetura *ResNet*, adaptado para a segmentação semântica de lesões de Esclerose Múltipla. Além disso, esse trabalho tem como objetivos secundários:

- Comparar métricas de precisão (Dice) entre modelos treinados com bases puramente reais *versus* bases híbridas (reais + sintéticas).
- Quantificar o *overhead* computacional introduzido pelo aumento do volume de dados, analisando o tempo de convergência durante o treinamento.
- Mensurar a demanda de recursos de hardware (uso de VRAM e tempo desprendidos) proporcional à inserção gradual de volumes sintéticos no *dataset*.

## 2 Fundamentos básicos

### 2.1 Sistema Imune

O sistema imune é uma organização de células e moléculas altamente especializadas em combater infecções que acometem o organismo; ao combate dessas infecções é dado o nome de resposta imune. Existem fundamentalmente duas principais formas de resposta imune a antígenos invasores, denominadas Resposta Imune Inata e Resposta Imune Adquirida (DELVES; ROITT, 2000).

#### 2.1.1 Resposta Imune Inata

O sistema imune inato constitui a primeira linha de defesa do organismo contra agentes patogênicos, atuando de forma rápida e inespecífica logo após a transposição das barreiras físicas e anatômicas. Esse sistema baseia-se em um repertório limitado de receptores de reconhecimento de padrões, capazes de identificar estruturas moleculares conservadas compartilhadas por uma ampla gama de microrganismos invasores (TURVEY; BROIDE, 2010). Entre seus principais componentes estão células especializadas, como macrófagos e neutrófilos, além de mediadores solúveis, incluindo citocinas, que coordenam e amplificam a resposta inflamatória inicial.

Além de sua atuação direta na contenção de patógenos, o sistema imune inato desempenha papel fundamental na ativação da resposta imune adaptativa, uma vez que, após a fagocitose, fragmentos antigênicos derivados dos invasores são processados e apresentados às células do sistema adaptativo, possibilitando o desenvolvimento de respostas imunológicas específicas.

#### 2.1.2 Resposta Imune Adquirida

O sistema imune adquirido, por outro lado, produz uma resposta altamente específica ao patógeno invasor. Após a apresentação do antígeno por células do sistema imune inato,

denominadas células apresentadoras de antígeno (APCs, do inglês *Antigen-Presenting Cells*), os linfócitos B e T passam a atuar de forma coordenada na produção de anticorpos específicos capazes de reconhecer e neutralizar estruturas moleculares características do patógeno.

Nesse contexto, os linfócitos T desempenham papel central ao reconhecer complexos formados por peptídeos antigênicos associados às moléculas do Complexo Principal de Histocompatibilidade (*Major Histocompatibility Complex – MHC*) expressas pelas APCs, além de mediar mecanismos efetores direcionados à eliminação do antígeno. Os linfócitos T são classicamente subdivididos em duas principais populações funcionais:

- **CD4<sup>+</sup>**: células responsáveis pela coordenação da resposta imune adquirida, ativando outras células de defesa, como macrófagos e linfócitos T CD8<sup>+</sup>.
- **CD8<sup>+</sup>**: células responsáveis pela eliminação direta de células infectadas ou alteradas, por meio do reconhecimento de antígenos apresentados por moléculas do MHC classe I.

## 2.2 Esclerose Múltipla

Dobson e Giovannoni (2019) descrevem a Esclerose Múltipla como uma doença autoimune mediada por células T, caracterizada por inflamação crônica do sistema nervoso central. A patologia da EM envolve lesões inflamatórias perivenulares que levam à desmielinização e formação de placas, predominantemente na substância branca do cérebro e da medula espinhal. A doença manifesta-se quando o sistema imune, sobretudo por meio de linfócitos T e B autorreativos, reconhece equivocadamente a bainha de mielina como um antígeno, iniciando uma resposta imunológica patológica direcionada a esse tecido, sendo a bainha de mielina uma membrana responsável por otimizar a passagem de impulsos nervosos pelas conexões neurais. O ataque imunológico contra a bainha de mielina gera placas (ou escleroses) que dificultam a passagem dos impulsos nervosos pelas áreas afetadas. Embora a visão tradicional sugira uma doença de duas fases, uma fase inflamatória inicial associada à forma remitente-recorrente e uma fase neurodegenerativa tardia associada às formas progressivas, a eficácia de terapias direcionadas às células B desafia essa concepção.

A EM é considerada uma doença que evolui ao longo de um espectro contínuo, em que inflamação e neurodegeneração coexistem desde os estágios iniciais. Cohen et al. (2010) demonstram que os principais tratamentos para a doença se concentram na redução da quantidade de surtos que produzem as lesões, principalmente através de terapias injetáveis como a Interferon Beta ou de medicação oral como o Fingolimode, que vêm se mostrando ainda mais eficaz na contenção dos surtos da doença.

## 2.3 Redes Neurais

Redes Neurais Artificiais (Redes Neuronais Artificiais do Português Europeu, ou *Artificial Neural Networks* em Inglês) são modelos computacionais inspirados no funcionamento do sistema nervoso central de organismos biológicos, em especial do cérebro humano. Esses modelos são compostos por unidades matemáticas interconectadas, denominadas neurônios artificiais, capazes de receber sinais de entrada, processá-los por meio de combinações ponderadas seguidas da aplicação de funções de ativação, e produzir sinais de saída. Quando organizados em camadas e conectados em grande número, esses neurônios permitem que a rede aprenda representações complexas dos dados, viabilizando a resolução de tarefas como classificação, regressão, reconhecimento de padrões e tomada de decisão.

As redes neurais podem ser organizadas segundo diferentes arquiteturas, que se distinguem principalmente pela forma de conexão entre os neurônios, pela profundidade do modelo e pelo tipo de dado que processam. Essa diversidade arquitetural possibilita a aplicação de redes neurais a uma ampla gama de problemas, desde dados estruturados até imagens, sinais e sequências temporais. A escolha da arquitetura adequada está diretamente relacionada à natureza do problema, ao volume e à qualidade dos dados disponíveis, bem como aos requisitos de desempenho e capacidade de generalização do modelo.

Entre as arquiteturas mais conhecidas, destacam-se as Redes Neurais Convolucionais (*Convolutional Neural Networks*, ou CNNs), introduzidas formalmente por LeCun et al. (1989), que se tornaram amplamente utilizadas em tarefas de visão computacional, como classificação, detecção e segmentação de imagens. As CNNs exploram operações

de convolução e compartilhamento de pesos, permitindo a extração hierárquica de características espaciais relevantes.

Outra classe importante de modelos são as Redes Neurais Recorrentes (*Recurrent Neural Networks*, ou RNNs), propostas por Rumelhart, Hinton e Williams (1986), projetadas para o processamento de dados sequenciais e séries temporais. Essas redes possuem conexões recorrentes que permitem a incorporação de informações de estados anteriores. Variantes como as redes LSTM (*Long Short-Term Memory*) e GRU (*Gated Recurrent Unit*) foram desenvolvidas para mitigar problemas relacionados ao desvanecimento e à explosão do gradiente, ampliando a capacidade de modelar dependências de longo prazo.

Mais recentemente, arquiteturas baseadas no mecanismo de atenção, como os Transformers, têm alcançado resultados expressivos, especialmente no processamento de linguagem natural, devido à sua capacidade de modelar relações globais entre elementos de uma sequência de forma paralela. Além dessas arquiteturas gerais, modelos especializados como os Autoencoders e a U-Net são amplamente empregados em tarefas de redução de dimensionalidade, aprendizado de representações latentes e segmentação de imagens médicas, respectivamente, sendo esta última particularmente relevante no contexto deste trabalho. A escolha da arquitetura adequada está diretamente relacionada à natureza do problema, ao volume de dados disponíveis e aos requisitos de desempenho e generalização.

## 2.4 Segmentação automática

O diagnóstico da Esclerose Múltipla pode envolver diferentes abordagens clínicas e laboratoriais; entretanto, exames de Ressonância Magnética (RM) do encéfalo e da medula espinhal, frequentemente associados ao uso de contraste à base de gadolínio, desempenham papel central na identificação de lesões características da doença Ömerhoca, Akkaş e İcen (2018). Nesse contexto, a segmentação das lesões em imagens de RM constitui uma etapa fundamental, não apenas para o estabelecimento do diagnóstico, mas também para o acompanhamento da progressão da doença e para a avaliação da resposta aos tratamentos.

Mais recentemente, Gabr et al. (2020) apresentaram um método de segmentação automática baseado em redes neurais convolucionais profundas, especificamente uma rede

totalmente convolucional (*Fully Convolutional Neural Network*, ou FCNN), aplicado à segmentação de tecidos cerebrais e lesões de Esclerose Múltipla em um amplo conjunto de dados de Ressonância Magnética. Os resultados indicaram que a abordagem fundamentada em *Deep Learning* alcançou elevada acurácia, com coeficientes de similaridade de Dice superiores a 0,95 para tecidos cerebrais e em torno de 0,70 para lesões, validando a eficácia do método em cenários multicêntricos.

## 2.5 Formato *NIFTI*

O formato *Neuroimaging Informatics Technology Initiative* (NIfTI) é amplamente empregado no armazenamento e processamento de imagens médicas e neuroimagem, especialmente em estudos envolvendo Ressonância Magnética. Arquivos no formato NIfTI (`.nii` ou `.nii.gz`) permitem a representação eficiente de dados volumétricos tridimensionais, possibilitando a codificação de informações espaciais provenientes de exames como Tomografia Computadorizada, Ressonância Magnética e técnicas avançadas de neuroimagem funcional e estrutural.

Uma das principais características do formato NIfTI é a incorporação de um cabeçalho estruturado que armazena metadados essenciais para a correta interpretação do volume de imagem. Esses metadados incluem informações sobre dimensões do volume, resolução espacial dos *voxels*, orientação no espaço físico, tipo de dado armazenado, além de matrizes de transformação que relacionam o espaço do *voxel* ao sistema de coordenadas do mundo real. Tais propriedades tornam o formato particularmente adequado para análises quantitativas e comparações intersujeitos.

Além disso, o formato NIfTI oferece suporte nativo à representação de imagens 4D, permitindo o armazenamento de séries temporais, como aquelas obtidas em estudos de Ressonância Magnética funcional (fMRI) ou sequências dinâmicas. Essa flexibilidade facilita a integração do formato com ferramentas computacionais amplamente utilizadas em neuroimagem, incluindo bibliotecas e softwares de análise estatística, visualização e segmentação automática de estruturas anatômicas e lesões.

Devido à sua padronização, simplicidade e ampla compatibilidade com diferentes plataformas e linguagens de programação, o formato NIfTI tornou-se um padrão em

pesquisas em neuroimagem, sendo amplamente adotado em *pipelines* de processamento automático e em bases de dados públicas utilizadas para fins científicos.

## 2.6 Formas de aquisição

No contexto da Ressonância Magnética (RM), existem diferentes formas de aquisição de imagens, também denominadas sequências de aquisição. Essas sequências correspondem a diferentes configurações dos parâmetros físicos do equipamento, como tempos de repetição e eco, que resultam em variações no contraste entre os tecidos biológicos. Dessa forma, uma mesma estrutura anatômica pode apresentar intensidades de sinal distintas, permitindo a ênfase de características específicas dos tecidos e de alterações patológicas.

No estudo da Esclerose Múltipla, a utilização de múltiplas sequências de RM é fundamental, uma vez que diferentes tipos de lesões e estágios da doença são melhor evidenciados por contrastes específicos. As sequências mais comumente empregadas nesse contexto são T1, T2 e FLAIR, cada uma fornecendo informações complementares relevantes tanto para o diagnóstico quanto para o acompanhamento da progressão da doença.

A Tabela 2.1 apresenta uma comparação entre essas sequências, destacando suas principais características e aplicações no contexto da Esclerose Múltipla. Adicionalmente, a Figura 2.1 ilustra um exemplo de um mesmo corte axial adquirido nas sequências T1, T2 e FLAIR, evidenciando as diferenças de contraste observadas entre elas.

