

Universidade do Minho
Escola de Engenharia
Mestrado em Engenharia Informática

SISTEMAS INTELIGENTES

Aprendizagem Profunda

Ano Letivo de 2023/2024

Classificação de dados sintéticos em imagem médica volumétrica (CT e MRI)

Grupo 3

PG53895 Joana Isabel Freitas Pereira
PG53645 Ana Rita Santos Poças
PG52686 João Delfim Da Cruz Pereira

30 de maio de 2024

Índice

1	Introdução	3
1.1	Contextualização	3
2	Desenvolvimento	4
2.1	Dataset	4
2.1.1	Formato das Imagens	4
2.1.2	Organização dos Dados	5
2.1.3	Divisão dos Dados	5
2.1.4	Balanceamento dos Dados	5
2.2	Treino e Avaliação do Modelo	6
2.2.1	Utilização do CUDA	6
2.2.2	Batch Size e Shuffle	6
2.2.3	Modelos Utilizados	7
2.2.4	Função de Loss e Otimização	7
2.2.5	Epochs e desempenho	7
2.3	Função de Avaliação	9
2.4	Visualização no ITK-SNAP	10
3	Desafios e Melhorias	12
3.1	Desafios	12
3.2	Melhorias	12
4	Conclusão	14

1 Introdução

Neste relatório, será abordado o projeto prático desenvolvido no âmbito da disciplina de Aprendizagem Profunda, pertencente ao perfil de Sistemas Inteligentes do Mestrado Integrado em Engenharia Informática da Universidade do Minho. O projeto, realizado no primeiro ano do segundo semestre do ano letivo 2023/2024, teve como objetivo o desenvolvimento de modelos de classificação que consigam determinar se um determinado exame médico é original ou sintético e baseado nesses modelos construir uma função de avaliação da qualidade de um exame sintético.

1.1 Contextualização

O avanço tecnológico e a crescente disponibilidade de dados têm permitido o desenvolvimento de sistemas de inteligência artificial cada vez mais sofisticados. No campo da saúde, a geração de exames médicos sintéticos surge como uma solução inovadora para a criação de dados que podem ser utilizados em pesquisas, treino de modelos de aprendizagem profunda e simulações, sem comprometer a privacidade dos pacientes.

Um exame sintético é um exame gerado artificialmente através de algoritmos de inteligência artificial e técnicas de geração de dados, como redes adversariais generativas (GANs). Estes exames imitam as características dos exames médicos reais, permitindo a criação de grandes volumes de dados de forma rápida e eficiente. No entanto, a qualidade desses exames sintéticos é crucial para garantir a sua eficácia e equivalência aos exames originais.

A importância deste projeto reside na necessidade de assegurar que os exames gerados automaticamente mantêm um padrão de qualidade comparável aos exames tradicionais, contribuindo para a fiabilidade e validade dos dados utilizados em contextos clínicos e de investigação. Além disso, o desenvolvimento de modelos robustos de classificação e avaliação de qualidade pode ter implicações significativas não só no âmbito da saúde, mas também noutras áreas onde a geração de dados sintéticos é relevante, como na produção de imagens médicas, simulações de tratamentos e na melhoria dos algoritmos de diagnóstico assistido por computador.

Assim, este relatório detalha todo o processo de desenvolvimento e implementação dos modelos de classificação, bem como a construção da função de avaliação, destacando os desafios encontrados e as soluções adotadas para alcançar os objetivos propostos.

2 Desenvolvimento

2.1 Dataset

O conjunto de dados utilizado neste projeto consiste em um dataset com imagens volumétricas de ressonância magnética do cérebro, fornecidas pela equipa docente. O dataset contém um total de 100 imagens, das quais 50 são sintéticas e 50 são reais. As imagens estão no formato nii.gz e seguem uma convenção de nomenclatura específica:

- As imagens sintéticas têm o prefixo "Synt" seguido do nome do arquivo, por exemplo: "synt1_BraTS-GLI-00048-001_0000.nii.gz".
- Por outro lado, as imagens reais não têm nenhum prefixo, sendo identificadas apenas pelo nome do arquivo, como por exemplo: "BraTS-GLI-00000-000_0000.nii.gz".

