

CENTRO UNIVERSITÁRIO FEI  
DIOGO FELIPE DE MELO SANTIAGO

**ANÁLISE DE IMAGENS DE RESSONÂNCIA MAGNÉTICA PARA CLASSIFICAÇÃO  
DE CARDIOMIOPATIA COM O APOIO DE MECANISMOS DE ATENÇÃO**

São Bernardo do Campo  
2025

DIOGO FELIPE DE MELO SANTIAGO

**ANÁLISE DE IMAGENS DE RESSONÂNCIA MAGNÉTICA PARA CLASSIFICAÇÃO  
DE CARDIOMIOPATIA COM O APOIO DE MECANISMOS DE ATENÇÃO**

Dissertação apresentada ao Centro Universitário FEI,  
como parte dos requisitos para a obtenção do título  
de Mestre em 2025. Orientado(a) pela Profa. Dra.  
Leila Cristina Carneiro Bergamasco

São Bernardo do Campo  
2025

Santiago, Diogo Felipe de Melo.

Análise de imagens de ressonância magnética para classificação de cardiomiopatia com o apoio de mecanismos de atenção / Diogo Felipe de Melo Santiago. São Bernardo do Camo, 2025.

68 f. : il.

Dissertação - Centro Universitário FEI.

Orientadora: Prof.<sup>a</sup> Dra. Leila Cristina Carneiro Bergamasco.

1. Radiomics. 2. Mecanismo de Atenção. 3. Transformers  
Cardiomiopatia. I. Bergamasco, Leila Cristina Carneiro, orient. II. Título.



# APRESENTAÇÃO DE DISSERTAÇÃO ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Mestrado

PGE-10

**Aluno(a):** Diogo Felipe de Melo Santiago

**Matrícula:** 123.114-1

**Título do Trabalho:** ANÁLISE DE IMAGENS DE RESSONÂNCIA MAGNÉTICA PARA CLASSIFICAÇÃO DE CARDIOMIOPATIA COM O APOIO DE MECANISMOS DE ATENÇÃO

**Área de Concentração:** Processamento de Sinais e Imagens

**Orientador(a):** Profª Drª Leila Cristina Carneiro Bergamasco

**Data da realização da defesa:** 28/03/2025

**ORIGINAL ASSINADA**

## Avaliação da Banca Examinadora:

A dissertação foi avaliada e aprovada pela banca examinadora por unanimidade, considerando sua relevância científica e contribuição para a área de estudo. O trabalho apresenta uma abordagem consistente, com fundamentação teórica adequada e uma metodologia bem estruturada, permitindo a obtenção de resultados significativos. Dessa forma, atestamos que a dissertação atende aos requisitos exigidos para a obtenção do título de Mestre

A Banca Julgadora abaixo-assinada atribuiu ao aluno o seguinte resultado:

APROVADO

REPROVADO

## MEMBROS DA BANCA EXAMINADORA

Profª Drª Leila Cristina Carneiro Bergamasco

Prof. Dr. Rafael Luiz Testa

Prof. Dr. Cléber Gimenez Corrêa

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

## **RESUMO**

A crescente disponibilidade de exames de imagem médica, como a ressonância magnética, gera um grande volume de dados, tornando sua análise complexa e desafiadora. Neste cenário, abordagens computacionais avançadas podem otimizar a interpretação dessas imagens e auxiliar no diagnóstico precoce de doenças cardiovasculares. Este trabalho consiste em unificar abordagens contemporâneas na avaliação da cardiomiopatia. Com o apoio da análise radiômica, a qual extrai informações das características estatísticas e de textura de uma imagem médica, e de características oriundas de uma rede neural clássica para visão computacional, como a *ResNet50*, é possível obter resultados promissores. Os resultados certificam que a união de informações de diversos âmbitos, acerca de um dado paciente, quando aliadas, podem culminar em resultados mais interessantes se comparados aos dados de forma isolada. O presente trabalho visa usar as abordagens citadas, baseado em literaturas prévias, efetuando uma aplicação inédita para o teste de cardiomiopatia, adaptando e propondo uma arquitetura mais robusta de forma a obter bons resultados.

**Palavras-chave:** Radiomics; Mecanismo de Atenção; Transformers; Cardiomiopatia.

## **ABSTRACT**

The increasing availability of medical imaging exams, such as magnetic resonance imaging, generates a large volume of data, making its analysis complex and challenging. In this scenario, advanced computational approaches can optimize the interpretation of these images and assist in the early diagnosis of cardiovascular diseases. This work aims to unify contemporary approaches in the evaluation of cardiomyopathy. With the support of radiomic analysis, which extracts information from the statistical and texture characteristics of a medical image, and features derived from a classical neural network for computer vision, such as ResNet50, promising results can be obtained. The results confirm that the combination of information from various domains regarding a given patient, when integrated, can lead to more interesting outcomes compared to analyzing data in isolation. This study aims to apply the aforementioned approaches, based on previous literature, in an innovative application for cardiomyopathy testing, adapting and proposing a more robust architecture to achieve better results.

**Keywords:** Radiomics; Attention Mechanism; Transformers; Cardiomyopathy.

## LISTA DE FIGURAS

Figura 1 – Exemplos de aplicação da análise radiômica na identificação de tumores. CE = <i>Contrast Enhanced</i> , HH = <i>high-high</i> (filtro passa-alta).....	17
Figura 2 – Aplicação de Vizinhos em Abordagens Radiômicas .....	18
Figura 3 – Representação visual da camada convolucional. O elemento central do <i>kernel</i> é aplicado no vetor de entrada, que é calculado e substituído pela ponderada dele mesmo e de quaisquer pixels próximos. ....	20
Figura 4 – Conexão de Salto da Arquitetura <i>ResNet</i> .....	22
Figura 5 – Fluxo do Bloco SE - Após blocos convolucionais normais, uma cama da de <i>squeeze</i> seguida e uma camada de <i>excitation</i> é aplicada e por fim os valores originais são escalonados pelo resultado. ....	24
Figura 6 – Módulo SE-Resnet, que estende o bloco ResNet tradicional com um bloco SE para recalibrar os canais da saída residual com base em informações globais, aprimorando a representação de características relevantes. ....	25
Figura 7 – Desempenho da SE-ResNet-50 com diferentes valores do fator de redução $r$ , demonstrando variação nos erros top-1 e top-5 e no número de parâmetros, em comparação com a arquitetura ResNet-50 original. ....	26
Figura 8 – Arquitetura <i>Transformers</i> - Modelo composto por codificador e decodifica- dor com camadas de atenção multi-cabeça, normalização, <i>feed-forward</i> e codificação posicional. ....	28
Figura 9 – Módulo de Atenção Multi-Cabeças .....	29
Figura 10 – Funil da Seleção da Literatura - Etapas do processo de busca, filtragem e seleção de artigos com base em critérios de inclusão, exclusão e qualidade, resultando no conjunto final de estudos analisados. ....	32
Figura 11 – Modelo de arquitetura de rede neural proposto por Jiang <i>et al.</i> (2021). A imagem ilustra: (A) o fluxo principal da rede; (B) um módulo de atenção de canal; e (C) um módulo de atenção combinada (canal e espacial). ....	33
Figura 12 – Arquitetura Proposta por Ai <i>et al.</i> (2023), que emprega um Módulo de Fusão Baseado em Autoatenção (Self-attention based Fusion Module). O modelo integra características radiômicas, extraídas diretamente das imagens, com características de aprendizado profundo de uma rede ResNet50 pré-treinada. Ambas as vertentes de características passam por um processo de seleção (F- test) antes de serem concatenadas e processadas pelo módulo de autoatenção para a classificação final.....	35
Figura 13 – Método proposto. <i>DenseNet</i> é o modelo base, a saída da rede densa é alimen- tada pela rede de atenção que prioriza as fatias e regiões e gera a classificação binária do status de metilação do MGMT. ....	36

Figura 14 – Arquitetura Proposta por Iranmehr, Gutta e Hadaegh (2022) para análise de mapas de características convolucionais, que inclui bloco SE e autoatenção, segmentação de regiões integrais (representadas pelas áreas em verde e vermelho), além de uma camada LSTM. O modelo também utiliza diferentes tipos de atenção para focar em regiões específicas dos dados. ....	37
Figura 15 – Fluxograma - Pré-Processamento e Arquitetura .....	40
Figura 16 – Fluxograma - Continuação.....	41
Figura 17 – Imagem médica da classe DCM e sua respectiva máscara .....	44
Figura 18 – Imagem médica da classe HCM e sua respectiva máscara .....	44
Figura 19 – Imagem médica da classe NOR e sua respectiva máscara.....	45
Figura 20 – Arquitetura proposta para classificação de cardiomiopatia utilizando características radiômicas e profundas. As características são processadas por testes-F, transformações lineares e concatenadas antes de passar por blocos convolucionais, caso a concatenação seja na dimensão da profundidade, e um módulo de autoatenção. ....	50
Figura 21 – Representação esquemática da recorrência do Módulo de Autoatenção. O diagrama ilustra o processamento recursivo na arquitetura proposta, recursividade esta que permite que as características aprendidas sejam refinadas ao longo de múltiplas iterações, aumentando a capacidade do modelo de capturar dependências complexas nos dados. ....	53
Figura 22 – Captura Diastólica CMD .....	54
Figura 23 – Captura Diastólica de CMH .....	54
Figura 24 – Captura Diastólica NOR .....	55
Figura 25 – Matriz de confusão do modelo base, apresentando os resultados da classificação binária. Observa-se 20 verdadeiros negativos, 9 verdadeiros positivos, 10 falsos positivos e 11 falsos negativos, totalizando 50 amostras com acurácia de 58%. ....	56
Figura 26 – Curva ROC obtida pelo modelo proposto, ilustrando a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos. ....	56
Figura 27 – Painéis adaptáveis gerados pela plataforma <i>CometML</i> durante o treinamento do modelo. A visualização superior apresenta a evolução da função de perda (loss) ao longo dos passos de treinamento, com picos iniciais que se estabilizam no decorrer do treino. Os painéis inferiores mostram as métricas finais obtidas: acurácia de treinamento de 100% e acurácia de teste de 58% (valor 0.58)., evidenciando um possível sobreajuste do modelo. ....	57
Figura 28 – Valores Coletados no Treino gerados pela plataforma <i>CometML</i> incluindo: função de perda (loss), acurácia de treino, e teste.....	57

Figura 29 – Composição Bloco SE. O diagrama ilustra o fluxo de processamento em três etapas principais: Squeeze (compressão), que reduz a dimensionalidade espacial através de Global Pooling; Excitation (excitação), composta por duas camadas totalmente conectadas (FC) com funções de ativação ReLU e Sigmoid; e Scaling (escalonamento), que recalibra os mapas de características originais. Este mecanismo permite que a rede aprenda interdependências entre canais e atribua pesos adaptativos para realçar características relevantes. ..... 58

## LISTA DE TABELAS

Tabela 1 – Resultados dos Artigos.....	30
Tabela 2 – Critérios de Inclusão e Exclusão .....	31
Tabela 3 – Questões de Aceitação .....	31
Tabela 4 – Classes ACDC -Descrição .....	43
Tabela 5 – Classes do ACDC .....	44
Tabela 6 – Estatísticas dos volumes e função do ventrículo esquerdo (média).....	46
Tabela 7 – Fonte: Componentes Utilizados .....	51
Tabela 8 – Métricas do Experimento - Modelo Base .....	55
Tabela 9 – Métricas ACDC - Adaptação do Modelo Original Negrito representa o modelo base .....	59
Tabela 10 – Métricas ACDC - Adaptação do Modelo Original + Máscaras Negrito representa maior assertividade.....	60
Tabela 11 – Métricas ACDC - Modelos Adaptados - Blocos Conv. + SE Negrito representa maior assertividade.....	60
Tabela 12 – Métricas SunnyBrook - Adaptação do Modelo Original Negrito representa o modelo base .....	61
Tabela 13 – Métricas SunnyBrook - Adaptação do Modelo Original Com Máscaras Negrito representa maior assertividade.....	61
Tabela 14 – Métricas SunnyBrook - Adaptação Adicionando Blocos Conv. e SE Negrito representa maior assertividade.....	62
Tabela 15 – Resultados dos testes estatísticos para acurácia dos modelos .....	63

## LISTA DE ABREVIATURAS E SIGLAS

ACDC	<i>Automated Cardiac Diagnosis Challenge</i>
Adam	<i>Adaptive Moment Estimation</i>
API	<i>Application Programming Interface</i>
AP	Aprendizado Profundo
AUC	<i>Area Under the ROC Curve</i>
BCE	<i>Binary Cross Entropy</i>
CBAM	<i>Convolutional Block Attention Module</i>
CAR	Cardiomiopatia
CMD	Cardiomiopatia Dilatada
CMH	Cardiomiopatia Hipertrófica
CNN	Rede Neural Convolucional
DCM	<i>Dilated Cardiomyopathy</i>
FE	<i>Fração de Ejeção</i>
FEI	Fundação Educacional Inaciana
GBM	<i>Glioblastoma Multifoma</i>
GLCM	<i>Gray-Level Cooccurrence Matrix</i>
HCM	<i>Hypertrrophic Cardiomyopathy</i>
IA	Inteligência Artificial
InCor	Instituto do Coração do Hospital das Clínicas da FMUSP
LLM	<i>Large Language Models</i>
LoG	<i>Laplace of Gaussian</i>
LSTM	<i>Long Short-Term Memory</i>
MINF	<i>Myocardial Infarction</i>
NOR	Normal

PET	Tomografia por Emissão de Pósitrons
RMC	Ressonância Magnética Cardiovascular
ROI	<i>Region of Interest</i>
ROC	<i>Receiver Operating Characteristic Curve</i>
RNP	Redes Neurais Profundas
RV	<i>Right Ventricular Abnormality</i>
SE	<i>Squeeze and Excitation</i>
TCIA	<i>Cancer Imaging Archive</i>
TC	Tomografia Computadorizada

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	12
1.1	OBJETIVOS .....	13
1.2	ESTRUTURA DO TRABALHO .....	14
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	15
2.1	RESSONÂNCIA MAGNÉTICA EM CARDIOMIOPATIA .....	15
2.2	ANÁLISE RADIÔMICA .....	16
<b>2.2.1</b>	<b>Características por Histograma .....</b>	16
<b>2.2.2</b>	<b>Características de Textura .....</b>	16
<b>2.2.3</b>	<b>Características de Modelo .....</b>	18
<b>2.2.4</b>	<b>Características de Transformação.....</b>	19
2.3	REDES NEURAIS CONVOLUCIONAIS .....	19
2.4	RESNET .....	21
<b>2.4.1</b>	<b>Extração de Características Profundas .....</b>	22
2.5	REDE SQUEEZE AND EXCITATION .....	23
<b>2.5.1</b>	<b>Squeeze: Informação Global .....</b>	23
<b>2.5.2</b>	<b>Excitation: Recalibração Adaptativa .....</b>	24
<b>2.5.3</b>	<b>Utilização em redes conhecidas: SE-Resnet .....</b>	25
2.6	ARQUITETURA TRANSFORMERS .....	27
2.7	OTIMIZADOR ADAM .....	29
<b>3</b>	<b>TRABALHOS RELACIONADOS .....</b>	30
3.1	REVISÃO SISTEMÁTICA DA LITERATURA .....	30
3.2	ANÁLISE RADIÔMICA COM ATENÇÃO .....	32
3.3	CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	38
<b>4</b>	<b>TRABALHOS RELACIONADOS .....</b>	39
4.1	BASES DE DADOS .....	41
<b>4.1.1</b>	<b>Conjunto de Dados ACDC .....</b>	42
<b>4.1.2</b>	<b>Conjunto de Dados Sunny .....</b>	45
4.2	PRÉ-PROCESSAMENTO .....	46
4.3	EXTRAÇÃO DE CARACTERÍSTICAS RADIÔMICAS .....	46
4.4	EXTRAÇÃO DE CARACTERÍSTICAS PROFUNDAS .....	47
4.5	CONCATENAÇÃO .....	47
4.6	MÓDULO DE ATENÇÃO SELETIVA .....	47
4.7	MÓDULO DE AUTOATENÇÃO .....	48

4.8	MÉTRICAS .....	49
4.9	ARQUITETURA PROPOSTA .....	49
4.10	CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	50
<b>5</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>51</b>
5.1	MATERIAIS .....	51
5.2	CENÁRIOS DE TESTE .....	52
5.3	EXPERIMENTOS MODELO BASE .....	52
5.4	EXPERIMENTOS MODELOS PROPOSTO .....	53
5.5	RESULTADOS ACDC .....	54
5.5.1	<b>Resultados dos Modelos Base - ACDC .....</b>	<b>55</b>
5.5.2	<b>Resultados dos Modelos Adaptados - ACDC .....</b>	<b>57</b>
5.5.3	<b>Resultados dos Modelos Adaptados - <i>SunnyBrook</i> .....</b>	<b>59</b>
5.5.4	<b>Testes Estatísticos .....</b>	<b>62</b>
5.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	63
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>64</b>
	<b>REFERÊNCIAS .....</b>	<b>66</b>

## 1 INTRODUÇÃO

A tecnologia está cada vez mais presente nas diversas áreas do conhecimento, trazendo uma infinidade de benefícios e facilitando o cotidiano contemporâneo. Entre os diversos campos impactados, a área médica destaca-se como uma das que mais se beneficiaram da inovação tecnológica. Desde meados dos anos 2000, a quantidade de dados gerados na medicina tem crescido exponencialmente, atingindo projeções de milhares de *exabytes* a partir de 2020 (Gantz; Reinsel, 2012). Esse cenário evidencia a importância de ferramentas que possam processar e analisar eficientemente grandes volumes de informações.

Exames de imagem, como a tomografia computadorizada (TC) e a ressonância magnética cardiovascular (RMC), tornaram-se essenciais na medicina moderna. Esses exames não apenas oferecem uma representação tridimensional detalhada de estruturas do corpo humano, mas também produzem dados que podem ser analisados de forma quantitativa. Em paralelo, a inteligência artificial (IA) trouxe avanços significativos à análise de imagens diagnósticas, proporcionando maior eficiência e precisão nos diagnósticos médicos (Argentiero *et al.*, 2022).

Nesse sentido, as redes neurais profundas, um dos principais pilares da IA, têm demonstrado alto desempenho em tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação. Essas redes conseguem aprender características discriminantes uma vez otimizadas a cerca do conjunto de dados em que foi treinada. Além disso, arquiteturas como o *transformers* ficaram populares por serem comumente usadas em redes gerativas auto-regressivas para geração sintética de texto, também conhecidas como *large language models* (LLM), tendo como seu exemplo mais conhecido o *ChatGPT*. Os *transformers* são arquiteturas que destacam-se pela capacidade de paralelismo e pelo uso de mecanismos de autoatenção, que permitem ao modelo focar nas partes mais relevantes dos dados de entrada (Russell; Norvig, 2020).

Adicionalmente, técnicas de processamento de imagem como a análise de textura já vem sendo utilizada por várias décadas em diversos domínios da medicina. A análise radiométrica emergiu como uma ferramenta poderosa na extração de informações quantitativas de imagens médicas, capturando padrões que muitas vezes passam despercebidos ao olho humano. Essa abordagem tem mostrado potencial em diversas áreas, como oncologia e cardiologia (Schofield *et al.*, 2019).

No domínio da análise de imagens cardíacas, a análise de textura aplicada à RMC tem avaliado o risco de arritmia pós-infarto do miocárdio. O uso de análise de textura para RMC em imagens sem contraste e com realce tardio de gadolínio em pacientes com cardiomiopatia para prever o resultado do exame é uma área particular de interesse (Schofield *et al.*, 2019).

A cardiomiopatia hipertrófica (CMH) é uma das cardiomiopatias mais comuns, frequentemente diagnosticada em jovens e indivíduos de meia-idade. Embora em muitos casos seja assintomática, a doença pode levar a condições graves, como insuficiência cardíaca e acidente vascular cerebral. Isso torna o diagnóstico precoce essencial para prevenir desfechos adversos

(Kwon *et al.*, 2022). Neste sentido a análise radiômica que consiste em extrair dados qualitativos de imagens médicas, incluindo, em muitos casos, a análise da textura dessas imagens (Lambin *et al.*, 2012). A análise radiômica pode auxiliar com o diagnóstico prévio afim de compreender e atuar nos casos de cardiomiopatia que demonstrem risco ao paciente.

Neste cenário, combinar técnicas de IA e análise radiômica representa uma estratégia promissora para a detecção de cardiomiopatias entre outras condições cardíacas. Estudos recentes, como o de Ai *et al.* (2023), demonstraram que a integração de características profundas e radiômicas podem melhorar significativamente o desempenho preditivo de modelos diagnósticos de câncer de pulmão via imagens de TC. Esses modelos baseiam-se em mecanismos de autoatenção para identificar padrões relevantes em dados concatenados, alcançando acurácia de até 82,35% e *area under the ROC curve* (AUC) de 0,74.

Assim, o presente trabalho propõe a implementação e validação de uma estratégia de fusão que combine características radiômicas e profundas, com o uso de mecanismos de autoatenção, para melhorar a classificação de cardiomiopatias. Além de avançar o estado da arte em diagnósticos médicos, esta pesquisa busca contribuir para a adoção de soluções mais eficazes e acessíveis no apoio à decisão clínica.

