# Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada

*Lucas Almeida da Silva*

# Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada

*Lucas Almeida da Silva*

**Orientador:** *Prof. Dr. Rafael Giusti*
**Coorientadora:** *Profa. Dra. Eulanda Miranda dos Santos*

Dissertação apresentada ao Instituto de Computação da Universidade Federal do Amazonas - ICOMP-UFAM, como parte dos requisitos para obtenção do título de Mestre em Informática.

**UFAM – Manaus – Junho/2023**

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S586a

Silva, Lucas Almeida da
Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada / Lucas Almeida da Silva . 2023
66 f.: il. color; 31 cm.

Orientador: Rafael Giusti
Coorientadora: Eulanda Miranda dos Santos
Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas.

1. Extração de quadros em tomografia. 2. Grad-CAM. 3. Aprendizado profundo. 4. Tomografia computadorizada. 5. Redes neurais convolucionais. I. Giusti, Rafael. II. Universidade Federal do Amazonas III. Título

# FOLHA DE APROVAÇÃO

## "Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada"

## LUCAS ALMEIDA DA SILVA

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Dr. Rafael Giusti - PRESIDENTE

Prof. Dr. Eduardo James Pereira Souto - MEMBRO INTERNO

Prof. Dr. Luis Cuevas Rodríguez - MEMBRO EXTERNO

Dr. Bernardo Bentes Gatto - MEMBRO EXTERNO

Manaus, 30 de junho de 2023

Trabalho desenvolvido com financiamento CAPES, processo Nº 88887.499518/2020-00.

# Resumo

A análise de imagens médicas em dados volumétricos normalmente é feita com a utilização de redes neurais convolucionais profundas 2D (CNN 2D), o que implica na análise independente e quadros individuais. Em grande parte, isso é devido aos desafios impostos pela natureza de dados tridimensionais, tais como: tamanho de volume variável, altos requisitos de memória (GPU e RAM), otimização de parâmetros, dentre outros. No entanto, lidar com os quadros individuais de forma independente em CNNs 2D descarta, deliberadamente, as informações temporais que constituem a profundidade dos volumes, o que pode resultar em baixo desempenho para a tarefa pretendida. Portanto, é importante desenvolver métodos que superem os requisitos computacionais impostos para que se aproveite as informações 3D.

Para isso, neste trabalho é proposto um método não supervisionado baseado em *Grad-Cam* para seleção dos segmentos mais relevantes em volumes de tomografia computadorizada, por meio da avaliação do mapa de ativação na última camada convolucional de uma CNN3D projetada para esse fim. Mesmo que o diagnóstico por métodos de Aprendizado de Máquina já mostre resultados promissores por meio do uso de redes neurais para avaliação de imagens radiológicas dos pulmões, a grande maioria dos métodos utiliza imagens pré-selecionadas por profissionais humanos para compor uma base de dados adequada, e isso se agrava quando são utilizados volumes de tomografia computadorizada, onde se faz necessária a separação de quadros mais significativos para avaliação clínica, uma vez que a análise de volume completo é computacionalmente cara e demorada.

Experimentos extensivos com volumes de tomografia computadorizada demonstraram o êxito da metodologia proposta. A eficácia do método Grad-Cam Slice Selection (GSS) se evidenciou ao superar técnicas atuais do estado da arte, tanto em termos de área sob a curva ROC (AUC) quanto de F1 Score, em todas as configurações testadas.

# Abstract

A common approach for analyzing medical images on volumetric data employs deep 2D convolutional neural networks (2D CNN), which imply the use of individual frames. This is largely attributed to the challenges posed by the nature of three-dimensional data: variable volume size, sufficient GPU and RAM allocation, parameter optimization, and more. However, handling the individual frames independently in 2D CNNs deliberately discards the temporal information that constitutes the depth of the volumes, which results in poor performance for the intended task. Therefore, it is important to develop methods that go beyond the computational requirements imposed in order to take advantage of 3D information.

For this, we propose an unsupervised method based on Grad-Cam to select key-frames in computed tomography volumes by evaluating the activation map in the last convolutional layer of a CNN3D designed for this purpose. The diagnosis of coronavirus disease was used as a case study for this first stage of the project. Even though the diagnosis by Machine Learning methods already shows promising results through the use of neural networks for the evaluation of radiological images of the lungs, the vast majority of methods use images pre-selected by human professionals to compose an adequate database and this is aggravated when computed tomography volumes are used, where it is necessary to separate more significant frames for clinical evaluation because the full volume analysis is computationally expensive and time-consuming.

Extensive experiments with computed tomography volumes demonstrated the success of the proposed methodology. The effectiveness of the Grad-Cam Slice Selection (GSS) method was shown to outperform current state-of-the-art techniques, both in terms of area under the ROC curve (AUC) and F1 Score, in all configurations tested.

# Agradecimentos

Agradeço a Deus pela vida que Ele me concedeu. Só cheguei aqui por conta Dele e é por Ele todas estas coisas.

Agradeço aos meus pais por todo o esforço investido na minha educação. Muito obrigado por terem me ensinado desde criança o caminho que deveria seguir.

Agradeço à minha esposa que sempre esteve ao meu lado, me incentiva e ajuda em tudo que preciso.

Agradeço aos meus irmãos pelo grande apoio nas batalhas do dia a dia.

Sou grato pelas várias oportunidades dadas pelos professores orientadores neste trabalho.

Professor Rafael, meu orientador, quem sempre esteve disponível para me ajudar em todos os passos, sendo um parceiro e amigo que quero manter mesmo com a conclusão deste trabalho.

Professora Eulanda, minha coorientadora, quem sempre me ajudou a quebrar qualquer barreira encontrada no projeto.

Quero agradecer à Universidade Federal do Amazonas e todo o seu corpo docente pela oportunidade a mim concedida.

Por fim um imenso obrigado à CAPES pelo financiamento deste projeto.

# Sumário

**Resumo** **5**

**Abstract** **7**

**Agradecimentos** **8**

**Lista de Tabelas** **11**

**Lista de Figuras** **13**

**1 Introdução** **1**

1.1 Contextualização e Motivação . . . . . 1

1.2 Estudo de Caso . . . . . 2

1.3 Objetivos e Hipótese . . . . . 3

1.4 Organização do texto . . . . . 4

**2 Fundamentação Teórica** **5**

2.1 Considerações Iniciais . . . . . 5

2.2 Aprendizado de Máquina . . . . . 5

2.2.1 Aprendizado de Máquina aplicado ao reconhecimento de Imagens . . . 6

2.3 *Aprendizado Profundo* . . . . . 8

2.3.1 Obtenção de conhecimento . . . . . 8

2.3.2 Convolução em aprendizado profundo . . . . . 12

2.3.3 Custo Computacional . . . 13
2.3.4 Função de ativação . . . 15
2.3.5 Interpretações das ativações com Grad-Cam . . . 16
2.4 Tomografia Computadorizada . . . 19
2.4.1 Aplicação no Aprendizado Profundo . . . 20
2.5 Considerações Finais . . . 21

**3 Trabalhos Relacionados** **23**
3.1 Considerações Iniciais . . . 23
3.2 Técnicas para Amostragem em TC . . . 24
3.2.1 Subset Slice Selection (SSS) . . . 24
3.2.2 Even Slice Selection (ESS) . . . 26
3.2.3 Spline Interpolated Zoom (SIZ) . . . 27
3.3 Técnicas para redução de dimensionalidade . . . 28
3.3.1 Análise de Componentes Principais . . . 29
3.3.2 Análise de Componentes Independentes . . . 30
3.3.3 t-Distributed Stochastic Neighbor Embedding . . . 32
3.4 Técnicas de extração de características significativas com Aprendizado profundo 34
3.4.1 Autocodificadores . . . 34
3.4.2 Seleção de quadros por modelos de detecção e segmentação . . . 35
3.4.3 Uso de Convoluções em profundidade para aprender características espaço-temporais . . . 36
3.5 Classificação com volumes de tomografia . . . 37
3.6 Considerações Finais . . . 37

**4 Abordagem Proposta** **39**
4.1 Considerações Iniciais . . . 39
4.2 Grad-Cam Slice Selection (GSS) . . . 40
4.3 Aplicação do Método a modelos de classificação . . . 46
4.3.1 Ajuste do tamanho da entrada . . . 46

4.3.2 Treino de modelos de aprendizado profundo . . . . . 48
4.4 Considerações Finais . . . . . 50

**5 Avaliação Experimental** **51**
5.1 Considerações Iniciais . . . . . 51
5.1.1 Dados . . . . . 51
5.1.2 Configuração dos Experimentos . . . . . 53
5.1.3 Modelos . . . . . 54
5.1.4 Métricas . . . . . 55
5.1.5 Resultados . . . . . 55

**6 Conclusão** **58**
6.1 Considerações Finais . . . . . 58
6.2 Principais Contribuições e Limitações . . . . . 59

# Lista de Tabelas

4.1 Template de arquitetura 3DCNN-C . . . . . 43

5.1 Propriedades da base *MosMed* (Morozov et al., 2020). . . . . 52

5.2 Resultados . . . . . 56

# Lista de Figuras


1.1 Exemplo de seleção de quadros significativos. . . . . 2

2.1 Processo de aprendizado de máquina. . . . . 7
2.2 Representações profundas aprendidas do modelo de classificação de digitos. 8
2.3 Representação do aprendizado pelo ajuste dos pesos. . . . . 9
2.4 O papel da função de perda no aprendizado. . . . . 10
2.5 O papel do otimizador no aprendizado. . . . . 12
2.6 Função de ativação sigmoid (esquerda) e tangente hiperbólica (direita). . . . . 16
2.7 Função rectified linear unit (ReLU)). . . . . 16
2.8 Exemplo de mapa de calor em um quadro de tomografia. . . . . 17
2.9 Planos anatômicos visualizados na tomografia. . . . . 20

3.1 Aplicação do método *SSS*. . . . . 25
3.2 Aplicação do método *ESS*. . . . . 26
3.3 Aplicação do método *SIZ*. . . . . 28

4.1 Aplicação de convolução em profundidade e pontual a um volume de tomografia computadorizada . . . . . 42
4.2 Esquema *Grad-Cam Slice Selection (GSS)*. . . . . 44
4.3 Amostras visuais do mapa de ativação em quadros de um volume de TC. . . . 45
4.4 Esquema *Grad-Cam Slice Selection (GSS)*. . . . . 49

5.1 Esquema Geral no qual GSS faz parte do processo para treino e inferência. . 54

5.2 Avaliação de 50 volumes anotados por especialistas. . . . . 57

Capítulo 1

# Introdução

## 1.1 Contextualização e Motivação

O campo de radiologia e diagnóstico por imagem evoluiu sobremaneira nos últimos anos. Os exames de imagem deixaram de ser somente qualitativos e de diagnósticos e passaram a fornecer informações quantitativas e de gravidade de doença. Diante disso, sistemas computadorizados de auxílio diagnóstico vêm sendo desenvolvidos com o objetivo dar suporte à decisão terapêutica. Com o advento da Inteligência Artificial (IA), do *Big Data* e do Aprendizado de Máquina (AM), caminha-se para a rápida expansão do uso dessas ferramentas no dia a dia dos médicos, tornando cada paciente único e levando a radiologia ao encontro do conceito de abordagem multidisciplinar e à medicina de precisão.

Embora volumes de tomografia computadorizada (TC) contenham mais informações que uma única imagem e já existam resultados na literatura que estabelecem que o uso de dados tridimensionais podem ser melhores para o aprendizado de máquina (Huang et al., 2017; Hosny et al., 2018; Li et al., 2016; Milletari et al., 2016), a maioria dos mecanismos de IA não os utilizam. Atualmente, modelos 3D exigem mais memória computacional para o qual o custo computacional envolvido no treino e inferência costuma ser impraticável para grande parte dos sistemas de informação (Ahmed et al., 2018).

A literatura já aborda formas para redução de dados temporais de vídeos, tais como a escolha de quadros equidistantes, extração de quadros em posições mais relevantes ou até mesmo formas de interpolação dos volumes, que serão apresentados na Seção 3.2. Porém, a maioria das técnicas aplica um processo de amostragem estratificada em que podem ser selecionados quadros não relevantes para o problema de AM.

A proposta deste trabalho é apresentar um mecanismo de classificação de volumes de tomografia que realize a extração de quadros significativos de dados volumétricos para serem utilizados por um modelo de aprendizado de máquina. A exemplo da Figura 1.1, um mecanismo de seleção seria responsável em passar para o modelo os quadros mais relevantes.

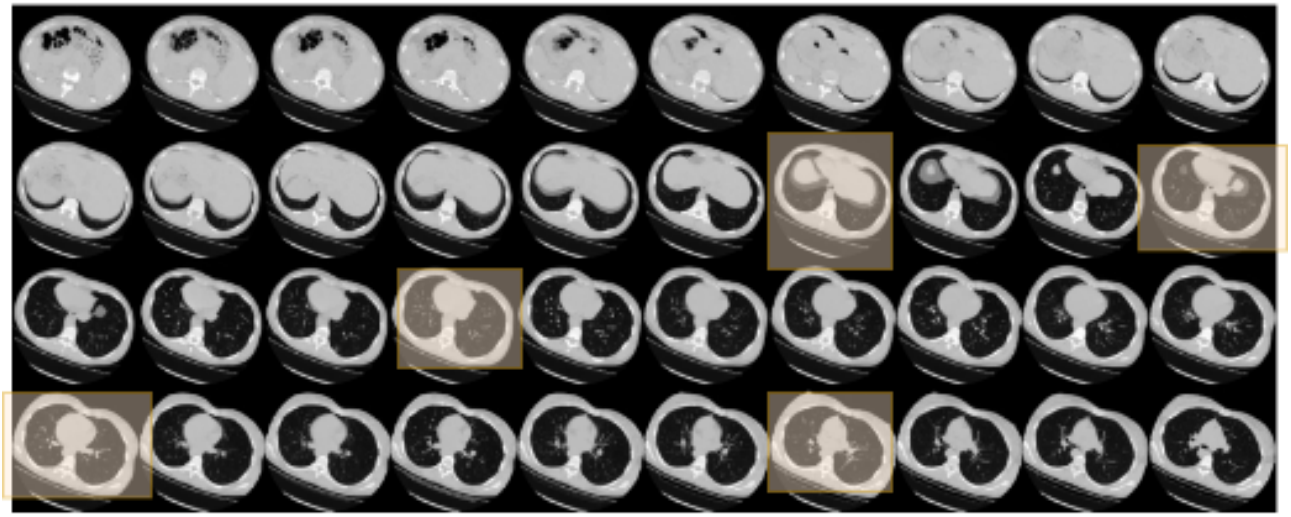

Figura 1.1: Exemplo de seleção de quadros significativos.

Isso pouparia recursos computacionais ao mesmo tempo que possibilitaria o trabalho com maiores resoluções espaciais de entrada, menores tempos de treino e permitiria que o modelo conseguisse extrair características mais relevantes que, sem a seleção, não poderia extrair devido a limitações de *hardware*. Como estudo de caso foi utilizado um conjunto de volumes de tomografia computadorizada dos pulmões para determinar se o paciente possui sequelas causada por Covid-19.

## 1.2 Estudo de Caso

A pandemia de Covid-19 (*Corona Virus Disease* 2019, a doença respiratória causada pelo tipo de coronavírus denominado Sars-CoV-2) destacou a importância da ciência como a principal ferramenta da humanidade na luta contra patologias que afetam a saúde humana. Graças aos esforços colaborativos de várias instituições em todo o mundo, várias vacinas foram desenvolvidas em tempo recorde. No entanto, a busca por tratamentos eficazes e pela erradicação da doença persiste.

Dentre todos os malefícios ocasionados pelo Sars-CoV-2, o dano aos pulmões é o mais preocupante (Del Rio et al., 2020), com consequências muitas vezes irreversíveis nos pacientes que conseguem se recuperar. Imagens radiológicas de tomografia computadorizada são comumente utilizadas por médicos para realizar o diagnóstico e prognóstico do paciente. Com a pandemia, ocorreu uma sobrecarga e ficou evidente que não existem profissionais suficientes para atender a grande demanda de análise das tomografias dos pacientes. Essa realidade se estende aos dias atuais onde profissionais qualificados são poucos em relação à demanda no campo da medicina. O aprendizado de máquina poderá ajudar nessa tarefa a partir de algoritmos inteligentes de aprendizado supervisionado, em especial, com aprendizado profundo.

Como estudo de caso é abordado o desafio de classificar pacientes com sequelas causadas pelo vírus com base em imagens de tomografia computadorizada dos pulmões. O objetivo é desenvolver uma solução que seja precisa, dentro de um limiar de aceitação em relação à eficácia, e que apresente um custo computacional aceitável por meio da abordagem proposta de seleção de quadros significativos. Ao enfrentar essas restrições, busca-se contribuir para o campo da medicina, oferecendo um método eficiente e eficaz para auxiliar os profissionais de saúde na identificação e no tratamento de pacientes.

## 1.3 Objetivos e Hipótese

A tomografia computadorizada é uma técnica de diagnóstico por imagem que produz uma grande quantidade de dados que constituirão um volume, composto de diversos quadros. No entanto, nem todos os quadros são igualmente relevantes para o diagnóstico clínico. Portanto, a identificação dos quadros mais significativos torna-se crucial para otimizar o processo de diagnóstico e reduzir a carga computacional associada. A Hipótese de Pesquisa deste trabalho estabelece que

> “*A avaliação do conhecimento obtido por meio de modelos de aprendizado profundo, em tarefas de classificação de imagens de tomografia computadorizada, possibilita a seleção de quadros significativos para avaliação clínica.*”

A hipótese central deste trabalho de pesquisa é baseada na premissa de que a utilização de modelos de aprendizado profundo, quando aplicados a tarefas de classificação de volumes de tomografia computadorizada, pode facilitar a identificação e seleção de quadros significativos para avaliação clínica por meio da avaliação do conhecimento obtido por esses modelos. Isso, por sua vez, contribuiria para a redução do custo computacional envolvido no processamento dessas imagens e até possíveis melhorias durante o aprendizado de máquina.

Isso sugere que, com o uso de modelos de aprendizado profundo, é possível identificar e selecionar esses quadros relevantes de maneira eficaz. Neste contexto, a pesquisa busca explorar os benefícios e possibilidades oferecidos por essas técnicas avançadas de aprendizado de máquina. Com base nessa hipótese, o objetivo de pesquisa é extrair os quadros mais significativos em volumes de tomografia computadorizada para redução de custo computacional em tarefas de classificação com aprendizado profundo.

A partir desse objetivo geral, alguns objetivos específicos foram estabelecidos:

- Definir modelos e técnicas de aprendizado de máquina para extrair quadros significativos em volumes de tomografia computadorizada.
- Especificar as melhores práticas para gerar instâncias a fim de servirem como insumos para avaliação por modelos de aprendizado de máquina a partir dos volumes de tomografia;

- Reduzir o custo computacional de processamento de imagens, dada a opção de se obter um subconjunto de dados significativos sem perda de desempenho do modelo;
- Prover método de extração de quadros significativos em volumes de tomografia computadorizada por meio de modelos profundos de aprendizado de máquina.

## 1.4 Organização do texto

Este documento está estruturado em cinco capítulos. O Capítulo 1, Introdução, apresentou a visão geral do trabalho e seus objetivos.

No Capítulo 2 são apresentados conceitos teóricos relevantes para a compreensão deste projeto de pesquisa tais como aprendizado de máquina, noções básicas de radiologia e a relação de aprendizado profundo com medicina.

Já no Capítulo 3 são abordadas algumas técnicas de amostragem em tomografia computadorizada em trabalhos relacionados.

O Capítulo 4 descreve a proposta deste trabalho e os métodos empregados.

No Capítulo 5 são mostrados a avaliação experimental e os resultados obtidos.

Finalmente, no Capítulo 6, são feitas considerações finais deste projeto, destacando-se as principais contribuições alcançadas e expondo o plano de trabalho para a continuidade da pesquisa.

Capítulo 2

# Fundamentação Teórica

## 2.1 Considerações Iniciais

Neste capítulo, é proporcionada uma análise abrangente da literatura e teorias existentes que são relevantes para o tema em estudo. O objetivo é estabelecer um contexto geral e fornecer uma compreensão aprofundada das principais técnicas empregadas nesta pesquisa. Além disso, são destacadas as lacunas no conhecimento atual, as quais este trabalho pretende preencher.

Ao longo desta seção, será oferecida uma breve introdução ao aprendizado de máquina e aos conceitos de aprendizado profundo aplicados a volumes de tomografia computadorizada, bem como outras referências relevantes à temática em questão. Vale ressaltar que esta introdução não tem a pretensão de ser exaustiva, ou seja, abordar todos os principais paradigmas e fundamentações envolvidos no campo de estudo. Em vez disso, busca-se definir os conceitos e notações que serão utilizados no decorrer do texto e oferecer uma visão geral do campo para facilitar a compreensão. Dessa forma, este capítulo serve como base teórica para a investigação desenvolvida ao longo desta dissertação.

## 2.2 Aprendizado de Máquina

*Machine Learning* ou Aprendizado de Máquina (AM) é uma área de pesquisa multidisciplinar, frequentemente tida como parte da Inteligência Artificial (IA), que trata do estudo sistemático de algoritmos e sistemas que são capazes de melhorar seu desempenho com a experiência (Alpaydin, 2020). Um algoritmo nesse domínio é capaz de aprender a partir de dados, através da captura de padrões e efetuando inferências com o intuito de solucionar um problema específico.

A forma tradicional para trabalhos com IA emprega o paradigma de programação clássico, o qual envolve a criação manual de regras com dados processados pelas mesmas a fim

de obter respostas como saída. Entretanto, em algoritmo de AM, os dados e as respostas associadas aos dados são obtidos de maneira automática (Francois, 2017). Por exemplo, na classificação de patologias em exames de tomografia computadorizada, um modelo de AM aprende padrões entre o conteúdo dos quadros e a existência ou inexistência de patologia. No que tange o contexto deste trabalho, os quadros mais significativos são obtidos analisando-se quais quadros contêm informação mais significativa para expressar esses padrões.

As técnicas clássicas de aprendizado de máquina incluem algoritmos como *Support Vector Machines* (SVM), *Naive Bayes* e *Random Forest*. Esses algoritmos são baseados em características extraídas "manualmente" e são treinados em dados rotulados para realizar previsões. No reconhecimento de imagens, técnicas clássicas de processamento de imagem são usadas para extrair características dos eexemplos, as quais são posteriormente utilizadas para treinar o algoritmo, por exemplo, em tarefas de reconhecimento de objetos, segmentação etc.

As técnicas de aprendizado profundo, por sua vez, se fundamentam em redes neurais artificiais e não necessitam de características obtidas manualmente. Em vez disso, o algoritmo aprende a extrair características diretamente dos dados por meio de múltiplas camadas de computação. Redes neurais convolucionais (CNNs) representam as técnicas de aprendizado profundo mais utilizadas para reconhecimento de imagem, em que o algoritmo é treinado em extensos conjuntos de dados para realizar alguma tarefa, como identificar objetos.