Sequência	Principais características	Aplicação em EM
<b>T1</b>	LCR escuro Boa definição anatômica	Visualização da anatomia Lesões crônicas (com gadolínio)
<b>T2</b>	LCR claro Alta sensibilidade à água	Identificação de lesões Destaca áreas desmielinizantes
<b>FLAIR</b>	Baseada em T2 Sinal do LCR suprimido	Detecção de lesões periventriculares Sequência mais usada na EM

Tabela 2.1: Comparação entre as sequências de RM T1, T2 e FLAIR no contexto da Esclerose Múltipla (EM).

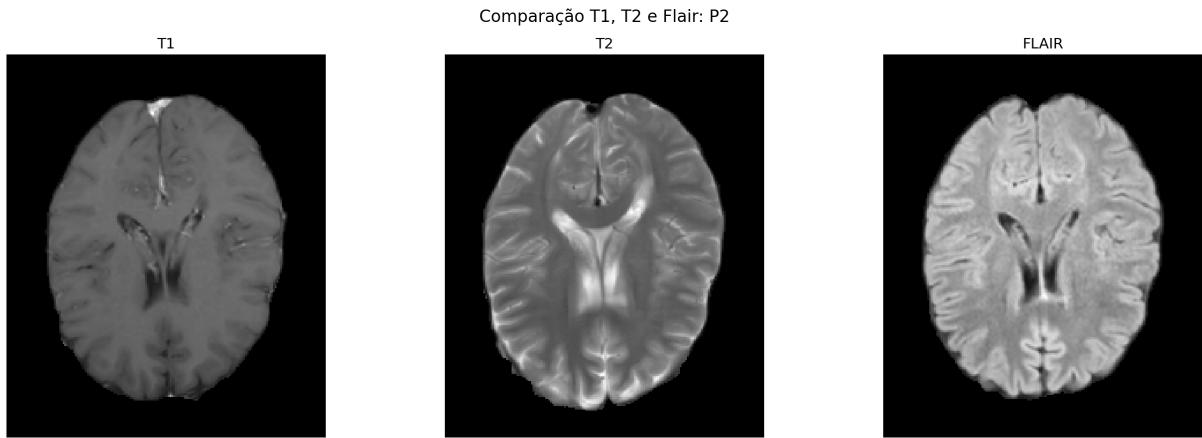


Figura 2.1: Exemplo das diferentes formas de aquisição de um paciente presente no *dataset* de Esclerose Múltipla utilizado nesse trabalho.

## 2.7 Geração de imagens e *Data augmentation*

Técnicas de *Data Augmentation* consistem na aplicação de transformações sistemáticas e controladas sobre os dados originais com o objetivo de ampliar artificialmente o conjunto de treinamento. Essa estratégia visa aumentar a diversidade amostral, melhorar a capacidade de generalização dos modelos e reduzir o risco de *overfitting*, especialmente em cenários caracterizados pela escassez de dados rotulados. No contexto de imagens, tais técnicas tradicionalmente envolvem transformações geométricas e fotométricas relativamente simples, como rotações, translações, inversões, escalonamentos e recortes. Apesar de sua simplicidade, essas abordagens continuam amplamente empregadas em trabalhos recentes, em razão do baixo custo computacional e da facilidade de implementação (CETIN et al., 2025).

Com o avanço de métodos modernos de aprendizado de máquina, em particular aqueles baseados em aprendizado profundo, modelos generativos passaram a desempenhar um papel relevante no aumento artificial de dados. Diferentemente das abordagens tradicionais, esses modelos são capazes de aprender a distribuição subjacente dos dados de treinamento e sintetizar novas amostras com características estatísticas semelhantes às dos dados reais. Entre essas abordagens, destacam-se as Redes Generativas Adversariais (*Generative Adversarial Networks*, ou GANs), amplamente utilizadas na geração de imagens sintéticas em diversos domínios.

Conforme ilustrado na Figura 2.2, uma GAN é composta por dois modelos trei-

nados simultaneamente: o Gerador e o Discriminador. O Gerador recebe como entrada um vetor de ruído amostrado de um espaço latente e tem como objetivo produzir amostras sintéticas que se aproximem da distribuição dos dados reais. O Discriminador, por sua vez, atua como um classificador binário, geralmente implementado por meio de uma rede neural convolucional, cuja função é distinguir entre amostras reais e aquelas geradas artificialmente.

Esse arranjo estabelece um processo de otimização adversarial, formulado como um jogo de soma zero, no qual o Gerador é treinado para maximizar a probabilidade de enganar o Discriminador, enquanto este busca minimizar sua taxa de erro. Ao longo do treinamento, espera-se que o Gerador aprenda representações cada vez mais fiéis da distribuição dos dados, resultando na geração de amostras sintéticas visual e estatisticamente coerentes com o conjunto original. (IMRAN; TERZOPOULOS, 2021) utilizou uma arquitetura de rede que incorpora um conjunto de discriminadores em uma rede VAE-GAN, usando conjuntos de dados dos domínios de visão computacional e imagens médicas, para gerar novas imagens realistas de dados médicos. (BARILE et al., 2021) demonstra a capacidade de GAN's para *Data Augmentation* em tarefas de classificação de Esclerose Múltipla.

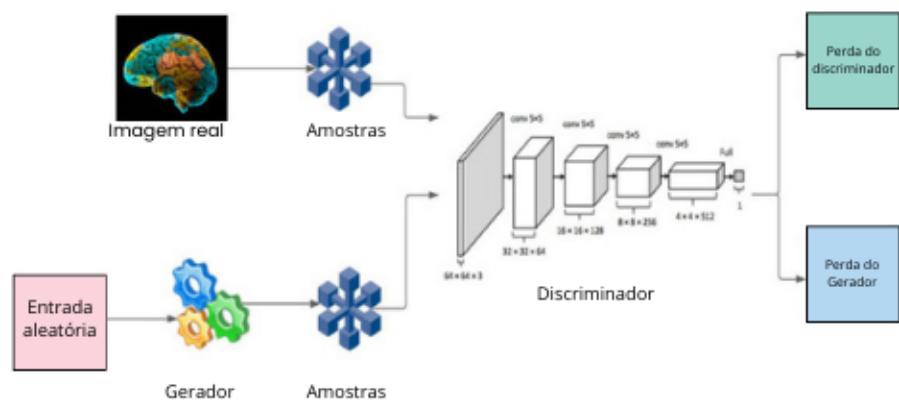


Figura 2.2: Estrutura padrão de uma GAN. Adaptado de Aggarwal, Mittal e Battineni (2021)

## 3 Materiais e Métodos

### 3.1 Base de dados

O artigo de Guarnera et al. (2025) apresenta o MSLesSeg, um novo banco de dados de acesso público projetado para avançar na pesquisa de segmentação de lesões de Esclerose Múltipla (EM). O conjunto de dados contém 115 exames de ressonância magnética (RM) de 75 pacientes, incluindo sequências  $T1$ ,  $T2$  e  $FLAIR$ , além de dados clínicos complementares. As imagens foram adquiridas em diferentes *scanners* com intensidades de campo de 1.5 e 3 Tesla. Após o pré-processamento, que converteu os arquivos do formato DICOM original, todas as sequências de imagem e as máscaras de segmentação foram disponibilizadas no formato NIFTI.

Um protocolo de anotação manual, validado por especialistas, foi utilizado para gerar máscaras de segmentação das lesões, estabelecendo um conjunto de dados de referência confiável para avaliação de algoritmos. O *dataset* foi dividido em conjuntos de treinamento e teste. O conjunto de treinamento é composto por 53 pacientes e o de teste por 22 pacientes. O objetivo é fornecer um recurso para o desenvolvimento e a avaliação comparativa (*benchmarking*) de novas soluções de segmentação automática baseadas em Inteligência Artificial.

A máscara atua sobre a MRI, classificando cada *pixel* como pertencente ou não à região de interesse (lesão). A Figura 3.1 mostra um recorte axial de uma das ressonâncias presentes no *dataset*, exibindo a anotação *pixel* a *pixel* feita pelos especialistas e a sobreposição entre a máscara e as lesões.

Na etapa de pré-processamento, optou-se pela decomposição dos volumes 3D em cortes 2D axiais. Tal estratégia visa mitigar a alta demanda por recursos computacionais associada a arquiteturas tridimensionais, bem como reduzir a complexidade na síntese de estruturas anatômicas, facilitando a convergência da rede generativa. Em contrapartida, ao não utilizar de um processamento volumétrico completo, o método possui a limitação intrínseca de não preservar a continuidade espacial entre fatias adjacentes (coerência *inter-*



Figura 3.1: Exemplo de sobreposição de máscara e imagem de um corte no eixo  $z$  de um paciente com lesões de Esclerose Múltipla identificadas.

*slice*), priorizando a qualidade da síntese *intra-slice* em detrimento da consistência global do eixo  $z$ .

Também optou-se por utilizar a forma de aquisição em T2, devido a um equilíbrio de quantidade de imagens presentes no *dataset* e a capacidade de fornecer ao modelo informações robustas sobre as lesões.

Nesta etapa, também foram excluídas imagens com forte dependência do eixo  $y$ , que correspondem a cerca de 17% da base de dados original, uma vez que, como demonstrado na Figura 3.2, estas apresentam posicionamentos oculares ou recortes cerebrais substancialmente distintos do padrão predominante no conjunto de dados. Tais variações introduzem discrepâncias anatômicas relevantes que, na ausência de informações volumétricas tridimensionais, podem comprometer significativamente a capacidade de generalização do modelo gerativo.

## 3.2 Ambiente computacional

No estágio de pré-processamento dos dados, empregaram-se as bibliotecas NiBabel e NumPy para a leitura e manipulação dos arquivos no formato NIFTI, possibilitando a

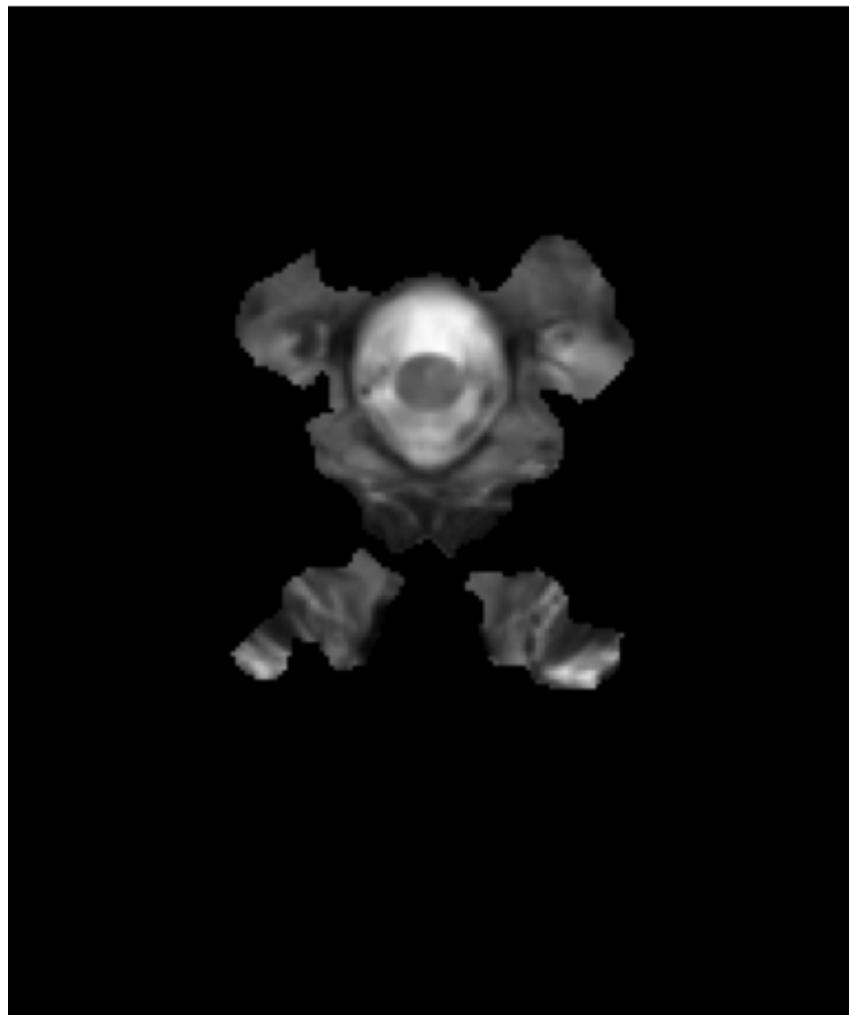


Figura 3.2: Exemplo de imagem fortemente dependente da posição ao longo do eixo  $y$ , na qual a ausência de informação espacial volumétrica pode levar o modelo a interpretações ambíguas, resultando em falhas de generalização ou colapso do aprendizado

extração e a conversão de cortes bidimensionais em imagens PNG. Para todas as instâncias do conjunto, selecionou-se o corte central no eixo  $z$ , tanto para as imagens quanto para as máscaras, as quais também foram convertidas para o formato PNG.