## 1.1 OBJETIVOS

Os objetivos deste trabalho consistem em propor e validar uma abordagem inovadora para a classificação de cardiomiopatias utilizando técnicas de inteligência artificial e análise radiômica. Estes objetivos foram definidos para responder às necessidades clínicas e avançar o estado da arte na área de diagnóstico médico. Com isso, o objetivo geral do presente projeto é desenvolver, implementar e validar um modelo de classificação de cardiomiopatias, utilizando a integração de características radiômicas e profundas mediadas por mecanismos de autoatenção. Os objetivos específicos, por sua vez, são elencados conforme os itens abaixo:

1. Identificar e extrair características radiômicas de imagens de RMC, capturando padrões texturais e estatísticos relevantes para a classificação das cardiomiopatias.
2. Projetar um pipeline de análise que combine eficientemente características radiômicas e profundas, garantindo a fusão informativa para modelos de aprendizado profundo.
3. Implementar um modelo baseado em redes neurais profundas que utilize mecanismos de atenção para priorizar regiões relevantes nas imagens, melhorando a acurácia e a interpretabilidade do modelo.
4. Validar a eficácia do modelo proposto utilizando métricas padrão, como acurácia, precisão, revocação, F1-score e AUC, em um conjunto de dados públicos e diversificado.
5. Comparar o desempenho do modelo proposto com técnicas existentes na literatura, identificando avanços e limitações em relação às abordagens tradicionais.

## 1.2 ESTRUTURA DO TRABALHO

Este trabalho está organizado em sete capítulos, cada um projetado para abordar diferentes aspectos da pesquisa e sua implementação. A seguir, apresenta-se a estrutura detalhada desta pesquisa:

O **Capítulo 1** apresenta o contexto geral do trabalho, destacando a relevância do uso de inteligência artificial e análise radiômica no diagnóstico de cardiomiopatias. São expostos os objetivos gerais e específicos da pesquisa, bem como a motivação para o desenvolvimento deste estudo. O **Capítulo 2** discute os conceitos teóricos que embasam o trabalho, incluindo princípios de redes neurais profundas, análise radiômica, mecanismos de autoatenção e sua aplicação em imagens médicas. Este capítulo também revisa estudos relacionados que contribuíram para o estado da arte na área. No **Capítulo 3** é estudado os trabalhos recentes, não superior a cinco anos passados, de outros autores que estão de alguma forma relacionados com tema, sendo estes trabalhos relacionados oriundos de uma minuciosa revisão sistemática da literatura. O **Capítulo 4** detalha o método proposto para a classificação de cardiomiopatias, incluindo a descrição do conjunto de dados utilizado, o processo de extração de características radiômicas, o desenvolvimento do modelo baseado em redes neurais profundas e as etapas de validação e experimentação. O **Capítulo 5** faz a proposição experimental que contempla os dados utilizados, informações do seu pré-processamento, hiperparâmetros planejados e demais informações a cerca dos experimentos realizados. Neste capítulo também é apresentado os resultados obtidos com a implementação do modelo proposto, acompanhados de análises quantitativas e qualitativas. Discute-se o desempenho do modelo em comparação com outras abordagens e suas implicações técnicas. Finalmente, o **Capítulo 6** sintetiza os principais achados da pesquisa, destacando as contribuições do trabalho para o estado da arte e suas potenciais aplicações. Além disso, são sugeridas direções para estudos futuros que possam expandir e aprimorar as técnicas apresentadas.

## 2 FUNDAMENTAÇÃO TEÓRICA

A proposta deste trabalho visa unificar informações das características radiômicas e características profundas oriundas de uma rede neural convolucional e, por conseguinte, propor uma arquitetura de autoatenção para identificar características discriminantes a fim de obter resultados relevantes na classificação de cardiomiopatias.

### 2.1 RESSONÂNCIA MAGNÉTICA EM CARDIOMIOPATIA

A RMC se tornou um dos principais métodos em imagem médica para avaliar a fisiologia e a função na doença cardíaca congênita. A RMC possui funções que normalmente são obtidas por ecocardiografia, tendo como exemplo a velocidade média em um vaso, mas também apresenta características muito especiais como calcular a velocidade em um voxel de um milímetro em qualquer lugar no espaço tridimensional do vaso. Usando a tecnologia atual de RMC, os investigadores podem obter percepções únicas sobre a função ventricular, por exemplo, deformação ventricular regional *in vivo*, movimento da parede e mecânica dos fluidos, por exemplo, visualização *in vivo* de perfis de velocidade além de aumentar a precisão de medidas padrão aceitas pela fisiologia ou função ventricular, por exemplo, débito cardíaco, volumes ventriculares e massa. Devido ao rápido avanço da tecnologia, muitas técnicas de RMC são experimentais ou ainda estão em fase de desenvolvimento clínico; no entanto, há muitas técnicas que são clinicamente úteis, por exemplo, medição do volume ventricular em pacientes com tamanho ventricular esquerdo limítrofe. Como muitas das técnicas experimentais atuais serão, sem dúvida, empregadas na prática clínica no futuro, os médicos devem estar cientes de todo o espectro de capacidades da RMC (Fogel, 2000).

Em muitos cenários clínicos, as limitações técnicas da ecocardiografia e a expressão fenotípica heterogênea tornaram tal avaliação difícil e a RMC emergiu como uma modalidade de imagem complementar útil para incorporar a ecocardiografia transtorácica de rotina. A RMC cardíaca é única em sua alta resolução espacial e temporal com excelente contraste entre o *pool* sanguíneo e o miocárdio, sem limitação de janela de imagem ou plano de imagem. A heterogeneidade fenotípica da miocardiopatia hipertrófica é bem reconhecida pela dificuldade em sua identificação. Isto se complica ainda mais pois nem todos os pacientes com hipertrófia ventricular esquerda têm CMH, enquanto uma fisiopatologia semelhante à CMH com obstrução dinâmica do trato de saída do ventrículo esquerdo pode ser observada sem hipertrófia do ventrículo esquerdo, em um subgrupo de pacientes com anormalidades na válvula mitral e/ou no músculo papilar (Pontone *et al.*, 2022).

A CMH é uma doença heterogênea com expressão morfológica complexa que requer uma caracterização precisa da doença para um planejamento terapêutico ótimo e estratificação de risco. A RMC cardíaca emergiu como um complemento útil para esses propósitos. Com a incorporação crescente de imagem multimodal na avaliação clínica da CMH, o entendimento sobre a importância de diferenças morfológicas sutis continuará a crescer, e pesquisas adicionais

definirão novos marcadores prognósticos e melhorarão as estratégias de tratamentos atuais (To; Dhillon; Desai, 2011).

## 2.2 ANÁLISE RADIÔMICA

A análise radiômica é um campo de pesquisa em rápida evolução, preocupada com a extração de informações quantitativas, incluindo padrões complexos que são difíceis de reconhecer ou quantificar pelo olho humano a cerca de imagens médicas, características estas chamadas de características radiômicas. As características radiômicas podem capturar características de tecidos e lesões, como forma e heterogeneidade, e podem, sozinhas ou em combinação com dados demográficos, histológicos, genômicos ou proteômicos, ser usadas para a resolução de problemas clínicos.

Mesmo que características radiômicas individuais podem se correlacionar com dados genômicos ou dados clínicos, o impacto da análise radiômica se torna mais relevante a medida que mais informações são por ela extraídas. Tipicamente centenas de características, uma fração do total que é extraído, contribuem para identificação de uma doença específica e é processada usando técnicas de aprendizado de máquina. As características radiômicas podem ser subdivididas em estatísticas, incluindo as baseadas em histograma e textura; modelo; transformação; e forma. A extração pode ser dada em regiões de interesse de 2-dimensões (2D) e 3-dimensões (3D) (Mayerhoefer *et al.*, 2020).

### 2.2.1 Características por Histograma

Os descritores estatísticos mais simples são baseados no histograma global de níveis de cinza e incluem média de nível de cinza, máximo, mínimo, variância e percentis. Como essas características são baseadas em análises de *pixel* único ou *voxel* único (3D), elas são chamadas de características de primeira ordem. Algumas características mais sofisticadas incluem assimetria e curtose que descrevem a forma da distribuição da intensidade dos dados: a assimetria reflete a assimetria da curva da distribuição de dados para a esquerda (assimetria negativa, abaixo da média) ou direita (assimetria positiva, acima da média), enquanto a curtose reflete a caudalidade de uma distribuição de dados em relação a uma distribuição gaussiana devido a *outliers*. Outras características incluem histograma entrópico e uniformidade, e também é comumente chamado de energia.

### 2.2.2 Características de Textura

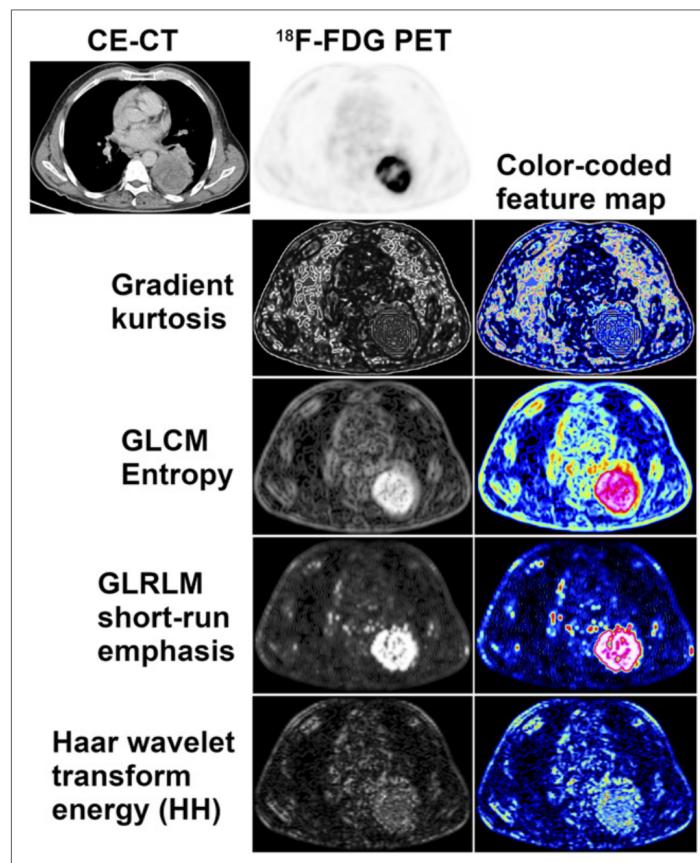
Uma abordagem simples para a descrição de textura radiômica é a análise do gradiente absoluto, que reflete o grau ou a abruptidate das flutuações de intensidade de nível de cinza em uma imagem. Para dois *pixels* ou *voxels* adjacentes, o gradiente é o máximo possível quando um for preto e o outro branco, enquanto se ambos os *pixels* forem pretos (ou ambos forem brancos)

o gradiente nessa localidade é zero. Similares aos histogramas, as características por gradiente incluem média, variância, assimetria e curtose.

Na Figura 1 são apresentadas imagens de TC e tomografia por emissão de pósitrons (PET) com contraste de um câncer de pulmão parcialmente necrótico no lobo inferior esquerdo. Os mapas de características radiômicas são gerados ao mover uma pequena janela retangular sobre a imagem de TC, calculando o valor da característica em cada posição. Esses mapas refletem diferentes aspectos da heterogeneidade do metabolismo da glicose no tumor. Cada mapa representa uma única característica radiônica, onde valores altos indicam intensidades elevadas no mapa de níveis de cinza.

Figura 1 – Exemplos de aplicação da análise radiônica na identificação de tumores.

CE = *Contrast Enhanced*, HH = *high-high* (filtro passa-alta)



Autor: (Mayerhoefer *et al.*, 2020)

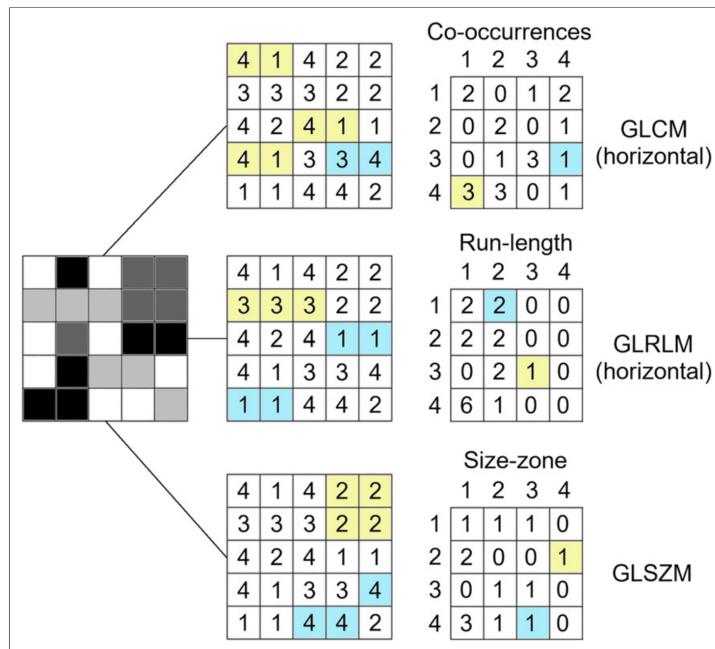
A matriz de coocorrência de níveis de cinza, do termo *gray-level cooccurrence matrix* (GLCM), é um histograma em níveis de cinza de segunda ordem. O GLCM captura relações espaciais entre pares de *pixels* ou *voxels* com níveis de cinza pré-definidos, em diferentes direções (horizontal, vertical ou diagonal para análise 2D, 13 direções para 3D) e com uma distância pré-definida entre os *pixels* ou *voxels*. As características de GLCM incluem entropia, uma medida composta pela medida da inhomogeneidade ou aleatoriedade dos níveis de cinza, momento angular de segunda ordem (também chamado de uniformidade ou energia), que reflete a homogeneidade

ou ordem dos níveis de cinza e contraste, que enfatiza as diferenças nos níveis de cinza entre *pixels* ou *voxels* pertencentes a um par de *pixel* ou *voxel*. A Figura 2 ilustra um matriz que atua como um contador dos pares de *pixel* ou *voxel*.

A matriz de comprimento de corrida de níveis de cinza, do termo *Gray-Level Run-length Matrix* (GLRLM), fornece informações sobre a distribuição espacial das sequências de *pixels* consecutivos com os mesmos níveis de cinza, em uma ou mais direções, em 2 ou 3 dimensões. As características da GLRLM incluem fração, que avalia a porcentagem de *pixels* ou *voxels* dentro da região de interesse que fazem parte das sequências e, portanto, reflete a granulosidade.

A matriz de zona do tamanho dos níveis de cinza, do termo *Gray-Level Size Zone Matrix* (GLSZM), é baseada no mesmo princípio do GLRLM porém aqui, contagens do número de grupos (chamados zonas) de *pixels* ou *voxels* vizinhos interconectados com o mesmo nível de cinza formam a base para a matriz, como visto na Figura 2. Uma textura mais homogênea resultará em uma matriz mais ampla e plana. A GLSZM não é calculada para diferentes direções, mas pode ser calculada para diferentes distâncias de *pixels* ou *voxels* que definem a vizinhança. As características da GLSZM podem ser calculadas em 2 dimensões (8 *pixels* vizinhos) ou 3 dimensões (26 *voxels* vizinhos) e, seguindo as definições da GLRLM, incluem fração (porcentagem de *pixels* ou *voxels* que fazem parte das zonas), ênfase em zonas pequenas e grandes, entre outras.

Figura 2 – Aplicação de Vizinhos em Abordagens Radiométricas



Autor: (Mayerhoefer *et al.*, 2020)

### 2.2.3 Características de Modelo

As análises baseadas em modelos visam interpretar características em níveis de cinza espaciais para caracterizar objetos ou formas. Um modelo parametrizado de geração de textura é

calculado e ajustado à região de interesse, do termo *region of interest* (ROI), e seus parâmetros estimados são utilizados como características radiômicas. O modelo autorregressivo é um exemplo de abordagem baseada em modelo e baseia-se na ideia de que o nível de cinza de um *pixel* é uma soma ponderada dos níveis de cinza de 4 *pixels* vizinhos. Além disso, o  $\sigma$ , o qual carrega informações sobre a variância do erro de previsão mínimo, mede a regularidade da textura. A análise fractal também produz recursos que podem ser usados na análise radiônica, em particular a dimensão fractal, que reflete a taxa de adição de detalhe estrutural com o aumento da magnificação, escala ou resolução e, portanto, serve como uma medida de complexidade. A lacunaridade, um recurso que mede a falta de invariância rotacional ou translacional, reflete a inhomogeneidade (Mayerhoefer *et al.*, 2020).

#### 2.2.4 Características de Transformação

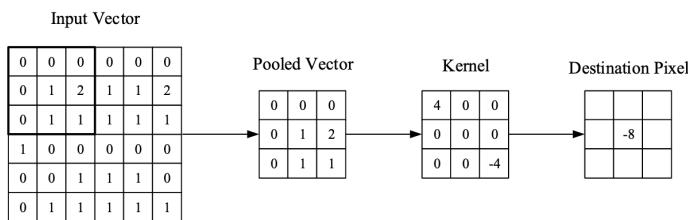
Métodos baseados em transformadas, incluindo transformadas de *Fourier*, *Gabor* e de *wavelets* de *Haar*, analisam padrões de níveis de cinza em um espaço diferente. A transformada discreta *wavelet* de *Haar*, por exemplo, analisa o conteúdo da frequência de uma imagem em diferentes escalas. A decomposição por *wavelets* de uma imagem é possível aplicando um par de filtros espelhados em quadratura, um filtro de passa-alta e um de passa-baixa. Embora o filtro de passa-alta destaque as mudanças no nível de cinza e, assim, enfatize detalhes da imagem, o filtro de passa-baixa suaviza a imagem em termos de nível de cinza, removendo detalhes da imagem. Após a decomposição do sinal, um conjunto de canais de frequência orientados espacialmente está disponível, o qual é usado para descrever a variabilidade local da imagem. As energias dentro dos canais de frequência são então usadas como características. A filtragem de passa-alta em ambas as direções, parte inferior da Figura 1, captura detalhes diagonais, a filtragem de passa-alta seguida por filtragem de passa-baixa captura bordas verticais, a filtragem de passa-baixa seguida da filtragem de passa-alta captura bordas horizontais, e a filtragem de passa-baixa em ambas as direções captura as frequências mais baixas, em diferentes escalas. Notavelmente, a transformação por *wavelets* pode ser usada não apenas para a geração de características radiômicas, mas também para segmentação de imagens ou como um passo de pré-processamento para análise de textura (Mayerhoefer *et al.*, 2020).

### 2.3 REDES NEURAIS CONVOLUCIONAIS

Os autores LeCun *et al.* (1989) pesquisaram sobre o reconhecimento de caracteres em documentos escritos à mão, criando a arquitetura LeNet, uma rede que consiste em camadas convolucionais seguidas por camadas de *pooling*, e finalizando com camadas totalmente conectadas. Os autores introduziram o *backpropagation* a redes convolucionais, permitindo o aprendizado automático e substituindo o uso de técnicas para escolha manual de coeficientes. LeCun *et al.* (1989), ainda estudando a classificação de caracteres, expandem a LeNet para a LeNet-5, uma rede que firmou a superioridade das redes convolucionais na tarefa de classificação de imagens se comparado a outros métodos da época.

Como o nome sugere, a camada convolucional desempenha um papel vital no funcionamento da rede neural convolucional (CNN). Os parâmetros das camadas concentram-se no uso de *kernels* treináveis. Esses *kernels* geralmente são pequenos em dimensionalidade espacial, mas se estendem por toda a profundidade da entrada. Quando os dados passam por uma camada convolucional, ela aplica cada filtro na dimensionalidade espacial da entrada para produzir um mapa de ativação 2D. À medida que se percorre a entrada, o produto escalar é calculado para cada valor nesse *kernel*. A partir disso, a rede aprenderá *kernels* que “disparam” quando detectam uma característica específica em uma determinada posição espacial da entrada. Esses são comumente conhecidos como ativações conforme descrito na Figura 3.

Figura 3 – Representação visual da camada convolucional. O elemento central do *kernel* é aplicado no vetor de entrada, que é calculado e substituído pela ponderada dele mesmo e de quaisquer pixels próximos.



Fonte: (O’Shea; Nash, 2015)

Cada *kernel* terá um mapa de ativação correspondente, que será empilhado ao longo da dimensão de profundidade para formar o volume completo de saída da camada convolucional. Como mencionado anteriormente, treinar redes neurais artificiais tendo imagens como entrada resulta em modelos demasiadamente grandes para serem treinados de forma eficaz. Isso ocorre devido à conexão totalmente densa dos neurônios padrão das redes neurais artificiais. Para mitigar esse problema, cada neurônio em uma camada convolucional é conectado apenas a uma pequena região do volume de entrada (O’Shea; Nash, 2015).

A dimensionalidade dessa região é comumente chamada de “tamanho do campo receptivo” do neurônio. A magnitude da conectividade ao longo da profundidade geralmente é igual à profundidade da entrada. Por exemplo, se a entrada para a rede for uma imagem de tamanho 64x64x3 (uma imagem colorida RGB com dimensionalidade de 64x64) e for definido o tamanho do campo receptivo como 6x6, é obtido um total de 108 pesos em cada neurônio dentro da camada convolucional (6x6x3, onde 3 é a magnitude da conectividade na profundidade do volume). Para colocar isso em perspectiva, um neurônio padrão visto em outras formas de redes neurais artificiais conteria 12.288 pesos cada.

Camadas convolucionais também conseguem reduzir significativamente a complexidade do modelo por meio da otimização de sua saída. Essas otimizações são feitas por meio de três hiperparâmetros: profundidade, *stride* e *zero-padding*.

A profundidade do volume de saída produzido pelas camadas convolucionais pode ser definida manualmente através do número de neurônios dentro da camada para uma mesma região

da entrada. Isso pode ser observado em outras formas de redes neurais artificiais, onde todos os neurônios da camada oculta estão diretamente conectados a cada neurônio anterior. Reduzir esse hiperparâmetro pode minimizar significativamente o número total de neurônios da rede, mas também pode reduzir as capacidades de reconhecimento de padrões do modelo.

Também é possível definir o *stride*, que determina como é configurada a profundidade em torno da dimensionalidade espacial da entrada para posicionar o campo receptivo. Por exemplo, se o *stride* for definido como 1, obtém-se um campo receptivo fortemente sobreposto, produzindo ativações extremamente grandes. Por outro lado, ao configurar o *stride* para um número maior reduz a sobreposição e gera uma saída com dimensões espaciais menores.

