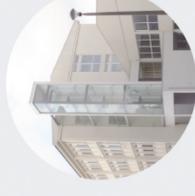
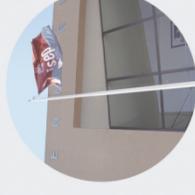


Padrões Acústicos de Voz na Detecção de Doença Coronária Aterosclerótica

MÉLISSA PATRÍCIO
julho de 2020



Padrões Acústicos de Voz na Detecção de Doença Coronária Aterosclerótica

Padrões Acústicos de Voz na Detecção de Doença Coronária
Aterosclerótica

MÉLISSA PATRÍCIO
Julho de 2020

MÉLISSA PATRÍCIO
Julho de 2020



Mestrado em Engenharia de Computação e Instrumentação Médica

Padrões Acústicos de Voz na Detecção de Doença Coronária Aterosclerótica

Mélissa Patrício

Dissertação apresentada no Instituto Superior de Engenharia do Porto para a obtenção
do grau de Mestre em Engenharia de Computação e Instrumentação Médica

Orientadores

PROF. DR. LUÍS COELHO

Porto, 2020

Agradecimentos

Em primeiro lugar quero agradecer ao meu orientador Prof. Luís Coelho por todo o seu apoio, pela sua disponibilidade no esclarecimento de dúvidas, pelos bons conselhos e por toda a sua dedicação na realização deste projeto.

Um muito obrigado à equipa de Angio-TC do Centro Hospitalar de Vila Nova de Gaia / Espinho, em especial ao Dr. Nuno Ferreira, Dr. Gustavo Morais, Técnico Wilson e Técnica Mónica pela simpatia com a qual sempre me receberam, pela disponibilidade e constante ajuda durante a recolha de todas as informações, sem as quais seria impossível atingir os objetivos delineados e pelos seus conselhos na realização deste projeto.

Aos meus pais, Jacinto e Isabel, pela paciência e apoio, por todos os esforços que realizaram ao longo destes últimos anos para me proporcionar esta oportunidade na minha vida pois sem eles, nada disto teria sido possível. À minha restante família pela compreensão das minhas ausências, por acreditarem sempre em mim e continuarem a cuidar de mim.

Aos meus colegas de curso, pelos bons momentos, por me terem acompanhado sempre e terem partilhado comigo momentos e aprendizagens desta vida académica. Aos meus amigos de longa data, pela compreensão, amizade e essencialmente, por terem continuado sempre presentes neste percurso, independentemente da distância.

Ao meu namorado Hugo pela amizade, paciência e compreensão. Ele me fez entender o verdadeiro sentido das palavras amor, companheirismo e cumplicidade.

Um sincero obrigado a todos!

Resumo

Atualmente, a doença cardiovascular é a principal causa de mortalidade em ambos os sexos, afetando maioritariamente os países industrializados. De entre todas as doenças cardiovasculares destaca-se inclusivamente, a Doença Arterial Coronária. Por se tratar de uma doença fortemente associada a inúmeros fatores de risco, ao longo das últimas décadas foram desenvolvidos modelos de estimativa do risco cardiovascular com base nestes. No entanto, estudos realizados identificam algumas limitações no uso destas ferramentas, revelando a necessidade de se investigar técnicas ou biomarcadores adicionais para melhorar as estimativas do risco cardiovascular. Recentemente na literatura científica, identificou-se uma possível relação entre os parâmetros do sinal voz e a doença coronária aterosclerótica. De facto, a análise acústica é uma ferramenta eficaz, não invasiva e de fácil aplicação para realizar avaliações objetivas da função vocal. Deste modo, a presente dissertação de mestrado tem como objetivo avaliar o potencial da análise da voz em identificar padrões acústicos associados à doença coronária, através da recolha de sinais de voz de pacientes diagnosticados com patologia cardíaca e de indivíduos saudáveis. Desenvolveu-se uma base de dados constituída pelas gravações de voz e registo de informações clínicas recolhidas de um total de 76 participantes. Tendo em conta os avanços tecnológicos da Inteligência Artificial e o sucesso das suas técnicas na área da medicina, para o processamento e análise dos dados recolhidos utilizou-se a técnica de *deep learning* devido à sua capacidade em identificar padrões e discriminar dados. Foram desenvolvidos diferentes modelos com base nas duas abordagens para a realização da análise acústica. No geral, os resultados obtidos foram satisfatórios para os diferentes dados recolhidos, inclusivamente com desempenhos de 70% a 88% de exatidão na classificação por parte dos modelos desenvolvidos.

Palavras-Chave: Doença Arterial Coronária; Análise Acústica; Redes Neurais Profundas.

Abstract

Actually, cardiovascular disease is the main cause of mortality in both genders, affecting mostly industrialized countries. Among all cardiovascular diseases, Coronary Artery Disease stands out particularly. Because it is a disease strongly associated with numerous risk factors, over the past few decades models of cardiovascular risk estimation based on these have been developed. However, studies carried out identify some limitations in the use of these tools, recognizing the need to investigate additional techniques or biomarkers to improve estimates of cardiovascular risk. Recently in the scientific literature, a possible relationship between the parameters of the voice signal and atherosclerotic coronary disease has been identified. In fact, acoustic analysis is an effective, non-invasive and easy to apply tool for performing objective assessments of vocal function. Thus, the present master's thesis aims to evaluate the potential of voice analysis in identifying acoustic patterns associated with coronary disease, through the collection of voice signals from patients diagnosed with cardiac pathology and from healthy individuals. A database was developed consisting of voice recordings and registration of clinical information collected from a total of 76 participants. Taking into account the technological advances of Artificial Intelligence and the success of its techniques in the field of medicine, the deep learning technique was used for the processing and analysis of the collected data due to its ability to identify patterns and discriminate data. Different models have been developed based on two approaches for performing acoustic analysis. In general, the results obtained were satisfactory for the different data collected, including performances of 70% to 88% of accuracy in the classification by the models developed.

Key-words: Coronary Artery Disease; Acoustic Analysis; Deep Neural Networks.

Índice

Agradecimentos	iii
Resumo	v
Abstract.....	vii
Índice	ix
Lista de Figuras	xiii
Lista de Tabelas.....	xv
Lista de Abreviaturas	xvii
Capítulo 1: Introdução	1
1.1. Enquadramento.....	1
1.2. Motivação.....	2
1.3. Formulação de Hipóteses	3
1.4. Objetivos	4
1.5. Estrutura da dissertação.....	4
Capítulo 2: Conceitos Básicos	5
2.1. Doença Arterial Coronária	5
2.1.1. Fisiopatologia.....	7
2.1.2. Fatores de risco.....	8
2.1.3. Avaliação do Risco Cardiovascular	10
2.2. Análise Acústica do Sinal de Voz.....	14
2.2.1. Domínio do tempo.....	16
2.2.2. Domínio das frequências	19
2.3. Estado da Arte	21

2.3.1. Análise acústica como meio de diagnóstico.....	21
Capítulo 3: Métodos.....	25
3.1. Procedimento.....	25
3.2. Base de dados.....	25
3.2.1. Critérios de Seleção e Exclusão dos participantes	25
3.2.2. Caracterização Geral da Amostra.....	26
3.2.3. Aspectos Éticos	27
3.3. Recolha de dados.....	27
3.3.1. Ferramentas	27
3.3.2. Procedimento.....	28
3.4. Processamento e análise de dados	29
3.4.1. <i>Deep Learning</i>	29
3.4.2. Avaliação da performance.....	37
3.4.3. Ferramentas	39
Capítulo 4: Implementação	43
4.1. Arquitetura geral	43
4.2. Pré-Processamento e desenvolvimento do modelo	44
4.2.1. Modelo no domínio do tempo	44
4.2.2. Modelo no domínio das frequências	48
4.3. Método de Avaliação	51
Capítulo 5: Resultados	55
5.1. Desenvolvimento da Base de Dados	55
5.2. Resultados obtidos	59
5.2.1. Dados recolhidos na emissão das vogais sustentadas.....	61
5.2.2. Dados recolhidos na leitura dos textos foneticamente balanceados	69
5.3. Comparação e discussão dos resultados.....	74
5.3.1. Impacto das informações clínicas adicionais	76
5.3.2. Principais técnicas e parâmetros utilizados nos modelos DL.....	77
5.3.3. Teste Adicional: exclusão de participantes fumadores	79

5.4. Síntese dos resultados.....	80
Capítulo 6: Conclusão.....	83
6.1. Considerações finais.....	83
6.2. Principais limitações e perspetivas futuras.....	85
Bibliografia	87
A. Parecer da Comissão de Ética do Centro Hospitalar de Vila Nova de Gaia/Espinho....	95
B. Consentimento Informado Livre e Esclarecido	97
C. Guião como instrumento para a recolha de dados.....	101
D. Evolução da ‘accuracy’ em cada categoria CAD-RADS.....	103

Lista de Figuras

Figura 2.1: Formação da placa aterosclerótica. Adaptado de [33].	8
Figura 2.2: Cálculo do risco a 10 anos de doença cardiovascular fatal em populações de países com alto risco cardiovascular. Adaptado de [12].	12
Figura 2.3: Representação gráfica dos parâmetros acústicos fornecidos pela análise acústica da voz como a frequência fundamental, shimmer e jitter. Adaptado de [57].	18
Figura 2.4: Relação entre a escala Mel e a frequência linear. Fonte: [61].	20
Figura 2.5: Exemplo de um banco de filtros triangulares para obtenção dos coeficientes Mel-Cepstrais. Fonte: [61].	20
Figura 3.1: Ilustração do conjunto de funcionalidades utilizadas do software Praat.	27
Figura 3.2: Analogia entre o modelo de um neurónio artificial e um neurónio biológico. Adaptado de [89].	31
Figura 3.3: Esquematização de uma rede neuronal profunda. Adaptado de [89].	32
Figura 3.4: Exemplo das funções de ativação mais populares em modelos de <i>deep learning</i> [91].	32
Figura 3.5: Esquematização da arquitetura típica de uma rede neuronal convolucional. Adaptado de [94].	34
Figura 3.6: Ilustração do software Spyder.	40
Figura 4.1: Esquematização dos passos gerais do desenvolvimento do modelo.	43
Figura 4.2: Exemplo representativo de uma arquitetura de uma <i>deep feed forward</i> .	47
Figura 4.3: Fluxograma representativo das fases do desenvolvimento do modelo no domínio do tempo.	48
Figura 4.4: Exemplo representativo de uma arquitetura de rede neuronal convolucional.	51
Figura 4.5: Fluxograma representativo das fases do desenvolvimento do modelo no domínio das frequências.	51
Figura 4.6: Esquematização do processo de validação cruzada <i>k-fold</i> estratificado com 5 interações ($k = 5$).	53
Figura 5.1: Descrição do grupo em análise e de controlo face ao sexo e faixa etária.	56
Figura 5.2: Amostra do estudo segundo o sistema de classificação CAD-RADS.	57
Figura 5.3: Fatores de risco associados aos participantes do estudo (em %).	58

Figura 5.4: Esquema das principais variáveis utilizadas em cada modelo de <i>deep learning</i> desenvolvido.	59
Figura 5.5: Desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação binária.	62
Figura 5.6: Desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclases.	65
Figura 5.7: Desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na leitura dos textos, no cenário de problema de classificação binária.	70
Figura 5.8: Desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclases.	71
Figura 5.9: Comparação dos resultados obtidos pelos modelos desenvolvidos entre amostras com e sem participantes fumadores, no cenário de problema de classificação binária.	80
Figura 5.10: Comparação dos resultados obtidos pelos modelos desenvolvidos quanto à ‘ <i>accuracy</i> ’ entre os dois conjuntos de dados recolhidos, no cenário de problema de classificação binária.	81
Figura 5.11: Comparação dos resultados obtidos pelos modelos desenvolvidos quanto à ‘ <i>accuracy</i> ’ entre os dois conjuntos de dados recolhidos, no cenário de problema de classificação multiclases.	81
Figura D.1: Evolução do desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na emissão das vogais sustentadas, quanto à avaliação individual das categorias CAD-RADS.	104
Figura D.2: Evolução do desempenho da ‘ <i>accuracy</i> ’ durante os processos de treino da validação cruzada <i>k-fold</i> com os dados recolhidos na leitura dos textos, quanto à avaliação individual das categorias CAD-RADS.	105

Lista de Tabelas

Tabela 2.1: Sistema de classificação CAD-RADS.....	6
Tabela 2.2: Fatores de risco cardiovasculares segundo Rupert (2012).	10
Tabela 2.3: Score de Cálcio (Método de Agatston). Adaptado de [42].	13
Tabela 2.4: Identificação de alguns softwares de análise acústica. Adaptado de [47], [52].	15
Tabela 3.1: Caraterização geral dos participantes face ao grupo da amostra e ao sexo.	26
Tabela 3.2: Representação de uma matriz de confusão. Adaptado de [105].....	37
Tabela 3.3: Listagem das bibliotecas Python instaladas para o desenvolvimento do projeto.	41
Tabela 4.1: Descrição dos parâmetros acústicos obtidos após análise acústica dos sinais de voz recolhidos.	45
Tabela 5.1: Caraterização geral dos participantes face ao grupo, faixa etária e sexo.	56
Tabela 5.2: Resumo das informações utilizadas como variáveis de entrada em cada modelo desenvolvido.	60
Tabela 5.3: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação binária.	61
Tabela 5.4: Resultados obtidos pelos modelos com os dados recolhidos na emissão das vogais sustentadas e sem os participantes fumadores, no cenário de problema de classificação binária.	63
Tabela 5.5: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclassess.	64
Tabela 5.6: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclassess.	66
Tabela 5.7: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 1.	66
Tabela 5.8: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 2.	67