O desenvolvimento dos modelos foi conduzido utilizando as linguagens Python (versão 3.9.12) e Triton (versão 3.2.0), em conjunto com as bibliotecas PyTorch (versão 2.6) e MONAI (versão 2.4). Os experimentos foram realizados em um ambiente computacional com uma CPU Intel I7-11500 com 16 núcleos e uma GPU V100 com 32GB de VRAM.

### 3.3 Modelo de segmentação

Nos últimos anos, diversas arquiteturas de redes neurais têm sido investigadas para tarefas de segmentação de imagens médicas, impulsionadas pela necessidade de diagnósticos automatizados precisos e análises quantitativas de estruturas anatômicas. Nesse cenário, as abordagens derivadas da U-Net, uma rede convolucional baseada no paradigma *encoder-decoder* introduzida por Ronneberger, Fischer e Brox (2015), consolidaram-se como o estado da arte. Trabalhos como os de Pinto et al. (2024) e Alom et al. (2018) evidenciam a eficácia dessa topologia na segmentação de imagens médicas em variados subdomínios.

Neste trabalho, adotou-se a arquitetura SegResNet, proposta por Myronenko (2018). Trata-se de uma rede *encoder-decoder* otimizada para segmentação de alto desempenho, que integra os princípios das Redes Residuais (*ResNets*). Diferentemente da U-Net convencional, a SegResNet estrutura seus estágios de codificação e decodificação através de blocos residuais (*residual blocks*). Essa abordagem beneficia-se das conexões de atalho internas em cada bloco, facilitando o fluxo de gradientes durante o treinamento e viabilizando a construção de redes mais profundas sem a degradação de desempenho típica de modelos convolucionais extensos.

Adicionalmente, a SegResNet emprega a Normalização de Grupo (*Group Normalization*) em detrimento da Normalização de Lote (*Batch Normalization*). Essa é uma escolha estratégica para a segmentação médica 3D, onde o tamanho do lote (*batch size*) é frequentemente restrinrido pela memória disponível na GPU. As conexões de salto (*skip connections*) entre o codificador e o decodificador são preservadas, garantindo a retenção

de detalhes espaciais e combinando a robustez na extração de características da ResNet com a precisão de localização essencial para o delineamento de estruturas anatômicas complexas. A Tabela 3.3 mostra a configuração utilizada para o modelo e uma breve explicação de cada parâmetro.

Parâmetro	Valor	Descrição
<code>spatial_dims</code>	2	Dimensão espacial dos volumes (2D)
<code>in_channels</code>	1	Número de canais da entrada (Escala de cinza)
<code>out_channels</code>	2	Número de classes de saída (Lesão + Fundo)
<code>init_filters</code>	56	Número de filtros na primeira camada de convolução
<code>dropout_prob</code>	0,2	Probabilidade de <i>dropout</i>

Tabela 3.1: Configuração do modelo de segmentação através da biblioteca MONAI.

### 3.3.1 Aumento de dados clássico

Considerando a elevada variabilidade anatômica interpaciente e a limitada disponibilidade de anotações densas *pixel-a-pixel*, foi implementado um *pipeline* clássico de aumento de dados estocástico (*data augmentation*), aplicado dinamicamente durante o treinamento. Inicialmente, as imagens passam por um processo de normalização de intensidade, no qual os valores dos *voxels* são reescalados para o intervalo [0, 1], com aplicação de *clipping* baseado em percentis para reduzir a influência de valores extremos.

Para mitigar o severo desequilíbrio entre regiões lesionadas e tecido saudável, adotou-se uma estratégia de amostragem balanceada por *patches*. Foram extraídos *patches* bidimensionais de dimensão  $128 \times 128$ , utilizando-se uma razão de amostragem de 7:1 entre centros localizados em regiões de lesão (*foreground*) e regiões de fundo (*background*), favorecendo a apresentação de exemplos informativos ao modelo durante o treinamento.

Adicionalmente, aplicaram-se transformações geométricas aleatórias com o objetivo de aumentar a invariância espacial do modelo. Essas transformações incluem rotações discretas de múltiplos de 90 graus, espelhamentos horizontais e verticais, bem como transformações afins contínuas (como rotação, translação e escala) aplicadas com probabilidade de 0,3. Por fim, foram introduzidas perturbações fotométricas para simular variações de aquisição, incluindo a injeção de ruído Gaussiano com média zero e desvio padrão  $\sigma = 0,01$ , além de variações aleatórias de intensidade e escala de brilho.

### 3.3.2 Formulação da Função de Perda

O treinamento do modelo foi conduzido por meio da minimização de uma função de perda híbrida, denotada por  $\mathcal{L}_{total}$ , projetada especificamente para lidar com o severo desequilíbrio de classes característico da segmentação de lesões pequenas em imagens de Ressonância Magnética. Essa função combina a perda Dice, sensível à sobreposição espacial, com a perda Focal, que enfatiza exemplos de difícil classificação. Formalmente, a função de perda total é definida como:

$$\mathcal{L}_{total}(y, \hat{y}) = \mathcal{L}_{Dice}(y, \hat{y}) + \mathcal{L}_{Focal}(y, \hat{y}). \quad (3.1)$$

**Perda Dice ( $\mathcal{L}_{Dice}$ )** Na tarefa de segmentação, o objetivo central consiste em maximizar a sobreposição entre o mapa de predição  $P$  produzido pelo modelo e a máscara de referência  $G$  (*ground truth*). Para essa finalidade, adotou-se como métrica principal o Coeficiente de Dice, originalmente proposto em Dice (1945) e posteriormente consolidado no contexto de segmentação de lesões médicas em Zijdenbos et al. (1994). Essa métrica quantifica a similaridade entre dois conjuntos de *pixels*, sendo definida como:

$$Dice(P, G) = \frac{2 \cdot |P \cap G|}{|P| + |G|},$$

onde  $|P|$  representa o número de *pixels* classificados como pertencentes à classe positiva pelo modelo,  $|G|$  corresponde ao número de *pixels* positivos na máscara verdadeira, e  $|P \cap G|$  denota a interseção entre ambos, isto é, os *pixels* corretamente segmentados.

Para viabilizar o uso do Coeficiente de Dice como função de perda no treinamento de redes neurais profundas, empregou-se uma formulação contínua e diferenciável. Seja  $p_i \in [0, 1]$  a probabilidade predita para o *pixel*  $i$  e  $g_i \in \{0, 1\}$  o valor correspondente na máscara de referência. O Dice suavizado é então expresso por:

$$Dice(P, G) = \frac{2 \sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \epsilon},$$

em que  $N$  denota o número total de *pixels* da imagem e  $\epsilon$  é um termo de regularização introduzido para evitar instabilidades numéricas decorrentes de divisões por zero.

Como o processo de treinamento visa à minimização de uma função objetivo, a perda Dice é definida como o complemento do coeficiente de similaridade, isto é:

$$\mathcal{L}_{Dice}(P, G) = 1 - Dice(P, G).$$

Dessa forma, valores de perda próximos de zero indicam elevada sobreposição entre predição e máscara verdadeira, enquanto valores mais elevados refletem discrepâncias significativas entre ambas.

**Perda Focal ( $\mathcal{L}_{Focal}$ )** Com o objetivo de lidar explicitamente com o desequilíbrio de classes e reduzir a influência de exemplos facilmente classificados, foi incorporada à função de perda a perda Focal. Essa abordagem atribui maior peso aos exemplos de difícil classificação (*hard examples*), atenuando a contribuição de *pixels* corretamente classificados com alta confiança. A formulação empregada inclui um fator de ponderação de classe  $\alpha$ , calculado dinamicamente a partir da razão entre o número de *pixels* negativos e positivos no conjunto de treinamento,  $w_{pos} = N_{neg}/N_{pos}$ , bem como um parâmetro de foco fixado em  $\gamma = 2,0$ . A perda Focal é definida como:

$$\mathcal{L}_{Focal} = -\alpha(1 - \hat{y}_t)^\gamma \log(\hat{y}_t), \quad (3.2)$$

onde  $\hat{y}_t$  representa a probabilidade predita associada à classe verdadeira do *pixel* em análise.

## 3.4 Geração artificial de imagens

A abordagem proposta para geração artificial de imagens baseia-se na combinação de um gerador condicional do tipo ResNet, com o objetivo de sintetizar imagens de Ressonância Magnética (MRI) a partir de mapas de segmentação de lesões. O *pipeline* é formulado como um modelo de *image-to-image translation*, no qual a entrada consiste em um mapa semântico binário que codifica a presença e a localização espacial das lesões, enquanto a saída corresponde a uma fatia de MRI sintética, apresentando contraste e texturas coerentes com a anatomia condicionada.

Essa formulação permite controlar explicitamente a distribuição espacial das lesões geradas, ao mesmo tempo em que preserva características anatômicas realistas, tornando o método particularmente adequado para aplicações de aumento de dados em tarefas de segmentação supervisionada.

### 3.4.1 *Variational Autoencoder*

Para a modelagem generativa das máscaras binárias de lesões e o aprendizado de sua distribuição subjacente, foi empregado um *Variational Autoencoder* (VAE) de natureza convolucional. O modelo é composto por duas sub-redes principais: um codificador probabilístico, também denominado *Inference Network*, que aproxima a distribuição posterior  $q_\phi(z|x)$ , e um decodificador gerador responsável por modelar a distribuição condicional  $p_\theta(x|z)$ .

A rede codificadora recebe como entrada uma máscara binária  $x \in \mathbb{R}^{H \times W \times C}$  e a processa por meio de uma sequência hierárquica de blocos convolucionais. Em contraste com abordagens baseadas em operações de *pooling* determinísticas, optou-se pelo uso de convoluções estratificadas (*strided convolutions*) com *stride* igual a 2, permitindo a redução progressiva da resolução espacial de forma aprendível. Cada bloco convolucional é composto por uma convolução 2D com *kernel*  $3 \times 3$  e *padding* unitário, seguida de uma camada de *Batch Normalization*, responsável por estabilizar a distribuição interna das ativações, e por uma função de ativação *LeakyReLU* com coeficiente  $\alpha = 0,2$ , adotada para mitigar o problema conhecido como *dying ReLU*.

Ao final do processo de codificação, o mapa de características é linearizado e projetado em duas camadas densas paralelas, responsáveis por estimar os parâmetros da distribuição latente: o vetor de médias  $\mu$  e o vetor de log-variâncias  $\log(\sigma^2)$ . O espaço latente  $z$  é modelado como uma distribuição Gaussiana multivariada, o que permite uma representação contínua e regularizada das máscaras de entrada.

Para possibilitar a retropropagação do erro através do processo estocástico de amostragem do espaço latente, foi empregado o truque de reparametrização (*reparameterization trick*). Considerando um ruído auxiliar  $\epsilon$  amostrado de uma distribuição normal padrão  $\mathcal{N}(0, I)$ , a variável latente é obtida segundo a seguinte expressão:

$$z = \mu + \sigma \odot \epsilon. \quad (3.3)$$

O decodificador realiza o mapeamento inverso, projetando o vetor latente  $z$  de volta para uma representação espacial de características profundas e, em seguida, aplicando uma sequência de convoluções transpostas (*transposed convolutions*) para recuperar gradualmente a resolução original da imagem. A arquitetura do decodificador espelha a do codificador, fazendo uso de *Batch Normalization* e funções de ativação *LeakyReLU*. A camada final aplica uma convolução transposta que projeta as características aprendidas para o número de canais original, produzindo *logits* não normalizados, os quais representam os parâmetros de uma distribuição de Bernoulli associada a cada *pixel* da máscara reconstruída.