O *zero-padding* é o processo de adicionar bordas à entrada, sendo um método eficaz para proporcionar maior controle sobre a dimensionalidade dos volumes de saída. É importante entender que, ao se usar dessas técnicas, a dimensionalidade espacial da saída das camadas convolucionais é alterada. Para este cálculo pode-se usar a seguinte Equação 2.1.

$$\frac{(V - R) + 2Z}{S + 1} \quad (2.1)$$

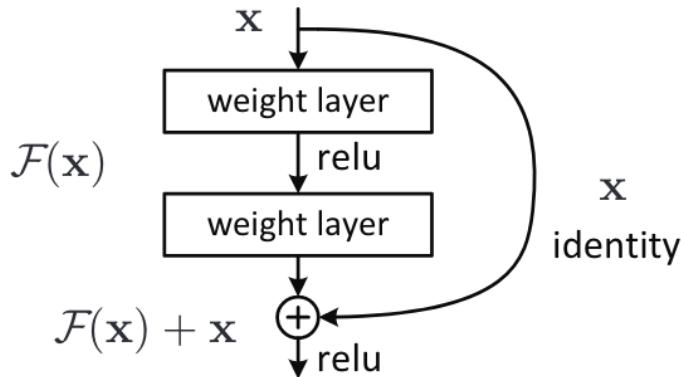
O termo  $V$  representa o tamanho do volume de entrada (altura x largura x profundidade),  $R$  representa o tamanho do campo receptivo,  $Z$  é a quantidade de *zero-padding* definida e  $S$  refere-se ao *stride*. Se o resultado calculado a partir desta equação não for um número inteiro, o *stride* foi configurado incorretamente, pois os neurônios não conseguirão se ajustar perfeitamente à entrada (O'Shea; Nash, 2015).

## 2.4 RESNET

A *ResNet* (Rede Residual) é uma arquitetura de rede neural profunda amplamente utilizada em tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação de imagens. Ela foi introduzida por He *et al.* (2015), e se destacou por ganhar a competição *ImageNet* em 2015 com uma precisão significativamente maior do que as arquiteturas anteriores. A principal característica da *ResNet* é sua capacidade de treinar redes muito profundas sem sofrer com o problema do “desvanecimento do gradiente” que afeta principalmente redes neurais profundas e impede o treinamento de progredir por conta de valores muito baixos de gradiente, problema este muito comum em redes tradicionais muito profundas. A *ResNet* introduz um conceito chave chamado bloco residual que é um componente da rede onde a entrada do bloco é somada à sua saída antes de passar para a próxima camada. Esta conexão direta é chamada de conexão de salto e pode ser observada na Figura 4.

A *ResNet* é composta por vários blocos residuais empilhados. Existem várias versões da *ResNet*, como *ResNet-18*, *ResNet-34*, *ResNet-50*, *ResNet-101*, e *ResNet-152*, onde os números indicam a profundidade total da rede, ou seja, o número de camadas.

Figura 4 – Conexão de Salto da Arquitetura *ResNet*



Fonte: (Ai *et al.*, 2023)

#### 2.4.1 Extração de Características Profundas

Para extrair características profundas das imagens de RMC, foi empregada a arquitetura pré-treinada de uma rede *ResNet50*. As redes *ResNet50* ficaram muito conhecidas em meados do ano de 2015 por vencer diversas competições em visão computacional, incluindo o primeiro lugar na competição de classificação de imagens *ILSVRC 2015*. As redes *ResNet*, do termo *Residual Networks*, inovaram em sua época trazendo uma nova forma de treinar modelos que possuem maior profundidade, chegando a mais de 100 camadas, com resultados superiores à outros modelos convolucionais competitivos, como o *VGG19*, sem sofrer sintomas comuns a redes neurais muito profundas como o *overfitting* e a saturação ou a ausência dos gradientes em tempo de treino. Os autores da *ResNet* sugeriram o uso de saltos de conexão entre as camadas da rede afim de manter os gradientes relevantes e controlados entre uma camada e outra. A *ResNet50* é um modelo de rede neural convolucional profunda de 50 camadas que compreende muitos blocos residuais. A cada bloco, se encontram módulos de convolução e uma conexão de salto que transfere a informação do bloco anterior para o próximo bloco. A conexão de salto ajuda a reter a informação semântica mais baixa aprendida nas camadas anteriores, que de outra forma se tornaria abstrata devido à conexão de longa cadeia. A conexão de salto também evita que o gradiente desapareça nas camadas mais profundas, fornecendo um caminho alternativo para a retropropagação. A informação da conexão de salto é adicionada à informação calculada em cada bloco (He *et al.*, 2015).

Várias abordagens de sucesso aplicaram redes convolucionais para extraírem características genéricas para tarefas de recuperação de imagens e obtiveram resultados promissores. Elas utilizam principalmente o poder das características locais para gerar uma representação de uma imagem genérica baseada em redes convolucionais pré-treinadas. As representações das camadas finais da rede convolucional são utilizadas para capturar características semânticas para o fim de nível de categoria à classificação que o modelo se dispõe (Alzu’bi; Amira; Ramzan, 2017).

## 2.5 REDE SQUEEZE AND EXCITATION

Para cada camada convolucional, um conjunto de filtros é aprendido para expressar padrões de conectividade espacial local ao longo dos canais de entrada. Em outras palavras, espera-se que os filtros convolucionais sejam combinações informativas ao fundir informações espaciais e baseadas nos canais dentro de campos receptivos locais. Ao empilhar uma série de camadas convolucionais intercaladas com não linearidades e redução de amostragem, as CNNs são capazes de capturar padrões hierárquicos com campos receptivos globais, funcionando como descrições de imagem poderosas (Hu; Shen; Sun, 2018).

Estudos de Hu, Shen e Sun (2018), demonstraram que o desempenho das redes pode ser aprimorado ao incorporar explicitamente mecanismos de aprendizado que ajudam a capturar correlações espaciais, sem a necessidade de supervisão adicional. Neste sentido, foi proposto um mecanismo que permite à rede realizar a recalibração de características, por meio do qual ela pode aprender a usar informações globais para enfatizar seletivamente características informativas e suprimir aquelas menos úteis. Este estudo intitulado *Squeeze and Excitation* (SE) Net, investiga um aspecto diferente no design arquitetural, a relação entre canais ao introduzir o bloco SE.

Os blocos SE podem ser construídos para qualquer transformação conforme a indica a Equação 2.2 onde  $\mathbf{U}$  representa a saída da transformação inicial da rede neural, antes da aplicação do bloco SE.

$$\mathcal{F}_{tr} : \mathbf{X} \rightarrow \mathbf{U}, \mathbf{X} \in R^{H' \times W' \times C'}, \mathbf{U} \in R^{H \times W \times C} \quad (2.2)$$

Sendo  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_C]$  os filtros de aprendidos onde  $\mathbf{v}_C$  se refere ao parâmetro de  $c$ -ésimo filtro. As saídas das convoluções podem ser escritas da seguinte forma:  $\mathcal{F}_{tr}\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_C]$ , sendo o termo  $\mathbf{u}_C$  descrito na equação 2.3.

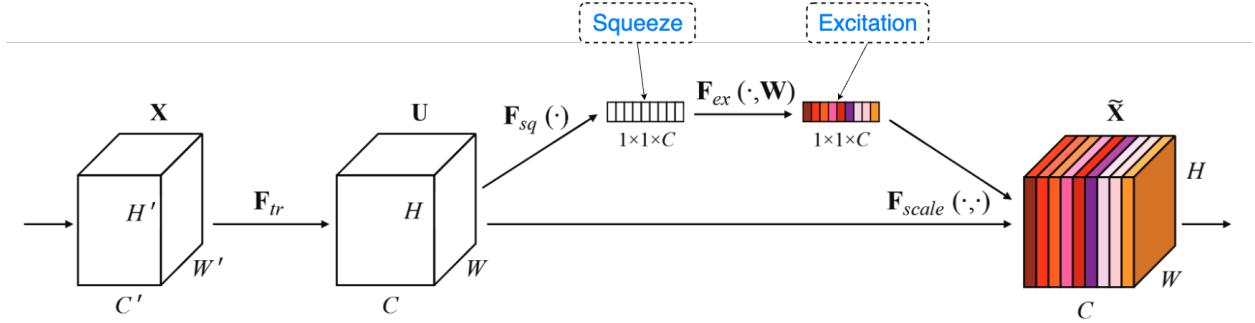
$$\mathbf{u}_c = \mathbf{v}_c * \mathbf{X} = \sum_{s=1}^{C'} \mathbf{v}_c^s * \mathbf{x}^s \quad (2.3)$$

O termo  $*$  denota convolução,  $\mathbf{v}_c$  é um filtro espacial 2D e representa um único canal de  $\mathbf{v}_c$  que atua no canal correspondente de  $\mathbf{X}$ . Tendo como o objetivo garantir que a rede seja capaz de aumentar sua sensibilidade a características informativas, para que possam ser exploradas por transformações subsequentes, e suprimir as menos úteis. Foi proposto um modelo explícito das interdependências de canal para recalibrar as respostas dos filtros em duas etapas — *squeeze* e *excitation* —, antes de serem alimentadas na transformação subsequente. Um diagrama de um bloco de construção SE é demonstrado na Figura 5.

### 2.5.1 Squeeze: Informação Global

Para lidar com o problema de explorar as dependências entre os canais, é considerado o primeiro sinal de cada canal nas características de saída. Cada um dos filtros aprendidos opera com um campo receptivo local e, consequentemente, cada unidade da saída da transformação  $\mathbf{U}$

Figura 5 – Fluxo do Bloco SE - Após blocos convolucionais normais, uma cama da de *squeeze* seguida e uma camada de *excitation* é aplicada e por fim os valores originais são escalonados pelo resultado.



Fonte: Adaptado de (Hu; Shen; Sun, 2018)

é incapaz de explorar informações contextuais fora dessa região. Esse problema se torna ainda mais grave nas camadas mais baixas da rede, cujos campos receptivos são menores.

Para mitigar esse problema, foi proposto comprimir as informações espaciais globais em um descriptor de canal. Isso é obtido por meio de uma instrução de *average pooling* global para gerar estatísticas específicas de cada canal. O cálculo do *average pooling* pode ser conferido na Equação 2.4.

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2.4)$$

### 2.5.2 Excitation: Recalibração Adaptativa

Para aproveitar as informações agregadas na operação de *squeeze*, uma segunda operação que tem como objetivo capturar completamente as dependências entre os canais. Para cumprir esse objetivo, a função deve atender a dois critérios: primeiro, ela precisa ser flexível (em particular, deve ser capaz de aprender uma interação não linear entre os canais) e, segundo, ela deve aprender uma relação não mutuamente exclusiva, pois é de interesse garantir que vários canais possam ser enfatizados, em vez de apenas uma ativação *one-hot*. Para atender a esses critérios, emprega-se um mecanismo de *gating* simples com uma ativação sigmoide conforme demonstrado na Equação 2.5:

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (2.5)$$

sendo que o termo  $\delta$  se refere à função de ativação ReLU (Nair; Hinton, 2010),  $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$  e  $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ . Para limitar a complexidade do modelo e auxiliar na generalização, o mecanismo de *gating* é parametrizado formando um gargalo com duas camadas totalmente conectadas ao redor da não linearidade, ou seja, uma camada de redução de dimensionalidade com os parâmetros  $W_1$  e razão de redução  $r$ , seguida de uma ReLU e, em seguida, uma camada de aumento de

dimensionalidade com os parâmetros  $W_2$ . A saída final do bloco é obtida ao redimensionar a saída da transformação  $U$  com as ativações conforme Equação 2.6.

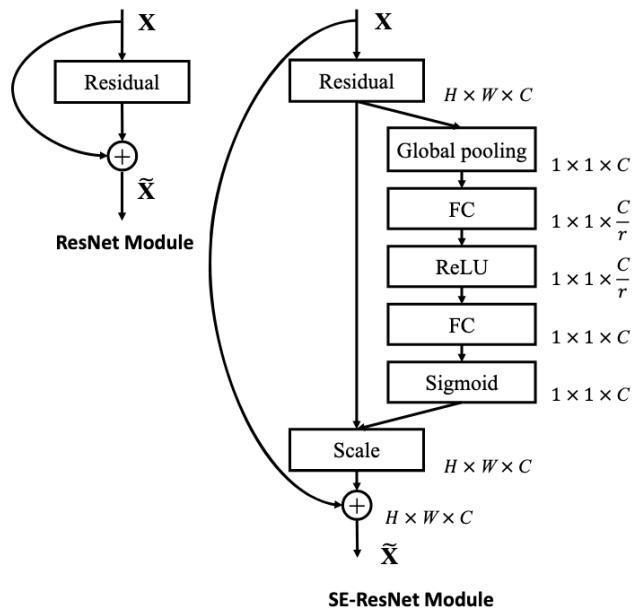
$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, s_c) = s_c \cdot \mathbf{u}_c \quad (2.6)$$

onde  $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_C]$  e  $\mathbf{F}_{scale}(\mathbf{u}_c, s_c)$  se referem a multiplicação no nível dos canais entre os mapas de características  $\mathbf{u}_c \in \mathbb{R}^{H \times W}$  e o valor escalar  $s_c$ .

### 2.5.3 Utilização em redes conhecidas: SE-Resnet

A flexibilidade do bloco SE significa que ele pode ser aplicado diretamente a transformações além das convoluções padrão. Para ilustrar esse ponto, os autores desenvolveram as redes SE integrando blocos SE em arquiteturas modernas com *designs* sofisticados. A Figura 6 demonstra o esquema de um módulo SE-ResNet. Nele, a transformação do bloco SE,  $F_{tr}$ , é considerada o ramo não identidade de um módulo residual. As operações de *squeeze* e *excitation* atuam antes da soma com o ramo identidade.

Figura 6 – Módulo SE-Resnet, que estende o bloco ResNet tradicional com um bloco SE para recalibrar os canais da saída residual com base em informações globais, aprimorando a representação de características relevantes.



Fonte: Adaptado de (Hu; Shen; Sun, 2018)

Para que o bloco SE proposto seja viável na prática, ele deve fornecer um equilíbrio eficaz entre a complexidade do modelo e o desempenho, o que é importante para escalabilidade. Foi definida uma razão de redução  $r$  como 16 em todos os experimentos, exceto onde indicado de outra forma.

O valor  $r$  é um hiperparâmetro importante pois permite variar a capacidade e o custo computacional dos blocos SE no modelo. O autor do trabalho conduz experimentos baseados

no SE-ResNet-50 para uma variedade de valores diferentes de  $r$ . A comparação vista na Figura 7, revela que o desempenho não melhora monotonicamente com o aumento da capacidade. Isso provavelmente é resultado do bloco SE ajustar em excesso as interdependências de canal do conjunto de treinamento. Definir  $r = 16$  alcançou um bom equilíbrio entre precisão e complexidade e, consequentemente, foi o valor utilizado pelos autores nos experimentos. O Algoritmo 1 reflete a implementação do bloco SE.

Figura 7 – Desempenho da SE-ResNet-50 com diferentes valores do fator de redução  $r$ , demonstrando variação nos erros top-1 e top-5 e no número de parâmetros, em comparação com a arquitetura ResNet-50 original.

Ratio $r$	top-1 err.	top-5 err.	Million Parameters
4	23.21	6.63	35.7
8	23.19	6.64	30.7
16	23.29	6.62	28.1
32	23.40	6.77	26.9
original	24.80	7.48	25.6

Fonte: Adaptado de (Hu; Shen; Sun, 2018)

---

#### Algorithm 1 Bloco Squeeze-and-Excitation (SE)

---

**Entrada:** Mapa de características  $\mathbf{U} \in \mathbb{R}^{H \times W \times C}$

**Saída:** Mapa de características recalibrado  $\tilde{\mathbf{U}} \in \mathbb{R}^{H \times W \times C}$

1: **Squeeze:** Executa o *average pooling* global para agregar as dimensões espaciais:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \quad \forall c \in \{1, \dots, C\}$$

2: **Excitation:** Usa duas camadas totalmente conectadas para modelar a interdependência dos canais:

$$\mathbf{s} = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z}))$$

**onde:**

- $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ : Camada de redução de dimensionalidade
- $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ : Camada de restauração da dimensão original
- $\delta$ : Ativação ReLU
- $\sigma$ : Ativação Sigmoide

3: **Reescala:** Recalibra o mapa de características usando as ativações:

$$\tilde{\mathbf{u}}_c = s_c \cdot \mathbf{u}_c, \quad \forall c \in \{1, \dots, C\}$$

4: **return**  $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \dots, \tilde{\mathbf{u}}_C]$

---

## 2.6 ARQUITETURA TRANSFORMERS

Os *transformers* consistem em uma arquitetura que possui como um de seus propósitos resolver as limitações das arquiteturas recorrentes e sua dificuldade em manter as relações entre pontos dentre as camadas recorrentes além das restrições vinculadas ao custo de computação sequencial. O modelo de arquitetura *transformers* se utiliza do mecanismo de atenção e este se tornou parte essencial de modelos eficazes de modelagem e transdução de sequências em diversas tarefas, permitindo a modelagem de dependências sem considerar a distância entre elas nas sequências de entrada ou saída. Os *transformers* como arquitetura descarta o uso de módulos de recorrência e se utiliza inteiramente do mecanismo de atenção para capturar as dependências globais entre entrada e saída. Os *transformers* também são responsáveis por um ganho significante em paralelismo em sua execução (Vaswani *et al.*, 2017).

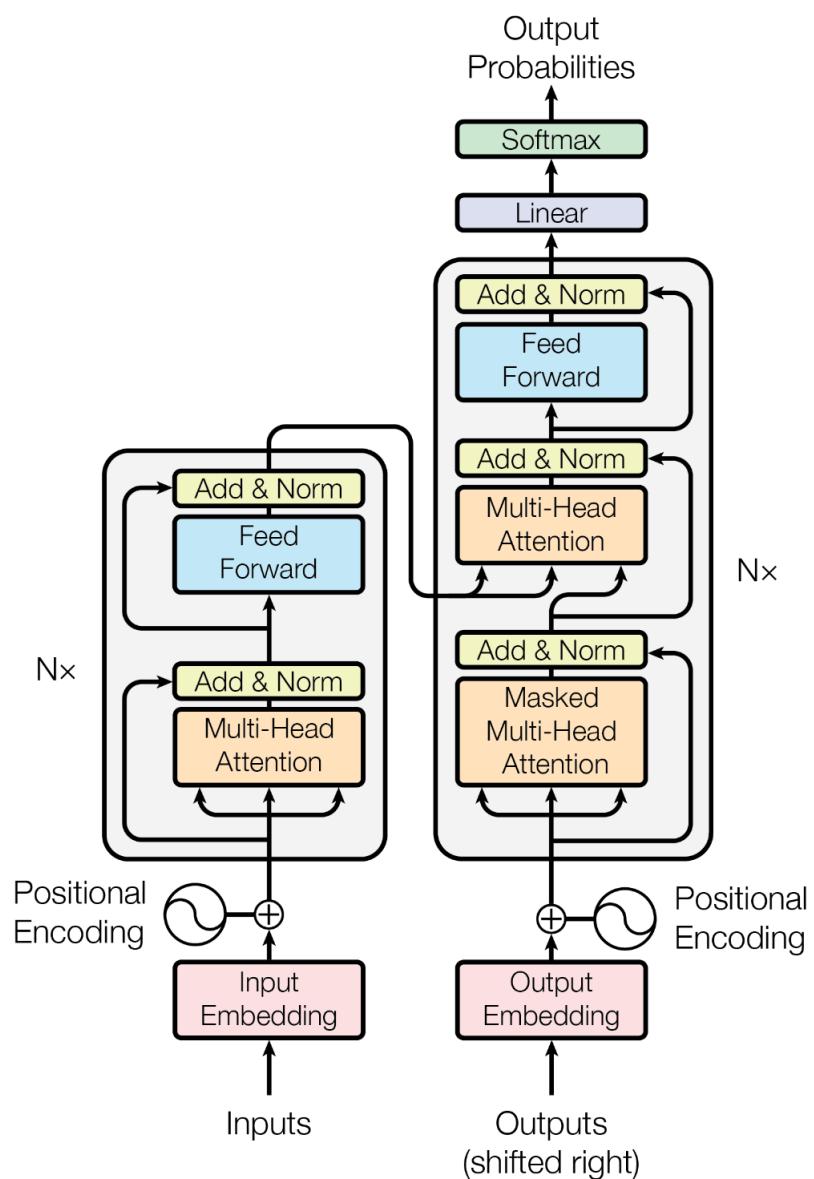
Os *transformers* são compostos por um *Encoder* e um *Decoder*, representados pelos os blocos da esquerda e direita respectivamente apresentados na Figura 8. Em ambos *Encoder* e *Decoder*, tem-se como bloco cerne da rede intitulado de atenção multi-cabeças. O mecanismo de atenção pode ser descrito mapeando uma *query* a um par de chave e valor, onde a *query*, a chave e o valor são todos vetores de saída. A saída é computada como uma soma ponderada dos valores onde o peso destinado a cada valor é computado por uma função de compatibilidade da *query* com a chave correspondente.

