



UNIVERSIDADE D
COIMBRA

Luís André Mendes Lopes

**A ECOGRAFIA VERSUS ELASTOGRAFIA NO
ESTUDO DA SARCOPENIA**

**Dissertação no âmbito do Mestrado em Engenharia Biomédica, de
Instrumentação Biomédica orientada pelo Professor Doutor Jaime Baptista
dos Santos, co-orientada pelo Professor Doutor José Silvestre Silva e pela
Professora Doutora Alexandra André e apresentada ao Departamento de
Física da Faculdade de Ciências e Tecnologias da Universidade de Coimbra.**

setembro de 2023



FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

A ecografia versus elastografia no estudo da sarcopenia

Supervisor:

Professor Doutor Jaime Santos

Co-supervisors:

Professor Doutor José Silvestre Silva

Professora Doutora Alexandra André

Jury:

Professor António Dourado

Professor Fernando Manuel dos Santos

Professor Jaime Batista dos Santos

Dissertation submitted in partial fulfillment for the degree of Master of Biomedical
Instrumentation in Biomedical Engineering.

Coimbra, setembro de 2023

Agradecimentos

Agradeço aos Professores Jaime Santos, José Silva e Alexandra André pela inestimável ajuda e apoio durante todo o processo de elaboração desta dissertação. A sua orientação e disponibilidade foram verdadeiramente apreciadas.

Gostaria de agradecer à minha família pelo seu apoio e preocupação durante os anos que passei na universidade e por nunca deixarem de acreditar no meu futuro.

Um enorme obrigado a todos os meus amigos que me apoiaram ao longo dos últimos cinco anos, pelas memórias e momentos partilhados, que guardarei sem dúvida no meu coração.

Esta conquista não teria sido possível sem as contribuições coletivas daqueles que me encorajaram, partilharam as minhas aspirações e me proporcionaram uma rede de segurança e coragem.

Resumo

Devido aos seus efeitos sobre a saúde e bem-estar da população idosa, a sarcopenia, perda de massa e função do músculo esquelético relacionada com a idade, tem atraído mais atenção. Nesta investigação, é explorada a utilização de métodos de imagem avançados e abordagens de *machine learning* para a classificação da sarcopenia, utilizando imagens de ecografia e elastografia do músculo reto femoral.

Este estudo compara o desempenho de modelos de classificação convencionais e modelos de *deep learning* de ponta, avaliando a capacidade de cada modelo em identificar a sarcopenia em vários grupos etários num conjunto de dados diversificado de 180 imagens recolhidas de 30 pessoas com idades compreendidas entre os 20 e os 75 anos. O conjunto de dados foi submetido a um pré-processamento cuidadoso, incluindo métodos de aumento de dados, de modo a aumentar a sua diversidade e potencial de generalização. A nossa investigação incluiu redes neurais, *Support Vector Machine*, *Logistic Regression*, *Stochastic Gradient Descent Classifier*, e *Support Vector Machine*. Além disso, explorámos as potencialidades de modelos de *deep learning*, como o *DenseNet 121*, o *VGG 16*, o *VGG 19*, o *ResNet 50* e o *Inception V3*. Cada modelo foi ajustado e comparado exaustivamente com base no seu desempenho de classificação. As nossas conclusões mostram que os classificadores de *machine learning* igualaram e alguns até ultrapassaram os equivalentes de *deep learning* em termos de precisão de classificação. Notavelmente, a Rede Neuronal emergiu como o desempenho mais fiável entre os modelos, alcançando um notável *F1 score* de 99,81% para o conjunto de dados de ecografia.

Este resultado destaca a capacidade das arquiteturas de *deep learning* para identificar corretamente a sarcopenia em imagens de elastografia e ecografia. Além disso, o desempenho competitivo da rede neural levanta a possibilidade de que a alta precisão nem sempre exige modelos complexos de *deep learning*, principalmente quando se trabalha com conjuntos de dados limitados.

Esta dissertação oferece algumas conclusões sobre a classificação da sarcopenia utilizando uma vasta gama de modelos, contribuindo para o crescente corpo de investigação na

intersecção da imagiologia médica e de *machine learning*. A eficácia observada nos modelos convencionais e de *deep learning* realça a potencial utilização destes métodos para auxiliar os profissionais de saúde na deteção e intervenção precoces da sarcopenia. Esta descoberta tem possíveis implicações para melhorar a qualidade de vida dos indivíduos mais velhos mediante melhores ferramentas de diagnóstico e cuidados individualizados à medida que a população mundial continua a envelhecer.

Palavras chave: Sarcopenia, reto femoral, ecografia, elastografia, *machine learning*, *deep learning*

Abstract

Due to its effects on the health and well-being of the aging population, sarcopenia, the age-related loss in skeletal muscle mass and function, has garnered more attention. In this research, we investigate the use of cutting-edge imaging methods and machine learning approaches for the classification of sarcopenia, using ultrasound and elastography images of the rectus femoris muscle.

Our study compares the performance of conventional machine learning models and cutting-edge deep learning models, evaluating each model's ability to identify sarcopenia in various age groups within a diverse dataset of 180 images collected from 30 people ranging in age from 20 to 75. The dataset underwent careful preprocessing, including data augmentation methods, to increase its diversity and potential for generalization. Neural networks, Nu Support Vector Machine, Logistic Regression, Stochastic Gradient Descent Classifier, and Support Vector Machine were all included in our investigation. Furthermore, we explored the possibilities of deep learning models such as DenseNet 121, VGG 16, VGG 19, ResNet 50, and Inception V3. Each model was adjusted and thoroughly compared based on its classification performance. Our findings show that the machine learning classifiers matched and some even surpassed the deep learning equivalents in terms of exceptional classification accuracy. Notably, the Neural Network emerged as the most reliable performer among the deep learning models, achieving a remarkable F1 score of 99.81% in the ultrasound dataset.

This outcome highlights the capability of deep learning architectures to correctly identify sarcopenia in images from elastography and ultrasound. Additionally, the neural network's competitive performance raises the possibility that high accuracy may not always require complex deep learning models, particularly when working with limited datasets.

This dissertation offers insights into the classification of sarcopenia using a wide range of models, contributing to the growing body of research at the intersection of medical imaging and machine learning. The observed effectiveness of both conventional and deep learning models emphasizes the potential use of these methods in aiding medical practitioners in the

early detection and intervention of sarcopenia. This finding has possible implications for enhancing the quality of life for older individuals through improved diagnostic tools and individualized care as the world's population continues to age.

Keywords: Sarcopenia, rectus femoris, ultrasound, elastography, *machine learning*, *deep learning*

"Intelligence is the ability to avoid doing work, yet getting the work done."

Linus Torvald

Índice

Lista de Acrónimos	xiii
Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Motivação do estudo	1
1.2 Objetivo	2
1.3 Estrutura	3
1.4 Contribuições Científicas	3
2 Sarcopenia	4
2.1 Definição	4
2.2 Fisiopatologia	5
2.2.1 Impacto de diferentes definições a nível mundial	6
2.3 Fatores de Risco	7
2.4 Diagnóstico / Técnicas de imagiologia	7
2.4.1 Avaliação da Força Muscular	8
2.4.2 Avaliação do desempenho Físico-Funcional	9
2.4.3 Avaliação da massa muscular	10
2.5 Tratamento	16
2.5.1 Exercício físico	16
2.5.2 Nutrição e intervenções farmacêuticas	17
3 Classificadores	18
3.1 <i>Machine Learning</i> Clássico	19
3.1.1 Definição	19

3.1.2	Classificação de imagens médicas a nível muscular	20
3.2	<i>Deep Learning</i>	21
3.2.1	Definição	22
3.2.2	Redes Neurais Convolucionais (CNNs)	23
3.2.3	Classificação de imagens médicas a nível muscular	26
4	Metodologia	30
4.1	Pré-processamento e aumento de dados	31
4.1.1	Janela deslizante	32
4.1.2	Aumento de dados	33
4.2	Classificação tradicional	33
4.2.1	Extração de caraterísticas	33
4.2.2	Normalização de caraterísticas	37
4.2.3	Seleção de caraterísticas	38
4.2.4	Classificadores implementados	39
4.2.5	Otimização dos modelos	42
4.3	<i>Deep Learning</i>	43
4.3.1	Aprendizagem por transferência	43
4.3.2	Função de perda	44
4.3.3	Algoritmos de otimização	45
4.3.4	Programador da taxa de aprendizagem	46
4.3.5	Mecanismo de atenção	47
4.3.6	Modelos implementados	49
4.4	Métricas utilizadas	54
4.4.1	Exatidão	54
4.4.2	Precisão	55
4.4.3	Especificidade	55
4.4.4	Sensibilidade	55
4.4.5	<i>F1 score</i>	56
4.4.6	Área sob a curva (AUC)	56
5	Resultados	57
5.1	Implementação em <i>Python</i>	57
5.2	Conjuntos de dados	59
5.2.1	Participantes	59

5.2.2	Método de aquisição	59
5.2.3	Ecografia	59
5.2.4	Elastografia	60
5.2.5	Ecografia e Elastografia	60
5.3	Preparação dos dados	60
5.4	<i>Machine Learning</i> Clássico	61
5.4.1	Tamanho e passo no método janela deslizante.	63
5.4.2	Algoritmo de normalização de caraterísticas (<i>Standard Scaler vs MinMax Scaler</i>)	66
5.4.3	Algoritmo de seleção de caraterísticas (por variância <i>vs</i> informação mútua)	68
5.4.4	Número de caraterísticas selecionadas	70
5.4.5	Avaliação final	73
5.5	<i>Deep Learning</i>	75
5.5.1	Tamanho de grupo e número de épocas	76
5.5.2	Função de perda (perda de entropia cruzada <i>vs</i> perda focal)	80
5.5.3	Algoritmo de otimização (Adam <i>vs</i> Descida de Gradiente Estocástico)	82
5.5.4	Programador de taxa de aprendizagem (redução de passo <i>vs</i> redução de cosseno)	84
5.5.5	Mecanismo de atenção (atenção suave <i>vs</i> atenção de múltiplas cabeças)	86
5.5.6	Avaliação final	88
5.6	Comparação entre <i>machine learning</i> clássico e <i>deep learning</i>	91
6	Conclusões e Trabalhos Futuros	93
6.1	Conclusões	93
6.2	Trabalho Futuro	94
	Referências	96
A	Documento submetido à RECAPD	105
B	Conjuntos de dados criados pelo método de janela deslizante	108
C	Hiperparâmetros otimizados nos classificadores de <i>machine learning</i>	114

Lista de Acrónimos

Adam *Adaptive Moment Estimation* - Estimaco Adaptativa de Momentos

AMD *Advanced Micro Devices*

ANOVA *Analysis of Variance* - Anlise de Varincia

AUC *Area Under the Curve* - rea Sob a Curva

BIA *Bio-electrical Impedance Analysis* - Anlise de Impedncia Bioeltrica

CNN *Convolutional Neural Network* - Redes Neurais Convulsionais

CPU *Central Processing Unit* - Unidade Central de Processamento

DCNN *Deep Convolutional Neural Network* - Redes Neurais Convulsionais Profundas

DEXA *Dual X-ray Absorptiometry* - Absorciometria Bifotnica de Raios-X

DL *Deep Learning*

EWGSOP *European Working Group on Sarcopenia in Older People*

GM Gastrocnmio Medial

GL Gastrocnmio Lateral

GLCM *Gray-Level Co-occurrence Matrix* - Matriz de Coocorrncia de Nvel Cinzento

GLDM *Gray Level Dependence Matrix* - Matriz de Dependncia de Nvel Cinzento

GLRLM *Gray Level Run Length Matrix* - Matriz de Comprimento de Execuo de Nvel Cinzento

GLSZM *Gray Level Size Zone* - Matriz de Zona de Tamanho de Nvel Cinzento

GPU *Graphics Processing Unit* - Unidade de Processamento Grfico

IA *Inteligência Artificial*

ML *Machine Learning*

NGTDM *Neighbouring Gray Tone Difference Matrix* - Matriz de Diferença de Tom Cinzento Vizinho

pQTC *Peripheral Quantitative Computed Tomography* - Tomografia Computadorizada Quantitativa Periférica

RAM *Random-Access Memory* - Memória de Acesso Aleatório

ReLU *Rectified Linear Unit* - Unidade Linear Retificada

RF *Reto Femoral*

RFE *Recursive Feature Elimination* - Eliminação Recursiva de Características

RGB *Red Green Blue*, Vermelho Verde Azul

RM *Ressonância Magnética*

RN *Rede Neuronal*

ROC *Receiver Operating Characteristic* - Característica de Funcionamento do Recetor

ROI *Region of Interest* - Região de Interesse

SCPT *Stair Climb Power Test*

SGD *Stochastic Gradient Descent* - Descida de Gradiente Estocástico

SPPB *Short Physical Performance Battery*

SVM *Support Vector Machine*

TC *Tomografia Coputorizada*

TIFF *Tagged Image File Format*

TUG *Timed-Up and Go*

VL *Vasto Lateral*

Lista de Figuras

2.1	Exemplo de Dinamómetro Digital de Mão.	9
2.2	Exemplo de aparelho utilizado na Análise de Impedância Bioelétrica	10
2.3	Exemplo de aparelhos utilizados em Tomografia Computadorizada e em Tomografia Computadorizada Quantitativa Periférica.	13
2.4	Exemplo de aparelho utilizado na Ressonância Magnética	14
2.5	Exemplo de aparelho utilizado na DEXA	15
2.6	Exemplo de aparelho utilizado nas técnicas de ultrassons	16
3.1	Categorias dos modelos de <i>machine learning</i> de acordo com a natureza dos dados de entrada (adaptado)	20
3.2	Diagrama de <i>Venn</i> das classes e subcategorias de <i>machine learning</i> (adaptado) .	22
3.3	Arquitetura simples de uma CNN (adaptado)	23
4.1	Organização geral das etapas realizadas.	31
5.1	Localização do músculo reto femoral assim como de outras estruturas numa imagem ecográfica da perna.	61
5.2	Comparação entre as imagens obtidas pelo ecógrafo e a região de interesse extraída da imagem anterior	61
5.3	Desempenho dos cinco classificadores eleitos para o conjunto de dados A.	64
5.4	Desempenho dos cinco classificadores eleitos para o conjunto de dados B.	64
5.5	Desempenho dos cinco classificadores eleitos para o conjunto de dados C.	65
5.6	Desempenho dos cinco classificadores eleitos para diferentes métodos de normalização de caraterísticas.	67
5.7	Desempenho dos cinco classificadores eleitos para diferentes métodos de seleção de caraterísticas.	69

5.8	Desempenho dos cinco classificadores eleitos para diferentes números de características selecionadas para o conjunto de dados A.	71
5.9	Desempenho dos cinco classificadores eleitos para diferentes números de características selecionadas para o conjunto de dados B.	71
5.10	Desempenho dos cinco classificadores eleitos para diferentes números de características selecionadas para o conjunto de dados C.	72
5.11	Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados A.	77
5.12	Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados B.	77
5.13	Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados C.	78
5.14	Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 32 para o conjunto de dados A.	79
5.15	Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 32 para o conjunto de dados B.	79
5.16	Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 64 para o conjunto de dados C.	80
5.17	Desempenho dos cinco modelos eleitos utilizando diferentes funções de perda. . .	81
5.18	Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados A.	82
5.19	Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados B.	83
5.20	Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados C.	83
5.21	Desempenho dos cinco modelos eleitos utilizando diferentes programadores de taxa de aprendizagem.	85
5.22	Desempenho dos cinco modelos eleitos utilizando diferentes mecanismos de atenção.	87

Lista de Tabelas

2.1	Comparação de técnicas de imagem para o diagnóstico da sarcopenia (adaptado)	12
4.1	Estatísticas calculadas pela função “statistics”	34
5.1	Valores de <i>F1 score</i> para os 10 modelos selecionados, com <i>Bayesian optimization</i> , <i>10-fold cross-validation</i> durante 100 ciclos, para os conjuntos de dados N.º 35 das Tabelas B.1, B.2 e B.3.	62
5.2	Hiperparâmetros dos classificadores eleitos.	63
5.3	Número ótimo de características (<i>K</i>) selecionadas para cada um dos classificadores e conjuntos de dados.	73
5.4	Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados A.	73
5.5	Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados B.	74
5.6	Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados C.	75
5.7	Caraterísticas dos conjuntos de dados utilizados para o treino, validação e teste dos modelos de <i>deep learning</i> eleitos.	76
5.8	Desempenho em percentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados A.	88
5.9	Desempenho em percentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados B.	89
5.10	Desempenho em percentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados C.	89
5.11	Desempenho de cada um dos modelos eleitos de <i>machine learning</i> clássico (ML) e <i>deep learning</i> (DL), com a sua melhor configuração para todos os conjunto de dados.	91

B.1	Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para imagens de ecografia.	108
B.2	Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para imagens de elastografia.	110
B.3	Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para a união das imagens de ecografia com as imagens de elastografia.	112
C.1	Classificadores otimizados e os respectivos hiperparâmetros com o seu intervalo de valores.	114

1 Introdução

1.1 Motivação do estudo

A sarcopenia pode ser definida como a perda de massa e função muscular esquelética [1]. A definição inicialmente proposta na literatura descreve a condição como sendo exclusivamente associada ao envelhecimento dos indivíduos. Ainda que não deixe de ser verdade, existe outro entendimento mais recente de que a sarcopenia pode ser de origem secundária, ou seja, além do envelhecimento, afetando indivíduos adultos. Outras causas podem ser de carácter físico, como um estilo de vida sedentário, podem estar ligadas a doenças inflamatórias ou endocrinológicas, bem como a fatores nutricionais. Assim, o envelhecimento pode não ser causa única, mas sim, associado a outros fatores [2].

A patologia em causa resulta numa menor qualidade de vida, num maior encargo financeiro e num aumento das despesas médicas [3], originando repercussões globais na saúde pública e nos sistemas de saúde. A massa muscular e a força são essenciais para a realização das tarefas diárias, incluindo andar, levantar pesos e manter a postura e o equilíbrio. As pessoas afetadas sentem frequentemente limitações na mobilidade, na independência funcional e até no envolvimento social à medida que estas capacidades se deterioram. A patologia também aumenta a probabilidade de cair, sofrer outras lesões [4], necessitando de tratamento, internamento hospitalar e reabilitação. Além disso, um sistema músculo-esquelético comprometido pode dificultar o tratamento de doenças crónicas pré-existentes, aumentando a necessidade de cuidados médicos [3].

A abordagem dos efeitos da sarcopenia tornou-se crucial à medida que a esperança de vida aumenta. Urge, assim, garantir a prestação de cuidados oportunos, seguros e efetivos aos cidadãos em risco de, ou diagnosticados com, sarcopenia, desde a sua deteção precoce ao tratamento especializado e centrado na pessoa. Neste contexto, as tecnologias de imagiologia médica afiguraram-se enquanto instrumentos de apoio à tomada de decisão dos profissionais de saúde, contribuindo para uma avaliação rigorosa e informada da composição, estrutura e

características mecânicas do tecido muscular.

Na literatura internacional, a ecografia e elastografia emergem enquanto métodos não invasivos que permitem avaliar as propriedades muscular da pessoa. A elastografia permite aos profissionais avaliar a rigidez tecidual, um sinal de saúde muscular, enquanto a ecografia produz imagens de alta resolução que permitem a avaliação da área da secção transversal do músculo. Esta dissertação explora o tema da classificação da sarcopenia, utilizando duas técnicas de imagem e empregando o músculo reto femoral como alvo representativo.

A classificação e o tratamento da sarcopenia pode ser revolucionado pela utilização da inteligência artificial, em particular das abordagens de *machine learning* e de *deep learning*. Esta nova estratégia resolve problemas significativos provocados pela subjetividade, tempo necessário e disponibilidade por parte do profissional de saúde, proporcionando um método mais preciso, eficaz e fiável de classificação desta patologia [5]. Também as fronteiras geográficas podem ser ultrapassadas por sistemas de classificação de sarcopenia alimentados por IA, garantindo um diagnóstico normalizado e preciso em todo o mundo.

O domínio da ciência informática conhecido como inteligência artificial (IA) dedica-se ao desenvolvimento de modelos capazes de realizar atividades que tradicionalmente requerem a inteligência humana. Os modelos estão agora habilitados a interpretar eficientemente grandes conjuntos de dados. Embora existam muitos subdomínios da IA, a *machine learning* e *deep learning* são os subdomínios mais importantes na classificação de imagens médicas [6]. A *machine learning* utiliza uma variedade de algoritmos preparados para reconhecer características específicas através da aquisição de padrões a partir de conjuntos de dados [6].

1.2 Objetivo

Este projeto visa comparar dois métodos de imagiologia, ecografia e elastografia, no campo da classificação da sarcopenia. É também realizada uma comparação entre *machine learning* e *deep learning* para esta classificação específica. Este projeto teve origem na escassez de estudos que contrastem claramente a ecografia e a elastografia na classificação da sarcopenia. Embora cada um destes métodos de imagem tenha demonstrado ser promissor na avaliação da qualidade do músculo, a sua acoplação a modelos computacionais para a classificação de sarcopenia tem sido pouco explorada.

O objetivo final desta dissertação visa clarificar tanto qual a técnica de imagiologia como o modelo computacional que resultam num melhor desempenho na classificação da sarcopenia.

1.3 Estrutura

Esta dissertação é iniciada pelo capítulo 2 que apresenta um enquadramento teórico sobre a patologia, definindo a patologia. O capítulo 3 expõe um enquadramento teórico sobre os classificadores tanto de *machine learning* como de *deep learning*, introduzindo também exemplos das suas aplicação na classificação de imagens médicas. De seguida, o capítulo 4 visa a exposição dos métodos utilizados no projeto apresentando o decorrer do desenvolvimento prático e o capítulo 5 com uma exposição e discussão de todos os resultados práticos. Por fim vem o capítulo 6 onde são destacadas as principais conclusões do estudo e clarificadas as restrições que condicionaram o estudo. São, ainda, apresentadas recomendações para futuras iniciativas de investigação.

1.4 Contribuições Científicas

Os resultados e a análise do estudo foram submetidos à 29^a Conferência Portuguesa de Reconhecimento de Padrões (RECPAD), promovida pela APRP (Associação Portuguesa de Reconhecimento de Padrões), como parte deste projeto de investigação. Esta 29^a edição, em 2023, é organizada pelo Instituto Superior de Engenharia de Coimbra (ISEC) e pelo Instituto Politécnico de Coimbra (IPC). O documento submetido encontra-se no Apêndice A.

2 Sarcopenia

2.1 Definição

Inicialmente, a sarcopenia era descrita apenas como uma atrofia muscular, ou seja, redução da massa muscular. A definição desta condição física foi posteriormente atualizada incluindo a função muscular no seu conceito. Deste modo, o termo sarcopenia, renovado, foi introduzido pela primeira vez em 1988 por Rosenberg, correspondendo à perda de massa e função muscular esquelética que ocorre com o envelhecimento [1]. A referida perda muscular provoca uma diminuição da força, da taxa metabólica e da capacidade aeróbica, levando a uma inevitável redução da capacidade funcional [7].

Esta patologia possui uma manifestação mais recorrente em idosos, uma vez que é nesta faixa etária que se observa uma maior redução da quantidade e tamanho das fibras que formam os músculos, da produção de hormonas, como estrogénio e testosterona, no organismo, além da redução da prática de atividade física. Apesar da sarcopenia ser mais evidente e frequente em idosos, esta pode também surgir nas restantes faixas etárias, uma vez que a diminuição de força muscular e o conseqüente declínio da capacidade funcional pode estar associado a fatores genéticos, de estilo de vida e ainda a uma série de condições não exclusivas ao processo de envelhecimento. São exemplos dessas condições: desnutrição, doenças crónicas, inatividade física ou mobilidade limitada [8], por conduzirem à atrofia muscular. As conseqüências funcionais destas condições são mais acentuadas em idosos do que em jovens, uma vez que, em geral, nos jovens a massa muscular é superior, previamente à instalação das referidas condições [9]. A distinção entre sarcopenia primária, relacionada com o envelhecimento, e sarcopenia secundária causada pelas diversas condições mencionadas anteriormente é difícil de realizar na prática, uma vez que é frequente pacientes mais idosos possuírem múltiplas condições [8].

O *European Working Group on Sarcopenia in Older People* (EWGSOP) definiu ainda três diferentes estádios para a sarcopenia. Segundo o grupo, esta distinção auxilia a gestão clínica, permitindo assim uma melhor seleção de tratamentos e uma definição de metas terapêuticas

adequadas às necessidades e preferências das pessoas[2].

Os diferentes estádios são:

Pré-sarcopenia: Decréscimo da massa muscular sem perda da força muscular ou desempenho físico.

Sarcopenia: Decréscimo da massa muscular associado à perda da força muscular ou redução do desempenho físico

Sarcopenia grave: Quando os três critérios estão presentes, ou seja, decréscimo da massa muscular associado à perda da força muscular aliado a uma redução do desempenho físico.

O mesmo grupo ainda classificou a sarcopenia como sendo primária, quando a causa exclusiva é o envelhecimento e secundária quando é estabelecida outra causa. As causas podem ser de natureza física, como, por exemplo, levar um estilo de vida sedentário; podem estar relacionadas a doenças inflamatórias ou endócrinas e ainda estar relacionadas à nutrição. No entanto, o envelhecimento pode não ser a única causa, mas sim estar ligado a outras causas [2].

O conhecimento sobre os mecanismos moleculares que conduzem à atrofia muscular não é muito vasto. Porém, uma vez que, a sarcopenia está associada a uma perda de força muscular sem afetar a velocidade ou propriedades contrateis do músculo, deduz-se que na patologia são afetados os sarcómeros em paralelo e preservando o número de sarcómeros em série derivado de uma perda proteica [10].

Para além das consequências primárias da sarcopenia, a perda de força e função no músculo esquelético, a sarcopenia ainda apresenta uma vasta lista de consequências secundárias. Algumas destas consequências são: risco aumentado de quedas, fraturas, fragilidade e hospitalizações [4]. Também desencadeia a perda de neurónios motores, a perturbação da dinâmica proteica, a disfunção mitocondrial e a deterioração da regeneração muscular [11]. Por outro lado, existem intervenções promissoras, que resultam num melhoramento substancial, estas intervenções irão ser abordadas nos próximos capítulos.

2.2 Fisiopatologia

O músculo esquelético é constituído por dois tipos de fibras, fibras do tipo I e fibras do tipo II. As primeiras são consideradas fibras lentas, resistentes à fadiga devido à maior densidade de mitocôndrias, capilares e conteúdo de mioglobina. Já as fibras do tipo II são notórias pela sua resposta rápida em consequência de um potencial glicolítico mais elevado e menor capacidade oxidativa [12].

Durante uma atividade física de baixa intensidade a função muscular provém essencialmente da atividade exercida pelas fibras do tipo I. Por outro lado, em momentos de maior intensidade física, as fibras do tipo II são igualmente ativas. [12].

O envelhecimento do corpo é muitas vezes acompanhado com uma perda de massa muscular com o decréscimo do número de fibras dos dois tipos. Sucede, também, uma atrofia muscular que afeta maioritariamente as fibras do tipo II [12], resultando na redução do seu diâmetro. Não obstante às alterações morfológicas identificadas a nível da musculatura esquelética, verifica-se, igualmente, o aumento de tecido adiposo tanto entre grupos musculares, tecido adiposo intermuscular, como entre os fascículos musculares (tecido adiposo intramuscular). O aumento do tecido adiposo é consequência direta da redução da capacidade oxidativa das fibras musculares, conduzindo a uma diminuição gradual da força muscular das pessoas [11].

De igual modo, também os fascículos musculares atrofiam ao longo do processo de envelhecimento, um fenómeno acompanhado pela diminuição do ângulo de inserção dos tendões. Estas reduções implicam uma perda significativa de sarcómeros, tanto em paralelo, provocando um decréscimo no potencial de gerar força, como em série, provocando uma redução da velocidade de encurtamento muscular [13].

2.2.1 Impacto de diferentes definições a nível mundial

Não obstante à sua relevância e representatividade clínica, a literatura internacional não é consensual no que respeita ao diagnóstico clínico de sarcopenia, sendo possível identificar uma miríade de definições, critérios diagnósticos, instrumentos de avaliação e pontos de corte. [14]. Esta diversidade no processo de diagnóstico ocorre devido à falta de padronização na avaliação da massa muscular esquelética funcional [15].

Os pontos de corte definidos nos diferentes estudos carecem de robustez avaliativa, dado que a sua precisão diagnóstica é influenciada pelo fenótipo dos indivíduos e características socioculturais locais. Assim, os pontos de corte definidos para populações caucasianas são de difícil aplicação em populações asiáticas. Por exemplo, a população asiática pode apresentar uma quantidade de gordura corporal, substancialmente superior, face à população ocidental para o mesmo índice de massa corporal [16]. Argumenta-se que, devido às diferenças nas composições corporais entre as populações, os pontos de corte devem ser ajustados às necessidades e características locais [17].

2.3 Fatores de Risco

São inúmeros os fatores que aumentam consideravelmente o risco de sarcopenia, podendo ser divididos em duas categorias major: i) fatores imutáveis (e.g., sexo, idade); e ii) fatores alteráveis (e.g., hábitos alimentares, sedentarismo). No que diz respeito ao envelhecimento, S. von Haehling *et al.* salientam que a diminuição da massa muscular decorrente do processo de envelhecimento fisiológico inicia-se, sensivelmente, a partir da quinta década de vida [18]. Já B. H. Goodpaster *et al.* identificaram que o risco de sarcopenia difere entre sexos, verificando-se uma redução da força muscular mais acentuada nos homens, independentemente da existência de diferenças na força de base entre os sexos [19].

No que respeita aos fatores de risco alteráveis, a literatura converge amplamente na sua mitigação enquanto estratégia preventiva ou de intervenção clínica. O envelhecimento é acompanhado de um natural declínio nas necessidades energéticas [20], que poderão conduzir à perda de peso, massa muscular e força, aumentando assim o risco de sarcopenia. De igual modo, um padrão alimentar deficitário e não individualizado nas necessidades nutricionais de cada indivíduo configura um fator significativo de risco, uma vez que, há evidências do benefício do consumo de proteínas tanto na força como na função muscular [21]. J. Gianoudis *et al.* também revelaram uma ligação entre o comportamento sedentário e a sarcopenia, independente da atividade física realizada, concluindo que o risco de sarcopenia aumenta cerca de 33% por cada incremento de uma hora em tempo sem qualquer atividade física.

Vários fatores aumentam o risco de sarcopenia. A perda das placas terminais dos neurónios motores relacionadas com a idade, que interfere com a comunicação nervo-músculo e dificulta a ativação e o recrutamento muscular [22]. A redução da síntese proteica e do crescimento muscular aumenta a probabilidade de sarcopenia devido à diminuição das hormonas anabólicas e ao desenvolvimento de resistência à insulina [23]. A obesidade, pode aumentar o risco ao causar inflamação crónica, resistência à insulina e desequilíbrios hormonais [24].

2.4 Diagnóstico / Técnicas de imagiologia

Atualmente, reconhece-se que o diagnóstico de sarcopenia está associado a um número significativo de quedas, com sequelas graves para as pessoas e necessidade de admissão hospitalar recorrente. Num estudo realizado por Francesco Landi *et al.* concluíram que a taxa de risco de queda para um indivíduo com sarcopenia era 3,23 [25], isto significa que uma pessoa com sarcopenia tem 3,23 vezes mais probabilidades de cair do que uma pessoa saudável.

Globalmente, estima-se que cerca de 10% das pessoas mais velhas apresentem um

diagnóstico de sarcopenia. Acompanhando as tendências de envelhecimento demográfico verificados na população mundial, é expectável que estes números aumentem exponencialmente [26]. Urge, assim, garantir que os cidadãos têm acesso a cuidados de saúde oportunos, seguros e informados por evidência científica recente neste âmbito, salvaguardando a necessidade de se identificar precocemente fatores de risco de sarcopenia, assim como de monitorizar regularmente a evolução desta síndrome em indivíduos já diagnosticados.

Apesar da sarcopenia envolver tanto a perda de massa muscular como da sua função, o seu diagnóstico incide comumente na perda de massa muscular, dado ser este um dado objetivo, passível de ser avaliado transversalmente nos diversos contextos clínicos. Porém, considerando que a avaliação deste parâmetro pode estar associada a custos mais elevados [14], dada a necessidade de técnicas e ferramentas diferenciadas, alguns autores destacam como alternativas a avaliação de outros parâmetros concomitantes (e.g., antropometria, desempenho físico, força muscular), que passamos a apresentar de seguida.

2.4.1 Avaliação da Força Muscular

A avaliação da força muscular é identificada na literatura como uma abordagem diagnóstica simples, rápida e de baixo custo, passível de ser realizada na maioria dos contextos clínicos. Porém, no que respeita à sarcopenia, alguns autores destacam a inexistência de padronização na sua implementação em cortes mais velhas, assim como posterior classificação de risco. Para resultados mais precisos seria melhor realizar esta medição nos membros inferiores, porém nem sempre é o mais prático. A medição de prensão das mãos apresenta-se como um substituto fiável por ter uma boa correlação com a força da perna [27].

A força de prensão das mãos deve ser medida com um dinamómetro bem calibrado, devem também ser utilizados dados de população de referência adequados. Em casos singulares, em que não é possível a medição da força de prensão da mão, devido a outras condições médicas que incapacitem a mão em certa medida, é sempre possível realizar os métodos de torque isométrico aos membros inferiores, para a medição da força dos membros inferiores [28].



Figura 2.1: Exemplo de Dinamómetro Digital de Mão [29].

2.4.2 Avaliação do desempenho Físico-Funcional

Entende-se por desempenho físico como uma função relacionada com a mobilidade, passível de ser avaliada objetiva, e que decorre da atividade intrínseca e simultânea de vários órgãos e sistemas do corpo humano [28].

Alguns dos testes utilizados de modo a medir o desempenho físico-funcional, incluem: a velocidade de marcha; *Short Physical Performance Battery*(SPPB); *Timed-Up and Go* (TUG) e *Stair Climb Power Test* (SCPT).

O teste mais frequentemente utilizado, pela sua simplicidade, é a velocidade de marcha. Para este teste são apenas necessários um cronómetro e um pavimento liso e desprovido de obstáculos. O grupo EWGSOP aconselha uma distância de 4 m e define uma velocidade de corte de 0,8 m/s, caracterizando valores inferiores como um mau desempenho físico [28].

O teste SPPB envolve vários fatores, permitindo determinar o equilíbrio, a força, a marcha e a resistência muscular. Os teste incluídos são: a velocidade de marcha, teste ao equilíbrio e por fim o *chair stand test* [30].

Por outro lado, o teste TUG, permite avaliar tanto a mobilidade como o equilíbrio dos indivíduos. Este método requer que os participantes se levantem de uma cadeira, caminhem cerca de 3 m, caminhem de volta e se voltem a sentar [28].

Por fim, temos o SCPT, em que é cronometrado o tempo necessário para o paciente subir um lance de 10 escadas, o mais rápido possível. Este teste permite assim avaliar a potência muscular dos membros inferiores.

2.4.3 Avaliação da massa muscular

Existe uma grande variedade de técnicas de modo a realizar a avaliação da massa muscular. Nestas técnicas pode ser avaliada a quantidade de massa muscular como massa muscular esquelética total ou massa muscular esquelética apendicular. Por outro lado, pode ser caracterizada como a razão entre a massa magra apendicular e a altura ao quadrado.