Essa distinção na nomenclatura é fundamental para diferenciar entre as imagens sintéticas e reais durante o processamento e análise dos dados.

2.1.1 Formato das Imagens

As imagens estão no formato nii.gz, que é uma extensão de arquivo comum para dados de imagem médica volumétrica. O ".nii" indica que é um arquivo de imagem em formato NIfTI (Neuroimaging Informatics Technology Initiative), enquanto o ".gz" indica que o arquivo foi comprimido usando o algoritmo gzip.

O formato NIfTI é amplamente utilizado em neuroimagem e outras áreas da medicina para armazenar dados de imagem volumétrica. Ele suporta imagens 3D e 4D (com o tempo como uma dimensão adicional), bem como metadados adicionais, como informações de orientação e calibração.

Este formato é particularmente adequado para imagens médicas volumétricas, pois permite armazenar informações tridimensionais detalhadas, como é o caso das ressonâncias magnéticas (RM).

2.1.2 Organização dos Dados

Para organizar os dados de forma adequada, foi criado um array chamado *imagens*, que contém os caminhos para todas as imagens do dataset. Este array foi construído de modo que as primeiras 50 entradas correspondem às imagens sintéticas, seguidas pelas 50 imagens reais.

Além disso, foi criado um array de labels chamado *correct_labels*, que atribui a cada imagem uma label indicando se ela é sintética (0) ou real (1). Assim, os primeiros 50 elementos deste array são 0, representando as imagens sintéticas, e os 50 elementos seguintes são 1, correspondendo às imagens reais.

2.1.3 Divisão dos Dados

Posteriormente, procedeu-se à divisão dos dados em conjunto de treino e conjunto de validação. Optou-se por não criar um conjunto de teste devido ao número limitado de imagens no dataset. A divisão foi realizada manualmente, selecionando-se 35 imagens sintéticas e 35 imagens reais para o conjunto de treino, e as 15 restantes de cada tipo foram alocadas para o conjunto de validação. Essa divisão foi cuidadosamente planejada para garantir uma distribuição equilibrada entre os tipos de imagens em ambos os conjuntos.

2.1.4 Balanceamento dos Dados

O balanceamento dos dados foi verificado para assegurar que ambos os conjuntos de treino e validação contêm uma quantidade equilibrada de imagens sintéticas e reais. A Figura 2.1 ilustra essa distribuição.

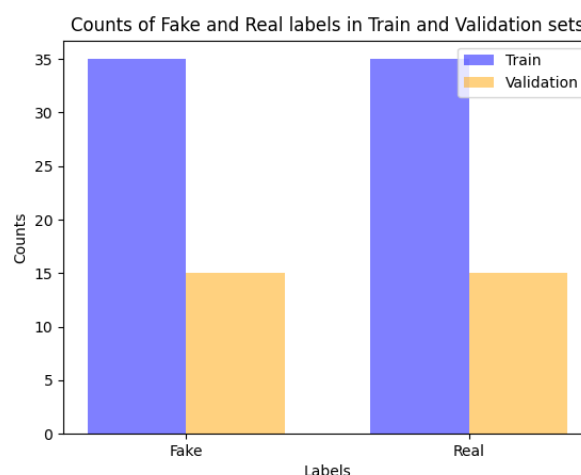


Figura 2.1: Distribuição das imagens sintéticas e reais nos conjuntos de treino e validação.

2.2 Treino e Avaliação do Modelo

Para o treino e avaliação do modelo, foram utilizadas técnicas específicas para garantir a eficiência e a precisão do modelo.

2.2.1 Utilização do CUDA

O treino foi realizado utilizando a GPU (CUDA) para acelerar o processo. Embora o conjunto de dados consistisse apenas em 100 imagens, é importante salientar que estas eram imagens volumétricas, o que requer um processamento intensivo. A GPU mostrou-se fundamental para lidar com a complexidade dessas imagens, permitindo um treino mais eficiente do modelo.