O mecanismo de atenção é particularmente chamado de “Atenção de Produto Escalar Dimensionado”, visto na Figura 9. A entrada consiste em *queries* e chaves de dimensão  $d_k$  e valores de dimensão  $d_v$ . São calculados os produtos escalares da consulta com todas as chaves e dividido por  $\sqrt{d_k}$ , para fins de controle dos valores em uma menor amplitude. Em seguida, aplica-se uma função *softmax* para obter as probabilidades sobre os valores. Na prática, é calculada a função de atenção em um conjunto de consultas simultaneamente, agrupadas em uma matriz  $Q$ . As chaves e os valores também são agrupados em matrizes  $K$  e  $V$ . Calcula-se a matriz de saídas como:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (2.7)$$

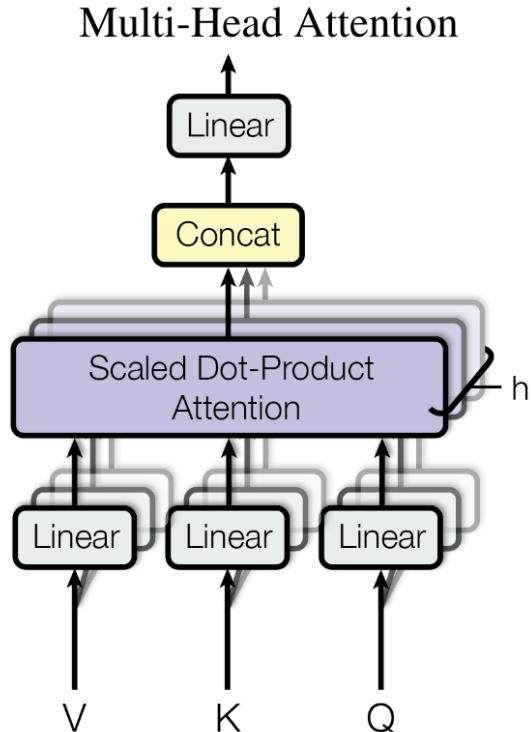
Dentre os pontos de vantagem do mecanismo de atenção, se destacam: o total de poder computacional por camada, o total de computação que pode ser paralelizada e o poder de capturar a relação de dependência entre dados que se encontram distantes espacialmente um do outro. Como benefício adicional, o mecanismo de atenção pode gerar modelos mais interpretáveis, sob o ponto de vista de como os *tokens* se correlacionam. Não apenas as cabeças de atenção individuais claramente aprendem a executar diferentes tarefas, muitas parecem exibir comportamentos relacionados à estrutura sintática e semântica das frases, no caso da aplicação em *tokens* oriundos de textos.

Figura 8 – Arquitetura *Transformers* - Modelo composto por codificador e decodificador com camadas de atenção multi-cabeça, normalização, *feed-forward* e codificação posicional.



Autor: Vaswani *et al.* (2017)

Figura 9 – Módulo de Atenção Multi-Cabeças



Autor:(Vaswani *et al.*, 2017)

## 2.7 OTIMIZADOR ADAM

Toda rede neural é treinada aplicando otimização em uma determinada função objetivo afim de minimizar o erro perante os dados de treinamento. Neste contexto, o presente trabalho escolhe o *adaptive moment estimation* (Adam) como otimizador dada sua adaptatividade.

O método Adam, introduzido por Kingma e Ba (2014), é um método popular para o treinamento de modelos de aprendizado profundo (AP). O Adam combina os benefícios de outras duas técnicas de otimização, o *AdaGrad* e o *RMSProp*. O Adam é um método de otimização estocástica eficiente que requer apenas gradientes de primeira ordem com pouca exigência de memória. O método calcula taxas de aprendizado adaptativas individuais para diferentes parâmetros a partir de estimativas dos primeiros e segundos momentos dos gradientes.

Algumas das vantagens do Adam podem ser elencadas: taxas de aprendizado adaptativas, que levam a uma convergência mais rápida e eficiente se comparadas com métodos de aprendizado fixos; robustez, que suportam gradientes esparsos de forma efetiva, cruciais para diversas aplicações de AP atuais; fácil de utilizar, pois requer menos ajustes de parâmetros a tornando amigável ao usuário e acessível a uma grande gama de tarefas.

### 3 TRABALHOS RELACIONADOS

Aplicar a unificação de técnicas radiômicas e aprendizado profundo na pesquisa de cardiomiopatia tem recebido cada vez mais atenção da comunidade científica. Neste capítulo serão apresentados trabalhos e metodologias correlatos ao esforço de outros autores a cerca deste tema. Será descrito como foi a metodologia utilizada na revisão sistemática da literatura (Petersen, 2015), assim como trabalhos que trazem ideias próximas ao proposto neste trabalho.

#### 3.1 REVISÃO SISTEMÁTICA DA LITERATURA

Para a fase exploratória dos trabalhos relacionados, foi utilizada a ferramenta *Parsifal*<sup>1</sup> com o objetivo de identificar estudos que aplicam a análise radiônica no contexto de cardiomiopatia. O objetivo inicialmente definido foi o seguinte: “Identificar estudos onde se aplica análise radiônica a cardiopatia. Como segunda opção, alguma outra doença de natureza cardíaca”. A pesquisa teve como objetivo responder as seguintes perguntas: quais desafios descritos nos estudos prévios e se há replicabilidade da proposta no contexto de cardiomiopatia. As palavras chaves selecionadas foram: *cardiac*, *cardiomyopathy* e *radiomics*. A palavra de busca selecionada foi: “*radiomics AND "data extraction" AND (cardiac OR cardiomyopathy)*” e a última atualização das pesquisas foram feitas em agosto de 2023. Os resultado do filtro de busca são apresentados na Tabela 1.

Tabela 1 – Resultados dos Artigos

Origem	Artigos
IEEE	6
PUBMED	19
Science@Direct	24

Fonte: Autor

A partir dos resultados foram aplicados os critérios de inclusão e exclusão conforme descritos na Tabela 2.

Critérios de inclusão também foram aplicados e são apresentados na Tabela 3. Dentre estes critérios, cada artigo é representado pela soma das pontuações obtidas pelos critérios e os artigos que obtiveram uma nota menor que 3 foram excluídos da revisão. Para as notas, ficou definido como pontuação: 1 para critério aceito, 0,5 para critério parcialmente aceito e 0 para critério não atingido. Uma vez aplicado os critérios e filtrado mediante pontuação, restaram 18 dos 49 artigos iniciais. Para fins de elucidar, o critério de mais de dois autores envolvem

<sup>1</sup> <https://parsif.al>

Tabela 2 – Critérios de Inclusão e Exclusão

Critério de Inclusão	Critério de Exclusão
Contém ressonância magnética?	Estudos Duplicados
O objeto de estudo é o coração?	Estudos Secundários ou Terciários
Usa análise de textura?	Estudos que não estão em PT ou EN
Utiliza análise radiométrica?	Leitura cinza
É cardiomiopatia?	Não aplica técnicas computacionais Não trata do coração Não usa RM Trabalho que atua somente com dados genômicos

Fonte: Autor

colaboração entre diferentes especialistas ou instituições, seguindo por vezes, padrões mais rigorosos e tendo maior visibilidade na comunidade científica.

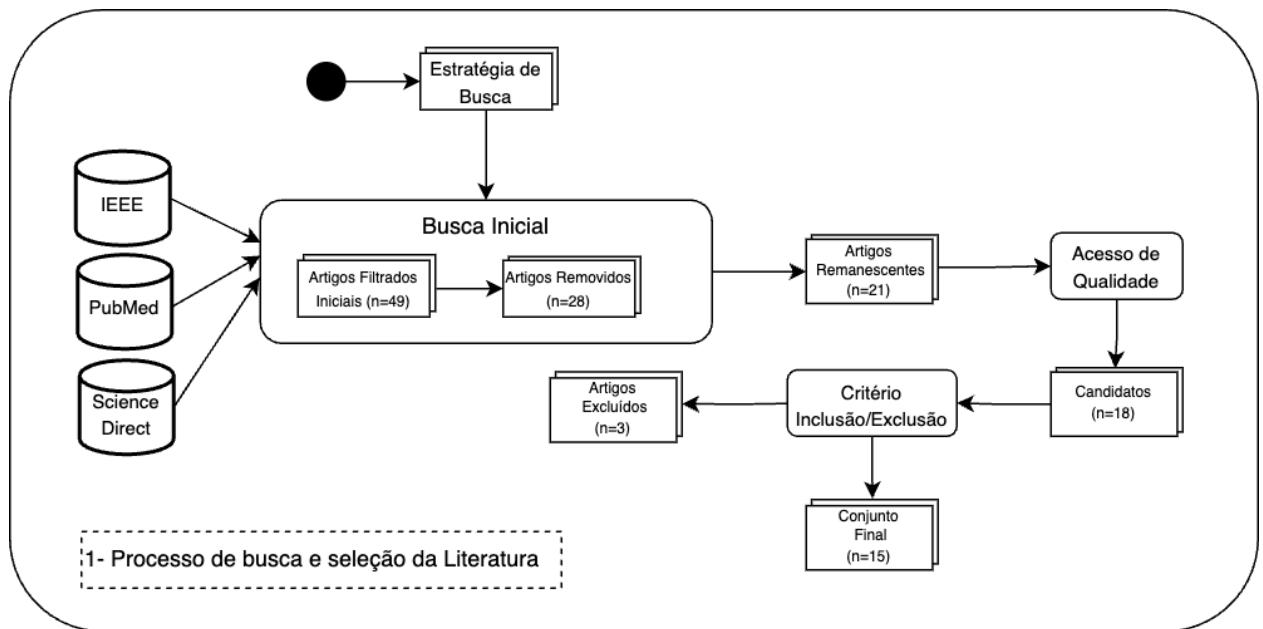
Tabela 3 – Questões de Aceitação

Questões
É descrita as limitações do estudo?
Há um experimento bem descrito para avaliar a proposta?
Há mais de 2 autores?
O trabalho apresenta resultados?
A introdução apresenta o problema de forma clara
É análise sistemática da literatura?

Fonte: Autor

O diagrama com o funil de seleção aplicado é apresentado na Figura 10. É possível verificar na imagem, por exemplo, que foram removidos 28 artigos dos 49 artigos selecionados inicialmente, restando 21. Aplicando os critérios de qualidade, foram selecionados 18 artigos. Dentre os artigos selecionados, 3 foram excluídos por não agregarem o suficiente ou por serem redundantes.

Figura 10 – Funil da Seleção da Literatura - Etapas do processo de busca, filtragem e seleção de artigos com base em critérios de inclusão, exclusão e qualidade, resultando no conjunto final de estudos analisados.



Fonte: Autor

### 3.2 ANÁLISE RADÔMICA COM ATENÇÃO

Jiang *et al.* (2021) conduziram um estudo abrangente que aproveita as capacidades do aprendizado de máquina para melhorar os processos diagnósticos em oncologia, especificamente focando no câncer de colo de útero em estágio inicial. Sua pesquisa, intitulada “Abordagem Radiômica Baseada em Ressonância Magnética com Aprendizado Profundo para Predição de Invasão Vascular em Câncer de Colo de Útero em Estágio Inicial”, foca na aplicação de técnicas de aprendizado profundo em dados de RMC multiparamétrica para prever a invasão vascular, um fator crítico para determinar a agressividade do câncer de colo de útero e informar estratégias de tratamento.

O estudo utilizou um conjunto substancial de dados compreendendo 1.070 imagens de ressonância magnética com contraste dinâmico T1 (DCE-T1) e 986 imagens de ressonância magnética ponderadas em T2 (T2WI) e coletadas de 167 pacientes diagnosticadas com câncer do colo do útero em estágio inicial entre janeiro de 2014 e agosto de 2018. Os pesquisadores empregaram uma nova estrutura de aprendizado profundo que integrou esses dois tipos distintos de varreduras em imagens de RMC para criar um modelo preditivo robusto. Implementando uma estratégia de aprendizado em conjunto com atenção, o modelo efetivamente distinguiu entre casos com invasão vascular e aqueles sem. Quatro modelos de CNN foram utilizados: VGGNet, GoogLeNet (Inception-v3), Residual Network (ResNet) e *DenseNet*. Um módulo padrão de SE e o módulo de atenção de bloco convolucional, do termo inglês *convolutional block attention module* (CBAM), foram introduzidos após cada bloco de convolução da rede *AdaptedVGG19* para

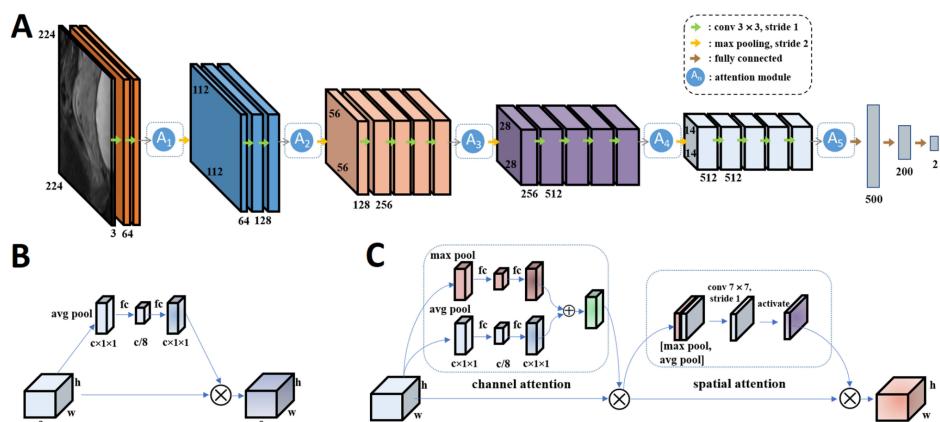
fornecer um *AdaptedVGG19-SE* e um *AdaptedVGG19-CBAM*, respectivamente. A arquitetura proposta é vista na Figura 11.

O desempenho preditivo dos modelos foi rigorosamente avaliado, com os modelos finais alcançando uma AUC de 0,911. Essa alta AUC indica uma forte capacidade do modelo para classificar corretamente a presença ou ausência de invasão vascular, com métricas de sensibilidade e especificidade também demonstrando precisão diagnóstica substancial.

Esta pesquisa demonstra o potencial de integrar algoritmos de aprendizado profundo com dados de imagem complexos para melhorar as avaliações diagnósticas pré-operatórias. Os achados de Jiang *et al.* (2021) sugerem que tais abordagens analíticas avançadas podem fornecer suporte substancial na tomada de decisões clínicas, potencialmente levando a planos de tratamento mais personalizados e melhores resultados para os pacientes.

No contexto da pesquisa contínua em imagens médicas e diagnóstico de câncer, a metodologia e os resultados fornecem um exemplo convincente do potencial da inteligência artificial para revolucionar os processos diagnósticos em oncologia. O uso de RMC multiparamétrica e modelos sofisticados de aprendizado de máquina exemplifica as abordagens inovadoras que estão sendo desenvolvidas para enfrentar os desafios da detecção precoce e precisa de câncer.

Figura 11 – Modelo de arquitetura de rede neural proposto por Jiang *et al.* (2021). A imagem ilustra: (A) o fluxo principal da rede; (B) um módulo de atenção de canal; e (C) um módulo de atenção combinada (canal e espacial).



Fonte: (Jiang *et al.*, 2021)

Em outro trabalho, Ren *et al.* (2023), desenvolveram um modelo avançado usando uma rede Bi-LSTM (memória de longo prazo bidirecional) combinada com um mecanismo de atenção de múltiplas cabeças, utilizando características radiométricas e imagens de TC de lesões, para aprimorar a diferenciação dos principais subtipos de adenocarcinoma pulmonar, integrando assinaturas radiométricas com características radiológicas de tomografias computadorizadas. O estudo recrutou 421 pacientes de três hospitais, confirmados com adenocarcinoma *in situ*, adenocarcinoma minimamente invasivo ou adenocarcinoma invasivo, com base na análise de 427 lesões.

A metodologia empregada envolveu a extração de características radiômicas usando o software ‘*PyRadiomics*’ das regiões de lesões identificadas em cada imagem de TC. As 100 principais características foram então selecionadas por meio do método de classificação de características de máxima relevância e mínima redundância. Um modelo preditivo foi subsequentemente desenvolvido empregando essas características juntamente com as características radiológicas, usando a estrutura Bi-LSTM e atenção múltipla para classificar as lesões.

O desempenho diagnóstico do modelo foi quantitativamente impressionantes relativos aos modelos bases testados, alcançando valores da área sob a curva (AUC) de 0, 985, 0, 94 e 0, 981 nos grupos de treinamento, teste e validação, respectivamente, com precisão correspondentes a 0, 92, 0, 976 e 0, 91. Além disso, comparações foram feitas com dois outros modelos — CNN em conjunto com atenção múltipla, e LSTM em conjunto com atenção múltipla — demonstrando que o modelo Bi-LSTM e atenção múltipla superou essas alternativas em precisão e acurácia no conjunto de testes.

Esta pesquisa destaca a potente combinação de técnicas avançadas de aprendizado de máquina com análises radiômicas e radiológicas detalhadas para refinar o processo diagnóstico para subtipos de adenocarcinoma pulmonar, potencialmente orientando abordagens de tratamento mais personalizadas baseadas na caracterização do subtipo.

Ai *et al.* (2023) conduziram um estudo inovador intitulado “Um Modelo de Fusão Baseado em Autoatenção de Características Radiômicas e Profundas para Previsão de Recorrência Precoce em Câncer de Pulmão de Células Não Pequenas”, que aborda o significativo desafio de prever a recorrência precoce deste câncer usando técnicas avançadas de aprendizado de máquina. Sua pesquisa aproveita o mecanismo de autoatenção para fundir características radiômicas manuais e características de aprendizado profundo extraídas de imagens de TC, com o objetivo de aumentar a precisão preditiva e robustez para recorrência precoce em câncer de pulmão de células não pequenas.

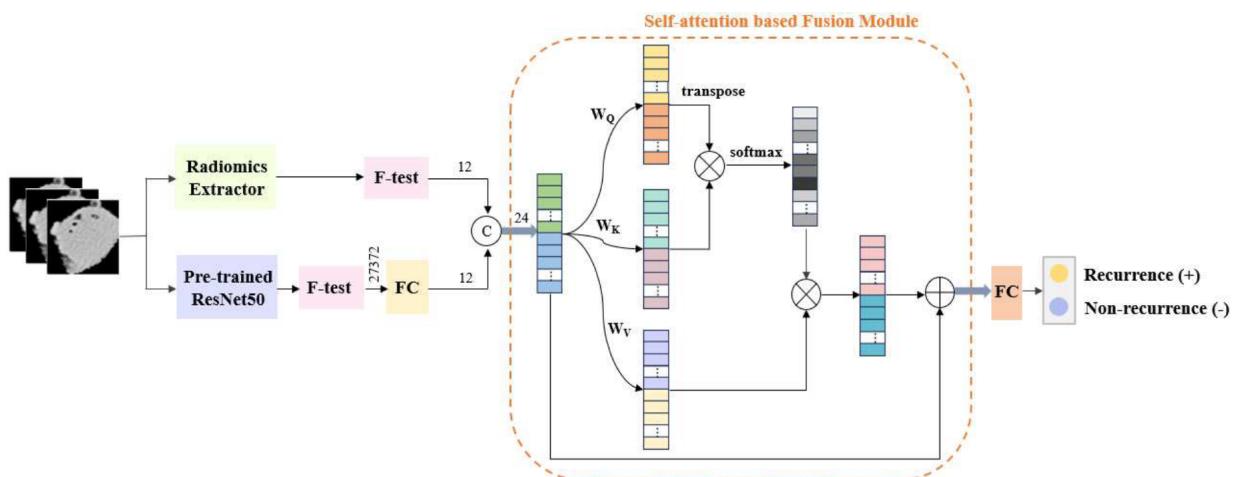
O estudo começou empregando diversas técnicas de aprendizado de máquina para extrair uma variedade de características manuais de imagens de TC, incluindo características de textura, forma e escala de cinza. Para capturar informações semânticas de alto nível e de representação, uma rede *ResNet50* pré-treinada foi utilizada para a extração de características profundas. Essas características extraídas foram então fundidas com um vetor de características extraído de dados de texturas de imagens radiômicas e unificado usando um módulo de fusão de autoatenção inovador desenvolvido pelos pesquisadores. Este módulo otimiza e pondera o vetor de características fundidas, aproveitando plenamente o mecanismo de autoatenção para melhorar as capacidades de previsão do modelo e sua arquitetura pode ser conferida na Figura 12.

Os resultados experimentais, avaliados no conjunto de dados público, *cancer imaging archive* (TCIA), demonstraram que o modelo proposto superou significativamente os métodos existentes na previsão de recorrência precoce. O modelo exibiu melhorias substanciais em precisão de classificação, sensibilidade, especificidade e AUC, destacando seu potencial para guiar o tratamento em estágio inicial e melhorar as taxas de sobrevivência para pacientes com

câncer de pulmão de células não pequenas.

Este estudo exemplifica a aplicação de técnicas avançadas de aprendizado de máquina em imagens médicas e oncologia, fornecendo um método robusto para a previsão de recorrência precoce que poderia impactar significativamente os resultados clínicos e as estratégias de tratamento em câncer de pulmão.

Figura 12 – Arquitetura Proposta por Ai *et al.* (2023), que emprega um Módulo de Fusão Baseado em Autoatenção (Self-attention based Fusion Module). O modelo integra características radiômicas, extraídas diretamente das imagens, com características de aprendizado profundo de uma rede ResNet50 pré-treinada. Ambas as vertentes de características passam por um processo de seleção (F-test) antes de serem concatenadas e processadas pelo módulo de autoatenção para a classificação final.



Fonte: (Ai *et al.*, 2023)

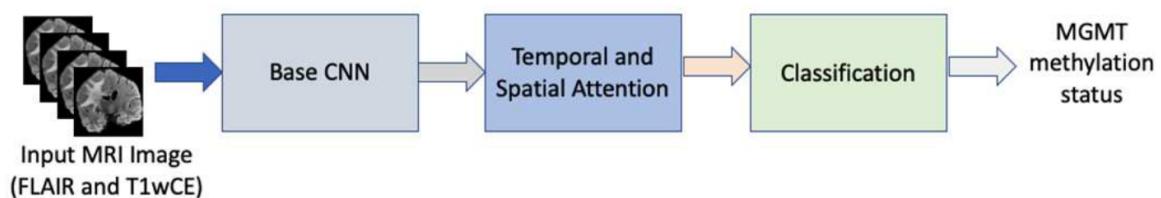
Iranmehr, Gutta e Hadaegh (2022) desenvolveram uma rede de aprendizado profundo inovadora que utiliza um mecanismo baseado em atenção para aprimorar a previsão do estado de metilação do gene MGMT em *glioblastoma multiforme* (GBM), o tipo mais agressivo de tumor cerebral. Pacientes com GBM possuem uma expectativa muito baixa, entre 18 e 24 meses e requerem tratamentos agressivos, como por exemplo, quimioterapia. Sua pesquisa, apresentada em "Improved Prediction of MGMT Methylation Status in Glioblastoma using a Deep Attention Network", destaca um avanço significativo nas capacidades diagnósticas não invasivas para GBM, que tipicamente tem uma taxa de sobrevivência de apenas 18 meses.