Tabela 5.9: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS +3.	68
Tabela 5.10: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS +3.	68
Tabela 5.11: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, no cenário de problema de classificação binária.	69
Tabela 5.12: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclass.	71
Tabela 5.13: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclass.	72
Tabela 5.14: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 1.	73
Tabela 5.15: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 2.	73
Tabela 5.16: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS 3.	74

Lista de Abreviaturas

ASSIGN	<i>CV risk estimation model from the Scottish Intercollegiate Guidelines Network</i>
AVC	Acidente Vascular Cerebral
apq3	<i>Three-point Amplitude Perturbation Quotient</i>
apq5	<i>Five-point Amplitude Perturbation Quotient</i>
cps	Ciclos por segundo
CAD-RADS	<i>Coronary Artery Disease Reporting and Data System</i>
CNN	<i>Convolutional Neural Network</i>
DAC	Doença Coronária Aterosclerótica
dda	<i>Average absolute difference between consecutives the amplitudes</i>
ddp	<i>Average absolute difference between consecutives periods</i>
dB	Decibel
DCT	<i>Discrete Cosine Transform</i>
DCV	Doenças Cardiovasculares
DFF	<i>Deep Feed Forward</i>
DGS	Direção Geral de Saúde
DL	<i>Deep Learning</i>
EEG	Eletroencefalograma
F0	Frequência Fundamental
HNR	<i>Harmonic Noise Rate</i>
Hz	Hertz
IA	Inteligência Artificial
MFCC	<i>Mel Frequency Cepstrum Coefficient</i>
ML	<i>Machine Learning</i>
MRFIT	<i>Multiple Risk Factor Intervention Trial</i>
NHR	<i>Noise Harmonic Rate</i>
OECD	<i>Organisation for Economic Co-operation and Development</i>
OMS	Organização Mundial de Saúde
PCR	Proteína C-Reativa

ppq5	<i>Five-point Period Perturbation Quotient</i>
QRISK	<i>Cardiovascular disease risk score for the United Kingdom (2007)</i>
QRISK2	<i>Cardiovascular disease risk score for the United Kingdom (2008)</i>
rap	<i>Relative Average Perturbation</i>
ReLU	<i>Rectified Linear Units</i> (Unidade Linear Retificada)
RGPD	Regulamento Geral de Proteção de Dados
RNA	Redes Neurais Artificiais
ROC	<i>Receiver Operating Characteristic</i>
SAMPA	<i>Speech Assessment Methods Phonetic Alphabet</i>
SCORE	<i>Systematic COronary Risk Evaluation</i>
TMF	Tempo Máximo de Fonação
UH	Unidades <i>Hounsfiels</i>

Introdução

Os modelos de análise baseados no sinal de voz têm um forte potencial de demonstrar serem uma ferramenta tecnológica eficaz na detecção de uma ampla gama de estados de saúde em conjunto com os recentes avanços nos sistemas de computação através da inteligência artificial [1]. Assim, este capítulo tem como objetivo fornecer uma visão geral sobre o tema subjacente ao desenvolvimento desta dissertação após um enquadramento do mesmo, bem como as motivações e os principais objetivos e por último, apresentação da organização do presente documento.

1.1. Enquadramento

Baseada no processamento digital de sinal, a análise acústica da voz é uma técnica não invasiva para avaliação da qualidade vocal. Por este motivo, trata-se de uma ferramenta eficaz e eficiente na prática clínica dado que esta auxilia no diagnóstico objetivo de distúrbios da voz, identifica alterações na função vocal, avalia tratamentos cirúrgicos e farmacológicos e ainda, processos de reabilitação [2], [3].

Para além disto, diversos estudos têm sido realizados nos quais identificaram características do sinal voz que estão associadas a diversas entidades patológicas incluindo, por exemplo, a perturbação do espectro do autismo e a doença de Parkinson, entre outros distúrbios neurológicos [4]–[6]. Adicionalmente, um estudo recentemente desenvolvido sugere uma possível relação entre os parâmetros do sinal voz e a Doença Arterial Coronária (DAC), conhecida também como doença coronária aterosclerótica [7].

Atualmente, sabe-se que a doença cardiovascular é a principal causa de morte nos países industrializados afetando frequentemente indivíduos com idade superior a 65 anos, quer do sexo masculino, quer do sexo feminino [8], [9]. De entre todas as doenças cardiovasculares, a DAC é a principal causa de mortalidade em ambos os sexos, correspondendo a 12% de todas as mortes em todo o mundo no ano de 2004 [10].

Existem diversos fatores de risco associados a esta patologia, dos quais se destacam: a dislipidemia, o tabagismo, a hipertensão arterial e ainda, a diabetes mellitus. Estes são comumente

usados para a avaliação do risco cardiovascular com recurso a determinadas ferramentas, como por exemplo, o modelo europeu de avaliação sistemática do risco coronário (SCORE) [11]. No entanto, alguns estudos demonstraram que a utilização destas ferramentas é pouco eficaz em determinados grupos particulares de pacientes, existindo mesmo a necessidade de se realizarem estudos que avaliem prospectivamente o uso destas ferramentas na prevenção de eventos cardiovasculares [11], [12]. Assim, seria útil procurar outro tipo de testes ou técnicas adicionais para melhorar a estimativa do risco cardiovascular, através da inclusão de biomarcadores adicionais. Na verdade, cada vez mais o uso da tecnologia tem o poder de revolucionar a prestação de cuidados de saúde ao identificar, por exemplo, patologia cardíaca antes que as suas manifestações clínicas se tornem evidentes [13], [14]. A evolução tecnológica tem proporcionado grandes avanços em diversas áreas, incluindo na medicina, onde se utilizam cada vez mais inúmeras técnicas da Inteligência Artificial (IA).

Por exemplo, a abordagem baseada em *deep learning* tem sido responsável por grandes avanços em áreas como por exemplo, o reconhecimento da voz [15], [16], áreas onde os investigadores da IA têm lutado nos últimos anos para ultrapassar alguns desafios inicialmente encontrados. É uma das técnicas implementadas pela IA mais utilizada atualmente, de última geração e no geral, as arquiteturas de *deep learning* resultam de Redes Neurais Artificiais (RNA) de múltiplas camadas com modelos de neurónios não lineares [1], [17]. A sua aplicação na área da medicina tem alcançado particular destaque ao permitir o desenvolvimento de algoritmos, que auxiliam na discriminação de dados, identificação de padrões e de outras informações importantes [18], muitas das vezes sem ser necessário proceder a uma extração explícita de *features*. A classificação e deteção automática de lesões da pele [19] e de patologias, como a retinopatia diabética [20]; deteção de imprecisões em sinais de eletroencefalograma (EEG) [17]; e ainda, avaliação automática da doença de Parkinson através da análise acústica [21], são apenas alguns exemplos que demonstram a aplicabilidade desta técnica na área da medicina.

1.2. Motivação

Segundo a Organização Mundial de Saúde (OMS), as Doenças Cardiovasculares (DCV) são consideradas como a causa nº 1 de mortalidade em todo mundo. Estima-se que, no ano de 2016, cerca de 17,9 milhões de pessoas morreram de causa cardiovascular, o que representa cerca de 31% da mortalidade global [22]. Com base nestes dados, é de extrema importância identificar indivíduos que apresentam elevado risco de desenvolver uma doença cardiovascular por forma a garantir que estes recebam o mais rápido possível o tratamento e monitorização adequada com intuito de prevenir eventuais situações fatais prematuras.

Como já foi referido, a análise acústica é uma técnica útil na avaliação de uma infinidade de condições de saúde, uma vez que pequenas alterações no estado físico e mental de um indivíduo podem afetar a sua capacidade de controlar o aparelho vocal de uma forma subconsciente, podendo estas alterações nas características acústicas do sinal de voz serem facilmente mensuráveis e analisadas [1]. Por este motivo, este método de diagnóstico tem sido amplamente utilizado e testado em diversas patologias destacando-se nomeadamente, um estudo desenvolvido muito recentemente por um conjunto de investigadores no qual analisaram o sinal de voz de pacientes com e sem diagnóstico de patologia cardíaca, verificando inclusive que existiam diferenças com significado estatístico entre os dois grupos relativamente às características acústicas analisadas [7].

Considerando este facto e em conjunto com as limitações apresentadas pelos investigadores do estudo anterior, como a necessidade de se realizarem estudos adicionais noutros idiomas para além do Inglês e ainda, a procura de novas técnicas que permitam melhorar a estimativa do risco cardiovascular para prevenir a mortalidade associada às doenças cardiovasculares, foram alguns dos fatores que despertaram o interesse em desenvolver o presente trabalho.

1.3. Formulação de Hipóteses

No seguimento das motivações anteriormente apresentadas e pelas reivindicações expostas no estudo de Maor *et al* [7] coloca-se a seguinte hipótese de investigação (H1), se “*É possível identificar padrões de patologia cardíaca utilizando a análise acústica do sinal de voz?*”.

Em caso afirmativo, é possível colocar hipóteses adicionais segundo diversas perspetivas, nomeadamente:

1. Na perspetiva clínica:

H2. “A severidade da patologia pode ser estimada?”

H3. “É possível uma análise diferencial?”

2. Na perspetiva de engenharia:

H4. “Que modelo matemático permite distinguir os casos patológicos?”

H5. “Que modelo matemático traduz a evolução da patologia?”

H6. “Quais as características do sinal de voz que melhor caracterizam a presença da patologia?”

3. Na perspetiva linguística:

H7. “O português europeu tem um comportamento idêntico ao inglês, para o qual já existem resultados?”

H8. “Quais as características linguísticas que melhor traduzem a presença de patologia?”

1.4. Objetivos

Na sequência das hipóteses de investigação acima mencionadas considera-se que este estudo apresenta três objetivos principais:

O1. Confirmar o potencial da análise da voz relativamente à possibilidade de se identificarem padrões acústicos associados à patologia cardíaca em indivíduos falantes do Português Europeu;

O2. Desenvolver uma base de dados de voz contendo registos de indivíduos com patologia cardíaca já diagnosticada;

O3. Desenvolver um algoritmo para análise dos dados de voz recolhidos com recurso às técnicas de *deep learning*, tendo em conta sucesso das técnicas da IA em vários campos, inclusivamente no desenvolvimento de algoritmos que sejam capazes de identificar padrões e discriminar dados [18].

1.5. Estrutura da dissertação

O presente trabalho está organizado num total de seis capítulos sendo que o presente, é o primeiro onde se realiza uma pequena introdução ao tema e ao trabalho desenvolvido.

No segundo capítulo faz-se um levantamento dos conceitos teóricos mais importantes relacionados com as doenças cardiovasculares, nomeadamente a DAC e ainda, sobre a análise acústica.

No capítulo três é realizada a descrição de toda a metodologia adotada neste estudo incluindo, a caracterização da amostra e o procedimento implementado para a aquisição dos dados pretendidos. Para além disto, é realizada uma breve descrição sobre o *software* utilizado para a análise acústica dos dados recolhidos bem como, os conceitos teóricos fundamentais para o desenvolvimento do algoritmo com recurso à técnica de deep learning.

No capítulo quatro encontram-se descritos todos os passos para a implementação dos algoritmos desenvolvidos e propostos para o tratamento de dados deste estudo.

De seguida, no capítulo cinco são apresentados todos os resultados obtidos por parte dos modelos desenvolvidos, bem como a discussão e comparação dos mesmos.

Por último, no sexto e último capítulo, estão descritas as principais conclusões retiradas da dissertação e ainda, são apresentadas algumas perspetivas futuras que visam ultrapassar barreiras e limitações encontradas para investigações e projetos futuros.

Conceitos Básicos

Neste capítulo são apresentados os principais conceitos teóricos associados às temáticas debatidas neste trabalho fazendo-se um enquadramento mais profundo e procurando apresentar uma perspetiva mais detalhada do problema para delinear melhor os objetivos. Na secção 2.1 é apresentada a problemática da doença arterial coronária, debatendo os fatores de riscos associados e os problemas associados aos modelos de prevenção primária para a estimativa do risco cardiovascular. Na secção 2.2 é explorada a temática da análise acústica do sinal de voz e por último, na secção 2.3 é realizada o estado da arte da utilização desta técnica.

2.1. Doença Arterial Coronária

As complicações decorrentes das DCV são a primeira causa de morbilidade e mortalidade em todo o mundo, no entanto têm-se verificado uma diminuição no que diz respeito à taxa de mortalidade devido em grande parte à intervenção precoce com recurso à angiografia e às terapias farmacológicas. Ainda assim, verifica-se que a prevalência dos fatores de risco associados à patologia permanecem excessivamente altos na população [23].