Durante o processo, enfrentámos algumas complicações com o uso da GPU, o que nos levou a desenvolver o trabalho num ficheiro Python em vez de notebooks. Apesar disso, o treino foi bem-sucedido e organizado, demonstrando a eficácia da utilização da GPU para este fim.

Para proporcionar uma visão mais completa do ambiente de treino, foi considerado incluir uma captura de ecrã com as informações relevantes da GPU utilizada, destacando os recursos disponíveis e as configurações utilizadas durante o treino. Isso contribuiria para uma compreensão mais detalhada do contexto em que o modelo foi desenvolvido.

2.2.2 Batch Size e Shuffle

Foi utilizado um batch size de 2 durante o treinamento. Um batch size pequeno permite uma atualização mais frequente dos pesos do modelo, o que pode levar a uma convergência mais rápida. No entanto, isso também pode introduzir mais ruído na estimativa do gradiente, o que pode afetar a estabilidade do treinamento.

Além disso, o parâmetro `shuffle=True` foi usado para embaralhar os dados de treino a cada época. Isso ajuda a garantir que o modelo não aprenda a memorizar a ordem dos dados, mas sim a generalizar melhor a partir dos padrões presentes nos dados.

Para verificar se este shuffle estava a ser eficaz, foi gerado o seguinte gráfico:

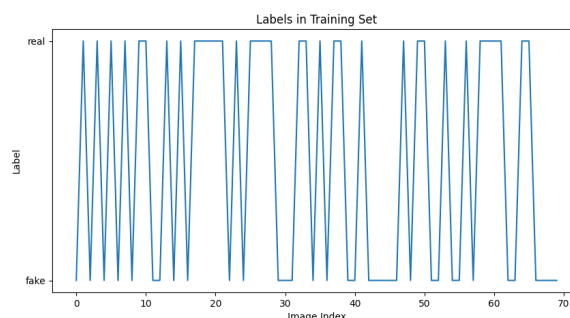


Figura 2.2: Distribuição dos rótulos no conjunto de treino após cada época de treino.

O gráfico mostra a distribuição aleatória dos rótulos (sintéticos e reais) no conjunto de treino. A variabilidade na distribuição ao longo das épocas sugere que o shuffle foi eficaz.

A mistura dos dados de treino é crucial para:

- **Evitar Overfitting:** Previne que o modelo memorize a ordem dos dados.
- **Aumentar a Robustez:** Expõe o modelo a diferentes sequências de dados, melhorando a sua capacidade de lidar com variações e ruídos nos dados.
- **Garantir a Imparcialidade:** Assegura que o modelo tenha uma visão equilibrada dos dados, não favorecendo certos padrões ou classes.
- **Acelerar a Convergência:** Facilita o processo de aprendizagem, ajudando o modelo a encontrar melhores soluções de forma mais rápida.

2.2.3 Modelos Utilizados

Foi utilizado o modelo DenseNet121 devido à sua capacidade de lidar eficientemente com dados volumétricos complexos, como as imagens 3D utilizadas neste projeto.

A configuração específica para a DenseNet121 foi:

- **spatial_dims=3:** Indica que estamos a lidar com dados volumétricos 3D.
- **in_channels=1:** Indica que as imagens têm um único canal de entrada (escala de cinza).
- **out_channels=1:** Indica que o modelo tem uma única saída, que representa a probabilidade de a imagem ser real.

2.2.4 Função de Loss e Otimização

A função de perda utilizada foi a *BCEWithLogitsLoss()*, apropriada para problemas de classificação binária como este caso. O otimizador escolhido foi o Adam, devido à sua eficácia e capacidade de lidar com problemas de gradientes dispersos.