Estes métodos diferem entre si na sua complexidade, na sua fiabilidade, exposição à radiação e custo total por exame. Na generalidade podemos agrupar todas as técnicas em três secções: análise de impedância bioelétrica (BIA), medidas antropométricas e técnicas de imagem corporal.

2.4.3.1 Análise de Impedância Bioelétrica (BIA)

A análise de impedância bioelétrica é uma técnica onde a medida da massa muscular é realizada indiretamente. No corpo humano, diferentes elementos possuem diferentes impedâncias elétricas, nomeadamente a água e tecidos isentos de gordura com uma baixa impedância e a massa gorda com uma alta impedância. Assim, a técnica BIA tira proveito deste conceito, fazendo passar uma pequena corrente elétrica pelo corpo. A impedância medida permite conhecer aproximadamente a quantidade de água e o rácio entre massa magra e massa gorda [31].

Devido ao seu modo de funcionamento, os resultados obtidos através desta técnica poderão estar a sobrestimar a massa gorda e conseqüentemente a massa muscular. Isto, por ser fortemente influenciada pelo nível de hidratação da pessoa. Por outro lado, enquanto vantagens desta técnica, destaca-se a sua simplicidade, baixo custo e segurança para a pessoa.



Figura 2.2: Exemplo de aparelho utilizado na Análise de Impedância Bioelétrica [32].

2.4.3.2 Medidas Antropométricas

A avaliação de medidas antropométricas afigura-se como uma das abordagens mais recorrentemente realizadas nos diversos contextos clínicos, sendo comumente realizada por diversos profissionais de saúde no acompanhamento da pessoa com risco de, ou diagnóstico, de sarcopenia. Entre os diferentes métodos envolvidos, destacam-se o cálculo do índice de massa corporal, a medição de perímetro abdominal, membros superiores e inferiores, assim como a medição da prega cutânea [31].

Não obstante à sua prática amplamente difundida e recorrente nos mais diversos contextos clínicos, o seu uso não é recomendado para o diagnóstico da sarcopenia, considerando o seu baixo nível de fiabilidade diagnóstico e inexistência de pontos de corte validados internacionalmente para a população mais velha. A margem de erro é acrescida pela variabilidade de avaliadores, não permitindo a deteção de variações ao longo de múltiplas medições ao longo do tempo. Também a prega cutânea apenas mede a gordura subcutânea e não a gordura visceral, a qual é um fator de risco independente para várias condições, como a diabetes, doenças cardiovasculares e neoplasias [31].

2.4.3.3 Técnicas de imagem corporal

As técnicas de imagem corporal permitem detetar a perda de massa muscular esquelética e diferem entre si na sua complexidade, no seu custo, na exposição a radiação e no seu grau de fiabilidade. Na literatura, destacam-se enquanto técnicas de imagiologia a: tomografia computadorizada, ressonância magnética, absorciometria bifotónica de raios-X e ultrassonografia. De forma geral, as características destas técnicas estão comparadas na Tabela 2.1, onde “-” significa a ausência do parâmetro na tecnologia e uma escala empírica de superioridade no parâmetro (desde “+” a “+++”).

Tabela 2.1: Comparação de técnicas de imagem para o diagnóstico da sarcopenia (adaptado) [33].

Tecnologias	Tomografia	Ressonância	Absorciometria	
	Computorizada	Magnética	bifotónica de raios-X	Ultrassons
Fiabilidade	+++	+++	++	+
Rapidez	+	+	+++	+++
Complexidade	+++	+++	+	++
Despesa	+++	+++	++	+
Exposição a raios-X	+++	-	+	-
Erro de estimativa da massa magra	<8.5%	6-8.5%	<5%	-
Campo de aplicação	Investigação	Investigação	Clínica / Investigação	Clínica

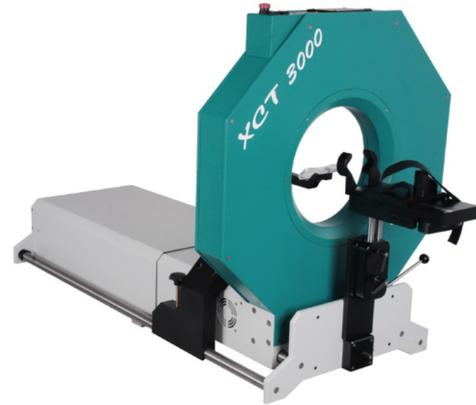
- **Tomografia Computadorizada (TC)**

A tomografia computadorizada é uma técnica imagiológica não invasiva, que consiste na captação de cortes transversais de diferentes quadrantes do corpo com raios-X. É de notar que devido aos níveis de atenuação para a densidade muscular, a tomografia computadorizada permite ainda identificar infiltrações de gordura nos músculos [33].

Deste modo, esta técnica permite analisar tanto alterações quantitativas como qualitativas na composição do corpo humano, isto é, avaliar tanto o volume do músculo esquelético assim como avaliar diretamente o conteúdo de gordura visceral abdominal [31]. A desvantagem mais significativa do uso deste método é a exposição a elevados níveis de radiação por parte da pessoa. De forma a contornar esta limitação, a tomografia computadorizada quantitativa periférica (pQTC) emerge na literatura como uma alternativa mais segura para a pessoa, uma vez que, ao utilizar um *scanner* de pequenas dimensões, as doses de radiação serão também menores [33]. Por outro lado, este método apresenta-se como dispendioso e de difícil acesso, conseqüentemente não é indicado para uso rotineiro, sendo mais utilizado em ambientes de investigação [31].



(a) Aparelho utilizado para TC [34]



(b) Aparelho utilizado para pQTC [35]

Figura 2.3: Exemplo de aparelhos utilizados em Tomografia Computadorizada e em Tomografia Computadorizada Quantitativa Periférica [34] [35].

- **Ressonância Magnética (RM)**

A ressonância magnética consiste na utilização de ondas de radiofrequência não ionizantes num campo magnético, permitindo criar uma imagem detalhada de segmentos/quadrantes do corpo humano.

Este método, assim como a tomografia computadorizada possibilita a avaliação da massa muscular esquelética da pessoa, distinguindo-a da gordura e/ou infiltração de gordura no tecido muscular. Este processo é conseguido analisando as oscilações na sequência dos pulsos de radiofrequência [33]. A ressonância magnética é ainda utilizada para estudar a composição das fibras musculares, apresentando-se como uma solução não invasiva [33].

Por fim, a ressonância magnética é vista de forma preferencial relativamente a outros métodos, uma vez que, não necessita de radiação ionizante e por ser um dos métodos mais precisos no estudo da composição corporal. No entanto, o seu elevado custo, falta de portabilidade, complexidade e inaplicabilidade em idosos portadores de aparelhos médicos mais antigos fazem com que esta técnica seja utilizada maioritariamente em ambientes de investigação [28], [33].



Figura 2.4: Exemplo de aparelho utilizado na Ressonância Magnética [36].

- **Absorciometria Bifotônica de Raios-X (DEXA)**

A DEXA baseia-se na atenuação de dois feixes de raios-X com diferentes níveis de energia de modo a conseguir fornecer uma estimativa face à densidade mineral óssea, massa gorda e massa sem gordura [33]. Esta análise pode ser realizada em todo o plano corporal ou, se necessário, num membro ou quadrante corporal específico. Apesar da medição de massa magra realizada por esta técnica incluir tecidos excluídos em outras técnicas, como a pele, tecidos conjuntivos e massa muscular esquelética, os seus resultados são consistentes e demonstram uma forte correlação com medidas antropométricas [33].

São reconhecidas enquanto desvantagens à DEXA a influência do estado de hidratação do corpo nas medições, a falta de portabilidade, a necessidade de profissionais de saúde com formação diferenciada na sua utilização e a falta de avaliações qualitativas sobre alterações da massa muscular [28]. Por outro lado, um tempo de medição curto, uma baixa dose de radiação e um reduzido custo apresentam-se como vantagens do uso deste método [31], constituindo a DEXA como uma técnica de referência a ser utilizada tanto em ambientes clínicos como ambientes de investigação.



Figura 2.5: Exemplo de aparelho utilizado na DEXA [37].

- **Ultrassonografia**

Técnicas de imagem que utilizem ultrassons permitem a produção de imagens das estruturas do corpo através da captação da reflexão das ondas sonoras de alta frequência previamente emitidas. Assim, este método não permite a medição da massa muscular, mas sim da espessura muscular [31]. É possível também conhecer a redução da massa magra e ainda, através da análise da escala de cinzentos para avaliar a intensidade do eco consegue-se avaliar alterações qualitativas, uma vez que a maior intensidade de eco indica um aumento de tecido fibroso intramuscular ou adiposo [33].

Erros de medição podem ocorrer, uma vez que este é teste dependente do examinador, como, por exemplo, no ângulo de orientação utilizado no transdutor e a tensão aplicada sobre o tecido [33]. Sendo que estes erros podem ser minimizados, aumentando a força de *feedback* e padronizando o procedimento. Por outro lado, é uma técnica de imagem não invasiva, de baixo custo e útil em doentes com mobilidade reduzida.

Dentro do uso de ultrassons existem várias técnicas. Os exemplos mais relevantes são a Ecografia em que acontece uma aplicação direta dos conceitos mencionados anteriormente e a Elastografia que envolve outros métodos.

A elastografia tira proveito dos ultrassons para mapear uma distorção causada no tecido. Esta distorção é denominada cisalhamento, isto é, a alteração da forma dos tecidos sem a alteração do volume dos mesmos e pode ser provocado por compressão manual ou por uma onda acústica [38].

Através desta técnica é possível medir a rigidez muscular tanto por deslocamento, deformação ou velocidade da onda de cisalhamento [38]. A elastografia é relevante no

estudo da sarcopenia uma vez que muitos estudos revelam que ocorre uma alteração na rigidez do músculo antes de ocorrer alterações na massa muscular [38]. Dentro das diferentes técnicas de elastografia a elastografia de onda de cisalhamento é a que se distingue mais no estudo da sarcopenia devido à sua independência do operador e por fornecer informações quantitativas [38].



Figura 2.6: Exemplo de aparelho utilizado nas técnicas de ultrassons [39].

2.5 Tratamento

Apesar do seu impacto negativo no bem estar e saúde das pessoas, a sarcopenia é considerada uma síndrome reversível. Ao longo das últimas décadas, verifica-se um interesse crescente por parte da comunidade científica e clínica internacional no desenvolvimento de intervenções terapêuticas estruturadas que permitam a mitigação de fatores de risco, ou reversão de quadros de sarcopenia já diagnosticados, na população mais idosa.

Não obstante ao volume e diversidade de respostas terapêuticas propostas, de natureza farmacológica e não farmacológica, é possível destacar o crescente interesse nos efeitos de programas de exercício físico regular e alterações dietéticas em idosos sarcopênicos. Deste modo, de seguida, serão aludidos a alguns dos mais recentes avanços nestes dois campos.

2.5.1 Exercício físico

Dentro da variedade de atividades físicas é considerada que a conciliação entre exercícios aeróbicos, como andar, correr ou nadar, e exercícios de resistência progressiva seja o melhor curso de ação [40].

O exercício aeróbico possui um impacto positivo na sarcopenia uma vez que, melhora o controle metabólico e a sensibilidade à insulina do músculo esquelético. Todavia, não apresenta um aumento tão elevado na melhoria da massa e força muscular como as atividades de resistência. O último, impulsiona a síntese proteica nas células do músculo esquelético, daí um maior nível de melhoria na massa e força muscular [41].

2.5.2 Nutrição e intervenções farmacêuticas

Outra forma de prevenção/tratamento da sarcopenia é uma alimentação adequada, dado que a perda de peso está irreversivelmente ligada à perda de massa muscular. Conseqüentemente, a estabilização do peso evita perdas desnecessárias de massa muscular, sendo este o objetivo tanto da prevenção como do tratamento à sarcopenia. Porém, para ser benéfica, a alimentação deverá ser equilibrada, fornecendo as quantidades de calorias e nutrientes adequadas à pessoa, tendo por base a sua idade, sexo, nível de atividade física, entre outros aspetos [42].

Atualmente não existem intervenções farmacêuticas viáveis para o atraso do progresso da sarcopenia [43]. Algumas fontes apresentam a hipótese de substituição hormonal, no entanto, esta apresenta resultados pouco consistentes[11].

3 Classificadores

Sarcopenia é uma condição clínica comum entre os idosos, caracterizada por uma perda gradual de massa muscular e de força, reduzindo a capacidade funcional e aumentando o risco de mobilidade reduzida e mortalidade. Como aludido anteriormente, dada a potencial reversibilidade desta síndrome, a detecção precoce da sarcopenia é crucial para a implementação de medidas preventivas.

Técnicas baseadas na imagem, tais como a ecografia e a elastografia, têm sido amplamente utilizadas para avaliar a composição muscular e a qualidade muscular, devido à sua facilidade, rapidez e custo reduzido. No entanto, uma análise manual destas imagens é suscetível a subjetividade e erro, tornando necessário a implementação de um método de classificação automatizado [44].

Nos últimos anos, uma combinação de técnicas de inteligência artificial, tais como *deep learning* e *machine learning*, mostrou ter o poder de classificar diversas patologias por meio de imagens médicas, tudo isto com elevado desempenho. Estes modelos conseguem aprender padrões e características complicadas associadas à sarcopenia, o que pode levar a classificações mais precisas e fiáveis. Por outro lado, a classificação de sarcopenia utilizando imagens produzidas pelas técnicas de ecografia e elastografia tem sido pouco explorada, carecendo em metodologias a seguir que resultem na classificação da patologia com elevado desempenho.

Este capítulo visa rever o estado da arte na classificação da sarcopenia utilizando imagens de ecografia e de elastografia, usando classificadores baseados em *deep learning* e *machine learning*, destacando as suas vantagens e desafios. Devido ao número reduzido de artigos e publicações versando métodos de classificação da sarcopenia, foram analisados artigos reportando a utilização de modelos para identificação de outras patologias a nível muscular ou até em situações de trauma muscular, de modo a compreender a aplicação de modelos computacionais na classificação de imagens médicas. Desta forma foi possível perceber o atual estado da arte e identificar tendências futuras e áreas de investigação necessárias.

3.1 *Machine Learning* Clássico

3.1.1 Definição

Machine learning é um ramo de inteligência artificial centrado-se no desenvolvimento de algoritmos e modelos capazes de aprender com os dados para melhorar o desempenho em determinadas tarefas. Os sistemas de aprendizagem automática destinam-se a adaptar-se e a generalizar a partir de instâncias, em contraste com a programação tradicional baseada em regras, que fornece instruções explícitas para cada caso [45].

Os algoritmos processam e analisam a informação de forma iterativa de modo a reconhecer padrões, correlações e tendências nos dados de entrada. Quando são fornecidos dados novos e não observados, estes padrões aprendidos são posteriormente utilizados para produzir previsões ou classificações [45]. De modo a maximizar o desempenho do modelo, é necessário selecionar o algoritmo mais adequado, as características corretas, treinar o modelo num dado conjunto de dados rotulados e ajustar os parâmetros do modelo.

O objetivo final de *machine learning* é criar modelos que possam evoluir e adaptar-se automaticamente à medida que novos dados ficam disponíveis, melhorando ao longo do tempo a sua precisão e eficácia [45]. Com a ajuda desta tecnologia, que possui aplicações em vários domínios, incluindo veículos autónomos e processamento de linguagem natural, os computadores podem agora realizar tarefas complexas e tomar decisões hábeis.

Os modelos podem ainda ser categorizados como possuindo uma aprendizagem supervisionada ou não supervisionada. De forma a realizar previsões ou categorizar dados, a aprendizagem supervisionada inclui o treino de um modelo em dados rotulados com entrada e saída fornecidas. Por outro lado, a aprendizagem não supervisionada centra-se na procura de padrões e estruturas nos dados sem resultados especificados e lida com dados não rotulados [45]. Uma representação esquemática destas categorias é ilustrada na Figura 3.1

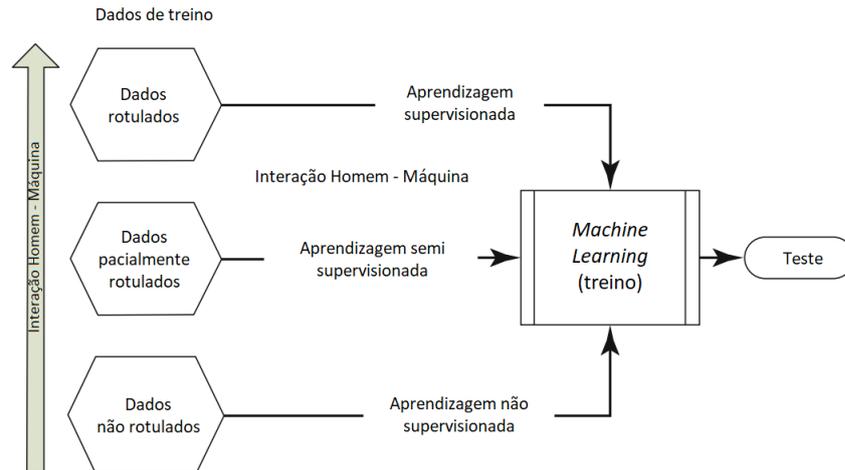


Figura 3.1: Categorias dos modelos de *machine learning* de acordo com a natureza dos dados de entrada (adaptado) [45].

3.1.2 Classificação de imagens médicas a nível muscular

Sun *et al.* [46] (2020), publicaram um estudo em que apresentaram um modelo funcional para a classificação de jovens saudáveis com diferentes níveis de exercício com base em múltiplas imagens de ecografia dos músculos tibial anterior e gastrocnémio lateral. Neste estudo participaram um total de 107 jovens saudáveis, dos quais foram retiradas imagens de ultrassons do músculo tibial anterior e do músculo gastrocnémio lateral em diversas posições. Os participantes foram divididos em dois grupos de acordo como o seu nível de exercício: praticantes regulares (mais de três vezes por semana) e praticantes irregulares. Primeiramente, os autores procederam ao mapeamento das características de morfologia muscular, da análise da frequência na obtenção de imagens e da textura das imagens. A posterior seleção de características foi realizada através de uma eliminação recursiva de características (RFE), ou seja, um algoritmo baseado em *Support Vector Machine* (SVM), SVM-RFE. De seguida, de modo a classificar o nível de exercício dos jovens, os autores utilizaram o modelo clássico SVM. Devido ao baixo número de participantes, optaram por uma *nested cross-validation* para realizar o aperfeiçoamento e teste do modelo. O modelo final apresentou uma precisão de 78,90%, sendo esta uma média da precisão para cada perna. Os autores concluíram que a estrutura muscular é influenciada por diferentes níveis de exercício, podendo ser classificada via imagens ecográficas.

Num outro estudo, foram utilizadas imagens de um ecógrafo em modo B combinadas com imagens de elastografia de onda de cisalhamento de vários músculos, de modo a classificar a lombalgia não específica (dor lombar sem causa específica) [47]. Participaram um total de 52 pacientes, divididos como tendo dor moderada ou severa. Os autores realizaram uma seleção

de características utilizando *ANOVA*, resultando em 800 características relevantes das quais se destacam: morfologia muscular, textura de imagem e elasticidade. A seguir aplicaram o modelo SVM para classificar as imagens como pertencentes a pacientes com dor moderada ou severa. De modo a estudar o desempenho do modelo, foi realizada uma *five-fold cross validation* combinada com uma pesquisa em grelha devido ao baixo número de participantes. Foi calculado o desempenho do modelo utilizando três grupos de características: relativas apenas ao ecógrafo em modo B, apenas à elastografia e dos dois conjuntos combinados. O melhor desempenho corresponde ao último grupo com uma precisão de 85%. Este estudo comprovou que a lombalgia não específica é uma patologia que apresenta diferenças musculares significativas em diversos locais, quando comparada com os músculos de indivíduos saudáveis. Além disso, foi constatado que modelos de classificação que consideram imagens de vários músculos apresentam desempenho superior em relação àqueles que consideram apenas imagens de um músculo.

Em 2023 foi publicado um artigo onde classificaram a especificação da disfunção muscular do pescoço através de imagens de elastografia com a ajuda de modelos de *machine learning* [48]. Neste estudo participaram um total de 38 mulheres adultas, resultando em 1099 imagens. Foram utilizados e comparados seis modelos diferentes: *Random Forest*, *Decision Trees*, SVM, *K Nearest Neighbor*, *Logistic Regression* e *Naive Bayes*. Para o treino e validação do modelo foram utilizadas 323 características onde foi aplicada uma *ten-fold cross-validation* a todo o conjunto de dados. Este processo foi repetido dez vezes com uma nova otimização hiperparamétrica utilizando um algoritmo de pesquisa em grelha e utilizando dez novas dobras aleatórias. Por fim, o teste e consequente desempenho foi calculado através de uma *100-cross-validation*. Aqui os modelos com melhor desempenho foi o SVM e o *Random Forest* com precisões 81,40% e 81,20% respectivamente. Uma vez que a dor no pescoço é uma condição complexa em que o diagnóstico é dependente da sensação de dor do paciente e da experiência do técnico, este método veio trazer uma alternativa precisa e objetiva para diagnosticar esta patologia.

3.2 *Deep Learning*

A *deep learning* é uma subcategoria de *machine learning*. O termo *machine learning* refere-se a uma variedade de métodos onde os algoritmos são subjugados a um treino de modo a reconhecer padrões nos dados e a gerar previsões. As redes neurais têm a capacidade de aprender automaticamente características e hierarquias complexas nas representações de dados para gerar previsões [49]. Essencialmente, *deep learning* utiliza redes neurais profundas de modo a realizar tarefas como o reconhecimento de imagens e de voz no âmbito do domínio mais

vasto de *machine learning*.

Apesar de *deep learning* pertencer ao ramo de *machine learning*, nesta dissertação foram estudados separadamente de forma a realizar uma comparação entre o desempenho de cada uma. No diagrama de *Venn* da Figura 3.2, a relação hierárquica entre os conceitos é sintetizada. Através da figura é possível perceber o conceito de *machine learning*, que se subdivide em diversos ramos e ainda, o enquadramento de *deep learning*, um ramo especializado dentro do conceito de *machine learning*.

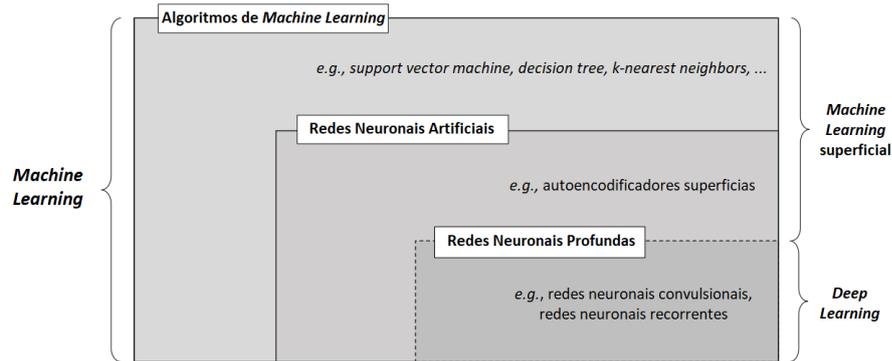


Figura 3.2: Diagrama de *Venn* das classes e subcategorias de *machine learning* (adaptado) [49].

3.2.1 Definição

Deep learning, como mencionado anteriormente, é um subcampo de *machine learning*, baseando-se na utilização de redes neurais artificiais complexas e de múltiplas camadas, muitas vezes conhecidas como redes neurais profundas [50]. Os modelos aprendem automaticamente representações e padrões complexos a partir de dados brutos, em contraste com os métodos típicos de *machine learning*, que dependem frequentemente de características criadas manualmente. Através da sua natureza em camadas, estas arquiteturas profundas destinam-se a extrair gradualmente características de nível superior, permitindo-lhes captar pequenos detalhes e abstrações complexas nos dados [50].

Dentro dos modelos de *deep learning* podemos encontrar modelos como redes neuronais convolucionais (CNN), que preenchem uma lacuna vital ao serem especialistas na análise de dados em grelha, como a análise de imagens. Funcionam utilizando camadas de convolução para identificar tendências regionais e agrupamento de camadas para extrair informações essenciais [51]. Podem distinguir características cada vez mais abstratas à medida que as camadas se tornam mais profundas, graças a este processo hierárquico de extração de características [51].

3.2.2 Redes Neurais Convolucionais (CNNs)

Utilizando imagens de ecografia e elastografia para classificar casos de sarcopenia, as redes neurais convolucionais (CNN) são um subconjunto de modelos de *deep learning* que se adequam bem à avaliação de imagens como dados de entrada [52]. Quando se trata de recolher dependências espaciais, selecionar automaticamente características relevantes e realizar tarefas de identificação de imagens, as CNNs têm mostrado grande sucesso.

Utilizando uma série de camadas progressivamente empilhadas, uma CNN transforma gradualmente a imagem de entrada numa representação significativa que permite uma categorização exata. A camada convolucional, que efetua a extração de características locais, é o conceito fundamental subjacente às CNN [52]. As CNN são compostas essencialmente por e camadas distintas: camadas de convolução, camadas de subamostragem e camadas totalmente conectadas. Na Figura 3.3 é apresentada a arquitetura básica de uma CNN.

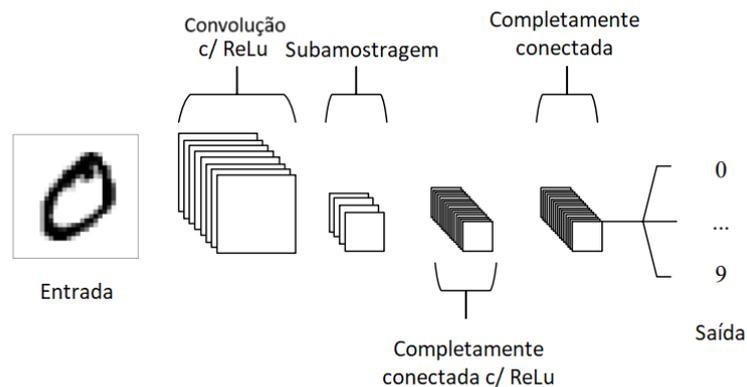


Figura 3.3: Arquitetura simples de uma CNN (adaptado) [52].

De seguida é apresentada uma visão geral das camadas presentes em modelos de *deep learning*, tanto as específicas a CNNs como as mais gerais, presentes em diversos modelos.

3.2.2.1 Camadas de Convolução

A convolução é um processo fundamental nas redes neurais convolucionais (CNN) que, neste contexto, permite a extração de características pertinentes de uma imagem de entrada. A matriz de dados de entrada é multiplicada por um filtro, também designado por kernel, para produzir um mapa de características que extrai as informações-chave da imagem [52].

Em comparação com os dados de entrada, a matriz de pesos do filtro é uma matriz bidimensional mais pequena. O tamanho inferior do filtro permite aplicá-lo metodicamente a várias áreas da imagem de entrada. Pode encontrar características significativas em qualquer

parte da imagem, aplicando o filtro várias vezes sobre a entrada. Cada multiplicação entre o filtro e a área correspondente da imagem de entrada produz um único valor. Estes valores são organizados num mapa de características, uma matriz bidimensional de saída. O mapa de características é uma representação filtrada da entrada que ajuda a identificar padrões ou características distintivas, chamando a atenção para os dados pertinentes [52].

3.2.2.2 Funções de Ativação

As camadas de ativação podem contribuir de forma não linear para o modelo, desempenhando uma função crucial nas redes neuronais. Isto permite que a rede aprenda padrões complicados e gere previsões com maior expressividade. Na arquitetura da rede, estas camadas são posicionadas após cada camada de convolução ou totalmente conectada [53].

As camadas de ativação são utilizadas para adicionar alterações, que podem ser não lineares aos resultados das camadas anteriores. Como resultado, a rede consegue reconhecer e aprender sobre padrões não lineares nos dados, sendo frequentemente cruciais para a realização de tarefas difíceis. A capacidade da rede de aprender modificações seria severamente limitada se não existissem camadas de ativação. As funções mais comuns são de unidade linear retificada, sigmoide e tangente hiperbólica [53].

As funções de ativação da unidade linear retificada (ReLU) estão entre as funções de ativação mais frequentemente utilizadas. A ReLU, essencialmente, introduz não linearidade ao modelo, definindo todos os valores negativos como zero enquanto mantém os valores positivos. A ReLU ajuda a mitigar o problema do desaparecimento do gradiente durante o treino e é computacionalmente eficiente [54], sendo que este problema é quando os neurónios nas camadas anteriores aprendem a uma velocidade inferior aos neurónios presentes nas camadas posteriores.

As funções de ativação sigmoide e tangente hiperbólica são duas funções comuns. A função sigmoide condensa os valores de entrada no intervalo $[0, 1]$, tornando-a adequada para questões de classificação binária onde os resultados podem ser expressos como probabilidades [54]. Tangente hiperbólica, por outro lado, condensa os valores de entrada numa gama de valores de saída entre -1 e 1 [54].

3.2.2.3 Camadas de Subamostragem

A saída dos mapas de características da camada convolucional tem o inconveniente de registar a localização exata das características na entrada. Isto implica que mesmo pequenas alterações na posição de uma características na imagem de entrada produzirão um mapa de características

diferente [52].

A redução da amostragem é uma estratégia frequentemente utilizada para ultrapassar este problema. Para tal, é produzido um sinal de entrada de baixa resolução que retém as características estruturais cruciais, mas não os pormenores finos que podem não ser tão úteis para a tarefa em causa. Esta redução da amostragem é efetuada utilizando camadas de agrupamento. Para produzir um novo conjunto com o mesmo número de mapas de características, estas camadas trabalham independentemente em cada mapa de características. A escolha de uma operação, como um filtro, a aplicar ao mapa de características é um passo no processo de agrupamento. A subamostragem média e a subamostragem máxima são os dois tipos de subamostragem, sendo que um determina o valor médio dentro de cada grupo e outro determina o valor máximo dentro de cada grupo [52].

3.2.2.4 Camadas de Eliminação

O sobreajuste, que ocorre quando um modelo fica excessivamente especializado nos dados de treino, com um mau desempenho em dados novos e não testados, pode afetar as redes neuronais treinadas em conjuntos de dados limitados. As camadas de eliminação são um método para reduzir o ajuste excessivo [55]. É um método de regularização que simula o treino simultâneo de várias redes neuronais com várias configurações.

Subconjuntos aleatórios de neurónios numa camada são eliminados ou brevemente ignorados durante o treino. Como resultado, a camada é tratada como se tivesse um número diferente de nós e conectividade com a camada abaixo. Essencialmente, cada atualização de camada durante o treino é efetuada utilizando uma configuração distinta. As camadas da rede adaptam-se às falhas das camadas anteriores, reforçando o modelo. A eliminação incentiva a rede a aprender realmente uma representação geral, simulando a ativação geral de uma determinada camada [55].

3.2.2.5 Camadas Totalmente Conectadas

As camadas totalmente conectadas, por vezes designadas por camadas densas, são as camadas de uma rede neuronal profunda em que cada entrada de uma camada está associada a cada unidade de ativação da camada seguinte. As últimas camadas de um projeto de rede neuronal profunda são normalmente camadas totalmente conectadas que combinam os dados recolhidos pelos níveis anteriores para criar o resultado final [52]. As camadas completamente conectadas requerem muita computação porque há muitos parâmetros que têm de ser aprendidos durante a fase de formação.

As camadas densas que vêm depois das camadas convolucionais numa CNN funcionam como uma ligação entre a saída e as características que foram extraídas. As características recuperadas são utilizadas para aprender padrões e relações complexas, sendo depois transformadas numa representação adequada para a tarefa em causa. Dependendo da dificuldade do problema e das decisões de conceção da arquitetura da rede, o número e a configuração destas camadas completamente conectadas podem mudar [52].

3.2.3 Classificação de imagens médicas a nível muscular

Num estudo realizado por Jisook Yi *et al.* (2022) [56] foram utilizadas imagens obtidas através do uso de ultrassonografia com escala de cinzentos e de elastografia por *shear-wave* de modo a detetar sarcopenia. Foram eleitas três redes neuronais convolucionais profundas (DCNNs) (*VGG 19*, *ResNeT 50* e *DenseNet 121*) para classificar a sarcopenia. Estes classificadores estavam pré-treinados com 1,28 milhões de imagens naturais com 1000 categorias de *ImageNet*. O treino para todos os modelos foi realizado utilizando uma descida de gradiente de *mini-batch* com otimizador de *Adam* e uma taxa de aprendizagem base de 0,0001 acoplado por um fator de 0,1 quando o gradiente estava preso num *plateau*. No global, a amostra do estudo era composta por 160 pessoas, tendo sido utilizadas imagens captadas do músculo reto femoral direito de cada indivíduo. Destes 160, apenas 141 (29 sarcopénicos) foram selecionados para o estudo, dos quais 17 serviram para validação do modelo, 20 para o teste e os restantes para o treino. Os resultados revelaram uma superioridade em precisão das imagens obtidas com elastografia por *shear-wave* (70,0 a 80,0%) comparadas com o uso de imagens de ultrassonografia por escala de cinzentos (65,0 a 75,0%). Entre os três DCNNs o que produziu melhor classificação foi o *VGG 19* com uma sensibilidade de 77,8%, especificidade de 72,7% e precisão de 75,0%.

Também no âmbito de classificar a sarcopenia foi criado um modelo de rede autocodificador neuronal convolucional para aplicações de entrada *multi-frame* [57]. Neste projeto foram usados 1100 vídeos de ultrassonografia de vários músculos para treino e teste, dos quais 17% possuíam sarcopenia. Em primeiro lugar foi treinado um autocodificado constituído por redes neuronais artificiais, capaz de reduzir e comprimir a dimensão de um espaço inicial, remover o ruído e ainda extrair as características robustas. A seguir, estas características foram introduzidas numa rede neuronal convolucional (CNN) para identificação da sarcopenia. De modo a perceber qual o melhor modelo, foi analisado o desempenho de três modelos pré-treinados: *ResNet*, *Xception* e *DenseNet* e um modelo CNN customizado. O desempenho dos modelos foi testado com e sem o uso do autocodificador. O uso de autocodificador acoplado ao modelo pré-treinado *DenseNet* emergiu como a abordagem de maior precisão diagnóstica (90,7%), tendo-se verificado um *F1*

score de 68,40%. Através do modelo proposto, os autores consideram ser possível fazer o rastreamento rápido e em larga escala da população, sendo a única desvantagem a necessidade de significativos recursos computacionais para conseguir treinar o autocodificador.

Outra abordagem para a classificação de sarcopenia baseado em imagens de ultrassons foi apresentada por Katakis [58]. Neste estudo foram utilizadas 2005 imagens correspondentes a quatro músculos diferentes de 210 indivíduos. Os músculos eram: tibial anterior, reto femoral, músculo gastrocnémio e músculos do braço anterior (bíceps braquial e músculo braquial anterior) e as imagens foram ainda divididas em três grupos: jovens e saudáveis (1), idosos e saudáveis (2) e idosos sarcopênicos (3). Foram usadas várias arquiteturas baseadas em deep learning para realizar a extração automática da área de secção transversal de cada músculo. Nesta parte do estudo, o modelo que apresentou melhor precisão foi TMUNet (95%). Os autores aplicaram o modelo ResNet 18 para classificar cada imagem quanto ao seu grupo demográfico por uma análise da textura das imagens do músculo (propriedade visual da imagem). Os autores verificaram uma precisão diagnóstica de 93,1% para o primeiro grupo, 78,4% para o segundo grupo, e 44,2% para o grupo de pessoas idosas com sarcopenia. Esta redução em precisão pode ter origem em vários aspectos como, por exemplo, a sarcopenia não afetar de igual modo todos os músculos, sendo que poderia ser mais evidente em alguns músculos do que em outros, conduzindo a uma classificação incorreta por parte do modelo. Também a coexistência de outras variáveis que influenciem a arquitetura dos músculos no segundo grupo (pessoas idosas saudáveis) poderá, em parte, justificar o diferencial de precisão verificado em contraste com o grupo três. Apesar do baixo desempenho, este processo de classificação de sarcopenia apresenta-se como promissor, pois mesmo que necessária a melhoria do modelo, foi demonstrado que é possível realizar a classificação através da textura do músculo.