Segundo o Serviço Nacional de Saúde em Portugal, cerca de 35 mil portugueses morrem anualmente por DCV, considerando ainda que estas poderiam ter sido evitadas com uma mudança simples nos hábitos alimentares [24]. Efetivamente no que toca à taxa de mortalidade por DCV em Portugal, para além desta ter diminuído nas últimas décadas, os dados mais recentes indicam que estes são bastante semelhantes aos observados por exemplo, na Europa Ocidental e na América do Norte [25].

De entre as doenças cardiovasculares, a DAC é a principal causa de mortes em ambos os sexos, correspondendo a 12% de todas as mortes em todo o mundo no ano de 2004. Ainda assim, segundo os dados da *Organisation for Economic Co-operation and Development* (OECD), é importante referir que a mortalidade por DAC também tem diminuído substancialmente nas últimas décadas [8], [10], [25].

Esta condição grave do coração, também denominada por “doença coronária” ou até “Doença Coronária Isquémica”, está fortemente associada à presença de aterosclerose nas artérias

que suprem o miocárdio, sendo este o processo fisiopatológico subjacente mais comum e que é facilmente identificado com recurso à imagiologia médica, como a angiografia coronária. No entanto, tem-se verificado que a inflamação das artérias coronárias também desempenha um papel crucial no desenvolvimento desta patologia [10], [26], [27].

Geralmente os pacientes com suspeita de DAC são avaliados através do eletrocardiograma (ECG) e, em casos positivos, é recomendável a realização de exames adicionais para avaliação da extensão e grau de oclusão coronária utilizando por exemplo, a angiografia coronária por Tomografia Computorizada (angio-TC). Adicionalmente, a angio-TC poderá também ser determinante no diagnóstico da patologia em estadios iniciais [28].

Através da utilização da angio-TC, é possível realizar um diagnóstico preciso da doença coronária, nomeadamente com recurso ao sistema de classificação CAD-RADS (*Coronary Artery Disease Reporting and Data System*). Este sistema foi desenvolvido com base em dados científicos, na experiência de profissionais especializados em imagem cardíaca e ainda, em conjunto com sociedades da área da cardiologia e radiologia. É importante adotar uma terminologia padrão para os relatórios com os resultados da angio-TC para diminuir a variabilidade de conceitos entre os profissionais e, adicionalmente, revela-se ainda importante para a investigação, bem como na melhoria da comunicação entre os profissionais de saúde e dos cuidados aos doentes com diagnóstico de DAC [29].

A classificação CAD-RADS é constituída por seis categorias, variando numa escala entre 0, caracterizada pela ausência de aterosclerose, até à escala 5 quando é revelada uma oclusão coronária total. Tal como é possível analisar na tabela 2.1, as diferentes categorias da classificação CAD-RADS estão diretamente relacionadas quanto à gravidade de estenose, identificada através da análise do achado imagiológico clinicamente mais relevante [29].

Tabela 2.1: Sistema de classificação CAD-RADS.

Categorias CAD-RADS	Grau de estenose coronária	Interpretação clínica
0	0% (sem estenose)	Ausência de DAC
1	1 – 24%	DAC não obstrutiva mínima
2	25 – 49%	DAC não obstrutiva leve
3	50 – 69%	Estenose moderada
4	(A) 70 - 99% (B) principal esquerda > 50% ou doença obstrutiva trivascular (70%)	Estenose Severa
5	100% (oclusão total)	Oclusão coronária total
N	Estudo não diagnosticável	DAC obstrutiva não deve ser excluída

A tabela 2.1 apresenta todas as categorias deste sistema de classificação para casos em que existem queixas de dor torácica estável, mas existe uma outra tabela para os casos que se verifique dor torácica aguda. Para além das categorias CAD-RADS apresentadas na tabela 2.1, estas podem ser ainda complementadas por outros modificadores adicionais, destacando-se: o modificador ‘N’, que indica a impossibilidade de existir um diagnóstico preciso; o modificador ‘S’ e ‘G’ quando são identificados no sistema coronário a presença de *stents* e excertos, respetivamente; e ainda, o modificador ‘V’ para assinalar a existência de placa vulnerável. Este último deve ser utilizado quando uma placa coronária apresentar duas ou mais características de alto risco (placa com baixa atenuação, remodelação positiva e calcificação irregular). Uma particularidade face ao primeiro modificador anteriormente apresentado (N), este pode também representar uma categoria da classificação CAD-RADS para além de ser usado como um simples modificador. Considera-se a categoria CAD-RADS N quando é necessária uma avaliação adicional para orientação do tratamento, dado que os segmentos não são totalmente interpretáveis, quer para um paciente sem estenose (0%), com estenose mínima (1 a 24%) ou leve (25 a 49%). No entanto, em situações em que a presença de estenose é moderada a grave, a estenose mais elevada deve ser classificada além do modificador N [29].

Sabe-se que dada a ocorrência de uma síndrome coronária aguda, esta rapidamente se transforma numa situação fatal e por outro lado, os riscos de recorrência de entre os sobreviventes destas síndromes são bastantes elevados. Tendo em conta estes dois aspetos, é fácil de compreender a extrema importância da prevenção destes eventos através de intervenções, quer ao nível individual, quer ao nível populacional [27]. Por este motivo, é fundamental o conhecimento do processo fisiopatológico desta patologia bem como, os fatores de risco associados e ainda, analisar quais os principais métodos aplicados quer na prevenção, quer no diagnóstico precoce desta grave condição do coração.

2.1.1. Fisiopatologia

Os mecanismos que influenciam o desenvolvimento da aterosclerose têm sido investigados desde muito cedo. Pensa-se que este processo complexo, inicia-se com a disfunção das células endoteliais que revestem as artérias coronárias (os vasos sanguíneos que fornecem o oxigénio e os nutrientes ao coração). Quando isto ocorre, as células endoteliais deixam de ser capazes de regular adequadamente o tónus vascular devido por exemplo, ao LDL elevado e modificado, aos radicais livres provocados pelo tabagismo, hipertensão e por muitos outros fatores de risco. Desta forma, formam-se as chamadas lesões ateroscleróticas que se infiltram progressivamente nas paredes dos vasos provocando uma resposta inflamatória. Estas lesões correspondem a espessamentos focais assimétricos da camada mais interna da artéria e resultam da acumulação de células, elementos do tecido conectivo, lípidos e outros detritos [26], [27], [30].

Para além disto, neste processo também se encontram envolvidos mediadores pró-inflamatórios como as citocinas, que promovem a formação de lesões e até, em alguns casos, a própria rutura da placa aterosclerótica como se pode observar na figura 2.1. Isto acontece, porque as citocinas provocam determinadas respostas como a indução de moléculas de adesão, o aumento do stress oxidativo e ainda, a diminuição da produção do óxido nítrico sendo que estas, por sua vez, estimulam o próprio desenvolvimento da aterosclerose [31].

Este processo que explica o desenvolvimento da doença aterosclerótica, é resultante da hipótese inflamatória proposta desde os meados da década de 90, no qual é evidente a forte relação entre a aterosclerose e os diversos fatores de riscos associados [27], [32].

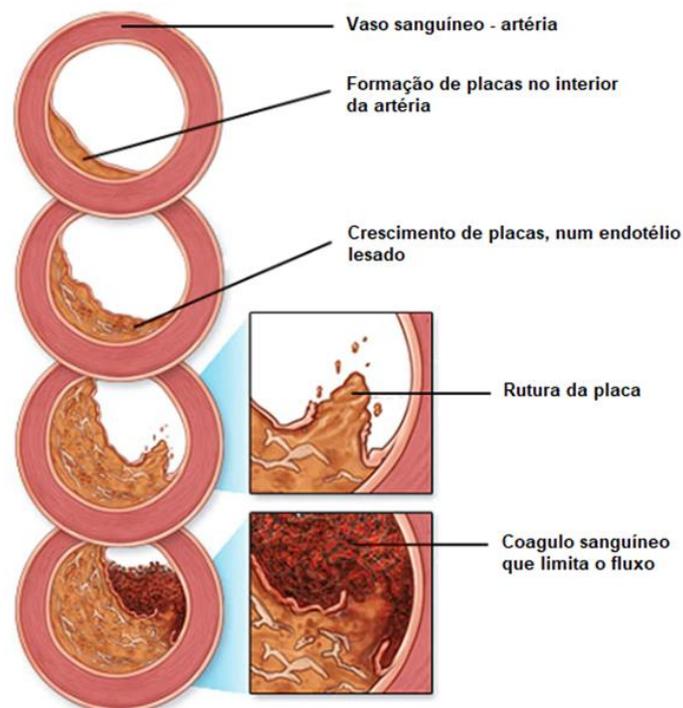


Figura 2.1: Formação da placa aterosclerótica. Adaptado de [33].

2.1.2. Fatores de risco

Nas últimas décadas têm sido realizados diversos estudos com intuito de se identificar quais os fatores de risco associados às doenças cardiovasculares, nomeadamente à doença coronária aterosclerótica. Destes destacam-se grandes estudos de coorte como, *Multiple Risk Factor Intervention Trial* (MRFIT) e *Framingham Heart Study*.

Através de uma análise conjunta entre estes últimos estudos de coorte e ainda, o *Chicago Heart Association Detection Project in Industry*, foram analisados um total de 390 000 participantes durante um *follow up* de 21 a 30 anos. Com o objetivo de determinar a frequência de exposição aos fatores de risco desta patologia, verificou-se que cerca de 87% a 100% dos

participantes com DAC fatal tinham pelo menos um fator de risco associado. No entanto, esta avaliação nestes três estudos de coorte anteriormente mencionados teve apenas em consideração os seguintes fatores de risco: valores de colesterol total superiores a 240 mg/dl; pressão arterial sistólica superior a 140 mmHg; pressão arterial diastólica superior a 90 mmHg, tabagismo e por último, diabetes *mellitus* [34]. É importante referir o grande interesse na realização deste tipo de estudos, uma vez que estes permitem o cálculo direto de medidas de risco e demonstram a relação temporal entre a exposição e a patologia em causa [35].

Ainda na década de 90, foram realizados diversos ensaios clínicos internacionais randomizados e através de uma análise realizada por Khot *et al.* [36] identificaram que 80% a 90% dos indivíduos com diagnóstico de DAC envolvidos nos ensaios apresentavam determinados fatores de risco convencionais: tabagismo, hipertensão, dislipidemia e diabetes *mellitus*.

Relativamente à diabetes *mellitus* acredita-se que as alterações metabólicas que ocorrem, nomeadamente a glicólise generalizada das proteínas e a resistência à insulina, são fatores importantes no aparecimento da aterosclerose difusa e acentuada nos vasos coronários, como as artérias de médio e pequeno calibre. Por outro lado, alguns estudos comprovaram a influência das hiperlipidemias no desenvolvimento da aterosclerose, sendo que isto pode estar relacionado em alguns casos com alterações genéticas e em outros, devido aos hábitos alimentares (por exemplo, pela ingestão excessiva de gorduras animais). No que diz respeito ao tabagismo, sabe-se que está relacionado com o desenvolvimento de várias formas clínicas de DAC e aterosclerose difusa em ambos os sexos [37]. Por último, quer a pressão arterial sistólica e a diastólica encontram-se relacionada positivamente com a DAC. Porém, por se tratar de um fator de risco modificável, com a intervenção farmacológica é possível reduzir síndromes coronárias com a diminuição da hipertensão arterial. Ademais, o colesterol e os triglicéridos também apresentam algum risco à população passíveis de intervenção terapêutica [38].

Para além destes fatores de risco, destacam-se ainda os fatores de risco não modificáveis, isto é, fatores que não são passíveis de qualquer intervenção terapêutica direta, destacando-se os seguintes fatores: a idade, o género, a história pessoal e/ou familiar e a etnia [38].

A idade é o fator de risco com particular interesse por se tratar de um forte preditor de um possível evento cardiovascular fatal. No entanto, em algumas situações é difícil perceber qual o verdadeiro processo que terá desencadeado esta patologia. Por outras palavras, é difícil discriminar se é devido ao envelhecimento da pessoa verificando-se por exemplo, alterações vasculares degenerativas ou, se é por causa do processo aterosclerótico relacionado com a própria idade. Adicionalmente, a idade é fortemente associada ao aumento da comorbidade [38]. Relativamente ao género sabe-se nomeadamente que, a menopausa é um importante fator de risco para o sexo feminino, isto porque as alterações hormonais e fisiológicas como por exemplo, a privação do estrogénio, são fatores que afetam e agravam o processo aterosclerótico nas mulheres pós-menopausa [39].

Apesar da identificação de todos estes fatores de risco, têm sido realizadas diversas pesquisas nos últimos anos com o objetivo de se identificar novos marcadores para o risco cardiovascular destacando-se, em particular, os biomarcadores genéticos [38].

A identificação dos fatores de risco (tabela 2.2) tem particular importância, na medida em que estes após serem estabelecidos, são incorporados nos diversos modelos para cálculo do risco cardiovascular que irão ser discutidos já de seguida.

Tabela 2.2: Fatores de risco cardiovasculares segundo Rupert (2012).