### 2.2.1 Aprendizado de Máquina aplicado ao reconhecimento de Imagens

O reconhecimento de imagem é um campo em rápido crescimento que tem tido inúmeros avanços nos últimos anos. A capacidade dos algoritmos de reconhecer com precisão objetos, pessoas e outros elementos em imagens tornou o AM uma ferramenta valiosa para várias aplicações.

O aprendizado de máquina para reconhecimento de imagens pode ser análogo à forma que humanos aprendem. Por exemplo, um pai ensinando seu filho a classificar um animal entre cachorro e gato. O pai mostrará um animal ao filho e pedirá que o classifique entre cão ou gato, mostrando a resposta certa quando a criança errar. Com a prática, haverá uma melhoria de desempenho e uma habilidade de **generalização** irá permitir que a criança possa classificar até mesmo exemplos nunca antes vistos.

Semelhantemente, um programa de aprendizado de máquina poderá aprender de maneira supervisionada a partir de imagens, analisando características como formas, cores, texturas e proporções. A avaliação será feita com base no seu poder de generalização e capacidade de classificar imagens nunca antes vistas. Isso ocorre por meio de um processo que se assemelha a recordações do cérebro humano, o qual consegue criar associações a partir de exemplos previamente vistos (Brink et al., 2017). A Figura 2.1 mostra o processo de apren-

dizado de máquina supervisionado, no qual imagens já categorizadas são apresentadas ao modelo, sendo este depois posto à prova.

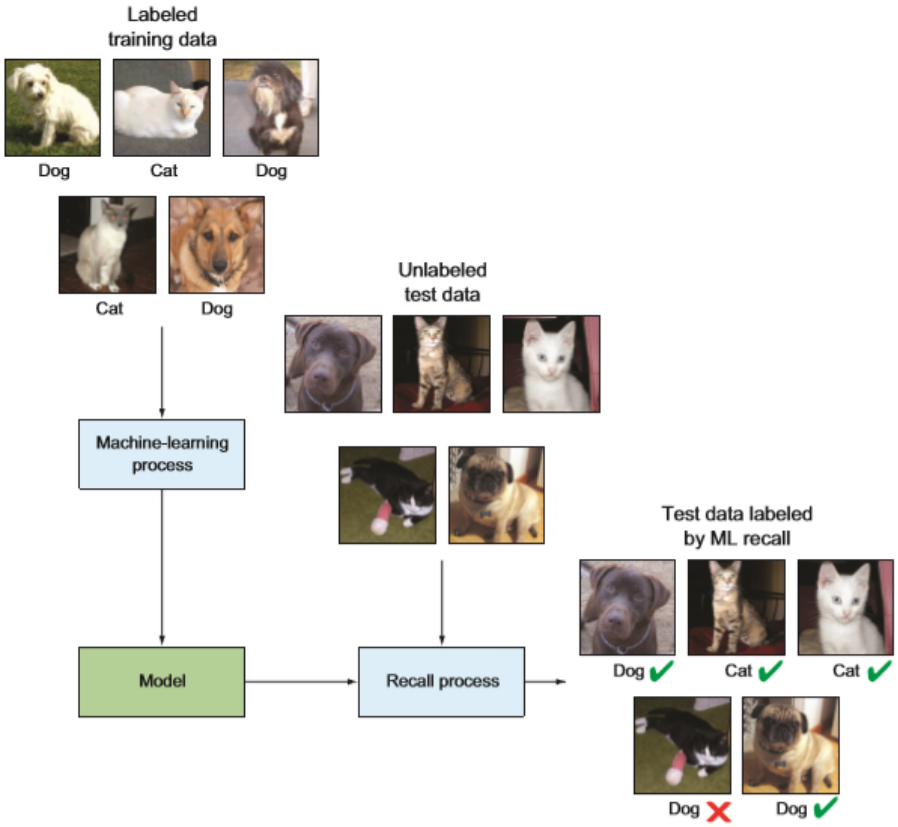


Figura 2.1: Processo de aprendizado de máquina.
Fonte: (Brink et al., 2017).

Em comparação com as técnicas clássicas de aprendizado de máquina, as técnicas de aprendizado profundo mostraram desempenho superior em tarefas de reconhecimento de imagem. Os algoritmos de aprendizado profundo são capazes de aprender relacionamentos mais complexos entre os dados e os rótulos. Além disso, os algoritmos de aprendizado profundo não exigem extração manual de atributos, tarefa esta que pode ser demorada e propensa a erros.

No entanto, os algoritmos de aprendizado profundo também têm algumas desvantagens. Eles exigem grandes quantidades de dados rotulados para treinamento, o que pode ser difícil e demorado de se obter. Além disso, os algoritmos de aprendizado profundo podem ser computacionalmente intensivos e exigir hardware especializado, tornando-os mais difíceis de implementar em comparação com as técnicas clássicas de aprendizado de máquina. Dessa forma, embora as técnicas de aprendizado profundo tenham mostrado desempenho superior, as técnicas clássicas de aprendizado de máquina classificadas ainda podem ser uma opção viável para determinadas aplicações devido à sua simplicidade e facilidade de implementação. Mais pesquisas nessa área poderão levar a algoritmos e técnicas aprimorados para reconhecimento de imagem que podem combinar os pontos fortes do aprendizado de máquina clássico e das técnicas de aprendizado profundo.

## 2.3 *Aprendizado Profundo*

O Aprendizado Profundo, também conhecido como Deep Learning, é uma técnica de aprendizado de máquina baseada em redes neurais artificiais compostas por várias camadas de processamento. Essas camadas são interconectadas, cada uma com um conjunto de neurônios que processam informações e transferem essas informações para a camada seguinte. Com o treinamento de dados, a rede neural é capaz de identificar padrões, reconhecer objetos, categorizar informações e, em geral, realizar tarefas cada vez mais complexas (Francois, 2017).

### 2.3.1 Obtenção de conhecimento

A Figura 2.2 mostra o funcionamento de uma rede neural profunda, exemplificando como as representações são apreendidas por algoritmos de aprendizado profundo. No caso demonstrado, a rede tem o objetivo de categorizar (ou classificar) o dígito expresso nos pixels da imagem de entrada.

Durante o processo de aprendizado, as camadas da rede fazem transformações no sinal de entrada, gerando mapas de características cada vez mais abstratos. Esses mapas possibilitam ao modelo reconhecer padrões complexos nos dados, aumentando a precisão da classificação. As informações assim obtidas podem ser empregadas para identificar os dados mais relevantes em um conjunto de dados.

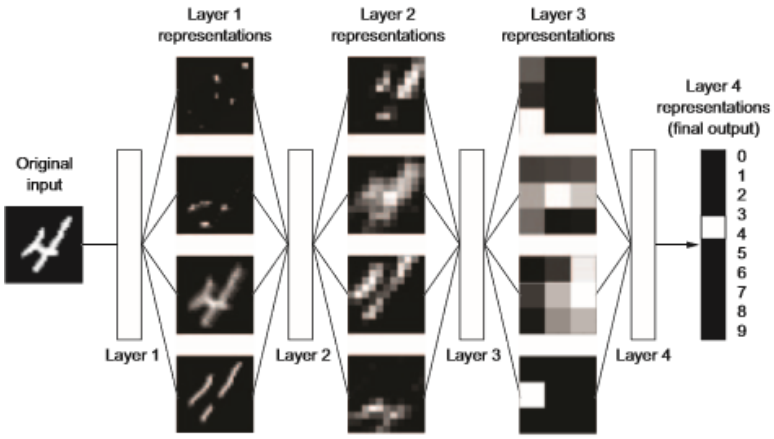


Figura 2.2: Representações profundas aprendidas do modelo de classificação de digitos.
Fonte: (Francois, 2017).

Esse processo pode ser conceitualmente comparado a uma "destilação de dados", na qual que a informação de entrada (*e.g.*, volumes de tomografia) é submetida a uma série de filtros especializados. Esses filtros extraem representações intermediárias cada vez mais abstratas à medida que a informação avança pela rede, culminando na camada de saída que realiza a classificação final do volume. No exemplo dado, a imagem do dígito se transforma, tornando-se distinta do dígito original e, ao mesmo tempo, tornando-se mais informativa para a rede.

Em problemas de classificação é objetivo desta técnica mapear as entradas (*inputs*) a rótulos (*labels*) por meio de um aprendizado derivado da observação de **dados iniciais** que determinam os parâmetros a serem aprendidos pela rede. A especificação do que cada camada faz com os dados de entrada é armazenada como um conjunto de *pesos* da camada (Goodfellow et al., 2016). Em termos técnicos pode-se dizer que as transformações nos dados de entrada feitas por cada camada são parametrizadas por seus pesos, como mostrados na Figura 2.3. Nesse contexto, o *aprendizado* se dá em encontrar o conjunto de valores dos pesos de todas as camadas da rede, de forma que conseguirá mapear corretamente um conjunto de valores de entrada às suas respectivas categoriais.

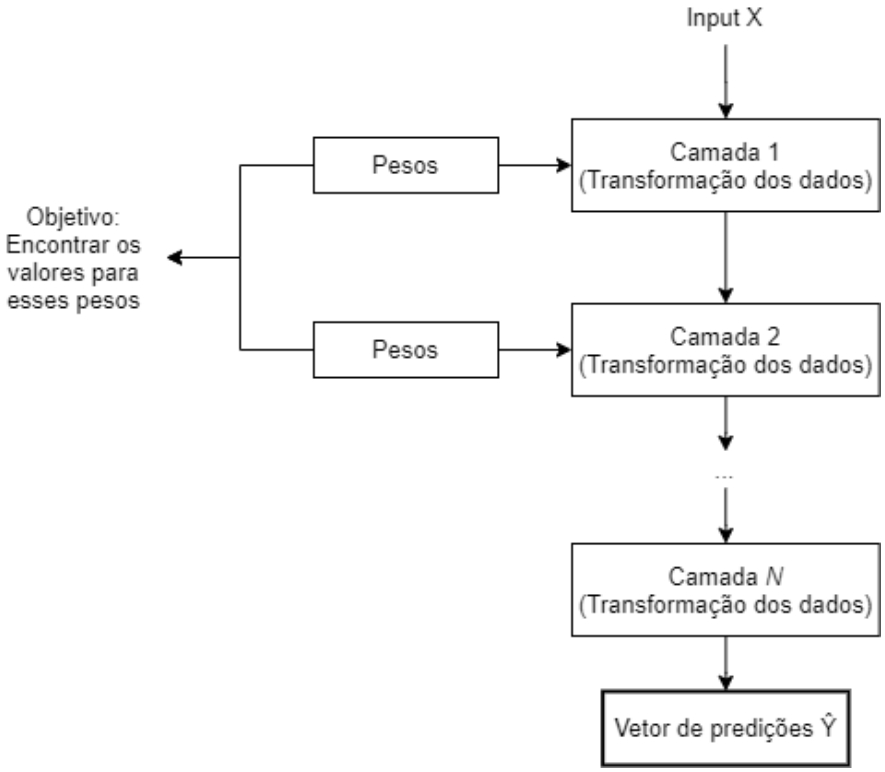


Figura 2.3: Representação do aprendizado pelo ajuste dos pesos.

Em cada iteração do processo de aprendizado é medido o quão longe a saída inferida está da saída esperada por meio da **função de perda** da rede. A função de perda produz um score que reflete o quão bem a rede foi capaz de aproximar a saída esperada da saída real, muitas vezes também chamadas de *ground truth*. Ela pode ser, por exemplo, a distância entre dois vetores ou o erro médio de uma variável-alvo numérica. A Figura 2.4 mostra esse processo.

A utilização dos pontos de perda como sinal de *feedback* para o ajuste dos valores dos pesos das camadas, com o objetivo de minimizar os pontos de perda, propiciará o aprendizado do modelo. Esse ajuste é feito por um otimizador que implementa um algoritmo de retropropagação ou *backpropagation*: o algoritmo central do aprendizado profundo responsável por atualizar os pesos de cada camada. É utilizado o conceito matemático de vetor gradiente que indica a direção na qual se obtém o crescimento máximo do valor de

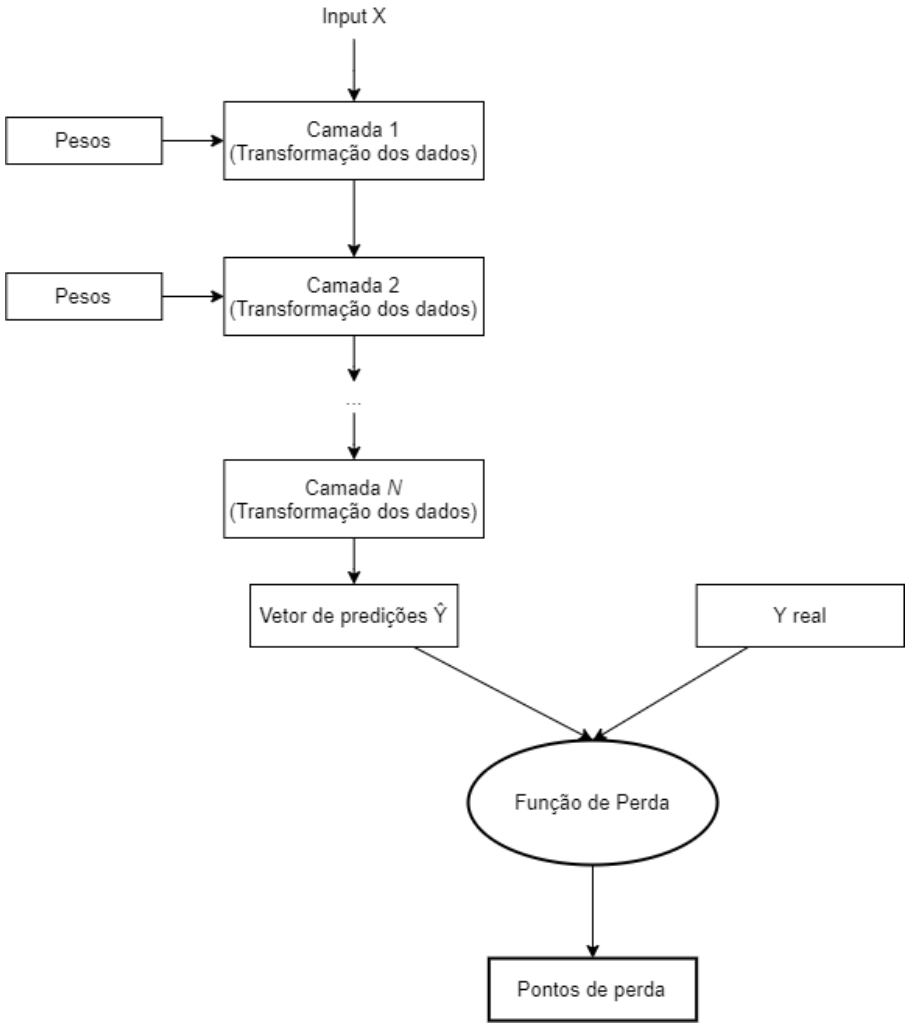


Figura 2.4: O papel da função de perda no aprendizado.

uma função qualquer. Neste contexto, o objetivo é encontrar o mínimo (local) da função de perda, reduzindo os pontos de perda em cada iteração (Goodfellow et al., 2016).

O processo de retropropagação envolve duas etapas: a propagação para frente *(forward pass)* e a propagação para trás *(backward pass)*. Durante a propagação para frente, a entrada é passada através da rede e a saída é calculada. Durante a propagação para trás, o erro é calculado a partir da diferença entre a saída da rede e a saída desejada, e esse erro é propagado de volta pela rede, ajustando os pesos das conexões para compensar a perda.

Para exemplificar esse processo, considere uma rede neural com $L$ camadas, na qual a camada $l$ tem $n^{(l)}$ unidades e cada unidade $i$ na camada $l$ é denotada por $a_i^{(l)}$. A saída da rede neural é dada por $h_\Theta(x)$, na qual $x$ é a entrada e $\Theta$ são os pesos da rede.

O processo de retropropagação inicia calculando o erro na camada de saída da rede. Seja $y$ o rótulo verdadeiro correspondente à entrada $x$, a função de perda é comumente definida como a função de erro quadrático médio (MSE), conforme demonstrado na Equação 2.1:

$$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\Theta(x^{(i)}) - y^{(i)})^2, \tag{2.1}$$

onde $m$ é o número de exemplos de treinamento.

O gradiente com relação aos pesos da última camada $L$ é então calculado conforme a Equação 2.2:

$$\delta^{(L)} = \nabla_{a^{(L)}} J(\Theta) \odot \sigma'(z^{(L)}), \tag{2.2}$$

onde $\sigma'(z^{(L)})$ é a derivada da função de ativação aplicada à entrada $z^{(L)}$ da última camada $L$, e $\odot$ denota o produto entre dois vetores.

O erro é então propagado para as camadas anteriores da rede neural usando a Equação 2.3:

$$\delta^{(l)} = ((\Theta^{(l)})^T \delta^{(l+1)}) \odot \sigma'(z^{(l)}), \tag{2.3}$$

onde $\Theta^{(l)}$ é a matriz de pesos entre as camadas $l$ e $l+1$, $z^{(l)}$ é a entrada da camada $l$, e $\sigma'(z^{(l)})$ é a derivada da função de ativação aplicada à entrada $z^{(l)}$.

O gradiente com relação aos pesos de cada camada é então calculado conforme a Equação 2.4:

$$\nabla_{\Theta^{(l)}} J(\Theta) = \delta^{(l+1)} (a^{(l)})^T, \tag{2.4}$$

onde $a^{(l)}$ é o vetor de ativações da camada $l$.

Com os gradientes calculados em relação aos pesos da rede neural em cada iteração do algoritmo de otimização, os pesos são então atualizados na direção oposta ao gradiente (descida de gradiente), de acordo com uma taxa de aprendizado que controla a magnitude da atualização, conforme a Equação 2.5:

$$\Theta^{(l)} := \Theta^{(l)} - \alpha \nabla_{\Theta^{(l)}} J(\Theta), \tag{2.5}$$

onde $\alpha$ é a taxa de aprendizado.

O processo de retropropagação é repetido por várias épocas até que o modelo seja treinado o suficiente para fazer previsões precisas em novos dados ou até que uma condição pré-estabelecida seja alcançada(Goodfellow et al., 2016).

A Figura 2.5 exemplifica o ciclo completo de aprendizagem. O processo todo é feito em um *loop* de treino no qual, em cada iteração, são ajustados os pesos de forma a minimizar os

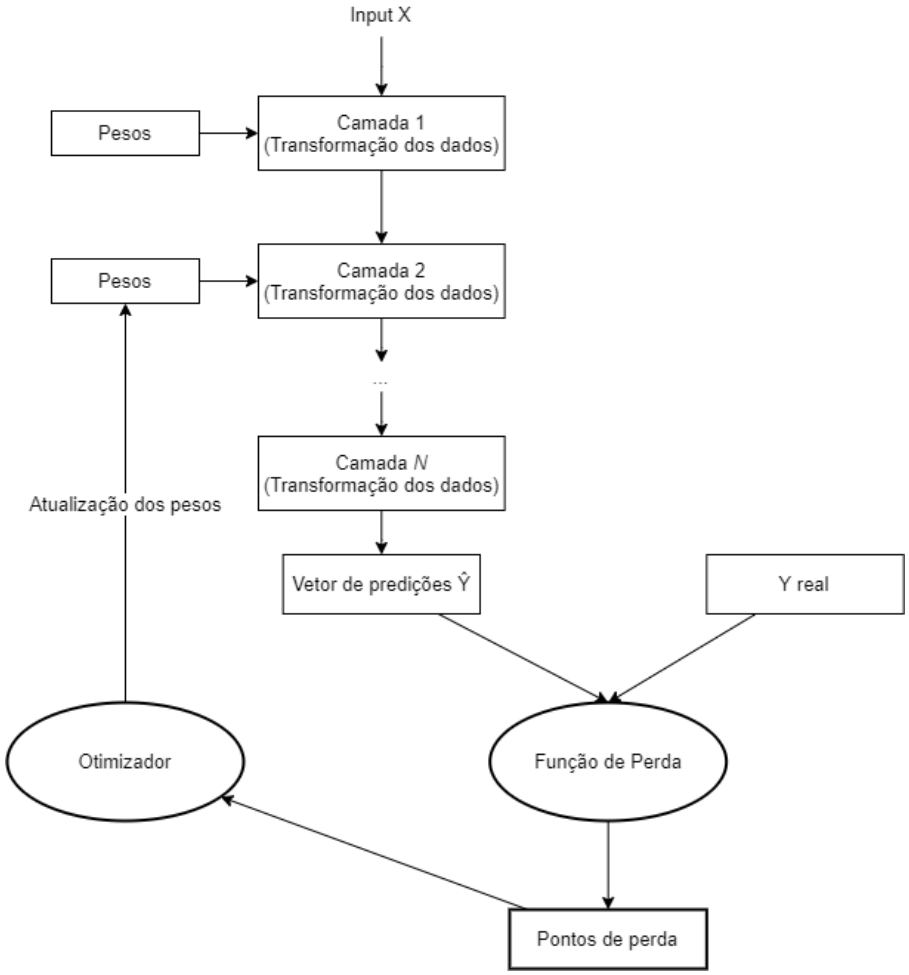


Figura 2.5: O papel do otimizador no aprendizado.

pontos de perda. O resultado final é um grafo de camadas com os pesos ajustados para resolução de um problema para um conjunto de entrada de tipo específico.

### 2.3.2 Convolução em aprendizado profundo

A convolução é uma operação matemática usada para "mesclar" duas funções. No aprendizado profundo, é usada em redes neurais convolucionais para gerar representações mais abstratas dos sinais de entrada, como mostrado na Equação 2.6,

$$f * g = \int_{-\infty}^{+\infty} f(t)g(x-t)dt, \tag{2.6}$$

na qual $f$ e $g$ são funções que representam os dados de entrada e o kernel (filtro) de convolução, respectivamente. Observe que essa equação se aplica apenas a um neurônio da rede. A saída dessa operação é um valor numérico que reflete a transformação do sinal $f$ de acordo com o filtro $g$.

A **convolução em profundidade ou *depthwise convolution*** é uma variação da operação de convolução padrão que opera em cada canal de um tensor de entrada independentemente. A operação padrão aplica um conjunto compartilhado de filtros a todos os canais de um tensor de entrada, enquanto a operação em profundidade aplica um conjunto separado de filtros a cada canal do tensor de entrada.

Já a operação de convolução em profundidade pode ser matematicamente representada como mostra a Equação 2.7,

$$f_i * g_i = \int_{-\infty}^{+\infty} f_i(t) g_i(x - t) dt, \tag{2.7}$$

na qual $f_i$ e $g_i$ são funções que representam o i-ésimo canal do tensor de entrada e o i-ésimo filtro, respectivamente. A saída da operação de convolução em profundidade é um conjunto de mapas de características, um para cada canal do dado de entrada.