2.2.5 Epochs e desempenho

O treino foi conduzido ao longo de várias épocas, começando com 10 épocas, depois 30 e, finalmente, 100 épocas. Constatou-se que o desempenho do modelo estabilizou apenas após cerca de 40 épocas, atingindo uma precisão de treino de 1. A escolha do modelo foi baseada na análise das métricas de desempenho, nomeadamente os gráficos de precisão e perda para ambos os conjuntos de treino e validação. A precisão foi calculada como a proporção de previsões corretas (após aplicar a função sigmoide e arredondar para 0 ou 1) sobre o total

de previsões. A perda foi monitorada para garantir que o modelo não estava sofrendo de sobreajuste. Através da análise das imagens referentes às 100 epochs é perceptível que as potenciais zonas de paragem seriam às 40 epochs, 50 epochs ou 65 epochs. Foram realizados modelos para estes diferentes parâmetros mas foi o de 50 epochs que foi escolhido como sendo o melhor modelo, uma vez que o treino chegou a 1 de *accuracy* e estabilizou nesse valor. Assim, foi decidido que se iria usar o modelo gerado com as 50 epochs tanto na análise dos resultados como na função de avaliação, que será abordada mais à frente.

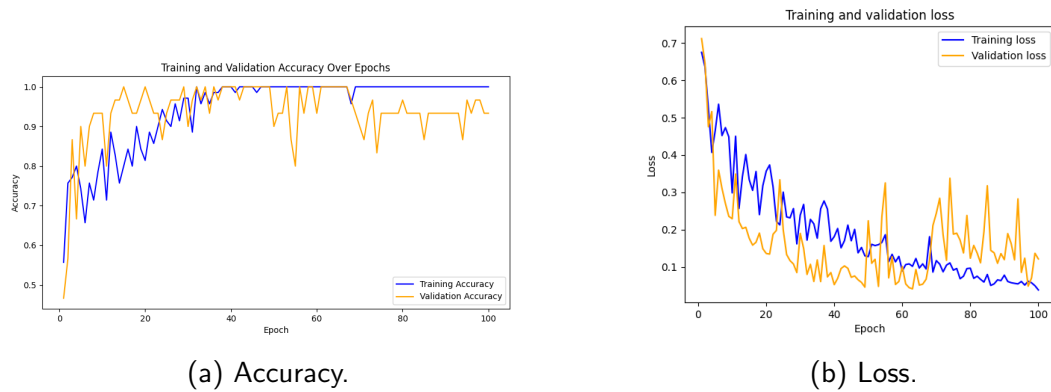


Figura 2.3: Epoch 100.

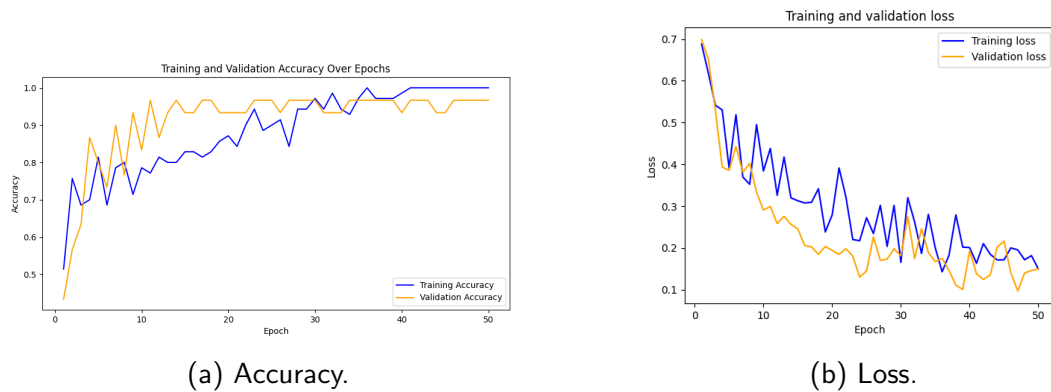


Figura 2.4: Epoch 50.