O estudo foca no gene MGMT, cujo estado de metilação é crucial para determinar a eficácia da quimioterapia em pacientes com GBM. As análises radiômicas tradicionais, embora úteis, muitas vezes não capturam os recursos intrincados necessários para uma previsão precisa da metilação. Iranmehr, Gutta e Hadaegh (2022) propõem um modelo que integra características radiômicas manuais com técnicas de AP, melhorando a extração de características e a precisão da previsão de GBM.

O modelo introduzido pela equipe utiliza uma combinação de mecanismos de atenção

espremida e sequencial para priorizar fatias e regiões relevantes dentro das imagens de RMC, respectivamente. Esse método não apenas melhora o foco em áreas significativas, mas também aprimora a interpretabilidade geral do modelo. O modelo proposto na Figura 13, consiste de três etapas: 1) Um modelo base para extrair as características, 2) uma rede de atenção temporal e espacial e 3) uma rede de classificação para prever se o exame é metilado ou não metilado.

Figura 13 – Método proposto. *DenseNet* é o modelo base, a saída da rede densa é alimentada pela rede de atenção que prioriza as fatias e regiões e gera a classificação binária do status de metilação do MGMT.



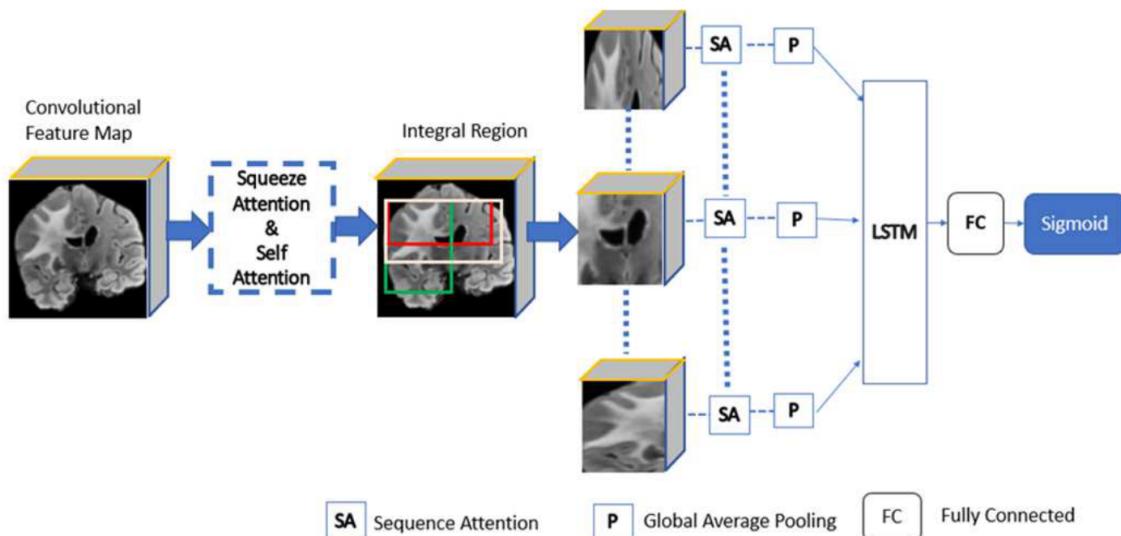
Fonte: (Iranmehr; Gutta; Hadaegh, 2022)

O modelo base consiste em uma *DenseNet* e segundo os autores, ela pode suportar milhares de camadas e ser resistente ao *overfitting*. A saída da rede densa é fornecida para a rede espremida e autoatenção conforme mostrado na Figura 14. Cada exame consiste em várias fatias e a atenção espremida priorizará fatias, usando o *average pooling* global seguido por duas camadas densas separadas e depois o produto escalar com a entrada inicial. A saída da atenção espremida é fornecida à rede de autoatenção. Após a aplicação da autoatenção, o mapa de atenção contém *pixels* com uma seção de maior importância em cada fatia. Com a rede de autoatenção, é possível enfatizar *pixels* e regiões diferentes independente da distância em que estes *pixels* se localizam na imagem. Múltiplas regiões com diferentes tamanhos, geradas a partir de regiões integrais, são então fornecidas à atenção sequencial. A rede de atenção sequencial adapta uma rede *Long Short-Term Memory* (LSTM), que é capaz de aprender dependências de longo prazo. A saída passa por uma camada densa e uma função sigmoide para fins de classificação.

Avaliado em várias métricas de classificação binária, o modelo alcançou a melhor AUC de 70,59, demonstrando sua superioridade em relação aos métodos existentes. Este trabalho fornece uma abordagem robusta e automática para capturar características críticas de imagens de ressonância magnética, avançando significativamente na previsão do estado de metilação em GBM em comparação com métodos anteriores. As implicações de tais avanços são profundas, potencialmente melhorando o planejamento de tratamentos personalizados e, em última análise, os resultados para os pacientes com glioblastoma.

Por fim, Yang *et al.* (2024) demonstram como o mecanismo de atenção auxilia as redes neurais a se concentrar em informações pertinentes nos dados, e sua integração com redes neurais convolucionais aprimora as capacidades de extração de características do modelo como um todo.

Figura 14 – Arquitetura Proposta por Iranmehr, Gutta e Hadaegh (2022) para análise de mapas de características convolucionais, que inclui bloco SE e autoatenção, segmentação de regiões integrais (representadas pelas áreas em verde e vermelho), além de uma camada LSTM. O modelo também utiliza diferentes tipos de atenção para focar em regiões específicas dos dados.



Fonte: Adaptado de (Iranmehr; Gutta; Hadaegh, 2022)

O trabalho destaca que o mecanismo de atenção pode ser categorizado em: mecanismo de atenção seletiva e mecanismo de autoatenção.

O mecanismo de atenção seletiva gera um conjunto de pesos de atenção ao prever diretamente a importância de cada parte dos dados em relação ao objetivo da tarefa, adicionando os pesos ponderadamente aos dados de entrada para realçar as informações importantes e suprimir as informações irrelevantes nos dados. A SENet é um trabalho representativo nos mecanismos de atenção seletiva. A SENet propõe um mecanismo de atenção no domínio dos canais que aprende o valor dos pesos para cada canal no mapa de características, a fim de realçar ou suprimir cada característica do canal. O mecanismo de atenção seletiva é orientado para os objetivos da tarefa e otimiza diretamente as características, mas carece da capacidade de extrair informações globais. O mecanismo de autoatenção consegue capturar informações globais, mas sofre com alta complexidade e dificuldades de otimização.

Yang *et al.* (2024) propuseram um algoritmo com módulo de busca de atenção diferenciável, assim a rede pode extrair as informações cognitivas e semântica das características da imagem. Em seguida, um espaço de busca multi-atenção é estabelecido englobando seis operações de atenção consolidadas em um único módulo de busca. Por fim, para aprimorar a eficiência, o espaço de busca, antes discreto, é convertido para contínuo, empregando um algoritmo de busca diferencial para identificar a rede de atenção ótima. Para validar os resultados, foi aplicada a classificação em imagens muito detalhadas.

Como resultado o método alcançou acurácia de 87,4% no conjunto de dados CUB-

200-2011 que consiste em mais de 11 mil imagens de pássaros distribuídas em 200 classes, apresentando um ganho de desempenho de 1,1%, 0,9% e 2,1% em relação ao módulo de atenção seletiva por canal, ao módulo de atenção seletiva espacial e ao módulo de autoatenção espacial, respectivamente. Por fim, os autores concluem que o módulo de atenção encontrado incorpora as vantagens de diferentes mecanismos de atenção e pode ajudar o modelo a melhor extrair e utilizar características discriminativas (Yang *et al.*, 2024).

### 3.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Os métodos de avaliação de imagens médicas, tanto via análise radiômica por extração de textura quanto por aprendizado profundo obtiveram êxito em análise de imagens médicas. Vale pontuar que estas técnicas de Aprendizado de Máquina, utilizando dados de textura podem ser insuficientes em capturar a complexidade e diversidade de informações a cerca do ventrículo como também pode ser afetada pela qualidade da imagem. Métodos de aprendizado profundo podem aprender automaticamente semânticas de alto nível e informações representativas de imagens de RMC sem necessidade de características extraídas manualmente. Métodos de aprendizado profundo tem obtido ótimos resultados em diagnósticos, segmentação, estadiamento e prognósticos, todavia o mesmo tem suas limitações como a baixa quantidade no conjunto de dados, um desafio geral na área médica, *overfitting* e pouca interpretabilidade do modelo em si (Pontone *et al.*, 2022).

Dadas as vantagens e limitações discutidas, este trabalho tem como proposta a fusão de uma arquitetura de aprendizado profundo baseada no mecanismo de atenção que integra características radiômicas e profundas para a identificação de cardiomiopatias.

## 4 TRABALHOS RELACIONADOS

A metodologia deste trabalho visa propor e investigar arquiteturas de AP que unificam características radiômicas e características profundas. O modelo base selecionado é uma versão adaptada do trabalho de Ai *et al.* (2023), mencionado nos trabalhos relacionados e o esquemático de sua arquitetura pode ser visto na Figura 12. Importante ressaltar que o trabalho em questão é aplicado no cenário de câncer de pulmão de células não pequenas em imagens de TC e este trabalho pretende aplicar o mesmo estudo no cenário de cardiomiopatia sendo este o modelo linha de base. Também serão aplicado adaptações a esta arquitetura original como: quantidade de execuções nos transformers, aumento dos vetores profundos e radiômicos, mudança na forma de concatená-los, entre outros.

Os experimentos foram divididos em duas fases manuais. Na primeira fase, foi aplicado o modelo base, ao conjunto de dados público selecionado. Na etapa de pré-processamento desta fase, foram aplicadas diversas técnicas de aprendizado de máquina para extrair características manuais de imagens de RMC e sua respectiva máscara, abrangendo textura, forma, escala de cinza, etc. Estas características serão chamadas, neste ponto em diante, de características radiômicas. Seguindo com o pré-processamento, as imagens de RMC foram processadas por uma rede *ResNet50* pré-treinada, extraíndo características que serão intituladas de características profundas. As características profundas encapsulam informações semânticas de alto nível e de representação das imagens de RMC. Estas características, radiômicas e profundas, são então fundidas em um vetor de características unificado e aplicado a um único bloco de autoatenção, e obtendo ao fim da arquitetura do modelo base, a classificação indicando se há ou não cardiomiopatia, ou seja, uma classificação binária.

Na segunda fase, diversas adaptações foram aplicadas ao modelo base com o intuito de aprimorar a robustez e a acurácia. Algumas das adaptações feitas consistem no aumento do tamanho dos vetores semânticos, também conhecidos na literatura como *embeddings*, seu tamanho foi variado também entre 12, 24, 48 e 64, valores testados de forma análoga a um hiperparâmetro, com a premissa de preservar mais das informações originais ao custo do aumento quase irrelevante de custo computacional.

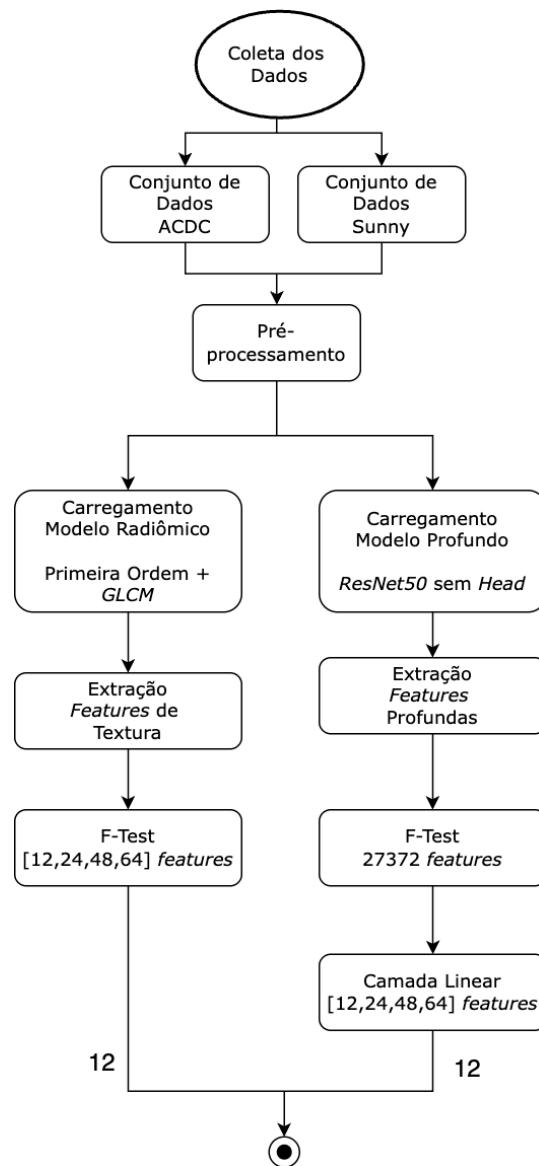
A concatenação dos *embeddings* também foi adaptada se comparada ao modelo original. Na versão original a concatenação se dá na última dimensão enquanto que nas versões adaptadas se dá na primeira dimensão. Em termos práticos, na versão original dois vetores 1x12 concatenados resultam em 1x24, na abordagem adaptada resultam em 2x12.

A mudança na forma de concatenar é feita de forma proposital pois a versão adaptada conta com uma camada convolucional inédita se comparada com a versão original, visando extraír informação relevante da junção dos *embeddings*. A versão adaptada adiciona também uma camada de atenção seletiva no formato de um bloco SE que precede o bloco de autoatenção. Outra mudança na versão adaptada é o bloco de autoatenção ser aplicado  $N$  vezes ao invés de uma única vez como acontece no modelo base trazendo uma perceptível melhora nos resultados.

Como o número de vezes que um bloco é executado aumenta respectivamente o tamanho da rede, será testado com alguns valores numéricos de forma incremental para dosar os resultados e evitar sobreajuste. Lembrando que ambos mecanismos de atenção tem propostas diferentes, enquanto o bloco de atenção seletiva atua na relação entre os canais dos *embeddings*, o bloco de autoatenção otimiza e pondera o vetor de características fundidas de forma eficaz. A intuição de aplicar mecanismos de atenção com propósitos diferentes vem do trabalho de Yang *et al.* (2024).

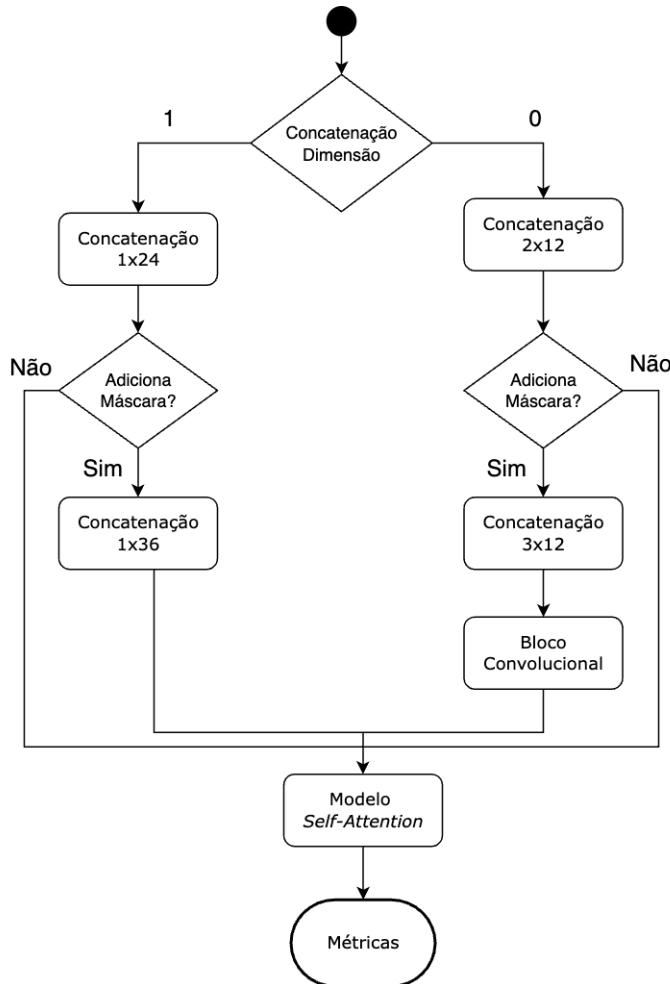
A Figuras 15 e 16 demonstram o fluxograma do projeto na fase de pré-processamento e o fluxograma da execução da arquitetura proposta respectivamente.

Figura 15 – Fluxograma - Pré-Processamento e Arquitetura



Fonte: Autor

Figura 16 – Fluxograma - Continuação



Fonte: Autor

#### 4.1 BASES DE DADOS

Duas bases de dados foram utilizadas no experimento: o primeiro conjunto de dados é o *automated cardiac diagnosis challenge* (ACDC)<sup>1</sup>. O segundo conjunto de dados é o SunnyBrook Cardiac Data (SCD) <sup>2</sup> (Radau *et al.*, 2009). Ambos conjuntos de dados são públicos e destinados a pesquisa.

O conjunto de dados ACDC é composto por 150 imagens sendo 100 para treino e 50 para testes distribuídas em 5 classes igualmente distribuídas: *dilated cardiomyopathy* (DCM), *hypertrophic cardiomyopathy* (HCM), *normal* (NOR), *myocardial infarction* (MINF) e *right Ventricular abnormality* (RV). O conjunto de dados SunnyBrook é composto por 45 imagens sendo e, respeitando as proporções determinadas no ACDC, foram escolhidas 30 imagens para treino e 15 para testes. As classes são NOR, IC, IC-I e HIP. Detalhes das classes são descritas nas seções que seguem destinadas a cada conjunto de dados.

<sup>1</sup> <https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html>

<sup>2</sup> <https://www.cardiacatlas.org/sunnybrook-cardiac-data>

#### 4.1.1 Conjunto de Dados ACDC

O conjunto de dados ACDC foi criado a partir de exames clínicos reais adquiridos no hospital universitário de Dijon (França). Os dados adquiridos foram totalmente anonimizados e tratados de acordo com as regulamentações estabelecidas pelo comitê ético local do hospital de Dijon. O conjunto de dados cobre várias patologias bem definidas com casos suficientes para (1) treinar adequadamente métodos de aprendizado de máquina e (2) avaliar claramente as variações dos principais parâmetros fisiológicos obtidos a partir de cine-RM (em particular volume diastólico e fração de ejeção). O conjunto de dados é composto por 150 exames (todos de pacientes diferentes) divididos em cinco subgrupos distribuídos de forma igualitária, sendo quatro patológicos e um grupo de sujeitos saudáveis. Cada exame é composto por um número variável de imagens que compõem o volume do coração. Dentre as cinco classes distintas, tem-se: DCM, HCM, NOR, MINF e RV. As classes DCM e HCM são interpretadas como passíveis de cardiomiopatia (CAR) e as demais; NOR, MINF e RV; são interpretadas como o coração em condições sem cardiomiopatia, os detalhes das classes podem ser verificados na Tabela 4.

As máscara de segmentação também acompanham o conjunto de dados, para possíveis aplicações de segmentação. Os valores dos rótulos variam de 0 a 3 e representam voxels pertencentes ao fundo (0), à cavidade do ventrículo direito (1), ao miocárdio (2) e à cavidade do ventrículo esquerdo (3). As Figuras 17, 18 e 19 representam as imagens e suas máscaras das classes DCM, HCM e NOR. Nas máscaras é possível ver três níveis de tons de cinza respectivos a segmentação citada.

Tabela 4 – Classes ACDC -Descrição

Condição	Descrição
<b>DCM</b>	A cardiomiopatia dilatada é uma condição em que o coração se torna dilatado e não consegue bombear o sangue de forma eficiente.  <b>Característica:</b> O ventrículo esquerdo está dilatado e com função sistólica reduzida de acordo com seu tamanho original.
<b>HCM</b>	A cardiomiopatia hipertrófica é uma condição onde há um espessamento anormal do músculo cardíaco, especialmente do ventrículo esquerdo.  <b>Característica:</b> A parede do ventrículo esquerdo está significativamente espessada, o que pode restringir o volume sanguíneo e o fluxo de saída.
<b>NOR</b>	Esta classe representa corações normais (sem cardiomiopatia) sem qualquer condição patológica.  <b>Característica:</b> Estruturas e funções cardíacas normais, sem dilatações ou hipertrofias significativas.
<b>MINF</b>	O infarto do miocárdio, ou ataque cardíaco, ocorre quando o fluxo sanguíneo para uma parte do coração é bloqueado por um período suficiente para causar danos ao músculo cardíaco.  <b>Característica:</b> Áreas do coração mostram cicatrizes ou fibrose devido ao infarto anterior, frequentemente visível nas imagens de ressonância magnética.
<b>RV</b>	A hipertrofia do ventrículo direito é uma condição em que o ventrículo direito do coração está aumentado.  <b>Característica:</b> Espessamento da parede do ventrículo direito, que pode ser devido a diversas condições, como hipertensão pulmonar ou doenças congênitas do coração.