Fatores de Risco Cardiovasculares	
Clássicos Não Modificáveis	Idade
	Género
	História pessoal e/ou familiar
	Etnia
Clássicos Modificáveis	Hipertensão Arterial
	Colesterol
	Diabetes <i>mellitus</i>
Novos	Proteína C-reativa (PCR)
	Biomarcadores Genéticos
	Marcadores bioquímicos: microalbuminúria, cistatina C, ácido úrico e homocisteína

Assim, a identificação de alguns fatores de risco são de extrema importância para a estratificação do risco, sendo que alguns destes (fatores de risco modificáveis) ainda apresentam o benefício adicional de poderem ser combatidos com recurso a uma determinada intervenção terapêutica [38].

2.1.3. Avaliação do Risco Cardiovascular

A prevenção de qualquer doença cardiovascular deve ser valorizada e implementada, independentemente do nível socioeconómico e dos cuidados de saúde disponibilizados à população. Acredita-se que deve ser adotada uma abordagem sistemática no que toca ao rastreio das DCV, nomeadamente, para a avaliação do risco cardiovascular. Por outras palavras, deve ser realizada um programa de triagem na população em geral, mas especialmente nos subgrupos direcionados como por exemplo, indivíduos com história familiar de doença cardiovascular prematura que apresentam um risco muito elevado [12].

O risco cardiovascular global resulta do risco que é determinado pela presença e interação de todos os fatores de risco num indivíduo. É de extrema importância realizar o cálculo do risco cardiovascular global para se identificar os indivíduos que apresentem riscos consideráveis e que

devem receber tratamento como uma medida de prevenção da doença cardiovascular e até, definir-se qual a melhor terapêutica a ser aplicada. Como resultado da prevenção primária da doença cardiovascular, este cálculo deve acontecer na presença de um indivíduo no qual está em risco de desenvolver doença aterosclerótica, ou seja, indivíduos que apresentam diversos fatores de risco, mas que ainda não existe um diagnóstico estabelecido da doença. Para além disto, nos indivíduos em que já é conhecido o risco e este é elevado, o cálculo do risco é por sua vez utilizado para avaliar o benefício das intervenções terapêuticas adotadas através da comparação de um valor obtido antes da intervenção terapêutica com um outro valor obtido após o tratamento [40].

Assim, foram desenvolvidas diversas escalas ou modelos que permitem avaliar o risco cardiovascular global, como por exemplo a escala de Framingham e SCORE (*Systematic COronary Risk Evaluation*). Com recurso à identificação simultânea de vários fatores de risco anteriormente discutidos, estes instrumentos possibilitam o cálculo do risco de um determinado indivíduo em questão. Independentemente do grau de gravidade do risco calculado, estes modelos são importantes porque motivam as populações a alterar os seus comportamentos de risco após a tomada de consciência das implicações que acarreta ter ou apresentar um determinado risco cardiovascular. Isto poderá desencadear uma melhor adesão às intervenções terapêuticas estabelecidas, bem como serem definidos intervalos específicos para uma vigilância terapêutica [40].

O modelo de risco mais utilizado na Europa é o SCORE, desenvolvido por várias sociedades europeias, como por exemplo o *European Society of Cardiology*, e com base nos dados obtidos em doze estudos de coorte realizados num número abrangente de países europeus com diferentes níveis de risco cardiovascular. A primeira grande desvantagem deste modelo é que este quantifica apenas eventos cardiovasculares fatais, avaliando a mortalidade por doença cardiovascular total. O SCORE é constituído por duas tabelas uma para o sexo feminino e outro para o sexo masculino, sendo estas ainda subdivididas em mais duas: uma destinada aos fumadores e a restante para os não fumadores. Relativamente aos outros fatores de risco, quer as idades, quer os valores de colesterol e da pressão arterial sistólica estão organizados por escalões como se pode observar na figura 2.2. Ainda assim, as tabelas das quais compreende as faixas etárias entre os 50 e os 65 anos apresentam mais detalhes por se considerar que esta faixa etária corresponde ao período crítico no qual ocorrem rapidamente mudanças no risco. Por último, existe também a possibilidade de quantificar o risco relativo [38], [40].

Por outro lado, é possível identificar algumas limitações comuns neste modelo e em outros, nomeadamente: falta de alguns fatores de risco e outros parâmetros importantes (por exemplo, quando à presença de hipertrigliceridemia, o risco é subestimado); quer em mulheres, quer em diabéticos, o risco calculado poderá ser pouco preciso; e ainda, não serem considerados novos fatores de risco [38], [40].

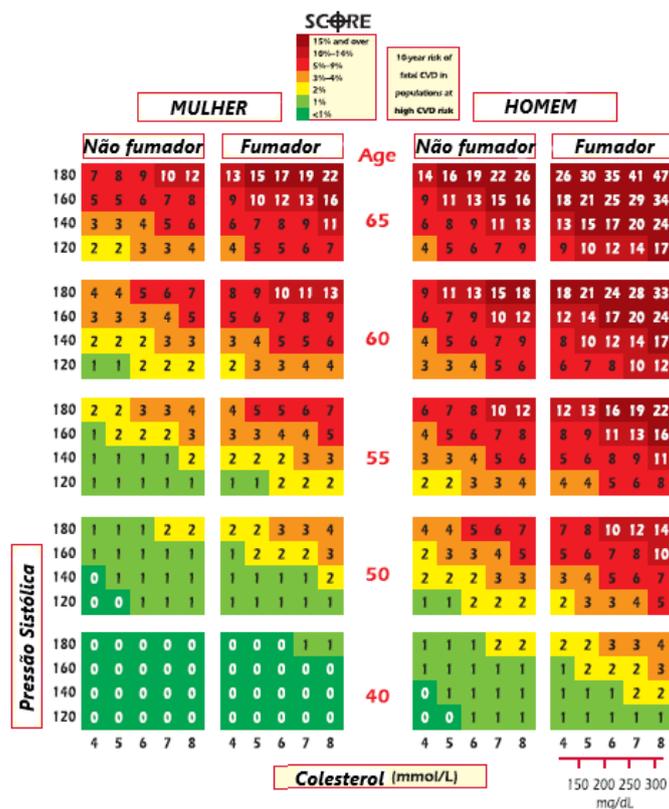


Figura 2.2: Cálculo do risco a 10 anos de doença cardiovascular fatal em populações de países com alto risco cardiovascular. Adaptado de [12].

Para além destas, muitas outras escalas de risco foram desenvolvidas nos últimos anos: escala de risco Reynolds, um outro modelo norte-americano desenvolvido especialmente para mulheres, no qual foram incluídos novos biomarcadores, como a Proteína C Reativa (PCR); e ainda, o modelo *ASSIGN* e *QRISK* desenvolvidos para serem aplicados à população do Reino Unido, sendo este último revisto em 2008 originando o modelo *QRISK2* [38].

Um outro importante método na avaliação do risco cardiovascular em doentes assintomáticos é o score de cálcio coronário, dado que a calcificação coronária quando é identificada resulta num importante marcador da presença de placas ateroscleróticas nos vasos coronários. Por outra palavras, existe uma boa correlação entre a quantificação de cálcio presente nas artérias coronárias e a carga aterosclerótica coronária total [41].

Existem diversos sistemas para a quantificação do score de cálcio em vasos coronários, mas o mais utilizado na prática clínica trata-se do Método de *Agatston*. Neste método utiliza-se a soma ponderada das lesões com densidade acima de 130 Unidades *Hounsfiels* (UH), multiplicando a área do cálcio por um determinado fator, dependendo da atenuação máxima da placa aterosclerótica. Ou seja, quando a atenuação máxima é inferior a 200 UH multiplica-se pelo fator 1; se atenuação máxima estiver entre 200 e 300 UH, multiplica-se pelo fator 2; entre 300 e 400 UH pelo fator 3; e por último, se for superior ou igual 400 UH, multiplica-se pelo fator 4. Quanto

à classificação final do score de cálcio, podem ser utilizados valores absolutos e valores com base em percentis ajustados ao sexo, idade e etnia, como é possível observar na tabela 2.3 [42].

Tabela 2.3: Score de Cálcio (Método de Agatston). Adaptado de [42].

Grau de calcificação coronária	Valores absolutos (Agatston)	Valores ajustados	Interpretação clínica
Ausente	0	0	Risco de eventos coronários futuros muito baixo
Discreto	1 a 100	≤ 75	Risco de eventos coronários futuros baixos
Moderado	101 a 400	76 a 90	Maior risco de eventos coronários futuros
Acentuado	> 400	> 90	Maior probabilidade de isquemia miocárdica

Apesar destes modelos para o cálculo do risco cardiovascular terem um papel importante no que toca à prevenção primária dos indivíduos e de no geral, serem amplamente utilizados como ferramentas que auxiliam na tomada de decisões clínicas, têm-se verificado que existe pouca evidência de que o uso destes modelos sejam eficaz [38]. Por exemplo, sabemos agora que a escala SCORE prevê o risco de mortalidade cardiovascular em 10 anos, mas na prática clínica ainda não está claramente evidente que a sua aplicação ajuda a prevenir os eventos cardiovasculares. Deste modo, um estudo realizado por Tomasik *et al.* [11] investigaram a existência de alguma avaliação da estimativa do risco calculado através do SCORE, nomeadamente na prevenção de eventos cardiovasculares graves em indivíduos sem qualquer diagnóstico de doença cardiovascular. Na verdade, após uma extensa pesquisa na literatura, não foi encontrado nenhum estudo prospetivo nem ensaios clínicos randomizados dos quais se realizasse uma avaliação da utilização da escala SCORE [11].

Adicionalmente, foram realizados diversos estudos em alguns países europeus, no qual avaliaram o uso das diretrizes europeias para prevenção das doenças cardiovasculares entre os médicos. No geral, os resultados demonstraram que apenas 36-57% dos profissionais utilizavam estas diretrizes na prática clínica, sendo que os restantes admitiam realizar outro tipo de avaliação com base na sua experiência. De entre estes estudos, a “limitação do tempo” foi a principal barreira apontada e além do mais, uma grande maioria acredita que as equações para o cálculo do risco global apresentam limitações devido em grande parte à ausência de importantes fatores de risco [12], [43]–[45].

Tendo em consideração estes fatores anteriormente referidos verifica-se assim, a crescente necessidade de se obterem novas técnicas ou testes simples e não invasivos que facilitem a triagem e melhorem a precisão do cálculo da estimativa do risco cardiovascular através da incorporação de marcadores adicionais da doença cardiovascular [7].

2.2. Análise Acústica do Sinal de Voz

A análise da voz para efeitos de diagnóstico em saúde, mais abrangente do que a foniatria, é realizada de forma muito variada, dependendo o método da patologia que se procura caracterizar. Quando se pretende analisar diretamente os elementos do aparelho fonador em atividade, tanto no caso dinâmico como estático, utilizam-se técnicas como os ultrassons, o raio-x, a tomografia axial computadorizada, a estroboscopia, a eletroglotografia, a análise acústica e a análise de imagem, entre outros.

A análise acústica é uma das primeiras ferramentas utilizadas quando se pretende realizar uma avaliação objetiva da voz, e por se tratar de um procedimento não invasivo e de fácil aplicação, tem sido amplamente utilizada nas últimas décadas [46]. Por outras palavras, este método de avaliação permite extrair dados com recurso a um determinado processamento computacional sendo que estes dados, correspondem a diferentes parâmetros acústicos que formam o sinal destacando-se por exemplo, a duração e amplitude [47].

Um dos grandes objetivos da utilização da análise acústica é quantificar e caracterizar um determinado sinal sonoro de forma não invasiva fornecendo dados quantitativos que, por sua vez, permitem identificar diferenças significativas na fala de indivíduos que apresentam alguma patologia em comparação com indivíduos saudáveis [46], [48].

No entanto, esta não é a única grande vantagem da utilização desta técnica. De facto, a análise acústica apresenta inúmeros benefícios destacando-se, nomeadamente: a avaliação da eficácia e o progresso de uma determinada intervenção terapêutica adotada; a quantificação de um conjunto de sintomas que poderão estar associados a um distúrbio da voz; ou ainda, promover a exploração dos mecanismos de fonação [49], [50].

Do ponto de vista do objeto da fonação, geralmente a emissão de vogais sustentadas, como por exemplo da vogal /a/, é o procedimento mais comum quando se recorre à análise acústica [49], no entanto têm sido realizados diversos estudos dos quais defendem existência de outros métodos que poderão ser ainda mais eficazes. Nomeadamente, Moon *et al.* [51] investigaram as diferenças obtidas entre os resultados da análise acústica usando dois métodos distintos: vogal sustentada e a leitura de um texto pré-determinado. Os seus resultados demonstraram que o método que implica a leitura de um texto é mais reflexível e real quer em termos da voz, quer em termos de padrões de entoação do indivíduo, mas ainda assim não excluem na totalidade o método da vogal sustentada. Desta forma, ambos os métodos poderão ser considerados como procedimentos eficazes da análise acústica até mesmo o método da vogal sustentada, devido à sua simplicidade, menor tempo dispensado e ainda, de fácil implementação [47], [51]. Sobre este método, Shu, Jiang e Willey [49] reforçam a importância de que se deve analisar o segmento mais estável da vogal sustentada, encontrando-se este na porção média do sinal recolhido. Adicionalmente, se o procedimento utilizado para análise acústica incluir um discurso

espontâneo, é considerado o método mais realista em termos fonéticos e prosódicos [47]. A existência de um guião ou protocolo para a gravação permite estabelecer uma homogeneidade de condições que servirão de referencial para a comparação entre os sinais recolhidos e as análises a realizar. Para um dado estudo será possível comparar diferentes informantes e, para diferentes estudos, poder-se-ão comparar resultados.