A convolução é bastante utilizada em aprendizado profundo por vários motivos, dentre os quais:

- Eficiência do modelo: pode ser usada para aumentar a eficiência dos modelos de aprendizado profundo, pois reduz o número de parâmetros em um modelo.
- Interpretabilidade do modelo: para aumentar a interpretabilidade dos modelos de aprendizado profundo, pois permite que cada canal de um tensor de entrada seja processado independentemente.
- Reconhecimento de objetos: pode ser utilizada na área de visão computacional para reconhecimento de objetos.

### 2.3.3 Custo Computacional

O aprendizado profundo tem sido amplamente adotado em vários campos, incluindo visão computacional, processamento de linguagem natural e reconhecimento de fala. Um dos principais desafios é equilibrar a compensação entre precisão e custo computacional.

A dimensão de entrada refere-se ao número de características em uma instância de dados usada para treinar um modelo de aprendizado profundo. Aumentar a dimensão de entrada pode levar a uma melhor representação dos dados, mas também aumenta o custo computacional. Isso ocorre porque dimensões de entrada maiores requerem mais recursos computacionais para processar, armazenar e manipular.

O custo computacional de um algoritmo se refere ao número de operações básicas como multiplicações, somas, dentre outras operações do algoritmo. Também refere-se à quantidade de recursos computacionais necessários para treinar um modelo de aprendizado profundo. Isso inclui a quantidade de memória, poder de processamento e tempo neces-

sário para treinar o modelo. Em relação à complexidade computacional, normalmente é expressa em função do tamanho da entrada $n$ na forma $T(n)$ (Cormen et al., 2009).

O exemplo abaixo exemplifica o custo de uma convolução de entrada $n = 9$. Assumindo um *kernel* $\mathbf{H} \in \mathbb{R}^{3\times3}$ e um dado de entrada $\mathbf{I} \in \mathbb{R}^{3\times3}$, passo igual a 1 e parâmetro preenchimento igual a zero, tem-se 9 multiplicações e 8 somas, portanto foram feitas um total 17 operações só nesta operação de convolução.

$$\mathbf{I} \circledast \mathbf{H} = \begin{bmatrix} i_{11} & i_{12} & i_{13} \\ i_{21} & i_{22} & i_{23} \\ i_{31} & i_{32} & i_{33} \end{bmatrix} \odot \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \sum_{ij} \begin{bmatrix} i_{11}h_{11} & i_{12}h_{12} & i_{13}h_{13} \\ i_{21}h_{21} & i_{22}h_{22} & i_{23}h_{23} \\ i_{31}h_{31} & i_{32}h_{32} & i_{33}h_{33} \end{bmatrix}$$
$$= i_{11}h_{11} + i_{12}h_{12} + i_{13}h_{13} + i_{21}h_{21} + i_{22}h_{22} + i_{23}h_{23} + i_{31}h_{31} + i_{32}h_{32} + i_{33}h_{33}$$

Em comparação com uma convolução 2D padrão, uma convolução 3D, como aquela usada em tomografia computadorizada, requer muito mais operações. A diferença pode ser significativa especialmente quando o volume de dados é grande. Considerando um dado de entrada de tamanho (512,512,100), o número total de operações realizadas, durante essa mesma operação de convolução, pode ser calculada pela seguinte fórmula:

$$\begin{aligned} \text{Número de operações} = \text{Tamanho do kernel}^2 \times \text{Número de canais de entrada} \times \\ \text{Número de saídas} \times \text{Tamanho da imagem} \end{aligned} \tag{2.8}$$

Onde o tamanho do *kernel* refere-se à altura ou largura do *kernel* (supondo que seja quadrado), o número de canais de entrada é o número de canais na imagem de entrada, o número de saídas é o número de filtros aplicados à imagem de entrada e o tamanho da imagem é a altura ou largura da imagem de entrada.

$$\text{Número de operações} = 3^2 \times 100 \times 510 \times 510 = 7,834,800,000$$

Usando-se a Equação 2.8 para calcular o número de operações para um volume de tomografia computadorizada de tamanho (512,512,100), obtém-se um total de 7.834.800.000 operações para um único volume em uma única convolução. Isso é significativamente maior do que o número de operações para uma convolução 2D comum, ilustrando o aumento do custo computacional à medida que a dimensão dos dados de entrada aumenta.

Além disso, em redes neurais profundas, o custo geralmente é avaliado pela soma do número de parâmetros de suas camadas. Esses parâmetros são os pesos que são aprendidos durante o treinamento. Portanto, quanto maior o número de parâmetros, maior o custo da rede. Como mostrado na Seção 2.3, à medida que a dimensão de entrada aumenta, o custo computacional dos modelos de aprendizado profundo também aumenta (Yue et al., 2020; Justus et al., 2018; Thompson et al., 2020).

Para mitigar o impacto do aumento da dimensão da entrada no custo computacional, será necessário empregar técnicas que envolvam o redimensionamento do dados de entrada ou a compactação do modelo. Essas abordagens podem ajudar a manter o custo computacional sob controle, mesmo quando lidando com grandes volumes de dados, como no caso de tomografias computadorizadas.

O impacto da dimensão de entrada no custo computacional dos modelos de aprendizado profundo pode ser significativo. Grandes dimensões de entrada podem levar a um aumento significativo no tamanho do modelo, tornando-o computacionalmente mais intensivo para treinar (LeCun et al., 2015; Tan e Le, 2019). Isso pode resultar em tempos de treinamento mais longos e um risco maior de overfitting, onde o modelo se ajusta excessivamente aos dados de treinamento, resultando em baixo desempenho nos dados de teste (Srivastava et al., 2014).

### 2.3.4 Função de ativação

Nas camadas convolucionais (incluindo a de convolução transposta), os filtros realizam convoluções em todos os canais, os valores são somados e a eles é acrescentado um viés. Porém, antes que o resultado seja colocado no mapa de características para ser enviado à próxima camada, uma função de ativação é aplicada sobre ele, assim como ocorre nos neurônios de camadas totalmente conectadas, típicos de redes neurais clássicas.

Essa função de ativação tem como objetivo acrescentar características não lineares ao problema, permitindo a modelagem de cenários mais complexos. Caso uma função de ativação linear fosse utilizada, toda a rede poderia ser reduzida a uma única camada convolucional, independente da profundidade da rede. Portanto, funções de ativação são um elemento extremamente importante das redes neurais artificiais. Elas basicamente decidem se um neurônio deve ser ativado ou não. Ou seja, se a **informação que o neurônio está recebendo é relevante ou deve ser ignorada**, como mostra a equação abaixo:

$$\mathrm{Y} = \mathrm{Activation}\left(\sum_i (\mathrm{weight}_i * \mathrm{input}_i) + \mathrm{bias}\right) \tag{2.9}$$

Tradicionalmente, duas funções de ativação foram mais utilizadas: a função *sigmoid* e a função tangente hiperbólica (*tanh*), ilustradas na Figura 2.6. Contudo, à medida que redes mais profundas foram criadas, percebeu-se que essas funções são especialmente suscetíveis ao problema do desaparecimento do gradiente, em que os gradientes das primeiras camadas da rede ficam tão pequenos que os seus parâmetros quase não são alterados.

Para evitar esse problema, a função ReLU, cujo gráfico é mostrado na Figura 2.7, começou a se popularizar e, na prática, tende a mostrar melhor desempenho para convergência dos modelos (Krizhevsky et al., 2012).

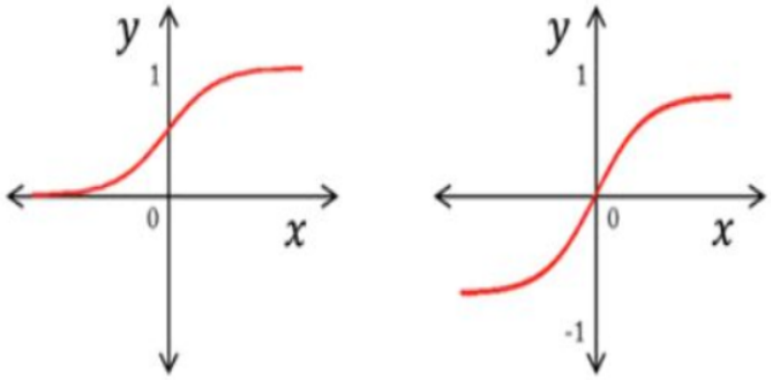


Figura 2.6: Função de ativação sigmoid (esquerda) e tangente hiperbólica (direita).
Fonte: (Khan et al., 2018) (Adaptado).

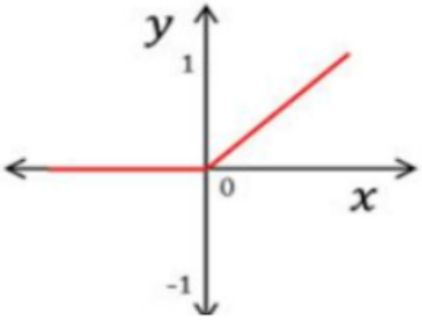


Figura 2.7: Função rectified linear unit (ReLU)).
Fonte: (Khan et al., 2018) (Adaptado).

### 2.3.5 Interpretações das ativações com Grad-Cam

As características que ativaram camadas intermediárias da CNN podem fornecer dicas úteis sobre a qualidade das representações aprendidas. A ideia é que os pesos (parâmetros) da rede são maiores nas "regiões" mais importantes da entrada. Essa informação da rede pode ser obtida avaliando-se o mapa de características após a última camada convolucional de uma CNN treinada para um problema de classificação. Esse mapa de atenção gerado pode dar um *insight* de quais regiões da imagem receberam maior atenção do modelo. Em outras palavras, **é possível visualizar quais partes da entrada contribuíram mais para a predição**. A Figura 2.8 mostra um exemplo de um modelo treinado com imagens bidimensionais para classificação de volumes de tomografia para existência de anomalias causadas por Covid-19 ou pulmões saudáveis. A imagem à esquerda equivale ao mapa de calor, a do centro mostra o mapa de calor sobreposto à imagem original e à direita tem-se a imagem do quadro do volume. Observa-se em verde as regiões da imagem que foram consideradas mais importantes para extração de características. Observe que elas correspondem à região do pulmão do paciente, enquanto outras regiões foram ignoradas.

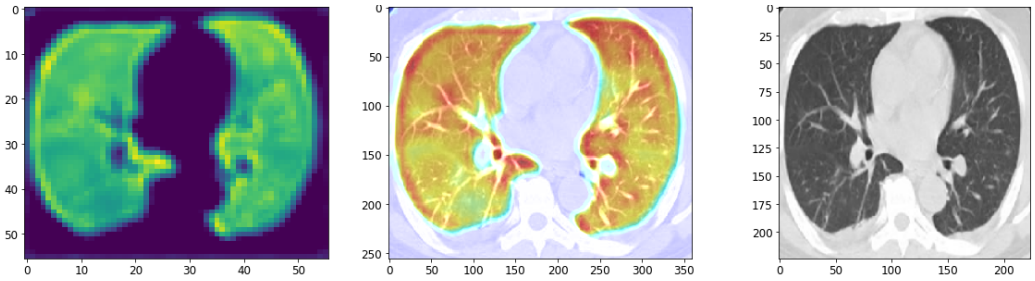

Figura 2.8: Exemplo de mapa de calor em um quadro de tomografia. (Ideal visualizar em cores)

Balduzzi et al. (2017) apresentaram a ideia de que a visualização das distribuições de gradientes fornecem informações úteis do comportamento de redes neurais profundas. A análise nesse trabalho mostrou que gradientes retropropagados dentro de uma CNN podem ser usados para identificar padrões específicos da imagem de entrada que maximizam a ativação de um determinado neurônio em uma camada da CNN. Em outras palavras, gradientes podem ser ajustados para gerar visualizações que ilustram os padrões que um neurônio usou para procurar informações úteis nos dados de entrada.

Baseado nessa ideia, Selvaraju et al. (2017) propuseram a técnica do **Gradient-weighted Class Activation Mapping (Grad-CAM)**, que utiliza os gradientes que fluem da última camada convolucional de uma rede neural para realizar uma tarefa específica e, assim, produzir um mapa de localização visual. Esse mapa destaca as regiões da imagem que são consideradas importantes pelo modelo. A técnica Grad-CAM generaliza o mapa de ativação de classe (CAM) da rede, conforme demonstrado no mesmo trabalho.

A técnica Grad-CAM é baseada em gradientes e utiliza as ativações da última camada convolucional $A^k$ de uma rede neural. A fórmula do Grad-CAM é ilustrada na Equação 2.10(Selvaraju et al., 2017).

$$L^c_{GradCAM} = ReLU(\sum_k \alpha^c_k A^k) \tag{2.10}$$

Nesta equação, $L^c_{GradCAM}$ representa o mapa de ativação da classe $c$, $A^k$ é a ativação da última camada convolucional para a k-ésima característica, e $\alpha^c_k$ é um peso atribuído à k-ésima característica para a classe $c$.

Os pesos $\alpha^c_k$ são calculados usando os gradientes da saída em relação às ativações da última camada convolucional:

$$\alpha^c_k = \frac{1}{Z} \sum_i \sum_j -\frac{\partial y^c}{\partial A^k_{i,j}} \tag{2.11}$$

Aqui, $y^c$ é a saída para a classe $c$, e $Z$ é uma constante de normalização.

O resultado final é um mapa de calor que mostra as regiões mais importantes da imagem para a saída da rede neural. Enquanto o CAM funciona melhor apenas com redes que possuem uma camada global de *pooling* médio (*Global Average Pooling*) antes da camada de classificação (Lin et al., 2013), o Grad-CAM pode ser aplicado a uma variedade maior de arquiteturas de rede neural, incluindo CNNs e redes com camadas totalmente conectadas (Dense Layers). Tornando-o mais adequado para comparação e avaliação de diferentes arquiteturas de aprendizado profundo (Chattopadhay et al., 2018).

A técnica Grad-CAM também pode ser aplicada a dados tridimensionais, como volumes médicos, para identificar as regiões mais importantes que influenciam a saída de uma rede neural. Essa variação é chamada de 3D Grad-CAM. A fórmula para calcular o mapa de ativação 3D Grad-CAM é semelhante à Equação 2.10, mas utiliza a ativação da última camada convolucional 3D.

Os pesos $\alpha_k^c$ para o 3D Grad-CAM são calculados de maneira análoga à Equação 2.11, mas com as derivadas parciais em relação às ativações da última camada convolucional 3D.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \sum_k \frac{\partial y^c}{\partial A_{i,j,k}^l} \tag{2.12}$$

Aqui, $y^c$ é a saída para a classe $c$, e $Z$ é uma constante de normalização.

O 3D Grad-CAM pode ser usado, por exemplo, para identificar as regiões mais importantes em um volume médico que contribuem para uma determinada decisão da rede neural, como a presença de uma patologia ou a classificação de uma imagem. Isso pode ajudar na interpretação de resultados e no desenvolvimento de modelos mais confiáveis e robustos. Dessa forma, o mapa de calor terá a seguinte forma de saída conforme as dimensões de entrada:

- Grad-CAM 2D: Para uma imagem de entrada com dimensões $(H, W, C)$, onde $H$ é a altura, $W$ é a largura e $C$ é o número de canais, a saída do mapa de calor do Grad-CAM 2D terá dimensões $(H', W')$. Essas dimensões $(H', W')$ são as dimensões do mapa de características da camada convolucional escolhida e geralmente são menores que as dimensões originais da imagem de entrada $(H, W)$ devido às operações de convolução e pooling.
- Grad-CAM 3D: Para um volume de entrada com dimensões $(H, W, D, C)$, onde $H$ é a altura, $W$ é a largura, $D$ é a profundidade e $C$ é o número de canais, a saída do mapa de calor do Grad-CAM 3D terá dimensões $(H', W', D')$. Essas dimensões $(H', W', D')$ são as dimensões do mapa de características da camada convolucional 3D escolhida e, como no Grad-CAM 2D, tendem a ser menores que as dimensões originais do volume de entrada $(H, W, D)$.

Grad-CAM pode ser usado para uma variedade de tarefas, incluindo:

- Interpretação do modelo: para interpretar o processo de tomada de decisão de modelos de aprendizado profundo, permitindo que pesquisadores e profissionais entendam melhor como esses modelos fazem previsões.
- Depuração de modelo: para depurar modelos de aprendizado profundo, particularmente no contexto de visão computacional, destacando as regiões de uma imagem nas quais o modelo está tomando decisões e identificar possíveis viés ou problemas na base de dados.
- Melhoria do modelo: para melhorar os modelos de aprendizado profundo, identificando as regiões de uma imagem que o modelo não está considerando ou está considerando incorretamente.

## 2.4 Tomografia Computadorizada

Na busca constante por um diagnóstico mais preciso e detalhado, o avanço tecnológico fez chegar com ainda mais vantagens o uso da tomografia. A Tomografia Computadorizada Volumétrica é uma técnica de obtenção de imagem que utiliza um feixe cônico de radiação associado a um receptor de imagens bidimensional. Nessa técnica, o conjunto fonte de raios X e receptor de imagens giram 360° em torno da região de interesse. Durante esse giro, múltiplas projeções bidimensionais em ângulos diferentes são obtidas e enviadas a um computador. Essas projeções contêm toda a informação necessária para compor a matriz da imagem em 3D. Cortes nos três planos (axial, coronal e sagital, como ilustrado na Figura 2.9) do paciente podem então ser obtidos a partir dessa imagem tridimensional. Os volumes são obtidos em diversos formatos como NII (*Neuroimaging Informatics Technology Initiative) ou DICOM (Digital Imaging and Communications*); este último é o padrão para armazenamento e transmissão de informações médicas, principalmente imagens, e laudos de radiologistas experientes responsáveis pela interpretação das imagens.

As imagens digitais são formadas por pequenos pontos, a menor unidade destas, que são pequenos quadrados com medidas laterais idênticas, largura ($x$) e altura ($y$), sendo chamado de pixel. Como a tomografia é um volume tridimensional, um novo plano é adicionado, a profundidade ($z$), constituindo então não mais um quadrado e sim um cubo, chamado voxel. Teoricamente, quanto menor o tamanho do voxel, mais nítida tende a ser a imagem, mas outros fatores como a qualidade do sensor, projeto do aparelho, estabilidade do paciente e software interferem na nitidez final. A granularidade desses voxels pode mensurada em *Hounsfield units* (HU), que é uma escala quantitativa que descreve a radiodensidade.

Dependendo do eixo em observação, a imagem resultante tem nomes diferentes, dependendo da posição do observador, como mostra a Figura 2.9, sendo:

- Plano horizontal, transverso ou axial - Eixo Z. Divide o corpo nas porções cranial (superior) e caudal (inferior)

- Plano coronal - Eixo Y. Divide o corpo nas porções anterior (frente) e posterior (costas).
- Plano sagital - Eixo X. Divide o corpo nas porções esquerda e direita.

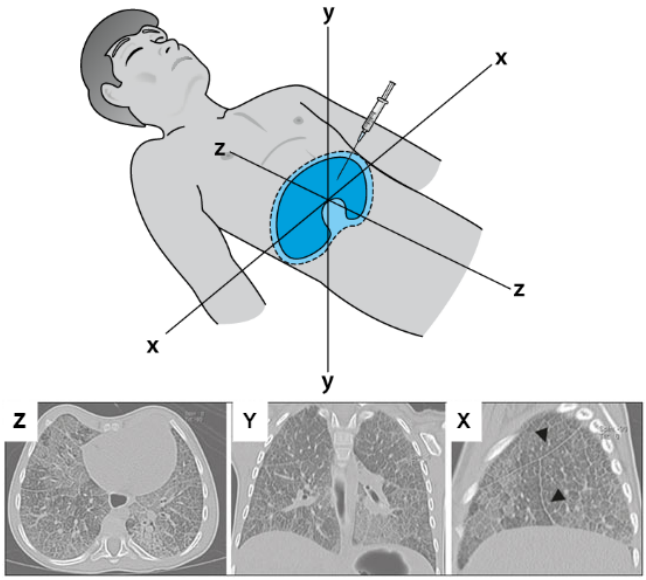


Figura 2.9: Planos anatômicos visualizados na tomografia.
Fonte: (Roberts, 2015) Adaptado.

### 2.4.1 Aplicação no Aprendizado Profundo

Os dados de tomografia computadorizada (TC) e raio-x são usados em aplicações de aprendizado profundo, mas é verdade que os dados de raio-x são mais comuns do que os de TC. Existem algumas razões pelas quais isso pode ocorrer:

- Disponibilidade de dados: Os dados de raio-x são mais fáceis de se obter do que os dados de TC, visto que os exames de raio-x são menos dispendiosos e tomam menos tempo do que os de tomografias. Além disso, a maioria dos hospitais e clínicas tem equipamentos de raio-x disponíveis, enquanto a disponibilidade de equipamentos de TC pode ser mais limitada.
- Tamanho dos dados: O tamanho de dados de TC é muito maior em comparação aos de raio-x. Os arquivos de imagem de TC podem chegar a vários gigabytes, enquanto as imagens de raio-x geralmente têm alguns megabytes. Isso significa que os dados de TC requerem muito mais espaço de armazenamento e recursos de processamento, o que pode tornar o treinamento de modelos de aprendizado profundo com esses dados mais desafiador.
- Dificuldade de processamento: As imagens de TC são complexas e têm mais informações que as imagens de raio-x, o que pode tornar o processamento dos dados mais difícil. Além disso, as imagens de TC geralmente têm artefatos de imagem que precisam ser removidos antes de poderem ser usados para treinar modelos de aprendizado profundo.

No entanto, é importante notar que, embora os dados de raio-x sejam mais comuns no uso de aprendizado profundo, os dados de TC têm suas próprias vantagens e podem ser usados em uma variedade de aplicações médicas. No caso do diagnóstico de doenças que afetam o pulmão, como Covid-19, os dados de TC podem ser mais úteis do que os dados de raio-x. Isso se deve em grande parte à capacidade da TC de produzir imagens em 3D com alta resolução, permitindo a identificação de opacidades pulmonares, que são comuns em pacientes com Covid-19.

O aprendizado profundo pode ser usado para analisar essas imagens e identificar padrões específicos de opacidades pulmonares associados à Covid-19. Esses padrões podem ser difíceis de detectar com o olho humano, mas são facilmente identificados por um algoritmo de aprendizado profundo treinado com dados de TC.

Além disso, a TC pode fornecer informações mais precisas sobre a extensão e localização da infecção. É crucial ressaltar a importância dessa abordagem no contexto do tratamento de patologias, pois a localização das áreas afetadas pode afetar a escolha do tratamento e o prognóstico do paciente.

Essas vantagens se estendem a diferentes aplicações e casos no uso de TC com métodos de aprendizado de máquina. Esse método pode se revelar extremamente benéfico em áreas como radiologia, oncologia e cardiologia, onde a precisão das imagens pode ser crucial para o diagnóstico e tratamento. Por exemplo, em um estudo sobre o uso de aprendizado profundo em imagens de TC para o diagnóstico de câncer de pulmão, os pesquisadores relataram uma precisão de diagnóstico significativamente maior do que com o uso de raio-x (Ardila et al., 2019).