Snider *et al.* [59], otimizaram um algoritmo de *deep learning* de classificação de imagens para detecção de corpos estranhos em tecido muscular a partir de imagens de ultrassons. Neste estudo foram apenas usadas imagens obtidas através da adição de fragmentos a modelos de coxa em gelatina (fantomas) e a tecido de suínos, recentemente eutanasiado. O modelo de rede neuronal convolucional (CNN) foi sucessivamente otimizado em sete aspectos distintos: ativadores de camadas, arquitetura do modelo, otimizador, parâmetros treináveis, taxa de desistência e por último número de ciclos. Para o modelo final foi selecionada a função de ativação *ReLU* para as camadas 2D CNN e a sigmoide para a camada 1D densa e totalmente conectada, tendo a camada densa o dobro do tamanho da camada final da CNN. Para a seleção hiperparamétrica utilizaram um otimizador de propulsores RMS, com cerca de 17 milhões de parâmetros treináveis, um aumento de dados de 10% de magnitude e uma taxa de desistência de 55% durante 100 ciclos.

Este modelo obteve um *F1 score* de 95% para as imagens de tecido fantoma e de 99% para as imagens de suíno. Resultou num algoritmo com uma precisão elevada para a identificação dos corpos, além disso, os autores ainda consideram relevante o uso desta estrutura em outras aplicações de imagem relevantes para o trauma.

Mais recentemente, os mesmos autores [60], apresentaram um estudo comparativo entre o modelo criado por eles e outros modelos existentes. Neste estudo foi usado novamente o fantoma em gelatina, mas agora modificado para simular a coxa de um homem adulto. O modelo foi mais uma vez otimizado com 250 iterações Bayesianas de otimização em seis hiperparâmetros chave. Após esta otimização, o modelo teve um desempenho de 97% em precisão e um *F1 score* de 97,65%. Foram escolhidos 11 modelos para comparação. Foi iniciado o treino dos 11 modelos com cinco ciclos utilizando 200 imagens. O número de classificadores foi reduzido para cinco, consoante os seus desempenhos, e então foram retreinados com o conjunto completo de dados, utilizando 100 ciclos. Aqui o modelo com o melhor desempenho foi o *MobileNetv2* com um *F1 score* de 3,1% superior ao modelo criado pelos autores e 3,2% superior em precisão. Em contrapartida, era mais lento a realizar a identificação dos corpos. Uma vez que o modelo é destinado à utilização em tempo real, o modelo criado pelos autores foi considerado pelos mesmos como tendo o melhor desempenho, possuindo o tempo de classificação mais rápido e uma das precisões mais elevadas.

Ritsche, [61], propôs um método para a segmentação automática da área de secção transversal anatómica em imagens de ultrassons dos músculos dos membros inferiores humanos, utilizando *deep learning*. As redes neuronais convolucionais foram treinadas e testadas com um total de 1772 imagens de ultrassons correspondentes a 153 participantes, com uma média de 38 anos. Foram treinados três modelos separados, um para cada músculo ou conjunto de músculos (reto femoral (RF), vasto lateral (VL) e gastrocnémio medial (GM) em conjunto com gastrocnémio lateral (GL)), todos baseados na arquitetura *U-Net*. Os modelos foram treinados com um máximo de 50 ciclos e um tamanho de lote igual a um. Para a otimização foi eleito o otimizador de *Adam*. No final, a saída do modelo é uma imagem de máscara binária que contém a área muscular prevista. Os resultados foram promissores com uma correlação intra classe com uma análise manual de 96% para RF, de 94% para VL e de 97% para GM/GL e ainda precisões de 98,40%, 97,10% e 97,30%, respetivamente. Os autores consideram a segmentação como eficiente, precisa, comparável à segmentação realizada manualmente e de fácil aplicação tanto em pesquisas quanto em ambientes clínicos.

Com o intuito de classificar miosite (inflamação muscular) por imagens de ultrassons do músculo foi adaptado um modelo baseado em *deep learning* [62]. Neste estudo foi utilizado um

total de 3214 imagens correspondentes a 80 indivíduos, 30 saudáveis, nove com miosite corporal, 14 com polimiosite e 14 com dermatomiosite. O modelo proposto foi obtido através da combinação das camadas convulsivas dos modelos *VGG 16* e *VGG 19*. Os modelos eram pré-treinados por *ImageNet* sendo usado para treino e validação a técnica *five-fold cross validation*. Para otimização dos parâmetros foi utilizado o otimizador de *Adam* com uma taxa de aprendizagem fixa em 0,00002. O modelo foi testado em quatro cenários diferentes: o primeiro destinado a realizar a distinção entre imagens de indivíduos saudáveis e doentes, o segundo entre os saudáveis e miosite corporal, o terceiro entre miosite corporal e as restantes doenças, por fim, no quarto cenário foi realizada a classificação entre todas as classes. O modelo proposto obteve uma precisão média de 93% no primeiro cenário, 96,01% no segundo, 91,74% no terceiro e 95,12% no quarto. Devido ao bom desempenho obtido, este método foi considerado, pelos autores, útil na obtenção de resultados precisos, rápidos e de baixo custo.

Jingxu Xu *et al.* (2020) [63], propuseram um método para classificar os pacientes como masculino ou feminino através de imagens ecográficas dos músculos. Participaram 107 jovens saudáveis, resultando em 1498 imagens dos músculos tibial anterior e gastrocnémio lateral numa variedade de posições. Em primeiro lugar foram comparados três modelos de *deep learning* na classificação das imagens como masculino ou feminino. Os modelos foram: *Alexnet*, *Googlenet* e *VGG 16* e obtiveram precisões de 83,30%, 84,40% e 96,70%, respetivamente. Visto que *VGG 16* obteve o melhor desempenho, este foi posteriormente otimizado com o otimizador de *Adam* com um tamanho de lote de 16, número máximo de iterações definido para 10000, e uma taxa de aprendizagem inicial de 0,0001. A precisão final deste modelo manteve-se nos 96,70% sendo o modelo com melhor desempenho. Este estudo contribuiu para a melhoria de modelos de classificação com imagens ecográficas músculo-esqueléticas, permitindo conhecer as regiões mais importantes da imagem que a rede neuronal profunda utiliza para a tarefa de classificação.

4 Metodologia

O capítulo sobre a metodologia aborda as técnicas utilizadas para responder aos objetivos do estudo, utilizando três conjuntos de dados de ecografia, elastografia e ecografia e elastografia combinados. A metodologia está dividida em duas partes: *machine learning* clássico e *deep learning*, cada uma das quais abrange uma fase diferente. Foi utilizado o método de janela deslizante para aumentar os dados de entrada dos modelos de *machine learning* clássico. Para escolher as opções mais promissoras, foram inicialmente testados vários classificadores, selecionando apenas cinco. Depois realizou-se uma otimização dos classificadores seguido de uma comparação de várias possibilidades nas fases de: extração, normalização e seleção de características.

Para *deep learning* o conjunto de dados foi melhorado utilizando abordagens de aumento de dados. O tamanho de grupo, as épocas, as funções de perda e os algoritmos de otimização foram todos incluídos na exploração de parâmetros. As configurações ideais foram encontradas testando metodicamente cada parâmetro e comparando as possibilidades disponíveis. Foi ainda testado a possível adição de programadores de taxas de aprendizagem e mecanismos de atenção.

O estudo comparativo das melhores configurações dos modelos nos três conjuntos de dados e dentro dos respectivos domínios, foi o colmatar da comparação dentro de cada domínio (*machine learning* clássico e *deep learning*). As vantagens e desvantagens de vários modelos foram destacadas por esta abordagem abrangente, que também facilitou a realização de um exame completo entre domínios. Finalmente foi realizada uma comparação entre os modelos de *machine learning* clássico e *deep learning* nos vários conjuntos de dados.

Nas próximas secções são descritos todos os métodos utilizados no decorrer deste projeto, assim como uma breve descrição dos modelos selecionados e métricas de eleição. Na generalidade, as etapas executadas estão presentes na Figura 4.1.

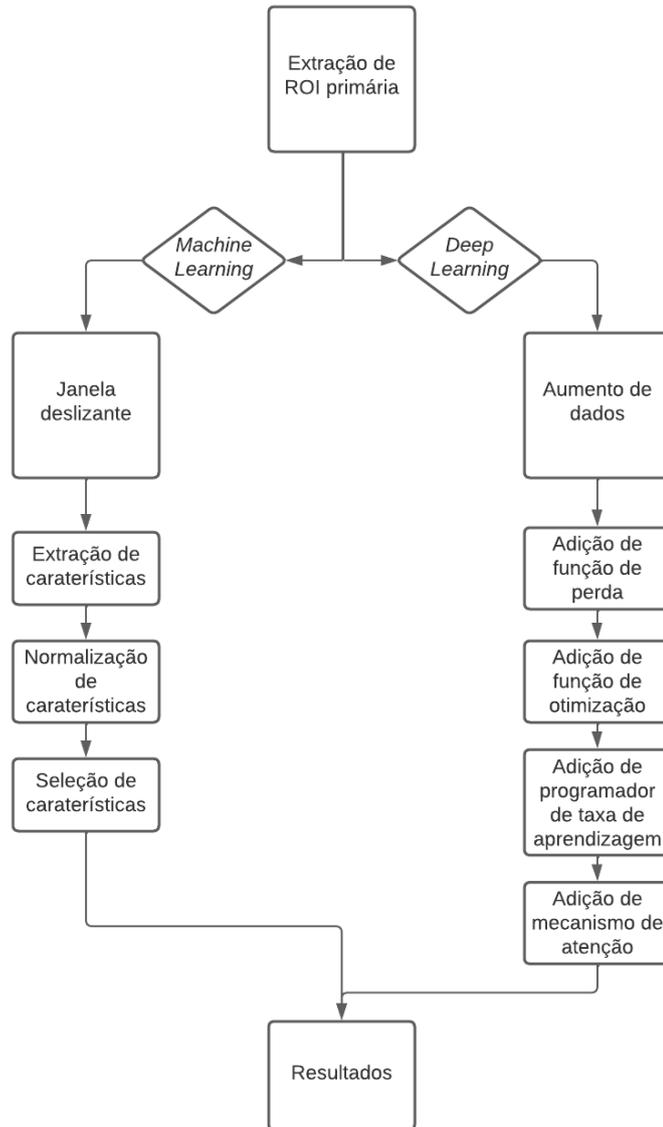


Figura 4.1: Organização geral das etapas realizadas.

4.1 Pré-processamento e aumento de dados

Na elaboração de um modelo geralmente subdivide-se os dados em três grupos: treino, validação e teste. Em primeiro lugar são utilizados os dados de treino como entrada do modelo com as suas devidas classes, aqui, o modelo apreende e define cada classe com uma gama de características. A seguir, utilizam-se os dados de validação como um teste preliminar, onde o modelo não sabe as classes e tenta definir os dados quanto à classe, este grupo é comumente utilizado para a validação dos parâmetros selecionados. Por fim, é utilizado o grupo de teste para avaliar o desempenho do modelo.

Devido ao número reduzido de imagens obtidas para o treino e teste dos modelos

desenvolvidos, foi necessário aumentar este número. Note-se que o uso de poucos dados de entrada leva a que as imagens de teste e validação sejam muito semelhantes às utilizadas no treino dos modelos. Assim, um número limitado de dados de entrada apresenta-se como sendo um dos principais problemas associados a *machine learning* clássico e *deep learning* uma vez que o risco de sobre ajuste aumenta consideravelmente. Um número elevado de hiperparâmetros apesar de ser benéfico na distinção de classes, poderá conduzir o modelo a ser pouco generalizado, ou seja, excessivamente adaptado aos dados de treino.

Existem várias técnicas que ajudam a aumentar os dados de entrada. Neste projeto foram utilizadas duas técnicas distintas, uma direcionada a modelos de *machine learning* clássico e outra a modelos de *deep learning*. Sendo assim as técnicas eleitas neste projeto foram: janela deslizante e aumento de dados.

4.1.1 Janela deslizante

O método janela deslizante, consiste na extração de várias regiões de interesse secundárias a partir de uma imagem (ROI primária) movendo uma janela de tamanho fixo através da imagem de uma forma sistemática, permitindo a extração de características locais ou o reconhecimento de padrões [64]. Ao examinar cada janela separadamente, este método permite a deteção de padrões e características locais que podem ser essenciais para previsões ou classificações precisas. Isto permite ter uma extração de características localizada. No entanto, um tamanho de janela fixo possui vários inconvenientes. Uma janela demasiado pequena pode não detetar padrões significativos e uma janela demasiado larga pode introduzir informações desnecessárias ou aumentar as exigências de processamento [64]. Na Figura 4.2 está ilustrado o processo de janela deslizante.

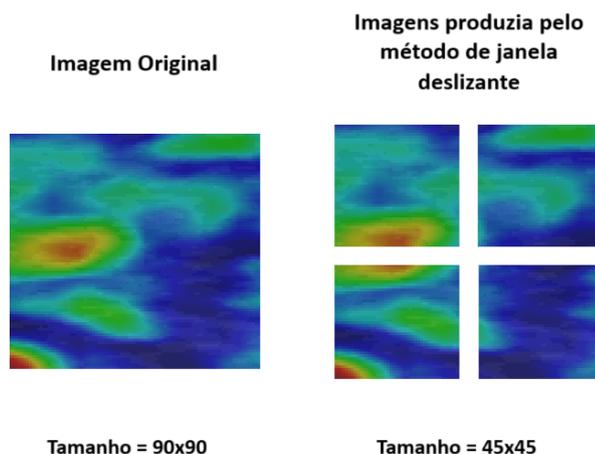


Figura 4.2: Ilustração do método de janela deslizante produzindo múltiplas imagens.

4.1.2 Aumento de dados

Através da aplicação de uma gama de modificações ou perturbações aos dados de entrada do modelo, é possível gerar exemplos novos igualmente válidos. Estas alterações aos dados introduzem variações, mantendo a classe subjacente.

Ao criar amostras baseadas nas amostras existentes, o aumento de dados aumenta a capacidade do modelo de processar dados previamente desconhecidos, resultando numa maior adaptação do modelo. A adição de variações realistas aos dados pode, também, criar robustez às variações, tornando o modelo mais resistente a diferentes vistas da imagem, condições de iluminação ou níveis de ruído [65]. Por outro lado, como resultado, as necessidades computacionais aumentam.

As técnicas de aumento de dados selecionadas para este projeto foram:

- Transformações geométricas, que alteram as propriedades espaciais dos dados através de operações de rotação, escala e inversão [65], tornando o modelo resistente a vários pontos de vista.
- A adição de ruído gaussiano aleatório aos dados de entrada, tornando o modelo capaz de lidar com entradas ruidosas, simulando alterações encontradas em circunstâncias do mundo real.

4.2 Classificação tradicional

4.2.1 Extração de características

Foram extraídas um total de 1097 características de cada imagem utilizando diversas bibliotecas. As imagens foram abertas e guardadas na forma de uma matriz em que os valores correspondem à intensidade de cada píxel no respetivo canal: vermelho, verde e azul (RGB). Uma vez que diversas estatísticas foram calculadas inúmeras vezes, foi criada uma função (“statistics”) que recebia como entrada uma matriz e devolvia os respetivos valores correspondentes a cada uma das estatísticas enumeradas na Tabela 4.1.

Tabela 4.1: Estatísticas calculadas pela função “statistics”

caraterísticas	
1	Média aritmética
2	Valor máximo
3	Valor mínimo
4	Valor mediano
5	Desvio padrão dos valores
6	Intervalo de valores
7	Variância dos valores
8	Assimetria dos valores
9	Curtose dos valores
10	Moda dos valores
11	Entropia dos valores

A função descrita foi aplicada sempre que uma caraterística resultava numa matriz de valores, uma vez que pretendíamos obter caraterísticas de um só valor. Assim, as primeiras caraterísticas extraídas foram obtidas aplicando a função “first_order_statistics” à matriz da imagem ainda com os três canais RGB e a seguir à matriz da imagem apenas com um canal, ou seja, com a imagem em tons de cinzento. As restantes caraterísticas foram extraídas ou pela aplicação direta de funções pertencentes a diferentes bibliotecas, ou através da aplicação de diversos filtros de modo a detetar vários elementos nas imagens. As caraterísticas extraídas estão numeradas a seguir, divididas segundo o seu tipo e biblioteca utilizada, sendo relevante mencionar que mais uma vez a função “first_order_statistics” foi aplicada sempre que uma caraterística resultou numa matriz de valores.

4.2.1.1 Caraterísticas de textura

Os elementos de imagem conhecidos como caraterísticas de textura são utilizados para descrever as caraterísticas de textura visual de uma imagem. Estas caraterísticas baseiam-se nas relações espaciais entre píxeis adjacentes numa imagem ou na distribuição dos valores de intensidade.

Neste projeto foram extraídas várias caraterísticas de textura, incluindo: a intensidade dos píxeis em tons RGB e em tons de cinzento, gradiente de magnitude, entropia, transformação da distância, espetro de frequência e de potência, entropia Shannon, força de desfoque, histograma de gradientes orientados, transformação de *Gabor*, matriz de adjacência do limiar,

padrão binário local, espectros de ordem superior, espectro de potência de *Fourier*, análise da textura da dimensão fractal, matriz de características estatísticas, estatísticas das diferenças de nível de cinza, características energéticas da textura de *Law's* e por fim as matrizes de coocorrência de nível cinzento, de comprimento de execução de nível cinzento, de zona de tamanho de nível cinzento, de diferença de tom cinzento vizinho e de dependência de nível cinzento, das quais foram retirados vários parâmetros.

Para a matriz de coocorrência de nível cinzento (GLCM) foram extraídos os seguintes parâmetros: contraste, dissimilaridade, homogeneidade, energia, correlação, segundo momento angular, autocorrelação, destaque do *cluster*, sombra de aglomerado, tendência de *cluster*, diferença média, entropia da diferença, variância da diferença, diferença inversa, momento de diferença inversa, momento de diferença inversa normalizada, diferença da inversa normalizada, primeira medida da informativa de correlação, segunda medida da informativa de correlação, variação inversa, média da união, energia da união, entropia da união, coeficiente máximo de correlação, probabilidade máxima, soma da média, soma da entropia, soma dos quadrados.

Relativamente à matriz de comprimento de execução de nível cinzento (GLRLM) temos: nível de cinzento não uniformizado, nível de cinzento não uniformizado normalizado, variação do nível de cinzento, ênfase na corrida de alto nível de cinzento, ênfase no longo prazo, ênfase no nível de cinzento alto a longo prazo, ênfase no nível cinzento baixo a longo prazo, ênfase na corrida de baixo nível de cinzento, execução de entropia, não uniformidade de comprimento de execução, comprimento de execução não uniformizado normalizado, percentagem de execução, execução de variância, ênfase de curto prazo, ênfase no nível de cinzento alto a curto prazo, ênfase em curtas corridas de baixo nível de cinza,

Por sua vez, da matriz de zona de tamanho de nível cinzento (GLSZM) foram retiradas: nível de cinzento não uniformizado, nível de cinzento não uniformizado normalizado, variação do nível de cinzentos, ênfase na zona de alto nível de cinzentos, ênfase na grande área, ênfase na grande área de alto nível de cinzentos, ênfase na grande área de baixo nível de cinzentos, ênfase na zona de nível cinzento baixo, zona de tamanho não uniformizado, zona de tamanho não uniformizado normalizado, ênfase em pequenas áreas, ênfase na pequena área de alto nível de cinza, ênfase na pequena área com baixo nível de cinzas, zona entropia, zona percentagem, variância de zona.

Foi, ainda, extraída a ocupação, rudeza, complexidade, contraste e força da matriz de diferença de tom cinzento vizinho (NGTDM). E por fim, temos a matriz de dependência de nível cinzento (GLDM) onde foi possível extrair os seguintes parâmetros: dependência entropia, dependência não uniformizada, dependência não uniformizada normalizada, variância

de dependência, nível de cinzento não uniformizado, variação do nível de cinzento, ênfase no alto nível de cinzento, ênfase na grande dependência, ênfase na grande dependência alto nível de cinzento, grande dependência na ênfase no baixo nível de cinzento, ênfase no baixo nível de cinzento, ênfase na pequena dependência, pequena dependência na ênfase de alto nível de cinzento e pequena dependência na ênfase no nível cinzento baixo.

4.2.1.2 Caraterísticas de morfologia

Uma série de medidas denominadas caraterísticas morfológicas foram retiradas de imagens descrevendo o tamanho e a forma dos objetos na imagem. Estas caraterísticas são frequentemente utilizadas na análise de imagens e em aplicações de visão por computador para categorizar e separar vários tipos de objetos.

Para objetos com limiar de *Otsu* as seguintes caraterísticas foram extraídas: n.º de objetos, área, perímetro, perímetro de *Crofton*, excentricidade, diâmetro, diâmetro *feret* máximo, *aspect ratio*, circularidade, compactação, solidez, centroide, centroide local, centroide pesado, momento, momento central, momento *Hu*, momento normalizado, momento pesado, momento pesado central, momento pesado *Hu*, momento pesado normalizado, orientação, covariação, tamanho do maior eixo, tamanho do menor eixo, intensidade média, intensidade máxima, intensidade mínima, número *Euler*, extensão, tensor de inércia, valores próprios do tensor de inércia e n.º de píxeis.

Do mesmo modo, para manchas presentes nas imagens, foram extraídas as seguintes caraterísticas: n.º de manchas, área, perímetro, perímetro de *Crofton*, excentricidade, diâmetro, diâmetro *feret* máximo, compactação, solidez, centroide, centroide local, centroide pesado, momento, momento central, momento *Hu*, momento normalizado, momento pesado, momento pesado central, momento pesado *Hu*, momento pesado normalizado, orientação, tamanho do maior eixo, tamanho do menor eixo, intensidade média, intensidade máxima, intensidade mínima, número *Euler*, extensão, tensor de inércia, valores próprios do tensor de inércia e n.º de píxeis.

Em adição a estas caraterísticas foi ainda extraído o eixo medial, a árvore máxima, o perímetro da imagem e de *Crofton* e por fim a caraterística de *Euler*.

4.2.1.3 Caraterísticas de forma

Um grupo de caraterísticas conhecido como caraterísticas de forma é utilizado para descrever as propriedades geométricas de um objeto. Estas caraterísticas podem ser utilizadas para categorizar vários itens ou medir como a forma de um objeto muda ao longo do tempo.

As características utilizadas para caracterizar estas propriedades foram: a identificação de arestas, o rácio de arestas, o índice de forma, a alongação da forma, os comprimentos do eixo principal e do eixo menor, o diâmetro máximo da forma, o perímetro da forma, a superfície da malha da forma, a relação da superfície do perímetro da forma, a superfície de píxeis da forma, a esfericidade e por fim os momentos da imagem, centrais, Hu e de *Zernike*.

4.2.2 Normalização de características

De forma a garantir que todas as características utilizadas num modelo se encontram na mesma escala, a normalização de características é um passo de pré-processamento essencial em *machine learning*. Para evitar problemas onde características distintas possuem intervalos de valores diferentes, resultando na predominância de algumas características em detrimento de outras durante a fase de aprendizagem do modelo, é necessária a normalização das características. A normalização também auxilia na melhoria da taxa de convergência do modelo, o que se traduz num tempo de treino mais curto e previsões mais precisas. Além disso, quando se trabalha com números grandes ou pequenos, a normalização pode ajudar a evitar problemas de instabilidade numérica.

Nesta dissertação foram comparadas duas formas distintas de normalização de características disponibilizadas pela biblioteca *Scikit-Learn* onde se analisou o desempenho dos modelos como resultado da aplicação de cada função. As duas funções comparadas foram: *StandardScaler* e *MinMaxScaler*.

4.2.2.1 *StandardScaler*

O método de normalização conhecido como *StandardScaler* dimensiona as características de forma a terem uma média igual a zero e uma variância unitária. A fórmula de normalização é a seguinte:

$$z = (X - u)/s \tag{4.1}$$

Onde X é a característica original, u é a média de X e s é o desvio padrão de X , z é a característica normalizada [66]. O *StandardScaler* funciona bem com muitas técnicas de *machine learning* que assumem que os dados são normalmente distribuídos, uma vez que o método de normalização realiza essa mesma suposição [66].

O *StandardScaler* tem a vantagem de ser aplicável a qualquer distribuição de dados. Além disso, comparado com outros métodos de normalização, é menos suscetível a valores atípicos.

4.2.2.2 *MinMaxScaler*

MinMaxScaler normaliza as características para um intervalo especificado, normalmente $[0, 1]$. O *MinMaxScaler* mantém a forma da distribuição original, adequado para dados que possuam intervalos conhecidos. Apresenta-se especialmente útil para algoritmos como redes neurais que precisam que os recursos de entrada estejam dentro de um determinado intervalo.

4.2.3 Seleção de características

Em *machine learning* é habitual selecionar apenas algumas das características extraídas, pois muitas delas não possuem uma correlação significativa com os dados do problema, apenas aumentam o tempo de execução do modelo. Assim, a escolha das características mais pertinentes para um determinado problema é uma etapa crucial. Foi eleita a função *SelectKBest* da biblioteca *Scikit-Learn* (*sklearn*) pela sua facilidade de uso e modo de funcionamento. A função seleciona as k melhores características comparando valores resultantes de funções estatísticas.

Foram ainda comparadas duas funções estatísticas, seleção de características por análise de variância e por informação mútua, de modo a verificar qual produzia melhor desempenho no modelo. A correlação entre as características e a classe é testada, comparada e organizada pelo melhor resultado pelas funções estatísticas. Posteriormente, os resultados provenientes das funções mencionadas são passados à função *SelectKBest* que seleciona as k melhores características consoante os seus resultados.

4.2.3.1 Seleção de características por análise de variância

A função *f_classif* calcula o *F-value*, também conhecido como o *F-value* da análise de variância (*ANOVA*) e os *p-values* correspondentes para cada característica, ilustrando a sua importância na descrição da variação na variável alvo [67].

O *F-value* é determinado pela comparação da variância dentro de cada grupo (ou seja, a variância dentro de cada categoria da variável-alvo) com a variância entre os grupos (ou seja, a variância das médias das várias classes ou categorias da variável alvo) [68].

Já o *p-value* é calculado utilizando os graus de liberdade e o resultado do *F-value* dos dados. Indica a probabilidade de observar um resultado de *F-value* extremo, se a hipótese nula for verdadeira. O pressuposto subjacente é que a característica e o resultado desejado não têm qualquer ligação visível, hipótese nula [68].

Em suma, um *F-value* mais elevado e um *p-value* mais baixo significam uma associação mais forte entre a característica e a variável alvo [68]. A principal vantagem desta função é a

sua eficiência computacional e a habilidade de conseguir receber tanto dados numéricos como categóricos. A ligação entre as características e a variável-alvo é considerada linear, no entanto, este pressuposto pode nem sempre ser verdadeiro.

4.2.3.2 Seleção de características por informação mútua

O método, *Mutual_info_classif*, classifica cada característica com base na sua dependência da variável-alvo. A informação mútua mede a quantidade de conhecimento que uma característica acrescenta à variável-alvo. À medida que a pontuação da informação mútua aumenta, a característica torna-se mais pertinente para a variável-alvo [69].

Este método consegue controlar as correlações não lineares entre as características e a variável-alvo. É também adequado para modelos não lineares em que se assume uma relação não linear entre a variável-alvo e as características. No entanto, em conjuntos de dados de elevada dimensão, onde a quantidade de características é vastamente superior à de amostras, é possível não ter um desempenho exemplar.

4.2.4 Classificadores implementados

Devido ao tipo de problema a ser explorado foi utilizado um subconjunto de *machine learning* supervisionado designado por classificação. Este subconjunto tem por objetivo classificar os dados em grupos ou classes pré-determinados utilizando padrões descobertos a partir de dados rotulados [70].

Foi eleita a biblioteca *Scikit-Learn* para implementar todos os modelos. Foram numa primeira instância selecionados 10 modelos, que seriam os mais adequados ao problema em questão, ou seja, classificadores binários. Devido ao elevado poder computacional necessário para otimizar cada modelo, estes foram comparados após uma otimização bayesiana com *10-fold cross-validation*. Foram selecionados os cinco classificadores com melhor desempenho para serem, posteriormente, utilizados na avaliação das imagens utilizando diferentes conjuntos de dados.

Nesta secção, serão analisados os classificadores implementados: Rede Neuronal, *Nu vector Support*, *Logistic Regression*, *Stochastic Gradient Descent* e *Support Vector Machine*. Cada classificador tem características técnicas e vantagens distintas que os tornam adequados para vários tipos de situações. Apresenta-se, de seguida, um resumo geral de cada classificador e explica-se o seu funcionamento, bem como vantagens e desvantagens pertinentes.

4.2.4.1 Rede Neuronal

A rede neuronal (RN) possui uma alimentação progressiva, é flexível e popular para aplicações de classificação e regressão. É constituída por várias camadas de nós, ou neurónios, ligados entre si por conexões ponderadas. Para captar interações complicadas entre os dados de entrada e os dados de saída, a RN utiliza funções de ativação não lineares. O modelo pode aprender representações de dados complexas graças às camadas ocultas, melhorando a generalização.

É possível considerar a camada de entrada, as camadas ocultas e a camada de saída de forma a compreender o funcionamento interno da RN. As camadas de entrada recebem as características correspondentes aos dados, de seguida as camadas ocultas realizam uma série de cálculos aos dados de entrada utilizando somas ponderadas e funções de ativação. Cada camada oculta é constituída por um número de nós (neurónios), cada um dos quais aplica uma soma ponderada e envia o total para uma função de ativação. Por último, mas não menos importante, a camada de saída cria o resultado previsto com base nas representações aprendidas. De modo a reduzir o erro entre os resultados previstos e os resultados efetivos, os pesos são ajustados iterativamente ao longo da formação da RN [71]. A rede neuronal utilizada neste projeto foi um *Multi-layer Perceptron*, uma rede neuronal superficial onde o número de camadas ocultas se apresenta como um hiperparâmetro a otimizar.

4.2.4.2 *Nu vector Support*

Uma variação do conhecido classificador *Support Vector Machine* (SVM) é designado por *Nu vector Support*. Oferece um método flexível para lidar com conjuntos de dados desequilibrados e permite controlar a largura da fronteira e o número de vetores de suporte. O *Nu vector Support* procura identificar um hiperplano que maximiza a margem entre várias classes, tal como o SVM. Os vetores de suporte são um subconjunto de amostras de treino que se considera estarem mais próximas do limite de decisão, a distância entre os vetores de suporte e o hiperplano é designada de largura de fronteira [72].

O *Nu vector Support* apresenta características semelhantes às do SVM, tais como o tratamento de dados separáveis não lineares utilizando funções de kernel. Difere na utilização de um parâmetro extra, valor de Nu . A quantidade de vetores de apoio e um limite superior da percentagem de erros de formação são controlados pelo parâmetro Nu . É possível gerir o compromisso entre a quantidade de vetores de apoio e a taxa de erro de classificação, variando o valor de Nu [72].

4.2.4.3 *Logistic Regression*

Logistic Regression é utilizada para prever resultados binários. Funciona através da análise da ligação entre as características de entrada e o resultado. O algoritmo atribui um peso a cada característica antes de combinar todas as características ponderadas para determinar uma pontuação. A função sigmoide, é então utilizada para converter esta pontuação numa probabilidade. A função sigmoide converte a pontuação num número entre 0 e 1, que é a probabilidade de ocorrência do resultado [73].

O algoritmo modifica os pesos durante o treino para reduzir a discrepância entre as probabilidades esperadas e os resultados reais. Após o treino, o modelo pode fazer previsões com base em novos dados. A função sigmoide, determina a pontuação para os novos dados e, em seguida, atribui uma classe com base num limiar pré-determinado [73]. Por exemplo, prevê a classe positiva se a probabilidade antecipada for superior a 0,5; caso contrário, prevê a classe negativa.

4.2.4.4 *Stochastic Gradient Descent*

O *Stochastic Gradient Descent* não é um classificador, mas sim uma abordagem de otimização, um método de treino de um modelo. No contexto deste projeto foram testados vários classificadores utilizando este algoritmo durante o seu treino de forma a seleccionar aquele que produz o melhor desempenho.

O algoritmo começa com pesos e desvios aleatórios ao longo da formação. Em seguida, escolhe uma amostra aleatória dos dados de treino. Calcula o gradiente da função de perda utilizando esta amostra, que mede a discrepância entre as etiquetas de classe esperadas e reais. Para reduzir a perda, os parâmetros do modelo são então alterados na direção oposta do gradiente [74].

Com amostras aleatórias distintas de cada vez, este processo é efetuado várias vezes. O objetivo do algoritmo é chegar a um conjunto de parâmetros de modelo ideais que produzam previsões fiáveis. Depois do treino, o modelo pode ser utilizado para classificar novos pontos de dados, calculando pontuações para cada classe e designando o rótulo projetado para a classe com a pontuação mais elevada [74].

4.2.4.5 *Support Vector Machine*

Support Vector Machine (SVM) é utilizado para problemas de classificação. Funciona através da identificação da fronteira ou linha ideal que divide vários grupos de pontos de dados. O objetivo é encontrar uma linha que maximize a margem, sendo a margem a distância entre a linha e os pontos de dados mais próximos para cada classe.

O SVM consegue lidar com dados que podem ser divididos em classes ao longo de uma linha reta. O *kernel*, um método para traduzir dados num espaço de dimensão superior, permite também tratar dados não lineares. Isto permite ao SVM mapear os dados para um espaço onde se tornam separáveis, permitindo identificar uma fronteira de decisão não linear [72].

O SVM estabelece o limite ideal durante a formação, procurando aumentar a margem e reduzir os erros [72]. Depois do treino, o SVM pode classificar novos pontos de dados identificando em que lado da fronteira eles se encontram.

4.2.5 Otimização dos modelos

Os hiperparâmetros dos modelos são opções ou configurações escolhidas antes de um modelo ser treinado e não são aprendidos a partir dos dados. Especificam como o modelo atua e funciona durante o processo de aprendizagem. A otimização dos hiperparâmetros de um modelo é uma etapa crucial na criação de modelos de *machine learning*, uma vez que pode melhorar significativamente o seu desempenho [75]. É possível obter uma melhor precisão, generalização e resiliência através da otimização dos hiperparâmetros, o que é crucial para a aplicação em ambientes reais. Consequentemente, a otimização dos hiperparâmetros é essencial para criar modelos que possam fazer previsões precisas e resolver problemas práticos. Neste projeto foi eleita como função de otimização a bayesiana.