### 3.4.2 Otimização Variacional e Termos de Regularização

O processo de treinamento do modelo gerativo foi formulado como um problema de maximização do limite inferior variacional da log-verossimilhança dos dados, conhecido como *Evidence Lower Bound* (ELBO). Nesse contexto, a função objetivo busca equilibrar a fidelidade de reconstrução das amostras observadas com a regularização do espaço latente, garantindo propriedades desejáveis de generalização e amostragem. A função de perda total  $\mathcal{L}$  é composta por um termo de reconstrução, um termo de divergência de Kullback–Leibler e um termo adicional de regularização baseado em conhecimento de domínio, conforme definido a seguir:

$$\mathcal{L}(x, \phi, \theta) = \mathcal{L}_{rec} + \beta \cdot \mathcal{L}_{KL} + \lambda_{area} \cdot \mathcal{L}_{area}. \quad (3.4)$$

O termo de reconstrução  $\mathcal{L}_{rec}$  quantifica o grau de similaridade entre a amostra gerada pelo decodificador e a entrada original. Considerando que os dados de entrada são máscaras binárias, a saída do modelo foi interpretada como os parâmetros de uma distribuição de Bernoulli multivariada, sendo a perda de reconstrução calculada por meio da *Binary Cross-Entropy* (BCE) aplicada aos *logits* produzidos pelo decodificador. Formalmente, esse termo é definido como:

$$\mathcal{L}_{rec} = - \sum_{i=1}^N [x_i \log(\sigma(y_i)) + (1 - x_i) \log(1 - \sigma(y_i))], \quad (3.5)$$

onde  $\sigma(\cdot)$  denota a função sigmoide,  $y_i$  corresponde aos *logits* de saída do decodificador e  $N$  representa o número total de *pixels* da máscara.

O termo de divergência de Kullback–Leibler  $\mathcal{L}_{KL}$  atua como um mecanismo de regularização do espaço latente, forçando a distribuição posterior aproximada  $q_\phi(z|x)$  a se aproximar de uma distribuição a priori Gaussiana isotrópica  $\mathcal{N}(0, I)$ . No entanto, em arquiteturas variacionais profundas, é comum a ocorrência do fenômeno conhecido como *posterior collapse*, no qual o decodificador passa a ignorar o código latente e a divergência KL tende prematuramente a zero. Para mitigar esse problema, foi empregada a técnica de *Free Bits*, que impõe um limite inferior  $\lambda_{free}$  à contribuição da divergência KL em cada dimensão latente  $j$ . Assim, o termo é definido como:

$$\mathcal{L}_{KL} = \sum_{j=1}^D \max(D_{KL}(q_\phi(z_j|x) || p(z_j)), \lambda_{free}), \quad (3.6)$$

garantindo que cada dimensão do vetor latente retenha uma quantidade mínima de informação, expressa em *nats*, e incentivando o uso efetivo de todo o espaço latente.

Além dos termos clássicos do ELBO, foi incorporada uma regularização adicional baseada em conhecimento de domínio, representada pelo termo  $\mathcal{L}_{area}$ . Esse termo visa evitar a geração de máscaras degeneradas, como mapas completamente vazios ou excessivamente preenchidos, penalizando desvios em relação a uma fração de área alvo previamente definida. A penalização é calculada por meio da norma L1 entre a média das probabilidades previstas e o valor de referência estabelecido, atuando como um mecanismo de controle estatístico da extensão espacial das lesões sintetizadas.

### 3.4.3 Estratégia de Treinamento

A otimização dos parâmetros do modelo foi realizada utilizando o algoritmo *AdamW*, que combina momentos adaptativos de primeira e segunda ordem com um termo explícito de *weight decay*, promovendo regularização L2 mais eficaz em redes profundas. Com o objetivo de aumentar a estabilidade do treinamento e favorecer a convergência do modelo

variacional, duas estratégias adicionais foram adotadas.

A primeira consiste no uso de  $\beta$ -*scheduling*, também conhecido como *annealing* do termo KL. Nessa estratégia, o coeficiente  $\beta$  associado à divergência KL é inicialmente definido como zero e incrementado de forma linear até o valor unitário ao longo das épocas iniciais de treinamento. Esse procedimento permite que o modelo priorize a aprendizagem da reconstrução dos dados nas fases iniciais, antes de ser progressivamente forçado a estruturar o espaço latente de acordo com a distribuição a priori.

A segunda estratégia envolve a aplicação de *gradient norm clipping*, na qual a norma dos gradientes é limitada a um valor máximo pré-definido durante a retropropagação. Essa técnica reduz o risco de explosão de gradientes, um problema recorrente em arquiteturas gerativas profundas, contribuindo para maior estabilidade numérica ao longo do processo de otimização.

## 3.5 Modelo Generativo Adversarial

Para a tarefa de tradução imagem-a-imagem, na qual mapas de segmentação binários  $x$  são mapeados para imagens sintéticas de Ressonância Magnética  $y$ , foi empregada uma Rede Generativa Adversarial Condisional (*Conditional Generative Adversarial Network* – cGAN). Essa abordagem é formulada como um jogo adversarial de soma zero entre dois modelos: um gerador  $G$ , responsável por sintetizar imagens que sigam a distribuição dos dados reais condicionadas às máscaras de entrada, e um discriminador  $D$ , treinado para distinguir entre pares reais  $(x, y)$  e pares sintéticos  $(x, G(x))$ .

O gerador adota uma arquitetura baseada em *ResNet*, organizada em três estágios principais: codificação, transformação e decodificação. No estágio de codificação, a resolução espacial da entrada é progressivamente reduzida por meio de convoluções com *stride* igual a 2, seguidas de Normalização de Instância (*Instance Normalization*) e funções de ativação ReLU. O núcleo da rede é composto por nove blocos residuais em sequência, os quais permitem o aprendizado de dependências globais profundas sem degradação do gradiente, favorecendo a estabilidade do treinamento. O estágio de decodificação emprega convoluções transpostas para restaurar a resolução original da imagem, fixada em  $256 \times 256$ , sendo a camada final seguida por uma função de ativação *Tanh*, responsável

por mapear os valores de saída para o intervalo  $[-1, 1]$ . Com o objetivo de estabilizar o processo de inferência e reduzir oscilações durante o treinamento adversarial, foi aplicada a técnica de Média Móvel Exponencial (*Exponential Moving Average – EMA*) aos pesos do gerador, utilizando um fator de decaimento  $\mu = 0,999$ .

O discriminador emprega uma arquitetura multiescala, composta por três discriminadores estruturalmente idênticos que operam sobre diferentes resoluções da imagem de entrada, correspondentes à resolução original, à metade e a um quarto da resolução original. Cada um desses discriminadores segue o paradigma *PatchGAN*, no qual regiões locais da imagem são classificadas como reais ou sintéticas, incentivando a coerência de texturas e detalhes de alta frequência. Para garantir a estabilidade do treinamento adversarial e mitigar problemas como colapso de modo, todas as camadas convolucionais do discriminador foram regularizadas por meio de Normalização Espectral (*Spectral Normalization*), assegurando o controle da constante de Lipschitz da função discriminadora.

### 3.5.1 Função Objetivo Composta

A otimização do modelo adversarial é conduzida por uma função de perda composta, que combina o objetivo adversarial com termos adicionais de fidelidade de reconstrução e coerência perceptual. A função de perda total associada ao gerador é definida como:

$$\mathcal{L}_{total} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{char}\mathcal{L}_{char} + \lambda_{perc}\mathcal{L}_{VGG} + \lambda_{fm}\mathcal{L}_{FM}. \quad (3.7)$$

O termo adversarial  $\mathcal{L}_{adv}$  foi formulado a partir da *Hinge Loss*, em substituição à entropia cruzada tradicional, uma vez que essa formulação penaliza apenas amostras que violam a margem de decisão, resultando em gradientes mais estáveis e robustos durante o treinamento. A função de perda do discriminador é dada por:

$$\mathcal{L}_D = \mathbb{E}_{(x,y)} [\max(0, 1 - D(x, y))] + \mathbb{E}_x [\max(0, 1 + D(x, G(x)))] , \quad (3.8)$$

enquanto a perda adversarial associada ao gerador é definida como:

$$\mathcal{L}_G = -\mathbb{E}_x [D(x, G(x))]. \quad (3.9)$$

Para garantir fidelidade no nível de *pixel*, foi incluída a Perda de Charbonnier  $\mathcal{L}_{char}$ , uma variante suavizada e diferenciável da norma  $\ell_1$ , reconhecida por sua robustez a *outliers* e por evitar instabilidades numéricas em regiões próximas de zero. Esse termo é definido como:

$$\mathcal{L}_{char}(G) = \mathbb{E}_{(x,y)} \left[ \sqrt{\|y - G(x)\|^2 + \epsilon^2} \right], \quad (3.10)$$

onde  $\epsilon = 10^{-3}$  é um parâmetro de suavização.

Adicionalmente, foram incorporados termos de perda perceptual e de *feature matching* com o objetivo de melhorar a qualidade visual das imagens geradas e reduzir o efeito de borramento associado a perdas puramente *pixel-a-pixel*. A Perda Perceptual  $\mathcal{L}_{VGG}$  é calculada como a distância L1 entre mapas de características extraídos de diferentes camadas de uma rede VGG-19 pré-treinada no conjunto ImageNet. Complementarmente, a Perda de *Feature Matching*  $\mathcal{L}_{fm}$  minimiza a discrepância estatística entre as ativações intermediárias do discriminador para imagens reais e sintéticas, forçando o gerador a produzir estruturas cujas estatísticas sejam consistentes com as observadas nos dados reais em múltiplas escalas.

A Tabela 3.5.1 apresenta os pesos (hiperparâmetros  $\lambda$ ) utilizados para ponderar a contribuição de cada componente na função de perda total da Rede Generativa Adversarial (GAN). Estes valores foram definidos empiricamente para balancear a qualidade visual, a consistência estrutural e a estabilidade do treinamento.

Componente da Perda	Símbolo	Peso
Reconstrução ( $L_1$ /Charbonnier)	$\lambda_{char}$	20.0
Feature Matching	$\lambda_{fm}$	5.0
Adversarial	$\lambda_{adv}$	5.0
Perceptual (VGG)	$\lambda_{perc}$	1.0

Tabela 3.2: Pesos utilizados para as funções de perda do modelo Generativo Adversarial

### 3.5.2 Detalhes de Implementação

O treinamento do modelo foi conduzido utilizando o otimizador Adam sob a regra de atualização em duas escalas de tempo (*Two-Time-Scale Update Rule* – TTUR), adotando-se uma taxa de aprendizado de  $1 \times 10^{-4}$  para o gerador e de  $2 \times 10^{-4}$  para o discriminador. Para reduzir o consumo de memória e acelerar o treinamento, foi empregada precisão mista automática (*Automatic Mixed Precision*). O aumento de dados aplicado durante o treinamento incluiu transformações geométricas, como rotações, espelhamentos e recortes, bem como perturbações não lineares de intensidade, aplicadas exclusivamente às imagens reais com o objetivo de aumentar a robustez do modelo frente a variações de contraste e condições de aquisição.

## 4 Resultados

Esse capítulo apresenta os resultados obtidos para cada um dos modelos que compõem o pipeline, bem como o modelo de segmentação que irá ser utilizado para verificação do impacto da metodologia de aumento de dados proposta no treinamento desse modelo.

### 4.1 Variational Autoencoder

O modelo foi treinado por 600 épocas, *batch size* utilizado foi 4. As Figuras 4.1, 4.2 e 4.3 mostram as perdas associadas ao conjunto de treino fornecido ao modelo. O melhor modelo contudo é definido pela menor *loss* em relação ao conjunto de validação, devido a possibilidade de *overfitting* em relação ao conjunto de treino disponível, isto é, a não generalização em relação a outras possíveis entradas.