Quanto ao plano anatômico a ser avaliado como insumo de modelos de IA, o plano axial é comumente utilizado na avaliação de dados de TC, devido à sua padronização e consistência na anatomia do corpo humano. As estruturas anatômicas no corpo humano são orientadas em relação ao plano axial, e, portanto, o uso desse plano facilita a localização e a identificação de estruturas específicas dentro da imagem, sem a necessidade de marcadores. Logo, é de suma importância em aplicações de visão computacional que buscam identificar estruturas específicas dentro de um conjunto de imagens de TC.

Além disso, a avaliação de dados de TC no plano axial fornece uma representação bidimensional do corpo humano, o que torna a análise mais fácil e acessível para algoritmos de visão computacional. Como resultado, a análise de dados de TC no plano axial permite a identificação e análise de estruturas e anomalias de maneira mais eficiente e precisa.

## 2.5 Considerações Finais

Este capítulo explorou os principais temas técnicos a serem utilizados neste trabalho. Tanto as técnicas clássicas de aprendizado de máquina quanto as técnicas de aprendizado profundo têm suas vantagens e desvantagens para o reconhecimento de imagens. Embora

as técnicas de aprendizado profundo tenham mostrado desempenho superior, as técnicas clássicas de aprendizado de máquina ainda podem ser uma opção viável para determinadas aplicações devido à sua simplicidade e facilidade de implementação. Mais pesquisas nessa área provavelmente levarão a algoritmos e técnicas aprimorados para reconhecimento de imagem que podem combinar os pontos fortes do aprendizado de máquina clássico e das técnicas de aprendizado profundo.

Quanto ao impacto do custo computacional, a dimensão de entrada dos dados nos modelos de aprendizado profundo pode ser significativa. Grandes dimensões de entrada podem levar a um aumento significativo no tamanho do modelo e torná-lo mais computacionalmente intensivo para realização de treinamentos. Para mitigar o impacto do aumento da dimensão de entrada no custo computacional, várias técnicas podem ser usadas, incluindo seleção de características, redução de dimensionalidade e compactação de modelo. Mais pesquisas são necessárias para entender completamente o impacto da dimensão de entrada no custo computacional dos modelos de aprendizado profundo e desenvolver métodos mais eficazes para equilibrar a compensação entre precisão e custo computacional.

Nesse contexto, o Grad-CAM é uma técnica valiosa para visualizar o processo de tomada de decisão de modelos de aprendizado profundo. Grad-CAM pode ser usado para interpretação, depuração e final melhoria dos modelos. Com a interpretação e quantificação das regiões mais relevantes, esta técnica pode ser usada para selecionar dados mais significativos que tiveram maior contribuição para tomada de decisão.

Capítulo 3

# Trabalhos Relacionados

## 3.1 Considerações Iniciais

Os volumes de TC frequentemente possuem muitos quadros, o que pode tornar o processamento desses dados demorado e intensivo em termos de recursos computacionais. Caso esses dados sejam empregados como insumos para modelos de aprendizado profundo, torna-se imprescindível o seu tratamento prévio. Conforme explanado na Seção 2.3.3, o custo computacional para treinar e processar dados de TC pode ser inviável para a maioria dos sistemas computacionais.

Dessa forma, a subamostragem de quadros no processo de pré-processamento é uma etapa crítica para modelos de aprendizado de máquina aplicados à TC, mas essa redução ainda assim devem representar adequadamente a anatomia ou patologia em questão. Da mesma maneira, a aplicação de técnicas para redução do volume de dados deve ser realizada para tornar viável o treinamento de modelos de aprendizado profundo. Algumas opções utilizam técnicas de amostragem, segmentação e redução de dimensionalidade:

- Reamostragem: consiste na seleção de quadros por meio de um algoritmo específico que seleciona o subconjunto de dados adequado.
- Segmentação: método que identifica e separa regiões de interesse nos volumes de TC, de forma semi-automática ou totalmente automática.
- Redução de dimensionalidade: abordagem que transforma os volumes de TC em uma representação de menor dimensionalidade, permitindo a análise de grandes conjuntos de dados de forma mais eficiente.

A seleção cuidadosa dos quadros pode aprimorar a eficiência do processamento de dados, além de reduzir a quantidade de dados desnecessários e elevar a qualidade dos resultados finais (Rajpurkar et al., 2018). Essa etapa de pré-processamento é crucial em modelos de

aprendizado de máquina, pois pode impactar significativamente a precisão e a eficácia dos modelos que utilizam dados de tomografia computadorizada.

Existem várias técnicas e métodos que visam extrair os dados mais significativos de uma base dados com os próprios modelos de aprendizado profundo. Alguns desses métodos utilizam técnicas de processamento de imagens, como filtragem e segmentação, para identificar regiões de interesse e, em seguida, extrair características dessas regiões. Outros métodos utilizam abordagens de aprendizado de máquina, como CNNs e autocodificadores, para identificar padrões complexos nas imagens.

No entanto, ainda há desafios significativos a serem superados na extração de quadros significativos de volumes de TC. A variação na aparência das imagens, bem como o tamanho e complexidade dos volumes, tornam a tarefa de identificação de características relevantes um processo difícil e demorado. Além disso, a falta de uma grande quantidade de dados rotulados pode dificultar o treinamento de modelos de aprendizado de máquina.

Este capítulo de trabalhos relacionados apresentará uma revisão da literatura existente sobre técnicas de extração de quadros significativos em volumes de TC, com um enfoque especial nos estudos que realizam a seleção de quadros de volumes de tomografia para servirem de insumos para modelos de aprendizado profundo. Serão apresentados os trabalhos mais relevantes na área, assim como as técnicas e métodos utilizados em cada um. Ademais, serão discutidas as limitações das abordagens existentes e as possíveis direções para futuras pesquisas.

## 3.2 Técnicas para Amostragem em TC

Amostragem é o processo de selecionar uma subpopulação de um conjunto de dados para análise. Em volumes de TC, pode ser utilizada para reduzir o tempo de processamento e a quantidade de dados a serem analisados. Essas técnicas de amostragem desempenham um papel crucial na geração de imagens de TC quando aplicadas em algoritmos de aprendizado profundo. A seleção de quadros, que envolve a escolha de um subconjunto de quadros do volume de TC para representar o conjunto completo. Esse subconjunto pode ser selecionado de várias maneiras, como escolher quadros equidistantes (Zunair et al., 2019), quadros que passam pelo centro do volume (Gao et al., 2017) ou pela realização de uma compressão ou ampliação do volume por meio de interpolações para normalizar e comprimir o volume desejável.

### 3.2.1 Subset Slice Selection (SSS)

*Subset Slice Selection* (SSS) é um método no qual o volume é dividido em três partes de tamanhos aproximadamente iguais e a quantidade desejada de quadros é selecionada do

início de cada parte (Zunair et al., 2020). Especificamente, o método amostra quadros das posições inicial, meio e final do volume conforme estabelecido no Algoritmo 1.

Embora o método proposto seja simples e fácil de implementar, ele pode levar a perdas semânticas significativas devido aos saltos temporais entre os quadros amostrados. Como resultado, o volume gerado pode não representar adequadamente o volume original. A Figura 3.1 ilustra o resultado da aplicação do SSS em um volume de TC. Na figura, é possível observar a amostragem de 12 quadros, sendo 4 do início, 4 do meio e 4 do final do volume.

Embora o método proposto possa ser útil para algumas aplicações, como a classificação de volumes de TC, ele não é indicado para aplicações de reconstrução, onde a precisão e a integridade do volume são críticas.

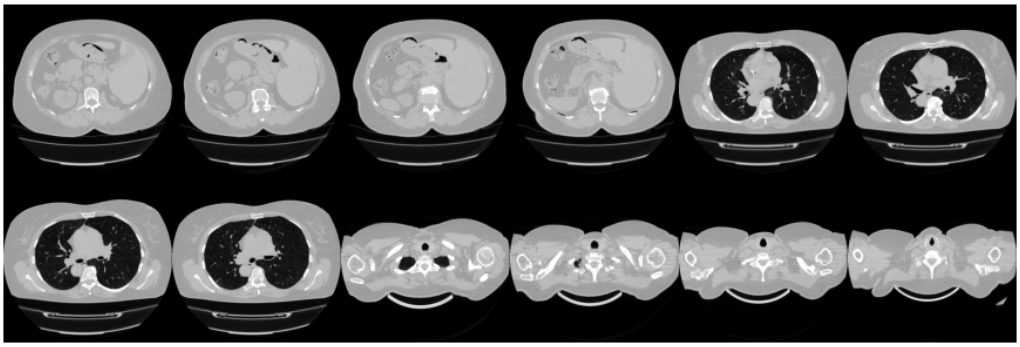

Figura 3.1: Aplicação do método *SSS*.
Fonte: (Morozov et al., 2020) (Adaptado).

**Algoritmo 1:** Método SSS

**Requisito:** Um volume 3D $V_0$ de tamanho W(*largura*) * H(*altura*) * D(*profundidade*)
**Assegura:** $V_f$ é um tensor de ranque 3

1: Defina a profundidade alvo constante de tamanho K múltiplo de 3.
2: Calcule os índices do volume particionado $i_{inicio} = 0$, $i_{meio} = \lfloor \frac{D}{2} \rfloor - \lfloor \frac{K}{2} \rfloor$, $i_{fim} = D - K$, respectivamente, início, meio e fim.
3: Inicialize o volume de saída $V_f$ com dimensão W * H * K
4: $counter = 0$
5: **para** cada índice $i_{parte}$ em $i_{inicio}, i_{meio}, i_{fim}$ **faça**
6:  **para** $i = 0$ Até $K/3$ **faça**
7:  Copie o quadro $V_0[:, :, i_{parte} + i]$ para o volume de saída $V_f[:, :, counter]$
8:  $counter++$
9:  **fim-para**
10: **fim-para**
11: Retorne Volume processado de saída $V_f$ de dimensão W * H * K

### 3.2.2 Even Slice Selection (ESS)

Técnica comumente usada para processamento de vídeos na qual é calculado um "fator de espaçamento" para que os quadros sejam equidistantes (Zhu et al., 2017; Fernando et al., 2017). Conforme o Algoritmo 2, tendo-se como $K$ o número de quadros desejados e $D$ o número de quadros total do volume, é calculado um fator de espaçamento $step = \frac{D}{K}$. Com base nesses parâmetros, seleciona-se uma sequência de $K$ quadros amostrados a cada *step* quadros.

A Figura 3.2 mostra um exemplo de uso da técnica de equidistância de quadros em um volume de TC. Observe que são selecionados quadros que mostram diferentes partes do pulmão. Diferentemente do SSS, essa técnica atenua as perdas semânticas dos dados temporais quando o fator de espaçamento é pequeno o suficiente. Caso contrário, é possível que informações importantes sejam perdidas, o que pode afetar negativamente o desempenho do modelo. Além disso, a amostragem de quadros igualmente espaçados pressupõe que todas as partes do vídeo são igualmente importantes, o que pode não ser verdade em muitos casos.

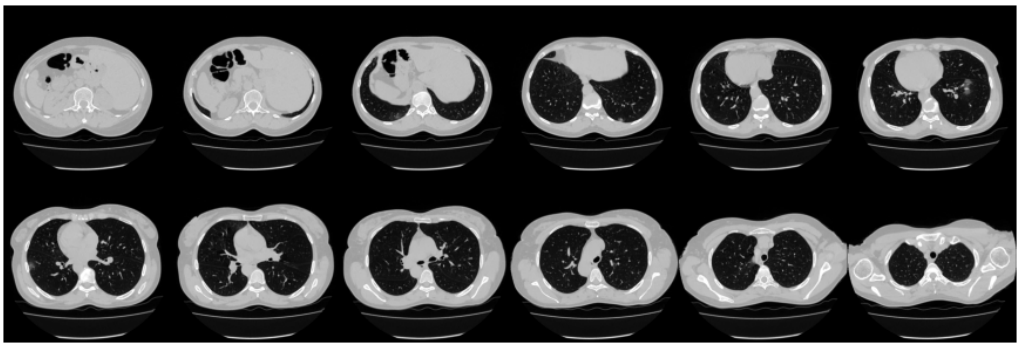

Figura 3.2: Aplicação do método *ESS*.
Fonte: (Morozov et al., 2020) (Adaptado).

**Algoritmo 2:** Método ESS

**Requisito:** Um volume 3D $V_0$ de tamanho W(*largura*) * H(*altura*) * D(*profundidade*)
**Assegura:** $V_f$ é um tensor de ranque 3

1: Defina a profundidade alvo constante de tamanho K.
2: Calcule o fator de espaçamento $D_F = \frac{D}{K}$
3: Inicialize o volume de saída $V_f$ com dimensão W * H * K
4: **para** $i = 0$ Até $K - 1$ **faça**
5: Calcule o índice do quadro de entrada $idx = \lfloor i \times D_F \rfloor$
6: Copie o quadro $V_0[:, :, idx]$ para o volume de saída $V_f[:, :, i]$
7: **fim-para**
8: Retorne Volume processado de saída $V_f$ de dimensão W * H * K

### 3.2.3 Spline Interpolated Zoom (SIZ)

As técnicas convencionais podem descartar regiões importantes dos dados volumétricos. A técnica de *Spline Interpolated Zoom* (SIZ) aborda esse problema ao ajustar o número de quadros do volume por meio da interpolação por splines cúbicas (Morozov et al., 2020), adequando-se a aplicações de aprendizado profundo.

O fator de espaçamento, $D_F$, é calculado como

$$D_F = \frac{D}{K}, \tag{3.1}$$

no qual $D$ é o número de quadros no volume original e $K$ é o número desejado de quadros no volume interpolado.

A interpolação por splines cúbicas utiliza funções polinomiais de grau 3, definidas por partes em cada intervalo entre os quadros originais. Cada função, $S_i(x)$, é definida no intervalo $[x_{i-1}, x_i]$ como

$$S_i(x) = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3, \tag{3.2}$$

sendo que $x$ representa a posição ao longo do eixo z do volume.

Os coeficientes $a_i, b_i, c_i$ e $d_i$ são calculados a partir dos valores de pixel nos quadros originais e do fator de espaçamento $D_F$. Eles garantem a continuidade e suavidade das funções em todo o intervalo. O coeficiente $a_i$ é obtido a partir dos valores de pixel dos quadros originais:

$$a_i = f(x_i) \tag{3.3}$$

Para encontrar os outros coeficientes, resolve-se um sistema de equações lineares formado por condições de continuidade da primeira e segunda derivadas das funções spline cúbica nos pontos $x_i$ e condições de contorno.

O SSZ tem-se mostrado eficaz em diversas aplicações de análise de volumes de TC, como na classificação de nódulos pulmonares (Gordaliza et al., 2019) e na detecção de lesões hepáticas (Kazlouski, 2019). Além disso, estudos recentes mostraram que essa técnica pode melhorar a precisão e a robustez de redes neurais treinadas com volumes de TC (Li et al., 2020; Yang et al., 2021).

Uma das principais desvantagens da técnica é que ela pode resultar em uma perda de resolução espacial. Isso acontece porque a interpolação por splines pode levar a uma suavização excessiva da imagem, o que pode afetar a precisão e a acurácia da análise (Gordaliza et al., 2019). Em alguns casos, a ampliação ou redução pode distorcer a estrutura do volume, resultando em uma representação imprecisa das informações. Esses proble-

mas podem afetar a interpretação dos dados e a avaliação por modelos de aprendizado de máquina.

A Figura 3.2 mostra o uso da técnica no mesmo volume, considerado o estado da arte e estando inclusive presente nas bibliotecas de processamento de arquivo *NII* (como *nibabel*).

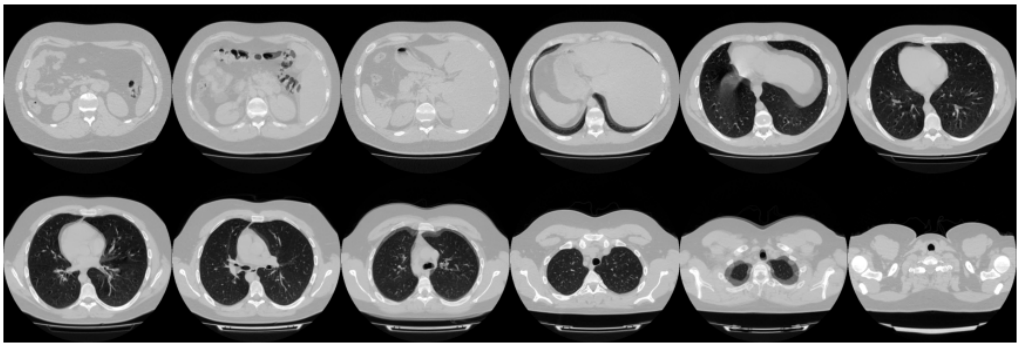

Figura 3.3: Aplicação do método *SIZ*.
Fonte: (Morozov et al., 2020) (Adaptado).

**Algoritmo 3:** Método SIZ

**Requisito:** Um volume 3D $V_0$ de tamanho W(*largura*) * H(*altura*) * D(*profundidade*)

**Assegura:** $V_f$ é um tensor de ranque 3

1: Defina a profundidade alvo constante de tamanho K.
2: Calcule o fator de espaçamento $D_F = \frac{1}{\frac{D}{K}}$
3: Aplique a interpolação por splines cúbicas para cada par de quadros adjacentes em $V_0$ considerando o fator de espaçamento $D_F$:
4: **para** $i = 1$ até $D - 1$ **faça**
5: Determine os coeficientes $a_i, b_i, c_i, d_i$ para a função $S_i(x)$ com base nos valores de pixel dos quadros originais e do fator de espaçamento $D_F$
6: Calcule os valores de pixel interpolados para as posições intermediárias entre os quadros $i$ e $i + 1$ usando a função $S_i(x)$
7: **fim-para**
8: Combine os quadros interpolados e os quadros originais para formar o volume processado de saída $V_f$ de dimensão W * H * K
9: Retorne $V_f$

## 3.3 Técnicas para redução de dimensionalidade

Os volumes de tomografia computadorizada contêm uma enorme quantidade de informações representadas por milhares ou milhões de voxels. Esses dados são de grande utilidade para auxiliar no prognóstico de pacientes. No entanto, a análise desses volumes pode ser desafiadora devido à sua complexidade. Para lidar com essa questão, técnicas de redução

de dimensionalidade podem ser empregadas. Essas técnicas têm o propósito de simplificar os dados, diminuindo sua complexidade e tornando a análise mais eficiente.

Existem diversas técnicas de redução de dimensionalidade que podem ser aplicadas em volumes de TC. Algumas dessas técnicas são descritas a seguir.

### 3.3.1 Análise de Componentes Principais

A Análise de Componentes Principais *(Principal Component Analysis - PCA)* é uma técnica de análise de dados que visa reduzir a dimensão dos dados, mantendo o máximo de informação possível. Ela é frequentemente usada em aplicações de aprendizado de máquina para selecionar as variáveis mais importantes em um conjunto de dados.

A ideia por trás da PCA é encontrar um conjunto de novas variáveis que são combinações lineares das variáveis originais. Essas novas variáveis são chamadas componentes principais e são ordenadas por importância, de modo que a primeira componente principal captura a maior variância nos dados, a segunda componente principal captura a segunda maior variância remanescente e assim por diante (Jolliffe e Cadima, 2016).

Para calcular as componentes principais, é necessário realizar uma transformação linear nos dados expressa pela equação

$$\mathbf{z} = \mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu}),$$

na qual $\mathbf{x}$ é um vetor de dados, $\boldsymbol{\mu}$ é o vetor de médias das variáveis e $\mathbf{W}$ é uma matriz de pesos, também chamada de matriz de carga. A matriz de carga é calculada a partir da matriz de covariância dos dados e contém as direções principais de variância nos dados.

As componentes principais encontram-se na matriz de carga, sendo que a primeira componente principal é dada pela primeira coluna da matriz, a segunda componente principal é dada pela segunda coluna e assim por diante. Cada componente principal é uma combinação linear das variáveis originais, de modo que é possível interpretá-las como novas variáveis que capturam a variação mais importante nos dados.

Attia et al. (2015) propõem o uso de PCA em conjunto com redes neurais para classificar imagens de ultrassom renal. O objetivo do estudo era desenvolver um método de classificação de imagens de ultrassom renal que pudesse ajudar na detecção precoce de doenças renais. Os autores utilizaram um conjunto de dados de imagens de ultrassom renal e aplicaram a PCA para reduzir a dimensionalidade das imagens e extrair as características mais relevantes. Em seguida, os dados reduzidos foram alimentados em uma rede neural para treinamento e classificação. Os resultados mostraram que a combinação de PCA e redes neurais pode ser eficaz na classificação de imagens de ultrassom renal com uma acurácia média de 97 % contra 95 % sem o uso da técnica.

O trabalho de (Nandi et al., 2015) apresenta um estudo sobre o uso da PCA em processamento de imagens médicas. Os autores discutem o papel da PCA na redução da dimensionalidade de dados de imagem, permitindo uma representação mais compacta e eficiente da informação. Através da redução de dimensão, é possível identificar as características mais importantes da imagem que contribuem para sua variância. O estudo apresenta exemplos de aplicação da PCA em imagens médicas, incluindo imagens de ressonância magnética, tomografia computadorizada e ultrassom. Em cada caso, a PCA é usada para diminuir a dimensionalidade dos dados de imagem, destacar as características mais relevantes e melhorar a qualidade da imagem.

Enquanto os resultados indicam que a PCA pode ser uma ferramenta útil em execução de imagens médicas, permitindo uma representação mais compacta e eficiente da informação, também fica evidente limitações quando aplicado a tomografia computadorizada:

- Perda de informações: Embora a técnica possa reduzir a dimensionalidade dos dados de imagem, ela também pode levar a uma perda de informações. Em alguns casos, essa perda pode ser significativa e resultar em uma representação menos precisa da imagem.
- Sensibilidade ao ruído: O método pode ser sensível ao ruído nos dados de imagem. Em imagens de TC, o ruído pode ser introduzido de várias fontes, incluindo frutos movimentos, ruído eletrônico e outros artefatos de aquisição. Se o ruído não for adequadamente tratado, pode afetar a precisão da análise PCA.
- Necessidade de pré-processamento: Pode exigir pré-processamento dos dados de TC para garantir que eles estejam em uma escala e orientação adequadas para análise. Essa tarefa pode requerer um tempo adicional e habilidades técnicas para realizar.

Em resumo, a técnica PCA pode ser útil para análise e processamento de imagens de TC, mas deve ser aplicada com cuidado, levando em consideração as limitações e desvantagens acima mencionadas.

### 3.3.2 Análise de Componentes Independentes

Análise de Componentes Independentes (ICA) é uma técnica estatística usada para separar um conjunto de sinais mistos em seus componentes independentes originais. Trata-se de uma técnica não-paramétrica que não requer conhecimento prévio sobre a distribuição dos sinais originais ou dos ruídos.