Desta forma, um determinado sinal acústico resultante da fonação de qualquer um dos métodos protocolares descritos, como por exemplo a emissão sustentada da vogal /i/, poderá ser captado por um microfone, de seguida ser digitalizado e posteriormente analisado com recurso a um programa computacional que permita observar e analisar parâmetros objetivos associados ao sinal em estudo.

Nas últimas décadas têm sido desenvolvidos diversos *softwares* de análise da voz possibilitando a análise e o processamento digital do sinal de voz, sendo possível obter vários parâmetros característicos. Existem diversas opções disponíveis, mas é evidente que a sua utilização está dependente de fatores como por exemplo, a natureza da recolha de dados ou o custo monetário envolvido [47]. Estes apresentam diversas funcionalidades destacando, inclusivamente: (1) análise da voz através da obtenção de parâmetros acústicos e da forma da onda; (2) análise espectral e de formantes; (3) filtragem do som; e ainda (4) *feedback* visual/auditivo para o utente. Na tabela 2.4 encontram-se descritos alguns dos principais softwares de análise acústica mais comumente mencionados na literatura e algumas das suas funcionalidades, sendo que estes distinguem-se tendo em conta fatores como: velocidade, compatibilidade, custos e tipo de medição disponível [47], [52].

Tabela 2.4: Identificação de alguns softwares de análise acústica. Adaptado de [47], [52].

Softwares de análise acústica do sinal de voz	
Gratuitos (<i>open source</i>)	Audacity (The Audacity Team)
	Praat (P. Boersma e D. Weenink)
	Speech Analyser (SIL International)
	Wavesurfer
Pagos	Dr. Speech, versão 4.0 (Tiger Electronics)
	Multi Dimensional Voice Program, MDVP (Kay Elemetrics)
	VoiceStudio (Seegnal)

A utilização destes programas têm vindo a aumentar cada vez mais devido em grande parte aos benefícios associados, destacando-se por exemplo [47], [53]: (1) proporcionam uma aproximação de diferentes métodos de avaliação da voz como por exemplo, a análise áudio-perceptiva e a acústica; (2) permitem obter dados normativos para diferentes realidades vocais de modo rápido e de fácil aplicação; (3) apresentam dados gráficos e numéricos relativamente à

qualidade vocal de um indivíduo; (4) permitem visualizar imagens e gráficos de fácil compreensão resultantes da análise acústica; (5) possibilitam a monitorização de uma determinada intervenção terapêutica ou ainda, a comparação de resultados vocais de diferentes abordagens terapêuticas em fases distintas do processo; e por último, (6) consideram-se uma ferramenta de deteção precoce de distúrbios vocais e laríngeos através, por exemplo, da avaliação de níveis de perturbação fonatória por comparação com os valores de referência.

Em suma, tanto na análise acústica como em qualquer outro método de avaliação e de diagnóstico, o objetivo é obter informações sobre um determinado estado de saúde de um indivíduo, sendo estas informações utilizadas posteriormente pelos clínicos em auxílio nas suas decisões quer em termos de intervenção terapêutica, quer para a realização de uma avaliação mais detalhada [48]. Os diversos parâmetros acústicos extraídos após um determinado processamento computacional, permitem definir propriedades físicas do sinal quer no domínio do tempo, quer no domínio das frequências [47] e, em ambos os casos, pode reportar-se a fenómenos de curta duração, por exemplo ocorrências intra fonémicas ou de transição fonémica, com durações de poucos milissegundos (ex.: oclusão ou ditongo), ou fenómenos de longa duração, considerando períodos de vários segundos (ex.: entoação ou espectro médio de longo termo). De seguida, são apresentados alguns parâmetros físicos do sinal de voz que podem ser analisados no domínio do tempo (secção 2.2.1) e no domínio das frequências (secção 2.2.2)

2.2.1. Domínio do tempo

Como referido anteriormente, a análise acústica permite descrever quantitativamente características do sinal de voz no domínio do tempo, inclusivamente quando se pretende avaliar a qualidade vocal. Conhecidos também por parâmetros acústicos, os mais habituais utilizados para a avaliação da função vocal são: a frequência fundamental, o *jitter* e o *shimmer*. Estes parâmetros surgem a partir dos sinais acústicos emitidos pela função laríngea e estão relacionados essencialmente, com os mecanismos de produção da voz envolvidos [46]. Na figura 2.3 é possível observar a representação gráfica das medidas de perturbação que podem ser obtidas através da análise acústica que irão abaixo ser descritas.

Frequência Fundamental

A frequência fundamental (F0) é determinada fisiologicamente pelo número de ciclos produzidos pelas cordas vocais num segundo, ou seja, é o parâmetro físico que resulta da vibração das cordas vocais por unidade de tempo [54], [55]. Considera-se ainda que está dependente das características físicas das cordas vocais nomeadamente pelo comprimento, massa, elasticidade e ainda, pela complacência destas estruturas [46]. É classificada como uma medida acústica no domínio do tempo e pode ser medida em ciclos por segundo (cps) ou em Hertz (Hz), sendo que esta última geralmente é a mais utilizada. Para além disto, considera-se uma medida acústica de

particular relevância na avaliação anatômica e funcional da laringe, dado que é capaz de demonstrar a eficiência do sistema fonatório, a biomecânica laríngea, incluindo por exemplo o comprimento natural da corda vocal e a sua interação com a aerodinâmica [47], [55].

Shimmer

O *shimmer* é uma medida de perturbação de curto termo (ciclo a ciclo) que demonstram a variabilidade em relação à amplitude do período glotal [48]. Esta medida de perturbação permite medir a variação da intensidade dos ciclos adjacentes da vibração das cordas vocais [47], [55]. Sabe-se ainda que o *shimmer* altera-se geralmente com a diminuição da resistência glótica e com lesões nas cordas vocais [54]. Segundo Teixeira e Gonçalves [56], o *shimmer* pode ser medido de quatro maneiras distintas:

- ***Shimmer absoluto (local, dB)*** é média absoluta do logaritmo de base 10 da diferença entre as amplitudes de períodos consecutivos, multiplicada por 20 e expressa em decibel (dB).
- ***Shimmer relativo (local)*** é a diferença absoluta média entre as amplitudes de períodos consecutivos, dividida pela amplitude média sendo geralmente expressa em decibel.
- ***Shimmer (apq3)*** ou Quociente de Perturbação de Amplitude de três pontos, é a diferença absoluta média entre a amplitude de um período e a média das amplitudes dos seus vizinhos, dividida pela amplitude média.
- ***Shimmer (apq5)*** ou Quociente de Perturbação de Amplitude de cinco pontos, é a diferença absoluta média entre a amplitude de um período, a média das suas amplitudes e dos seus quatro vizinhos mais próximos, dividida pela amplitude média.

Todas estas medidas serão usadas no presente trabalho, como também uma outra medida conhecida por *shimmer* (dda), que é a diferença absoluta média entre as diferenças consecutivas das amplitudes de períodos consecutivos.

Jitter

Tal como o *shimmer*, o *jitter* trata-se também de uma medida de perturbação que mede a variabilidade entre ciclos da frequência do período glotal fundamental [56]. Por exemplo, existem diversos investigadores que defendem um intervalo de valores “normais” ou de referência em jovens adultos que se situa entre os 0,5 e 1,0% no que diz respeito ao parâmetro *jitter* para fonações sustentadas [47], [55]. Acredita-se que este parâmetro acústico é afetado principalmente devido à falta de controlo na vibração das cordas vocais [54]. Segundo Teixeira e Gonçalves [56], o *jitter* pode ser avaliado de quatro formas diferentes:

- ***Jitter absoluto (local, absolute)*** é a diferença absoluta média entre períodos glotais consecutivos expressa em segundos (s) ou μ s.

- **Jitter relativo (local)** é a diferença absoluta média entre períodos consecutivos, dividida pelo período médio, em percentagem (%).
- **Jitter (rap)** ou Perturbação Média Relativa, é a diferença absoluta média entre um período, a sua média e dos seus dois vizinhos mais próximos, dividido pelo período médio em %.
- **Jitter (ppq5)** ou Quociente de Perturbação do Período de cinco pontos, é a diferença absoluta média entre um período, a sua média e dos seus quatro vizinhos, dividido pelo período médio em %.

Todas estas medidas serão utilizadas no presente trabalho, porém outra medida será tida em consideração designada por *jitter* (ddp). Este parâmetro é definido pela diferença absoluta média entre diferenças consecutivas entre períodos consecutivos, dividido pelo período médio.

Estes três últimos parâmetros acústicos acima referidos, são bastantes confiáveis por serem capazes de detetar pequenas variações no que toca à qualidade da voz. Isto poderá ser particularmente importante quando se pretende diferenciar vozes saudáveis e vozes patológicas [46], [57]. Para efeitos de análise da qualidade vocal e saúde da voz estabelece-se que uma voz saudável não deve ter jitter local superior a 1%, nem shimmer local superior a 3% [55].

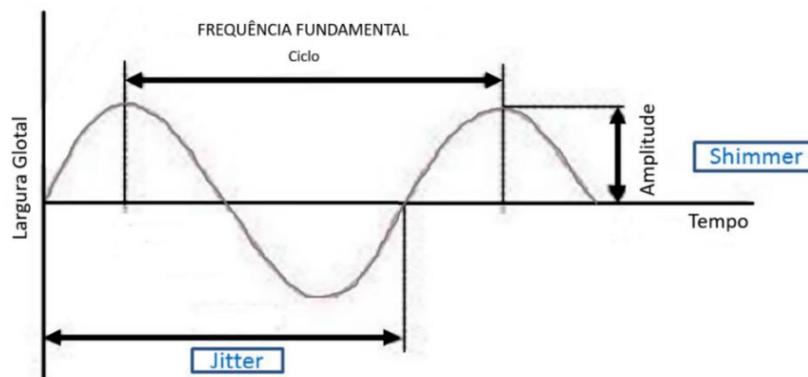


Figura 2.3: Representação gráfica dos parâmetros acústicos fornecidos pela análise acústica da voz como a frequência fundamental, *shimmer* e *jitter*. Adaptado de [57].

Parâmetros Harmónicos

Relativamente aos parâmetros harmónicos do sinal de voz, destacamos os seguintes parâmetros: autocorrelação, relação harmónico ruído (HNR) e a relação ruído harmónico (NHR).

Autocorrelação relaciona-se com a repetição de porções semelhantes da fala detetadas ao longo do sinal de voz. Ou seja, quanto maior a autocorrelação, maior é a repetição de casos semelhantes no sinal.

Por último, as medidas NHR e HNR são inversamente proporcionais e apresentam uma relação direta face à qualidade vocal, permitindo desta forma avaliar a presença de ruído no sinal de voz e determinar a qualidade vocal. Quando o valor de NHR for pequeno e em contrapartida,

o valor de HNR for grande, melhor será a qualidade vocal [58]. Por outras palavras, são uteis para quantificar o ruído presente no sinal e por outro lado, determinar o grau de rouquidão na prática clínica [46], [59]. Um trato vocal saudável apresenta, por exemplo na fonação de vogais sustentadas, um nível de ruído bastante baixo.

2.2.2. Domínio das frequências

Os Coeficientes Mel Cepstrais, em inglês denominados por *Mel Frequency Cepstrum Coefficient* (MFCC), são um tipo de características cuja extração é baseada no sistema auditivo periférico, onde se fazem ajustes à intensidade em função da frequência e onde se compartimentam gamas espectrais. É um método bastante conhecido e utilizado atualmente em tarefas tais como o reconhecimento da voz, a classificação automática entre vozes saudáveis e patológicas e ainda, em várias tarefas de reconhecimento de padrões acústicos [3], [60].

Face à frequência de sons em sinais de voz, sabe-se que a percepção humana não é capaz de seguir uma relação linear. Ou seja, para cada som com uma frequência real f (medida em Hz), um som subjetivo é calculado numa escala designada por “escala Mel” [60]. Assim, este método baseia-se no sistema de percepção humana, sendo estabelecida uma relação logarítmica entre a escala de frequência real (Hz) e a escala de frequência percetiva ou linear (mels) [2], [61]. É possível representar esta relação matematicamente através da equação 2.1 onde, A e B são constantes obtidas experimentalmente e ainda, F_{mel} e F_{Hz} correspondem à frequência medida na escala Mel e à frequência linear em Hz, respetivamente [61]:

$$F_{mel} = A \log \left[1 + \frac{F_{Hz}}{B} \right] \quad (2.1)$$

A escala Mel utiliza como unidade de medida de frequência o Mel sendo definido como referência, a frequência de 1 kHz (com 40 dB acima do limiar mínimo da audição do ouvido humano) correspondendo a 1000 mels [62]. Considerando os seguintes valores 2595 e 700 para A e B respetivamente, é possível representar a equação 2.1 graficamente num plano cartesiano e visualizar na figura 2.4 a relação logarítmica entre a escala Mel e a frequência real [61].