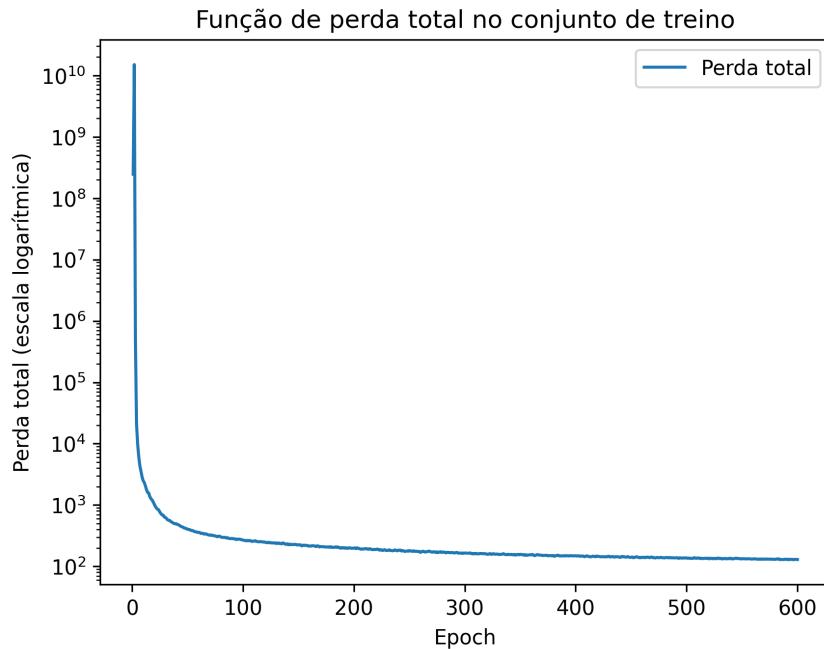


Figura 4.1: Valores da função de perda total no conjunto de treino.

Os resultados em relação ao conjunto de validação presente nas Figuras 4.4, 4.5 e 4.6 demonstram esse comportamento, visto que após certa época, mesmo com a redução da perda dentro do conjunto de treino há um aumento dentro do conjunto de validação,

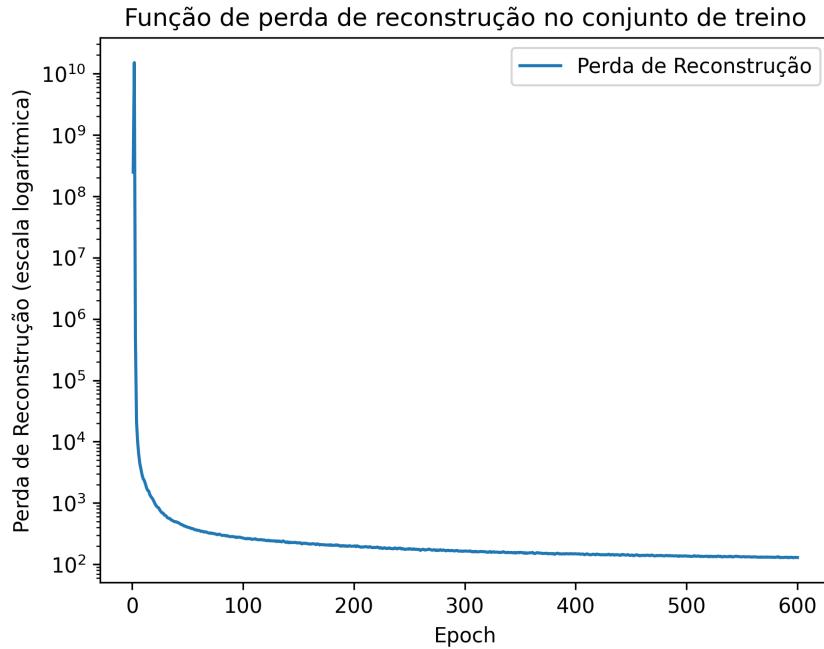


Figura 4.2: Valores da função de perda de reconstrução no conjunto de treino.

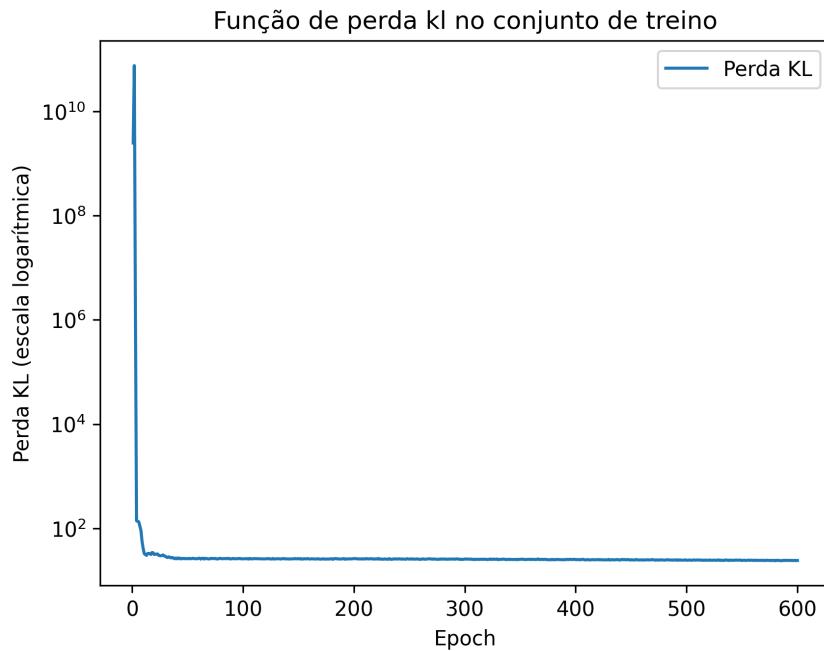


Figura 4.3: Valores da função de perda kl de reconstrução no conjunto de treino.

evidenciando que a melhora em relação ao conjunto de treino se deve a um ajuste excessivo em relação a esses dados.

Com o modelo do VAE foi possível gerar máscaras falsas que se aproximem em forma e distribuição espacial das máscaras reais feitas pelos especialistas, a Figura 4.7 mostra três máscaras geradas pelo método proposto, que poderiam servir de entrada para a inferência da Rede Generativa.

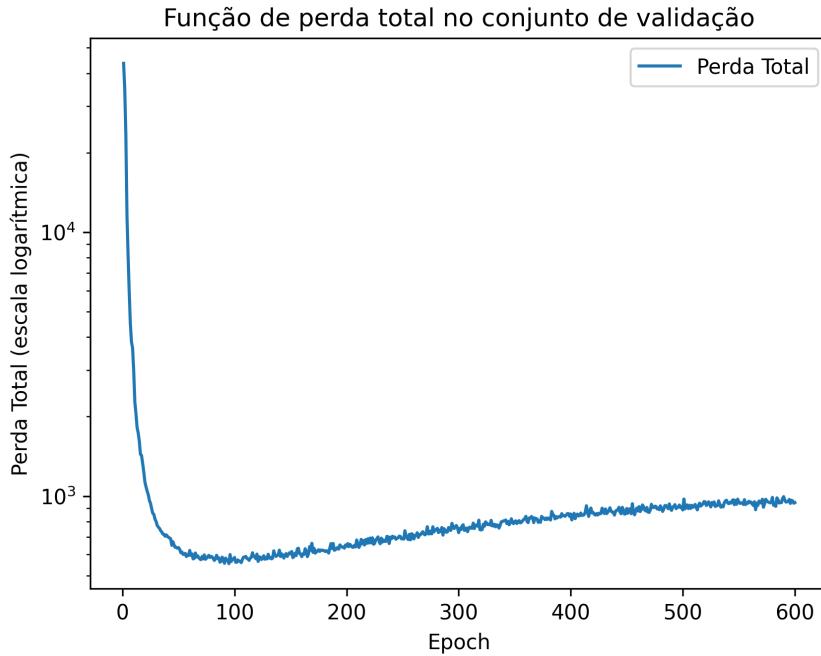


Figura 4.4: Valores da função de perda total no conjunto de validação.

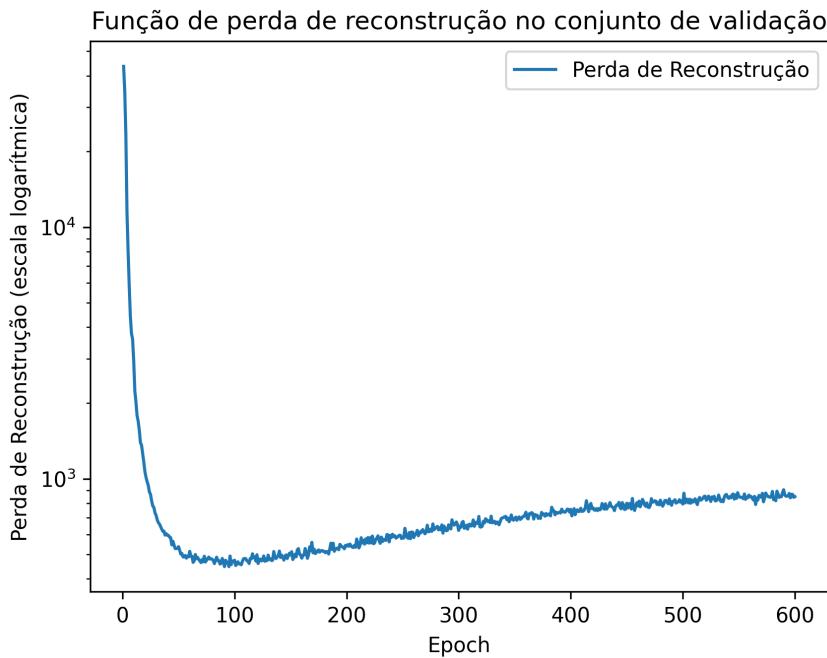


Figura 4.5: Valores da função de perda de reconstrução no conjunto de validação.

## 4.2 GAN

As Figuras 4.8 e 4.9 apresentam, respectivamente, a evolução das funções de perda total do Gerador e do Discriminador ao longo das épocas. Observa-se uma tendência consistente de minimização na perda do Gerador, evidenciando sua capacidade progressiva de sintetizar amostras verossímeis, indistinguíveis das reais. Simultaneamente, a esta-

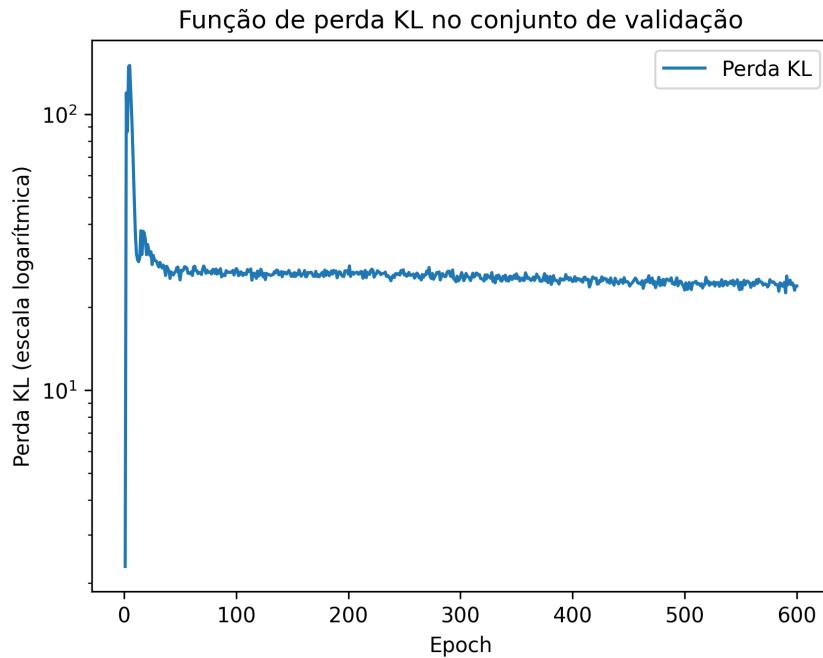


Figura 4.6: Valores da função de perda de reconstrução no conjunto de validação

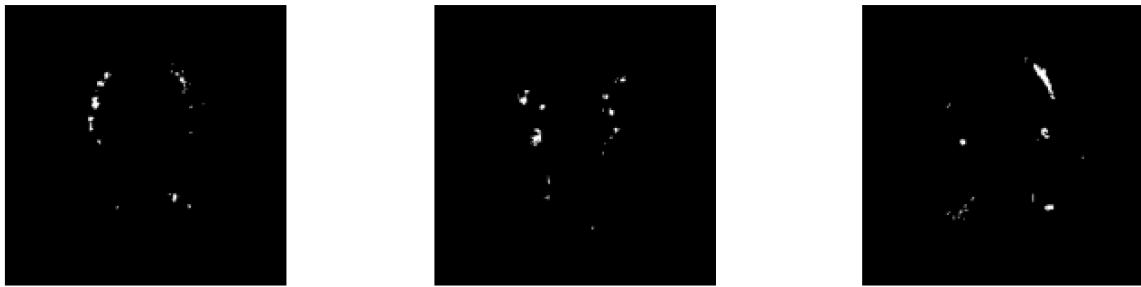


Figura 4.7: Exemplo de máscaras falsas geradas pelo Variational Autoencoder

bilidade observada na perda do Discriminador indica que o modelo adversário mantém sua robustez, não sendo superado trivialmente. A partir da época 170, o sistema exibe um comportamento de convergência, sugerindo que o treinamento atingiu um ponto de estabilidade competitiva entre as redes.