A ICA é baseada na ideia de que um conjunto de sinais mistos pode ser representado como uma combinação linear dos sinais independentes originais (Oja e Hyvarinen, 2000). Ainda, busca encontrar uma matriz de mistura que, quando aplicada aos sinais independentes, produza os sinais mistos observados. Essa matriz de mistura pode ser representada como:

$$\mathbf{X} = \mathbf{AS}$$

onde $\mathbf{X}$ é uma matriz de dados contendo os sinais mistos, $\mathbf{A}$ é uma matriz de mistura que combina os sinais independentes, e $\mathbf{S}$ é uma matriz contendo os sinais independentes originais.

O objetivo da ICA é encontrar a matriz de separação $\mathbf{W}$ que permite recuperar os sinais independentes originais a partir dos sinais mistos observados. A matriz de separação é dada por:

$$\mathbf{W} = \mathbf{A}^{-1}$$

Uma vez encontrada a matriz de separação, é possível recuperar os sinais independentes originais a partir dos sinais mistos observados, aplicando a matriz de separação aos sinais mistos:

$$\mathbf{S} = \mathbf{WX}$$

Os sinais independentes originais são ordenados por ordem de importância, de modo que o primeiro sinal independente captura a maior parte da variância nos dados, e os subsequentes capturam a variação restante, em ordem decrescente de importância.

Du et al. (2021) apresentam o uso da ICA em imagens de ressonância magnética funcional (fMRI) para o diagnóstico de alterações nas áreas funcionais do cérebro em pacientes com infarto cerebral. Os autores propõem que a ICA pode ser utilizada para identificar áreas funcionais do cérebro que estão relacionadas com a patologia da lesão do infarto cerebral. Em particular, a ICA permite a identificação de componentes independentes que representam padrões de atividade cerebral associados a diferentes áreas funcionais. Aplicaram-na em dados de fMRI de pacientes com infarto cerebral e compararam as áreas funcionais identificadas pela ICA com as áreas afetadas pela lesão da patologia. Os resultados mostraram que a ICA foi capaz de identificar áreas funcionais do cérebro que estavam relacionadas com a patologia da lesão do infarto cerebral.

Calhoun e Adali (2012) usaram a ICA para analisar imagens de ressonância magnética funcional (fMRI) de vários sujeitos. A fMRI é uma técnica que produz imagens tridimensionais do cérebro e é amplamente usada para estudar a atividade cerebral. Os pesquisadores empregaram a ICA para descompor os sinais fMRI em componentes independentes. Esses componentes independentes representam redes cerebrais distintas que mostram ativação sincronizada ao longo do tempo. Ao usar a ICA, os pesquisadores conseguiram identificar essas redes cerebrais sem a necessidade de um modelo pré-definido, o que é uma grande vantagem dessa técnica.

Além disso, a ICA também pode ser usada para remover artefatos e ruídos em volumes de TC, tais como artefatos de metal, sinais de respiração, efeitos de iluminação desigual, entre outros. Isso é possível porque a ICA é capaz de separar sinais independentes e identificar aqueles que não são relevantes para a análise. Para isso, o volume de TC é decomposto em componentes independentes, e os componentes que correspondem a artefatos e ruídos são removidos. O volume resultante é então reconstruído apenas com os componentes independentes relevantes.

Embora seu uso seja plenamente factível com dados de tomografia, existem algumas limitações que dificultam a aplicação da técnica com este tipo de dado:

- Sensibilidade a ruídos: A TC pode gerar ruídos que podem afetar a precisão da ICA. Esses ruídos podem ser gerados por uma variedade de fontes, incluindo movimentos do paciente, equipamentos com mau funcionamento e outros fatores externos. Essa presença pode resultar em componentes que não são significativos para a análise, reduzindo a eficácia da ICA.
- Dependência do pré-processamento: A precisão da ICA depende muito do pré-processamento dos dados de TC. A remoção inadequada do ruído de fundo ou a normalização inadequada dos dados pode resultar em uma análise imprecisa.
- Limitações na identificação de estruturas anatômicas: A ICA pode ser útil na separação de sinais em componentes independentes, mas pode não ser capaz de identificar com precisão estruturas anatômicas específicas nos dados de TC. Isso pode limitar a utilidade da ICA em análises que requerem a identificação precisa dessas estruturas (Calhoun et al., 2013).

Em resumo, a ICA é uma técnica poderosa para a separação de sinais complexos em componentes independentes, podendo fornecer informações valiosas para a análise de dados de imagens médicas, desde que sejam consideradas suas limitações e desvantagens em cada caso específico.

### 3.3.3 t-Distributed Stochastic Neighbor Embedding

O *t-Distributed Stochastic Neighbor Embedding* (t-SNE) é um algoritmo de redução de dimensionalidade não linear frequentemente utilizado em visualização de dados. É especialmente útil para a visualização de dados complexos, nos quais as relações entre os pontos de dados não são lineares (Van der Maaten e Hinton, 2008).

O t-SNE funciona mapeando cada ponto de dados de alta dimensionalidade em um ponto de baixa dimensionalidade, ao mesmo tempo em que preserva as distâncias entre os pontos. Isso é feito através de uma abordagem probabilística, em que o algoritmo calcula a semelhança entre cada par de pontos de dados de alta dimensão e constrói uma distribuição de probabilidade a partir dessa semelhança.

Posteriormente, o algoritmo cria uma distribuição de probabilidade para os pares de pontos de baixa dimensionalidade, de modo que a probabilidade de dois pontos serem vizinhos no espaço de baixa dimensão seja proporcional à sua similaridade na matriz de afinidade de alta dimensão.

O objetivo do algoritmo é minimizar a divergência entre as duas distribuições de probabilidade, preservando assim a distribuição de probabilidade sobre os pares de pontos de alta dimensão na distribuição de probabilidade sobre os pares de pontos de baixa dimensão. A divergência é medida pela distância de Kullback-Leibler (KL) entre as duas distribuições de probabilidade.

O t-SNE utiliza uma distribuição de probabilidade t-student para medir a similaridade entre os pontos de baixa dimensionalidade. O algoritmo emprega uma abordagem iterativa para ajustar a distribuição de probabilidade em cada iteração, com o objetivo de minimizar a divergência KL entre as distribuições de probabilidade.

Liu et al. (2021) empregaram t-SNE para visualizar as características extraídas dos volumes de TC de pacientes com Covid-19. A partir dos volumes de TC, foram extraídas características relevantes por meio de uma rede neural convolucional e, em seguida, a técnica t-SNE foi aplicada para reduzir a dimensionalidade dos dados e permitir a visualização das informações em duas dimensões. Os resultados demonstraram que a técnica t-SNE foi capaz de agrupar as imagens de acordo com a similaridade das características extraídas, o que pode auxiliar os médicos no processo de diagnóstico da Covid-19. Além disso, a utilização da técnica t-SNE permitiu visualizar como o modelo de aprendizado profundo (COVIDNet) fazia suas decisões de classificação e ofereceu *insights* valiosos para a tomada de decisão clínica.

Outra possível aplicação é a análise e validação das características extraídas. Em (Cheng et al., 2022) a técnica t-SNE foi utilizada para a visualização de características extraídas de imagens de ultrassonografia, para a classificação de lesões em quatro grupos de acordo com sua similaridade permitindo a classificação de lesões com tamanhos variados e, posteriormente, a análise de características extraídas por t-SNE para cada grupo de lesões.

Embora a técnica t-SNE seja amplamente utilizada na visualização de dados de alta dimensionalidade, ela pode apresentar algumas limitações e desvantagens quando utilizada com dados de tomografia computadorizada. Principalmente quando possuem alta dimensionalidade e podem ser difíceis de serem reduzidos sem perda de informações importantes, a maioria dos trabalhos que empregam a técnica a utilizam a fim de analisar e obter melhor compreensão visual das características obtidas por modelos de aprendizado profundo com a síntese das informações obtidas.

## 3.4 Técnicas de extração de características significativas com Aprendizado profundo

Com o avanço do aprendizado profundo, técnicas têm sido desenvolvidas para selecionar automaticamente os dados mais relevantes, permitindo uma representação mais compacta e de fácil interpretação. Nesse contexto, duas abordagens distintas são utilizadas na literatura com dados de tomografia computadorizada: o uso de autocodificadores e a seleção por meio de modelos de detecção e segmentação.

Os autocodificadores são uma técnica de aprendizado de máquina que pode ser utilizada para extrair características relevantes dos dados de imagem, permitindo uma representação mais compacta e de fácil interpretação. Já a seleção por modelos de detecção e segmentação utiliza algoritmos de aprendizado de máquina para identificar e isolar regiões de interesse em volumes de tomografia, permitindo a seleção das imagens mais informativas para uma determinada análise ou aplicação.

Ambas as abordagens têm suas vantagens e limitações, e seu uso depende do contexto e do conjunto de dados em questão. O objetivo deste trabalho é explorar as técnicas de seleção de quadros com Aprendizado de Profundo, investigando sua eficácia em diferentes aplicações de processamento de imagens médicas e examinando as principais questões e desafios que surgem ao utilizá-las.

### 3.4.1 Autocodificadores

Autocodificadores apresentam uma abordagem efetiva para a extração de características relevantes dos dados de imagem, possibilitando a criação de uma representação compacta e interpretável, traduzida em um espaço de menor dimensionalidade conhecido como espaço latente.

Volumes de tomografia computadorizada requerem pré-processamento para normalização e segmentação antes de servirem como dados de treinamento para o autocodificador. A normalização é crucial para uniformizar a escala dos dados e eliminar possíveis ruídos e artefatos presentes na imagem. A segmentação, por sua vez, visa isolar as estruturas de interesse na imagem, descartando fundos desnecessários.

O autocodificador, quando treinado com os volumes de tomografia computadorizada normalizados e segmentados, tem como entrada o volume de tomografia computadorizada e como saída a reconstrução do mesmo volume, agora porém com uma dimensionalidade reduzida. Essa redução ocorre no espaço latente, uma representação intermediária que compacta a informação original em uma estrutura de menor dimensão. Durante o treinamento, o autocodificador ajusta os pesos de suas camadas ocultas para minimizar o erro de reconstrução entre a entrada original e a saída reconstruída.

Após o treinamento, o espaço latente - que carrega a representação condensada dos volumes - pode ser aplicado em várias tarefas, como visualização, análise e classificação dos volumes de tomografia computadorizada. Por exemplo, esta representação pode ser utilizada como entrada para outra rede neural encarregada da classificação de doenças (Jia et al., 2015; Silva et al., 2020). Adicionalmente, o espaço latente pode ser visualizado tridimensionalmente para identificar padrões e relações entre diferentes volumes de tomografia computadorizada.

Embora os autocodificadores sejam uma técnica eficaz para a redução de dimensionalidade de volumes de tomografia computadorizada, existem algumas limitações e desvantagens que devem ser consideradas:

- Complexidade de treinamento: O treinamento de autocodificadores pode ser computacionalmente intensivo e exigir grandes quantidades de dados para atingir um desempenho satisfatório. Além disso, a escolha do tamanho e número de camadas ocultas pode afetar significativamente o desempenho do autocodificador.
- Problemas de generalização: Autocodificadores podem apresentar problemas de generalização, ou seja, podem ter dificuldades em reconstruir corretamente volumes de tomografia computadorizada que não foram vistos durante o treinamento. Isso pode ocorrer devido à falta de diversidade dos dados de treinamento ou à complexidade dos dados de teste.
- Sensibilidade a ruído e artefatos: Podem ser sensíveis a ruídos e artefatos em volumes de tomografia computadorizada, o que pode afetar negativamente o desempenho do modelo. É importante que os dados de entrada sejam pré-processados e limpos antes do treinamento do autocodificador.
- Limitações na interpretação: Embora a representação latente obtida com autocodificadores possa ser útil para a análise e classificação de volumes de tomografia computadorizada, a interpretação dos resultados pode ser desafiadora. Isso ocorre porque a representação latente pode não ter uma interpretação clara ou significado biológico.
- Complexidade do modelo: Autocodificadores podem ser modelos complexos e difíceis de interpretar. O tamanho da representação latente pode ser difícil de interpretar e pode exigir uma análise cuidadosa para identificar as características relevantes.

Em resumo, embora os autocodificadores sejam uma técnica poderosa para a redução de dimensionalidade de volumes de tomografia computadorizada, é importante considerar as limitações e desvantagens da técnica ao aplicá-la em problemas de imagem médica.

### 3.4.2 Seleção de quadros por modelos de detecção e segmentação

A seleção de quadros em volumes de tomografia pode ser realizada por meio de modelos de detecção e segmentação. Esses modelos utilizam algoritmos de aprendizado de máquina

para identificar e isolar as regiões de interesse nos volumes de tomografia, o que permite a seleção das imagens mais representativas e informativas para uma determinada análise ou aplicação. Também consiste em métodos nos quais um algoritmo primeiro varre cada fatia e determina se um objeto de interesse está presente.

Para a detecção de regiões de interesse, os modelos utilizam técnicas de detecção de objetos, que buscam identificar regiões com características específicas presentes nas imagens. Essas técnicas podem ser baseadas em métodos de CNNs ou de aprendizado profundo, que são capazes de extrair automaticamente as regiões de interesse. Por exemplo, um modelo de IA pode ser empregado em tomografias computadorizadas. Em seguida, selecionam-se as quadros que apresentam o órgão desejado e descartam-se os demais a exemplo do trabalho de (Cheng et al., 2022) que utiliza YoLo para esse fim.

Já para a segmentação das regiões de interesse, os modelos utilizam técnicas de segmentação de imagens que buscam dividir as imagens em regiões com características semelhantes. Podem ser baseadas em métodos de CNNs, como a segmentação por máscaras (Mask R-CNN), que utiliza uma rede neural para gerar máscaras para detectar quais quadros exibem segmentos de órgãos como pulmões, cérebro ou qualquer outro órgão alvo (da Cruz et al., 2020; Ray et al., 2018).

Embora essa abordagem se adapte aos dados, ela não leva em consideração a relação entre as quadros, ou seja, não tenta usar informações temporais ao selecionar as melhores quadros. Muitas vezes, descarta-se os subconjuntos iniciais e finais de quadros, uma vez que a maior parte do órgão alvo está tipicamente no “meio” do volume.

### 3.4.3 Uso de Convoluções em profundidade para aprender características espaço-temporais

Modelos que empregam convoluções em profundidade costumam apresentar baixo custo computacional e muitas vezes seu desempenho é comparável a modelos de maior custo. Além disso, o ganho relacionado à interpretabilidade do modelo tem grande potencial de uso com técnicas para interpretação da inferência de modelos de aprendizado profundo. Essa técnica também foi utilizada com sucesso em (Valadão et al., 2023) ao mostrar que informações temporais podem trazer grande ganho nas informações aprendidas por modelos de aprendizado profundo.

O artigo inspirado por este trabalho, propõe em aproveitar as características temporais dos dados e discute um *framework* de treinamento chamado *Spatio-Temporal Backpropagation* (STBP) para redes neurais de impulsos (SNNs) que considera simultaneamente os domínios espaciais e temporais. O *framework* é avaliado em conjuntos de dados estáticos e dinâmicos usando arquiteturas totalmente conectadas e convolucionais, e os resultados mostram que ele supera os algoritmos existentes de última geração em redes de impulsos ao aproveitar de características espaço-temporais em dois momentos, primeiro aproveitando-se de características espaciais e depois após a convolução da dimensão temporal.

## 3.5 Classificação com volumes de tomografia

Atualmente, existem duas abordagens comuns para a classificação de volumes de TC: a primeira considera cada quadro do volume como um dado individualmente único, enquanto a segunda utiliza o volume inteiro para consumo dos modelos. O uso de arquiteturas de redes neurais convolucionais (CNN) 2D, em que os quadros são tratados como dados independentes, é uma das formas mais populares para avaliar volumes de TC (Gao et al., 2017; Gentili, 2019; Hamadi et al., 2019; Kavitha et al., 2019; Ronneberger et al., 2015; Tabarcea et al., 2019; Grewal et al., 2018). No entanto, estudos recentes indicam que o uso de dados volumétricos produz resultados ainda melhores ao preservar as características de profundidade dos dados.

Os autores de Gu et al. (2018) propuseram uma CNN 3D profunda com estratégias de predição multiescala para a detecção de nódulos pulmonares a partir de imagens segmentadas. A CNN 3D teve um desempenho muito melhor com recursos mais avançados do que a CNN 2D. Os autores de Nasrullah et al. (2019) mostraram que o uso de uma rede neural convolucional 3D para a classificação de nódulos pulmonares em TC produziu resultados significativamente melhores que o uso de uma CNN 2D, alcançando uma precisão de 94.17% em comparação com 87,4% da CNN 2D. Da mesma forma, Zhu et al. (2018) propuseram uma abordagem de classificação de TC baseada em uma rede neural convolucional 3D, que considera o volume inteiro como entrada para diagnóstico de câncer com a base de dados LIDC-IDRI (Ozdemir et al., 2019). Eles mostraram que a abordagem 3D superou a abordagem 2D na detecção de nódulos pulmonares, alcançando uma acurácia de 90.44% em comparação com 86.7% da abordagem 2D.

Embora esses estudos tenham mostrado resultados promissores, a maioria foi focado em tarefas específicas de classificação de lesões ou nódulos em TC. Poucos estudos têm explorado a classificação de volumes de TC em categorias mais amplas, como saudável ou com doença. Nesse sentido, este trabalho tem como objetivo contribuir para a literatura explorando a classificação de volumes de TC em categorias mais amplas com o uso de técnicas de aprendizado profundo.

## 3.6 Considerações Finais

Esta seção examinou técnicas variadas relacionadas à seleção de quadros e à classificação de volumes de TC através de redes neurais convolucionais 2D e 3D. Cada abordagem possui suas peculiaridades, vantagens e desvantagens, requisitando uma análise cautelosa na hora de decidir a técnica mais adequada a uma específica aplicação em processamento de imagem médica.

A tomografia computadorizada oferece uma vasta gama de informações valiosas para avaliação clínica. Contudo, a literatura sobre a utilização de volumes de TC para classificação em aprendizado de máquina ainda é relativamente limitada, principalmente no que tange

a estratégias eficazes para a geração de instâncias a partir de dados brutos. Assim, o tratamento e a redução dos dados de entrada tornam-se essenciais para o manuseio de bases de dados de TC.

Também foram exploradas técnicas de redução de dimensionalidade e seleção de características com base em aprendizado de máquina, que podem auxiliar a análise de volumes de dados complexos como os de TC, apesar das suas limitações e desvantagens.

Quanto à classificação de volumes de TC, embora a abordagem com redes neurais convolucionais 2D seja prevalente, estudos recentes sugerem que redes neurais convolucionais 3D podem oferecer resultados superiores, preservando as características de profundidade dos dados.

Em resumo, a seleção criteriosa de técnicas de amostragem, redução de dimensionalidade e seleção de características é essencial em algoritmos de aprendizado profundo aplicados à imagem de TC. Estas escolhas são diretamente influenciadas pelos objetivos do estudo e pelas características dos dados. A combinação ideal destas técnicas pode agilizar e aprimorar a análise de volumes de TC, um aspecto crucial na prática clínica.

Este panorama do estado da arte das técnicas de seleção e amostragem de quadros em volumes de tomografia computadorizada destaca a necessidade de mais pesquisas e melhorias nestas técnicas, abrindo caminho para avanços significativos no campo da imagem médica.

Capítulo 4

# Abordagem Proposta

## 4.1 Considerações Iniciais

A tomografia computadorizada se destaca pela abundância de informações obtidas, as quais excedem em muito as provenientes de um raio-X convencional. Conforme discutido na Seção 2.4, a alta resolução espacial dessas imagens tridimensionais facilita a localização de anomalias, mesmo as menores, bem como áreas impactadas por patologias, contribuindo para um prognóstico mais preciso do paciente (Ruppert et al., 2020).

Para preservar essa riqueza de dados oriunda da alta resolução, uma seleção criteriosa de quadros relevantes em volumes de tomografia computadorizada pode ser utilizada para redução do custo computacional. Métodos convencionais de amostragem de quadros, entretanto, seguem rotinas específicas e não empregam algoritmos capazes de se adaptar aos dados para a seleção de quadros. Esse procedimento pode resultar na inclusão de informações irrelevantes no volume final. Adicionalmente, as técnicas atuais que fazem uso de aprendizado profundo podem sofrer de baixa interpretabilidade, alta sensibilidade a ruídos ou ainda descartar informações volumétricas relevantes, tornando a escolha de dados por quadro do volume uma tarefa complexa.

Diante deste cenário, é proposta uma abordagem fundamentada no Grad-CAM. Como visto na Seção 2.3.5, o Grad-CAM é uma técnica de interpretação para redes neurais convolucionais que possibilita a identificação das partes da entrada que são mais relevantes para alguma tarefa, como identificação de doenças em exames de CT.

O método proposto é constituído de uma rede convolucional projetada para captura do mapa de ativação de um volume de tomografia computadorizada, onde a saída é uma matriz de três dimensões que representa as regiões de interesse em cada quadro do volume. Isso é feito pela adaptação da técnica Grad-Cam que utiliza os gradientes obtidos ao avaliar a classe de interesse na rede classificadora, fluindo para a última camada convolucional a fim de mapear as regiões de ativação de uma imagem. Em conjunto com a técnica de

convolução em profundidade vista na Seção 2.3.2 é possível extrair as características mais relevantes do volume e calcular quais quadros foram mais importantes para o modelo de aprendizado profundo realizar a classificação.

A abordagem proposta utiliza uma CNN3D, projetada especificamente para esse fim, para analisar as imagens de tomografia e selecionar automaticamente os quadros mais relevantes, que podem fornecer informações importantes para o diagnóstico e acompanhamento de doenças. A seleção de quadros é feita a partir da avaliação do mapa de ativação na última camada convolucional da CNN3D, permitindo a identificação dos quadros que contêm as informações mais significativas. Após ampla avaliação, observamos que a técnica proposta superara as limitações dos métodos convencionais de amostragem de quadros e fornece resultados mais precisos e confiáveis.

## 4.2 Grad-Cam Slice Selection (GSS)

O método proposto, chamado *Grad-Cam Slice Selection* (GSS), utiliza Grad-Cam (*Gradient-weighted Class Activation Mapping*) para identificar as regiões do dado de entrada que são mais importantes de para a realização de uma previsão por meio de uma rede neural, como visto na Seção 2.3. Para isso, o Grad-CAM calcula o gradiente da pontuação da classe alvo em relação aos mapas de características da camada convolucional mais próxima da saída da rede. Dessa forma é possível identificar quais informações mais contribuíram para a classificação e produzir um mapa de calor que mapeia as regiões mais relevantes.