Primeiro são definidos o espaço de pesquisa dos hiperparâmetros e as suas distribuições correspondentes. Os hiperparâmetros podem ser contínuos, inteiros ou categóricos. Depois disso, é inicializado um modelo probabilístico da função objetivo baseado em processos gaussianos, utilizado para avaliar a eficácia de determinadas combinações de hiperparâmetros. À medida que novas observações são feitas, o modelo é modificado. Este processo continua até que o número máximo de iterações seja atingido, ou uma solução viável seja descoberta [75]. Neste projeto os modelos foram testados através da métrica *F1 score* durante 50 iterações.

Foi selecionada a otimização bayesiana devido à sua eficácia em navegar rapidamente num espaço de pesquisa considerável e em lidar com funções objetivo com ruído. Ao contrário de outras funções, a otimização bayesiana utiliza os resultados com hiperparâmetros testados na seleção dos hiperparâmetros a serem testados posteriormente, permitindo alcançar o conjunto ideal de hiperparâmetros num conseqüente menor número de iterações [75]. Além disso, ao adicionar estimativas de incerteza ao procedimento de pesquisa, consegue lidar com funções objetivo com ruído, comuns em tarefas de classificação binária. Em geral, é um método fiável e eficaz para otimizar hiperparâmetros em modelos *machine learning*.

4.3 *Deep Learning*

Deep learning inclui a formação de redes neuronais artificiais com numerosas camadas para reconhecer padrões e representações complexas nos dados, inspirando-se na estrutura e no funcionamento do cérebro humano [50]. Supera as técnicas convencionais de *machine learning* numa vasta gama de tarefas, graças à sua capacidade de extrair automaticamente características pertinentes e realizar uma aprendizagem de ponta a ponta.

Neste projeto serão aplicados modelos conhecidos como redes neurais convulsionais (CNNs) que foram criados para a análise de dados em grelha, como imagens. As camadas de convolução e de subamostragem, por exemplo, são camadas especializadas utilizadas pelas CNNs para aprender automaticamente representações hierárquicas de informações visuais [52]. Os filtros são aplicados aos dados de entrada, permitindo-lhes reconhecer padrões locais e, em última análise, aprender propriedades mais complexas e abstratas.

Associado ao conceito de CNNs, vários métodos são importantes, ao melhorarem e facilitarem o desempenho dos modelos. A aprendizagem por transferência que é uma estratégia de *deep learning* que utiliza o conhecimento de pré-treino de uma tarefa para melhorar o desempenho numa atividade relacionada. A aprendizagem por transferência permite que os modelos generalizem com êxito para novas tarefas, mesmo com poucos dados de treino, utilizando modelos pré-treinados ou características aprendidas [76].

Outro conceito importante em *deep learning* são os mecanismos de atenção. Estes permitem aos modelos efetuar cálculos de forma seletiva, concentrando-se em subconjuntos específicos de dados de entrada. Isto dá ao modelo a capacidade de avaliar a importância relativa de várias características ou regiões, captando com êxito padrões complexos e gerindo ligações de longo alcance [77].

Esta secção aborda as técnicas e conceitos mencionados anteriormente, como a aprendizagem por transferência, o mecanismo de atenção e, por último, os modelos de *deep learning* utilizados, salientando as suas principais características e vantagens.

4.3.1 Aprendizagem por transferência

A aprendizagem por transferência é uma técnica útil no âmbito de *deep learning* que aproveita o conhecimento adquirido com modelos pré-treinados para melhorar o desempenho dos modelos em novas tarefas ou conjuntos de dados com menos dados rotulados. Na aprendizagem por transferência, um modelo é inicialmente treinado num conjunto de dados de grande escala, frequentemente com milhões de exemplos rotulados e, no caso deste projeto, em *ImageNet dataset*

[78], uma base de dados pública popular. Este modelo pré-treinado capta informações úteis, sendo a sua aplicação bem generalizada de forma a poder ser aplicado numa vasta gama de atividades relacionadas.

A aprendizagem por transferência permite a transferência de conhecimentos da tarefa de origem para uma tarefa de destino, mesmo que as duas atividades sejam diferentes. Os pesos adquiridos durante o treino neste conjunto de dados podem ser utilizados para categorizar outro conjunto de dados completamente não relacionado, em vez de inicializar aleatoriamente os pesos desde o início. Esta transferência pode ser utilizada como base do modelo ou apenas para extração de características. A primeira implica pegar num modelo pré-treinado e remover as camadas originais totalmente conectadas, ajustar os pesos do modelo e adicionar uma nova camada de classificador que satisfaça a nova tarefa. A extração de características, por outro lado, implica a utilização do modelo pré-treinado como um extrator de características fixo, em que as representações aprendidas são introduzidas num novo classificador adaptado à tarefa-alvo [76]. Dado à sua elevada utilidade, todos os modelos de *deep learning* explorados nesta dissertação implementam este método.

4.3.2 Função de perda

Os modelos de *deep learning* incluem funções de perda como um componente fundamental, particularmente para tarefas baseadas na classificação. Estas funções fornecem uma medida da eficácia do funcionamento do modelo, quantificando a diferença entre os resultados projetados e os rótulos verdadeiros. As funções de perda funcionam como o objetivo de otimização do processo de formação do modelo, orientando o algoritmo de forma a reduzir os erros e melhorar as previsões do modelo [79].

Assim, nesta dissertação foram comparados os desempenhos dos modelos quando aplicadas duas funções de perda distintas: perda de entropia cruzada e perda focal.

4.3.2.1 Perda de entropia cruzada

Uma função de perda comum para problemas de classificação é a perda de entropia cruzada, normalmente designada por perda logarítmica. O objetivo é procurar discrepâncias entre as previsões do modelo e as classes reais do conjunto de dados [79]. A variação entre a probabilidade prevista pelo modelo e as probabilidades ideais associadas às classes corretas é calculada como a perda de entropia cruzada. O modelo produz previsões mais precisas e mais próximas da realidade quando existe uma diferença menor entre as probabilidades corretas e as probabilidades projetadas, o que resulta num valor de perda mais baixo [79].

4.3.2.2 Perda focal

Foi criada uma função de perda específica denominada perda focal para lidar com o desequilíbrio e o problema das amostras difíceis de classificar em tarefas de classificação. Acrescenta um elemento modulador que diminui a contribuição da perda de instâncias corretamente categorizadas, dando às amostras problemáticas mais atenção [80]. A perda focal promove o modelo para se concentrar em exemplos difíceis, reduzindo a influência de dados negativos simples, melhorando o desempenho geral.

4.3.3 Algoritmos de otimização

Os algoritmos de otimização são essenciais para treinar modelos de *deep learning* para reduzir a função de perda e melhorar o desempenho. A atualização iterativa dos parâmetros de um modelo com base nos gradientes da função de perda, relativamente a esses parâmetros, é o objetivo de um procedimento de otimização. O objetivo deste algoritmo é identificar o conjunto ideal de parâmetros que minimizam a perda e melhoram a generalização do modelo para novos dados.

4.3.3.1 Estimação adaptativa de momentos (Adam)

Ao calcular médias móveis exponenciais dos gradientes (primeiros momentos) e dos gradientes ao quadrado (segundos momentos), o Adam mantém uma taxa de aprendizagem adaptativa para cada parâmetro. De modo a fornecer estimativas fiáveis, a correção de enviesamento também está incluída. O sistema modifica a taxa de aprendizagem em função do primeiro e segundo momentos estimados, o que lhe permite adaptar-se a diferentes necessidades de taxa de aprendizagem para vários parâmetros e convergir mais rapidamente. O Adam é conhecido pela sua fiabilidade, eficácia e forte desempenho numa variedade de tarefas de *deep learning* [81].

4.3.3.2 Descida de Gradiente Estocástico (SGD)

Este algoritmo modifica os parâmetros do modelo com base nos gradientes da função de perda calculados em pequenos lotes de dados de treino. Utilizando o mini lote escolhido, determina os gradientes da função de perda relativamente aos parâmetros e atualiza os parâmetros deduzindo a taxa de aprendizagem multiplicada pelos gradientes, repetindo este processo iterativamente até que uma condição de convergência seja satisfeita, o que pode demorar várias épocas. Apesar de ser direto e simples de utilizar, pode ser sensível à taxa de aprendizagem e pode convergir lentamente ou ficar preso em soluções não ideais [82].

4.3.4 Programador da taxa de aprendizagem

A taxa de aprendizagem é um hiperparâmetro fundamental que afeta o tamanho do passo de otimização durante o treino de modelos de *deep learning*. A qualidade da solução final e a velocidade de convergência são diretamente afetadas pela taxa de aprendizagem [83]. No entanto, definir uma taxa de aprendizagem ideal pode ser difícil porque nem todos os cenários de formação podem ser adequados para uma determinada taxa de aprendizagem. A capacidade de modificar dinamicamente a taxa de aprendizagem durante o treino com programadores de taxa de aprendizagem permite uma otimização mais eficaz e eficiente.

Nesta dissertação foram eleitos dois programadores de aprendizagem distintos a serem comparados: redução da taxa de aprendizagem de passo e redução de cosseno.

4.3.4.1 Redução da taxa de aprendizagem de passo

Um método essencial para o treino de redes neurais é a alteração da taxa de aprendizagem por etapas, que envolve o ajuste sistemático da taxa de aprendizagem durante o treino, de modo a maximizar a convergência e o desempenho do modelo. O tamanho do passo das atualizações dos parâmetros é determinado pela taxa de aprendizagem, um hiperparâmetro crucial, com grande impacto nos resultados do treino. No final de cada passo, o programador de passos aumenta a taxa de aprendizagem por um fator especificado, dividindo o treino em épocas distintas. A abordagem suporta a regularização implícita e a convergência estável, mas é necessário ter em conta a sensibilidade dos parâmetros e a falta de adaptabilidade [84]. A chave para utilizar eficazmente as vantagens do programador da taxa de aprendizagem de passo em várias aplicações de *deep learning* é a seleção cuidadosa dos parâmetros através da experimentação.

4.3.4.2 Redução do cosseno

Outra opção frequentemente utilizada para modificar a taxa de aprendizagem durante o treino passa pela utilização do programador da taxa de aprendizagem por redução do cosseno. Com base na programação da redução do cosseno, a taxa de aprendizagem é reduzida de forma cíclica. O objetivo é criar uma curva semelhante ao cosseno, começando com uma taxa de aprendizagem maior, baixando-a gradualmente até um valor mínimo e, em seguida, aumentando-a novamente. O programador é utilizado para atualizar a taxa de aprendizagem durante o treino em cada época [83].

Ao diminuir e aumentar ciclicamente a taxa de aprendizagem, é possível que a taxa de aprendizagem explore várias áreas do cenário de perdas. Este comportamento cíclico ajuda o processo de otimização a evitar respostas menos que ideais e pode mesmo levar à descoberta de

opções superiores nas fases posteriores da formação.

4.3.5 Mecanismo de atenção

Os modelos de *deep learning* têm agora a capacidade de se concentrar seletivamente nas partes mais pertinentes da entrada, graças ao desenvolvimento de mecanismos de atenção. Os modelos podem utilizar processos de atenção para dar vários níveis de ênfase a determinadas áreas ou aspectos de uma entrada. Os mecanismos de atenção melhoram a capacidade do modelo para reconhecer padrões importantes e produzir previsões precisas, atendendo seletivamente a regiões instrutivas. Isto é especialmente útil na classificação de imagens médicas, uma vez que é crucial localizar e identificar características específicas de patologias. Segundo Zhaoyang et al. [77] (2021), podemos categorizar o mecanismo de atenção em quatro categorias não necessariamente independentes: suavidade da atenção, forma dos dados de entrada, representações de entrada, representações de saída.

A suavidade da atenção, permite alterações constantes e suaves nos pesos da atenção. Quer a atenção se concentre em elementos sequenciais ou em regiões espaciais das imagens, a forma como a atenção é aplicada depende da estrutura dos dados de entrada. Enquanto as representações de saída registam as representações combinadas ou ponderadas produzidas pelo mecanismo de atenção, as representações de entrada codificam os dados para a atenção, tais como os encaixes de palavras ou os atributos da imagem. Estes fatores trabalham em conjunto para melhorar a capacidade do modelo de seleção de informação crítica e facilitar os cálculos ou previsões subsequentes.

Neste projeto, dadas as restrições computacionais e de tempo útil, foram apenas comparados dois tipos de mecanismo de atenção de forma a perceber a sua relevância no desempenho do modelo. Os mecanismos de atenção comparados foram: atenção suave e atenção de múltiplas cabeças, por serem considerados os mais relevantes na extração de características de imagens. Apesar da comparação ter sido realizada com apenas dois mecanismos, é apresentada em seguida uma ideia geral de todos os mecanismos disponíveis.

4.3.5.1 Suavidade da atenção

O grau de flexibilidade na determinação da relevância ou peso de várias características na entrada é referido como a suavidade de um mecanismo de atenção. A atenção suave permite que o modelo dê ênfase a certos elementos em detrimento de outros, porque os pesos são contínuos e podem variar suavemente. O modelo pode captar relações e padrões complexos nos dados devido à sua flexibilidade. Atenção rígida (*hard attention*), pelo contrário, utiliza pesos discretos ou binários,

o que resulta numa atribuição de atenção mais direcionada e discriminatória. A atenção dá prioridade a uma característica em detrimento de outras, ignorando efetivamente as restantes, reduzindo a capacidade do modelo para detetar pequenas flutuações que outras características poderiam representar.

Outras possíveis suavidades são: atenção global ou local. Ao calcular os pesos da atenção, a atenção global tem em conta toda a sequência de entrada, permitindo ao modelo prestar atenção a todos os elementos ao mesmo tempo. A atenção local, por outro lado, limita a atenção a uma determinada área ou janela da entrada, permitindo que o modelo se concentre em padrões regionais e reduza a complexidade computacional [77].

4.3.5.2 Forma dos dados de entrada

Os dados de entrada ao mecanismo de atenção são as próprias características extraídas. Estas *features* podem ser representadas em termos de item ou de localização [77]. As características de entrada para a atenção por item são itens explícitos, como píxeis distintos. O sistema de atenção atribui a cada item um peso de importância, que também avalia a relevância dos elementos de entrada individuais. Isto permite que o modelo se concentre em pormenores minuciosos e padrões regionais. Em contraste, a atenção em termos de localização tem em conta áreas mais extensas ou posições espaciais nos dados de entrada permitindo ao modelo registar dados contextuais e dependências geográficas. O sistema de atenção atribui pesos de valor variáveis a locais ou regiões. Isto permite que o modelo reconheça estruturas e conexões amplas na entrada.

4.3.5.3 Representações de entrada

Quando se trata de extrair e utilizar informações pertinentes dos dados de entrada, as representações de entrada dos processos de atenção são muito importantes. Diferentes representações de entrada, tais como distintiva, auto, co-atenção e hierárquica, são frequentemente utilizadas nos mecanismos de atenção [77].

A atenção distintiva concentra-se em biomarcadores específicos ou em qualidades de imagem únicas para a atividade em causa. Permite que o modelo preste atenção a estas características distintivas, que podem ser um sinal de padrões ou propriedades cruciais para a tarefa de categorização. Por outro lado, a auto-atenção permite que o modelo reconheça dependências na própria entrada com base nas ligações entre os vários itens de entrada, calculando as pontuações de atenção.

A co-atenção é a atenção simultânea a vários itens de entrada. Tem em conta as ligações entre as várias entradas e calcula as pontuações de atenção em conformidade. Por fim, o objetivo

da atenção hierárquica é identificar as dependências na entrada em vários níveis de tamanho para que o modelo se concentre tanto na informação local como na global.

4.3.5.4 Representações de saída

Segundo Zhaoyang et al. [77], a representação de saída produzida pelo módulo de atenção pode ser de saída única, de múltiplas cabeças ou multidimensional. Uma única soma ponderada das características de entrada é calculada num mecanismo de atenção de saída única, produzindo uma representação consolidada, ou seja, uma característica. Com este método, os dados mais importantes da entrada são captados numa representação sucinta e consolidada. O mecanismo de atenção de múltiplas cabeças, por outro lado, calcula vários conjuntos de pesos de atenção e representações de saída em simultâneo. Ao concentrar-se em pontos de vista distintos da entrada, o modelo pode aprender várias representações. A representação final de saída é criada pela concatenação ou combinação destas representações. O método de múltiplas cabeças dá ao modelo uma flexibilidade adicional e a capacidade de reconhecer padrões e relações intrincados nos dados.

Finalmente, a representação de saída pode ocasionalmente ser multidimensional, o que significa que incorpora dados de uma variedade de modalidades ou dimensões. Num mecanismo de atenção multidimensional, por exemplo, o módulo de atenção pode prestar atenção a vários canais de entrada ou mapas de características, permitindo ao modelo aprender representações que refletem relações espaciais e entre canais.

4.3.6 Modelos implementados

Foi selecionado um subconjunto de modelos *deep learning* com base no estado de arte deste projeto. Devido à elevada necessidade de recursos para o treino destes modelos, não possível realizar uma análise mais abrangente quanto aos melhores modelos a serem selecionados. Por este motivo foram selecionados aqueles que obtiveram melhor desempenho nos artigos mencionados no estado de arte. Assim os modelos selecionados foram: *DenseNet 121*, *VGG 16*, *VGG 19*, *ResNet 50* e *Inception V3*. O objetivo deste subconjunto é classificar os dados de acordo com grupos ou classes específicos com base em padrões encontrados em dados rotulados.

Devido à sua adaptabilidade, a biblioteca *Pytorch* foi escolhida para implementar todos os modelos. Cada classificador é adequado para uma variedade de situações devido às suas vantagens e características técnicas únicas. Cada classificador será descrito em termos gerais, juntamente com uma descrição do seu funcionamento.

4.3.6.1 DenseNet 121

Numa rede convolucional densa (DenseNet), é efetuada uma ligação de alimentação entre cada camada e cada uma das outras. Neste modelo existem $\frac{L(L+1)}{2}$ ligações diretas, diferenciando-se das redes convulsionais habituais que possuem L camadas e L ligações, uma entre cada camada e a camada seguinte. Os mapas de caraterísticas de todas as camadas anteriores são utilizados como entradas para cada camada e os seus próprios mapas de caraterísticas são utilizados como entradas para todos os níveis posteriores [85].

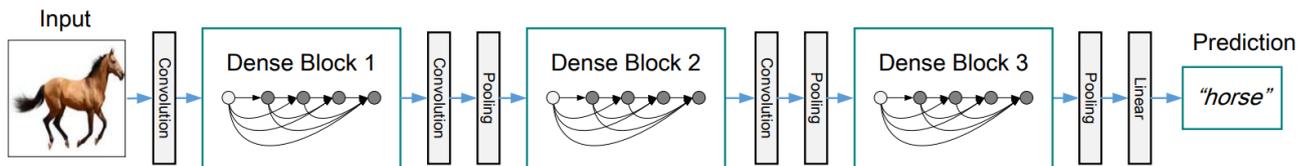


Figura 4.3: DenseNet com três camadas densas [85]

O modelo *DenseNet 121* é constituído por quatro blocos densos, cada um com várias camadas densas. Cada uma camada densa é formada por uma camada de normalização, uma função de ativação de unidade linear retificada (ReLU) e uma camada convolucional 3x3. As camadas de transição são utilizadas entre os blocos densos para regular a dimensionalidade e facilitar a redução da amostragem. Os blocos densos são ligados em série. Para produzir previsões de classe, a camada de classificação final utiliza o agrupamento médio global e uma camada totalmente ligada. É possível visualizar esta arquitetura na Figura 4.3

De forma a minimizar as dimensões espaciais, a imagem de entrada passa primeiro por uma camada convolucional inicial e depois por uma camada de subamostragem máxima. Em seguida, passa para os blocos densos. A saída de cada camada densa é combinada com todas as saídas anteriores em cada bloco denso para formar um padrão de conectividade densa. Como resultado deste procedimento de concatenação, a reutilização de caraterísticas e o fluxo de informação eficiente são encorajados e tornados possíveis para todas as camadas anteriores.

A quantidade de mapas de caraterísticas entre blocos densos é reduzida pelas camadas de transição. Estas são constituídas por uma camada para normalização de grupos, uma camada para redução da dimensionalidade e uma camada denominada subamostragem média para reduzir a amostragem das dimensões espaciais. Isto facilita o fluxo de informação e controla a complexidade do modelo.

O agrupamento médio global, que condensa as dimensões espaciais num tamanho fixo, surge após o último bloco denso. Uma camada totalmente conectada com ativação *softmax* é então

adicionada para criar a distribuição de probabilidade sobre as classes da imagem de entrada.

Este modelo tem uma série de vantagens. A rede pode propagar características complexas devido ao padrão de conectividade densa, que também suporta o fluxo de gradientes. Resolve o problema do desaparecimento de gradientes [85] e promove a reutilização de características, melhorando o desempenho do modelo e acelerando a aprendizagem. Além disso, demonstra uma forte capacidade de generalização, o que a torna útil mesmo com poucos dados de treino.

4.3.6.2 VGG 16 e VGG 19

Tanto o *VGG 16* como o *VGG 19* aderem ao conceito de arquitetura geral da rede *VGG*. São constituídos por camadas convulsionais empilhadas que são seguidas de camadas totalmente ligadas para classificação. A profundidade das redes e o número de camadas convulsionais têm um efeito significativo.

Enquanto o *VGG 19* tem 19 camadas, incluindo 16 camadas convulsionais e três camadas totalmente ligadas, o *VGG 16* tem 16 camadas, incluindo 13 camadas convulsionais. Em comparação com a *VGG 16*, a *VGG 19* contém mais três camadas convulsionais, o que a torna mais profunda e talvez capaz de captar características mais complexas.

São utilizados filtros 3x3 nas camadas convulsionais. A captura bem-sucedida de padrões e detalhes locais nas imagens de entrada é possível graças ao emprego de filtros de tamanho reduzido [86]. As camadas de subamostragem máxima são utilizadas após uma sequência de camadas convulsionais com ativação ReLU na imagem de entrada, a fim de minimizar as dimensões espaciais. As camadas de convolução captam detalhes de alto e baixo nível, extraíndo características a vários níveis de abstração. A saída é então transformada num vetor antes de ser classificada por camadas totalmente conectadas. A arquitetura do modelo *VGG 16* pode ser visualizada na Figura 4.4, sendo que como foi mencionado, a arquitetura do modelo *VGG 19* apenas difere na quantidade de camadas.

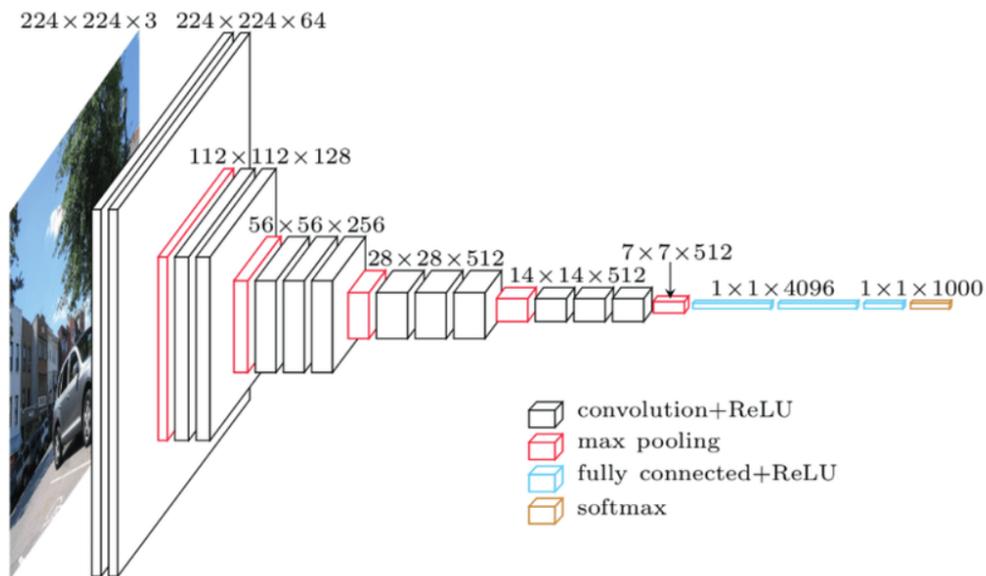


Figura 4.4: Arquitetura do modelo *VGG 16* [86]

4.3.6.3 *ResNet 50*

O termo *ResNet 50* significa: rede residual com 50 camadas. Camadas convulsionais, camadas de agrupamento, camadas totalmente conectadas e conexões de salto (também conhecidas como conexões residuais) estão entre as 50 camadas que compõem a arquitetura desta rede neuronal convolucional profunda. A principal inovação deste modelo é a adição de blocos residuais, que melhoram o fluxo de gradiente e ajudam a resolver o problema do gradiente de desaparecimento em redes profundas.

Após a normalização em grupo e a ativação ReLU, uma sucessão de camadas convulsionais é aplicada à imagem de entrada para a processar. Os blocos residuais, apresentam conexões de salto que omitem uma ou mais camadas e aumentam a saída do bloco ao incluir a entrada original. A rede pode aprender mapeamentos residuais graças a esta ligação de salto, tornando a otimização da rede mais simples [87]. O esqueleto da rede é constituído pelos restantes blocos, os quais são colocados uns sobre os outros.

As camadas de subamostragem máxima são utilizadas para minimizar as dimensões espaciais à medida que a entrada flui através dos blocos restantes. Após a transformação num vetor, as camadas totalmente conectadas são alimentadas com os mapas de características resultantes. Utilizando a ativação *softmax*, as camadas totalmente ligadas combinam a informação aprendida para produzir previsões de classe. A arquitetura de um modelo *ResNet 34* pode ser visualizado na Figura 4.5, esta arquitetura é semelhante à do modelo *ResNet 50* apenas variando na quantidade de camadas.

A principal vantagem deste modelo é a possibilidade de ser treinado com a adição de conexões residuais, melhorando a aprendizagem e o desempenho das características. As conexões de salto resolvem o problema do gradiente de desaparecimento e melhoram o fluxo de gradiente, permitindo um treino mais eficiente.

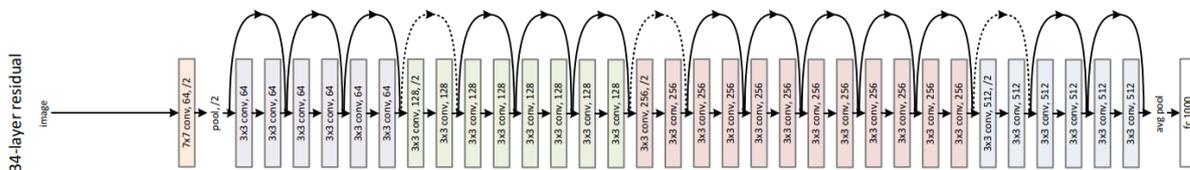


Figura 4.5: Arquitetura do modelo *ResNet 34* [87]

4.3.6.4 Inception V3

Inception V3 foi criado como membro da família de modelos *Inception*. É composto por vários módulos de incepção concebidos para captar características a várias escalas. Os filtros de convolução de vários tamanhos são utilizados em paralelo nas convoluções fatorizadas, que recolhem informações locais e globais.

As camadas de convolução são utilizadas para processar a imagem de entrada, cada uma utilizando um filtro de tamanho diferente. Os elementos fundamentais do modelo são os módulos de incepção. Cada módulo de incepção é constituído por trajetos paralelos com várias operações, tais como camadas de convolução de diferentes tamanhos e uma camada de subamostragem máxima após um dado número de camadas convulsionais paralelas [88]. A concatenação dos mapas de características de saída destas vias paralelas ao longo da dimensão do canal permite ao modelo captar uma variedade de dados. Esta arquitetura pode ser explorada na Figura 4.6

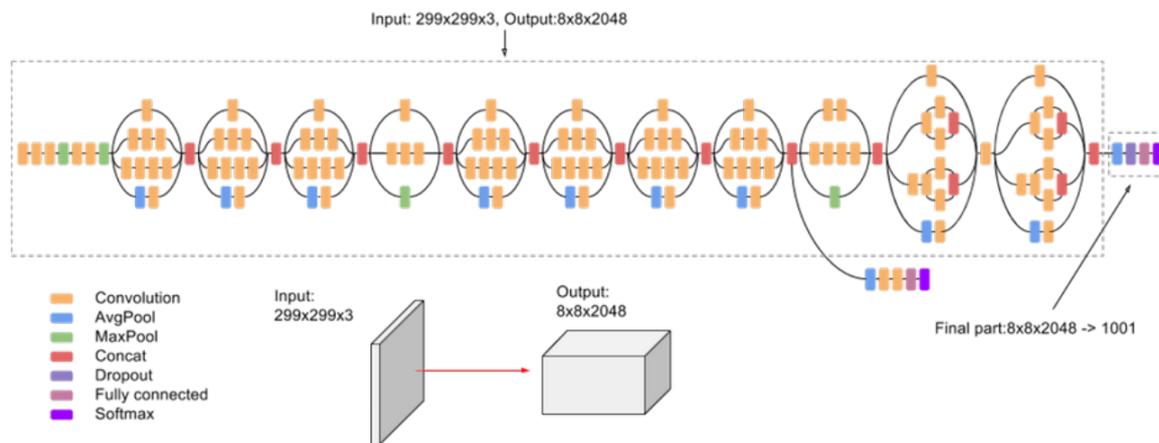


Figura 4.6: Arquitetura do modelo *Inception V3* [88]

Os classificadores auxiliares também são utilizados em camadas intermédias, durante a formação, para ajudar a evitar problemas de gradiente de desaparecimento. Em várias profundidades da rede, esses classificadores auxiliares promovem a aprendizagem de características mais significativas. Os classificadores auxiliares são eliminados durante a inferência, e as previsões finais são criadas usando os resultados da camada superior.

A rede consegue captar com êxito informações locais e globais quando são utilizados módulos de injeção com vias paralelas. Tal minimiza a complexidade computacional, mantendo ou melhorando o desempenho através da utilização de convoluções fatorizadas.

4.4 Métricas utilizadas

Determinar com exatidão o desempenho de um modelo é fundamental tanto em *machine learning* como em *deep learning*. As métricas são utilizadas para avaliar o desempenho de um modelo através do contraste entre os resultados projetados e os resultados reais. A eleição das métricas a serem utilizadas é influenciada pela tarefa em causa, pelo tipo de dados e pelo resultado pretendido. Neste projeto, foram selecionadas seis métricas, amplamente utilizadas, da biblioteca *Scikit-Learn*, incluindo exatidão, precisão, especificidade, sensibilidade, *F1 score* e área sob a curva (AUC).

Na descrição das métricas são utilizadas as nomenclaturas: verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). VP são exemplos de casos em que o modelo classificou com sucesso a classe positiva da imagem e VN são os casos em que o modelo classificou corretamente as classes negativas. Por outro lado, FP e FN são os casos em que o modelo classificou a imagem com a classe positiva quando deveria ter classificado como classe negativa e casos em que o modelo classificou a imagem como pertencendo à classe negativa quando a imagem pertencia à classe positiva, respetivamente. No contexto deste projeto, a classe positiva corresponde aos casos sarcopénicos e a classe negativa são os casos em que não se verifica a existência de sarcopenia.

4.4.1 Exatidão

A percentagem de previsões corretas entre todas as previsões é conhecida como exatidão. Mostra a eficácia com que o modelo consegue classificar as amostras. Por outro lado, para conjuntos de dados desequilibrados, onde a proporção de amostras em cada classe não é a mesma, a exatidão pode não ser a melhor métrica.

$$Exatidão = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.2)$$

Em que VP é o número de verdadeiros positivos, VN é o número de verdadeiros negativos, FP é o número de falsos positivos e FN é o número de falsos negativos [89]. Os valores de exatidão variam entre 0 e 1, onde 1 significa que o modelo classificou corretamente todos os casos.

4.4.2 Precisão

A precisão é o rácio entre os casos positivos verdadeiros e todas as previsões positivas. Mostra quantas das amostras que se esperava que apresentassem resultados positivos realmente o fizeram. Quando é importante reduzir os falsos positivos, é uma métrica eficaz.

$$Precisão = \frac{VP}{VP + FP} \quad (4.3)$$

Onde VP é o número de verdadeiros positivos e FP é o número de falsos positivos [89]. O valores variam entre 0 e 1, um valor mais próximo de 1 indica a presença de menos falsos positivos.

4.4.3 Especificidade

A percentagem de verdadeiros negativos entre todos os negativos reais é o que se designa por especificidade. A especificidade é calculada da seguinte forma:

$$Especificidade = \frac{VN}{VN + FP} \quad (4.4)$$

Em que VN é o número de verdadeiros negativos e FP o número de falsos positivos [89]. A gama de valores possíveis varia entre 0 e 1. Uma maior especificidade indica menos previsões de falsos positivos.

4.4.4 Sensibilidade

A sensibilidade é a fração de amostras positivas reais classificadas como positivas. Quando o custo de um falso negativo é substancial, esta é uma métrica de eleição.

$$Sensibilidade = \frac{VP}{VP + FN} \quad (4.5)$$

Onde VP é o número de verdadeiros positivos e FN é o número de falsos negativos [89]. Quanto menor a presença de previsões de falsos negativos, maior a sensibilidade, onde os seus valores variam no intervalo [0 1].

4.4.5 *F1 score*

Quando existe uma distribuição desigual das classes nos dados, é frequentemente utilizado o *F1 score*, uma vez que combina a precisão e a recuperação numa única estatística. Assim *F1 score* é a média harmónica da sensibilidade e da precisão.

$$F1\ score = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (4.6)$$

Em que *VP* é o número de verdadeiros positivos, *FP* é o número de falsos positivos e *FN* é o número de falsos negativos [89]. A gama de valores possíveis para esta métrica varia entre 0 e 1, quanto mais próximo de 1 melhor o equilíbrio entre falsos positivos e falsos negativos.

4.4.6 Área sob a curva (AUC)

AUC é uma estatística utilizada para avaliar a eficácia dos modelos de classificação binária. A área sob a curva da Característica de Funcionamento do Recetor (ROC) é representada pela AUC. A curva ROC é a taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos ($1 - \textit{Especificidade}$) para vários valores de limiar. AUC caracteriza assim o desempenho total do modelo em termos da sua capacidade de distinguir entre amostras positivas e negativas, sendo representado por um único número. Um melhor desempenho é indicado por uma AUC mais elevada [89]. Os valores desta métrica variam entre 0,5 e 1, onde 0,5 denota uma classificação aleatória, enquanto 1 denota uma classificação perfeita.

5 Resultados

5.1 Implementação em *Python*

Este projeto foi desenvolvido utilizando um computador portátil Lenovo Ideapad 330S-15IKB, com um CPU Intel Core i7-8550U 2.00GHz, 8.00GB de RAM e um GPU AMD Radeon(TM) 535. O desenvolvimento compreendeu o processamento das imagens, assim como a criação e teste de todo o código.

A programação foi efetuada usando linguagem de programação o *Python* devido ao seu ecossistema rico, flexível e ampla gama de recursos de documentação. Esta linguagem possui uma vasta gama de bibliotecas direcionadas aos problemas em questão, *machine learning* clássico e *deep learning*, permitindo uma implementação mais direta e de fácil compreensão. Também pela sua ampla comunidade *online*, usufrui de um apoio superior às suas bibliotecas, facilitando a resolução de possíveis problemas. O editor utilizado tanto para os modelos de *machine learning* clássico como para o aumento de dados foi o *Visual Studio Code (VS Code)* [90], um editor desenvolvido pela *Microsoft*. Para o desenvolvimento e teste dos modelos de *deep learning* recorreu-se à plataforma *Google Colab* [91] que permite a elaboração de *notebooks* em *Python*, disponibilizando placas gráficas com um elevado poder computacional.