2.3 Função de Avaliação

Por fim, foi implementada uma função de avaliação para classificar imagens sintéticas volumétricas utilizando o modelo treinado. É de ressaltar que, como a divisão das imagens foi feita apenas em dois grupos (treino e validação), para este teste foram usadas apenas as imagens de validação.

Esta função carrega o modelo previamente treinado a partir do ficheiro *classification_model_40.pth* e aplica-o às imagens. O modelo é transferido para a GPU, se disponível, para acelerar o processamento.

A função de ativação sigmoid é aplicada ao output do modelo para converter os resultados em probabilidades, que variam de 0 a 1. Com base nessas probabilidades, as imagens são classificadas em três categorias:

- **Good Images:** Imagens com probabilidade maior ou igual a 0.5.
- **Bad Images:** Imagens com probabilidade entre 0.1 e 0.5.
- **Very Bad Images:** Imagens com probabilidade inferior a 0.1.

Os resultados são gravados no ficheiro *results_{número_epochs}.txt*, que contém as categorias e, abaixo de cada uma, os nomes das imagens pertencentes a essa categoria. A estrutura do ficheiro de resultados é organizada de forma clara, como ilustrado na imagem seguinte.

```
1 Good Images:
2
3 Bad Images:
4 synt_BraTS-GLI-00048-001_0000.nii.gz
5 synt_BraTS-GLI-00051-000_0000.nii.gz
6 synt_BraTS-GLI-00052-000_0000.nii.gz
7 synt_BraTS-GLI-00054-000_0000.nii.gz
8 synt_BraTS-GLI-00058-000_0000.nii.gz
9 synt_BraTS-GLI-00058-001_0000.nii.gz
10 synt_BraTS-GLI-00059-000_0000.nii.gz
11 synt_BraTS-GLI-00059-001_0000.nii.gz
12 synt_BraTS-GLI-00061-000_0000.nii.gz
13 synt_BraTS-GLI-00061-001_0000.nii.gz
14
15 Very Bad Images:
16 synt_BraTS-GLI-00048-001_0000.nii.gz
17 synt_BraTS-GLI-00053-000_0000.nii.gz
18 synt_BraTS-GLI-00053-001_0000.nii.gz
19 synt_BraTS-GLI-00056-000_0000.nii.gz
20 synt_BraTS-GLI-00056-001_0000.nii.gz
21 synt_BraTS-GLI-00058-000_0000.nii.gz
22 synt_BraTS-GLI-00059-001_0000.nii.gz
23 synt_BraTS-GLI-00062-000_0000.nii.gz
```

Figura 2.5: Distribuição dos rótulos no conjunto de treino após cada época de treino.

2.4 Visualização no ITK-SNAP

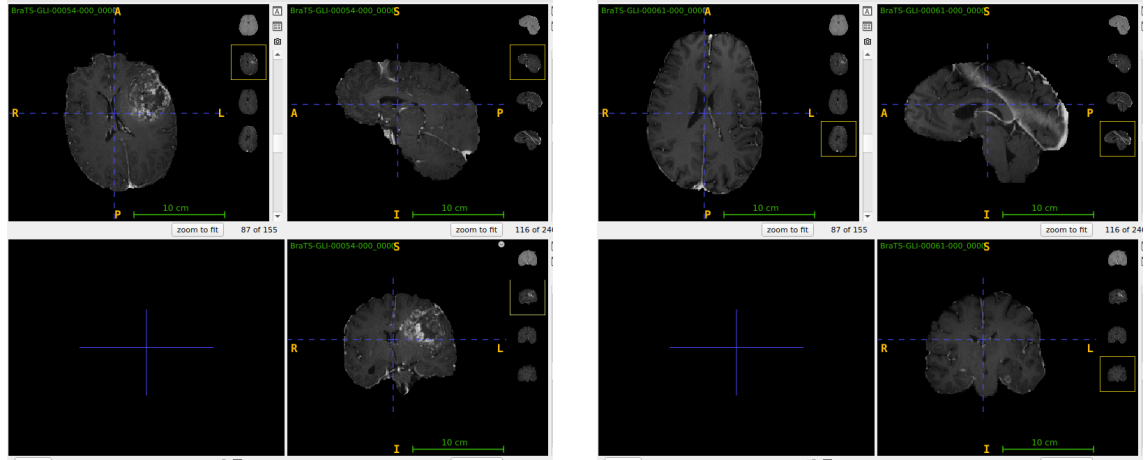
Após o treino do modelo e a avaliação do seu desempenho, decidimos realizar uma análise mais detalhada das previsões do modelo utilizando a aplicação ITK-SNAP. Esta ferramenta é amplamente utilizada para visualização, segmentação e registo de imagens médicas em formato volumétrico.