A convolução em profundidade acontece em uma camada convolucional que utiliza um único filtro para cada canal de entrada, produzindo uma saída com o mesmo número de canais da entrada. Isso significa que cada filtro é responsável apenas por detectar características em seu canal correspondente. Ao separar o processo de detecção de características dessa maneira, a convolução em profundidade é capaz de capturar características mais precisas e refinadas em cada canal. Essa técnica pode ajudar a reduzir o número de parâmetros na rede enquanto ainda captura informações espaciais, da mesma forma que produzirá mapas de características para cada canal do dado avaliado.

Após a aplicação da convolução em profundidade, os mapas de características resultantes são combinados usando a convolução pontual, ou convolução ponto a ponto (*pointwise convolution*), que aplica um filtro $1 \times 1$ a cada pixel em cada canal. O uso desse tipo de convolução garantirá que os gradientes avaliados com Grad-CAM estarão relacionados aos respectivos quadros do volume.

Essa relação quadro a quadro dos gradientes pode ser exemplificada pelo seguinte exemplo de convolução em profundidade que ocorrerá em um volume de tomografia computadorizada.

1. Seja $X$ um tensor de volume de tomografia computadorizada de entrada com dimensões $(H, W, C)$, onde $H$ e $W$ são a altura e a largura do volume e $C$ é o número de quadros do volume de entrada.
2. Seja $J$ um tensor de filtro com dimensões $(J_h, J_w, C, M)$, onde $J_h$ e $J_w$ são a altura e a largura do filtro e $M$ é o número de canais de saída desejados.
3. A convolução em profundidade de $X$ com $J$ pode ser escrita como:

$$Y(y, x, g) = \sum_{i=0}^{J_h-1} \sum_{j=0}^{J_w-1} \sum_{k=0}^{C-1} J(i, j, k, g) \cdot X(y + i, x + j, k)$$

   Onde $Y$ é o tensor de saída com dimensões $(H - J_h + 1, W - J_w + 1, M)$, e $y$, $x$ e $g$ são índices que variam de 0 a $H - J_h$, 0 a $W - J_w$ e 0 a $M - 1$, respectivamente. A operação de convolução em profundidade é aplicada para cada canal de entrada separadamente, utilizando um filtro único para cada canal.

Para combinar as características resultantes a convolução ponto a ponto será aplicada:

1. Seja $Y'$ um tensor de entrada com dimensões $(H', W', M')$.
2. Seja $P$ um tensor de filtro com dimensões $(1, 1, M', M'')$.
3. A convolução de ponto de $Y'$ com $P$ pode ser escrita como:

$$Z(i, j, m'') = \sum_{m'=0}^{M'-1} P(0, 0, m', m'') \cdot Y'(i, j, m')$$

   Onde:

   - $Z$ é o tensor de saída com dimensões $(H', W', M'')$.
   - $i$, $j$ são índices que variam de 0 a $H' - 1$ e 0 a $W' - 1$, respectivamente, representando as posições espaciais (altura e largura) do tensor de saída.
   - $m''$ é um índice que varia de 0 a $M'' - 1$, representando o canal de saída.
   - $m'$ é um índice que varia de 0 a $M' - 1$, representando o canal de entrada;.
   - $P(0, 0, m', m'')$ é o valor do filtro 1x1 aplicado entre o canal de entrada $m'$ e o canal de saída $m''$.
   - $Y'(i, j, m')$ é o valor do tensor de entrada $Y'$ na posição espacial $(i, j)$ e no canal de entrada $m'$.

Na convolução em profundidade, o número de canais de entrada $C$ é o mesmo que o número de canais de saída desejados para cada filtro. Já na convolução de ponto, o número de canais de entrada $M'$ pode ser diferente do número de canais de saída desejados $M''$. Esta etapa permite que a rede misture e combine as características detectadas em cada canal para criar uma representação mais complexa da entrada, aproveitando, dessa forma, as características temporais dos dados. A Figura 4.1 ilustra esse processo.

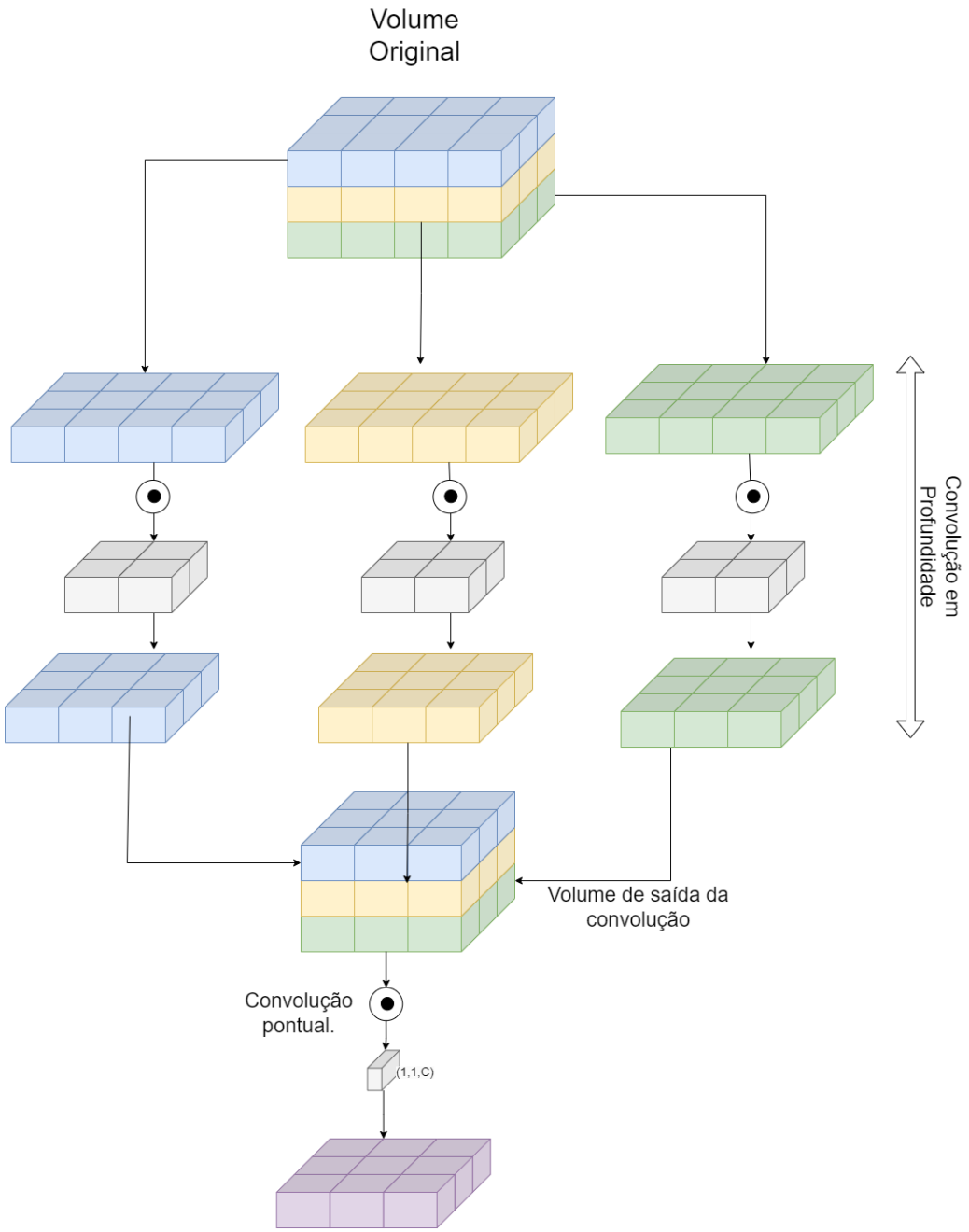


Figura 4.1: Aplicação de convolução em profundidade e pontual a um volume de tomografia computadorizada

A técnica proposta consiste na utilização da técnica Grad-CAM aplicada a uma rede que usa convolução em profundidade. Utilizando o mapa de gradiente produzido pelo Grad-CAM para avaliar os canais de saída de uma camada convolucional em profundidade. Dessa forma, é possível selecionar os canais de saída mais importantes para previsão.

Essa abordagem apresenta diversas vantagens, como a eficiência computacional, uma vez que reduz o número de parâmetros na rede, a interpretabilidade do modelo ao destacar os recursos mais importantes para a previsão e a possibilidade de melhorar o desempenho em tarefas específicas ao permitir que a rede se concentre nas informações mais relevantes da entrada.

Para sucesso da técnica na escolha de quadros mais significativos é importante que o modelo responsável em selecionar os quadros tenha como dimensões de entrada volumes com o máximo possível de quadros no eixo $z$ o que deve ser computacionalmente intensivo. Para tanto torna-se necessário a redução das dimensões espaciais do volume. Para solucionar esse problema, é possível utilizar a técnica *Spline Interpolated Zoom*, explicada na Seção 3.2.3, para redimensionar os dados de entrada. Dentre as técnicas de amostragem relacionadas, essa técnica permite o aumento ou redução do tamanho da imagem com menor perda de informação, utilizando a interpolação por splines para obter novos valores de pixel em um conjunto de pontos discretos. Dessa forma, é possível reduzir a resolução para redução de carga computacional.

No entanto, é importante ressaltar que o redimensionamento também deve ser cuidadosamente avaliado, pois se a imagem for reduzida ou aumentada em excesso, poderá causar grandes distorções ou perda de informações relevantes. É necessário avaliar cuidadosamente o tamanho final dos dados de entrada, considerando a capacidade do hardware disponível e a importância das informações perdidas na redução da resolução.

Outra questão é que as operações e camadas escolhidas deverão garantir o mapeamento dos quadros do volume, isto é, convoluções temporais deverão ser feitas somente após a camada convolucional na qual serão extraídas os gradientes para o Grad-Cam. Conforme explicado na Seção 2.3, o algoritmo de retropropagação ainda assim ajustará os pesos de toda a rede levando as informações da camada de saída, que incluem convoluções temporais. A arquitetura desenvolvida foi adaptada do trabalho de (da Silva et al., 2021) para utilizar convoluções em profundidade e camadas de *Pooling* espaciais e pode ser visualizada na Tabela 4.1. Essa arquitetura foi utilizada pelo alto desempenho em detecções de anomalia para dados temporais, sua adaptação será referenciada como 3DCNN-C.

| **Primeira Camada** |
|---|
| 1. Depthwise Conv<br>input = (largura, altura, profundidade)<br>2. SpatialMaxPooling<br>3. BatchNorm |
| Segunda Camada |
| Terceira Camada |
| .... |
| Última Camada Convolucional<br>output = (x,y, profundidade) |
| Pointwise Convolution |
| Flatten |
| Dense |
| Output (Softmax/Sigmoid) |

Tabela 4.1: Template de arquitetura 3DCNN-C

Com o modelo treinado para uma determinada tarefa de classificação, as informações mais relevantes podem ser avaliadas pelo Grad-Cam. Uma matriz será obtida representando o mapa de calor na forma $(x, y, z)$ onde x e y vão equivaler as dimensões espaciais e z equivale

a profundidade do volume, como a profundidade será mantida ao mapear cada quadro do volume, será possível avaliar a contribuição de cada quadro na classificação do volume. Esse processo pode ser melhor visualizado na Figura 4.2.

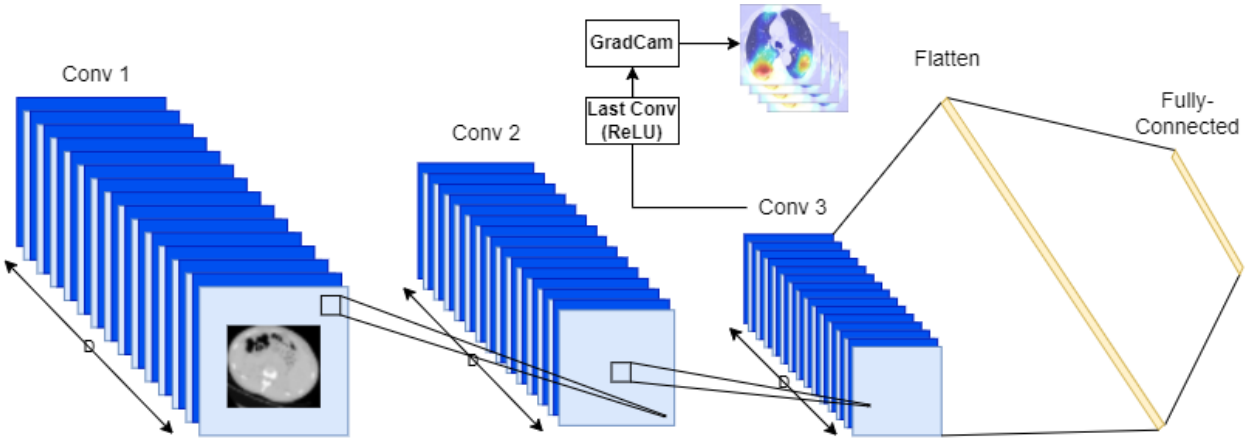


Figura 4.2: Esquema *Grad-Cam Slice Selection (GSS)*.

Com a matriz (mapa de calor) representando as macro-regiões de interesse do volume, para cada quadro $z$ é dado um *score*. Conforme a Equação 4.1 será a soma dos elementos da matriz de ativação deste quadro para, enfim, selecionar os *k* quadros mais relevantes.

$$score(z) = \sum_{i=1}^{x} \sum_{j=1}^{y} M_{i,j,z} \tag{4.1}$$

A escolha da soma como métrica agregadora é justificada por sua simplicidade e capacidade de capturar a relevância global de cada quadro para a tarefa de classificação. Ao somar os valores do mapa de calor, está-se de fato agregando a informação dos pontos de ativação relevantes em todo o quadro, proporcionando um panorama completo de seu impacto na decisão do modelo. Um *score* elevado indica que o quadro em questão apresenta um grande número de características relevantes, conforme identificado pelo mapa de calor gerado pelo Grad-CAM. Portanto, pode-se inferir que quadros com scores mais altos contribuíram significativamente para a classificação realizada pela rede.

A Figura 4.3 mostra uma representação visual de exemplo do mapa de calor, juntamente com a pontuação de cada quadro. Uma observação importante a ser destacada é que a representação visual não é necessária para este cálculo, utiliza-se somente a matriz obtida pelo *Grad-Cam* do quadro para computação do *score* e posterior ordenação para obter-se os *k* quadros mais significativos conforme o Algoritmo 4.

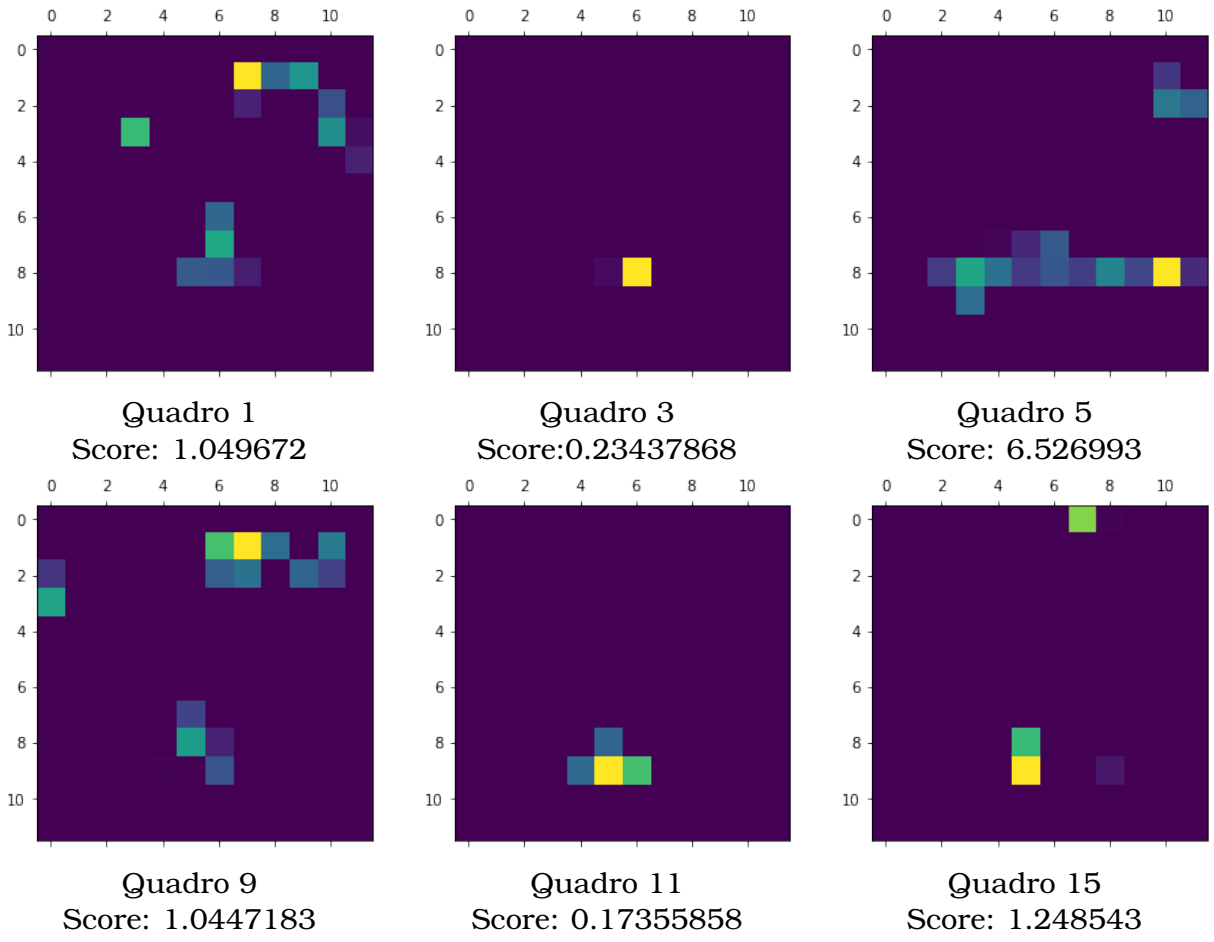


Figura 4.3: Amostras visuais do mapa de ativação em quadros de um volume de TC.

**Algoritmo 4:** Cálculo e Ordenação dos Scores

```
scores ← [ ]
N ←M.shape[2] {Obtém o número de quadros na matriz M}
para z de 1 até N faça
    score_z ← 0
    para i de 1 até x faça
        para j de 1 até y faça
            score_z ← score_z + M_{i,j,z} {Calcula o score para o quadro z}
        fim-para
    fim-para
    scores.append(score_z) {Adiciona o score à lista de scores}
fim-para
lista_indices_ordenada ← argsort(scores)[:: −1] {Lista de índices dos quadros mais relevantes ordenada}
retorne lista_indices_ordenada
```

## 4.3 Aplicação do Método a modelos de classificação

Com os quadros escolhidos, pode-se utilizar um modelo de uma rede convolucional cuja entrada será ajustada para trabalhar com a nova profundidade, agora determinada pelo número $k$ de quadros escolhidos. Com a redução da profundidade do volume original (eixo $z$) pode-se aumentar as dimensões de largura ($x$) e altura ($y$) do volume para o mais próximo possível do tamanho original, promovendo maior atenção aos detalhes dos dados.

### 4.3.1 Ajuste do tamanho da entrada

A adoção da maior resolução possível após a seleção dos quadros mais relevantes oferece inúmeros benefícios. Em primeiro lugar, possibilita que o modelo de aprendizado profundo opere com granularidades menores, favorecendo uma análise mais minuciosa e precisa dos quadros significativos. Isso pode levar a um melhor desempenho do modelo e a uma qualidade superior das informações extraídas. Com isso em mente, os experimentos deste trabalho priorizaram a utilização da resolução espacial mais próxima possível do volume original, ajustando o número de quadros de acordo com os recursos computacionais disponíveis.

Mesmo em condições nas quais seja relevante um meio termo nas reduções espaciais e temporais do volume pode-se regular o aumento de resolução conforme o número de quadros escolhidos também e ajustar a resolução espacial com base no número de quadros selecionados, com isso é possível equilibrar a complexidade computacional e a qualidade das informações obtidas. Segue a fundamentação do ajuste das dimensões do volume conforme os $k$ quadros selecionados:

Supondo a existência de ajuste do tamanho de entrada (largura, altura e número de quadros) que se adapte aos recursos de hardware disponíveis, tem-se:

1. Seja $\boldsymbol{V}$ um volume de tomografia computadorizada.
2. Sendo $\boldsymbol{V}$ constituído de $\boldsymbol{n}$ quadros de imagens de CT em tons de cinza. Tem-se o tamanho ($\boldsymbol{T}$) de $\boldsymbol{V}$:

$$V = (largura, altura, canais, quadros)$$

$$T(V) = largura \times altura \times 1 \times n$$

$$T(V) = largura \times altura \times n$$

Em cada volume será selecionado um subconjunto de $k$ quadros significativos.

1. O novo tamanho será de:

$$V = largura \times altura \times k$$

2. Dessa forma o tamanho da entrada irá ser reduzido na razão ***r*** que é simplesmente o tamanho original em razão do tamanho após a seleção dos quadros

$$r = \frac{n}{k}$$

Considerando $T_0$ como o tamanho inicial e n o número de quadros de um volume de CT, e $w$ e $h$ a largura e altura da imagem em pixels, temos:

$$T_0 = w \times h \times n$$

- Sendo $T_1$ o tamanho após a seleção de um subconjunto de $k$ quadros.

$$T_1 = w \times h \times k$$

- Existe uma constante $\alpha$ que podemos aumentar a resolução espacial do volume sem aumento do tamanho total da entrada:

$$V = (largura \times \alpha, altura \times \alpha, k)$$

Calculando $\alpha$:

$$T_2 = (w \times \alpha) \times (h \times \alpha) \times k$$

Para mesmo tamanho:

$$T_2 = T_0$$

$$(w \times \alpha) \times (h \times \alpha) \times k = w \times h \times n$$

$$\alpha^2 \times k = n$$

$$\alpha^2 = \frac{n}{k}$$

Como $\frac{n}{k}$ = r

$$\alpha = \sqrt{r}$$

Com a redução da profundidade do volume de entrada, é possível trabalhar com menores granularidades, podendo utilizar uma resolução de imagem maior na constante $\sqrt{r}$ com o volume final sendo:

$$V = (largura, altura, k)$$

$$V = (largura \times \sqrt{r}, altura \times \sqrt{r}, k)$$

Isso permite aumentar a resolução espacial da imagem sem aumentar o custo computacional inicial, o que é especialmente importante quando se trabalha com dados clínicos e reconhecimento de possíveis regiões de interesse.

### 4.3.2 Treino de modelos de aprendizado profundo

Como mencionado na Seção 3.5, uma rede neural convolucional profunda é uma abordagem de aprendizado profundo que tem se mostrado eficaz na classificação de volumes de tomografia computadorizada. A principal razão para isso é sua capacidade de extrair características espaciais em três dimensões, o que permite uma análise mais precisa e detalhada dos dados.