Para a construção do código foi criado um ambiente virtual onde foram instaladas diversas bibliotecas, todas elas utilizando *Python Package Installer*, sendo uma ferramenta que permite instalar e gerir bibliotecas. Foram utilizadas as seguintes bibliotecas:

- ***OS*** [92]: módulo incorporado que permite a interagir com o sistema operativo, incluindo operações com ficheiros e diretorias.
- ***Pandas*** [93]: biblioteca de análise de dados que permite a exploração, análise e exibição de dados.
- ***NumPy*** [94]: biblioteca de computação numérica que possibilita a utilização facilitada de matrizes e matrizes multidimensionais, possuindo também uma vasta gama de funções

matemáticas.

- ***SciPy* [95]**: construída sobre a base de *Numpy*, serve para uma computação científica mais abrangente, tendo funcionalidades adicionais para a álgebra linear, integração, entre outras.
- ***SimpleITK* [96]**: biblioteca para análise de imagens médicas, proporcionando uma forma simples e intuitiva de visualizar, aplicar filtros, segmentar e extrair características das imagens.
- ***Open Source Computer Vision* [97]**: biblioteca para processamento de imagem, facilitando a detecção de objetos e extração de características de uma imagem.
- ***Scikit-Image* [98]**: biblioteca que permite o processamento de imagens em *Python*, possuindo uma grande variedade de funcionalidades para a manipulação de imagens como: filtragem, segmentação e extração de características.
- ***Pyradiomics* [99]**: biblioteca para extrair características radiômicas de imagens médicas, fornecendo um conjunto abrangente de ferramentas para quantificar características baseadas em imagens que podem ser utilizadas para análise de imagens médicas e aprendizagem de máquinas.
- ***Pyfeats* [100]**: biblioteca para extração de características, produzindo uma coleção de características de domínio temporal, de frequência e estatísticas.
- ***Scikit-Learn* [101]**: biblioteca de *machine learning*, fornecendo uma ampla variedade de ferramentas como classificação, regressão, avaliação de modelos e ainda possibilita o pré-processamento e seleção de características.
- ***Scikit-Optimize* [102]**: biblioteca que possibilita a otimização de modelos de *machine learning*, possuindo diversos algoritmos para a afinação de hiperparâmetros como, por exemplo, bayesianos.
- ***PyTorch* [103]**: biblioteca direcionada ao *deep learning*, oferecendo apoio à construção e treino de redes neurais.

5.2 Conjuntos de dados

5.2.1 Participantes

Agruparam-se as imagens em três conjuntos de dados, cada conjunto de dados foi extraído de um grupo de 30 participantes, todos do sexo feminino, sendo 10 jovens, 10 idosos ativos e 10 idosos sedentários. A idade dos participantes variou entre os 20 e os 75 anos, a altura entre 1,58 e 1,70 metros e o peso entre 60 e 75 kg. São classificados como jovens ativos aqueles que praticam esporte pelo menos uma vez por semana, idosos ativos os que participam em caminhadas pelos menos duas vezes por semana e ginásio pelo menos uma vez por semana. Por fim, idosos sedentários são aqueles que não realizam nenhum tipo de atividade física.

5.2.2 Método de aquisição

Um procedimento uniforme foi usado para adquirir as imagens para os três conjuntos de dados (Ecografia, elastografia, e Ecografia e elastografia). Um ecógrafo Acuson NX3 Elite da Siemens, fazendo uso de uma sonda linear VF12-4, foi utilizado no procedimento de aquisição.

De modo a manter a consistência em todas as aquisições, manteve-se um fator de qualidade sempre superior a 65, melhorando a precisão da classificação e da análise, bem como a qualidade da imagem.

O procedimento de imagiologia incluiu uma avaliação ecográfica bilateral em plano transversal do músculo reto femoral. As imagens foram retiradas num ponto 30 cm acima da rótula. Foi utilizada uma sonda linear multifrequência com largura de banda de 7,5 MHz a 13,5 MHz para o processo de aquisição. A resolução e a profundidade de penetração proporcionadas por esta sonda permitiram a recolha precisa de imagens das estruturas musculares. Foi possível obter múltiplas imagens de todos os participantes na localização do músculo alvo de forma consistente e fiável devido à postura padronizada.

As imagens obtidas foram depois guardadas em formato *Tagged Image File Format*(TIFF) para posterior estudo. A utilização de uma técnica de captura uniforme garante que as fotografias em cada conjunto de dados têm qualidades e características comparáveis, minimizando potenciais desvios que poderiam prejudicar o procedimento de classificação subsequente.

5.2.3 Ecografia

Existem 89 regiões de interesse (ROIs) primárias no conjunto de dados de ecografia. Destas, 46 ROIs primárias eram de pessoas idosas com sarcopenia, enquanto 43 ROIs primárias eram de pessoas jovens consideradas saudáveis. O tamanho das ROIs primárias (ROIs selecionadas

pelo profissional na área) variam de 38x49 píxeis até 99x105 píxeis. A este conjunto de dados foi atribuída a designação de conjunto de dados A.

5.2.4 Elastografia

O conjunto de dados de elastografia contém 91 ROIs primárias, incluindo 42 ROIs primárias de pessoas jovens e saudáveis e 49 ROIs primárias de pessoas idosas com sarcopenia. As ROIs primárias obtidas possuem um tamanho de 40x49 píxeis até 90x87 píxeis e definido como conjunto de dados B.

5.2.5 Ecografia e Elastografia

O conjunto de dados combinados de ecografia e elastografia possui 180 ROIs primárias, das quais, 85 destas ROIs primárias provinham de pessoas jovens consideradas saudáveis e 95 destas ROIs primárias provinham de pessoas idosas a quem tinha sido dado um diagnóstico de sarcopenia. Por consequência, o tamanho das ROIs primárias neste conjunto de dados varia de 38x49 píxeis até 99x105 píxeis. No decorrer da dissertação será expresso pela letra C, passando a ser o conjunto de dados C.

5.3 Preparação dos dados

As imagens extraídas dos participantes correspondem a uma grande área da coxa, mais concretamente ao músculo reto femoral e, uma vez que a classificação da sarcopenia se efetua apenas na massa muscular, esta é selecionada, eliminando a parte de imagem correspondente ao tecido adiposo, osso e outras eventuais estruturas presentes não correspondentes ao músculo. Para este efeito foi extraída uma região de interesse (ROI), que para efeitos de distinção foi designada por ROI primária. Esta foi extraída das imagens ecográficas e de elastografia, onde a delimitação da área foi realizada por um profissional qualificado. Na Figura 5.1 é possível verificar a localização do músculo reto femoral numa imagem ecográfica. A seguir através da Figura 5.2 ilustra o género de imagens obtidas com a ajuda do ecógrafo (a) e ainda a ROI selecionada (b).

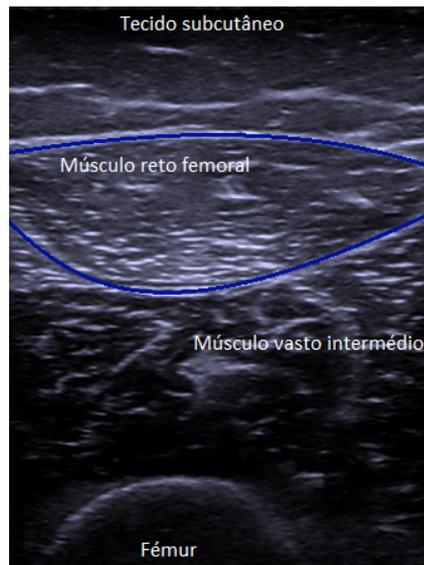


Figura 5.1: Localização do músculo reto femoral assim como de outras estruturas numa imagem ecográfica da perna.



(a) Imagem ecográfica do músculo reto femoral extraída utilizando um ecógrafo da Siemens (b) Região de interesse primária extraída por um profissional (não está proporcional à imagem 4.1(a))

Figura 5.2: Comparação entre as imagens obtidas pelo ecógrafo e a região de interesse extraída da imagem anterior

5.4 *Machine Learning* Clássico

De modo a selecionar o melhor classificador e os parâmetros utilizados no pré-processamento, foi escolhido um procedimento baseado em evidências tanto da investigação desenvolvida no estado da arte como nos testes realizados durante este projeto. Foram alvo de teste diversos métodos, comparando lado a lado diferentes abordagens para cada passo, sendo a abordagem testada sempre substituída por aquela que resultou num desempenho superior.

O procedimento eleito, mediante uma análise exploratória, foi selecionado apenas para efeitos de teste, sem qualquer fundamento substantivo. Uma janela de 32 píxeis e um passo de 23 píxeis para o método de janela deslizante (conjunto de dados n.º 35 das Tabelas B.1, B.2 e B.3) foi escolhida como conjunto de dados para o procedimento base por apresentar um número significativo de ROIs e um tamanho considerável. O algoritmo *Standard Scaler* foi escolhido para normalização, com 35 características selecionadas e uma distribuição de 20% para o grupo de teste e 80% para o grupo de treino. Por fim, a métrica de eleição foi o *F1 score*, uma vez que demonstra ser a métrica mais completa, ao refletir tanto a precisão como a sensibilidade.

Na seleção dos classificadores a utilizar, como mencionado na metodologia, foram comparados 10 dos modelos disponíveis pela biblioteca *Scikit-Learn* com uma otimização dos hiperparâmetros, obtendo-se os desempenhos representados na Tabela 5.1. A otimização efetuada foi a *Bayesian optimization* realizada com *10-fold cross-validation*, durante 100 ciclos. Os hiperparâmetros otimizados e o espaço eleito para os mesmos encontram-se no Apêndice C, na Tabela C.1. A comparação foi realizada para três conjuntos de dados distintos: A, B e C.

Tabela 5.1: Valores de *F1 score* para os 10 modelos selecionados, com *Bayesian optimization*, *10-fold cross-validation* durante 100 ciclos, para os conjuntos de dados N.º 35 das Tabelas B.1, B.2 e B.3.

Classificador	F1 score (%)		
	A	B	C
Rede Neuronal	87,06	92,00	87,42
<i>Nu Support Vector</i>	85,06	94,12	87,01
<i>Logistic Regression</i>	86,05	90,20	83,44
<i>Stochastic Gradient Descent</i>	85,39	90,20	84,00
<i>Support Vector Machine</i>	85,71	88,00	85,71
<i>Random Forest</i>	81,40	87,50	84,93
<i>Gaussian Process</i>	79,07	85,71	83,22
<i>k-Nearest Neighbors</i>	77,08	85,11	82,12
<i>Decision Tree</i>	73,27	76,60	81,29
<i>Naive Bayes</i>	72,09	80,77	65,36

Entre os 10 classificadores otimizados e testados foram selecionados os 5 classificadores com melhor desempenho para os três conjuntos de dados, realizando uma análise global. Assim, ao procedimento base, foram adicionados como classificadores a Rede Neuronal, o *Nu Support Vector*, o *Logistic Regression*, o *Stochastic Gradient Descent* e o *Support Vector Machine*. Os

hiperparâmetros selecionados através da otimização estão presentes na Tabela 5.2.

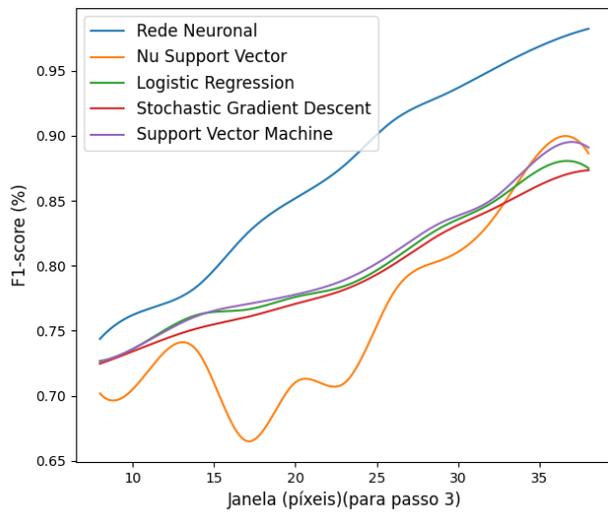
Tabela 5.2: Hiperparâmetros dos classificadores eleitos.

Classificador	Hiperparâmetros
Rede Neuronal	<ul style="list-style-type: none"> • $activation = relu, hidden_layer_sizes = 25,$ • $learning_rate = constant, max_iter = 284$
<i>Nu Support Vector</i>	<ul style="list-style-type: none"> • $degree = 5, kernel = linear, nu = 0.25$
<i>Logistic Regression</i>	<ul style="list-style-type: none"> • $C = 1.18, max_iter = 200, solver = saga$
<i>Stochastic Gradient Descent</i>	<ul style="list-style-type: none"> • $alpha = 1e - 06, eta0 = 1.0, learning_rate = invscaling,$ • $loss = hinge, penalty = l1$
<i>Support Vector Machine</i>	<ul style="list-style-type: none"> • $C = 3.83, degree = 1, gamma = 0.43, kernel = poly$

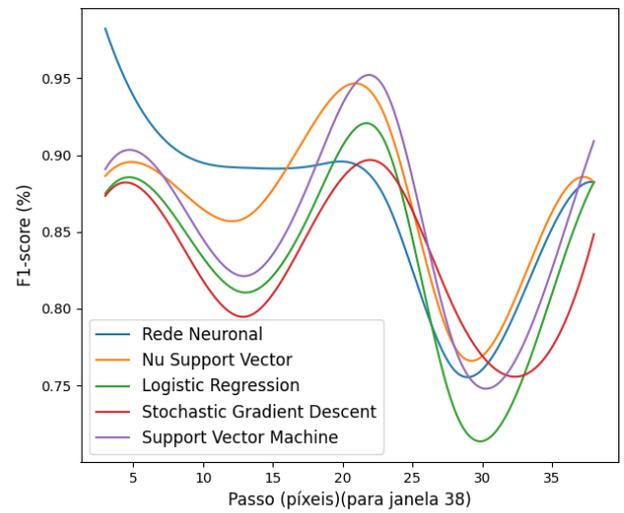
5.4.1 Tamanho e passo no método janela deslizante.

De modo a perceber a influência do tamanho da janela e do passo selecionado aquando da utilização do método janela deslizante, foram criados 51 conjuntos de dados distintos. O tamanho da janela foi variado de 5x5 píxeis até 38x38 píxeis, em incrementos de 3 píxeis, uma vez que a imagem mais pequena possui 38x49 píxeis. O tamanho do passo foi variado desde 3 píxeis até ao tamanho da janela em incrementos de 5 píxeis, sendo que quando o tamanho do passo iguala o tamanho da janela não há sobreposição de ROIs secundárias. A descrição destes conjuntos de dados criados encontra-se no Apêndice B, nas Tabelas B.1, B.2 e B.3.

Os desempenhos dos classificadores para estes *datsets* encontram-se nas Figuras 5.3, 5.4 e 5.5.

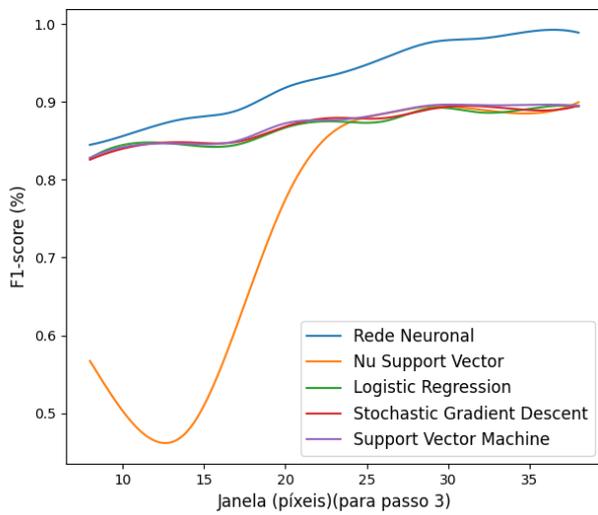


(a) Variação da janela para passo fixo

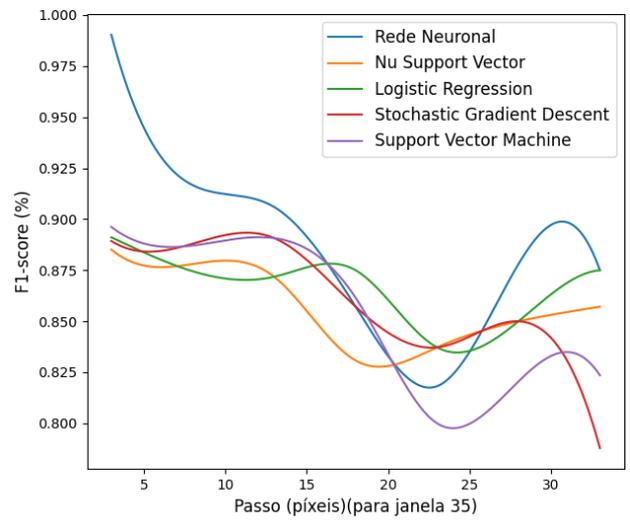


(b) Variação do passo para janela fixa

Figura 5.3: Desempenho dos cinco classificadores eleitos para o conjunto de dados A.

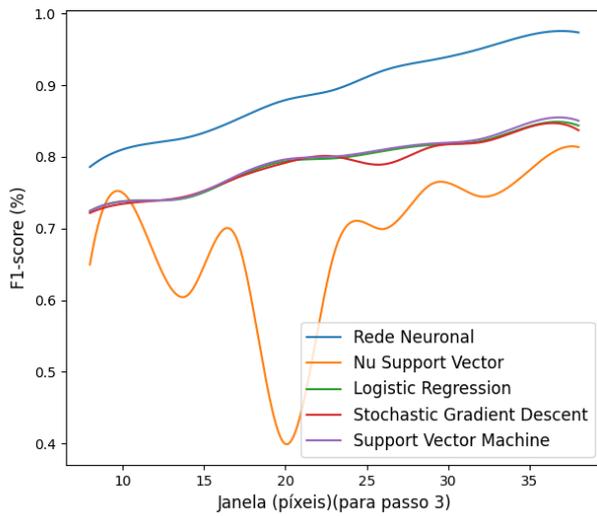


(a) Variação da janela para passo fixo

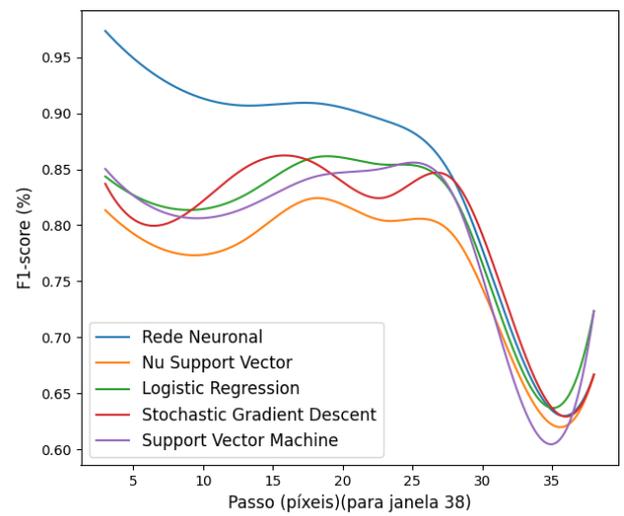


(b) Variação do passo para janela fixa

Figura 5.4: Desempenho dos cinco classificadores eleitos para o conjunto de dados B.



(a) Variação da janela para passo fixo



(b) Variação do passo para janela fixa

Figura 5.5: Desempenho dos cinco classificadores eleitos para o conjunto de dados C.

Observando os gráficos, nota-se que cada classificador, apesar de algumas oscilações, apresenta uma melhoria gradual do desempenho associada ao aumento do tamanho da janela (Figuras 5.3a, 5.4a e 5.5a). Isto é compreensível, uma vez que quanto maior for o tamanho dos dados de entrada, mais características são suscetíveis de serem extraídas. Aumentar o tamanho da janela também permite a incorporação de informações mais variadas e talvez irrelevantes em cada janela. Esta imprevisibilidade adicional pode causar confusão e tornar mais difícil o modelo discernir entre padrões importantes e oscilações aleatórias. Como resultado, ocorrem oscilações de desempenho devido às previsões do modelo tornarem-se mais irregulares.

Por outro lado, o desempenho torna-se mais irregular à medida que o tamanho do passo aumenta. Os tamanhos de passo menores resultam numa maior sobreposição, melhorando o desempenho, ao permitir que o modelo reúna dados mais precisos e pormenorizados sobre os padrões subjacentes. Estes padrões subjacentes são mais difíceis de captar com um tamanho de passo maior, mas observa-se alguma melhoria em alguns passos. Isto deve-se ao facto de um tamanho de passo maior permitir ao modelo captar uma perspetiva mais ampla dos dados, concentrando-se em padrões de nível superior e tendências globais, em vez de pormenores minuciosos. Este ponto de vista mais alargado pode aumentar a capacidade de generalização do modelo e ajudá-lo a gerar previsões com maior robustez, o que melhoraria o desempenho.

Com o aumento do tamanho do passo, observamos um declínio mais consistente no desempenho dos conjuntos de dados B e C. Isto pode ser explicado pelo facto de a quantidade de ROIs secundárias produzidas, diminuir com o aumento do tamanho do passo, tornando

muito provável a ocorrência de sobre ajuste dos dados.

Concluí-se que os classificadores têm o melhor desempenho com ROIs maiores e um grupo de treino significativo, ou seja, um passo menor. As oscilações mais abruptas também podem ser atribuídas à maior sensibilidade dos classificadores ao tamanho do grupo de treino.

Na experimentação inicial, foram testados diferentes tamanhos de janela e de passo para identificar a combinação que produzia o melhor desempenho para cada classificador. Esta abordagem visava encontrar a configuração ideal que maximizasse o desempenho da classificação. No entanto, à medida que a avaliação avançava, tornou-se evidente que alguns classificadores, que inicialmente demonstravam um desempenho superior com um conjunto de dados menor, chegavam a um ponto de sobre ajuste mais cedo quando alterado o número de características selecionadas (comparação final dos métodos do modelo na dissertação) que quando aplicados a conjuntos de dados maiores. Esta limitação impediu a realização de todo o potencial do modelo, uma vez que os modelos sobre ajustados não obtiveram uma boa generalização do modelo, diminuindo o desempenho.

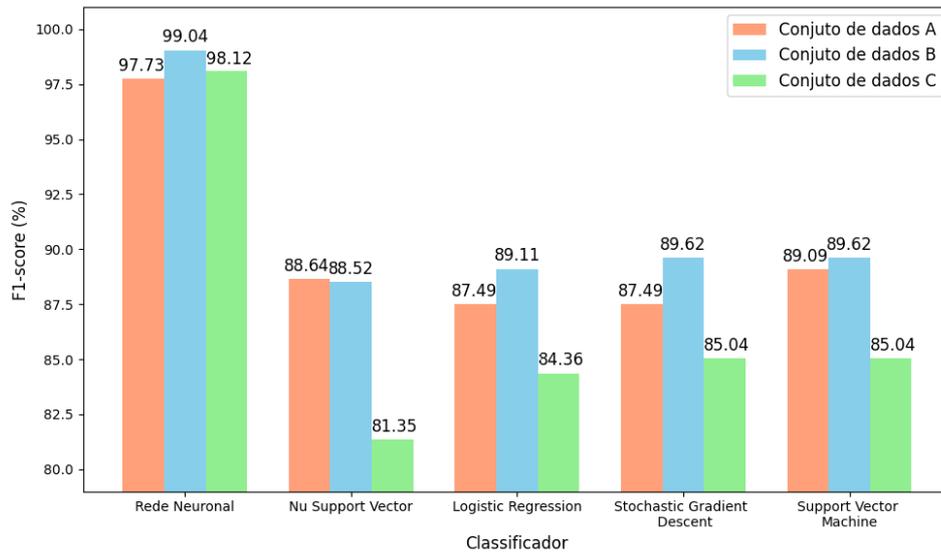
Para resolver este problema, foi adotada uma abordagem diferente. Em vez de selecionar a janela e o tamanho do passo com base apenas no desempenho inicial do classificador, a janela e o tamanho do passo foram escolhidos com base no modelo que apresentou o melhor desempenho global, tendo em conta o tamanho do conjunto de dados. Ao dar prioridade ao modelo com o melhor desempenho, independentemente do seu desempenho inicial em conjuntos de dados menores, o objetivo era selecionar uma configuração que se adaptasse melhor ao conjunto de dados maior e atenuasse o risco de sobre ajuste.

Finalmente, a janela e o passo foram selecionados com base no desempenho da Rede Neuronal que no caso do conjunto de dados A obteve um *F1 score* de 98,22%, por isso foi selecionada a janela com 38x38 píxeis e um passo com 3 píxeis. Do mesmo modo, para o conjunto de dados B obteve o melhor desempenho de 99,04%, logo foi eleita a janela 35x35 píxeis e um passo de 3 píxeis e por fim para o conjunto de dados C o melhor desempenho foi de 97,35% sendo por isso selecionada a janela 38x38 píxeis e um passo de 3 píxeis.

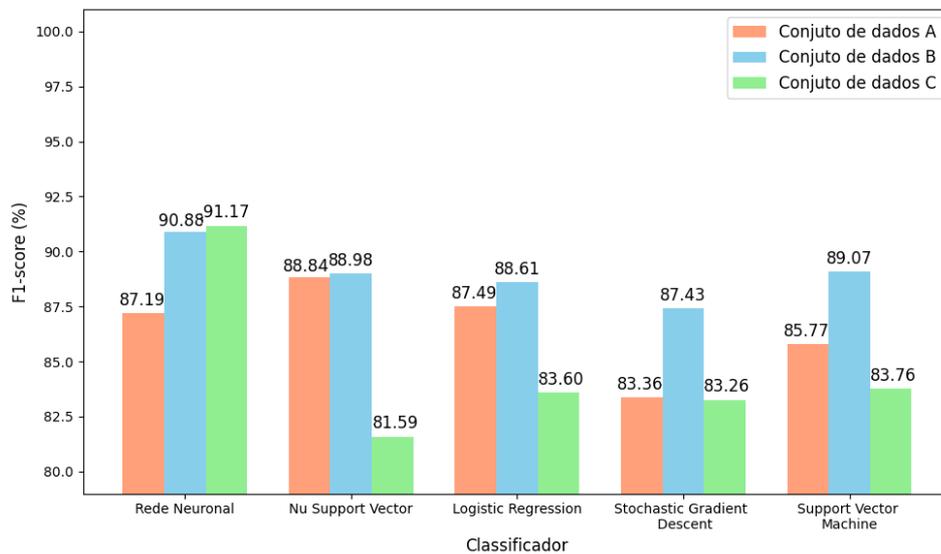
5.4.2 Algoritmo de normalização de caraterísticas (*Standard Scaler vs MinMax Scaler*)

A comparação de várias técnicas de normalização de caraterísticas é essencial porque permite a seleção do método mais adequado para um determinado conjunto de dados e objetivo de modelação. A normalização de caraterísticas é importante porque promove uma convergência mais rápida durante o treino e impede o domínio de caraterísticas com escalas superiores,

colocando as características num intervalo comparável. Desta forma, garante-se uma representação imparcial e evita-se enviesamentos nas previsões do modelo, melhorando o desempenho em geral. Foram assim comparados dois algoritmos de normalização: *standard scaler* e *minMax scaler*. Os desempenhos dos classificadores, para cada um dos métodos, encontram-se no ilustrados na Figura 5.6.



(a) *Standard Scaler*



(b) *MinMax Scaler*

Figura 5.6: Desempenho dos cinco classificadores eleitos para diferentes métodos de normalização de características.

Da análise dos gráficos conclui-se que a mudança de uma abordagem para outra normalmente

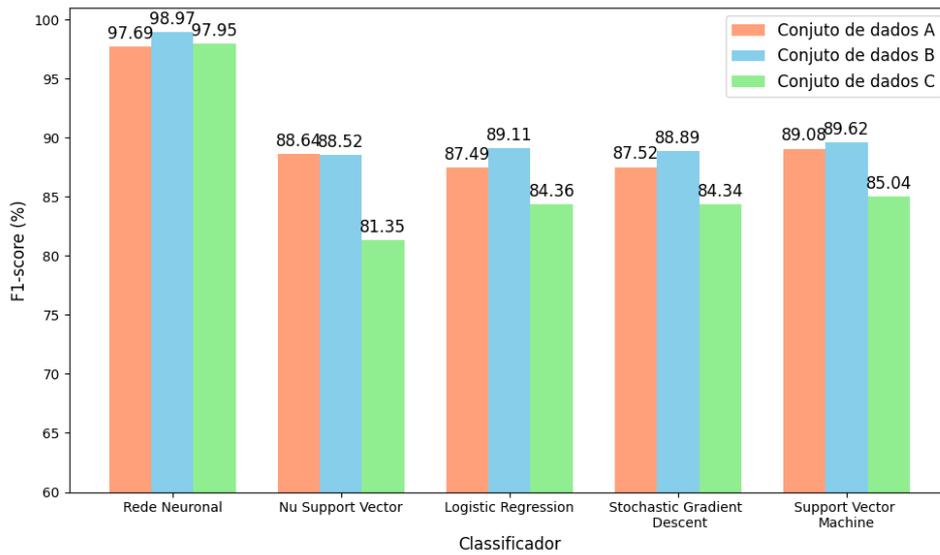
tem pouco ou nenhum efeito sobre o desempenho do classificador. A maior diferença reside na Rede Neuronal, devendo-se ao modo de processamento das características pelo mesmo.

Devido à dependência da RN de uma otimização baseada em gradiente, o *standard scaler* que normaliza os dados de modo a terem média zero e variância unitária, é especialmente útil para este modelo. Ao padronizar os recursos, a RN pode aprender pesos ideais mais rapidamente, levando a um melhor desempenho. O intervalo de valores pode ser comprimido pelo *minmax scaler*, que dimensiona os recursos para um intervalo predeterminado, o que pode resultar em perda de informação para a RN. Como resultado, quando comparado com o *minmax scaler*, a RN pode ter um desempenho muito melhor utilizando o *standard scaler*.

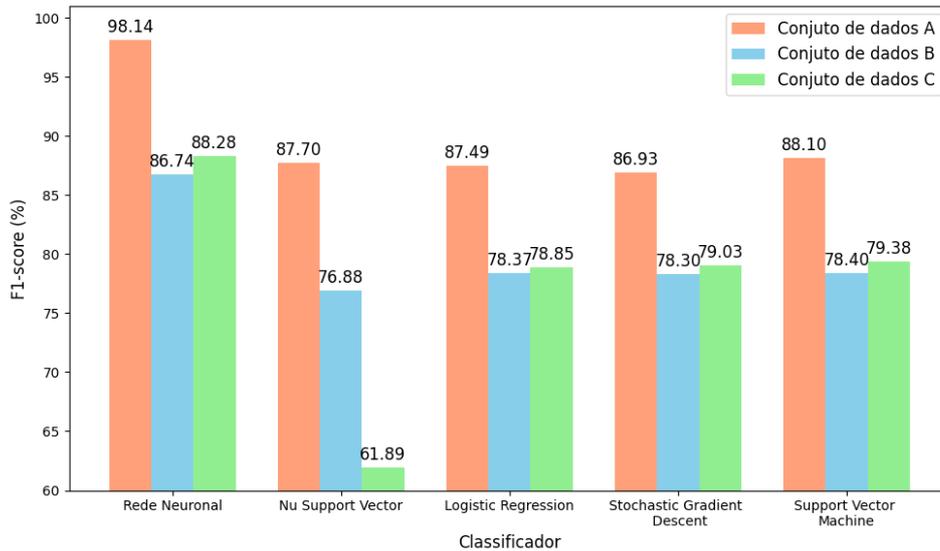
É claro da observação da Figura 5.6 que o classificador com melhor desempenho é a Rede Neuronal com um *F1 score* de 97,73%, para o método de normalização *Standard Scaler*, para o conjunto de dados A, de 99,04% para o conjunto de dados B e para o conjunto de dados C um *F1 score* de 98,12%. Em função dos resultados, foi selecionado o método *Standard Scaler* como método de normalização de características dos três conjuntos de dados.

5.4.3 Algoritmo de seleção de características (por variância *vs* informação mútua)

A comparação de várias técnicas de seleção de características pode ajudar a criar melhores modelos e a aumentar a eficácia dos seus processos de computação. Selecionando as características mais pertinentes é possível melhorar os modelos preditivos, evitando o sobre ajuste do modelo. Além disso, ao identificar as características que melhor descrevem os dados consegue-se reduzir o custo de computação. Os resultados dos classificadores utilizando uma seleção por variância e por informação mútua são representados na Figura 5.7.



(a) Por análise de variância



(b) Por informação mútua

Figura 5.7: Desempenho dos cinco classificadores eleitos para diferentes métodos de seleção de características.

As técnicas *f_classif* e *mutual_info_classif* utilizadas para a seleção de características, em regra, produzem resultados similares entre modelos. Apesar de utilizarem estratégias diferentes, ambos os métodos procuram encontrar características úteis para o objetivo de modelação atual. O desempenho comparável, resulta da sua capacidade de gerir vários conjuntos de dados, eliminar a duplicação e a informação irrelevante e ajustar-se às propriedades específicas do conjunto de dados. No final, a semelhança mostra que ambas as técnicas extraem bem características discriminativas, produzindo modelos com desempenho equivalente.

A diferença de desempenho entre a seleção por informação mútua e por variância para a classificação do conjunto de dados B e conseqüentemente o grupo C, leva a crer que a medida de informação mútua pode não conseguir captar com exatidão os padrões ou relações específicas necessários para a classificação nesta modalidade. Poderá ter sido mais fácil detetar características importantes nas imagens de elastografia utilizando uma seleção por variância, que se concentra nas variações dos valores médios.

Finalmente, como mostrado pela Figura 5.7, a Rede Neuronal foi o classificador mais eficaz, obtendo um *F1 score* de 97,69% para o método de seleção por variância para o conjunto de dados A, um *F1 score* de 98,97% para o conjunto de dados B e um *F1 score* de 97,95% para o conjunto de dados C. O método baseado em variância foi eleito para a seleção de características dos três conjuntos de dados.

5.4.4 Número de características selecionadas

O número de características utilizadas pode influenciar o desempenho do método, pelo que é útil comparar o desempenho produzido por diferentes quantidades de características selecionadas, para facilitar a determinação do equilíbrio ideal entre simplicidade e precisão e orienta os processos de tomada de decisão. Também ajuda a avaliar a eficácia das estratégias de seleção de características. Ao comparar a quantidade de características, obtêm-se informações importantes sobre a relevância das características e da interpretação dos modelos. Foi comparada uma seleção desde uma até 1097 características em incrementos de 10. Os desempenhos dos modelos para estas seleções encontram-se representados nas Figuras 5.8, 5.9 e 5.10.