Nesta secção, será apresentada uma seleção de imagens para visualização no ITK-SNAP. Optou-se por examinar duas imagens reais com probabilidades altas, bem como duas imagens sintéticas: uma com probabilidade muito baixa e outra com probabilidade alta. O objetivo é validar visualmente as previsões do modelo e verificar se as probabilidades atribuídas correspondem às características visuais das imagens.

A seguir, serão fornecidas informações sobre cada imagem selecionada, seguidas das próprias imagens visualizadas no ITK-SNAP.

Filename: BraTS-GLI-00054-000_0000.nii.gz, Label: 1, Probability: 0.9735157489776611

Filename: BraTS-GLI-00061-000_0000.nii.gz, Label: 1, Probability: 0.968421995639801



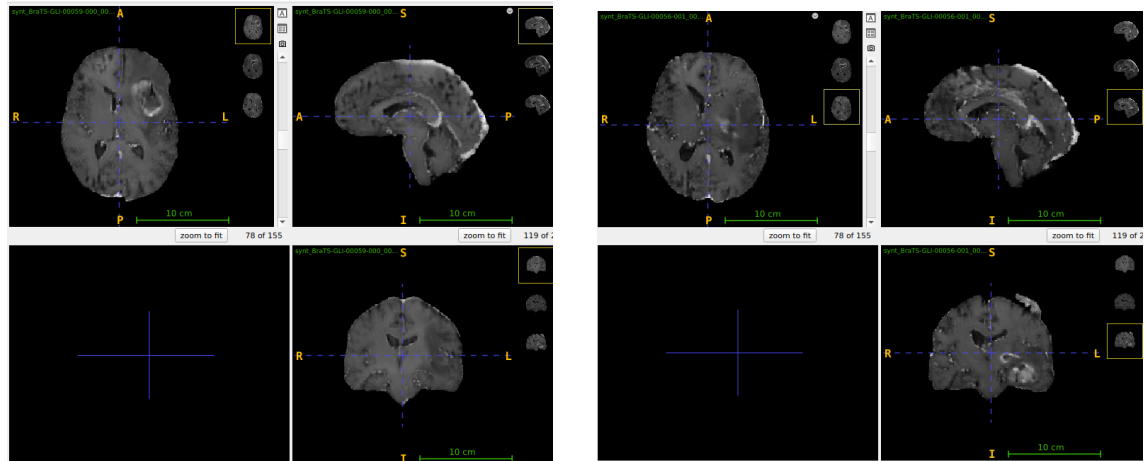
(a) BraTS-GLI-00054-000_0000

(b) BraTS-GLI-00061-000_0000

Figura 2.6: Imagens Volumétricas Reais

Filename: synt_BraTS-GLI-00059-000_0000.nii.gz, Label: 0, Probability: 0.09096892923116684

Filename: synt_BraTS-GLI-00056-001_0000.nii.gz, Label: 0, Probability: 0.004515563603490591



(a) BraTS-GLI-00059-000_0000

(b) BraTS-GLI-00056-001_0000

Figura 2.7: Imagens Volumétricas Sintéticas

Através das imagens, é perceptível que, apesar de todas as probabilidades serem baixas a diferença é significativa entre as imagens com probabilidades muito baixas e aquelas com probabilidades mais próximas de 0.5. Estas diferenças confirmam a eficácia do modelo em distinguir corretamente entre imagens reais e sintéticas, mesmo quando as probabilidades atribuídas são baixas.