A Figura 4.10 demonstra a comparação da textura cerebral criada pela GAN com a respectiva imagem real em relação a uma mesma máscara. Já a figura demonstra um conjunto de texturas produzidas a partir de máscaras artificiais de lesão, bem como a sobreposição da lesão na textura.

Os resultados também demonstram que o Gerador tende a se confundir mais quando as áreas de interesse são muito pequenas, produzindo alguns artefatos na textura final, artefatos esses que podem ser identificados na imagem 4.12. O fenômeno observado

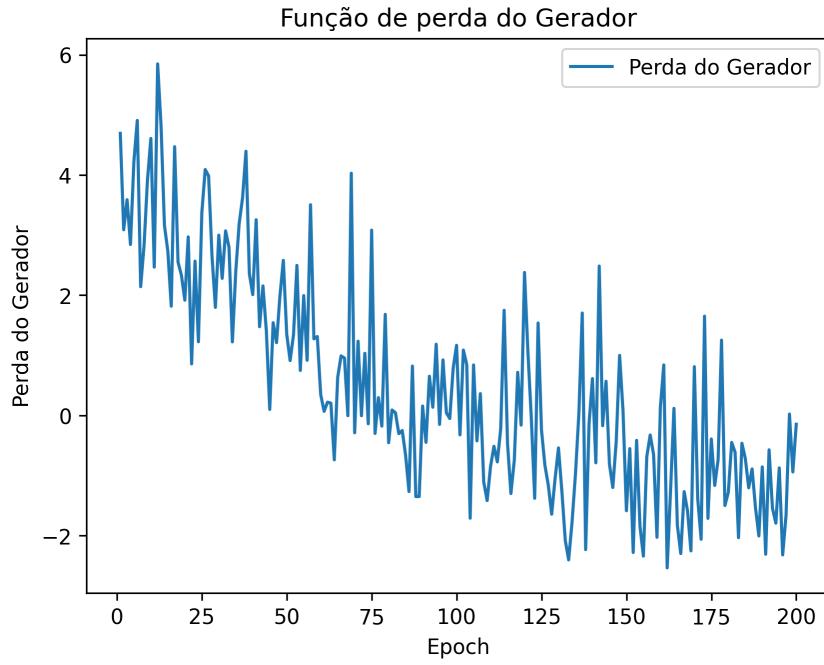


Figura 4.8: Perda total do Gerador ao longo das épocas.

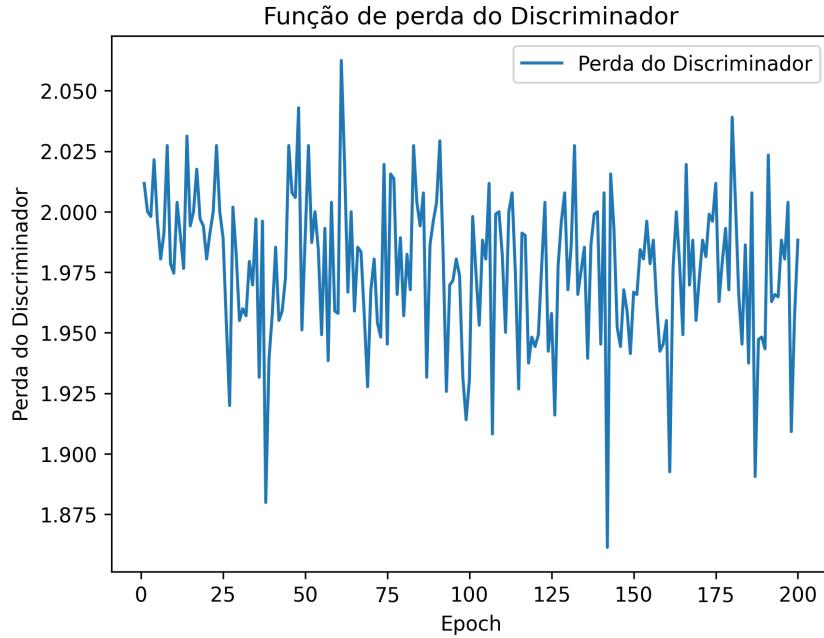


Figura 4.9: Perda total do discriminador ao longo das épocas.

é inerente à arquitetura de geração condicional proposta. Devido às operações de *downsampling*, a representação espacial de lesões diminutas tende a se dissipar nos mapas de características mais profundos. Consequentemente, o discriminador falha em penalizar a ausência dessas estruturas, validando a textura gerada com base na coerência global da imagem, uma vez que a contribuição da lesão para a distribuição estatística total é marginal.

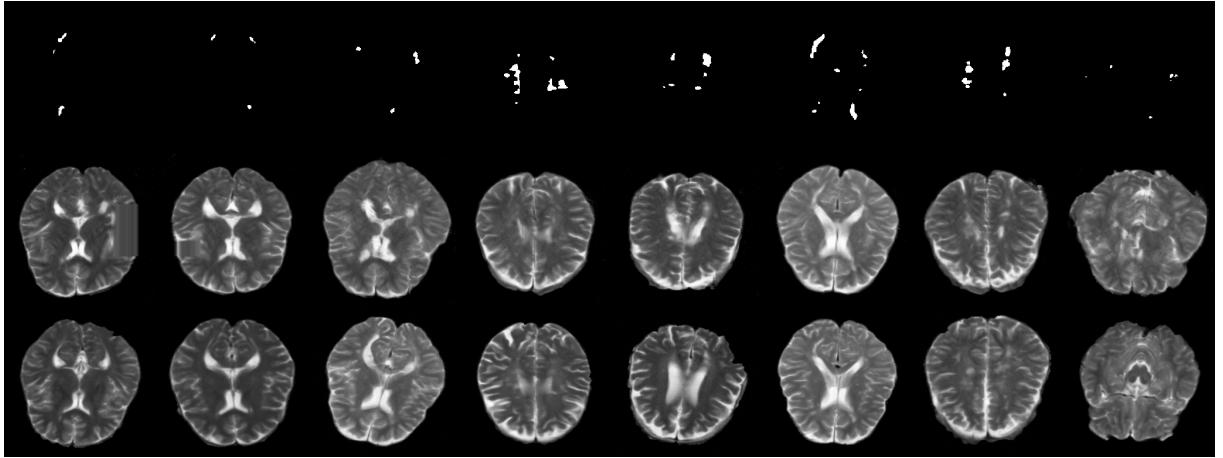


Figura 4.10: Comparação de texturas reais do cérebro com texturas geradas pela Rede Generativa Adversarial, sendo, a primeira linha a máscara (vinda diretamente do dataset original), segunda linha a textura falsa e a terceira o corte real.

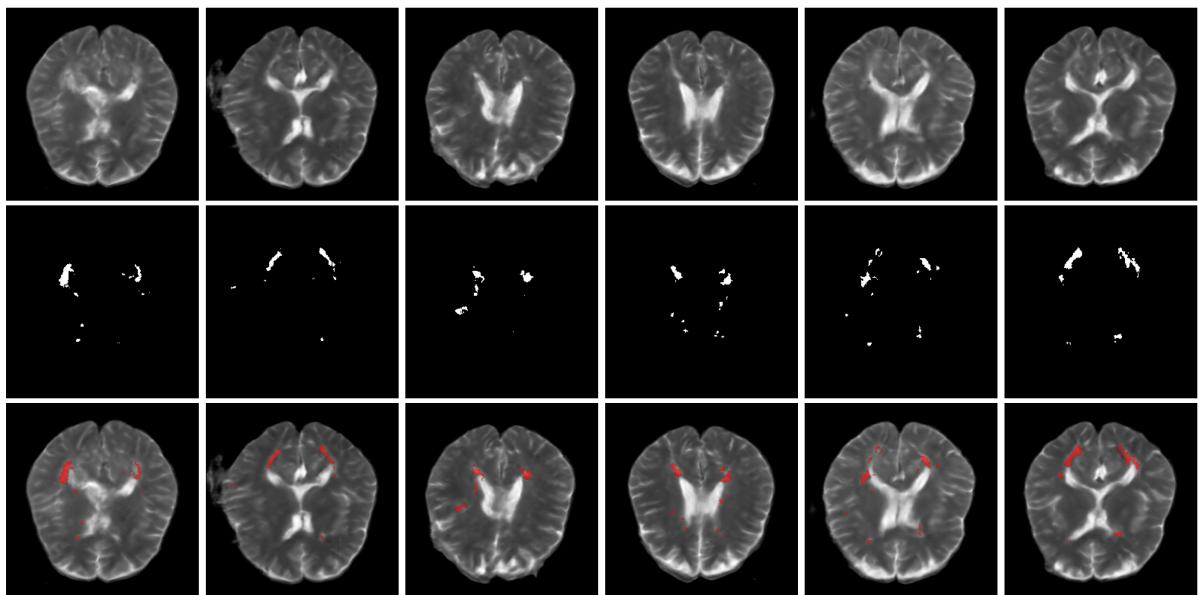


Figura 4.11: Comparação de textura gerada a partir de máscaras falsas produzidas através do Variational Autoencoder.

### 4.3 Modelo de segmentação

O modelo de segmentação foi treinado utilizando 2600 épocas de treinamento, com um *batch size* de 32, onde a cada 65 épocas foi executada uma etapa de validação. O alto número de épocas foi utilizado para compensar o *batch size* menor, visto limitações na VRAM da GPU. Os resultados do modelo treinado em cima da base de dados padrão estão apresentados na figura 4.13, esses resultados demonstram que o treinamento com os dados previamente disponíveis demonstram já uma capacidade de generalização próxima do esperado para tarefas de segmentação de Esclerose Múltipla, como descrito por (COM-

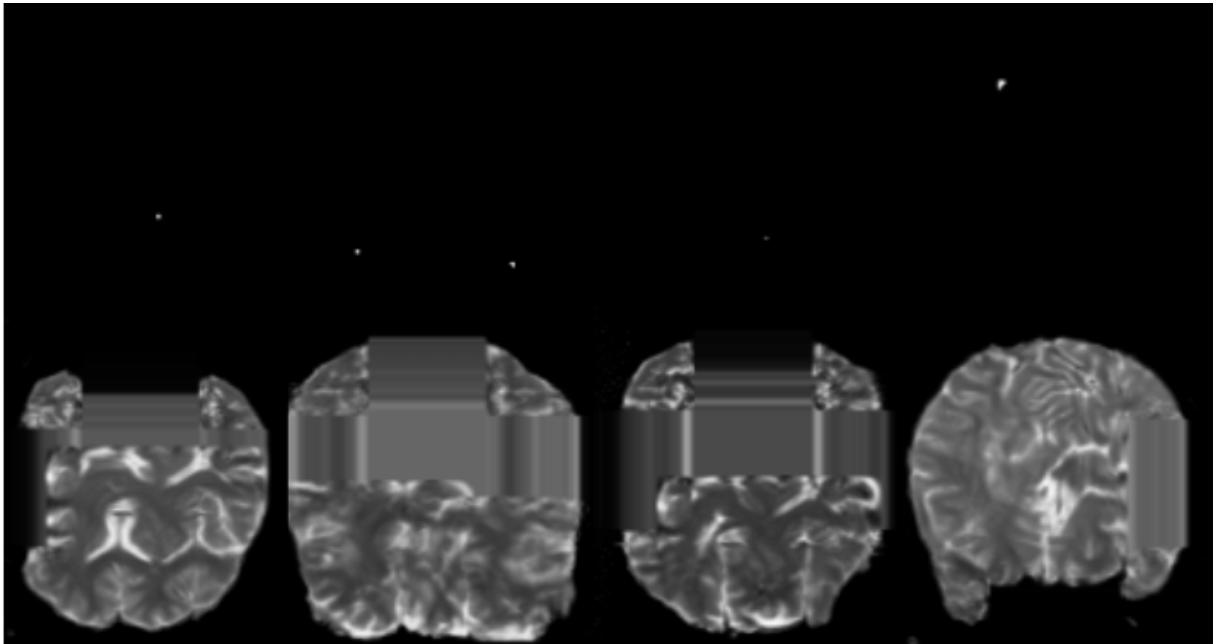


Figura 4.12: Imagens com artefatos visuais devido ao pequeno tamanho das lesões contidas na máscara de condicionamento

MOWICK et al., 2018)

A incorporação de imagens sintéticas geradas pela Rede Generativa resultou em um incremento considerável na métrica Dice final, conforme evidenciado no gráfico da Figura 4.16. Outra métrica relevante a ser avaliada é o número de lesões falsas marcadas na imagem, que seria a precisão(*ou precision*), a Tabela 4.1 mostra o resultado geral dessa métrica para cada modelo, bem como o resultado separado por quartis. Ressalta-se que a avaliação utilizou o mesmo conjunto de validação empregado no modelo treinado apenas com o dataset original, garantindo a comparabilidade dos resultados. A Tabela 4.2 sumariza o desempenho obtido com diferentes volumes de data augmentation. O melhor resultado foi alcançado com a adição de 750 imagens; volumes superiores não resultaram em ganho de desempenho, sugerindo uma saturação no aprendizado. Isso ocorre, primordialmente, devido à persistência da dificuldade em segmentar lesões pequenas. Tal limitação é ilustrada nos boxplots e análises gráficas das Figura 4.15, que evidenciam a dificuldade de generalização do modelo para estruturas menores. Esse comportamento corrobora os achados da Seção 4.2, onde observou-se que a incapacidade da GAN em sintetizar fielmente lesões minúsculas impede que os dados artificiais contribuam para a resolução deste problema específico.