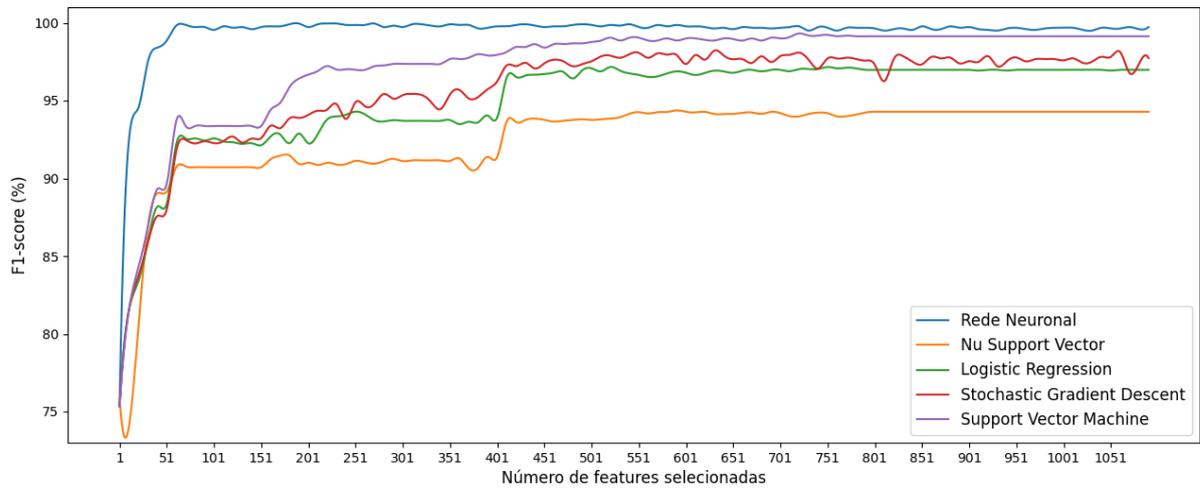


Figura 5.8: Desempenho dos cinco classificadores eleitos para diferentes números de caraterísticas selecionadas para o conjunto de dados A.

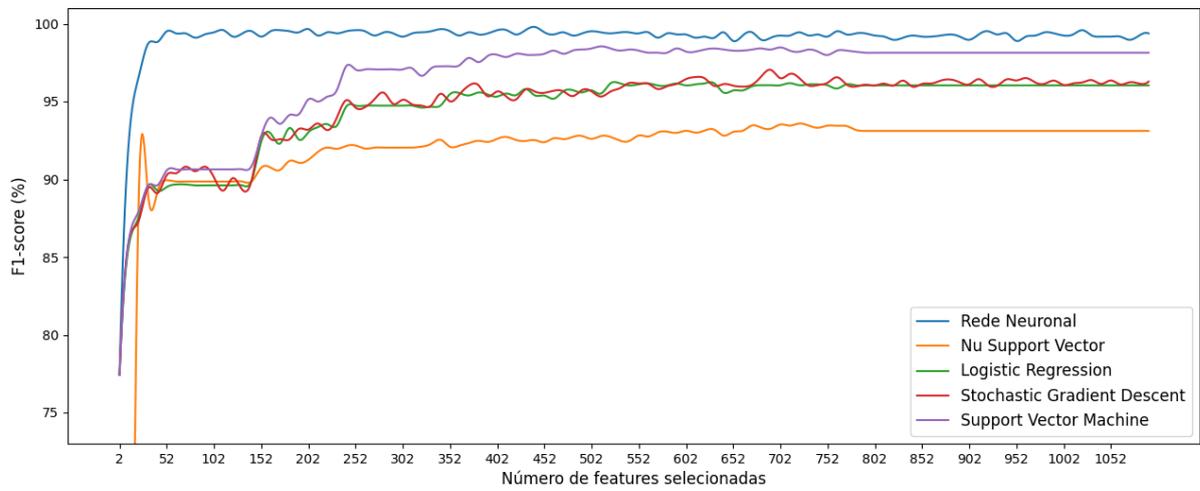


Figura 5.9: Desempenho dos cinco classificadores eleitos para diferentes números de caraterísticas selecionadas para o conjunto de dados B.

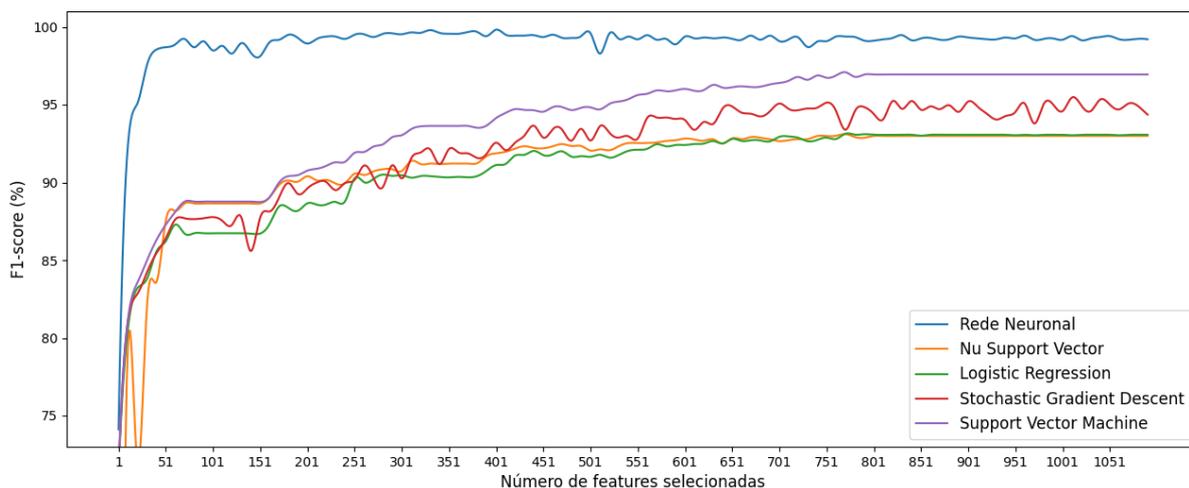


Figura 5.10: Desempenho dos cinco classificadores eleitos para diferentes números de caraterísticas selecionadas para o conjunto de dados C.

Analisando os gráficos, é possível observar o impacto que o número de caraterísticas escolhidas tem no desempenho de um modelo. Também é possível deduzir que o número de caraterísticas ideais varia de classificador para classificador. Cada classificador pode atribuir pesos diferentes a vários componentes dos dados e pode apresentar uma sensibilidade diferente a determinados subconjuntos de caraterísticas. A complexidade do problema também pode ter um impacto no número ideal de caraterísticas. Apenas a RN necessita de um número reduzido de caraterísticas para atingir o seu melhor desempenho. Deve-se à sua flexibilidade no ajuste dos seus pesos internos e funções de ativação não lineares, que permitem captar interações complexas entre caraterísticas, reduzindo a necessidade de um número de caraterísticas superior.

Mesmo quando se selecionam 1000 caraterísticas, não se observa sobre ajuste devido ao grande número de ROIs secundárias presentes nos conjuntos de dados utilizados. A abundância de ROIs secundárias garante que os classificadores não se ajustem demais a uma instrução, evitando assim o sobre ajuste.

A sarcopenia, uma doença que é difícil de detetar e classificar, exige frequentemente uma grande quantidade de caraterísticas para uma descrição precisa da classe. Este facto deve-se à complexidade e variabilidade das caraterísticas associadas à sarcopenia, que requerem uma representação abrangente para uma descrição precisa da classe.

Em suma, mesmo com um grande número de caraterísticas selecionadas, o risco de sobre ajuste é reduzido pela abundância de ROIs secundárias nos conjuntos de dados. Além disso, devido à complexidade da sarcopenia enquanto doença, é frequentemente necessário um elevado

número de características para captar com precisão todas as suas propriedades variadas.

O número ótimo de características selecionadas para cada um dos classificadores em cada um dos conjuntos de dados é representado na Tabela 5.3.

Tabela 5.3: Número ótimo de características (K) selecionadas para cada um dos classificadores e conjuntos de dados.

Classificador	Conjunto A		Conjunto B		Conjunto C	
	K	$F1$ score (%)	K	$F1$ score (%)	K	$F1$ score (%)
<i>Rede Neuronal</i>	60	99,82	50	99,52	70	99,22
<i>Nu Support Vector</i>	540	94,15	670	93,46	670	92,92
<i>Logistic Regression</i>	470	96,84	520	96,16	700	92,94
<i>Stochastic Gradient Descent</i>	510	97,84	690	97,03	750	95,13
<i>Support Vector Machine</i>	520	99,04	500	98,41	720	96,80

5.4.5 Avaliação final

Serão avaliados os classificadores utilizando uma série de métricas que permitem determinar os modelos com melhor desempenho, bem como as vantagens e desvantagens dos classificadores. Além disso, irá ser avaliado o desempenho dos classificadores em cada conjunto de dados separadamente. Através deste estudo, pode-se avaliar a robustez dos modelos em circunstâncias do mundo real e compreender até que ponto são generalizáveis em vários conjuntos de dados. Pode-se, também, avaliar diferenças ou padrões que possam surgir ao comparar o desempenho entre conjuntos de dados, o que também ajuda a compreender melhor as capacidades dos classificadores.

Tabela 5.4: Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados A.

Classificadores	Exatidão	Precisão	$F1$ score	Especificidade	Sensibilidade	AUC
<i>Rede Neuronal</i>	99,84	99,64	99,81	99,71	99,82	99,86
<i>Nu Support Vector</i>	94,84	95,49	94,03	96,58	92,62	94,60
<i>Logistic Regression</i>	97,24	97,24	96,84	97,86	96,44	97,15
<i>Stochastic Gradient Descent</i>	98,12	98,17	97,85	98,58	97,54	98,06
<i>Support Vector Machine</i>	99,16	98,91	99,04	99,15	99,18	99,16

Em termos de desempenho, como ilustrado na Tabela 5.4, o classificador da Rede Neuronal

tem o melhor *F1 score* (99,81%) e exatidão (99,84%). Tem, assim, um desempenho elevado tanto em termos de sensibilidade como de precisão. Em termos de exatidão e de *F1 score*, o SVM também tem um bom desempenho, demonstrando a sua eficácia na classificação correta das classes.

Os classificadores *Stochastic Gradient Descent Classifier* e *Logistic Regression* obtêm resultados competitivos em termos de exatidão e *F1 score*. Os valores AUC em todos os classificadores são consistentemente elevados, demonstrando uma capacidade discriminativa significativa.

Não é possível realizar uma comparação direta com outros estudos, pois para além do conjunto de dados ser distinto, a utilização destes classificadores para a classificação da sarcopenia é rara. No entanto, existem artigos onde foi aplicado o SVM de forma a identificar vários aspetos a nível muscular mediante imagens de ecografia e elastografia. Para o classificador também utilizado em outros projetos [46], [48], [47], a precisão da classificação variou entre 78% e 85%. Em contrapartida, nesta investigação, foram atingidos valores consistentes acima de 97%. A importância de variáveis adicionais, incluindo a composição do conjunto de dados, os métodos de pré-processamento e a afinação do modelo, que podem ter um impacto significativo nos resultados, é realçada pela discrepância dos resultados. É evidente que, apesar das semelhanças na estratégia, o desempenho pode diferir significativamente, dependendo do processo em estudo.

Tabela 5.5: Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados B.

Classificadores	Exatidão	Precisão	<i>F1 score</i>	Especificidade	Sensibilidade	AUC
Rede Neuronal	99,65	99,80	99,59	99,87	99,17	99,59
<i>Nu Support Vector</i>	94,25	91,53	93,32	93,57	95,19	94,38
<i>Logistic Regression</i>	96,75	95,90	96,16	96,99	96,42	96,71
<i>Stochastic Gradient Descent</i>	97,27	96,07	96,79	97,09	97,52	97,31
<i>Support Vector Machine</i>	98,61	98,89	98,34	99,20	97,80	98,50

Com a exatidão mais elevada (99,65%) e *F1 score* (99,59%), a Rede Neuronal continua a ser o classificador com melhor desempenho para o conjunto de dados B, mantendo um elevado nível de precisão e sensibilidade, evidenciado na Tabela 5.5. Com uma exatidão superior a 98% e um valor de *F1 score* competitivos, o *Support Vector Machine* possui um desempenho elevado.

Em comparação, os classificadores *Nu Support Vector* e *Logistic Regression* têm classificações de sensibilidade mais baixas, sugerindo uma dificuldade em detetar casos

positivos (com sarcopenia). Os valores AUC dos classificadores continuam a apresentar valores de elevado desempenho.

Tabela 5.6: Desempenho em percentagem de cada classificador eleito com a sua melhor configuração para o conjunto de dados C.

Classificadores	Exatidão	Precisão	<i>F1 score</i>	Especificidade	Sensibilidade	AUC
Rede Neuronal	99,31	99,46	99,20	99,59	98,93	99,26
<i>Nu Support Vector</i>	93,88	94,14	92,84	95,65	91,57	93,61
<i>Logistic Regression</i>	93,88	93,35	92,90	94,97	92,46	93,72
<i>Stochastic Gradient Descent</i>	95,68	95,12	95,01	96,28	94,90	95,59
<i>Support Vector Machine</i>	97,25	97,25	96,81	97,92	96,38	97,15

Para o conjunto de dados C, na Tabela 5.6, concluímos uma vez mais que a rede Neuronal continua a funcionar a um nível elevado com um *F1 score* elevado (99,20%) e exatidão (99,31%). Apresenta uma excelente sensibilidade e precisão. O *Support Vetor Machine* produz um desempenho equilibrado, exato e com *F1 score* elevado.

Quando comparados com outros classificadores, os valores de *F1 score* dos classificadores *Nu Support Vetor* e *Logistic Regression* são mais baixos. A capacidade dos classificadores para distinguir entre classes é demonstrada pelos elevados valores AUC dos classificadores.

Em conclusão, o classificador da Rede Neuronal apresenta continuamente o desempenho mais elevado em todos os conjuntos de dados, atingindo frequentemente os melhores valores de *F1 score*. Além disso, são demonstrados classificadores concorrentes como *Support Vetor Machine* e o *Stochastic Gradient Descent Classifier*.

5.5 *Deep Learning*

De modo a treinar e testar os modelos eleitos, foram aplicadas diversas técnicas de aumento de dados às ROIs primárias obtidas. Nomeadamente, foi aplicada uma inversão horizontal, rotações em seis ângulos diferentes (90°, 45°, 15° e respetivos simétricos) e por fim a adição de um ruído gaussiano. As características dos conjuntos de dados obtidos por este método encontram-se listados na Tabela 5.7.

Tabela 5.7: Características dos conjuntos de dados utilizados para o treino, validação e teste dos modelos de *deep learning* eleitos.

Conjunto de Dados	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
A	22 784	11 776	11 008	52:48
B	23 296	12 544	10 752	54:46
C	46 080	24 320	21 760	53:47

De modo a selecionar o melhor modelo, foi eleito um procedimento baseado em evidências tanto da investigação desenvolvida no estado da arte como dos testes realizados neste trabalho. Foram alvo de teste diversos métodos, comparando lado a lado diferentes abordagens para cada passo, fixando os restantes métodos, sendo a abordagem testada sempre substituída por aquela que resultou num desempenho superior.

De forma similar ao capítulo anterior, o procedimento eleito, mediante uma análise exploratória, foi efetuado apenas para efeitos de teste, sem qualquer fundamento substantivo. Como função de perda foi eleita a perda por Entropia Cruzada e um otimizador de *Adam*. O tamanho de grupo para os dados de treino, validação e teste foi o primeiro a ser explorado, em simultâneo com o número de ciclos. Por fim, a métrica de eleição foi o *F1 score*, uma vez que demonstra ser a métrica mais completa, ao refletir tanto a precisão como a sensibilidade.

Na seleção dos modelos a utilizar, como mencionado na metodologia, foram comparados 5 dos modelos disponibilizados pela biblioteca *Pytorch*. Estes modelos são: *DenseNet 121*, *VGG 16*, *VGG 19*, *ResNet 50* e por fim *Inception V3*.

5.5.1 Tamanho de grupo e número de épocas

Foram investigados os efeitos de vários tamanhos de grupo na funcionalidade de vários modelos de *deep learning*. Um parâmetro que afeta o comportamento de convergência, a eficácia da generalização e a eficiência computacional.

Primeiro, foi analisado como o tamanho do grupo afeta o tempo de treino. Como há menos iterações a cada época, grupos maiores podem acelerar o treino, mas também ocupam mais memória, o que pode necessitar de um elevado poder computacional. Embora o treino possa demorar mais tempo com tamanhos de grupo menores, o comportamento de generalização e convergência pode ser melhor.

Em segundo lugar, foram avaliados fatores como a taxa de convergência e as flutuações à medida que era analisado o comportamento de convergência de modelos com vários tamanhos

de grupo.

Nas Figuras 5.11, 5.12 e 5.13, é representado o desempenho dos diferentes modelos para vários tamanhos de grupo.

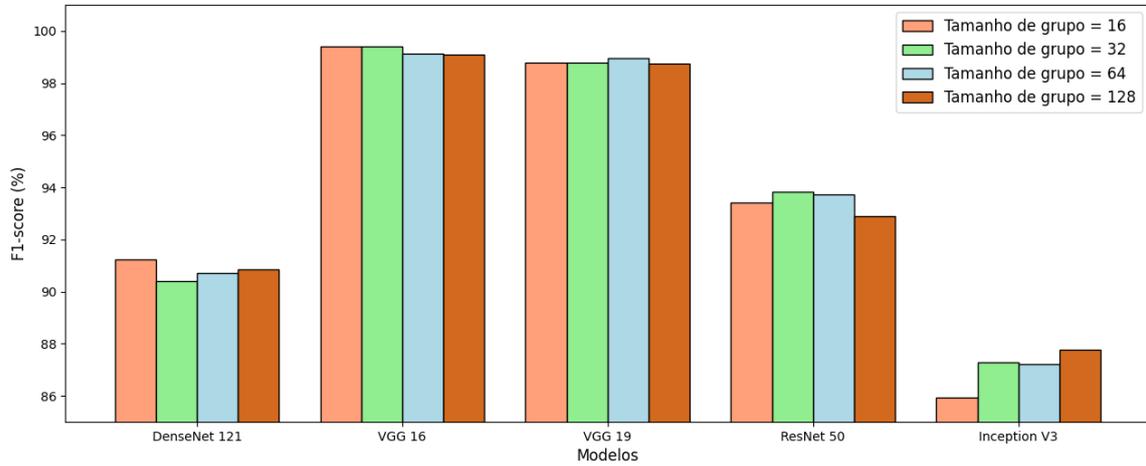


Figura 5.11: Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados A.

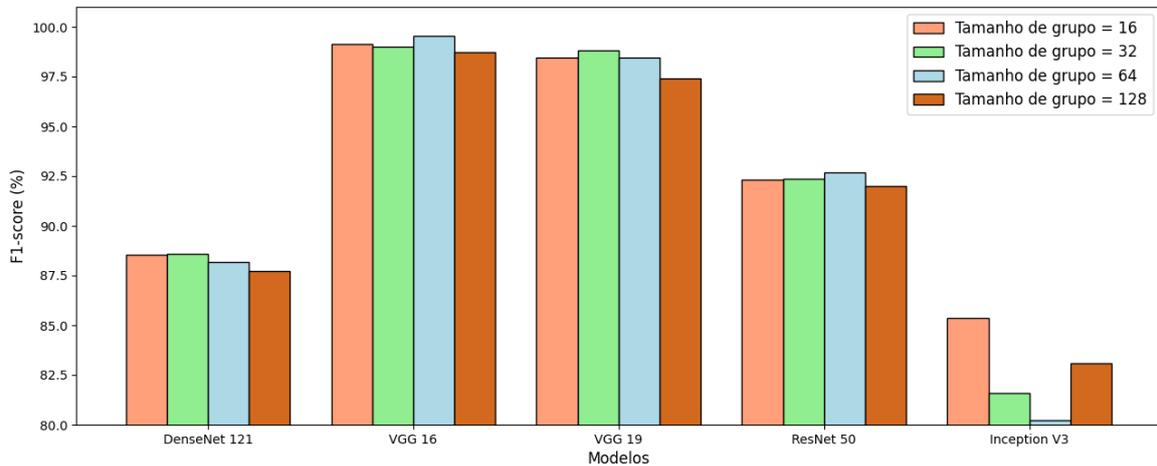


Figura 5.12: Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados B.

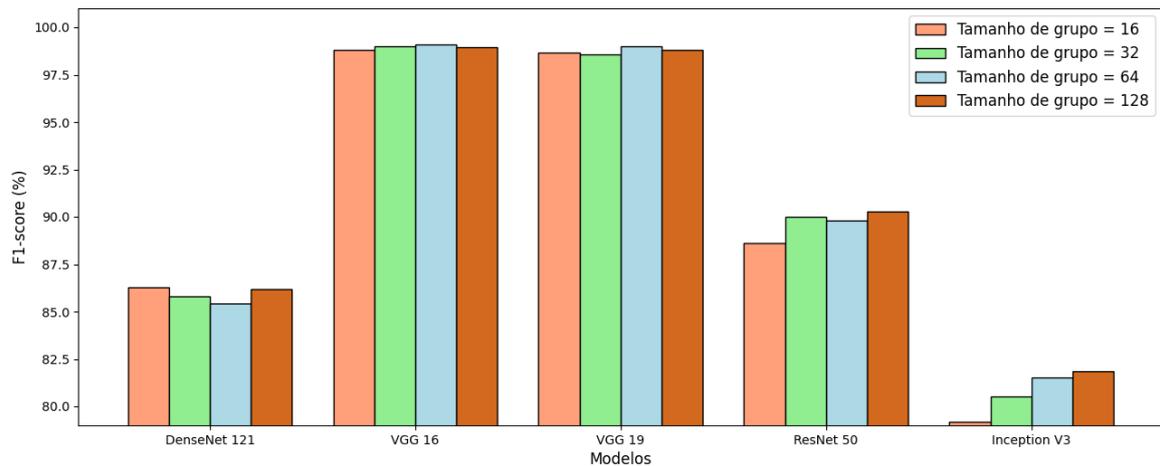


Figura 5.13: Desempenho dos diferentes modelos eleitos para diferentes tamanhos de grupo de treino para o conjunto de dados C.

Analisando os traçados, facilmente se conclui de que no conjunto de dados utilizado, o tamanho de grupo não provoca grande alteração nos desempenhos dos diferentes modelos. Apesar disso, o tempo de treino de tamanhos de grupo maiores foi bastante inferior relativamente aos restantes. Por outro lado, os tamanhos de grupo menores possuem uma convergência do desempenho para o seu valor mais elevado em menos épocas.

De forma a conjugar os dois aspetos, rapidez de convergência e de treino, foi eleito o tamanho de grupo igual a 32 quando utilizados os conjuntos de dados A e B, e um tamanho de grupo 64 quando utilizado o conjunto de dados C. Nas Figuras 5.14, 5.15 e 5.16 estão representados os resultados do treino dos diferentes modelos para estes tamanhos de grupo. Os valores de desempenho obtidos foram calculados com um grupo de validação ao fim de cada época.

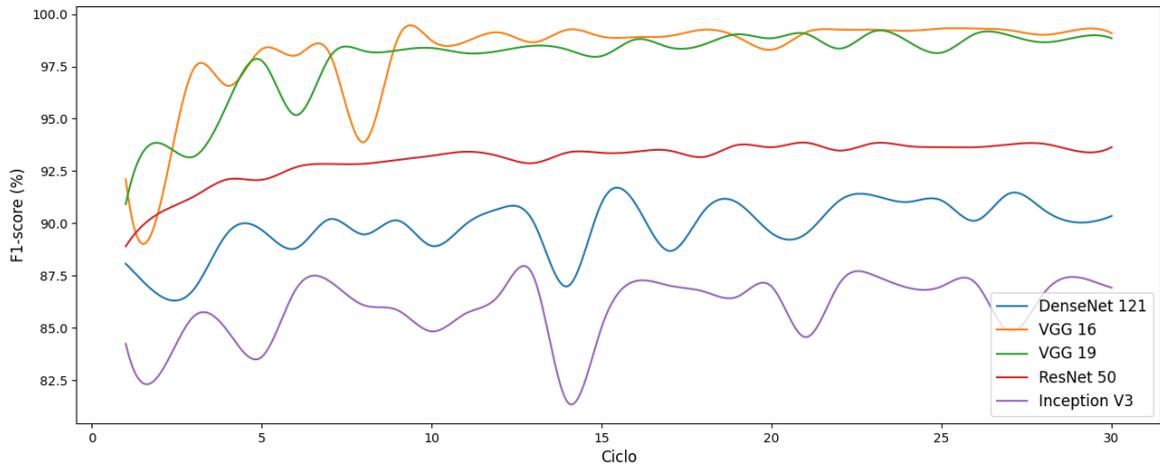


Figura 5.14: Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 32 para o conjunto de dados A.

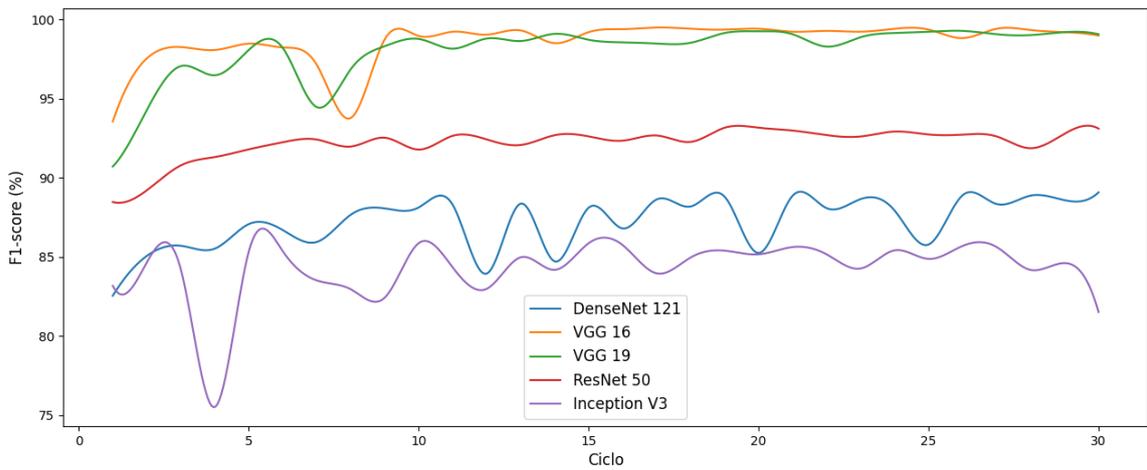


Figura 5.15: Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 32 para o conjunto de dados B.

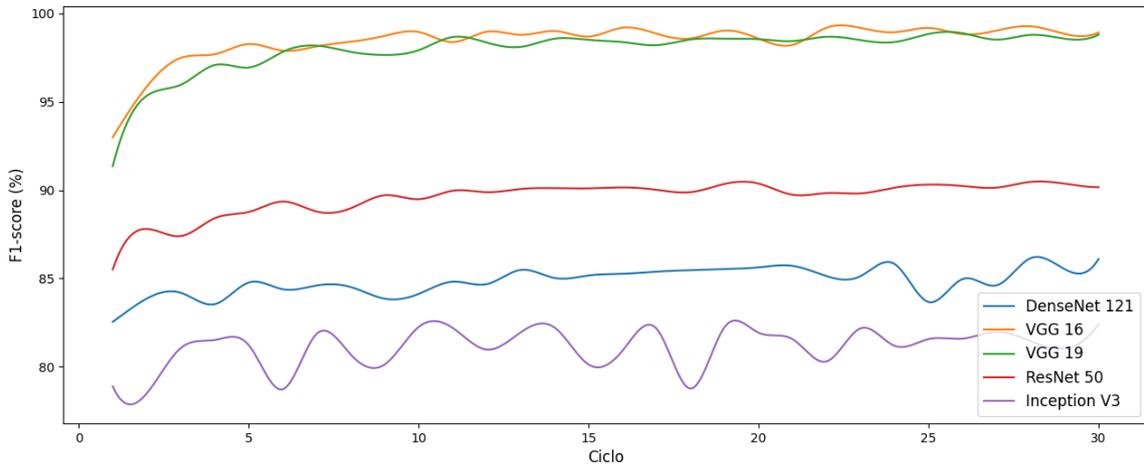


Figura 5.16: Progressão do treino dos diferentes modelos para um tamanho de grupo igual a 64 para o conjunto de dados C.

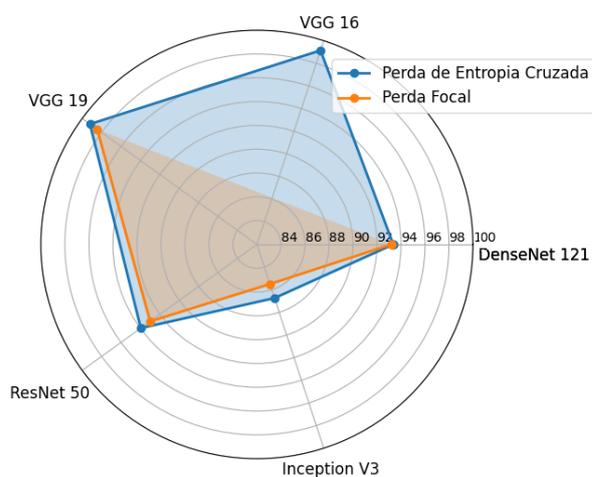
Analisando os gráficos, percebe-se que os modelos atingem a convergência em poucas épocas de treino. Este efeito, deve-se possivelmente ao conjunto de dados utilizado. As imagens podem não possuir padrões muito complexos, o que ao utilizar um modelo com um elevado grau de complexidade não encontra grande dificuldade em caracterizar as classes.

Por fim, de modo a aumentar a eficiência computacional, uma vez que não provoca uma diminuição no desempenho dos classificadores, foi eleito um número de épocas igual a 20 para os posteriores testes.

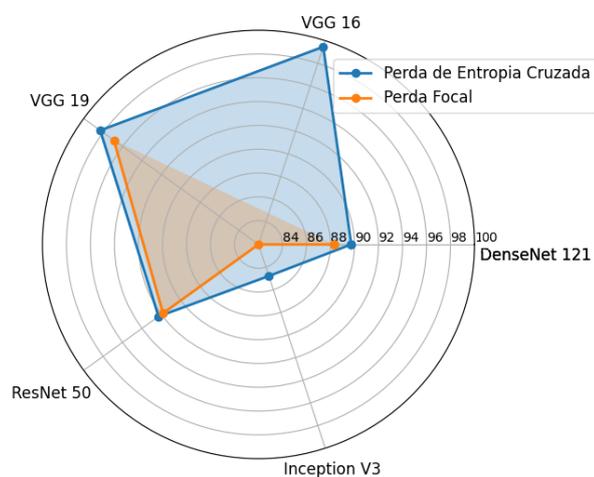
5.5.2 Função de perda (perda de entropia cruzada *vs* perda focal)

As funções de perda são elementos cruciais no treino de redes neurais, ao medirem a diferença entre as saídas esperadas e os rótulos reais, auxiliando o modelo a atingir um melhor desempenho.

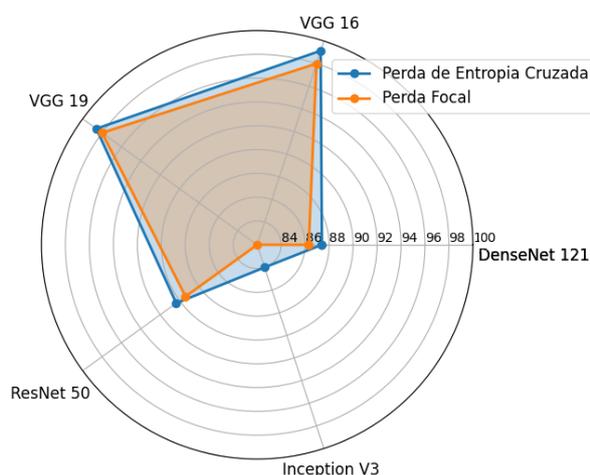
Foram examinados os efeitos da perda de entropia cruzada e a perda focal na generalização do modelo, na convergência do treino e na resistência a dados ruidosos ou distorcidos. Com esta comparação será possível eleger a função de perda mais adequada para as bases de dados utilizadas, desenvolvendo um conhecimento profundo das compensações e vantagens associadas a cada função de perda. Isto resultará, em última análise, num melhor desempenho do modelo e em previsões mais exatas.



(a) Desempenho para o conjunto de dados A.



(b) Desempenho para o conjunto de dados B.



(c) Desempenho para o conjunto de dados C.

Figura 5.17: Desempenho dos cinco modelos eleitos utilizando diferentes funções de perda.

Analisando os gráficos da Figura 5.17, verifica-se que em todos os modelos e todos os conjuntos de dados a função de perda que obteve desempenhos superiores foi a perda de entropia cruzada. Esta é uma função comum, eleita pelo seu elevado desempenho. Uma vez que a utilização da função de perda focal é mais adequada quando os dados de entrada não estão balanceados, um desempenho superior na utilização da função de perda de entropia cruzada é expectável, visto que os dados de entrada são equilibrados entre as duas classes.

É possível ainda verificar uma descida abrupta no desempenho do modelo *VGG 16* quando utilizada a perda focal nos conjuntos de dados A e B, este facto é justificado com a falta de diversidade de características, uma vez que no conjunto de dados C já obteve um desempenho

considerável. Isto porque a perda focal consegue atingir uma eficácia superior quando encontra exemplos difíceis de classificar.

Apesar da superioridade, em termos de desempenho, demonstrada pela perda de entropia cruzada, este não foi o motivo mais relevante para a sua seleção, mas sim a rapidez de convergência dos modelos quando aplicada esta função de perda. Assim, nas próximas comparações será utilizada a função de perda de entropia cruzada.

5.5.3 Algoritmo de otimização (Adam vs Descida de Gradiente Estocástico)

As funções de otimização desempenham um papel significativo na modelação do processo de formação da rede neuronal, influenciando o modo como os pesos do modelo são ajustados para minimizar a função de perda.

Um melhor método de treino é identificado analisando as funções de otimização, permitindo alinhar o processo de formação com a procura de uma convergência global ótima. Consequentemente, permite identificar o algoritmo que ajuda a produzir modelos mais confiáveis, precisos e generalizados.

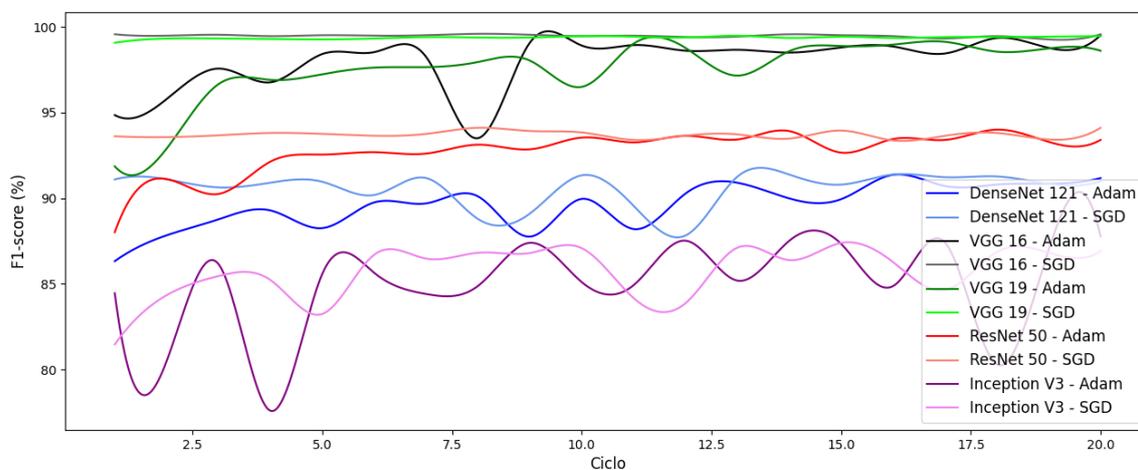


Figura 5.18: Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados A.