A arquitetura de uma rede neural convolucional profunda é composta por diversas camadas, incluindo camadas convolucionais, de *pooling* e totalmente conectadas. As camadas convolucionais em profundidade são responsáveis por detectar as características espaciais dos dados, como bordas, texturas e padrões. As camadas de convolução pontual aprendem as características temporais dos dados, e finalmente, as camadas totalmente conectadas são responsáveis por realizar a classificação propriamente dita.

Dada a complexidade do modelo, é necessário selecionar os quadros e reduzir o dado de entrada para o treino de um modelo mais complexo, que realizará a classificação dos volumes de TC.

Dessa forma, duas redes convolucionais serão utilizadas para realização dos experimentos: uma responsável por selecionar os quadros relevantes e outra para classificação do volume que explorará dados com maiores dimensões espaciais. Sumarizando, os seguintes passos foram efetuados para a obtenção dessa rede, na qual será aplicado o Grad-CAM para obtenção dos quadros significativos:

1. Normalizar os dados volumétricos.
2. Realizar uma pré-seleção dos quadros com uma das técnicas descritas na Seção 3.2 levando-se em conta os recursos computacionais disponíveis. A pré-seleção inicial tem como objetivo reduzir o volume de dados e permitir o treinamento do modelo de aprendizado profundo de maneira eficiente. A etapa subsequente, que envolve a seleção dos quadros mais relevantes com a técnica Grad-CAM, aprimorará essa pré-seleção, garantindo uma análise mais detalhada e precisa dos quadros significativos.
3. Treino de uma rede neural convolucional para classificação preservando a dimensão de profundidade até a camada que será extraída o mapa de calor.

Após esse processo, os volumes da base de dados que serão utilizados com maior resolução espacial são amostrados com os seguintes passos:

1. Entrada do volume na rede.
2. Extração da matriz representativa do mapa de calor da última camada convolucional em profundidade.
3. Cálculo dos *scores* como no Algoritmo 4.

4. Escolha dos $k$ quadros mais relevantes.

Esse processo pode ser condensado no seguinte algoritmo para treino do modelo responsável pela seleção dos quadros:

**Algoritmo 5:** Treino modelo GSS

**Requisito:** Base de dados com volumes de alta resolução espacial

$GSS_{Model}$ ←3DCNN-C
**para** cada época de treinamento **faça**
  **para** para cada batch de volumes da Base de Dados **faça**
    batch = normalize(batch)
    batch = SIZ(batch)
    train($GSS_{Model}$, batch)
  **fim-para**
**fim-para**
Retorne $GSS_{Model}$

E no Algoritmo 6 para treino do modelo responsável em classificar o volume:

**Algoritmo 6:** Algoritmo para treino com GSS

**Requisito:** Base de dados com volumes de alta resolução espacial
**Assegura:** O número de quadros a ser selecionado $k$ é menor que o número de quadros do volume

$GSS_{Model}$ ←Custom CNN Treinada no Algoritmo 5
$CNN_{Model}$ ←CNN a ser treinada com os quadros selecionados.
**para** cada época de treinamento **faça**
  **para** para cada batch de volumes do Dataset **faça**
    batch = normalize(batch)
    $quadros_{selecionados}$ = GSS(batch,$GSS_{Model}$, $k$) {Aplicação da seleção *GSS* para selecionar os $k$ quadros mais significativos de cada volume em batch com um modelo $GSS_{Model}$}
    batch = batch[$quadros_{selecionados}$]
    train($CNN_{Model}$, batch)
  **fim-para**
**fim-para**
Retorne $CNN_{Model}$

Quanto a etapa de inferência pode ser sintetizado na Figura 4.4.

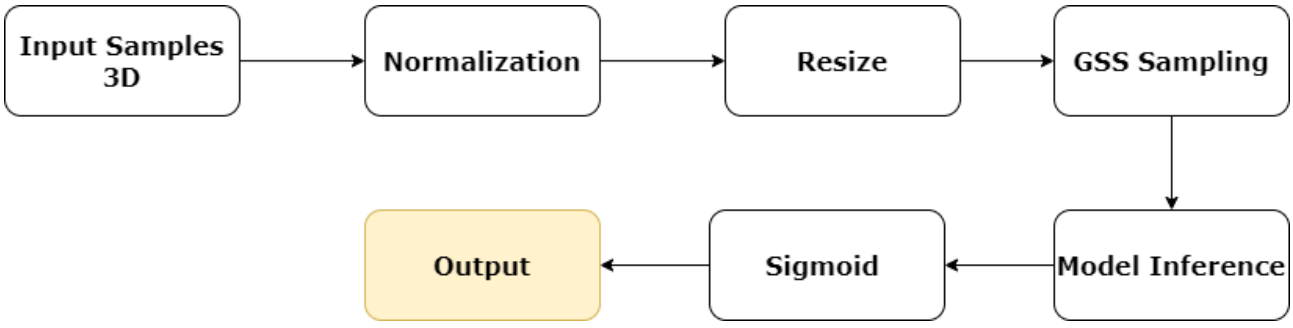


Figura 4.4: Esquema *Grad-Cam Slice Selection (GSS)*.

## 4.4 Considerações Finais

A técnica proposta de utilizar o Grad-CAM em conjunto com uma rede que emprega convolução em profundidade apresenta várias vantagens em comparação a outras abordagens de aprendizado profundo. Ao utilizar o mapa de gradiente produzido pelo Grad-CAM para selecionar os canais de saída mais relevantes de uma camada convolucional em profundidade, é possível reduzir o número de parâmetros na rede, tornando-a mais eficiente computacionalmente. Além disso, essa abordagem aumenta a interpretabilidade do modelo, destacando os recursos mais importantes para a previsão, e pode aprimorar o desempenho em tarefas específicas, permitindo que a rede se concentre nas informações mais relevantes da entrada e aumentando seu potencial para obter melhores resultados no treino.

No entanto, é crucial considerar o redimensionamento das dimensões de largura e altura para garantir a viabilidade do treinamento no hardware selecionado, além de escolher adequadamente as operações e camadas para garantir o mapeamento dos quadros do volume. A arquitetura proposta na Tabela 4.1 pode servir como ponto de partida para a implementação dessa técnica em outras aplicações.

Capítulo 5

# Avaliação Experimental

## 5.1 Considerações Iniciais

Neste capítulo, será mostrada a avaliação experimental da técnica *Gradcam Slice Selection* (GSS), aplicada à classificação de volumes de tomografia computadorizada. A proposta é analisar a performance desta técnica, assim como sua aplicabilidade em diferentes contextos médicos e comparação com outros métodos do estado da arte. A avaliação tem como base a análise quantitativa e qualitativa dos resultados obtidos com a técnica proposta, além de um estudo das métricas de desempenho.

O capítulo está organizado em diferentes seções, que incluem uma descrição detalhada dos dados utilizados nos experimentos, a configuração experimental, as métricas de avaliação e, por fim, os resultados obtidos. Para atingir os objetivos propostos, este estudo utiliza dois conjuntos de dados distintos: MosMedData e CQ500, que fornecem imagens de tomografia computadorizada dos pulmões e do crânio, respectivamente. A escolha desses conjuntos de dados tem como objetivo avaliar a aplicabilidade da técnica GSS em diferentes contextos médicos e garantir uma análise mais abrangente de seu desempenho. Além disso, serão analisadas diferentes configurações de treinamento e amostragem, a fim de identificar as melhores práticas para a aplicação da técnica GSS em problemas reais.

Nas seções seguintes, serão apresentados os detalhes do processamento dos dados e das configurações experimentais, assim como uma análise dos resultados obtidos. A discussão dos resultados possibilitará compreender as vantagens e limitações da técnica GSS, bem como sua eficácia em comparação com outros métodos do estado da arte.

### 5.1.1 Dados

Os experimentos foram conduzidos utilizando os conjuntos de dados *MosMedData: Chest CT Scans with COVID-19 Related Findings* (Morozov et al., 2020) e CQ500 (Chilamkurthy

et al., 2018). O conjunto de dados Mosmed contém tomografias computadorizadas dos pulmões com e sem achados relacionados à COVID-19. Algumas das imagens foram anotadas com máscaras binárias de pixels que representam regiões de interesse, como opacificação em vidro fosco e consolidações. Os volumes de TC foram obtidos entre 1º de março de 2020 e 25 de abril de 2020, a partir de hospitais municipais em Moscou, Rússia. Mais detalhes sobre a base de dados podem ser encontrados na Tabela 5.1.

| **Propriedade** | **Valor** |
|---|---|
| Origem | Rússia |
| Número de Estudos | 1110 |
| Número de Pacientes | 1110 |
| Resolução | 512 x 512 |
| Número de quadros(min./max.) | 31/72 |
| Distribuição por sexo(%) (M/F/O) | 42/56/2 |
| Distribuição por idade, anos (min./média/max.) | 18/47/97 |

Tabela 5.1: Propriedades da base *MosMed* (Morozov et al., 2020).

Embora o foco deste estudo de caso seja a classificação de volumes de tomografia computadorizada, é relevante avaliar a aplicabilidade do método em outros contextos médicos. A base de dados CQ500 é composta por 491 TCs cranianas, com uma média de 297 quadros por volume, totalizando 193.317 quadros. Esses dados, disponibilizados publicamente, incluem imagens DICOM *(Digital Imaging and Communications in Medicine)* anonimizadas. O conjunto de dados é fornecido pelo Centro de Pesquisa Avançada em Imagem, Neurociências e Genômica (CARING) em Nova Delhi, Índia. Essas informações podem ser utilizadas para treinar e testar algoritmos de análise de imagens médicas na detecção de anormalidades cranianas, como lesões, tumores e aneurismas.

Os conjuntos de dados Mosmed e CQ500 foram selecionados para avaliação devido ao menor número de quadros por volume em comparação com outras bases, possibilitando maior agilidade na obtenção de resultados e melhor ajuste dos modelos. A divisão dos dados seguiu a abordagem de *holdout*, com 70% treino, 15% validação, 15% teste

No que se refere ao processamento dos dados brutos, os voxels foram armazenados em unidades Hounsfield (HU) com valores entre -1024 e 2000 HU. Foi adotado o valor de 400 HU como limite superior, uma vez que os ossos apresentam diferentes radiointensidades acima deste valor. Para processar os volumes de tomografia, os valores de HU foram limitados para o intervalo de -1000 a 400 HU. Em seguida, foram os volumes de dados normalizados, garantindo que a distribuição dos valores de intensidade estivesse na mesma escala. O processamento dos volumes de tomografia envolveu os seguintes procedimentos:

- Rotação de 90 graus para uniformizar a orientação dos quadros;
- Normalização das unidades HU, resultando em valores entre 0 e 1;
- Redimensionamento de largura, altura e profundidade, levando em consideração a entrada do modelo e os recursos computacionais disponíveis.

Em resumo, as bases de dados Mosmed e CQ500 foram escolhidas por favorecerem maior agilidade na obtenção de resultados e melhor ajuste dos modelos. A divisão dos dados foi realizada com *holdout*, seguindo a proporção de 70% treino, 15% validação e 15% teste. O pré-processamento dos dados incluiu rotação em 90 graus, normalização das unidades HU para valores entre 0 e 1 e redimensionamento de largura, altura e profundidade, conforme a entrada do modelo e os recursos disponíveis. A avaliação quantitativa e qualitativa da técnica proposta será abordada nas seções seguintes, onde também será discutido seu desempenho em diferentes configurações de treinamento e amostragem, além de sua eficiência computacional e interpretabilidade do modelo.

### 5.1.2 Configuração dos Experimentos

Os experimentos foram executados em um ambiente equipado com uma placa de vídeo NVIDIA A100 de 40 GB de memória, que proporciona alto desempenho de computação. Para otimização, foi utilizado o algoritmo *Adam* com uma taxa de aprendizado de $10^{-4}$, que foi determinada experimentalmente para alcançar um equilíbrio entre rapidez de convergência e precisão do modelo. Além disso, para inicializar os pesos, foi empregado o método de *Inicialização Glorot Normal* (Glorot e Bengio, 2010), que é amplamente utilizado em redes neurais convolucionais.

A arquitetura do modelo consiste em 5 camadas convolucionais, seguindo o modelo de arquitetura apresentado na Seção 4.3.2 para o modelo que é responsável por extrair os quadros mais significativos, o 3DCNN-C. A função de ativação usada nas camadas de convolução é a *Rectified Linear Activation Function (ReLU)*, e, finalmente, a última camada, que se trata de um problema binário, usa a função de ativação *sigmoid*.

Para o problema em questão, a dimensão de entrada foi escolhida como (128,128,40), levando em consideração as características das bases de dados Mosmed e CQ500 e os recursos computacionais disponíveis. A escolha de 40 quadros se baseia na média do intervalo de quadros presente na base Mosmed, que varia entre 31 e 72 quadros. Essa seleção permite uma representação equilibrada das informações contidas nos volumes, sem sobrecarregar os recursos computacionais.

Essa entrada gera um mapa de calor por meio do Grad-Cam na forma (12,12,40), ou seja, cada quadro do volume original é dividido em uma matriz (12,12), como exemplificado na Tabela 4.3, que representa a contribuição daquela região na decisão do modelo. A Figura 5.1 mostra um esquema geral da técnica aplicada nos experimentos.

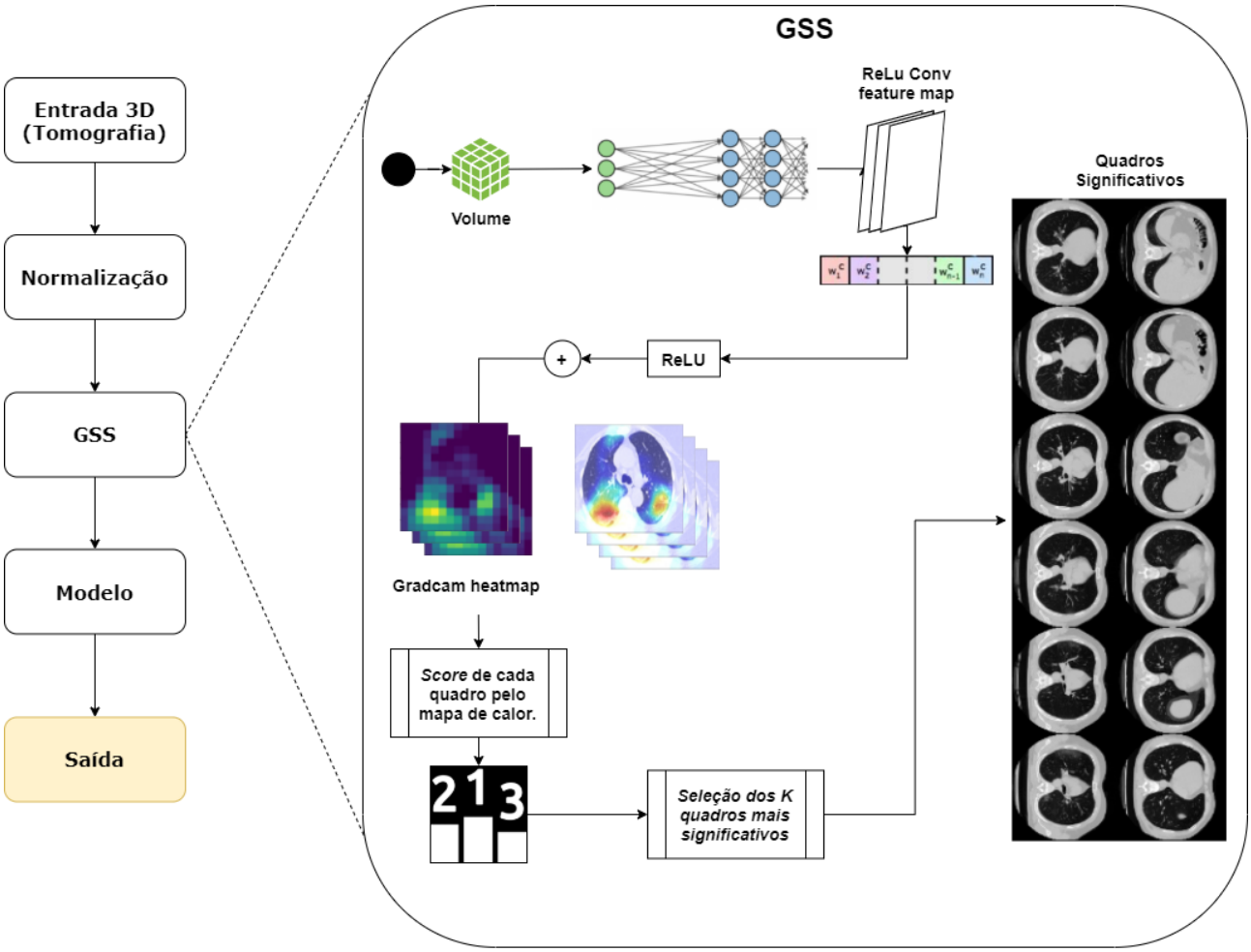


Figura 5.1: Esquema Geral no qual GSS faz parte do processo para treino e inferência.

### 5.1.3 Modelos

Neste trabalho, são explorados dois modelos principais de redes neurais convolucionais 3D (3D-CNNs): uma arquitetura personalizada de 3D-CNN (Silva et al., 2020) e o amplamente reconhecido modelo *Convolutional 3D* (C3D) (Tran et al., 2015).

- 3D-CNN (Silva et al., 2020): A arquitetura foi desenvolvida para se adequar especificamente à tarefa de detecção de anomalias em dados tridimensionais. A arquitetura é composta por cinco camadas de convolução 3D, cada uma seguida por uma camada de *Max-pooling* e *Batch Normalization*. Baseado em experimentações, as camadas de *Batch Normalization* superaram o desempenho das camadas de *Dropout*. Adicionalmente, foram adotados a Inicialização Glorot Normal e a função de ativação ReLU nas camadas de convolução. O modelo empregado se equipararou ao estado da arte em termos de acurácia ao mesmo tempo o com menor custo computacional dentre os modelos testados no trabalho.
- Modelo C3D: amplamente utilizado para análise de dados de vídeo, o modelo C3D é capaz de capturar características temporais e espaciais de volumes de vídeo através

da execução de convoluções tridimensionais (largura, altura e tempo). Ele tem sido amplamente empregado para tarefas de reconhecimento de ação em vídeos, demonstrando ser um modelo eficaz para extração de características em contextos tridimensionais.

### 5.1.4 Métricas

A tarefa é avaliada como um problema de classificação binária. Embora outras métricas tenham sido utilizadas, a forma de comparação dos modelos seguirá o padrão observado na revisão bibliográfica, sendo *Area Under the ROC Curve* (AUC) com a adição de F1 Score que pode trazer uma boa expressão do desempenho dos modelos.

- *Area Under the ROC Curve* (AUC): A curva ROC (Receiver Operating Characteristic) permite avaliar a performance de um modelo de classificação binária ao variar o limiar de classificação. O eixo X representa a taxa de falsos positivos (especificidade), enquanto o eixo Y representa a taxa de verdadeiros positivos (sensibilidade). Já AUC (Area Under the ROC Curve) é uma métrica numérica que indica a qualidade geral do modelo e é obtida pela integral da curva ROC, avaliando a relação entre verdadeiros positivos e falsos positivos para todos os possíveis limiares de classificação. O valor da AUC varia de 0 a 1, onde 1 indica um modelo de classificação perfeito e 0.5 um modelo aleatório.
- *F1 Score*: É uma métrica que considera tanto a precisão quanto a revocação do modelo. A precisão é a proporção de verdadeiros positivos em relação ao total de resultados positivos, enquanto a revocação é a proporção de verdadeiros positivos em relação ao total de positivos reais. Seu valor varia entre 0 e 1, onde 1 é o melhor valor possível. A fórmula matemática para o cálculo do F1 Score é:

$$F1 = 2 \times \frac{Precisao \times Revocacao}{Precisao + Revocacao}$$

A escolha dessas duas métricas (AUC e F1 Score) permite uma avaliação abrangente dos modelos de classificação binária. A AUC oferece uma medida geral da capacidade do modelo de distinguir entre as duas classes em todos os limiares de classificação, enquanto o F1 Score fornece uma avaliação mais detalhada da precisão e revocação do modelo, priorizando o equilíbrio entre as duas. Juntas, essas métricas oferecem uma visão robusta do desempenho dos modelos, adequada para a comparação dos resultados obtidos.

### 5.1.5 Resultados

A Tabela 5.2 apresenta os resultados de um modelo de classificação binária de volumes de tomografia computadorizada usando diferentes métodos de seleção de quadros (SSS, ESS,

SIZ e GSS) e diferentes configurações, como resolução (512x512 e 224x224) e número de quadros selecionados (30 e 12). Os resultados são avaliados em dois conjuntos de dados distintos, Mosmed e CQ500, e com dois modelos de aprendizado profundo, C3D e a mesma arquitetura (da Silva et al., 2021). As métricas de avaliação utilizadas são AUC (área sob a curva ROC) e F1 Score.

| **Método** | **Resolução** | **Número de Quadros Selecionados** | **Mosmed** | | | | **CQ500** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | **C3D** | | **CNN da Silva et al. (2021)** | | **C3D** | | **CNN da Silva et al. (2021)** | |
| | | | **AUC** | **F1 Score** | **AUC** | **F1 Score** | **AUC** | **F1 Score** | **AUC** | **F1 Score** |
| SSS | 512x512 | 30 | 0,693 | 0,845 | 0,788 | 0,816 | 0,710 | 0,760 | 0,760 | 0,760 |
| | | 12 | 0,617 | 0,845 | 0,733 | 0,851 | 0,680 | 0,756 | 0,730 | 0,743 |
| | 224x224 | 30 | 0,671 | 0,775 | 0,721 | 0,398 | 0,660 | 0,496 | 0,710 | 0,584 |
| | | 12 | 0,522 | 0,781 | 0,707 | 0,742 | 0,640 | 0,687 | 0,690 | 0,689 |
| ESS | 512x512 | 30 | 0,715 | 0,825 | 0,763 | 0,855 | 0,720 | 0,782 | 0,770 | 0,776 |
| | | 12 | 0,683 | 0,800 | 0,727 | 0,701 | 0,700 | 0,701 | 0,750 | 0,724 |
| | 224x224 | 30 | 0,671 | 0,800 | 0,737 | 0,834 | 0,690 | 0,755 | 0,740 | 0,748 |
| | | 12 | 0,652 | 0,795 | 0,625 | 0,701 | 0,670 | 0,685 | 0,720 | 0,702 |
| SIZ | 512x512 | 30 | 0,674 | 0,846 | 0,754 | 0,857 | 0,760 | 0,806 | 0,810 | 0,808 |
| | | 12 | 0,697 | 0,815 | 0,737 | 0,870 | 0,740 | 0,800 | 0,790 | 0,795 |
| | 224x224 | 30 | 0,703 | 0,787 | 0,736 | 0,838 | 0,730 | 0,780 | 0,780 | 0,780 |
| | | 12 | 0,615 | 0,809 | 0,693 | 0,797 | 0,710 | 0,751 | 0,760 | 0,755 |
| **GSS** | **512x512** | **30** | **0,753** | **0,865** | **0,804** | **0,838** | **0,800** | **0,819** | **0,850** | **0,834** |
| | | **12** | **0,732** | **0,866** | **0,775** | **0,881** | **0,780** | **0,827** | **0,830** | **0,829** |
| | **224x224** | **30** | **0,690** | **0,769** | **0,737** | **0,853** | **0,770** | **0,810** | **0,820** | **0,815** |
| | | **12** | **0,628** | **0,845** | **0,714** | **0,795** | **0,750** | **0,772** | **0,800** | **0,786** |

Tabela 5.2: Resultados

Analisando os resultados, pode-se tirar as seguintes conclusões:

- O método GSS apresenta consistentemente os melhores resultados em comparação com os outros métodos de seleção de quadros, tanto em termos de AUC quanto de F1 Score. Isso é verdade em todas as configurações e para ambos os modelos de aprendizado profundo (C3D e CNN da Silva et al. (2021)). Portanto, o método GSS é o mais promissor para a tarefa de detecção de anomalias em volumes de tomografia.
- O método SIZ apresenta melhores resultados do que SSS e ESS em geral, mas ainda não consegue superar os resultados obtidos pelo GSS.
- Os resultados mostram que selecionar 30 quadros tem um desempenho geral melhor do que selecionar 12 quadros. Isso indica que ter mais quadros disponíveis para o modelo pode levar a uma melhor detecção de anomalias.
- A resolução de 512x512 geralmente produz melhores resultados em comparação com a resolução de 224x224, o que sugere que uma resolução maior fornece informações mais valiosas para o modelo na detecção de anomalias.
- O modelo CNN (da Silva et al., 2021) tende a ter resultados melhores do que o modelo C3D em todas as configurações e conjuntos de dados.