Além disso, avaliou-se também o custo computacional associado a inclusão de

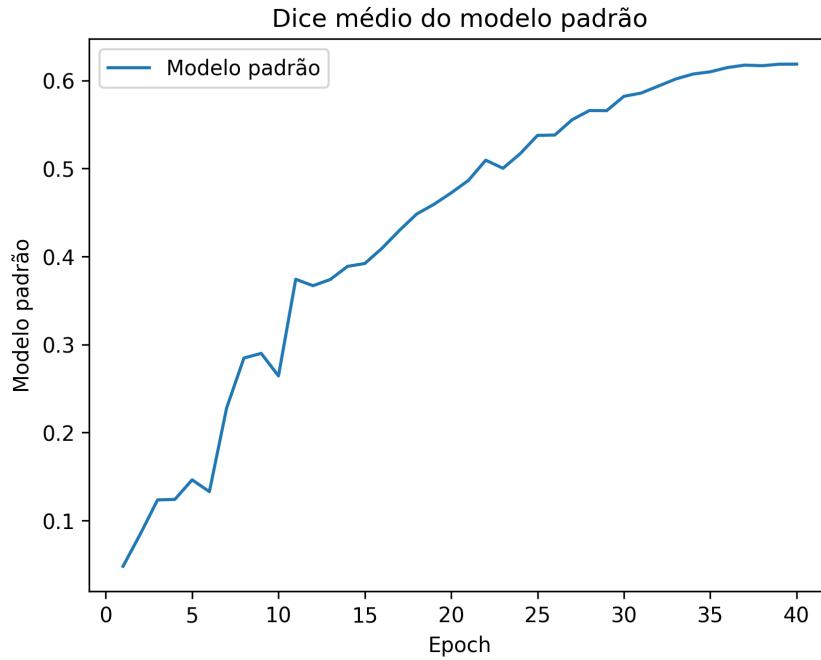


Figura 4.13: Métrica de Dice associadas ao modelo treinado com conjunto de dados padrão

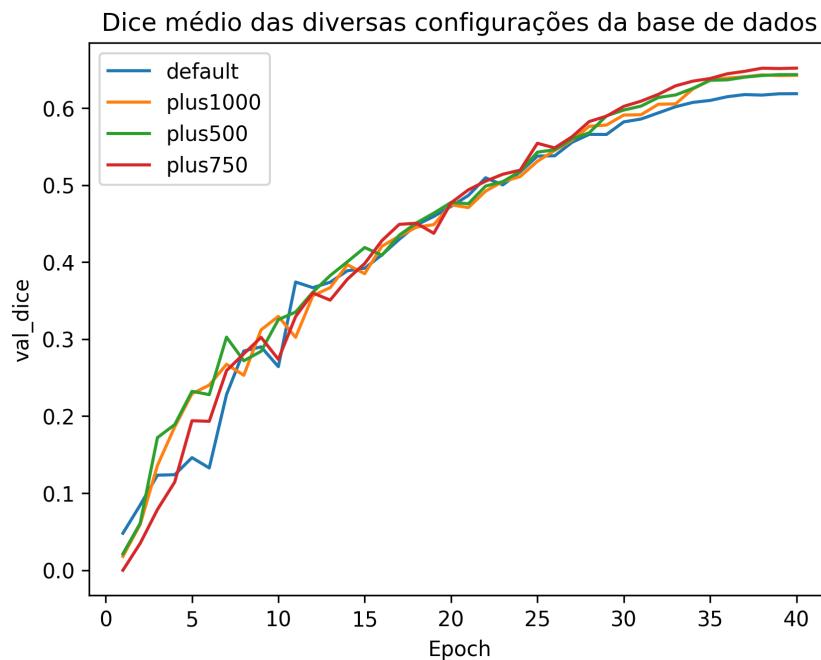


Figura 4.14: Comparação entre diferentes modelos, onde o default é o treinado no dataset padrão, um com 500 entradas extras e outro com 1000 entradas extras

novas imagens no dataset, a Tabela 4.3 demonstra que, embora o tempo de treinamento não mude consideravelmente entre cada suíte de treinos, o custo de memória se expande conforme aumenta-se o número de imagens presentes na base de treino.

O tempo de treinamento, como demonstrado na Tabela 4.3 não teve um aumento significativo porque nesse trabalho optou-se por definir o critério de parada do treinamento

Modelo	Ruído Médio Global	Ruído Q1	Ruído Q2	Ruído Q3	Ruído Q4
Padrão	23,44%	52,21%	27,91%	20,40%	11,83%
+500	22,68%	50,25%	28,38%	19,22%	10,86%
+750	21,41%	44,66%	26,11%	19,69%	11,11%
+1000	23,31%	50,33%	29,60%	21,44%	11,13%

Tabela 4.1: Precisão em relação as lesões demarcadas em porcentagem(Ruído), sendo a taxa de lesões segmentadas que não existe nenhuma correspondencia na máscara real.

Modelo	Dice	Dice Global Médio	Ruído Total Médio
Padrão	62,028%	80,85%	23,44%
+500	64,034%	82,14%	22,68%
+750	65,012%	82,51%	21,41%
+1000	64,038%	81,33%	23,31%

Tabela 4.2: Valores obtidos para o as principais métricas nas diferentes configurações de treino

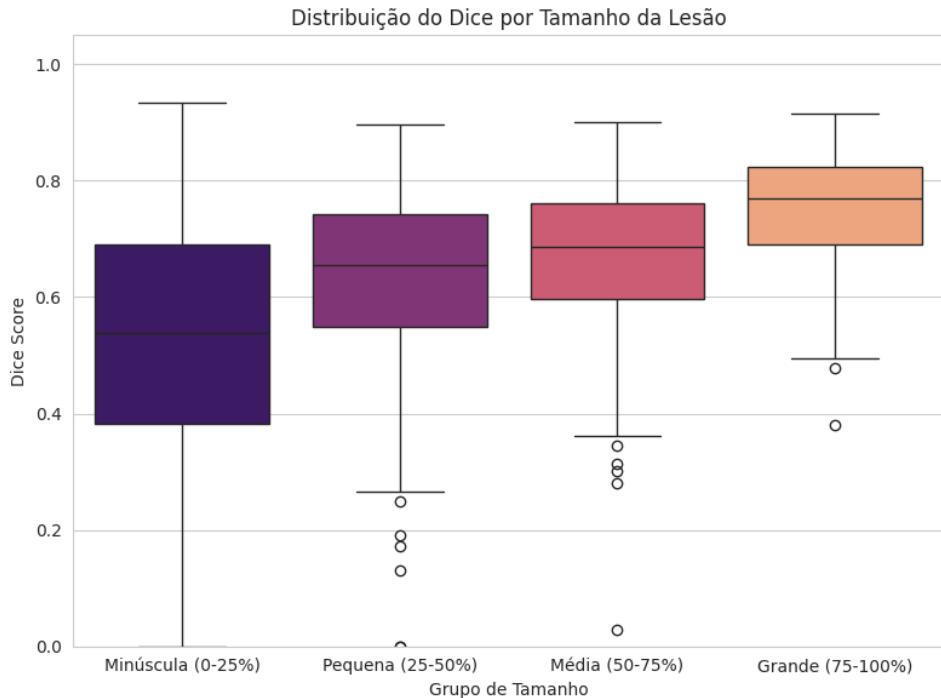


Figura 4.15: Boxplot da métrica de dice médio dos quartis do conjunto de validação, avaliado no melhor modelo obtido

com base em um número fixo de iterações globais, em detrimento de um número fixo de épocas completas. Dessa forma, a carga computacional total permaneceu constante, uma vez que o número de atualizações de pesos da rede independe do tamanho total do conjunto de dados.

O consumo de memória, por sua vez, comporta-se de maneira distinta em relação à disponibilidade de dados. Observa-se que o aumento na alocação de recursos não ocorre

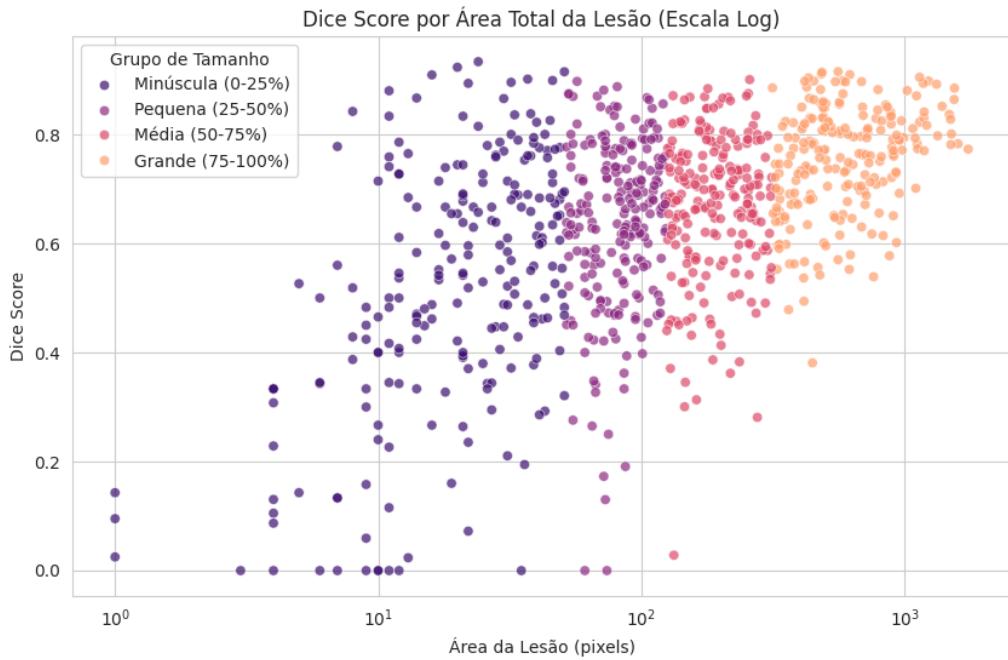


Figura 4.16: Gráfico do dice médio de cada item no conjunto de validação separado por quartis relacionados ao tamanho das lesões, gráfico avaliado no melhor modelo obtido

Modelo	Tempo de Treinamento(s)	Quantidade de VRAM alocada(GB)
Default	2342	28.567
+500	2358	29.293
+750	2374	29.567
+1000	2386	29.878

Tabela 4.3: Recursos computacionais necessários para execução de cada modelo

de forma proporcional à expansão da base de imagens, visto que o armazenamento dos dados brutos não representa o fator preponderante no uso de VRAM. O custo principal está associado à alocação de buffers para o cálculo de gradientes durante o processo de *backpropagation*. A Figura 4.17 evidencia que os picos de utilização de VRAM não são constantes, apresentando variabilidade conforme a etapa de treinamento. Ressalta-se que a incorporação de volumes maiores de imagens sintéticas foi inviabilizada pela exaustão da memória disponível para o cache interno da biblioteca *Pytorch*, resultando na interrupção do processo.