3 Desafios e Melhorias

Durante o desenvolvimento do projeto, foram enfrentados diversos desafios que impactaram o desempenho e a eficácia do modelo. Além disso, foram identificadas algumas áreas em que o projeto poderia ser melhorado para obter resultados ainda mais robustos e precisos.

3.1 Desafios

De seguida estão listados os principais desafios encontrados durante o processo de desenvolvimento:

- **Problemas de Dependência (PyTorch e MONAI):** Um dos principais desafios encontrados foi a gestão de dependências, principalmente com as bibliotecas PyTorch e MONAI. A compatibilidade entre diferentes versões dessas bibliotecas pode ser um problema, especialmente ao tentar reproduzir resultados em diferentes ambientes. Encontrar e resolver conflitos de dependências exigiu tempo e esforço adicionais, afetando o ritmo do desenvolvimento do projeto.
- **Configuração CUDA:** Outro desafio significativo foi a configuração e utilização do CUDA para treinar o modelo na GPU. Embora o CUDA ofereça vantagens significativas em termos de desempenho, a configuração inicial foi bastante complexa.
- **Quantidade dos Dados:** A quantidade limitada de dados disponíveis para treino e validação do modelo também representou um desafio. Com apenas 100 imagens no conjunto de dados, o modelo pode não ter sido capaz de capturar toda a complexidade e variabilidade dos exames médicos sintéticos e reais.

3.2 Melhorias

A seguir, são apresentadas as principais melhorias propostas:

- **Testar Outros Modelos (ResNet):** Uma das melhorias propostas é explorar outros modelos de redes neurais convolucionais, como a ResNet. A ResNet é conhecida por sua capacidade de lidar eficientemente com problemas de vanishing gradient, o que poderia resultar em melhores resultados de classificação, especialmente com conjuntos de dados pequenos.

- **Cross Validation:** A utilização de validação cruzada pode ajudar a avaliar o desempenho do modelo de forma mais robusta, especialmente com um conjunto de dados limitado. A validação cruzada divide o conjunto de dados em várias partes e realiza múltiplos treinos e testes, permitindo uma avaliação mais abrangente do desempenho do modelo.
- **Data Augmentation:** A aplicação de técnicas de aumento de dados pode ajudar a aumentar a variabilidade dos dados de treino e melhorar a capacidade do modelo de generalizar para novos dados. Técnicas de zoom podem ser aplicadas para gerar variações adicionais das imagens existentes, enriquecendo assim o conjunto de dados de treino.

4 Conclusão

Este projeto abordou o desenvolvimento e avaliação de modelos de classificação para distinguir entre exames médicos reais e sintéticos, com o objetivo de garantir a qualidade dos exames gerados por algoritmos de inteligência artificial. Utilizando técnicas avançadas de aprendizagem profunda e processamento de imagens volumétricas, foi possível criar um modelo robusto e uma função de avaliação eficaz.

Os resultados obtidos demonstram a viabilidade e importância de utilizar dados sintéticos em pesquisas e treinos, desde que sejam mantidos padrões de qualidade elevados. A implementação de um modelo baseado na arquitetura DenseNet121 provou ser eficiente, e a utilização de recursos como a GPU (CUDA) facilitou o processamento intensivo necessário para lidar com as imagens 3D.

O projeto também destacou a importância de enfrentar desafios técnicos, como problemas de dependência de bibliotecas e a configuração do ambiente de treino, bem como a necessidade de explorar métodos adicionais para melhorar a robustez e precisão do modelo.

Em suma, este trabalho contribuiu para o avanço na geração e avaliação de exames médicos sintéticos, oferecendo uma base sólida para futuras pesquisas e aplicações práticas na área da saúde. A capacidade de gerar dados de alta qualidade tem o potencial de revolucionar a pesquisa médica e melhorar os algoritmos de diagnóstico assistido por computador, beneficiando pacientes e profissionais da saúde.