Em resumo, o método GSS com 30 quadros selecionados e resolução de 512x512, utilizando o modelo proposto neste trabalho, apresenta os melhores resultados para a detecção de anomalias em volumes de tomografia. Esse resultado era esperado, uma vez que essa

configuração possui a maior resolução e número de quadros. A escolha das configurações adequadas quanto ao equilíbrio entre resolução e número de quadros deve ser avaliada de acordo com o tipo de problema a ser abordado, a fim de melhorar o desempenho dos modelos de aprendizado profundo nesta tarefa específica. Caso a patologia esperada seja muito sutil ou pequena, é recomendado priorizar maiores resoluções espaciais; por outro lado, quando o aspecto tridimensional é mais relevante, deve-se utilizar um maior número de quadros.

Dentre outros fatores, esse aspecto pode ser esclarecido pela Figura 5.2, que exibe a avaliação de 50 volumes de tomografia computadorizada de pacientes com sequelas de Covid-19, anotados por especialistas humanos. Realizou-se uma análise dos quadros e das partições do volume onde ocorriam mais anotações. A distribuição das regiões de interesse não é uniforme, e os quadros que mais contribuem para a avaliação clínica encontram-se em posições variadas do volume. Dessa forma, a seleção de quadros também não pode ser equilibrada, como implementado em outras técnicas. Observa-se que em poucos volumes sequer haviam regiões de interesse (*ROI*) nos quadros de início e fim. Ao dividir os volumes em 10 partes, nota-se que, a partir da segunda partição, a maior região de interesse pode estar em qualquer posição do volume, principalmente entre a terceira e a sexta posição.

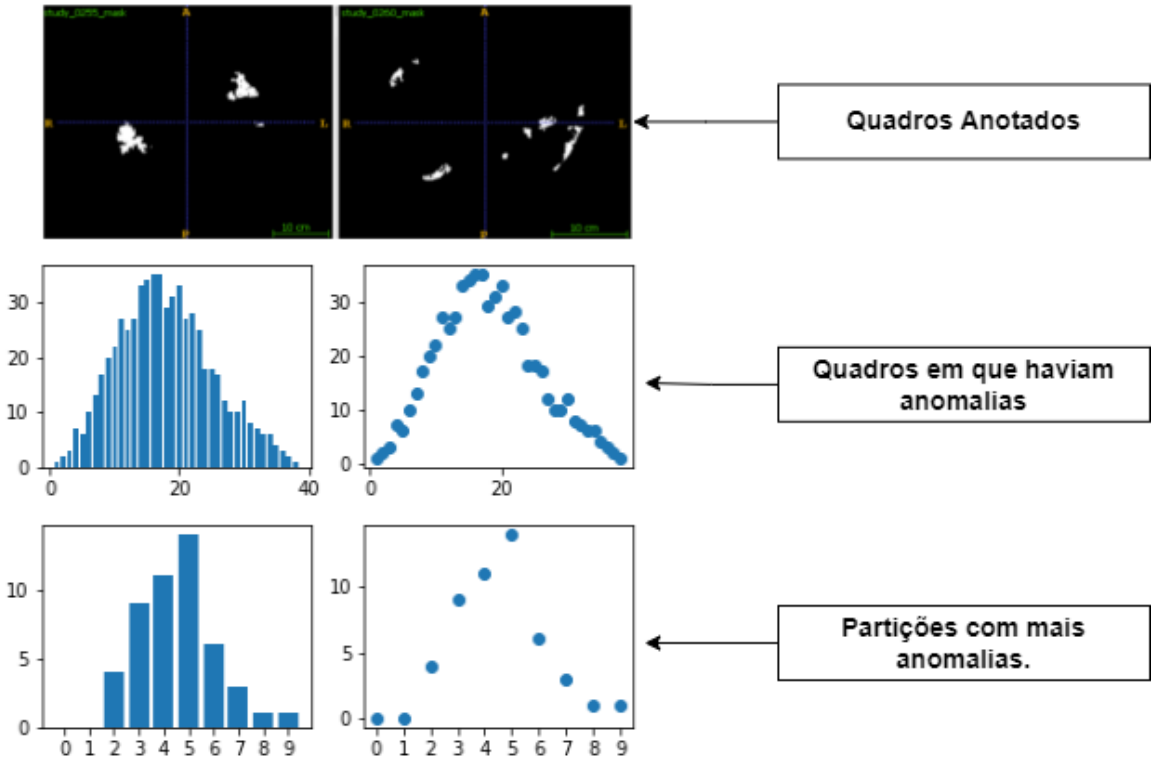


Figura 5.2: Avaliação de 50 volumes anotados por especialistas.

Capítulo 6

# Conclusão

## 6.1 Considerações Finais

Este trabalho propôs o método Grad-Cam Slice Selection (GSS) para selecionar quadros mais significativos em volumes de tomografia computadorizada, visando reduzir o tamanho dos dados e melhorar o desempenho de algoritmos de aprendizado profundo na detecção de anomalias. A técnica GSS foi avaliada experimentalmente utilizando os conjuntos de dados MosMedData e CQ500, e comparada o estado da arte.

Os resultados demonstraram que o método GSS apresentou desempenho superior em comparação com os outros métodos de seleção de quadros, tanto em termos de área sob a curva ROC (AUC) quanto de F1 Score. Isso foi observado em todas as configurações testadas e para ambos os modelos de aprendizado profundo (C3D e CNN adaptado de da Silva et al. (2021)). Além disso, o método GSS se mostrou eficiente em diferentes configurações de treinamento e amostragem, melhorando a interpretabilidade do modelo e sua eficiência computacional.

A utilização do GSS para selecionar os quadros mais significativos em volumes de tomografia computadorizada oferece várias vantagens no âmbito do aprendizado profundo. Entre elas, destaca-se a capacidade de extrair informações relevantes de regiões de interesse, proporcionando uma representação compacta e informativa dos dados. Isso permite que os modelos de aprendizado profundo possam aprender características discriminativas de forma mais eficiente, resultando em um melhor desempenho na detecção de anomalias em TC.

Além disso, a seleção de quadros significativos proporcionada pelo método GSS facilita o treinamento e a validação dos modelos, permitindo um ajuste mais rápido e preciso dos parâmetros e melhorando a eficiência computacional. Essa vantagem é particularmente importante em cenários com recursos computacionais limitados, onde o tempo de treinamento e a capacidade de processamento são fatores críticos ou limitados.

Para trabalhos futuros, seria interessante explorar a integração do método proposto com outras técnicas de aprendizado profundo, como modelos generativos, a fim de aprimorar ainda mais o desempenho e a robustez do processo de extração. Além disso, avaliar o método proposto em uma ampla variedade de tomografias computadorizadas de diferentes modalidades e populações de pacientes seria valioso para validar sua generalização ainda mais.

Em resumo, o método GSS proposto nesta dissertação de mestrado demonstrou ser uma abordagem promissora para a seleção de quadros em volumes de tomografia computadorizada, contribuindo para a melhoria do desempenho dos modelos de aprendizado profundo na detecção de anomalias e fornecendo um equilíbrio eficiente entre a redução de dimensionalidade e a preservação das informações relevantes dos dados. As contribuições deste trabalho têm potencial para impactar positivamente o desenvolvimento de algoritmos de análise de imagens médicas e aprimorar a detecção e o diagnóstico de condições médicas em tomografia computadorizada.

## 6.2 Principais Contribuições e Limitações

A proposta apresentada nesta dissertação tem como principal contribuição a capacidade de extrair quadros significativos em volumes de tomografia computadorizada de forma mais inteligente, que não causa deformações no volume resultante. Os resultados mostram que a abordagem proposta baseada em redes neurais convolucionais e Grad-CAM é altamente precisa e pode ser aplicada em diversas aplicações.

No entanto, é importante ressaltar que a abordagem foi testada apenas em volumes de TC de pulmões e anomalias cranianas. Portanto, é pertinente investigar sua aplicação em outros segmentos da medicina. Seria desejável realizar estudos adicionais, envolvendo especialistas de diferentes domínios da área da saúde, para avaliar a aplicação deste método em outras imagens radiológicas.

Outra limitação é que o método GSS foi avaliado apenas em modelos de aprendizado profundo baseados em C3D e CNN adaptado. Uma expansão da análise para incluir outros tipos de arquiteturas de aprendizado profundo poderia proporcionar uma compreensão mais abrangente das potencialidades e limitações do método proposto.

Além disso, seria relevante investigar a eficácia do GSS quando aplicado a conjuntos de dados com diferentes níveis de qualidade e resolução das imagens. Essa análise adicional poderia fornecer informações valiosas sobre a robustez do método em cenários clínicos variados, onde a qualidade das imagens de TC pode ser influenciada por diferentes fatores, como equipamentos e protocolos de aquisição.

Por fim, vale considerar a investigação de estratégias de otimização para tornar o método GSS ainda mais eficiente computacionalmente, possibilitando sua aplicação em tempo real ou em cenários com grandes volumes de dados.

Em suma, apesar das limitações, o método GSS demonstrou ser uma abordagem promissora para selecionar quadros em volumes de tomografia computadorizada. As contribuições deste trabalho têm o potencial de impactar positivamente o desenvolvimento de algoritmos de análise de imagens médicas e aprimorar a detecção e o diagnóstico de condições médicas em tomografia computadorizada. A investigação de suas aplicações em outros contextos médicos e a avaliação de sua eficácia em combinação com outras técnicas de aprendizado profundo podem impulsionar ainda mais os avanços na área de análise de imagens médicas.

# Referências Bibliográficas


Ahmed, E., Saint, A., Shabayek, A. E. R., Cherenkova, K., Das, R., Gusev, G., Aouada, D., e Ottersten, B. (2018). A survey on deep learning advances on different 3d data representations. *arXiv preprint arXiv:1808.01462*. 1

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press. 5

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961. 21

Attia, M. W., Abou-Chadi, F., Moustafa, H. E.-D., e Mekky, N. (2015). Classification of ultrasound kidney images using pca and neural networks. *Int J Adv Comput Sci Appl*, 6(4):53–57. 29

Balduzzi, D., Frean, M., Leary, L., Lewis, J., Ma, K. W.-D., e McWilliams, B. (2017). The shattered gradients problem: If resnets are the answer, then what is the question? Em *International Conference on Machine Learning*, páginas 342–350. PMLR. 16

Brink, H., Richards, J. W., Fetherolf, M., e Cronin, B. (2017). *Real-world machine learning*. Manning. 6, 7

Calhoun, V. D. e Adali, T. (2012). Multisubject independent component analysis of fmri: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. *IEEE reviews in biomedical engineering*, 5:60–73. 31

Calhoun, V. D., Potluru, V. K., Phlypo, R., Silva, R. F., Pearlmutter, B. A., Caprihan, A., Plis, S. M., e Adalı, T. (2013). Independent component analysis for brain fmri does indeed select for maximal independence. *PloS one*, 8(8):e73309. 32

Chattopadhay, A., Sarkar, A., Howlader, P., e Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. Em *2018 IEEE winter conference on applications of computer vision (WACV)*, páginas 839–847. IEEE. 18

Cheng, J., Wang, H., Li, R., Li, X., Zhou, X., Yang, X., Wang, Y., Xiong, L., Fan, H., Wang, T., et al. (2022). A two-stage multiresolution neural network for automatic diagnosis of hepatic echinococcosis from ultrasound images: A multicenter study. *Medical Physics*, 49(5):3199–3212. 33, 36

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., Mahajan, V., Rao, P., e Warier, P. (2018). Development and validation of deep learning algorithms for detection of critical findings in head ct scans. *arXiv preprint arXiv:1803.05854*. 51

Cormen, T. H., Leiserson, C. E., Rivest, R. L., e Stein, C. (2009). *Introduction to algorithms*. MIT press. 14

da Cruz, L. B., Araújo, J. D. L., Ferreira, J. L., Diniz, J. O. B., Silva, A. C., de Almeida, J. D. S., de Paiva, A. C., e Gattass, M. (2020). Kidney segmentation from computed tomography images using deep neural network. *Computers in Biology and Medicine*, 123:103906. 36

da Silva, L. A., dos Santos, E. M., Araújo, L., Freire, N. S., Vasconcelos, M., Giusti, R., Ferreira, D., Jesus, A. S., Pimentel, A., Cruz, C. F., et al. (2021). Spatio-temporal deep learning-based methods for defect detection: An industrial application study case. *Applied Sciences*, 11(22):10861. 43, 56, 58

Del Rio, C., Collins, L. F., e Malani, P. (2020). Long-term health consequences of covid-19. *Jama*, 324(17):1723–1724. 2

Du, N., Zhang, Z., Xiao, Y., Jiang, L., et al. (2021). Fast independent component analysis algorithm-based functional magnetic resonance imaging in the diagnosis of changes in brain functional areas of cerebral infarction. *Contrast Media & Molecular Imaging*, 2021. 31

Fernando, B., Bilen, H., Gavves, E., e Gould, S. (2017). Self-supervised video representation learning with odd-one-out networks. Em *Proceedings of the IEEE conference on computer vision and pattern recognition*, páginas 3636–3645. 26

Francois, C. (2017). Deep learning with python. 6, 8

Gao, X. W., Hui, R., e Tian, Z. (2017). Classification of ct brain images based on deep learning networks. *Computer methods and programs in biomedicine*, 138:49–56. 24, 37

Gentili, A. (2019). Imageclef2019: Tuberculosis-severity scoring and ct report with neural networks, transfer learning and ensembling. Em *CLEF (Working Notes)*. 37

Glorot, X. e Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. Em *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, páginas 249–256. JMLR Workshop and Conference Proceedings. 53

Goodfellow, I., Bengio, Y., e Courville, A. (2016). *Deep learning*. MIT press. 9, 10, 11

Gordaliza, P. M., Vaquero, J. J., Sharpe, S., Gleeson, F., e Munoz-Barrutia, A. (2019). A multi-task self-normalizing 3d-cnn to infer tuberculosis radiological manifestations. *arXiv preprint arXiv:1907.12331*. 27

Grewal, M., Srivastava, M. M., Kumar, P., e Varadarajan, S. (2018). Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans. Em *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, páginas 281–284. IEEE. 37

Gu, Y., Lu, X., Yang, L., Zhang, B., Yu, D., Zhao, Y., Gao, L., Wu, L., e Zhou, T. (2018). Automatic lung nodule detection using a 3d deep convolutional neural network combined with a multi-scale prediction strategy in chest cts. *Computers in biology and medicine*, 103:220–231. 37

Hamadi, A., Cheikh, N. B., Zouatine, Y., Menad, S. M. B., e Djebbara, M. R. (2019). Imageclef 2019: Deep learning for tuberculosis ct image analysis. Em *CLEF (Working Notes)*. 37

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., e Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510. 1

Huang, X., Shan, J., e Vaidya, V. (2017). Lung nodule detection in ct using 3d convolutional neural networks. Em *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, páginas 379–383. IEEE. 1

Jia, T., Zhang, H., e Bai, Y. (2015). Benign and malignant lung nodule classification based on deep learning feature. *Journal of Medical Imaging and Health Informatics*, 5(8):1936–1940. 35

Jolliffe, I. T. e Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202. 29

Justus, D., Brennan, J., Bonner, S., e McGough, A. S. (2018). Predicting the computational cost of deep learning models. Em *2018 IEEE international conference on big data (Big Data)*, páginas 3873–3882. IEEE. 14

Kavitha, S., Nandhinee, P., Harshana, S., S, J. S., e Harrinei, K. (2019). Imageclef 2019: A 2d convolutional neural network approach for severity scoring of lung tuberculosis using ct images. Em *CLEF (Working Notes)*. 37

Kazlouski, S. (2019). Imageclef 2019: Ct image analysis for tb severity scoring and ct report generation using autoencoded image features. Em *CLEF (Working Notes)*. 27

Khan, S., Rahmani, H., Shah, S. A. A., e Bennamoun, M. (2018). A guide to convolutional neural networks for computer vision. *Synthesis Lectures on Computer Vision*, 8(1):1–207. 16

Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105. 15

LeCun, Y., Bengio, Y., e Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444. 15

Li, B., Zhang, T., e Xia, T. (2016). Vehicle detection from 3d lidar using fully convolutional network. *arXiv preprint arXiv:1608.07916*. 1

Li, X., Zhou, Y., Du, P., Lang, G., Xu, M., e Wu, W. (2020). A deep learning system that generates quantitative ct reports for diagnosing pulmonary tuberculosis. *Applied Intelligence*, páginas 1–12. 27

Lin, M., Chen, Q., e Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*. 18

Liu, B., Liu, P., Dai, L., Yang, Y., Xie, P., Tan, Y., Du, J., Shan, W., Zhao, C., Zhong, Q., et al. (2021). Assisting scalable diagnosis automatically via ct images in the combat against covid-19. *Scientific reports*, 11(1):4145. 33

Milletari, F., Navab, N., e Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. Em *2016 fourth international conference on 3D vision (3DV)*, páginas 565–571. IEEE. 1

Morozov, S. P., Andreychenko, A. E., Blokhin, I. A., Gelezhe, P. B., Gonchar, A. P., Nikolaev, A. E., Pavlov, N. A., Chernina, V. Y., e Gombolevskiy, V. A. (2020). Mosmeddata: data set of 1110 chest ct scans performed during the covid-19 epidemic. *Digital Diagnostics*, 1(1):49–59. 12, 25, 26, 27, 28, 51, 52

Nandi, D., Ashour, A. S., Samanta, S., Chakraborty, S., Salem, M. A., e Dey, N. (2015). Principal component analysis in medical image processing: a study. *International Journal of Image Mining*, 1(1):65–86. 30

Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., e Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors*, 19(17):3722. 37

Oja, E. e Hyvarinen, A. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430. 30

Ozdemir, O., Russell, R. L., e Berlin, A. A. (2019). A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans. *IEEE transactions on medical imaging*, 39(5):1419–1429. 37

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P., et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. *PLoS medicine*, 15(11):e1002686. 23

Ray, S., Kumar, V., Ahuja, C., e Khandelwal, N. (2018). An automatic method for complete brain matter segmentation from multislice ct scan. *arXiv preprint arXiv:1809.06215*. 36

Roberts, B. W. (2015). Ct-guided intra-abdominal abscess drainage. *Radiologic technology*, 87(2):187CT–203CT. 20

Ronneberger, O., Fischer, P., e Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Em *International Conference on Medical image computing and computer-assisted intervention*, páginas 234–241. Springer. 37

Ruppert, A. M., Sroussi, D., Khallil, A., Giot, M., Assouad, J., Cadranel, J., e Gounant, V. (2020). Detection of secondary causes of spontaneous pneumothorax: Comparison between computed tomography and chest x-ray. *Diagnostic and interventional imaging*, 101(4):217–224. 39

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., e Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. Em *Proceedings of the IEEE international conference on computer vision*, páginas 618–626. 17

Silva, F., Pereira, T., Frade, J., Mendes, J., Freitas, C., Hespanhol, V., Costa, J. L., Cunha, A., e Oliveira, H. P. (2020). Pre-training autoencoder for lung nodule malignancy assessment using ct images. *Applied Sciences*, 10(21):7837. 35, 54

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., e Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958. 15

Tabarcea, A., Rosca, V., e Iftene, A. (2019). Imageclefmed tuberculosis 2019: Predicting ct scans severity scores using stage-wise boosting in low-resource environments. Em *CLEF (Working Notes)*. 37

Tan, M. e Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. Em *International conference on machine learning*, páginas 6105–6114. PMLR. 15

Thompson, N. C., Greenewald, K., Lee, K., e Manso, G. F. (2020). The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*. 14

Tran, D., Bourdev, L., Fergus, R., Torresani, L., e Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. Em *Proceedings of the IEEE international conference on computer vision*, páginas 4489–4497. 54

Valadão, M., Silva, L., Serrão, M., Guerreiro, W., Furtado, V., Freire, N., Monteiro, G., e Craveiro, C. (2023). Mobilenetv3-based automatic modulation recognition for low-latency spectrum sensing. Em *2023 IEEE International Conference on Consumer Electronics (ICCE)*, páginas 1–5. IEEE. 36

Van der Maaten, L. e Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11). 32

Yang, J., Huang, X., He, Y., Xu, J., Yang, C., Xu, G., e Ni, B. (2021). Reinventing 2d convolutions for 3d images. *IEEE Journal of Biomedical and Health Informatics*. 27

Yue, Z., Ma, L., e Zhang, R. (2020). Comparison and validation of deep learning models for the diagnosis of pneumonia. *Computational Intelligence and Neuroscience*, 2020. 14

Zhu, H., Liu, X., Mao, X., e Wong, T.-T. (2017). Real-time deep video deinterlacing. *arXiv preprint arXiv:1708.00187*. 26

Zhu, W., Liu, C., Fan, W., e Xie, X. (2018). Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. Em *2018 IEEE winter conference on applications of computer vision (WACV)*, páginas 673–681. IEEE. 37

Zunair, H., Rahman, A., e Mohammed, N. (2019). Estimating severity from ct scans of tuberculosis patients using 3d convolutional nets and slice selection. Em *CLEF (Working Notes)*. 24

Zunair, H., Rahman, A., Mohammed, N., e Cohen, J. P. (2020). Uniformizing techniques to process ct scans with 3d cnns for tuberculosis prediction. Em *International Workshop on PRedictive Intelligence In MEdicine*, páginas 156–168. Springer. 25