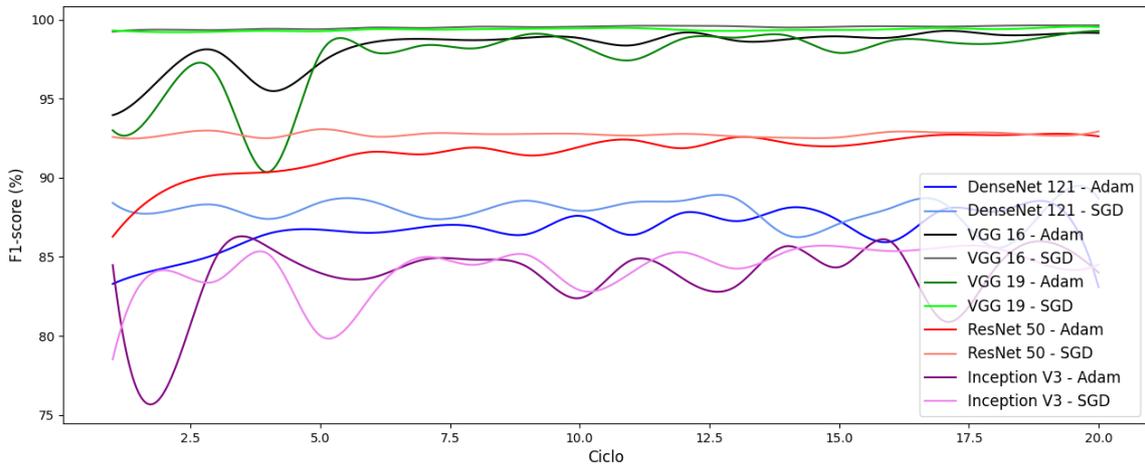


Figura 5.19: Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados B.

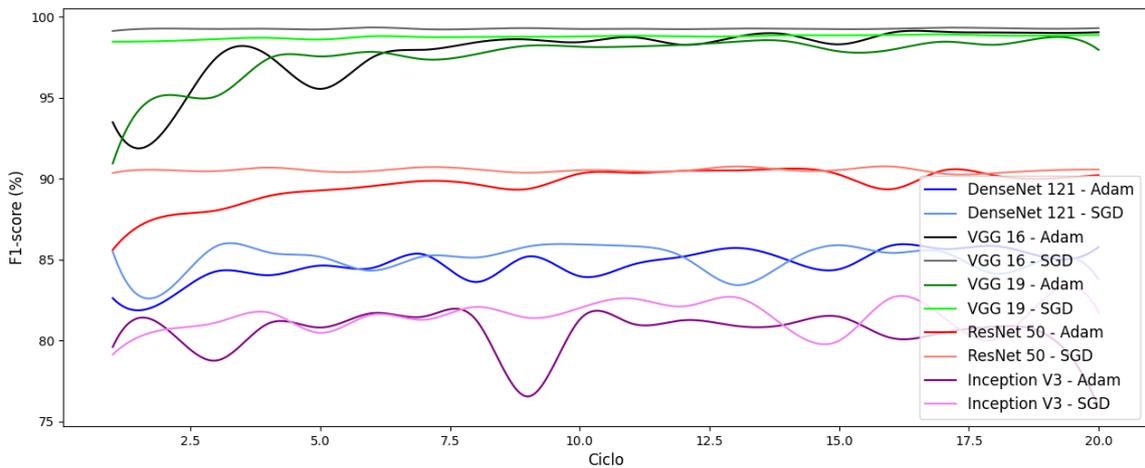


Figura 5.20: Progressão do treino dos diferentes modelos com diferentes algoritmos de otimização para o conjunto de dados C.

Através da análise das Figuras 5.18, 5.19 e 5.20, é possível compreender a capacidade de atingir níveis de desempenho superiores num período de tempo mais curto do método de otimização SGD, apresentando um desempenho superior em todos os conjuntos de dados. Apesar do desempenho final obtido pelas duas técnicas de otimização ser muito comparável, a abordagem SGD demonstrou superioridade, ao atingir um nível de desempenho superior muito mais cedo no período de treino. Esta obtenção precoce de um desempenho elevado pode ser

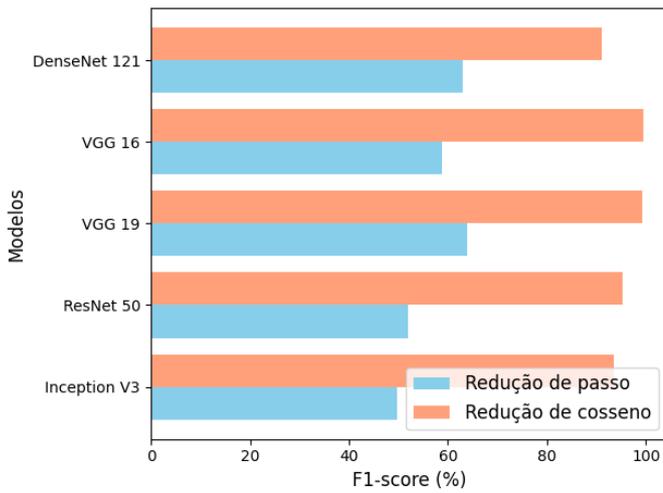
explicada devido à capacidade inerente do SGD de modificar rapidamente os parâmetros do modelo com base em secções mais pequenas dos dados de treino.

A abordagem de otimização SGD também mostrou uma vantagem em termos de velocidade de convergência. Ao usar o SGD, a convergência do processo de treino foi mais rápida. Com base nestas observações, este método foi eleito para todos os modelos em todos os conjuntos de dados.

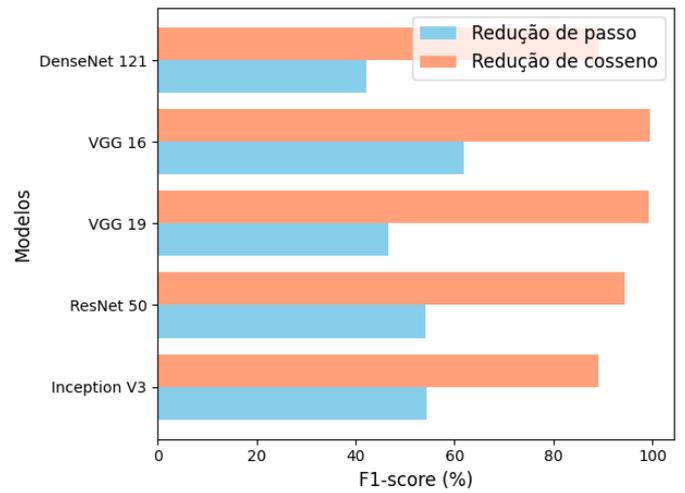
5.5.4 Programador de taxa de aprendizagem (redução de passo *vs* redução de cosseno)

A análise dos programadores de taxa de aprendizagem, uma extensão lógica do estudo das funções de otimização relacionadas com perdas e ganhos, é abordada nesta secção. A velocidade a que os pesos do modelo são ajustados durante o treino é determinada pelos programadores, de modo a maximizar a convergência e evitar oscilações indesejáveis.

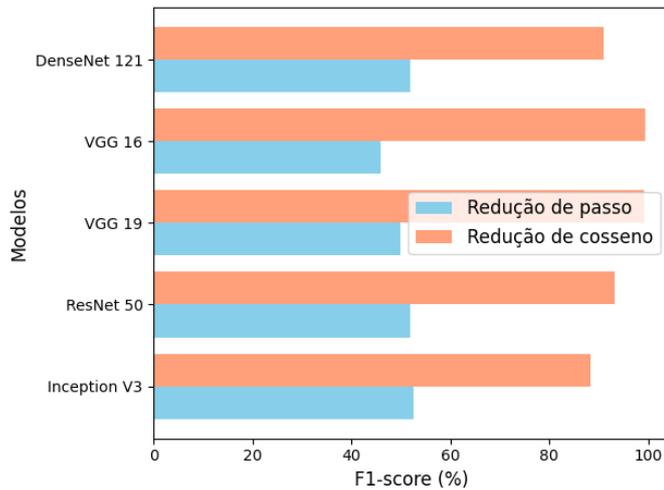
É possível escolher a abordagem mais adequada ao contexto da situação em causa depois de se compreenderem as vantagens e desvantagens de cada método. Uma previsão mais precisa da convergência do modelo, levando a tempos de treino mais eficazes e modelos finais com melhor desempenho e generalização.



(a) Desempenho para o conjunto de dados A.



(b) Desempenho para o conjunto de dados B.



(c) Desempenho para o conjunto de dados C.

Figura 5.21: Desempenho dos cinco modelos eleitos utilizando diferentes programadores de taxa de aprendizagem.

Todos os modelos em avaliação tiveram um desempenho inferior após a aplicação do programador de redução de passo como evidenciado nos gráficos da Figura 5.21. Embora os modelos tenham apresentado diferentes níveis de perda de desempenho, a tendência geral mostrou uma queda constante em toda a linha. Este resultado pode estar relacionado com a rigidez e a predefinição dos ajustes baseados em etapas, que podem não estar em perfeita sintonia com os padrões de aprendizagem das redes neurais.

Por outro lado, a aplicação do programador de redução de cosseno melhorou desempenho de alguns modelos. O programador da taxa de aprendizagem mostrou valores das métricas de desempenho aprimoradas para os modelos *ResNet 50* e *Inception V3*. O desempenho dos

restantes modelos, no entanto, não foi afetado por este programador.

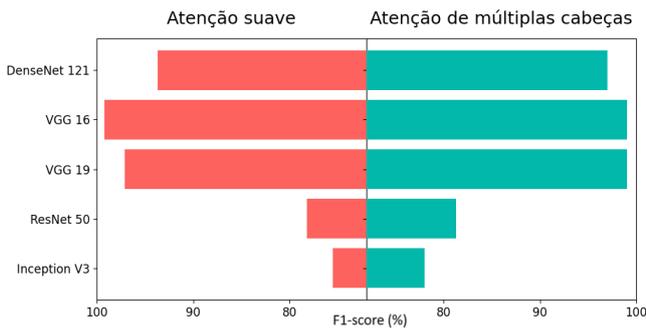
Esses resultados levaram à conclusão de que o programador de taxa de aprendizagem de redução de cosseno é uma opção superior para melhorar os modelos. É um método flexível que se adapta aos padrões de aprendizagem exibidos pelas redes neurais, ao proporcionar aumentos de desempenho para determinados modelos, mantendo o desempenho de outros.

5.5.5 Mecanismo de atenção (atenção suave *vs* atenção de múltiplas cabeças)

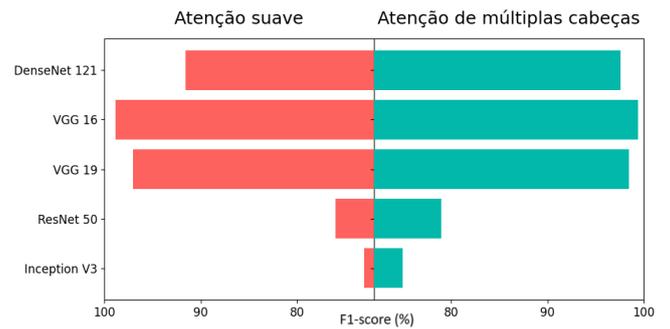
Nesta secção são analisados os mecanismos de atenção, um desenvolvimento crucial no domínio das redes neuronais. Ao escolherem concentrar-se em informação pertinente nos dados de entrada, os mecanismos de atenção desempenham um papel importante na melhoria da capacidade de aprendizagem dos modelos.

Apesar do elevado poder computacional fornecido pela plataforma *Google Colab*, o tempo necessário para comparar diversos mecanismos de atenção seria maior do que o disponível. Assim, elegeram-se aqueles que se consideraram mais pertinentes para a classificação de imagens: atenção suave e atenção de múltiplas cabeças.

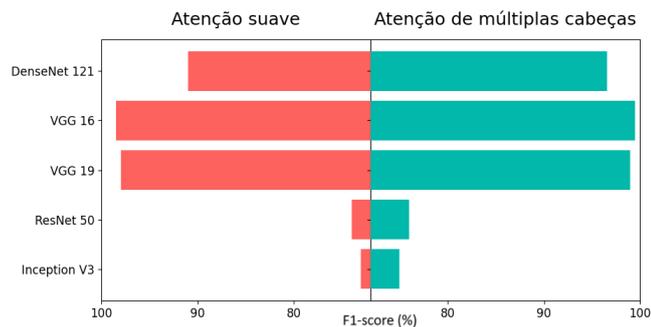
A atenção suave e a atenção de múltiplas cabeças destacam-se como sendo especialmente adequadas para tarefas de categorização de imagens, entre os muitos processos de atenção. A atenção suave permite que o modelo se concentre dinamicamente em áreas específicas de uma imagem, permitindo-lhe identificar características e detalhes importantes. A atenção de múltiplas cabeças aumenta a capacidade do modelo para reconhecer padrões complicados ao ter em conta várias partes da imagem ao mesmo tempo, melhorando os resultados da classificação. Como objetivo, pretende-se avaliar as vantagens e desvantagens relativas aos mecanismos de atenção via uma análise comparativa.



(a) Desempenho para o conjunto de dados A.



(b) Desempenho para o conjunto de dados B.



(c) Desempenho para o conjunto de dados C.

Figura 5.22: Desempenho dos cinco modelos eleitos utilizando diferentes mecanismos de atenção.

Analisando os gráficos da Figura 5.22, percebe-se que o desempenho do modelo *DenseNet 121* foi visivelmente melhorado após a integração do mecanismo de atenção de múltiplas cabeças. Esta melhoria mostra que o mecanismo de atenção reuniu e sintetizou com sucesso informações pertinentes em muitas cabeças de atenção, apoiando as capacidades melhoradas de convergência e generalização do modelo. O desempenho do modelo, por outro lado, foi comparável à linha de base quando o mecanismo de atenção suave foi implementado.

O mecanismo atenção de múltiplas cabeças mostra um ligeiro ganho no desempenho quando usado com os modelos *VGG 16* e *VGG 19*. Embora a melhoria não tenha sido grande, mostrou que o mecanismo conseguiu influenciar estes modelos com alguns conhecimentos úteis. No geral, com e sem o mecanismo, o desempenho destes modelos manteve-se essencialmente consistente.

Curiosamente, a adição do mecanismo de atenção aos modelos *ResNet 50* e *Inception V3* resultou num desempenho inferior. Este resultado levanta a possibilidade de a adição de mecanismos de atenção ter criado ruído ou interferência que dificultou os processos de aprendizagem dos modelos. Os modelos que não tinham um mecanismo de atenção tiveram, portanto, um melhor desempenho.

Estes resultados levaram à adoção de uma estratégia seletiva para a integração de mecanismos

de atenção. Assim, para os modelos *DenseNet 121*, *VGG 16* e *VGG 19*, em que se verificaram vantagens definitivas ou ligeiras melhorias, foi adotado o mecanismo de atenção de múltiplas cabeças. Optou-se por excluir completamente o mecanismo de atenção para os modelos *ResNet 50* e *Inception V3*, devido ao seu impacto negativo no desempenho dos modelos. O objetivo deste procedimento de seleção metódico foi maximizar as vantagens dos mecanismos de atenção, minimizando os potenciais inconvenientes para determinadas arquiteturas de modelos.

5.5.6 Avaliação final

A seguir são apresentados os desempenhos de cada modelo para cada um dos conjuntos de dados. Nesta avaliação foi utilizada a configuração eleita através das comparações realizadas anteriormente, de forma a obter os melhores desempenhos possíveis. A avaliação é apoiada por várias métricas, para que seja possível uma análise completa dos modelos, uma vez que, cada uma das quais fornece informações únicas sobre a categorização de imagens médicas.

Tabela 5.8: Desempenho em percentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados A.

Modelos	Exatidão	Precisão	<i>F1 score</i>	Especificidade	Sensibilidade	AUC
<i>DenseNet 121</i>	96,95	98,88	96,94	98,83	95,08	96,98
<i>VGG 16</i>	99,01	99,09	99,02	99,06	98,96	99,01
<i>VGG 19</i>	99,32	99,27	99,33	99,24	99,40	99,32
<i>ResNet 50</i>	95,28	97,31	95,36	97,22	93,49	95,35
<i>Inception V3</i>	93,09	91,26	93,50	90,10	95,85	92,98

Entre os desempenhos representados na Tabela 5.8, o desempenho do *VGG 19* destaca-se com o maior *F1 score* (99,33%) e precisão (99,32%). Tanto os exemplos positivos como os negativos (casos com patologia e sem patologia) podem ser classificados com este modelo. Com um *F1 score* de 96,94% e uma precisão de 96,95%, o *DenseNet 121* vem em segundo lugar. Apresenta um desempenho equilibrado em termos de sensibilidade e precisão.

Em comparação com outros modelos, o *DenseNet 121* e o *ResNet 50* têm valores de sensibilidade mais baixos, o que sugere que possuem dificuldade em detetar casos positivos. Todos os modelos apresentam uma boa capacidade de distinguir imagens com sarcopenia e sem sarcopenia graças aos seus valores AUC comparativamente elevados.

Tabela 5.9: Desempenho em porcentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados B.

Modelos	Exatidão	Precisão	<i>F1 score</i>	Especificidade	Sensibilidade	AUC
<i>DenseNet 121</i>	97,47	97,83	97,63	97,51	97,43	97,47
<i>VGG 16</i>	98,41	97,17	98,55	96,60	99,96	98,28
<i>VGG 19</i>	99,06	99,28	99,12	99,16	98,96	99,06
<i>ResNet 50</i>	93,91	93,46	94,32	92,43	95,20	93,82
<i>Inception V3</i>	88,15	87,01	89,13	84,51	91,36	87,94

O *VGG 19* continua a ter um bom desempenho, com o *F1 score* mais elevado (99,12%) e a exatidão (99,06%), como evidenciado na Tabela 5.9. Além disso, destaca-se em termos de sensibilidade e precisão. Mais uma vez, o *VGG 16* e o *DenseNet 121* têm um bom desempenho, com excelentes valores de *F1 score* e exatidão.

Em comparação com outros modelos, o *Inception V3* é menos exato e possui o *F1 score* mais baixo. Os valores AUC são consistentemente elevados em todos os modelos, o que demonstra a sua adequação à classificação dos dados.

Tabela 5.10: Desempenho em porcentagem de cada modelo eleito com a sua melhor configuração para o conjunto de dados C.

Modelos	Exatidão	Precisão	<i>F1 score</i>	Especificidade	Sensibilidade	AUC
<i>DenseNet 121</i>	96,44	96,08	97,63	95,61	97,19	96,40
<i>VGG 16</i>	99,08	98,83	99,12	98,70	99,42	99,06
<i>VGG 19</i>	98,64	98,36	98,71	98,17	99,07	98,62
<i>ResNet 50</i>	92,72	91,19	93,22	89,81	95,35	92,58
<i>Inception V3</i>	87,31	85,26	88,35	82,48	91,68	87,08

Na Tabela 5.10, com uma precisão de 98,83% e um *F1 score* de 99,12%, o *VGG 16* apresenta-se como o modelo com melhor desempenho. Tem um desempenho elevado em termos de precisão e sensibilidade, o que o torna adequado para uma classificação precisa. Também os modelos *DenseNet 121* e *VGG 19* possuem um bom desempenho, com um *F1 score* competitivo e níveis de precisão que ultrapassam os 97%. Em termos de precisão e *F1 score*, o *Inception V3* tem um desempenho inferior aos restantes modelos. À semelhança dos outros conjuntos de dados, os valores AUC continuam a ser elevados, realçando a capacidade dos modelos para distinguir as classes com exatidão.

Em conclusão, os modelos *VGG 19* e *VGG 16* apresentam regularmente um desempenho superior em todos os conjuntos de dados (A, B e C), obtendo frequentemente as melhores classificações de *F1 score*. Outro modelo com elevada eficácia, o *DenseNet 121*, supera consistentemente a concorrência. Os valores elevados de AUC dos modelos mostram que têm uma grande capacidade de discriminar entre doentes com sarcopenia e instâncias sem sarcopenia.

É possível, agora, realizar um comparação de resultados com diferentes estudos no que diz respeito à categorização de imagens a nível dos músculos utilizando as mesmas modalidades de imagem e até os mesmos modelos.

Por exemplo, num estudo onde foram utilizadas as arquiteturas *DenseNet 121*, *VGG 19* e *ResNet 50* registaram-se níveis de precisão que variam entre 65% e 75% no contexto de imagens de elastografia [56]. Por outro lado, neste projeto, conseguiu-se produzir resultados de precisão que variam entre 93% e 99%. Esta discrepância demonstra como a complexidade dos conjuntos de dados pode produzir variações substanciais no desempenho dos modelos. Foram obtidos resultados semelhantes quando a mesma patologia foi classificada utilizando imagens de ecografia. Os resultados apresentados revelaram um aumento considerável dos níveis de precisão, para valores entre 97% e 99%, quando estudos anteriores apenas conseguiram atingir taxas de precisão na ordem dos 65% a 75%. Isto sublinha mais uma vez como as propriedades do conjunto de dados podem afetar significativamente a eficácia dos modelos, mesmo quando são utilizadas as mesmas arquiteturas.

A utilização de imagens de ecografia e do modelo *DenseNet* num estudo diferente sobre a sarcopenia [57] produziu um *F1 score* de 68,40%. No presente trabalho, foi obtido um *F1 score* significativamente mais elevada, de 96,94%, para a mesma patologia, demonstrando mais uma vez como a seleção do conjunto de dados, o modelo e os ajustamentos podem produzir resultados consideravelmente diferentes.

Por último, mas não menos importante, o modelo *VGG 16* demonstrou uma precisão de 96,70% no domínio da classificação de patologias por ecografia numa outra investigação [63]. Por outro lado, nesta dissertação, a mesma arquitetura produziu uma taxa de precisão de aproximadamente 99% quando utilizada para classificar a sarcopenia. Este é um exemplo de como a alternância entre várias patologias pode resultar em alterações significativas de desempenho, mesmo dentro do mesmo modelo.

Embora a utilização das mesmas arquiteturas e modalidades de imagiologia crie um ponto de partida semelhante, é evidente que a especificidade de cada conjunto de dados, a complexidade da fisiopatologia específica e a calibração cuidadosa dos modelos podem resultar em resultados

significativamente diferentes. Isto realça a importância de métodos de investigação abrangentes que tenham em conta as propriedades técnicas e subtis dos dados que estão a ser utilizados.

5.6 Comparação entre *machine learning* clássico e *deep learning*

A Tabela 5.11 mostra o desempenho de vários modelos de *machine learning* clássico (ML) e *deep learning* (DL) para os conjuntos de dados distintos (A, B e C), com cada modelo a utilizar as suas configurações ótimas. O valor de *F1 score*, foi a estatística de eleição para a avaliação dos modelos.

Tabela 5.11: Desempenho de cada um dos modelos eleitos de *machine learning* clássico (ML) e *deep learning* (DL), com a sua melhor configuração para todos os conjunto de dados.

		<i>F1 score</i> (%)		
		A	B	C
	Modelos			
	Rede Neuronal	99,81	99,59	99,20
	<i>Nu Support Vector</i>	94,03	93,32	92,84
ML	<i>Logistic Regression</i>	96,84	96,16	92,90
	<i>Stochastic Gradient Descent</i>	97,85	96,79	95,01
	<i>Support Vector Machine</i>	99,04	98,34	96,81
	<i>DenseNet 121</i>	96,94	97,63	97,63
	<i>VGG 16</i>	99,02	98,55	99,12
DL	<i>VGG 19</i>	99,33	99,12	98,71
	<i>ResNet 50</i>	95,36	94,32	93,22
	<i>Inception V3</i>	93,50	89,13	88,35

A escolha do modelo e da combinação de conjuntos de dados afeta consideravelmente a eficácia da classificação de patologias, como se conclui da análise dos resultados apresentados na Tabela 5.11. Determinadas combinações de modelos e conjuntos de dados revelam-se frequentemente como opções mais fiáveis e robustas, oferecendo uma avaliação equilibrada da precisão e da sensibilidade.

Com um *F1 score* de 99,81% para o conjunto de dados A, 99,59% para o conjunto de dados B e 99,20% para o conjunto de dados C, a Rede Neural supera consistentemente os outros modelos de ML em todos os conjuntos de dados. Com um *F1 score* de 99,04% para o conjunto de dados A, 98,34% para o conjunto de dados B e 96,81% para o conjunto de dados C, o SVM também

tem um bom desempenho. A redução verificada nos valores de *F1 score* para os outros modelos de ML nos conjuntos de dados, é justificada por um declínio na capacidade de generalização do modelo.

Em *deep learning* (DL) obtiveram-se valores de *F1 score* consistentemente elevados, para o modelo *VGG 19*, em todos os conjuntos de dados (99,33% para o conjunto de dados A: 99,12% para o conjunto de dados B e 98,71% para o conjunto de dados C), destacando-se entre os restantes modelos. Por outro lado, o *VGG 16* apresenta também um desempenho elevado. O *DenseNet 121* apresentou um desempenho em termos de *F1 score* de 96,94% para o conjunto de dados A, 97,63% para o conjunto de dados B e 97,63% para o conjunto de dados C. Embora o *ResNet 50* apresente um declínio progressivo nos valores de *F1 score* em todos os conjuntos de dados, mantém, no entanto, um *F1 score* elevado, indicando alguma sensibilidade às diferenças dos conjuntos de dados. Com a diminuição dos valores de *F1 score* em todos os conjuntos de dados, a *Inception V3* apresenta um padrão semelhante que pode indicar a sua suscetibilidade a variações nas propriedades dos dados.

Em suma, com desempenhos consistentemente elevados em muitos conjuntos de dados, destacam-se a Rede Neuronal, o *VGG 19* e o *VGG 16*. O *DenseNet 121* e o SVM também apresentam um bom desempenho.

Mesmo que outros modelos e combinações tenham um bom desempenho, a rede neural na categoria ML e o *VGG 19* na categoria DL produzem consistentemente os melhores resultados. Na comparação da utilização dos diferentes conjuntos de dados, apesar muito semelhantes, os desempenhos de todos os modelos, a utilização de, apenas, imagens de ecografia (conjunto de dados A) apresenta-se como a opção que produz um desempenho mais elevado. O modelo de Rede Neuronal e o modelo *VGG 19* acabam por se destacar como as melhores opções para uma classificação fiável da patologia quando combinados com os respetivos conjuntos de dados. A escolha entre os dois é realizada com base numa análise aprofundada dos requisitos do projeto e da viabilidade computacional considerada ideal para o contexto clínico ou de investigação específico. Sendo que a Rede Neuronal é o modelo com menor dispêndio de recursos computacionais.

6 Conclusões e Trabalhos Futuros

6.1 Conclusões

Este estudo sobre a classificação da sarcopenia utilizando técnicas de imagem médica não invasivas produziu resultados promissores. Incluiu a utilização de modelos de *machine learning* e de *deep learning* e uma análise comparativa dos resultados em vários conjuntos de dados, nomeadamente utilizando imagens de ecografia, elastografia e o conjunto das duas.

As diferenças em desempenho, utilizando diferentes métodos aplicados aos modelos, permite concluir que *machine learning* provou ter um desempenho superior na classificação de sarcopenia, produzindo regularmente resultados elevados. Apresentou um *F1 score* mais baixo de 92,84% pelo *Nu Support Vector* no conjunto de dados C e superior a 94% no conjunto de dados A em todos os classificadores. Em contrapartida, em *deep learning*, o *Inception V3* apresenta um valor de *F1 score* mais baixo, de 88,35% no conjunto de dados C.

O modelo Rede Neural demonstrou a melhor eficiência por o seu desempenho realçar a sua capacidade de classificação, atingindo um desempenho médio de 99% em todas as métricas. Este desempenho foi ainda consistente em todos os conjuntos de dados, demonstrando fiabilidade e precisão na classificação.

Quando vários modelos foram comparados, verificou-se que algumas arquiteturas de *deep learning*, como o *VGG 19*, demonstraram um desempenho elevado (99%), em todos os conjuntos de dados. O elevado desempenho destes modelos verificou-se mesmo antes da otimização dos mesmos, ao contrário de modelos de *machine learning*, em que a maioria obteve um *F1 score* inferior a 90% na primeira otimização e um máximo atingido pelo *Nu Support Vector*, de 94,12%.

Os elevados valores de desempenho verificados nos modelos, atingindo valores de *F1 score* maioritariamente superiores as 95%, provam a utilidade da utilização destes modelos para a deteção de patologias em estados prematuros, redução de diagnósticos incorretos e consistência nos diagnóstico. Essencialmente, este trabalho reforça a ideia de que a *machine learning* e *deep learning* continuam a desempenhar um papel crucial no avanço da categorização de patologias e

na interpretação de imagens médicas. O elevado desempenho da Rede Neuronal é um exemplo motivador de como esta tecnologia tem potencial para ter um impacto significativo no sector dos cuidados de saúde e não só.

A semelhança nos desempenhos dos modelos entre os conjuntos de dados em análise é evidente, apresentando variações de apenas 1% na maioria dos casos. No entanto, é de salientar que a utilização de apenas imagens ecográficas (conjunto de dados A) demonstra ser a alternativa que proporciona o melhor nível de desempenho. Esta escolha parece estabelecer uma ligação mais próxima com as características únicas da situação em causa, resultando potencialmente num maior alinhamento entre as características das imagens e os princípios aprendidos do modelo.

Como resultado, embora exista uma semelhança no desempenho entre os modelos, apresentando uma variação máxima de 5%, à excessão de um desempenho inferior do modelo *Inception V3*, conclui-se que uma seleção cuidadosa do tipo de imagem utilizada aponta para uma estratégia que pode maximizar o desempenho como um todo. Esta observação realça a importância do estudo efetuado, a comparação de duas técnicas de imagiologia. Destaca também a necessidade de considerar cuidadosamente o contexto ao escolher o tipo de dados a utilizar em aplicações de análise médica, provado pelo uso de imagens ecográficas em detrimento de imagens de elastografia no modelo *Inception V3*, onde demonstrou um aumento de 4% na métrica *F1 score*.

Concluindo, esta dissertação apresenta um estudo comparativo completo de vários modelos e conjuntos de dados, na sua exploração da categorização da sarcopenia utilizando abordagens de imagens médicas obtidas por técnicas de imagiologia não invasivas. Este trabalho examina a eficácia de vários modelos de *machine learning* e *deep learning* no contexto da categorização da sarcopenia, comparando-os sistematicamente. A importância deste trabalho é ainda mais realçada pelo destaque na escassez de informação atualmente disponível sobre estas metodologias no contexto da classificação da sarcopenia.

6.2 Trabalho Futuro

À medida que o campo da análise de imagens médicas se desenvolve, abrem-se novas direções para a investigação. A disponibilidade limitada de recursos computacionais exige um estudo orientado de métodos de otimização que garantam um desempenho ótimo sem colocar exigências onerosas ao *hardware*.

O desempenho dos modelos pode ser melhorado mediante uma investigação exaustiva de diversos mecanismos de atenção e da exploração de várias arquiteturas da camada final

completamente conectada dos modelos. Além disso, aumentar o conjunto de dados e talvez incluir dados de ambos os sexos e de uma maior gama de idades, poderia melhorar a generalização dos modelos.

Referências

- [1] I. H. Rosenberg, “Sarcopenia: Origins and clinical relevance,” *The Journal of Nutrition*, vol. 127, pp. 990S–991S, 1997.
- [2] A. J. C.-J. *et al.*, “Sarcopenia: European consensus on definition and diagnosis: Report of the european working group on sarcopenia in older people,” *Age and Ageing*, vol. 39, pp. 412–423, 2010.
- [3] O. Bruyère, C. Beaudart, O. Ethgen, J. Y. Reginster, and M. Locquet, “The health economics burden of sarcopenia: a systematic review,” *Maturitas*, vol. 119, pp. 61–69, 2019.
- [4] L. A. S. *et al.*, “Associations of sarcopenia definitions, and their components, with the incidence of recurrent falling and fractures: The longitudinal aging study amsterdam,” *The Journals of Gerontology: Series A*, vol. 73, pp. 1199–1204, 2018.
- [5] V. C. *et al.*, “Sarcopenia: imaging assessment and clinical application,” *Abdominal Radiology 2021 47:9*, vol. 47, pp. 3205–3216, 2021.
- [6] T. Panch, P. Szolovits, and R. Atun, “Artificial intelligence, machine learning and health systems,” *Journal of Global Health*, vol. 8, 2018.
- [7] R. A. F. *et al.*, “Sarcopenia: an undiagnosed condition in older adults. current consensus definition: prevalence, etiology, and consequences. international working group on sarcopenia,” *Journal of the American Medical Directors Association*, vol. 12, pp. 249–256, 2011.
- [8] A. T. *et al.*, “Sarcopenia,” *Joint Bone Spine*, vol. 86, pp. 309–314, 2019.
- [9] B. Dawson-Hughes and H. A. Bischoff-Ferrari, “Defining sarcopenia,” *Nutritional Influences on Bone Health*, pp. 13–20, 2016.

- [10] “Scielo - brazil - mecanismos de perda muscular da sarcopenia mecanismos de perda muscular da sarcopenia.” <https://www.scielo.br/j/rbr/a/j45dnGRLSgbR7M8dbdsnbvq/>. Accessed: 2023-08-28.
- [11] T. L. *et al.*, “Sarcopenia: etiology, clinical consequences, intervention, and assessment,” *Osteoporosis International : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, vol. 21, pp. 543–559, 2010.
- [12] T. N. Kim and K. M. Choi, “Sarcopenia: Definition, epidemiology, and pathophysiology,” *Journal of Bone Metabolism*, vol. 20, p. 1, 2013.
- [13] M. C. Kostek and M. J. Delmonico, “Age-related changes in adult muscle morphology,” *Current Aging Science*, vol. 4, pp. 221–233, 2011.
- [14] N. Veronese, C. Beaudart, and S. Sabico, *Sarcopenia : research and clinical implications*. 2021.
- [15] A. J. Cruz-Jentoft, “Sarcopenia, the last organ insufficiency,” *European Geriatric Medicine*, vol. 7, pp. 195–196, 2016.
- [16] J. W. *et al.*, “Asians have lower body mass index (bmi) but higher percent body fat than do whites: comparisons of anthropometric measurements,” *The American journal of clinical nutrition*, vol. 60, pp. 23–28, 1994.
- [17] L. K. C. *et al.*, “Sarcopenia in asia: consensus report of the asian working group for sarcopenia,” *Journal of the American Medical Directors Association*, vol. 15, pp. 95–101, 2014.
- [18] S. von Haehling, J. E. Morley, and S. D. Anker, “An overview of sarcopenia: facts and numbers on prevalence and clinical impact,” *Journal of Cachexia, Sarcopenia and Muscle*, vol. 1, pp. 129–133, 2010.
- [19] B. H. G. *et al.*, “The loss of skeletal muscle strength, mass, and quality in older adults: The health, aging and body composition study,” *The Journals of Gerontology: Series A*, vol. 61, pp. 1059–1064, 2006.
- [20] P. Wakimoto and G. Block, “Dietary intake, dietary patterns, and changes with agean epidemiological perspective,” *The Journals of Gerontology: Series A*, vol. 56, pp. 65–80, 2001.