Relativamente ao processo de obtenção dos coeficientes MFCC, este inclui diversas etapas destacando-se: a utilização de um banco de filtros que permite a conversão do sinal para a escala Mel e de seguida, o cálculo do “*cepstrum*” através da aplicação da transformada discreta do cosseno (*DCT*) [62]. Numa fase inicial, é aplicada a Transformada Discreta de Fourier obtendo-se o espectro do sinal para posteriormente, este ser submetido ao banco de filtros triangulares [63]. Assim, uma série de filtros passa-banda com largura de banda e espaçamento constantes numa escala de frequência mel (figura 2.5), foram projetados com intuito de simular a filtragem passa-banda, tal como se pensa que acontece no sistema auditivo [60]. Para além disto, a utilização destes filtros triangulares tem a vantagem de reduzir a dimensionalidade dos dados do sinal de voz.

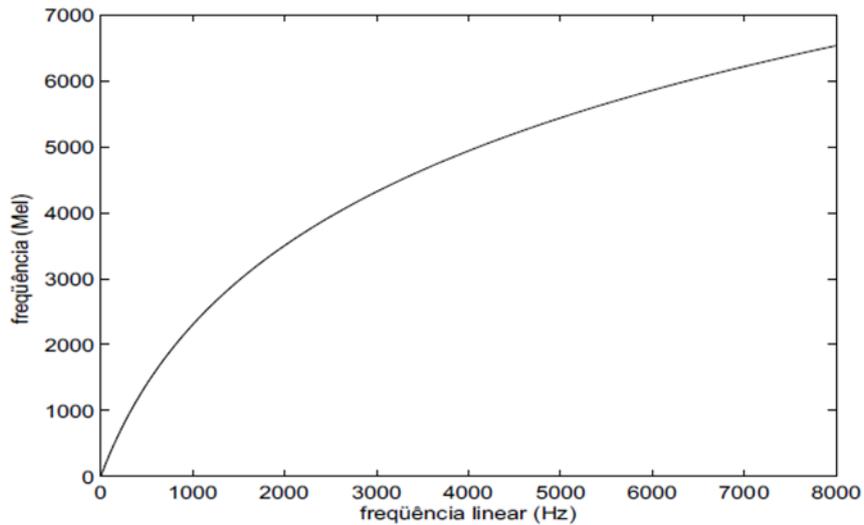


Figura 2.4: Relação entre a escala Mel e a frequência linear. Fonte: [61].

Quanto menor for o número de filtros usados neste processo, mais compacta será a informação extraída [61]. Em seguida, é calculado o logaritmo do espectro filtrado e aplica-se a transformada discreta do cosseno, que permite transformar o espectro na base de log Mel para o domínio dos tempos. No final desta operação obtêm-se os coeficientes de cepstro na frequência Mel [61], [64]. Na equação 2.2 encontra-se descrito este processo de obtenção dos coeficientes MFCC [62]:

$$c_n = \sum_{k=1}^K \log(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], n = 1, \dots, L \quad (2.2)$$

onde, L corresponde ao número de coeficientes e S_k são os coeficientes de potência da saída do k -ésimo filtro.

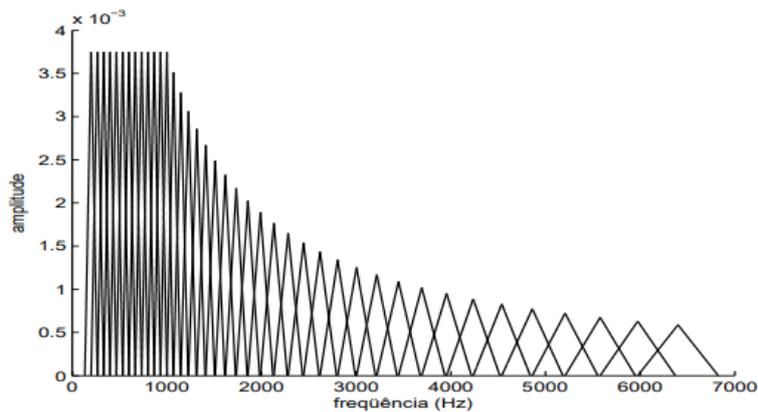


Figura 2.5: Exemplo de um banco de filtros triangulares para obtenção dos coeficientes Mel-Cepstrais.

Fonte: [61].

2.3. Estado da Arte

2.3.1. Análise acústica como meio de diagnóstico

Na literatura científica, é possível encontrar inúmeros estudos que confirmam a análise acústica como sendo uma técnica bastante valiosa e útil no diagnóstico de distúrbios da voz e de patologias da laringe [48], [65]–[68].

Adicionalmente, é também possível encontrar alguns estudos que evidenciam o uso desta ferramenta na avaliação e monitorização de uma determinada intervenção terapêutica através da comparação dos resultados antes e após o tratamento inclusive, em distúrbios da voz [69]–[71]. Por exemplo, Lu *et al.* [69] avaliaram os efeitos da terapia vocal em cerca de quarenta e cinco pessoas que apresentavam distúrbios da voz através da comparação com um grupo de controlo constituído por cinquenta mulheres saudáveis. Para tal, foi realizada uma análise acústica sendo extraídas algumas medidas acústicas (F0, *jitter*, *shimmer*, HNR) e o tempo máximo de fonação (TMP) usando o *software Praat*, a partir da emissão sustentada da vogal /i/ durante 5 segundos antes e após a realização da terapia vocal. Os resultados deste estudo apontaram uma melhoria nos pacientes após o tratamento, sendo que todos os valores das medidas acústicas e do TMP neste grupo foram estatisticamente piores comparativamente ao grupo de controlo antes do tratamento (exceto a F0) [69].

Por outro lado, existem ainda outros estudos recentes que sugerem que as características do sinal de voz poderão ser associadas com outras patologias, como por exemplo indivíduos que sofreram AVC isquémico, ataxia de Friedreich, doença de Alzheimer e outras doenças neurológicas ou neurodegenerativas [50], [72]–[77]. Por exemplo, a doença de Parkinson e a Esclerose Múltipla, são duas doenças neurodegenerativas comuns que afetam a fala dos indivíduos resultando em determinados distúrbios, como a disartria. Assim, um estudo desenvolvido recentemente, Vizza *et al.* [77] avaliaram a análise acústica como um método de processamento do sinal de voz com o objetivo de identificar distúrbios da voz em doentes com diagnóstico de, pelo menos uma das duas doenças neurodegenerativas referidas, em comparação com indivíduos saudáveis. Deste estudo verificou-se, através da aplicação de testes estatísticos, que a F0, *shimmer*, *jitter* e o HNR são características estatisticamente significativas, podendo ser utilizadas como bons indicadores no diagnóstico da disartria no grupo de pacientes estudados [77]. Simultaneamente, verificou-se no estudo desenvolvido por Meilán *et al.* [75] que as medidas *shimmer* e o HNR são capazes de identificar pacientes com a doença de Alzheimer com uma precisão de 84,8%.

Ainda em relação às doenças neurodegenerativas, Hahn e Wang [21] investigaram uma nova abordagem no que diz respeito à avaliação automática da doença de Parkinson usando dados acústicos e dados articulatórios acusticamente invertidos. Esta nova abordagem baseia-se na

possibilidade de padrões de movimento da mandíbula, língua e lábios poderem ser usados juntamente com dados acústicos para a estimativa automática da gravidade da doença de Parkinson. Devido à dificuldade na recolha de dados articulatórios, estes podem ser inversamente mapeados a partir dos dados acústicos, sendo utilizada esta técnica neste estudo [21]. Para além disto, uma das particularidades deste estudo é a utilização de técnicas de *deep learning*, ou seja, utilização de uma rede neuronal profunda para avaliação da doença de Parkinson, tendo em conta os bons resultados obtidos por exemplo, no reconhecimento da voz [78]–[80] em comparação com outras técnicas. De facto, neste estudo realizado por Hahm e Wang [21] concluíram na realização do seu estudo que a rede neuronal profunda apresentou resultados superiores comparativamente com a Regressão por Vetores de Suporte (SVR) na avaliação automática da doença de Parkinson.

Por outro lado, um estudo recentemente desenvolvido por Murton *et al.* [81] utilizaram análise acústica do sinal de voz com intuito de monitorizar doentes com insuficiência cardíaca que se encontravam em tratamento, verificando-se após este que os doentes apresentavam um aumento da F0 e uma diminuição face à variação da proeminência do pico cepstral. Este estudo coloca a hipótese de que o edema característico da insuficiência cardíaca poderá afetar inclusivamente, as cordas vocais e os pulmões e por sua vez, alterar o processo de fonação e a respiração. Foram incluídos cerca de 10 doentes com insuficiência cardíaca aguda, realizando em cada dia as gravações acústicas com recurso ao procedimento padrão: emissão de vogais sustentadas, leitura de dois textos determinados e ainda, a fala espontânea [81]. Assim sendo, segundo Murton *et al.* [81] a monitorização através do sinal de voz apresenta um forte potencial por se tratar de um método não invasivo e de fácil aplicação, sugerindo que estes biomarcadores sejam indicadores do edema relacionado à insuficiência cardíaca. Um outro estudo desenvolvido por Pareek e Sharma [82] identificaram diferenças estatísticas significativas nos parâmetros de voz, após avaliação de indivíduos com doença coronária com um grupo de controlo (sem doença). As variações entre os dois grupos foram analisadas com recurso ao espectrograma e à análise de outros parâmetros do sinal de voz (como por exemplo, o Jitter) com o software MDVP, e a partir de amostras recolhidas dos participantes que realizaram a emissão sustentada da vogal /a/ durante aproximadamente 4 segundos [82].

Recentemente, Maor *et al.* [7] estudaram a hipótese de que determinadas características do sinal de voz poderiam estar associadas com a presença de uma patologia cardíaca nomeadamente, da doença arterial coronária. Este estudo incluiu cerca de 138 pacientes referenciados para a angiografia coronária no qual, um grupo de pacientes apresentava o diagnóstico da doença e os restantes eram saudáveis, tendo sido usados os MFCC para extrair as características [7]. Posto isto, Maor *et al.* [7] identificaram cinco parâmetros vocais associados à DAC com *p-value* inferior a 0,05 com recurso à análise da regressão logística binária univariada. Para além disto, através da regressão logística binária multivariada, identificaram duas características vocais que se

encontravam independentemente associados à patologia (*p-value* de 0,009 e 0,02) principalmente quando os pacientes falavam de uma experiência emocional durante a recolha das suas vozes. Uma possível explicação apresentada por este estudo relaciona-se com a associação entre a voz e a aterosclerose ser mediada pela hipersensibilidade do sistema adrenérgico ao stress, uma vez que a relação entre o stress, o sistema adrenérgico e a aterosclerose já é bem conhecida. Assim, com este trabalho as características vocais anteriormente identificadas são apresentadas como um índice de stress subjacente e de atividade adrenérgica, fatores que se correlacionam à patologia cardíaca. No entanto, os autores alertam ainda para algumas limitações deste estudo como a necessidade de se realizarem mais pesquisas em populações maiores e diversificadas bem como a sua aplicação em outros idiomas para confirmar os resultados obtidos, como mencionado anteriormente [7].

Métodos

Ao longo deste capítulo são descritos os principais critérios de seleção dos participantes neste estudo, sendo realizada a caracterização geral da amostra. Adicionalmente, apresentam-se as ferramentas utilizadas para a recolha de dados justificando a seleção destas e o respetivo procedimento utilizado; e por último, é explicado como os dados recolhidos foram tratados e analisados através de uma técnica da IA.

3.1. Procedimento

De uma forma geral, para o desenvolvimento do presente trabalho foram definidas as seguintes etapas principais:

1. **Criação da base de dados.** Desenvolver uma base de dados através da realização das recolhas de sinais de voz em pacientes com diagnóstico de doença coronária aterosclerótica e de indivíduos saudáveis;
2. **Extração de características.** Realizar a análise acústica dos sinais de voz recolhidos e incluídos na base de dados;
3. **Criação de modelo de decisão.** Desenvolver um modelo de aprendizagem automática através de uma técnica da IA, como principal ferramenta de análise e processamento dos dados recolhidos;
4. **Identificação de padrões.** Investigar possíveis padrões acústicos relacionados com a patologia cardíaca através da exploração de modelos de aprendizagem automática que sejam capazes de proceder à discriminação entre indivíduos com e sem doença coronária aterosclerótica.

3.2. Base de dados

3.2.1. Critérios de Seleção e Exclusão dos participantes

A base de dados desenvolvida é composta por gravações de pessoas saudáveis, que constituem o grupo de controlo, e gravações de pessoas com patologia, que constituem o grupo em análise. Para a seleção dos participantes neste estudo foram utilizados como critérios de inclusão, para ambos os grupos que constituem a base de dados: (1) idade igual ou superior a 18 anos; (2) indivíduos falantes do Português Europeu; e (3) indicação clínica para a realização de uma Angiografia Coronária por Tomografia Computorizada (Angio-TC). Através deste último critério de seleção, foi possível obter a confirmação de diagnóstico de doença coronária aterosclerótica nos indivíduos do grupo em análise e por outro lado, obteve-se a confirmação da ausência de doença coronária aterosclerótica nos indivíduos do grupo de controlo.