Fonte: Adaptado de (Bernard *et al.*, 2018)

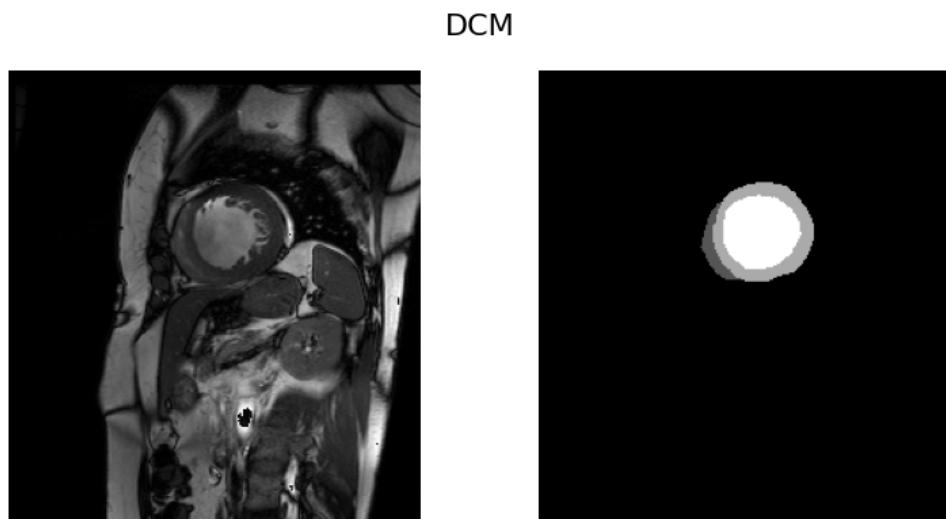
Os dados tem composição balanceada e sua distribuição pode ser conferida na Tabela 5. É possível notar, que após o agrupamento das classes como especificado, estas tornam-se minimamente desbalanceadas, em uma proporção de 40/60. As informações de cada paciente também é composta pelas seguintes informações: peso, altura, bem como o instante da fase diastólica e sistólica (Bernard *et al.*, 2018) que neste trabalho não são utilizadas dado o fato de ser tratado apenas as imagens médicas.

Tabela 5 – Classes do ACDC

<b>Grupo</b>	<b>Quantidade</b>	<b>C/ CM</b>	<b>S/ CM</b>
NOR	30	0	30
DCM	30	30	0
HCM	30	30	0
MINF	30	0	30
RV	30	0	30
<b>Total:</b>	<b>150</b>	<b>60</b>	<b>90</b>

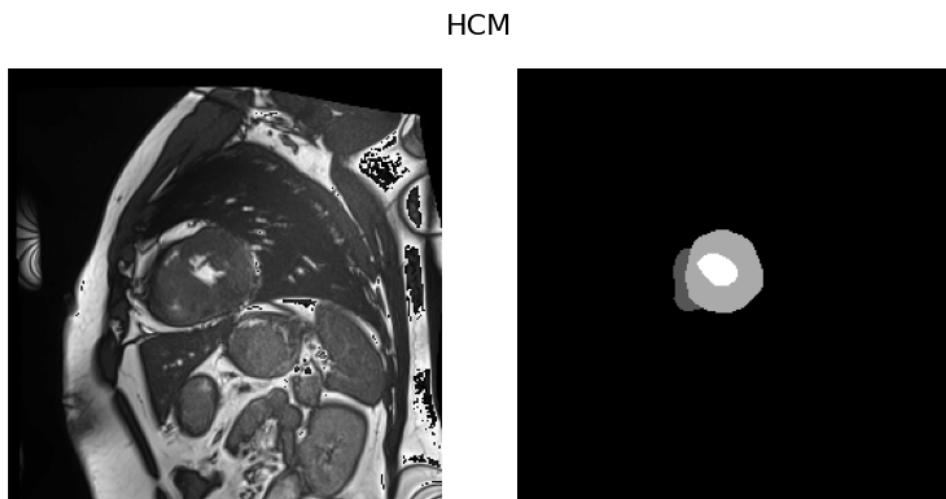
Fonte: Autor

Figura 17 – Imagem médica da classe DCM e sua respectiva máscara



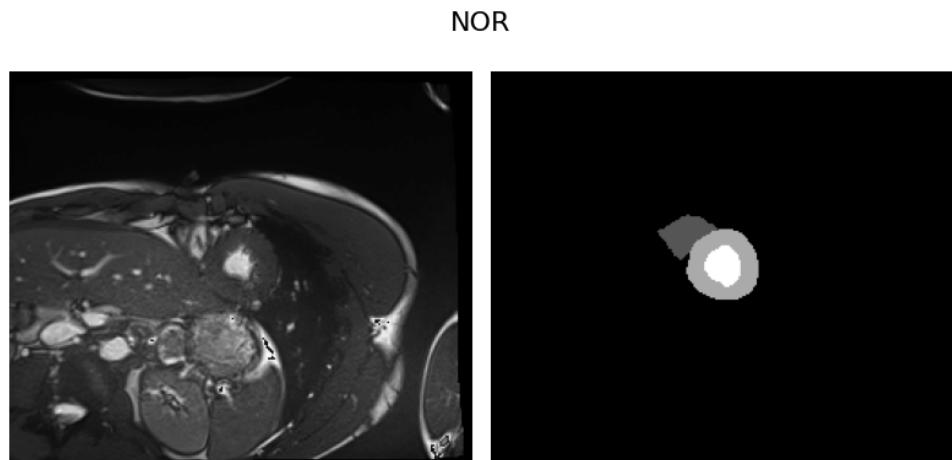
Fonte: Autor

Figura 18 – Imagem médica da classe HCM e sua respectiva máscara



Fonte: Autor

Figura 19 – Imagem médica da classe NOR e sua respectiva máscara



Fonte: Autor

#### 4.1.2 Conjunto de Dados Sunny

O conjunto de dados Sunny, também conhecidos como *2009 Cardiac MR Left Ventricle Segmentation Challenge data*, consiste em 45 imagens de RM de um grupo misto de pacientes e patologias: saudáveis, hipertrofia, insuficiência cardíaca com infarto e insuficiência cardíaca sem infarto. Um subconjunto desses dados foi utilizado pela primeira vez no desafio de segmentação automatizada do miocárdio a partir de imagens de RM em eixo curto, realizado em um workshop do MICCAI em 2009 (Radau *et al.*, 2009).

Há quatro grupos patológicos neste conjunto de dados, que foram classificados no seguinte formato:

1. Insuficiência cardíaca com infarto (**IC-I**): grupo com *fração de ejeção* (FE) < 40% e evidência de realce tardio com gadolínio (Gd);
2. Insuficiência cardíaca sem infarto (**IC**): grupo com FE < 40% e sem realce tardio com gadolínio;
3. Hipertrofia do ventrículo esquerdo (**HIP**): grupo com FE normal (> 55%) e uma razão de massa do ventrículo esquerdo sobre a área da superfície corporal > 83 g/m<sup>2</sup>;
4. Sem cardiomiopatia (**NOR**): grupo com FE > 55% e sem hipertrofia.

Para o presente trabalho foram utilizadas as classes **HIP** e **NOR** para classificação de cenário com e sem cardiomiopatia respectivamente. A Tabela 6 demonstra como se encontram distribuídos os 45 pacientes e os valores relacionados as atividades cardiovasculares utilizados para fins de classificação dos pacientes:

Tabela 6 – Estatísticas dos volumes e função do ventrículo esquerdo (média).

	<b>NOR (n=9)</b>	<b>HIP (n=12)</b>	<b>IC (n=12)</b>	<b>IC-I (n=12)</b>
<b>Vol. Diastólico Final (ml)</b>	115.69 (36.89)	114.39 (50.46)	233.67 (63.21)	244.92 (86.02)
<b>Vol. Sistólico Final (ml)</b>	43.10 (14.74)	43.11 (24.50)	158.28 (56.34)	174.34 (90.64)
<b>Fração de Ejeção (%)</b>	62.93 (3.65)	62.72 (9.22)	33.09 (13.07)	32.01 (12.27)
<b>Massa Ventrículo Esq. (g)</b>	130.27 (32.69)	175.87 (85.70)	193.69 (39.01)	201.32 (45.24)

Fonte: Adaptado de (Radau *et al.*, 2009).

#### 4.2 PRÉ-PROCESSAMENTO

Na fase de pré-processamento, foi utilizado o conjunto de dados ACDC. O ACDC originalmente tem cinco classes distintas e foi agrupado de forma que DCM e HCM indicam a classe CAR e os demais rótulos NOR, MINF e RV, são classificados como sem cardiomiopatia. O conjunto de *frames* utilizados são definidos na fase diástólica e a quantidade de *frames* é variada para cada paciente de acordo com consta no conjunto de dados. As características radiométricas são extraídas utilizando a biblioteca *PyRadiomics*<sup>3</sup>. Utilizando esta ferramenta, pode-se enviar como entrada o volume 3D das imagens cardíacas e suas respectivas máscaras e obter as características manuais extraídas abrangendo textura, forma, escala de cinza, etc. O resultado é um vetor com 78 características radiométricas. As características profundas são extraídas através de uma rede ResNet50 sem a última camada responsável pela classificação. Este processo é aplicado para cada fatia do volume 3D tanto para imagem cardíaca quanto para sua respectiva máscara tendo como resultado final vetores de tamanho 100.352.

Seguindo a receita do modelo base, é aplicado o *F-Test* para redução de dimensionalidade das características profundas de 100.352 para 27.372 e das características radiométricas de 78 para  $\text{EMBED}_{\text{size}} \in \{24, 48, 64\}$ . O valor de  $\text{EMBED}_{\text{size}}$  varia nas versões adaptadas, sendo considerado um hiperparâmetro, e garante que todos os *embeddings* possuam o mesmo valor. A concatenação é feita na última dimensão para o teste com modelo original e suas variações, e na primeira dimensão para as versões adaptadas. Vale ressaltar que diferente da versão do modelo base que só possui um vetor de características profundas, nas versões adaptadas pode-se ter um vetor adicional referente a máscara com a região de interesse.

#### 4.3 EXTRAÇÃO DE CARACTERÍSTICAS RADIÔMICAS

Foram extraídas características radiométricas da fase da diástole<sup>4</sup>, representada por um conjunto de fatias variando entre 6 e 18 *frames*, de cada paciente usando GLCM e estatísticas baseadas em histograma. Será aplicado o filtro *laplace of gaussian* (LoG) com cinco valores diferentes em cada parte para suavizar as imagens e realçar as bordas. Foram calculadas caracte-

<sup>3</sup> <https://pyradiomics.readthedocs.io>

<sup>4</sup> A diástole corresponde à parte do ciclo cardíaco caracterizada pelo relaxamento muscular e enchimento dos ventrículos (Brieler *et al.*, 2017)

rísticas GLCM como contraste, entropia, correlação, homogeneidade e energia para cada filtro LoG. Também são calculadas características de intensidade como média, variância, média dos percentis (10 e 90), desvio robusto da média absoluta, curtose e assimetria usando estatísticas de primeira ordem. Foram obtidos 78 características radiômicas para cada paciente dentro da quantidade de fatias extraídas na fase diastólica.

#### 4.4 EXTRAÇÃO DE CARACTERÍSTICAS PROFUNDAS

Para extrair características profundas das imagens de RMC, foi utilizada a arquitetura pré-treinada de um *ResNet50* sem sua última camada totalmente conectada e com os pesos congelados. A *ResNet50* é um modelo de rede neural convolucional profunda com 50 camadas que compreende muitos blocos residuais. Cada bloco contém módulos de convolução e uma conexão de salto que transfere a informação do bloco anterior para o próximo bloco. A conexão de salto ajuda a reter a informação semântica mais básica aprendida nas camadas anteriores, que de outra forma se tornaria abstrata devido à conexão de longa cadeia. A conexão de salto também evita o desaparecimento do gradiente nas camadas mais profundas ao fornecer um caminho alternativo para a retropropagação. A informação da conexão de salto é adicionada à informação calculada em cada bloco (Ai *et al.*, 2023). Ao todo são 100.352 características coletadas da saída deste modelo.

#### 4.5 CONCATENAÇÃO

Em posse das características radiômicas, é aplicado um *F-Test* tanto às 78 características radiômicas quanto às 100.352 características profundas, reduzindo cada um dos vetores ao espaço dimensional especificado. No método de fusão convencional, simplesmente é concatenado os dois vetores de características como na Equação 4.1, onde *Concat* simplesmente concatena os dois vetores. Unificando os vetores, o resultado é baseado na dimensão escolhida de concatenação escolhida. O resultado é enviado aos mecanismos de atenção, sendo este o bloco de autoatenção no modelo original e o bloco convolucional seguido do bloco de autoatenção no modelo adaptado.

$$F_{hd} = \text{Concat}(F_h, F_d, \text{dim}) \quad \text{dim} \in \{0, 1\} \quad (4.1)$$

#### 4.6 MÓDULO DE ATENÇÃO SELETIVA

A atenção seletiva desempenha um papel crucial na melhoria do desempenho de modelos de aprendizado profundo ao permitir que eles foquem em características mais relevantes nas representações internas. Especificamente, o bloco SE é projetado para recalibrar adaptativamente os canais das representações, atribuindo pesos maiores às características mais informativas. Esse mecanismo opera em duas etapas principais: primeiro, reduz a dimensionalidade das características globais (*squeeze*) para capturar o contexto geral, e, em seguida, aplica um mecanismo de excitação (*excitation*) para ajustar os pesos de cada canal com base na importância relativa. Ao

integrar blocos SE em uma arquitetura, o modelo pode aprimorar sua capacidade de discriminar padrões relevantes e suprimir ruídos ou informações irrelevantes, resultando em maior precisão e eficiência no processamento dos dados (Hu; Shen; Sun, 2018).

Sendo o bloco SE adicionado apenas na versão adaptada, visando aprimorar os resultados, o vetor resultante da concatenação é  $3 \times 64$  se assumirmos  $\text{EMBED}_{size}$  igual a 64 e a adição da máscara. Uma camada convolucional 1D é aplicada, com número de canais igual a 16, *kernel* igual a 3, *stride* igual a 1 e *padding* igual a 1 para não diminuir as dimensões após o processamento. Tendo como resultado da primeira convolução 16 canais, é aplicado o bloco SE para identificar os canais mais relevantes e escalar a entrada original. O bloco SE possui o parâmetro  $r$  de redução de dimensionalidade a ser configurado e foi utilizado o valor 16 conforme os autores do artigo original sugerem. Outra convolução 1D é aplicada com um único canal, assim tem-se vetor resultante de uma única dimensão que pode servir como entrada para o próximo bloco, o de autoatenção.

#### 4.7 MÓDULO DE AUTOATENÇÃO

O mecanismo de autoatenção foi empregado com a responsabilidade de aprender a importância de cada característica e capturar suas dependências de longo alcance. Como ilustrado na Figura 20, o módulo de autoatenção é utilizado para mapear uma consulta ( $Q$ ), chave ( $K$ ) e valor ( $V$ ) para um valor de atenção. São utilizadas 24 características concatenadas  $F_{hd}$  no modelo original e  $\text{EMBED}_{size}$  características nas versões adaptadas. Do vetor resultante, é projetada cada característica em três matrizes aprendíveis: matriz chave  $K$ , matriz consulta  $Q$  e matriz de valor  $V$  pelo produto escalar com suas respectivas matrizes de peso  $W_Q$ ,  $W_K$  e  $W_V$ . Logo os valores  $Q$ ,  $K$ ,  $V$  são denotados como  $W_Q F_{gd}$ ,  $W_K F_{gd}$ ,  $W_V F_{gd}$  respectivamente, onde  $W_Q$ ,  $W_K$  e  $W_V$  representam a transformação linear para as matrizes  $Q$ ,  $K$  e  $V$ .

O módulo de autoatenção é definido na Equação 2.7, em que  $d_k$  é a dimensão do valor de  $K$ . Sem utilizar operações recorrentes ou convolucionais em específico, o módulo de autoatenção pode modelar as dependências de longo prazo entre as características de entrada. Este módulo calcula de forma adaptativa os pesos entre as características com base em sua importância e relevância, capturando de forma mais abrangente as associações entre características radiométricas e profundas. Tal processo realça a capacidade expressiva das características fundidas e permite que o modelo foque mais precisamente nas características mais informativas que auxiliam a diagnosticar imagens médicas e reduzir a influência de características irrelevantes nesta previsão. Além disso, o modelo pode alojar dinamicamente atenção a diferentes amostras de imagens de RMC. Essa flexibilidade permite com que o modelo se adapte melhor à representação de características de diferentes amostras, melhorando a precisão e a generalização.

A função de perda utilizada utilizada no modelo é a função de entropia cruzada binária, do termo *binary cross entropy* (BCE), que é calculada pela Equação. 4.2 onde  $\mathcal{L}_{bce}$  denota o BCE,  $N$  denota o número de imagens de RMC vistas em tempo de treinamento,  $r$  denota a classe alvo de CAR e  $\hat{r}$  o valor previsto pelo modelo de CAR, 1 indica indícios de CAR e 0 sua ausência.

$$\mathcal{L}_{bce} = -\frac{1}{N} \sum_{i=1}^N (r_i \ln \hat{r}_i + (1 - r_i) \ln(1 - \hat{r}_i)) \quad (4.2)$$

## 4.8 MÉTRICAS

Para medir o desempenho da solução serão aplicadas as seguintes métricas: acurácia, precisão, revocação expressas pelas Equações 4.3, 4.4 e 4.5 respectivamente.

Também foi calculada a AUC, que consistem em um gráfico com a habilidade de validar um classificador binário. Esta curva é criada plotando, em gráfico, a taxa de verdadeiros positivos (TPR) contra a taxa de falsos positivos (FPR) em diversos limites contínuos de decisão. A equação da AUC é apresentada na Equação 4.6 em formato contínuo embora seja prática comum discretizá-lo.

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

$$Precisão = \frac{TP}{TP + FP} \quad (4.4)$$

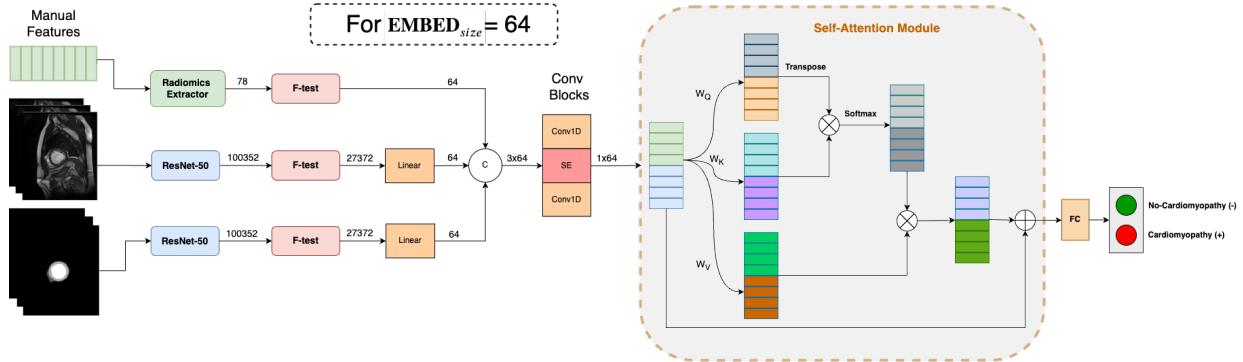
$$Revocação = \frac{TP}{TP + FN} \quad (4.5)$$

$$AUC = \int_{x=0}^1 TPR(FPR^{-1}(x)) \, d(x) \quad (4.6)$$

## 4.9 ARQUITETURA PROPOSTA

O esquemático da arquitetura proposta é observado na Figura 20. As características profundas são extraídas das imagens e máscaras via ResNet50 e sofrem uma transformação linear com o valor de 27.372 conforme trabalho original. As características radiômicas são extraídas via aplicação de diversos modelos estatísticos resultando no valor de 78. Um conjunto de F-Test é aplicado nos vetores para garantir que possuam o mesmo tamanho, tamanho este decidido pelo parâmetro  $EMBED_{size}$ . As características selecionadas são concatenadas e enviadas ao módulo de convolução composto respectivamente por: convolução 1D, bloco SE e convolução 1D. A primeira convolução converte o número de canais para 16, o bloco SE mantém a dimensionalidade e a segunda convolução converte os canais de 16 para 1. O processo segue no módulo se autoatenção e, por fim, uma camada linear com um único neurônio faz a classificação binária. A arquitetura proposta neste trabalho tem como base o trabalho de (Ai *et al.*, 2023).

Figura 20 – Arquitetura proposta para classificação de cardiomiopatia utilizando características radiômicas e profundas. As características são processadas por testes-F, transformações lineares e concatenadas antes de passar por blocos convolucionais, caso a concatenação seja na dimensão da profundidade, e um módulo de autoatenção.



Fonte: Autor

#### 4.10 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Como considerações, destacam-se o uso de base de dados pública, com dados de pacientes incluindo o conjuntos de fatias que identificam a ação de sístole e diástole do coração. É aplicado pré-processamento do qual são extraídas as características radiômicas, características estas que analisam textura, níveis e variações dos tons de cinza incluindo diversas estatísticas como de primeira ordem e o GLCM. Ainda em fase de pré-processamento é extraída as características profundas oriundas do modelo de visão *ResNet50*, com os pesos treinados na base de dados da *ImageNet*.

A adaptação inicia ao se adicionar os *embeddings* das máscaras também ao modelo, poder variar o vetor tamanho dos vetores de características, pois no modelo original eram apenas 12. Os vetores são então extraídos via *F-Test*, concatenados e inseridos no novo bloco convolucional. Este bloco possui um bloco SE responsável pela atenção seletiva, dando importância aos canais mais pertinentes. Por fim o módulo de autoattenção, confere as relações intrínsecas das transformações anteriores, identificando as características mais determinantes independente de espacialidade e a classificação é apresentada.

## 5 RESULTADOS E DISCUSSÕES

Este capítulo traz informações relacionadas com as ferramentas utilizadas no experimento, como linguagem, bibliotecas utilizadas, entre outros. Assim como, informações sobre o consumo de dados das bases ACDC e SunnyBrook, fase de pré-processamento, inferência nos modelos implementados, sendo estes, modelo linha de base e derivados, e modelos adaptados. Por fim, são apresentados os resultados considerando as métricas apresentadas na seção 4.8 utilizando as bases de dados ACDC e SunnyBrook e cada uma das abordagens: o modelo base e o modelo proposto.