- [21] S. M. R. *et al.*, “Does nutrition play a role in the prevention and management of sarcopenia?,” *Clinical Nutrition*, vol. 37, pp. 1121–1132, 2018.
- [22] M. D. *et al.*, “Motoneuron loss is associated with sarcopenia,” *Journal of the American Medical Directors Association*, vol. 15, pp. 435–439, 2014.
- [23] J. E. Morley, S. D. Anker, and S. von Haehling, “Prevalence, incidence, and clinical impact of sarcopenia: facts, numbers, and epidemiology—update 2014,” *Journal of Cachexia, Sarcopenia and Muscle*, vol. 5, pp. 253–259, 2014.
- [24] S. C. Shaw, E. M. Dennison, and C. Cooper, “Epidemiology of sarcopenia: Determinants throughout the lifecourse,” *Calcified Tissue International 2017 101:3*, vol. 101, pp. 229–247, 2017.
- [25] F. L. *et al.*, “Sarcopenia as a risk factor for falls in elderly individuals: Results from the ilsirente study,” *Clinical Nutrition*, vol. 31, pp. 652–658, 2012.
- [26] G. S. *et al.*, “Prevalence of sarcopenia in the world: A systematic review and meta-analysis of general population studies,” *Journal of Diabetes and Metabolic Disorders*, vol. 16, pp. 1–10, 2017.
- [27] F. L. *et al.*, “Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia,” *Journal of Applied Physiology (Bethesda, Md. : 1985)*, vol. 95, pp. 1851–1860, 2003.
- [28] A. C.-J. *et al.*, “Sarcopenia: revised european consensus on definition and diagnosis,” *Age and Ageing*, vol. 48, pp. 16–31, 2019.
- [29] “Dinamômetro digital de mão saehan - fisioterapia - carci: Tudo para fisioterapia e reabilitação física.” <https://www.carcioficial.com.br/dinamometro-digital-de-mao-saehan/>. Accessed: 2022-11-28.
- [30] “Short physical performance battery (sppb) | national institute on aging.” <https://www.nia.nih.gov/research/labs/leps/short-physical-performance-battery-sppb>. Accessed: 2022-11-30.
- [31] G. Rubbieri, E. Mossello, and M. D. Bari, “Techniques for the diagnosis of sarcopenia,” *Clinical Cases in Mineral and Bone Metabolism*, vol. 11, p. 181, 2014.
- [32] “Cuidado personal: Omron medidor de grasa corporal hbf-306int.” <https://www.dfauruguay.com/es/categorias/electronica/cuidado-personal/omronmedidordegrasacorporalhbf-306int-detail.html>. Accessed: 2022-11-28.

- [33] G. S. *et al.*, “Imaging of sarcopenia,” *European Journal of Radiology*, vol. 85, pp. 1519–1524, 2016.
- [34] “Rcc disponibiliza novo aparelho de tomografia computadorizada aquilionone genesis | blog – radiologia clínica de campinas.” <https://www.radiologiaclinicadecampinas.com.br/site/blog/28/2019/04/rcc-disponibiliza-novo-aparelho-de-tomografia-computadorizada-aquilionone-genesis/>. Accessed: 2022-11-28.
- [35] “Xct 3000 - technical data and product information - novotec medical and stratec medizintechnik.” <https://www.galileo-training.com/ch-deutsch/produkte/p25/xct-3000.html>. Accessed: 2022-11-28.
- [36] “Magnetom sola cardiovascular edition - siemens healthineers.” <https://www.siemens-healthineers.com/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner/magnetom-sola-cardiovascular-edition>. Accessed: 2022-11-28.
- [37] “Densitômetro ósseo dexa - lunar idxa - ge healthcare - de feixe tipo leque.” <https://www.medicalexpo.com/pt/prod/ge-healthcare/product-70717-428895.html>. Accessed: 2022-11-28.
- [38] S. B. *et al.*, “Usability and pitfalls of shear-wave elastography for evaluation of muscle quality and its potential in assessing sarcopenia: A review,” *Ultrasound in Medicine & Biology*, vol. 46, pp. 2891–2907, 2020.
- [39] A. Chauhan, “The emerging role of two-dimensional us shear-wave elastography in chronic liver disease,” *Radiology*, vol. 289, pp. 416–417, 2018.
- [40] S. Yu, K. Umaphysivam, and R. Visvanathan, “Sarcopenia in older people,” *International Journal of Evidence-Based Healthcare*, vol. 12, pp. 227–243, 2014.
- [41] S. C. Y. *et al.*, “Clinical screening tools for sarcopenia and its management,” *Current Gerontology and Geriatrics Research*, vol. 2016, 2016.
- [42] R. C. *et al.*, “Current nutritional recommendations and novel dietary strategies to manage sarcopenia,” *The Journal of Frailty & Aging*, vol. 2, p. 38, 2013.
- [43] C. G. D. Souza, “Pharmacological treatment of sarcopenia,” *Revista Brasileira de Ortopedia*, vol. 56, p. 425, 2021.

- [44] M. R. *et al.*, “Use of artificial intelligence in the imaging of sarcopenia: A narrative review of current status and perspectives,” *Nutrition*, vol. 89, p. 111227, 2021.
- [45] R. Li and M. E. Naqa, eds., *Machine Learning in Radiation Oncology*. Springer International Publishing, 2015.
- [46] S. Sun, W. Xue, and Y. Zhou, “Classification of young healthy individuals with different exercise levels based on multiple musculoskeletal ultrasound images,” *Biomedical Signal Processing and Control*, vol. 62, p. 102093, 2020.
- [47] X. Yu, X. Xu, Q. Huang, G. Zhu, F. Xu, Z. Liu, L. Su, H. Zheng, C. Zhou, Q. Chen, F. Gao, M. Lin, S. Yang, M. H. Chiang, and Y. Zhou, “Binary classification of non-specific low back pain condition based on the combination of b-mode ultrasound and shear wave elastography at multiple sites,” *Frontiers in Physiology*, vol. 14, p. 1176299, 2023.
- [48] F. P. *et al.*, “Specification of neck muscle dysfunction through digital image analysis using machine learning,” *Diagnostics*, vol. 13, p. 7, 2023.
- [49] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, pp. 685–695, 2021.
- [50] N. Rusk, “Deep learning,” *Nature Methods 2016 13:1*, vol. 13, pp. 35–35, 2015.
- [51] G. C. *et al.*, “Machine learning and deep learning in medical imaging: Intelligent imaging,” *Journal of Medical Imaging and Radiation Sciences*, vol. 50, pp. 477–487, 2019.
- [52] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, pp. 943–947, 2015.
- [53] R. H. *aet al.*, “Efficient hardware architecture of softmax layer in deep neural network,” *International Conference on Digital Signal Processing, DSP*, vol. 2018, 2019.
- [54] S. Kiliçarslan, K. Adem, and M. Çelik, “An overview of the activation functions used in deep learning algorithms,” *Journal of New Results in Science*, vol. 10, pp. 75–88, 2021.
- [55] S. Park and N. Kwak, “Analysis on the dropout effect in convolutional neural networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10112 LNCS, pp. 189–204, 2017.

- [56] J. Y. *et al.*, “Deep learning based sarcopenia prediction from shear-wave ultrasonographic elastography and gray scale ultrasonography of rectus femoris muscle,” *Scientific Reports*, vol. 12, p. 3596, 2022.
- [57] E. P. *et al.*, “An autoencoder convolutional neural network framework for sarcopenia detection based on multi-frame ultrasound image slices,” *IFIP Advances in Information and Communication Technology*, vol. 627, pp. 209–219, 2021.
- [58] S. K. *et al.*, “Muscle cross-sectional area segmentation in transverse ultrasound images using vision transformers,” *Diagnostics 2023*, vol. 13, p. 217, 2023.
- [59] E. J. Snider, S. I. Hernandez-Torres, and E. N. Boice, “An image classification deep-learning algorithm for shrapnel detection from ultrasound images,” *Scientific Reports 2022 12:1*, vol. 12, pp. 1–12, 2022.
- [60] E. N. Boice, S. I. Hernandez-Torres, and E. J. Snider, “Comparison of ultrasound image classifier deep learning algorithms for shrapnel detection,” *Journal of Imaging 2022*, vol. 8, p. 140, 2022.
- [61] P. R. *et al.*, “Deepacs: Automatic segmentation of anatomical cross-sectional area in ultrasound images of human lower limb muscles using deep learning,” *MedRxiv*, p. 2021.12.27.21268258, 2021.
- [62] E. Uçar, “Classification of myositis from muscle ultrasound images using deep learning,” *Biomedical Signal Processing and Control*, vol. 71, 2022.
- [63] J. X. *et al.*, “Automatic classification of male and female skeletal muscles using ultrasound imaging,” *Biomedical Signal Processing and Control*, vol. 57, p. 101731, 2020.
- [64] T. H. *et al.*, “Conditional sliding windows: An approach for handling data limitation in colorectal histopathology image classification,” *Informatics in Medicine Unlocked*, vol. 23, p. 100565, 2021.
- [65] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, pp. 91–99, 2022.
- [66] “sklearn.preprocessing.standardscaler — scikit-learn 1.3.0 documentation.” <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. Accessed: 2023-07-15.

- [67] “sklearn.feature_selection.f_classif — scikit-learn 1.3.0 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html. Accessed: 2023-07-15.
- [68] L. Sthle and S. Wold, “Analysis of variance (anova),” *Chemometrics and Intelligent Laboratory Systems*, vol. 6, pp. 259–272, 1989.
- [69] “sklearn.feature_selection.mutual_info_classif — scikit-learn 1.3.0 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html. Accessed: 2023-07-15.
- [70] I. Maglogiannis, *Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. Frontiers in artificial intelligence and applications, IOS Press, 2007.
- [71] H. R. *et al.*, “Multilayer perceptron: Architecture optimization and training,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, p. 26, 2016.
- [72] “1.4. support vector machines — scikit-learn 1.3.0 documentation.” <https://scikit-learn.org/stable/modules/svm.html#svm-mathematical-formulation>. Accessed: 2023-07-15.
- [73] “1.1. linear models — scikit-learn 1.3.0 documentation.” https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression. Accessed: 2023-07-15.
- [74] “1.5. stochastic gradient descent — scikit-learn 1.3.0 documentation.” <https://scikit-learn.org/stable/modules/sgd.html#sgd-mathematical-formulation>. Accessed: 2023-07-15.
- [75] R. T. *et al.*, “Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020,” pp. 3–26, 2021.
- [76] I. Kandel and M. Castelli, “Transfer learning with convolutional neural networks for diabetic retinopathy image classification. a review,” *Applied Sciences 2020*, vol. 10, p. 2021, 2020.
- [77] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing*, vol. 452, pp. 48–62, 2021.

- [78] J. D. *et al.*, “Imagenet: A large-scale hierarchical image database,” pp. 248–255, 2010.
- [79] Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels,” in *Advances in Neural Information Processing Systems* (S. B. *et al.*, ed.), vol. 31, Curran Associates, Inc., 2018.
- [80] T.-Y. L. *et al.*, “Focal loss for dense object detection,” pp. 2980–2988, 2017.
- [81] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2014.
- [82] I. S. *et al.*, “On the importance of initialization and momentum in deep learning,” 2013.
- [83] S. Noppitak and O. Surinta, “dropcyclic: Snapshot ensemble convolutional neural network based on a new learning rate schedule for land use classification,” *IEEE Access*, vol. 10, pp. 60725–60737, 2022.
- [84] L. W. *et al.*, “Convolutional neural network with automatic learning rate scheduler for fault classification,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021.
- [85] G. H. *et al.*, “Densely connected convolutional networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, 2016.
- [86] T. Bezdán and N. B. Džakula, “Convolutional neural network layers and architectures,” *International Scientific Conference on Information Technology and Data Related Research*, pp. 445–451, 2019.
- [87] K. H. *et al.*, “Deep residual learning for image recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 2015.
- [88] S. R. *et al.*, “Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification,” *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.
- [89] J. Lever, M. Krzywinski, and N. Altman, “Points of significance: Classification evaluation,” *Nature Methods*, vol. 13, pp. 603–604, 2016.
- [90] “Visual studio code - code editing. redefined.” <https://code.visualstudio.com/>. Accessed: 2023-07-15.

- [91] “colab.google.” <https://colab.google/>. Accessed: 2023-07-15.
- [92] “os — miscellaneous operating system interfaces — python 3.11.4 documentation.” <https://docs.python.org/3/library/os.html>. Accessed: 2023-07-15.
- [93] “pandas - python data analysis library.” <https://pandas.pydata.org/>. Accessed: 2023-07-15.
- [94] “Numpy.” <https://numpy.org/>. Accessed: 2023-07-15.
- [95] “Scipy.” <https://scipy.org/>. Accessed: 2023-07-15.
- [96] “Simpleitk - home.” <https://simpleitk.org/>. Accessed: 2023-07-15.
- [97] “Opencv - open computer vision library.” <https://opencv.org/>. Accessed: 2023-07-15.
- [98] S. V. D. W. *et al.*, “Scikit-image: Image processing in python,” *PeerJ*, vol. 2014, 2014.
- [99] J. J. V. G. *et al.*, “Computational radiomics system to decode the radiographic phenotype,” *Cancer Research*, vol. 77, pp. e104–e107, 2017.
- [100] “pyfeats · pypi.” <https://pypi.org/project/pyfeats/>. Accessed: 2023-07-15.
- [101] “scikit-learn: machine learning in python — scikit-learn 1.3.0 documentation.” <https://scikit-learn.org/stable/>. Accessed: 2023-07-15.
- [102] “scikit-optimize: sequential model-based optimization in python — scikit-optimize 0.8.1 documentation.” <https://scikit-optimize.github.io/stable/>. Accessed: 2023-07-15.
- [103] “Pytorch.” <https://pytorch.org/>. Accessed: 2023-07-15.

Apêndice A

Documento submetido à RECAPD

000 **Ultrasound Versus Elastography** 001 **in the Study of Sarcopenia**

002
003 Luís Lopes
004 AndreLuis_lopes2000@hotmail.com
005 Alexandra André
006 alexandra.andre@estescoimbra.pt
007 Jaime B. Santos
008 jaime@deec.uc.pt
009 José Silvestre Silva
010 jose.silva@academiamilitar.pt
011

Department of Physics, University of Coimbra, Portugal
Coimbra Health School, Coimbra, Portugal

CEMMPRE, ARISE, Department of Electrical and Computers Engineering, University of Coimbra, Portugal
CINAMIL & Portuguese Military Academy, Portugal
LIBPhys-UC & LA-Real, University of Coimbra, Portugal

012 **Abstract**

013
014 Using medical images from the rectus femoris muscle acquired through
015 ultrasound and elastography, the present work analyses the classification
016 of sarcopenia using modern deep learning architectures and conventional
017 machine learning models. The dataset, consists of 180 medical images
018 collected from 30 people with ages ranging from 20 to 75.

019 The study explores a variety of models, including deep learning models
020 like DenseNet 121, VGG16, VGG19, ResNet50, and Inception V3,
021 as well as traditional models like logistic regression and neural networks.
022 The performance of the neural network model is in line with deep learning
023 models. The Neural Network achieved the best performance with an
024 F1_score of 99.81%. This work provides insights into the function of
025 classical and deep learning methodologies as diagnostic tools for early
026 intervention and enhanced care of an aging global population, shedding
027 light on their potential to distinguish sarcopenia properly.

028 **Keywords:** Sarcopenia, ultrasound, elastography, traditional classifiers,
029 Deep Learning

030 **1 Introduction**

031 Sarcopenia's impact on muscle health, particularly in an aging popula-
032 tion, has boost research into detection and intervention strategies. Med-
033 ical imaging methods like ultrasonography and elastography offer non-
034 invasive insights into muscle characteristics [2] [1].

035 Models based on deep learning and machine learning techniques are
036 able to identify patterns and specific characteristics of sarcopenia from
037 large amounts of data, which can lead to more accurate classifications.
038 Automating the classification process can be faster and more efficient than
039 the manual process, saving time and resources [3].

040 This work explores the potential for classifying sarcopenia using ma-
041 chine learning and deep learning models. The dataset, comprising 180
042 images from the rectus femoris muscle of 30 participants aged 20 to 75,
043 undergoes thorough preprocessing. Traditional machine learning models
044 and advanced deep learning architectures, including VGG16 and others,
045 are evaluated. By comparing these approaches, the research seeks to illu-
046 minate effective pathways for early sarcopenia diagnosis and intervention,
047 ultimately contributing to improved healthcare for an aging population.

048 **2 Methodology**

049 The methodology used in this study applies a methodical approach to data
050 preparation, preprocessing, and data augmentation. It also makes use of
051 conventional classification approaches and deep learning Convolutional
052 Neural Networks (CNNs). The following is a summary of the methodol-
053 ogy's essential phases.
054
055

056 **2.1 Data Preparation**

057 The rectus femoris muscle images from the dataset, were acquired by
058 ultrasonography and elastography. Thirty volunteers, ranging in age from
059 20 to 75, provided the data points needed to create a sample that was
060 representative of each age group. A region of interest (ROI) used as input
061 to the models was extracted from the acquired images, as shown in Figure
062 1.

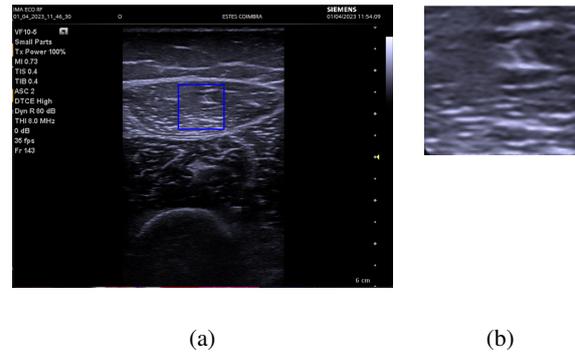


Figure 1: Comparison between the images obtained by the ultrasound scanner and the region of interest extracted from the previous image (not in real sizes).

056 **2.2 Preprocessing and Data Augmentation**

057 Recognizing the critical role of data volume in traditional machine learn-
058 ing, it was employed the sliding window technique to increase the dataset
059 size. This approach effectively expanded the machine learning input by
060 segmenting the images, thereby augmenting the number of samples avail-
061 able for analysis.

062 In contrast, the approach to augmenting the input data for deep learn-
063 ing models was done by introducing variations to the input images: Hor-
064 izontal flips were applied to introduce mirror images, Gaussian noise was
065 added to simulate real-world variability, and rotations were executed at
066 six distinct angles (-90, -45, -15, 15, and 90 degrees).

067 **2.3 Traditional Classification**

068 The conventional classification techniques included feature extraction,
069 normalization, and selection. In order to compare how well different clas-
070 sifiers performed in identifying sarcopenia from the acquired images, the
071 classifiers' performance was analyzed.

072 Throughout the optimization process, several functionalities and ap-
073 proaches were systematically explored. At several stages of optimization,
074 these functions, which are essential to the classification process, were as-
075 sessed and compared. They had a crucial role in determining the classif-
076 iers' learning behavior and discriminating abilities, which affected their
077 ability to distinguish between images with and without sarcopenia. This
078 thorough analysis allowed the understanding how several factors affected
079 the performance of the classifiers and ultimately helped to correctly iden-
080 tify sarcopenia in the images.

081 **2.4 Deep Learning**

082 Transfer learning was used in the pre-trained models like DenseNet 121,
083 VGG16, VGG19, ResNet50, and Inception V3. To increase feature rele-
084 vance, attention-related mechanisms were included.

085 This research included a wide range of features and properties es-
086 sential to model optimization. A thorough investigation and comparison
087 of variables including batch size, loss functions, optimization techniques,
088 and learning rate schedulers was conducted. The model's learning dy-
089 namics and convergence were considerably impacted by the selection of
090 these components. This research intended to identify the ideal configura-

tion that would enable the accurate diagnosis of sarcopenia in the images by methodically evaluating and contrasting these elements.

3 Results and Discussion

3.1 Datasets

We used datasets: A, B, and C. The dataset A has 89 principal Regions of Interest (ROIs) from ultrasound scans; 46 of these ROIs were from elderly adults with sarcopenia and 43 from young, healthy people.

The dataset B was 91 primary ROIs from elastography including 42 from young, healthy individuals and 49 from elderly persons with sarcopenia.

With 137 principal ROIs, Dataset C consists of ultrasonography and elastography data, where 85 of these come from young, healthy people in their 20s, and 95 from older people with sarcopenia.

The sliding window technique became a key technique for dataset analysis in the area of machine learning. The most effective combinations were found by a methodical investigation of various window widths and step sizes. The dataset A performed better with a 38 point window and 3 point step size. While dataset B showed improved outcomes with a 35-point window and a 3-point step size. On the other hand, the combined dataset C showed the best results with a window size of 38 and a step size of 3. The datasets are presented in Table 1 (see ML rows).

Several data augmentation techniques were employed in the effort to fully use the potential of the models. These techniques aimed to increase the input's diversity in order to produce input that was more like real-world scenarios and so increase the model's adaptability. The details about these datasets are in Table 1 (see DL rows).

Table 1: Datasets used for both machine learning (ML) and deep learning (DL) models.

	Datasets	Number of images	Ratio (unhealthy:healthy)
ML	A	12 509	45:55
	B	8 612	42:58
	C	19 460	43:57
DL	A	22 784	52:48
	B	23 296	54:46
	C	46 080	53:47

3.2 Results

In the current work, it is present a comparison of five classic classifiers and five Deep Learning Models. The computing requirements of testing different models has an impact on the choice of classifier in the deep learning context. In order to optimize, a base set of functions was initially created. Then, in each step, different functions were compared, and the one that produced the highest performance was chosen, modifying the basic set of functions. In machine learning, the hyperparameters were first optimized using Bayes search, and then the functions for feature normalization, feature selection, and feature quantity were compared. Following deep learning model tuning, a number of features and properties were assessed. A thorough evaluation and comparison of batch size, loss functions, optimization methods, and learning rate schedulers was conducted. The usage of attention processes was also evaluated. The final performance of each classifier is present in Table 2.

Table 2: Classifier's final performance using datasets A, B and C.

Model	F1_score (%)		
	A	B	C
Neural Network	99.81	99.59	99.20
Nu Support Vector	94.03	93.32	92.84
Logistic Regression	96.84	96.16	92.90
Stochastic Gradient Descent Classifier	97.85	96.79	95.01
Support Vector Machine	99.04	98.34	96.81
DenseNet 121	96.94	97.63	97.63
VGG 16	99.02	98.55	99.12
VGG 19	99.33	99.12	98.71
ResNet 50	95.36	94.32	93.22
Inception V3	93.50	89.13	88.35

A number of reasons may have contributed to machine learning classifiers' superior performance. Notably, given that classical machine learning is more effective with smaller, simpler datasets, the dataset's size and complexity may have been better suited for it.

It's possible that hyperparameter adjustment, which is frequently important, also had an impact. The machine learning models were subjected to refined hyperparameter optimization, resulting in configurations that take advantage of the features in the dataset. Due to the high computational power required by the deep learning models, the hyperparameter tuning was not as rigorous as the one made for the machine learning classifiers.

While typical machine learning classifiers performed more effectively when it came to categorizing cases of sarcopenia, an interesting finding regarding the effectiveness of deep learning models was made. Some deep learning models, including VGG16 and VGG19, showed high performance without requiring intensive optimization efforts. These models repeatedly demonstrated their promise as reliable classifiers across a variety of datasets by achieving an F1_score of 99%. They did have the capacity to reach comparable results more quickly, highlighting the importance of their inbuilt feature extraction abilities. It is also worth mentioning that the models ResNet 50 and Inception V3 reached a better performance without the attention mechanism. This finding could be related to their architecture that facilitates the propagation of features through different layers, thus not needing an attention mechanism.

Finally, both machine learning and deep learning models had comparable performance across all datasets. Overall, dataset A resulted in the best performances and dataset C in the lowest. In most models, the decrease in performance isn't relevant, but in models like Inception V3 is noticeable. This could be due to difficulties in defining elastography images, leading to suboptimal performance in datasets containing elastography images.

4 Conclusions and Future Work

In this research, it was possible to define sarcopenia in both ultrasound and elastography images with an average F1_score above 90% and in some models reaching 99%. These results show the model's ability to enhance the diagnostic of sarcopenia, facilitating more effective healthcare interventions in the context of an aging population.

It also reveals a complex interaction of elements that contributed to the advantage of machine learning classifiers superiority. Hyperparameter optimization, model complexity, data complexity, and computational resources all have difficult relationships that highlight how versatile model performance is.

To elect the best method for classifying sarcopenia, future studies may incorporate more thorough deep learning model exploration, including various architectures and configurations. Additionally, increasing the focus on hyperparameter optimization and domain-specific modifications has the potential to improve classification.

5 Acknowledgements

This research is sponsored by national funds through FCT – Fundação para a Ciência e a Tecnologia, under the project UIDB/00285/2020 and LA/P/0112/2020

References

- [1] Giuseppe Sergi *et al.* Imaging of sarcopenia. *European Journal of Radiology*, 85:1519–1524, 2016. ISSN 0720-048X. doi: 10.1016/J.EJRAD.2016.04.009.
- [2] Roger A. Fielding *et al.* Sarcopenia: an undiagnosed condition in older adults. current consensus definition: prevalence, etiology, and consequences. international working group on sarcopenia. *Journal of the American Medical Directors Association*, 12:249–256, 2011. ISSN 1538-9375. doi: 10.1016/J.JAMDA.2011.01.003. URL <https://pubmed.ncbi.nlm.nih.gov/21527165/>.
- [3] Shengfeng Liu *et al.* Deep learning in medical ultrasound analysis: A review. *Engineering*, 5:261–275, 4 2019. ISSN 2095-8099. doi: 10.1016/J.ENG.2018.11.020.

Apêndice B

Conjuntos de dados criados pelo método de janela deslizante

Tabela B.1: Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para imagens de ecografia.

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
1	8	3	41 429	19 884	21 545	48:52
2	8	8	6 284	3 020	3 264	48:52
3	11	3	37 736	18 040	19 696	48:52
4	11	8	5 741	2 740	3 001	48:52
5	14	3	34 221	16 288	17 933	48:52
6	14	8	5 230	2 489	2 741	48:52
7	14	13	2 125	1 020	1 105	48:52
8	17	3	30 884	14 628	16 256	47:53
9	17	8	4 734	2 246	2 488	47:53
10	17	13	1 967	924	1 043	47:53
11	20	3	27 725	13 060	14 665	47:53
12	20	8	4 265	2 006	2 259	47:53
13	20	13	1 784	849	935	48:52
14	20	18	999	484	515	48:52
15	23	3	24 744	11 584	13 160	47:53
16	23	8	3 837	1 792	2 045	47:53

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
17	23	13	1 561	743	818	48:52
18	23	18	931	452	479	49:51
19	23	23	608	277	331	46:54
20	26	3	21 941	10 200	11 741	46:54
21	26	8	3 376	1 577	1 799	47:53
22	26	13	1 407	669	738	48:52
23	26	18	805	372	433	46:54
24	26	23	560	259	301	46:54
25	29	3	19 316	8 908	10 408	46:54
26	29	8	3 043	1 430	1 613	47:53
27	29	13	1 256	570	686	45:55
28	29	18	720	334	386	46:54
29	29	23	488	232	256	48:52
30	29	28	358	183	175	51:49
31	32	3	16 869	7 708	9 161	46:54
32	32	8	2 660	1 223	1 437	46:54
33	32	13	1 143	522	621	46:54
34	32	18	639	293	346	46:54
35	32	23	437	214	223	49:51
36	32	28	320	160	160	50:50
37	35	3	14 600	6 600	8 000	45:55
38	35	8	2 318	1 048	1 270	45:55
39	35	13	981	457	524	47:53
40	35	18	575	265	310	46:54
41	35	23	393	190	203	48:52
42	35	28	279	127	152	46:54
43	35	33	230	94	136	41:59
44	38	3	12 509	5 584	6 925	45:55
45	38	8	2 011	902	1 109	45:55
46	38	13	842	393	449	47:53
47	38	18	501	235	266	47:53
48	38	23	338	159	179	47:53

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
49	38	28	255	109	146	43:57
50	38	33	219	90	129	41:59
51	38	38	172	76	96	44:56

Tabela B.2: Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para imagens de elastografia.

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
1	8	3	31 751	15 384	16 367	48:52
2	8	8	4 827	2 345	2 482	49:51
3	11	3	28 452	13 688	14 764	48:52
4	11	8	4 431	2 158	2 273	49:51
5	14	3	25 335	12 090	13 245	48:52
6	14	8	3 930	1 864	2 066	47:53
7	14	13	1 630	789	841	48:52
8	17	3	22 400	10 590	11 810	47:53
9	17	8	3 476	1 660	1 816	48:52
10	17	13	1 450	675	775	47:53
11	20	3	19 647	9 188	10 459	47:53
12	20	8	3 092	1 446	1 646	47:53
13	20	13	1 293	610	683	47:53
14	20	18	729	348	381	48:52
15	23	3	17 076	7 884	9 192	46:54
16	23	8	2 678	1 220	1 458	46:54
17	23	13	1 152	553	599	48:52
18	23	18	637	301	336	47:53
19	23	23	461	222	239	48:52
20	26	3	14 687	6 678	8 009	45:55
21	26	8	2 363	1 091	1 272	46:54
22	26	13	983	465	518	47:53

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
23	26	18	579	278	301	48:52
24	26	23	404	200	204	50:50
25	29	3	12 480	5 570	6 910	45:55
26	29	8	2 015	911	1 104	45:55
27	29	13	861	381	480	44:56
28	29	18	510	239	271	47:53
29	29	23	348	165	183	47:53
30	29	28	259	110	149	42:58
31	32	3	10 455	4 560	5 895	44:56
32	32	8	1 677	743	934	44:56
33	32	13	743	331	412	45:55
34	32	18	473	215	258	45:55
35	32	23	290	123	167	42:58
36	32	28	221	98	123	44:56
37	35	3	8 612	3 648	4 964	42:58
38	35	8	1 449	637	812	44:56
39	35	13	637	288	349	45:55
40	35	18	381	170	211	45:55
41	35	23	252	107	145	42:58
42	35	28	196	90	106	46:54
43	35	33	161	77	84	48:52
44	38	3	6 951	2 834	4 117	41:59
45	38	8	1 176	481	695	41:59
46	38	13	520	232	288	45:55
47	38	18	304	133	171	44:56
48	38	23	223	98	125	44:56
49	38	28	170	82	88	48:52
50	38	33	137	67	70	49:51
51	38	38	113	55	58	49:51

Tabela B.3: Dados dos conjuntos de dados utilizados na comparação do tamanho da janela e do passo no método de janela deslizante para a união das imagens de ecografia com as imagens de elastografia.

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
1	8	3	73 180	35 268	37 912	48:52
2	8	8	11 111	5 365	5 746	48:52
3	11	3	66 188	31 728	34 460	48:52
4	11	8	10 172	4 898	5 274	48:52
5	14	3	59 556	28 378	31 178	48:52
6	14	8	9 160	4 353	4 807	48:52
7	14	13	3 755	1 809	1 946	48:52
8	17	3	53 284	25 218	28 066	47:53
9	17	8	8 210	3 906	4 304	48:52
10	17	13	3 417	1 599	1 818	47:53
11	20	3	47 372	22 248	25 124	47:53
12	20	8	7 357	3 452	3 905	47:53
13	20	13	3 077	1 459	1 618	47:53
14	20	18	1 728	832	896	48:52
15	23	3	41 820	19 468	22 352	47:53
16	23	8	6 515	3 012	3 503	46:54
17	23	13	2 713	1 296	1 417	48:52
18	23	18	1 568	753	815	48:52
19	23	23	1 069	499	570	47:53
20	26	3	36 628	16 878	19 750	46:54
21	26	8	5 739	2 668	3 071	46:54
22	26	13	2 390	1 134	1 256	47:53
23	26	18	1 384	650	734	47:53
24	26	23	964	459	505	48:52
25	29	3	31 796	14 478	17 318	46:54
26	29	8	5 058	2 341	2 717	46:54
27	29	13	2 117	951	1 166	45:55
28	29	18	1 230	573	657	47:53
29	29	23	836	397	439	47:53

N.º	Janela	Passo	Total de ROIs	ROIs com patologia	ROIs saudáveis	Rácio
30	29	28	617	293	324	47:53
31	32	3	27 324	12 268	15 056	45:55
32	32	8	4 337	1 966	2 371	45:55
33	32	13	1 886	853	1 033	45:55
34	32	18	1 112	508	604	46:54
35	32	23	727	337	390	46:54
36	32	28	541	258	283	48:52
37	35	3	23 212	10 248	12 964	44:56
38	35	8	3 767	1 685	2 082	45:55
39	35	13	1 618	745	873	46:54
40	35	18	956	435	521	46:54
41	35	23	645	297	348	46:54
42	35	28	475	217	258	46:54
43	35	33	391	171	220	44:56
44	38	3	19 460	8 418	11 042	43:57
45	38	8	3 187	1 383	1 804	43:57
46	38	13	1 362	625	737	46:54
47	38	18	805	368	437	46:54
48	38	23	561	257	304	46:54
49	38	28	425	191	234	45:55
50	38	33	356	157	199	44:56
51	38	38	285	131	154	46:54

Apêndice C

Hiperparâmetros otimizados nos classificadores de *machine learning*

Tabela C.1: Classificadores otimizados e os respectivos hiperparâmetros com o seu intervalo de valores.

Classificador	Hiperparâmetros e respetivo intervalo de valores
<i>Logistic Regression</i>	$C \in [10^{-6}, 10^6]$ $solver \in \{lbfgs, liblinear, newton-cg, newton-cholesky, sag, saga\}$ $max_iter \in [200, 500]$
<i>Stochastic Gradient Descent</i>	$eta0 \in [0.01, 1.0]$ $loss \in \{hinge, log_loss, modified_huber, squared_hinge, perceptron\}$ $penalty \in \{l1, l2, elasticnet, None\}$ $alpha \in [10^{-6}, 10^{-2}]$ $learning_rate \in \{constant, optimal, invscaling, adaptive\}$
<i>Nu Support Vector</i>	$nu \in [0.01, 0.8]$ $kernel \in \{linear, rbf, poly, sigmoid\}$ $degree \in [1, 8]$
<i>k-Nearest Neighbors</i>	$n_neighbors \in [1, 50]$ $weights \in \{uniform, distance\}$ $p \in \{1, 2\}$ $algorithm \in \{auto, ball_tree, kd_tree, brute\}$

<i>Gaussian Process</i>	$optimizer \in \{fmin_l_bfgs_b, callable, None\}$
<i>Multi-layer Perceptron</i>	$hidden_layer_sizes \in [1, 100]$
	$activation \in \{identity, logistic, tanh, relu\}$
	$learning_rate \in \{constant, invscaling, adaptive\}$
	$max_iter \in [200, 500]$
<i>Support Vector Machine</i>	$C \in [10^{-6}, 10^3]$
	$gamma \in [10^{-6}, 10^1]$
	$degree \in [1, 8]$
	$kernel \in \{linear, poly, rbf, sigmoid\}$
<i>Naive Bayes</i>	$var_smoothing \in [10^{-6}, 10^{-1}]$
<i>Decision Tree</i>	$criterion \in \{gini, entropy\}$
	$max_depth \in [2, 20]$
	$min_samples_split \in [2, 20]$
<i>Random Forest</i>	$n_estimators \in [10, 100]$
	$criterion \in \{gini, entropy\}$
	$max_depth \in [2, 20]$
	$min_samples_split \in [2, 20]$