Relativamente aos critérios de exclusão, foram tidos em consideração os seguintes: (1) indivíduos iletrados; (2) apresentar défice cognitivo; (3) história pessoal de enfarte do miocárdio, intervenção coronária percutânea ou de cirurgia de revascularização do miocárdio, e (4) história pessoal de distúrbios da voz primários e/ou secundários, causados por patologia neuromuscular ou outra. Estes mesmos critérios de seleção foram aplicados aos casos de controlo, com a exceção do diagnóstico de doença coronária aterosclerótica.

Posto isto, foram selecionados para este estudo os sujeitos que se enquadraram na pesquisa conforme os critérios de inclusão e exclusão.

3.2.2. Caraterização Geral da Amostra

Pretende-se com o presente estudo, avaliar a existência ou ausência de diferenças entre padrões acústicos que se possam relacionar com a doença cardíaca, tratando-se assim de um estudo comparativo. Desta forma, a amostra que constitui a base de dados deste estudo divide-se em dois grupos: grupo constituído apenas por indivíduos com diagnóstico de doença coronária e o grupo de controlo, composto por indivíduos saudáveis. Assim, e com base nos critérios de inclusão e exclusão, a amostra é composta por um total de 76 participantes, sendo 22 (29%) do grupo de controlo e 54 (71%) do grupo de indivíduos com doença coronária.

Na tabela seguinte (tabela 3.1) é apresentada uma análise descritiva dos participantes, tendo em conta ao grupo da amostra a que pertencem e ao sexo.

Tabela 3.1: Caraterização geral dos participantes face ao grupo da amostra e ao sexo.

Sexo	Indivíduos saudáveis		Indivíduos com diagnóstico de DAC		Total (n)
	n	(%)	n	(%)	
Masculino	15	20%	24	32%	39
Feminino	7	9%	30	39%	37
Total	22	29%	54	71%	76

3.2.3. Aspetos Éticos

Antes de se iniciar a recolha dos dados para o desenvolvimento da referida base de dados, foi imprescindível obter a autorização necessária para a realização deste projeto e obter parecer favorável da respetiva Comissão de Ética do Centro Hospitalar de Vila Nova de Gaia/Espinho (CHVNG/E) e que se encontra no Anexo A.

3.3. Recolha de dados

3.3.1. Ferramentas

A gravação dos ficheiros áudio decorreu numa sala com pouco ruído ambiental (inferior a 50 dB) e com os participantes sentados confortavelmente. Relativamente aos materiais necessários para o desenvolvimento deste estudo utilizou-se: (1) um computador portátil da marca HP e modelo Pavilion 15-cc5xx e (2) *software* para gravação e análise do sinal de voz Praat, versão 6.0.49.

O Praat trata-se de um programa computacional que permite analisar, sintetizar e manipular a voz, possuindo versões para os vários sistemas operacionais. É útil para este estudo ao permitir a gravação e consequentemente extração e análise de parâmetros acústicos, visualizar uma representação da forma da onda sonora, criar espectrogramas e ainda, curvas de *pitch* e de intensidade [83]. Por exemplo, é possível visualizar algumas destas funcionalidades na figura 3.1 após a gravação da emissão sustentada da vogal /a/ durante alguns segundos. Inclusivamente em b. e c., observa-se respetivamente a representação da onda sonora e o “*voice report*” da mesma, que consiste num relatório de parâmetros acústicos (como o *pitch* ou *shimmer*).

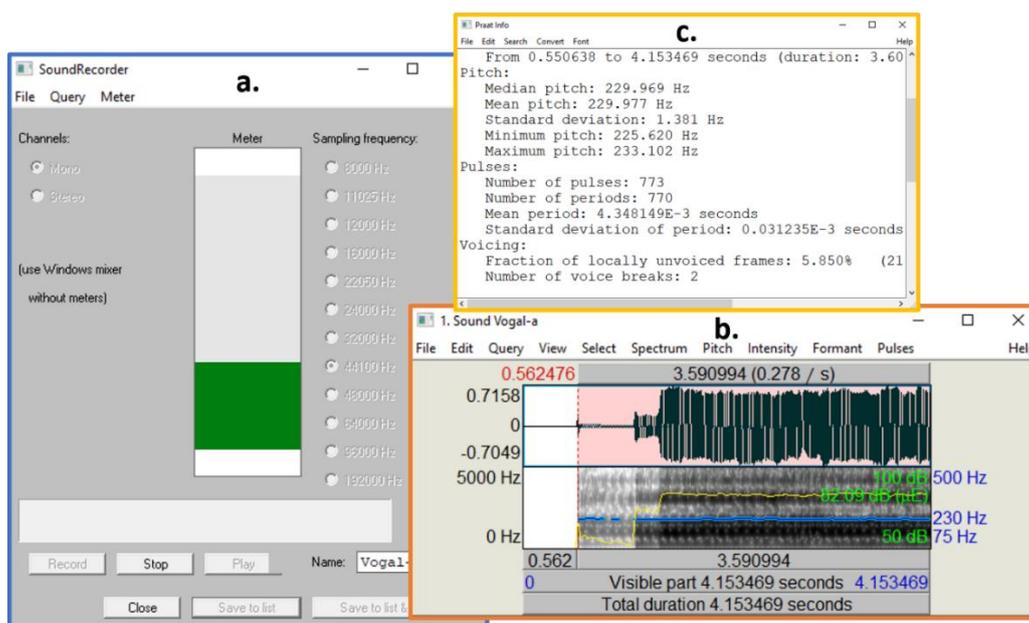


Figura 3.1: Ilustração do conjunto de funcionalidades utilizadas do *software* Praat.

3.3.2. Procedimento

As gravações do sinal de voz foram efetuadas numa sala do serviço de Cardiologia do CHVNG/E - Unidade de Diagnóstico e Intervenção Cardiovascular, entre os dias 14 de outubro e 29 de novembro de 2019.

Foi explicado a cada participante qual o objetivo da realização deste estudo e ainda, todas as instruções necessárias a serem realizadas durante o decorrer do procedimento, de modo a que todos os participantes fossem devidamente informados. Foi dada resposta a qualquer dúvida que dentro deste âmbito tenha sido colocada. Foi importante transmitir as seguintes informações aos participantes: (1) que a participação era voluntária e que podia a qualquer momento optar por desistir de participar no presente estudo, (2) que iria ser garantida a confidencialidade e o anonimato de todas as informações recolhidas, no qual o sinal de voz iria ser apenas utilizado para alcançar os objetivos propostos no âmbito desta investigação e que (3) a participação neste estudo não provocava qualquer dano físico, emocional ou financeiro. Todas as informações mais relevantes foram transmitidas aos participantes através da leitura e assinatura voluntária do Consentimento Informado Livre e Esclarecido para a participação nesta investigação, disponível para consulta no Anexo B.

Adicionalmente, foi essencial realizar um breve questionário aos participantes, antes de se proceder à recolha de dados, de forma a despistar patologias cardiovasculares prévias ou distúrbios de linguagem, com o objetivo de serem garantidas as condições necessárias para a realização da recolha de dados.

Um guião foi cuidadosamente pensado e elaborado antes da realização das gravações, para que o procedimento fosse realizado em apenas uma única sessão e com uma duração média de 5 minutos, encontrando-se a descrição e respetivos passos detalhados no Anexo C. Desta forma, foi solicitado aos participantes deste estudo, numa fase inicial, para realizarem a fonação sustentada das vogais /a/, /i/ e /u/ durante um período contínuo de aproximadamente 5 segundos. De seguida, foram apresentados aos participantes dois pequenos textos (aproximadamente 75 a 85 palavras) sendo solicitado aos mesmos a leitura de ambos, com um tom natural e intensidade regular.

Estes textos foram especialmente desenvolvidos para esta pesquisa com o objetivo de despertar ao participante do estudo, por um lado emoções positivas e por outro, emoções negativas. Pretende-se assim, contornar as questões éticas caso fosse solicitado aos participantes para expor experiências das suas vidas e de outro ponto de vista, transformar o procedimento da recolha de dados mais simples e rápido. Além do mais, ambos os textos são caracterizados por serem foneticamente equilibrados, ou seja, textos que são representativos de um conjunto alargado dos fonemas e dos formatos silábicos que constituem o Português Europeu segundo o

inventário IPA/SAMPA¹ [84], [85]. A utilização de um texto foneticamente equilibrado apresenta diversas vantagens, nomeadamente a sua aproximação ao discurso espontâneo [47]. A existência de um número vasto de fonemas permite fornecer ao sistema de *machine learning* uma maior variedade acústica e tem por objetivo facilitar a identificação de padrões característicos de patologia ou distintivos da “normalidade”.

Todas as informações recolhidas dos participantes deste estudo foram pseudonomizadas tendo em conta o Regulamento Geral de Proteção de Dados (RGPD) para que não fosse possível atribuir um sujeito aos dados de forma direta. Para tal, no momento da recolha dos ficheiros áudio, os dados de cada participante foram identificados como por exemplo: P01, P02 e assim sucessivamente, considerando o número total de participantes. Adicionalmente, no momento da realização das gravações procedeu-se à anotação num ficheiro CSV a identificação do participante, a data e hora da realização do Angio-TC, a data de nascimento e ainda, sexo do participante em estudo. Deste modo, após a gravação do ficheiro áudio e da elaboração do relatório clínico do exame realizado, o clínico experiente do serviço de cardiologia (elo de ligação ao estudo) registou no mesmo ficheiro CSV a classificação do participante em relação à ausência ou presença de doença coronária e tendo em conta o sistema de classificação CAD-RADS. Para além destas informações clinicamente relevantes, outras informações clínicas foram registadas pelo clínico experiente no referido ficheiro: o score de cálcio, o número de vasos com doença obstrutiva e ainda, os principais fatores de risco associados.

3.4. Processamento e análise de dados

Tal como referido anteriormente na secção *Objetivos* do Capítulo 1, para o processamento e análise dos dados recolhidos pretende-se recorrer a uma das técnicas mais utilizadas atualmente e de última geração da IA: *deep learning*. Sendo assim, propõe-se o desenvolvimento de um algoritmo/modelo que efetue automaticamente a distinção de indivíduos saudáveis e de indivíduos com patologia cardíaca após um processo de treino e de aprendizagem através dos dados recolhidos.

3.4.1. Deep Learning

As diversas tecnologias da IA têm revelado um enorme poder em vários aspetos da sociedade moderna. Por exemplo, os sistemas de *machine learning* (ML) são utilizados para identificar objetos em imagens ou para transcrever fala em texto, entre outras aplicações. Na verdade, cada vez mais estas aplicações utilizam especificamente uma técnica designada por *deep learning* (DL) [86]. Esta trata-se de um ramo específico da técnica ML e usa algoritmos inspirados em sistemas

¹ *Speech Assessment Methods Phonetic Alphabet*. É um alfabeto fonético de leitura ótica desenvolvido originalmente sob o projeto ESPRIT 1541, SAM (*Speech Assessment Methods*) em 1987-89.

neurônais biológicos que consistem em redes neuronais artificiais (RNA) multicamadas, no qual a unidade básica de processamento é o neurónio artificial [87]. Por outras palavras, trata-se de um modelo simples adotado por estes algoritmos com intuito de simular o mesmo mecanismo básico característico de um neurónio biológico [87].

Até ao momento, de entre os avanços mais importantes desta técnica destaca-se o reconhecimento automático de voz e de imagem permitindo por exemplo, a aplicação desta com sucesso em problemas de classificação de objetos, reconhecimento da face e ainda, na tradução de idiomas [17]. Os resultados obtidos com a utilização desta técnica são geralmente superiores aos reportados em outras técnicas.

No geral, o DL apresenta duas propriedades-chaves: (1) multicamadas não-lineares de neurónios artificiais e (2) aprendizagem supervisionada ou não-supervisionada na apresentação dos recursos em cada camada [88]. No entanto, existem outros parâmetros importantes no que toca ao desenvolvimento de um modelo DL que serão discutidos de seguida com maior pormenor nomeadamente, as funções de ativação e as métricas utilizadas na avaliação do desempenho destas redes.

Modelo do Neurónio Artificial

Como referido anteriormente, o elemento fundamental de uma típica rede neuronal é o neurónio, inspirado pelo neurónio biológico do sistema nervoso humano. Esta analogia encontra-se esquematizada na figura 3.2 sendo possível observar que se pretende simular a transferência da informação a partir de um neurónio. Particularmente, no modelo de um neurónio artificial identifica-se [89]:

1. O conjunto de sinais de entrada em que, a cada um está associado um determinado peso sináptico. Assim, a ligação entre o sinal de entrada e o neurónio tem um peso, representando deste modo as sinapses de um neurónio biológico;
2. Um operador que realiza a soma do valor do viés com as entradas ponderadas através dos respetivos pesos sinápticos para se produzir uma ativação, análogo ao axónio;
3. A função de ativação é responsável por restringir a amplitude da saída do neurónio (como por exemplo, num intervalo entre $[0,1]$) para se produzir o sinal de saída.

Relativamente ao viés e à função de ativação, estas podem ser parametrizadas pelo utilizador enquanto que os pesos sinápticos são ajustados automaticamente durante o processo de aprendizagem da rede. Cada neurónio artificial atua como uma só unidade de processamento e o sinal produzido por uma unidade é transmitido para outras unidades que se encontram organizadas em camadas, distinguindo-se três camadas principais tal como se pode visualizar na figura 3.3 [89].