Por fim, este capítulo apresenta os resultados da prova de conceito realizada com os algoritmos desenvolvidos, tanto o modelo base quanto os modelos adaptados, com o intuito de compará-los aos objetivos traçados nesta dissertação. Os testes realizados buscam avaliar a eficiência do algoritmo sob as diversas adaptações discutidas.

### 5.1 MATERIAIS

A Tabela 7 apresenta as ferramentas que foram utilizadas nesse projeto. Destaca-se que foram utilizadas majoritariamente ferramentas open-source como a linguagem *python*, bibliotecas como *pyradiomics*, *torch*, *numpy*, entre outras. Para registro dos experimentos, foi utilizada a ferramenta CometML<sup>1</sup>. No CometML é possível armazenar valores como: erro do lote no treinamento, acurácia do conjuntos de validação no treinamento, métricas resultantes como acurácia, hiperparâmetros como taxa de aprendizado, épocas, etc.

Tabela 7 – Fonte: Componentes Utilizados

Item	Descrição
Computador	<i>Macbook M1 Pro</i>
Memória	16gb
Versão <i>Python</i>	3.11.0
Versão <i>pyradiomics</i>	3.0.1
Versão <i>torch</i>	2.2.1
Versão <i>torchvision</i>	0.17.1
Versão <i>numpy</i>	1.26.4
Versão <i>scikit-learn</i>	1.4.1.post1
Versão <i>comet-ml</i>	3.47.4

Fonte: Autor

---

<sup>1</sup> <https://www.comet.com>

## 5.2 CENÁRIOS DE TESTE

Este trabalho visou realizar os experimentos em duas bases distintas com imagens de RMC, informações a cerca do paciente e rótulos que indicam presença ou não de cardiomiopatia. Testes iniciais foram feitos no conjunto de dados ACDC que possui 30 casos de CMH, 30 casos de cardiomiopatia dilatada (CMD) e 90 casos sem cardiomiopatia, resultando em 60 casos de cardiomiopatia e 90 casos sem cardiomiopatia. Da base ACDC se pretende utilizar apenas as fatias da fase diastólica. O segundo conjunto de dados é o *SunnyBrook*, possuindo imagens na fase diastólica contendo as classes 9 classes NOR (sem cardiomiopatia) e 12 HIP (hipertrofia do ventrículo esquerdo).

O modo de operação será o mesmo para ambos os conjuntos de dados com exceção do pré-processamento devido à forma de como o conjunto de dados *SunnyBrook* disponibiliza as máscaras.

Por fim, foram realizados três testes: 1) utilizando o modelo base no conjunto de dados do ACDC; 2) utilizando o modelo adaptado no conjunto de dados do ACDC; 3) utilizando o modelo adaptado no conjunto de dados do *SunnyBrook*. Todos os testes foram avaliados com as métricas acurácia, precisão, revocação e AUC em ambos os conjuntos de dados.

## 5.3 EXPERIMENTOS MODELO BASE

Uma prova de conceito foi aplicada ao conjunto de dados ACDC utilizando o modelo de linha de base para avaliação inicial, avaliação esta também utilizada como linha de base. O modelo base foi implementado seguindo o artigo e sua implementação é conferida na Figura 12. O conjunto de dados para treino é composto por 100 exames de pacientes, coletando apenas as fatias das imagens da fase diastólica. Características de primeira ordem e GLCM são extraídas, utilizando a biblioteca *PyRadiomics*, resultando em 78 valores que compõem as características radiométricas. Para extração das características profundas, foi utilizada uma rede *ResNet50* congelada sem sua última camada linear, responsável pela classificação originalmente de 1000 classes oriundas do conjunto de dados *ImageNet*, obtendo como resultado final um vetor com 100.352 características profundas.

Neste experimento foi treinado o modelo utilizando como função objetivo a entropia cruzada binária, taxa de aprendizado de 0,0001, otimizador *Adam*, tamanho de lote 1 e o treinamento se deu com aproximadamente 60 épocas, em que foi verificado que o erro se torna estável em tempo de treinamento. Também foi empregada a estratégia de aleatorizar as entradas no modelo em tempo de treinamento. Ao vetor resultante da saída do modelo é aplicada a função sigmoide, conforme Equação 5.1, função esta que limita os valores de sua entrada entre 0 e 1. Para fins de classificação, foi considerado valores maiores que 0,5 são considerados com cardiomiopatia e menores ou iguais a 0,5 são considerados sem cardiomiopatia.

$$\text{sigmoide}(x) = \frac{1}{1 + e^{-x}} \quad (5.1)$$

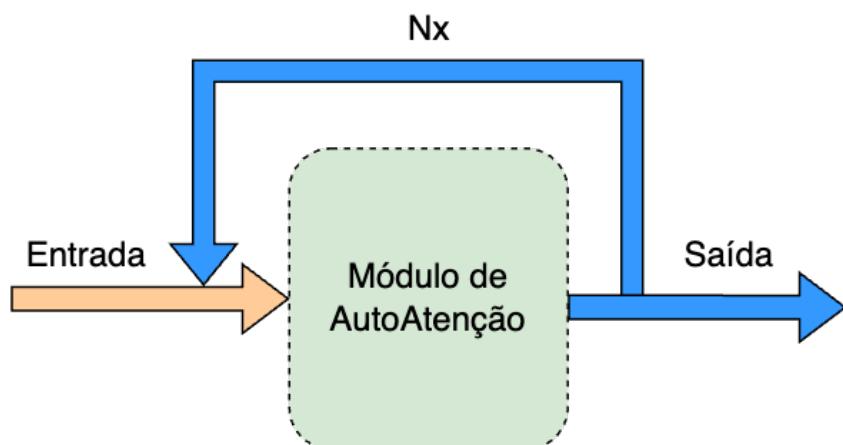
#### 5.4 EXPERIMENTOS MODELOS PROPOSTO

Os modelos adaptados refletem os experimentos desmembrados do modelo base, adaptações estas com a finalidade de verificar se mudanças, seja nos hiperparâmetros, seja em partes da arquitetura, podem trazer resultados promissores em relação aos resultados base. As mudanças podem ser elencadas em:

1. Mudanças do  $\text{EMBED}_{size}$  em que foram testados os valores [24, 48, 64] na tentativa de preservar mais informações originais;
2. No caso das versões adaptadas, também há a adição do consumo das respectivas máscaras que são futuramente concatenadas as características radiônicas e profundas;
3. A concatenação é feita de forma diferente, assumindo por exemplo o valor de  $\text{EMBED}_{size}$  igual a 12, na versão original tem-se como resultado  $1 \times 24$  porém na versão adaptada essa concatenação é feita na primeira dimensão, resultando neste exemplo em  $2 \times 12$ . Isto se dá por apresentar um novo bloco ao modelo original o bloco convolucional, composto de convolução e blocos SE.

A parte restante da arquitetura segue de acordo com a original, exceto que também podemos variar  $N$  vezes o bloco de autoatenção, ou seja, sua saída volta sendo sua entrada  $N$  vezes e este valor de  $N$  também pode ser considerado como um dos hiperparâmetros do experimento. Os valores de  $N$  utilizados foram [1, 2, 4, 6]. Estas mudanças não mudam a estrutura da arquitetura inicial fazendo com que o módulo autoatenção seja executado  $N$  vezes conforme trabalho de (Vaswani *et al.*, 2017) e pode ser visualizada na Figura 21.

Figura 21 – Representação esquemática da recorrência do Módulo de Autoatenção. O diagrama ilustra o processamento recursivo na arquitetura proposta, recursividade esta que permite que as características aprendidas sejam refinadas ao longo de múltiplas iterações, aumentando a capacidade do modelo de capturar dependências complexas nos dados.

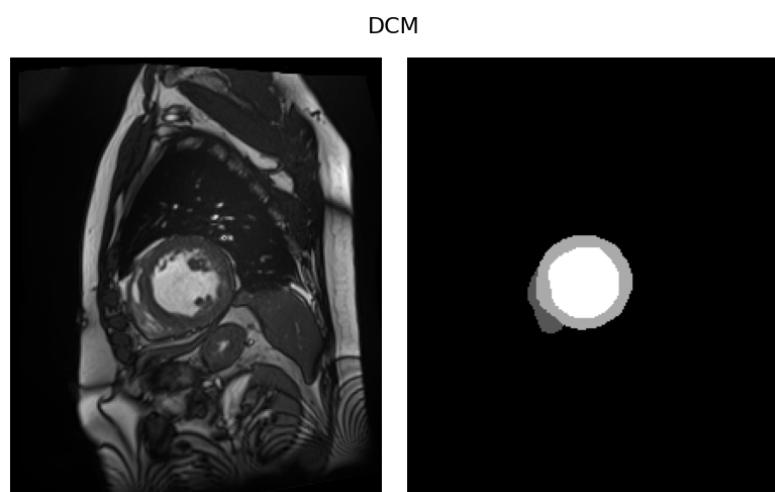


Fonte: Autor

## 5.5 RESULTADOS ACDC

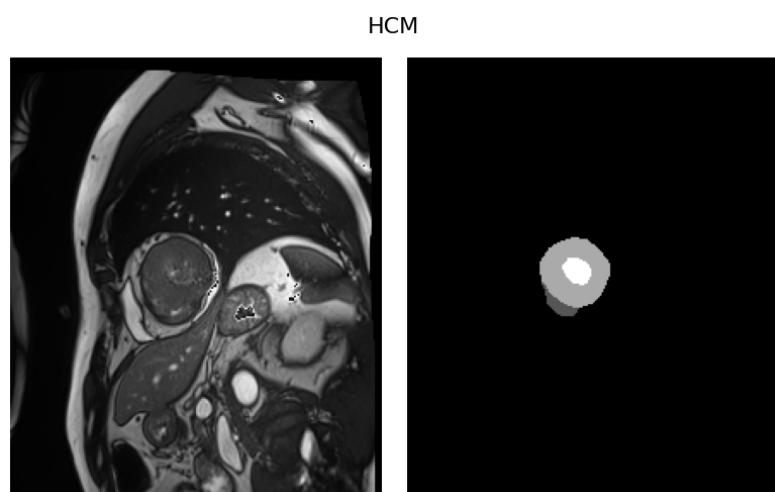
O conjunto de dados ACDC foi a referência inicial, principalmente para a coleta dos resultados com o modelo base. Este conjunto de dados público já se encontra separado com 100 exames para treino e 50 para testes. A Figura 22 e 23 são exemplos respectivamente, de imagens de DCM e HCM capturadas na diástole com suas respectivas máscaras, lembrando que ambas representam cardiomiopatia hipertrófica. Na Figura 24 temos a imagem de coração em estado sem anomalia (NOR) e sua respectiva máscara. As imagens demonstradas fazem parte do conjunto real de treinamento.

Figura 22 – Captura Diastólica CMD



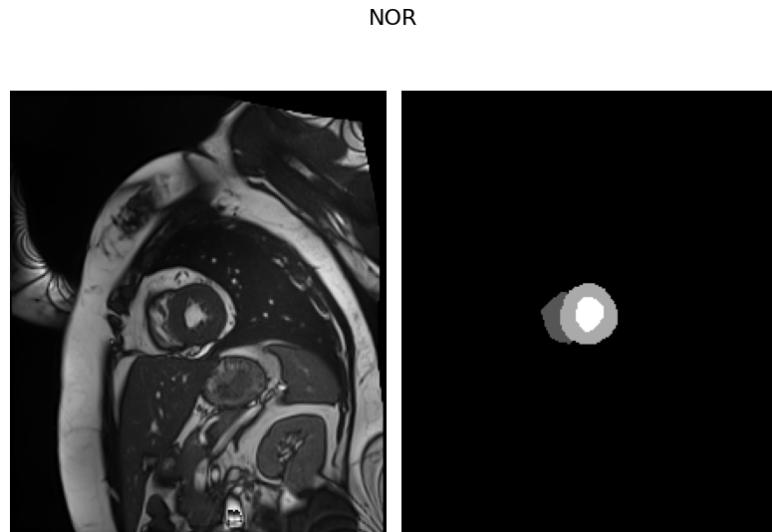
Fonte: Autor

Figura 23 – Captura Diastólica de CMH



Fonte: Autor

Figura 24 – Captura Diastólica NOR



Fonte: Autor

### 5.5.1 Resultados dos Modelos Base - ACDC

Com os dados pré-processados e previamente armazenados, o treinamento é efetuado com os vetores radiônicos e profundos. As métricas parciais foram sendo salvas em tempo de treinamento. Para o modelo base os hiperparâmetros são:  $\text{EMBED}_{size}$  igual a 12,  $N$  igual 1, ou seja, apenas um bloco de auto atenção, dimensão de concatenação 1, taxa de aprendizado 0,0001, otimizador Adam, tamanho de lote 1 e aproximadamente 60 épocas. Uma vez treinado, foi feita a inferência no conjunto de dados de teste, aplicando a função sigmoide e definindo 0,5 como o valor de corte onde os valores acima deste limite são classificados como CAR e os demais como sem cardiomiopatia.

As métricas resultantes podem ser conferidas na Tabela 8. A matriz de confusão é apresentada na Figura 25 e um gráfico ilustrando da *receiver operating characteristic curve* (ROC) é apresentado na Figura 26.

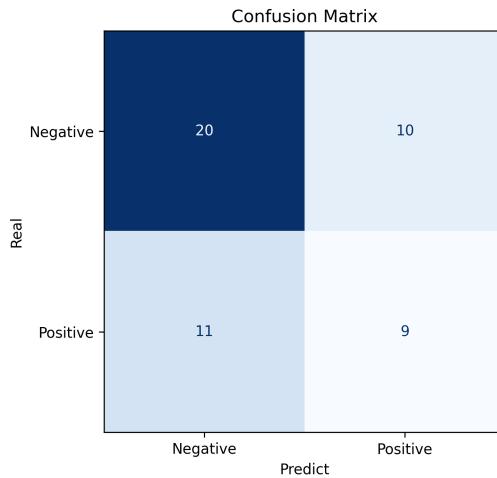
Tabela 8 – Métricas do Experimento - Modelo Base

Métrica	Valor
Acurácia	0.58
Precisão	0.47
Revocação	0.45
AUC	0.55

Fonte: Autor

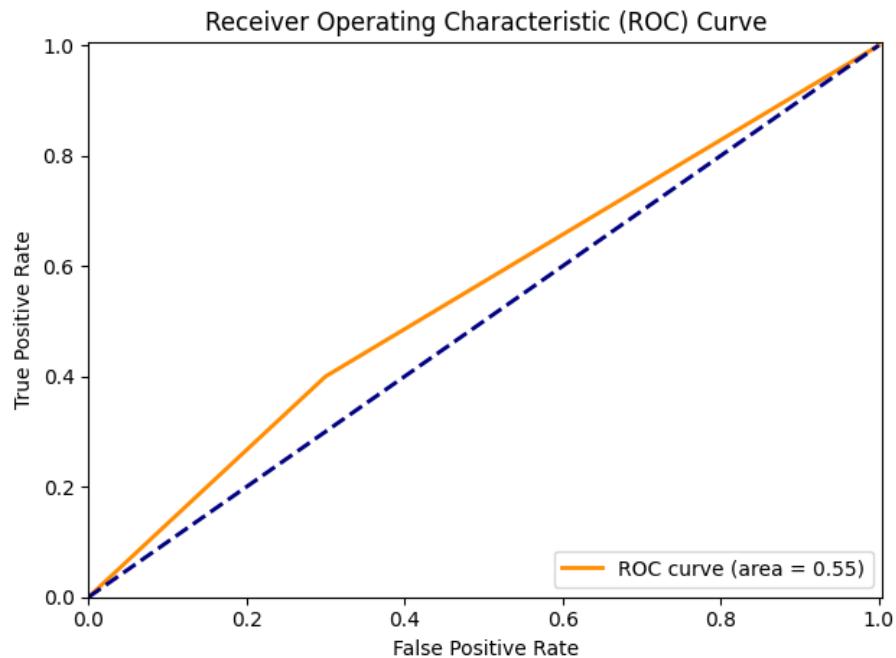
Para o registro dos resultados foi utilizada a ferramenta *CometML*. O *CometML* pode registrar, em tempo de execução, informações como erro por lote, erro por época, acurácia de validação, etc, podem ser armazenados enquanto o processo de treinamento é executado. Por

Figura 25 – Matriz de confusão do modelo base, apresentando os resultados da classificação binária. Observa-se 20 verdadeiros negativos, 9 verdadeiros positivos, 10 falsos positivos e 11 falsos negativos, totalizando 50 amostras com acurácia de 58%.



Fonte: Autor

Figura 26 – Curva ROC obtida pelo modelo proposto, ilustrando a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos.



Fonte: Autor

Finalmente métricas como acurácia e matriz de confusão podem ser armazenadas também. O serviço é acessado por uma *Application Programming Interface* (API) externa. As Figuras 27 e 28 demonstram respectivamente os painéis adaptáveis com alguns dos hiperparâmetros e métricas coletadas durante e ao terminar o treino. É possível notar que o modelo se sobre-ajusta nos dados de treino, com acurácia 1, predizendo corretamente todos os valores porém, nos dados

de teste a assertividade cai drasticamente para 0,58. Também é possível notar que após 2500 passos no treinamento, o erro de treinamento estabiliza e os demais processamentos não se fazem necessários.

Figura 27 – Painéis adaptáveis gerados pela plataforma *CometML* durante o treinamento do modelo. A visualização superior apresenta a evolução da função de perda (loss) ao longo dos passos de treinamento, com picos iniciais que se estabilizam no decorrer do treino. Os painéis inferiores mostram as métricas finais obtidas: acurácia de treinamento de 100% e acurácia de teste de 58% (valor 0.58),, evidenciando um possível sobreajuste do modelo.



Fonte: Autor

Figura 28 – Valores Coletados no Treino gerados pela plataforma *CometML* incluindo: função de perda (loss), acurácia de treino, e teste.

NAME	LAST			MIN			MAX		
	VALUE	STEP	TIME	VALUE	STEP	TIME	VALUE	STEP	TIME
loss	0.0269495	19019	12/24/04:17 AM	0	19010	12/24/04:17 AM	83.0093	2070	12/24/04:
test_acc	0.58	19019	12/24/04:17 AM	0.58	19019	12/24/04:17 AM	0.58	19019	12/24/04:
train_acc	1	19019	12/24/04:17 AM	0.67	950	12/24/04:16 AM	1	19019	12/24/04:

Fonte: Autor

### 5.5.2 Resultados dos Modelos Adaptados - ACDC

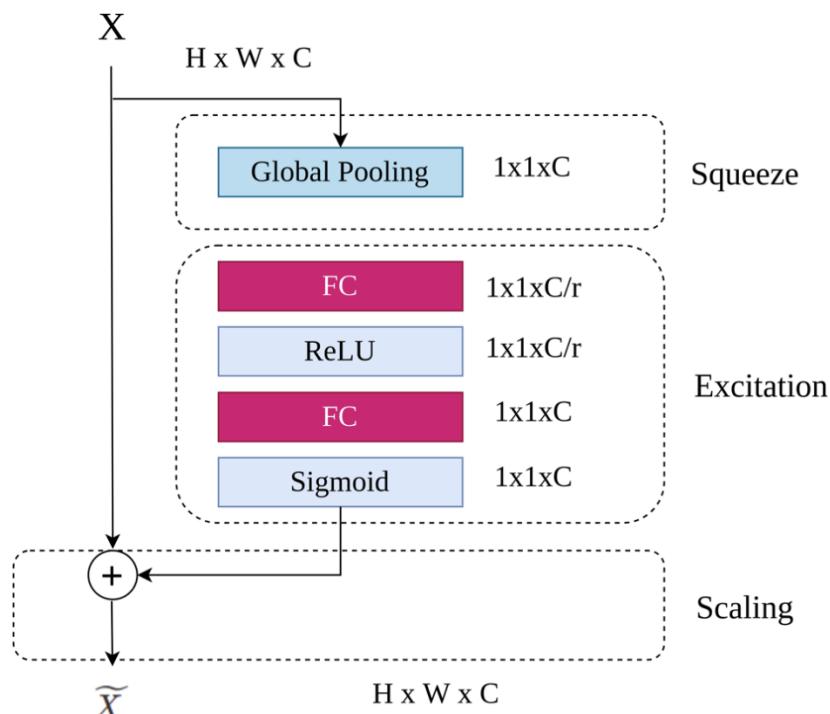
Para as versões adaptadas, se buscou mudar os hiperparâmetros e até mesmo a arquitetura do modelo com o intuito comparar os resultados com a versão base. Os primeiros experimentos foram aplicados variando o  $\text{EMBED}_{size}$ , também foi aplicado  $N$  vezes o bloco de autoatenção ao invés de uma única vez. Os experimentos foram executados gerando a combinatória destes hiperparâmetros para cobrir o máximo de cenários possíveis dentro da metodologia.

Também foi introduzido, nas versões adaptadas, um módulo convolucional, que antecede o módulo de autoatenção. Este módulo convolucional é composto de três blocos: uma camada

convolucionais 1D, um bloco SE e outro bloco convolucional 1D. Com a possibilidade de se ter que adicionar opcionalmente um terceiro vetor de características, sendo este a máscara, foi possível identificar relações intrínsecas entre mais este vetor. A concatenação agora na primeira dimensão faz cada vetor de característica se tornar um canal.

No módulo convolucional, uma primeira camada de convolução 1D aumenta o número de canais para 16, ou seja, 16 filtros são aplicados para extração de informação espacial. O bloco SE representa a camada de atenção seletiva que dá importância aos canais mais importantes, identificando a relevância de cada canal e finaliza escalando os canais na entrada original com um valor escalar otimizado em tempo de treinamento. Detalhes podem ser vistos na Figura 29. O bloco SE neste projeto é adaptado para ser 1D, diferente do trabalho original que visa sua aplicação em modelos de visão computacional já consolidados. O valor de  $r$  no bloco SE ficou fixo em 16, os autores do trabalho original fazem testes empíricos e confirmam que o valor de 16 é o que traz melhores resultados.