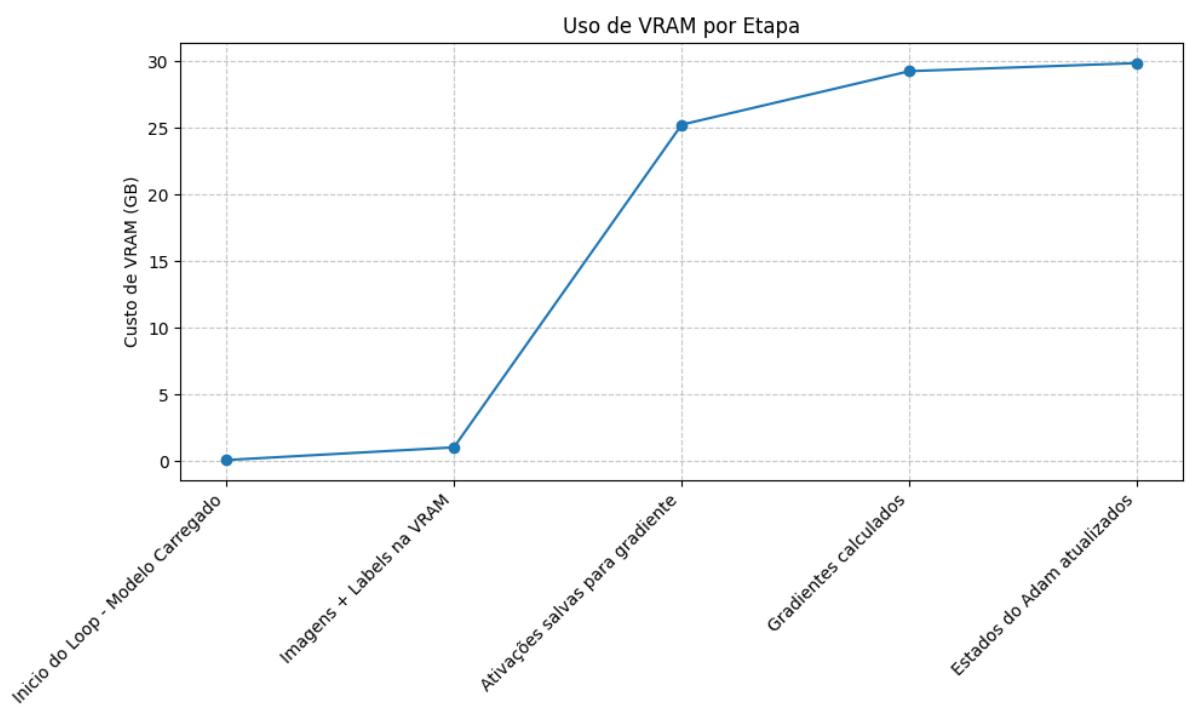


Figura 4.17: Gráfico do PICO de uso de VRAM durante cada etapa do treinamento, esse gráfico desconsidera o cache utilizado pelo *Pytorch* devido a alta volatilidade dessa informação durante diversos momentos no treinamento

## 5 Conclusão

O presente trabalho envolveu a implementação de um pipeline completo de data augmentation para segmentação de texturas relacionadas a identificação de lesões no contexto da Esclerose Múltipla. O primeiro passo envolveu a construção e utilização de um Variational Autoencoder para capturar a representação de imagens de lesões mapeadas por especialistas e converter essa representação em um espaço contínuo e probabilístico, sendo capaz de gerar novas máscaras de lesões próximas as reais. A partir disso foi construída uma Rede Generativa Adversarial (GAN) que recebe como entrada uma máscara e é capaz de a partir dessa entrada construir a textura cerebral correspondente. Os novos dados produzidos pela GAN foram então utilizados no treinamento de um modelo de segmentação de lesões de Esclerose Múltipla a fim de verificar a possibilidade de melhora na segmentação de um conjunto de validação padrão com a utilização de dados produzidos artificialmente.

Dessa forma, diferentemente de trabalhos que avaliam modelos generativos apenas sob o aspecto visual ou perceptual, este trabalho propõe um protocolo experimental sistemático para investigar o impacto da inclusão de imagens sintéticas, geradas de forma condicional, no desempenho de um modelo de segmentação de lesões de Esclerose Múltipla. A contribuição reside na análise controlada do efeito da quantidade de dados sintéticos, considerando tanto métricas de segmentação quanto o custo computacional associado.

Os diversos experimentos apresentados demonstraram a capacidade de generalização dos modelos propostos. A inclusão de dados artificiais no treinamento, embora aumentem o custo computacional do mesmo, aumentaram de maneira significativa a qualidade das segmentações automáticas. No entanto, considerando que o foco deste trabalho como sendo a análise do impacto de dados sintéticos em um cenário 2D, mais próximo de aplicações com restrição computacional ou disponibilidade de dados 3D limitada, a comparação com a literatura, que majoritariamente se concentra em segmentação volumétrica, se torna restrita.

## 5.1 Trabalhos futuros

Trabalhos futuros devem priorizar a mitigação das limitações inerentes à geração condicional, especificamente no que tange à baixa capacidade de generalização do modelo gerativo em cenários com lesões de pequenas dimensões — um dos maiores desafios na segmentação de Esclerose Múltipla. Uma estratégia promissora consiste na inclusão de imagens de indivíduos saudáveis (controles) ao conjunto de treinamento da GAN. Tal abordagem visa refinar a representação de tecidos normais, auxiliando o modelo a discriminar melhor casos com baixa carga lesional e reduzindo a ocorrência de artefatos nas imagens.

Outra possibilidade envolve a utilização de modelos Generativos baseados em difusão, como modelos de *Flow Matching*, que são amplamente suportados pela literatura moderna como mais simples de treinar em relação a Redes Generativas Adversariais, embora dependam, em muitos casos, de uma quantidade mais ampla de dados.

Adicionalmente, deve-se investigar o potencial das Redes Generativas na mitigação de outros desafios inerentes à Ressonância Magnética, notadamente a baixa generalização frente à variabilidade inter-scanner. Uma abordagem promissora envolve o treinamento de modelos capazes de realizar a adaptação de domínio, sintetizando imagens que mimetizem as características de aquisição de equipamentos distintos da base original.

Além disso, também deve-se buscar um pipeline que consiga se aproveitar da informação volumétrica presente em grande parte das bases de dados. Permitindo assim, a comparação direta com literatura moderna na área de Segmentação de Esclerose Múltipla.

## Bibliografia

- AGGARWAL, A.; MITTAL, M.; BATTINENI, G. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, v. 1, n. 1, p. 100004, abr. 2021. ISSN 26670968. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2667096820300045>>.
- ALOM, M. Z. et al. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1802.06955>>.
- BARILE, B. et al. Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. *Computer Methods and Programs in Biomedicine*, v. 206, p. 106113, jul. 2021. ISSN 01692607. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169260721001887>>.
- BROWNLEE, W. J. et al. Diagnosis of multiple sclerosis: progress and challenges. *The Lancet*, Elsevier, v. 389, n. 10076, p. 1336–1346, Apr 2017. ISSN 0140-6736. Disponível em: <[https://doi.org/10.1016/S0140-6736\(16\)30959-X](https://doi.org/10.1016/S0140-6736(16)30959-X)>.
- CETIN, O. et al. Enhancing precision in multiple sclerosis lesion segmentation: A U-net based machine learning approach with data augmentation. *NeuroImage: Reports*, v. 5, n. 1, p. 100235, mar. 2025. ISSN 26669560. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2666956025000030>>.
- COHEN, J. A. et al. Oral Fingolimod or Intramuscular Interferon for Relapsing Multiple Sclerosis. *New England Journal of Medicine*, v. 362, n. 5, p. 402–415, fev. 2010. ISSN 0028-4793, 1533-4406. Disponível em: <<http://www.nejm.org/doi/abs/10.1056/NEJMoa0907839>>.
- COMMOWICK, O. et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. *Scientific Reports*, v. 8, n. 1, p. 13650, set. 2018. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-018-31911-7>>.
- DELVES, P. J.; ROITT, I. M. The Immune System. *New England Journal of Medicine*, v. 343, n. 1, p. 37–49, jul. 2000. ISSN 0028-4793, 1533-4406. Disponível em: <<http://www.nejm.org/doi/10.1056/NEJM200007063430107>>.
- DICE, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, v. 26, n. 3, p. 297–302, jul. 1945. ISSN 0012-9658, 1939-9170. Disponível em: <<https://esajournals.onlinelibrary.wiley.com/doi/10.2307/1932409>>.
- DOBSON, R.; GIOVANNONI, G. Multiple sclerosis – a review. *European Journal of Neurology*, v. 26, n. 1, p. 27–40, 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/ene.13819>>.
- GABR, R. E. et al. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study. *Mult. Scler.*, SAGE Publications, v. 26, n. 10, p. 1217–1226, set. 2020.

- GUARNERA, F. et al. MSLesSeg: baseline and benchmarking of a new Multiple Sclerosis Lesion Segmentation dataset. *Scientific Data*, v. 12, n. 1, p. 920, maio 2025. ISSN 2052-4463. Disponível em: <<https://www.nature.com/articles/s41597-025-05250-y>>.
- IMRAN, A.-A.-Z.; TERZOPOULOS, D. Multi-adversarial variational autoencoder nets for simultaneous image generation and classification. In: \_\_\_\_\_. *Deep Learning Applications, Volume 2*. Singapore: Springer Singapore, 2021. p. 249–271. ISBN 978-981-15-6759-9. Disponível em: <[https://doi.org/10.1007/978-981-15-6759-9\\_11](https://doi.org/10.1007/978-981-15-6759-9_11)>.
- LECUN, Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, dez. 1989. ISSN 0899-7667, 1530-888X. Disponível em: <<https://direct.mit.edu/neco/article/1/4/541-551/5515>>.
- LLADÓ, X. et al. Segmentation of multiple sclerosis lesions in brain mri: A review of automated approaches. *Information Sciences*, v. 186, n. 1, p. 164–185, 2012. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025511005548>>.
- MYRONENKO, A. *3D MRI brain tumor segmentation using autoencoder regularization*. arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1810.11654>>.
- ÖMERHOCA, S.; AKKAŞ, S. Y.; İÇEN, N. K. Multiple sclerosis: Diagnosis and differential diagnosis. *Noro Psikiyatration Ars.*, Turk Noropsikiyatri Dernegi, v. 55, n. Suppl 1, p. S1–S9, 2018.
- PINTO, D. M. et al. Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em Swin-Unet. In: *Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2024)*. Brasil: Sociedade Brasileira de Computação - SBC, 2024. p. 471–482. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/28841>>.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: NAVAB, N. et al. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. v. 9351, p. 234–241. ISBN 9783319245737 9783319245744. Disponível em: <[http://link.springer.com/10.1007/978-3-319-24574-4\\_28](http://link.springer.com/10.1007/978-3-319-24574-4_28)>.
- ROSA, A. P. D. et al. Consensus of algorithms for lesion segmentation in brain MRI studies of multiple sclerosis. *Scientific Reports*, v. 14, n. 1, p. 21348, set. 2024. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-024-72649-9>>.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, n. 6088, p. 533–536, out. 1986. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/323533a0>>.
- TURVEY, S. E.; BROIDE, D. H. Innate immunity. *Journal of Allergy and Clinical Immunology*, v. 125, n. 2, Supplement 2, p. S24–S32, 2010. ISSN 0091-6749. 2010 Primer on Allergic and Immunologic Diseases. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0091674909010835>>.
- WHO. *Multiple sclerosis*. 2025. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/multiple-sclerosis>>.

ZIJDENBOS, A. et al. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Transactions on Medical Imaging*, v. 13, n. 4, p. 716–724, dez. 1994. ISSN 02780062. Disponível em: <<http://ieeexplore.ieee.org/document/363096/>>.