Figura 29 – Composição Bloco SE. O diagrama ilustra o fluxo de processamento em três etapas principais: Squeeze (compressão), que reduz a dimensionalidade espacial através de Global Pooling; Excitation (excitação), composta por duas camadas totalmente conectadas (FC) com funções de ativação ReLU e Sigmoid; e Scaling (escalonamento), que recalibra os mapas de características originais. Este mecanismo permite que a rede aprenda interdependências entre canais e atribua pesos adaptativos para realçar características relevantes.



Fonte: Adaptado de (Lafraxo *et al.*, 2024)

A Tabela 9 apresenta o conjunto de experimentos na versão original e versões adaptadas. A Tabela 10 demonstra os experimentos adicionando a máscara como um terceiro vetor de

características. A Tabela 11 apresenta a adição de blocos convolucionais e SE. Alguns pontos podem ser notados no experimento, o treinamento quase sempre sofre sobre-ajuste ficando muito próximo de 100% de assertividade, isso pode se justificado dado o fato de haver poucos exemplares para treinamento. Outro ponto notado é que quanto maior o tamanho do vetor de características (Emb) melhor costuma ser o resultado. Quantidades maiores no número de blocos de autoatenção (N\_Attn) também melhoraram os resultados, exceto no caso em que são configurados em 6. A justificativa se deve ao fato de tornar o modelo mais complexo se comparados a quantidade de dados limitadas para treino, prejudicando a sua capacidade de generalização. Os melhores resultados são conferidos nas versões que adicionam o bloco SE e com quatro blocos de autoatenção, sendo que o melhor experimento obteve 78% de assertividade.

Tabela 9 – Métricas ACDC - Adaptação do Modelo Original  
Negrito representa o modelo base

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
<b>24</b>	<b>1</b>	<b>0,58</b>	<b>0,47</b>	<b>0,45</b>	<b>0,46</b>	<b>0,55</b>
24	2	0,58	0,48	0,60	0,53	0,58
24	4	0,54	0,44	0,55	0,49	0,54
24	6	0,50	0,41	0,55	0,47	0,51
48	1	0,62	0,52	0,55	0,54	0,61
48	2	0,64	0,54	0,65	0,59	0,64
48	4	0,66	0,56	0,70	0,62	0,67
48	6	0,64	0,54	0,70	0,61	0,65
64	1	0,62	0,52	0,65	0,58	0,62
64	2	0,64	0,54	0,70	0,61	0,65
64	4	0,66	0,57	0,65	0,60	0,66
64	6	0,64	0,54	0,65	0,59	0,64

Fonte: Autor

### 5.5.3 Resultados dos Modelos Adaptados - *SunnyBrook*

Os experimentos no conjunto de dados *SunnyBrook* foram os mesmos aplicados ao conjunto de dados ACDC. No conjunto de dados *SunnyBrook*, houve uma etapa extra de pré-processamento para gerar as máscaras. Foi necessário interpolar os valores de um arquivo de texto que vem anexo com os dados. A máscara também é composta apenas pela região de interesse com valores entre 0 e 1.

Como o *SunnyBrook* tem poucos dados, apenas 45, sendo 30 para treino e 15 para teste e removendo as classes de infarto (IC e IC-I), sobram 21 registros sendo 13 para treino e 8 para teste. Foi considerado que o treinamento neste conjunto de dados não seria benéfico, pela quantidade limitada. Foi aplicada apenas a inferência considerando os 21 registros, sendo 12 casos de CMH e 9 casos em condições sem cardiomiopatia. Os modelos utilizados são os previamente treinados no ACDC. Os resultados são conferidos nas tabelas 12, 13 e 14. O modelo

Tabela 10 – Métricas ACDC - Adaptação do Modelo Original + Máscaras  
 Negrito representa maior assertividade

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
24	1	0,44	0,35	0,45	0,39	0,44
24	2	0,46	0,38	0,55	0,45	0,48
24	4	0,48	0,36	0,40	0,38	0,47
24	6	0,48	0,38	0,45	0,41	0,47
48	1	0,52	0,44	0,70	0,54	0,55
48	2	0,58	0,48	0,55	0,51	0,57
48	4	0,62	0,52	0,60	0,56	0,62
48	6	0,60	0,50	0,50	0,50	0,58
64	1	0,58	0,48	0,60	0,53	0,58
64	2	0,56	0,45	0,50	0,48	0,55
<b>64</b>	<b>4</b>	<b>0,64</b>	<b>0,54</b>	<b>0,65</b>	<b>0,59</b>	<b>0,64</b>
64	6	0,62	0,53	0,45	0,49	0,59

Fonte: Autor

Tabela 11 – Métricas ACDC - Modelos Adaptados - Blocos Conv. + SE  
 Negrito representa maior assertividade

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
24	1	0,66	0,56	0,70	0,62	0,67
24	2	0,68	0,58	0,75	0,65	0,69
24	4	0,70	0,59	0,80	0,68	0,72
24	6	0,68	0,59	0,65	0,62	0,67
48	1	0,70	0,58	0,90	0,71	0,73
48	2	0,70	0,62	0,65	0,63	0,69
48	4	0,74	0,65	0,75	0,70	0,74
48	6	0,70	0,60	0,75	0,67	0,71
64	1	0,64	0,54	0,65	0,59	0,64
64	2	0,72	0,64	0,70	0,67	0,72
<b>64</b>	<b>4</b>	<b>0,78</b>	<b>0,76</b>	0,65	0,70	0,76
64	6	0,76	0,65	<b>0,85</b>	<b>0,74</b>	<b>0,77</b>

Fonte: Autor

base consta em laranja e o melhor resultado consta em azul, indo de encontro aos resultados obtidos no ACDC. É possível notar que o conjunto de dados, por ser pequeno, compromete a classificação pois mesmo que o modelo possua um pequeno aprimoramento de uma arquitetura para outra, pode não ser suficiente para converter uma nova classificação, o que da a aparência de que alguns modelos possuem a mesma capacidade, o que não é a realidade.

Tabela 12 – Métricas SunnyBrook - Adaptação do Modelo Original

Negrito representa o modelo base

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
<b>24</b>	<b>1</b>	<b>0,48</b>	<b>0,54</b>	<b>0,58</b>	<b>0,56</b>	<b>0,46</b>
24	2	0,52	0,60	0,50	0,55	0,53
24	4	0,52	0,60	0,50	0,55	0,53
24	6	0,52	0,57	0,67	0,62	0,50
48	1	0,48	0,56	0,42	0,48	0,49
48	2	0,48	0,55	0,50	0,52	0,47
48	4	0,52	0,58	0,58	0,58	0,51
48	6	0,52	0,60	0,50	0,55	0,53
64	1	0,52	0,57	0,67	0,62	0,50
64	2	0,57	0,64	0,58	0,61	0,57
64	4	0,57	0,67	0,50	0,57	0,58
64	6	0,52	0,57	0,67	0,62	0,50

Fonte: Autor

Tabela 13 – Métricas SunnyBrook - Adaptação do Modelo Original Com Máscaras

Negrito representa maior assertividade

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
24	1	0,48	0,54	0,58	0,56	0,46
24	2	0,52	0,62	0,42	0,50	0,54
24	4	0,57	0,67	0,50	0,57	0,58
24	6	0,52	0,58	0,58	0,58	0,51
48	1	0,52	0,58	0,58	0,58	0,51
48	2	0,57	0,64	0,58	0,61	0,57
48	4	0,62	0,67	0,67	0,67	0,61
48	6	0,52	0,58	0,58	0,58	0,51
64	1	0,57	0,62	0,67	0,64	0,56
64	2	0,62	0,67	0,67	0,67	0,61
<b>64</b>	<b>4</b>	<b>0,71</b>	<b>0,80</b>	0,67	<b>0,73</b>	<b>0,72</b>
64	6	0,62	0,64	<b>0,75</b>	0,69	0,60

Fonte: Autor

Tabela 14 – Métricas SunnyBrook - Adaptação Adicionando Blocos Conv. e SE  
 Negrito representa maior assertividade

<b>Emb</b>	<b>N_Attn</b>	<b>Acc</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>AUC</b>
24	1	0,67	0,73	0,67	0,70	0,67
24	2	0,71	0,80	0,67	0,73	0,72
24	4	0,76	0,77	0,83	0,80	0,75
24	6	0,71	0,75	0,75	0,75	0,71
48	1	0,71	0,80	0,67	0,73	0,72
48	2	0,76	0,82	0,75	0,78	0,76
48	4	0,76	0,73	0,92	0,81	0,74
48	6	0,76	0,89	0,67	0,76	0,78
64	1	0,76	0,82	0,75	0,78	0,76
64	2	0,81	0,90	0,75	0,82	0,82
<b>64</b>	<b>4</b>	<b>0,86</b>	<b>0,85</b>	<b>0,92</b>	<b>0,88</b>	<b>0,85</b>
64	6	0,81	<b>0,90</b>	0,75	0,82	0,82

Fonte: Autor

#### 5.5.4 Testes Estatísticos

Para validar estatisticamente os resultados obtidos com os diferentes modelos avaliados, foi realizado um teste t de uma amostra (*one-sample t-test*) com valor de referência fixado em 0,58, correspondente à acurácia observada no modelo base. A Tabela 15 resume os valores médios de acurácia, desvios padrão, estatísticas t e p-valores associados a cada cenário avaliado: modelo base, conjunto ACDC e conjunto Sunny. O objetivo do teste foi verificar se as acurárias médias obtidas pelos modelos diferem significativamente do valor de referência, considerando um nível de significância  $\alpha = 0,05$ .

Os resultados mostram que todos os modelos apresentam diferenças estatisticamente significativas em relação ao valor de referência, uma vez que os p-valores obtidos são inferiores ao limiar de 0,05. O modelo base, apesar de apresentar média de acurácia próxima ao valor de referência (0,577), teve p-valor de aproximadamente  $2,10 \times 10^{-2}$ , suficiente para rejeitar a hipótese nula. Já os modelos ACDC e SUNNY apresentaram desempenhos substancialmente superiores, com acurárias médias de 0,78 e 0,83 respectivamente, além de t-estatísticas elevadas e p-valores extremamente baixos, reforçando a significância estatística dos ganhos obtidos com a abordagem proposta. Esses resultados indicam que o uso combinado de características radiométricas e profundas, aliado a mecanismos de atenção, resulta em modelos mais eficazes para a tarefa de classificação de cardiomiopatias.

Tabela 15 – Resultados dos testes estatísticos para acurácia dos modelos

Modelo	Média Acurácia	Desvio Padrão	T-Estatística	P-valor
BASE	0.58	0.01	-2.52	$2.10 \times 10^{-2}$
ACDC	0.78	0.00	106.80	$7.18 \times 10^{-28}$
SUNNY	0.82	0.03	41.50	$4.17 \times 10^{-20}$

Fonte: Autor

**Nota:** O valor de referência do teste ( $H_0$ ) é 0,58. Todos os modelos apresentaram **significância estatística** ( $p\text{-value} < 0,05$ ), indicando rejeição da hipótese nula.

## 5.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Os resultados obtidos, aplicando a arquitetura original proposta por Ai *et al.* (2023) e a aplicação de diversas modificações, como a forma de concatenar para propiciar um módulo convolucional de extrair informações inter-relacionadas entre características extraídas de diferentes técnicas, a adição de um mecanismo de atenção seletiva na forma de um bloco SE antecedente ao mecanismo de autoatenção, seguiram um padrão de resultados esperados. Conforme o modelo aumenta, seus resultados também melhoraram mas com ganhos mínimos dependendo do caso. Adicionar técnicas relevantes como o bloco SE aprimorou os resultados significativamente bem como o uso dos hiperparâmetros testados de forma empírica na busca de melhores resultados.

A atenção seletiva procura gerar um conjunto de pesos ao prever diretamente a importância de cada parte dos dados, aprendendo o valor dos pesos para cada canal no mapa de características com o objetivo de realçar ou suprimir as características de cada canal. O mecanismo de autoatenção envolve a utilização da correlação entre os conteúdos das camadas ocultas para calcular os pesos de atenção, que são então aplicados para agregar informações das camadas ocultas em uma escala global. Seguindo a intuição de utilizar diferentes mecanismos de atenção, conforme proposto pelos autores (Yang *et al.*, 2024), tal abordagem possibilitou melhorias significativas comparadas ao modelo base.

Outras adaptações também surtiram efeito, como aumentar o tamanho das características utilizadas, aplicar mais blocos de autoatenção, a adição das máscaras, etc. Os resultados comparando o modelo base com as versões adaptadas também podem indicar que há arquiteturas que podem funcionar melhor em certos cenário como o caso do modelo que aplicado no contexto de pulmão obteve bons resultados porém no contexto de coração os resultados não representaram a mesma qualidade.

Os experimentos indicam que há espaço para exploração e melhorias, como por exemplo, outros mecanismos de atenção, podem trazer resultados ainda mais promissores para o âmbito de cardiomiopatia hipertrófica.

## 6 CONCLUSÃO

Este trabalho apresentou uma abordagem inovadora para a classificação de cardiomopatias utilizando imagens de RMC combinadas com mecanismos de atenção, promovendo a integração entre características radiômicas e profundas. A proposta foi construída com base em fundamentos teóricos sólidos, utilizando técnicas modernas de aprendizado profundo, como a arquitetura SE Net e mecanismos de autoatenção, para explorar características discriminantes presentes em imagens médicas.

Os resultados obtidos destacaram a eficiência do modelo proposto na identificação de padrões associados a cardiomiopatias, superando abordagens tradicionais em termos de acurácia e generalização. A fusão de informações radiômicas, que capturam detalhes texturais e estatísticos, com características profundas extraídas de redes convolucionais, mostrou-se eficaz para lidar com a complexidade intrínseca das cardiomiopatias. Além disso, o uso de mecanismos de atenção também permitiu priorizar as regiões mais relevantes das imagens, promovendo maior expressividade do modelo.

A utilização do mecanismo SE-ResNet e da atenção espacial no domínio de imagens médicas representou uma abordagem inovadora ao priorizar informações relevantes no mapa de características. Essa técnica não apenas melhorou o desempenho do modelo, mas também ofereceu uma forma mais intuitiva de explicar as decisões tomadas pela IA, um aspecto crucial para sua aceitação na prática clínica.

Embora os resultados sejam promissores, este trabalho também expôs limitações, como a necessidade de bases de dados mais robustas e diversificadas, além da dificuldade em interpretar os modelos de aprendizado profundo, um desafio recorrente na área médica. Contudo, o modelo proposto representa um avanço significativo na aplicação de inteligência artificial em diagnósticos médicos, contribuindo para o desenvolvimento de ferramentas mais precisas e acessíveis para o suporte à decisão clínica, especialmente no diagnóstico precoce e acompanhamento de cardiomiopatias. Essas ferramentas podem melhorar a eficiência dos profissionais de saúde, reduzindo o tempo de análise manual e permitindo intervenções mais rápidas e precisas.

O desenvolvimento e avaliação do modelo proposto trazem contribuições relevantes para a literatura de aprendizado profundo aplicado à medicina, ao demonstrar como características radiômicas e mecanismos de atenção podem ser combinados de forma eficaz.

Por fim, os achados desta pesquisa abrem caminhos para futuras investigações, como a aplicação de mecanismos de atenção em outras doenças cardíacas e a integração de dados multimodais, incluindo genômicos e clínicos, para uma abordagem ainda mais abrangente e personalizada na medicina. O uso de técnicas avançadas, como aquelas desenvolvidas neste estudo, reforça o papel da inteligência artificial como uma aliada indispensável na medicina de precisão.

Como trabalhos futuros, sugere-se a exploração da combinação de imagens médicas com dados clínicos e genéticos para o desenvolvimento de modelos multimodais, capazes de

fornecer diagnósticos mais holísticos e personalizados. Além disso, o uso de técnicas como o aprendizado federado pode ser investigado como forma de viabilizar a colaboração entre instituições preservando a privacidade dos dados sensíveis.

Outra possibilidade é aplicar a abordagem proposta em diferentes conjuntos de dados e domínios, especialmente aqueles com características temporais mais complexas ou diferentes níveis de granularidade. Nesse contexto, a utilização de modelos convolucionais 3D, como a ResNet-3D e variantes similares, pode ser avaliada para melhor capturar padrões espaciais e temporais em volumes médicos tridimensionais. Também é recomendável investigar outras técnicas de seleção de variáveis e modelos de aprendizado de máquina, incluindo redes neurais profundas, transformers e métodos de ensemble mais sofisticados, com o objetivo de comparar o desempenho e robustez da solução.

Adicionalmente, seria relevante avaliar a escalabilidade do modelo em cenários com grandes volumes de dados, bem como sua integração em sistemas preditivos em tempo real. Por fim, uma investigação mais aprofundada sobre a interpretabilidade dos modelos gerados pode contribuir significativamente para sua aplicação prática em ambientes clínicos e de pesquisa.

## REFERÊNCIAS

- AI, Y. *et al.* A Self-Attention Based Fusion Model of Radiomics and Deep Features for Early Recurrence Prediction in NSCLC. In: **2023 IEEE 12th Global Conference on Consumer Electronics (GCCE)**. [S.l.: s.n.]: IEEE, 2023. p. 833–837. ISBN 9798350340181.
- ALZU'BI, A.; AMIRA, A.; RAMZAN, N. Content-based image retrieval with compact deep convolutional features. v. 249, p. 95–105, 2017. ISSN 09252312.
- ARGENTIERO, A. *et al.* The Applications of Artificial Intelligence in Cardiovascular Magnetic Resonance—A Comprehensive Review. 2022.
- BERNARD, O. *et al.* Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? v. 37, n. 11, p. 2514–2525, 2018. ISSN 0278-0062, 1558-254X.
- BRIELER, J. *et al.* Cardiomyopathy: An Overview. v. 96, n. 10, 2017.
- FOGEL, M. Assessment of Cardiac Function by Magnetic Resonance Imaging. v. 21, n. 1, p. 59–69, 2000. ISSN 1432-1971.
- GANTZ, J.; REINSEL, D. THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. 2012.
- HE, K. *et al.* **Deep Residual Learning for Image Recognition**. 2015.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-Excitation Networks. In: **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 7132–7141. ISSN 2575-7075. Disponível em: <https://ieeexplore.ieee.org/document/8578843/?arnumber=8578843>.
- IRANMEHR, A.; GUTTA, S.; HADAEGH, A. Improved Prediction of MGMT Methylation Status in Glioblastoma using a Deep Attention Network. In: **2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCMEE)**. [S.l.: s.n.]: IEEE, 2022. p. 1–6. ISBN 978-1-66547-095-7.
- JIANG, X. *et al.* MRI Based Radiomics Approach With Deep Learning for Prediction of Vessel Invasion in Early-Stage Cervical Cancer. v. 18, n. 3, p. 995–1002, 2021. ISSN 1557-9964.
- KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. [object Object], 2014. Disponível em: <https://arxiv.org/abs/1412.6980>.
- KWON, S. *et al.* Comparison of mortality and cause of death between adults with and without hypertrophic cardiomyopathy. Nature Publishing Group, v. 12, n. 1, p. 6386, 2022. ISSN 2045-2322. Disponível em: <https://www.nature.com/articles/s41598-022-10389-4>.
- LAFRAXO, S. *et al.* SEDARU-net: A squeeze-excitation dilated based residual U-Net with attention mechanism for automatic melanoma lesion segmentation. 2024. ISSN 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-024-20080-7>.
- LAMBIN, P. *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. v. 48, n. 4, p. 441–446, 2012. ISSN 09598049. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959804911009993>.

- LECUN, Y. *et al.* Handwritten Digit Recognition with a Back-Propagation Network. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.]: Morgan-Kaufmann, 1989. v. 2.
- MAYERHOEFER, M. E. *et al.* Introduction to Radiomics. v. 61, n. 4, p. 488–495, 2020. ISSN 0161-5505, 2159-662X.
- NAIR, V.; HINTON, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. 2010.
- O'SHEA, K.; NASH, R. **An Introduction to Convolutional Neural Networks**. 2015. Disponível em: <http://arxiv.org/abs/1511.08458>.
- PETERSEN, K. Guidelines for conducting systematic mapping studies in software engineering: An update. 2015.
- PONTONE, G. *et al.* Clinical applications of cardiac computed tomography: A consensus paper of the European Association of Cardiovascular Imaging-part II. v. 23, n. 4, p. e136–e161, 2022. ISSN 2047-2412.
- RADAU, P. *et al.* Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI. 2009. Disponível em: <https://www.midasjournal.org/browse/publication/658>.
- REN, J. *et al.* A Bi-LSTM and multihead attention-based model incorporating radiomics signatures and radiological features for differentiating the main subtypes of lung adenocarcinoma. v. 13, n. 7, p. 4245–4256, 2023. ISSN 2223-4292.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th. ed. [S.l.: s.n.]: Prentice Hall, 2020.
- SCHOFIELD, R. *et al.* Texture analysis of cardiovascular magnetic resonance cine images differentiates aetiologies of left ventricular hypertrophy. v. 74, n. 2, p. 140–149, 2019. ISSN 00099260. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0009926018305531>.
- TO, A. C.; DHILLON, A.; DESAI, M. Y. Cardiac Magnetic Resonance in Hypertrophic Cardiomyopathy. v. 4, n. 10, p. 1123–1137, 2011. ISSN 1936878X.
- VASWANI, A. *et al.* **Attention Is All You Need**. 2017.
- YANG, M. *et al.* Neural Network Design Based on Differentiable Attention Module Search. In: **2024 43rd Chinese Control Conference (CCC)**. [S.l.: s.n.], 2024. p. 8643–8648. ISSN 1934-1768. Disponível em: <https://ieeexplore.ieee.org/document/10662130/?arnumber=10662130>.