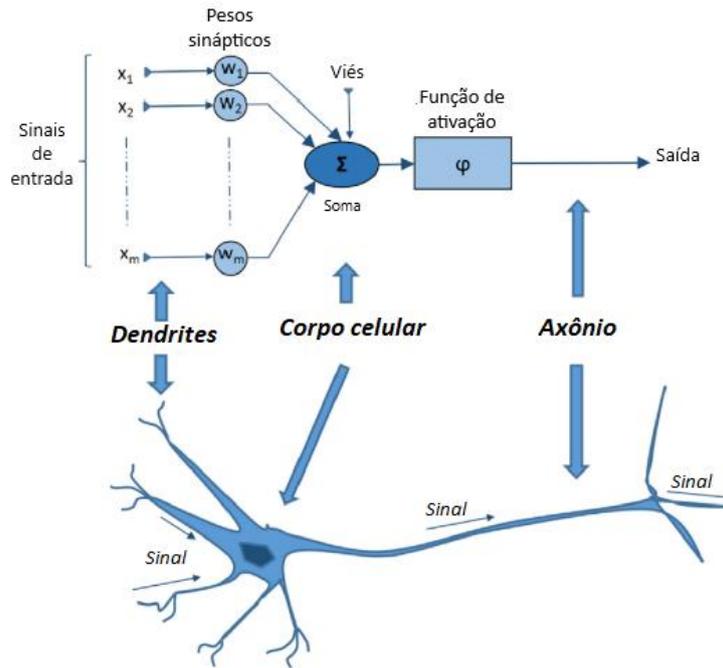


Figura 3.2: Analogia entre o modelo de um neurônio artificial e um neurônio biológico. Adaptado de [89].

A primeira camada executa o *input* do conjunto de dados de entrada e é designada de “camada de entrada” enquanto que, a “camada de saída” fornece o resultado e apresenta os respectivos valores finais. Relativamente às camadas intermédias também designadas de camadas ocultas, não produzem diretamente os *outputs* desejáveis, mas são responsáveis pelo cálculo de representações intermédias a partir dos dados de entrada, importantes para a extração de recursos de alto nível [87], [89]. Quando uma RNA é composta por diversas camadas ocultas intermédias são frequentemente designadas de redes neuronais profundas, principal característica das redes desenvolvidas pela técnica DL, bem como o tipo de arquitetura de rede mais usada.

Os modelos de decisão baseados em redes neuronais artificiais foram criados há cerca de três décadas, mas originalmente possuíam algumas limitações que travaram a sua disseminação. Por um lado, o algoritmo *back-propagation* utilizado para o treino das redes, navegava no espaço de soluções utilizando a técnica de *gradient descent*, mas nem sempre era eficaz em encontrar uma solução satisfatória. Por outro lado, os modelos matemáticos para a construção da rede, ao serem baseados em perceptrões multicamada, impunham limitações ao tipo de relações que podiam ser estabelecidas entre as variáveis.

As redes de aprendizagem profunda tornaram-se populares ao ultrapassarem os dois principais problemas descritos. Por um lado, desenvolveram-se novos algoritmos, mais eficazes na busca de uma solução ótima e ainda, criaram-se modelos de neurónios e de interligação entre estes que são mais complexos e oferecem a possibilidade de caracterizar dados cujas relações são hiperdinâmicas.

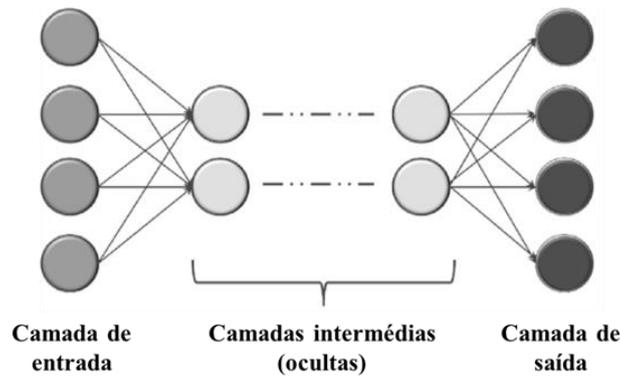


Figura 3.3: Esquemática de uma rede neuronal profunda. Adaptado de [89].

Funções de Ativação

A função de ativação, representada por $g(a)$, é um dos componentes chave das redes neurais profundas e são usadas para introduzir não-linearidade à rede. Por outras palavras, o papel de uma função de ativação é de seleccionar quais os recursos que devem ser transmitidos para o sinal de saída, sendo que geralmente ocorre uma restrição da amplitude do sinal do neurónio de saída num determinado intervalo finito [87], [90]. Portanto, uma melhor extração de características pode ser obtida através da seleção apropriada das funções de ativação [88]. Na figura 3.4 estão representados alguns exemplos de funções de ativação mais utilizadas.

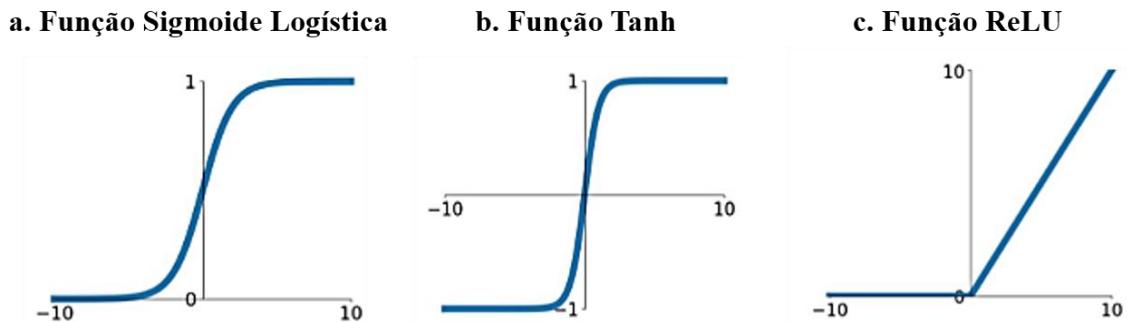


Figura 3.4: Exemplo das funções de ativação mais populares em modelos de *deep learning* [91].

A função sigmoide logística (figura 3.4 a.) transforma as variáveis em valores que variam entre 0 e 1, sendo comumente utilizada para produzir uma distribuição de Bernoulli. Esta função é definida através da seguinte equação [88]:

$$g(a) = \frac{1}{1 + e^{-a}} \quad (3.1)$$

Geralmente, uma função sigmoide logística é escolhida como função de ativação em problemas de classificação binária, ou seja, um problema que envolve apenas duas classes uma vez que permite reduzir os valores do sinal de saída ao intervalo (0,1) [90].

Uma outra função sigmoide bastante conhecida, trata-se especificamente da Função *Tanh* (figura 3.4 b.), que tal como a função de ativação anterior, depende da exigência do gradiente e é definida através da seguinte função [88], [92]:

$$g(a) = \tanh a = \frac{2}{1 + e^{-2a}} - 1 \quad (3.2)$$

A função Unidade Linear Retificada, mais conhecida por função ReLU, é considerada uma das funções de ativação mais popular em redes neuronais profundas. Esta função de ativação encontra-se descrita através da Equação 3.3 e é perfeitamente linear para sinais de entrada positivas transmitindo-as sem alterações, enquanto que as entradas negativas são avaliadas como zero, tal como se pode visualizar através da figura 3.4 c. [88], [90].

$$g(a) = \max(0, a) \quad (3.3)$$

Para problemas de classificação que envolvem mais do que duas classes, a função de ativação utilizada é a função *softmax*. Esta função calcula as probabilidades de cada classe sobre todas as classes envolvidas, sendo comumente utilizada na camada de saída da rede. Assim, a função *softmax* encontra-se descrita na seguinte equação [88], [92]:

$$g(a) = \frac{e^{a_i}}{\sum_j e^{a_j}} \quad (3.4)$$

Arquiteturas de Rede

A estrutura de uma rede neuronal é definida pela forma com que os neurónios estão conectados entre si. Tendo em conta esta característica, consideram-se dois principais grupos: as redes com apenas conexões *feedforward* e redes com conexões *feedforward* e *feedback*. É importante referir que nas redes com conexões *feedforward* o fluxo da informação ocorre apenas no sentido da camada de entrada para a camada de saída, ou seja, as conexões são sempre unidireccionais. Em contrapartida, as redes com conexões *feedforward* e *feedback* caracterizam-se por possuir conexões de realimentação, verificando-se que o processamento da informação pode ser realizado da saída para a entrada e existindo ainda, ligações entre neurónios da mesma camada com neurónios das camadas anteriores. Adicionalmente, diversas arquiteturas podem ser desenvolvidas através do aumento do número de camadas, aumentando consequentemente a complexidade da rede [93].

Para além da arquitetura já mencionada, dependendo das aplicações específicas e em grande parte do conjunto de dados utilizados no processo de aprendizagem, outras arquiteturas foram desenvolvidas destacando-se em particular neste trabalho, a rede neuronal convolucional (CNN) [89].

A CNN é uma das redes mais comumente aplicada no processamento e análise de imagens, ao permitirem extrair pequenas características das imagens de entrada. Estas consistem em várias

camadas sucessivas de neurónios no qual ocorre o processamento de dados, com intuito de se encontrar características representativas da imagem de entrada [88], [89].

Neste tipo de arquitetura, é possível identificar dois tipos particulares de camadas (figura 3.5): a camada de convolução e a camada de *pooling*. Na camada de convolução, ocorre a convolução de diferentes filtros convolucionais, deslocando estes na imagem de entrada, passo a passo, para se construir um mapa de recursos.

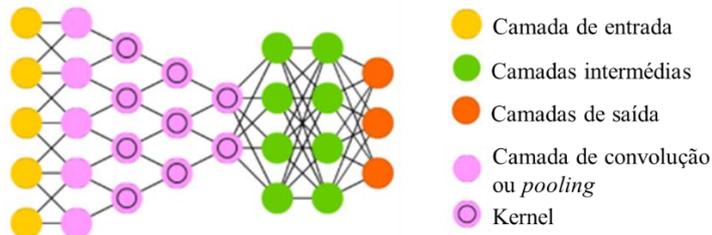


Figura 3.5: Esquemática da arquitetura típica de uma rede neuronal convolucional. Adaptado de [94].

Como os filtros convolucionais compartilham os mesmos parâmetros em cada pequena porção da imagem, ocorre uma grande redução do número de hiperparâmetros de cada imagem. Na camada de *pooling*, os resultados da convolução são resumidos, uma vez que utiliza a média ou o máximo dos recursos em vários locais do mapa de recursos, reduzindo deste modo a variância e capturando os recursos essenciais [88]. Em resumo a camada convolucional serve para detetar (múltiplos) padrões em sub-regiões multiponto no campo de entrada usando regiões recetivas, a camada de pooling serve para reduzir progressivamente o tamanho espacial da representação, reduzir o número de parâmetros e a quantidade de computação na rede e, portanto, também para controlar o excesso de ajuste (*overfitting*).

Função de Perda

No geral, os algoritmos DL sofrem na sua maioria por um processo de otimização que, por outras palavras, se refere ao processo de minimizar ou maximizar uma determinada função $f(x)$ modificando o valor de x . A função que se pretende minimizar ou maximizar é frequentemente designada por função objetiva, mas também podemos designá-la de função de custo ou de perda, quando se pretende minimizá-la [95].

Esta função de perda da qual a rede deseja minimizar quanto possível relaciona-se com a diferença entre o alvo e a saída da rede. Desta forma, a função de perda fornece uma métrica de distância e a rede pretende fazer com que esta distância seja o menor possível [90].

Posto isto, uma outra medida de desempenho bastante popular trata-se da função de perda logarítmica, também conhecida como ‘perda de regressão logística’ ou ‘perda de entropia cruzada’ (*cross-entropy loss*), utilizada nomeadamente em problemas de classificação e é definida como [90]:

$$L(y, p) = -\frac{1}{m} \sum_{i=1}^m \sum_{c=1}^C y_{i,c} \cdot \log(p_{i,c}), \quad (3.5)$$

onde m é o número de casos/exemplos, C é o número de rótulo possíveis (classes), $y_{i,c}$ é o indicador que identifica se o rótulo c é ou não a classificação correta para o exemplo i e $p_{i,c}$ é a probabilidade do modelo de atribuir o rótulo c ao exemplo i . Por exemplo, para um problema de classificação binária a aplicação da equação 3.5 resulta na seguinte equação simplificada [90]:

$$L(y, p) = -\frac{1}{m} \sum_{i=1}^m (y_i \log p_i + (1 - y_i) \log(1 - p_i)) \quad (3.6)$$

É importante referir que a escolha da função de perda está diretamente relacionada com a função de ativação utilizada na camada de saída do modelo desenvolvido, uma vez que a escolha da função de perda permite calcular o erro de forma adequada. Idealmente, um determinado modelo prevê de forma correta todas as probabilidades se a entropia cruzada calculada apresentar o valor de 0,0. Por outras palavras, para que um modelo apresente bons resultados a entropia cruzada deverá ser minimizada, tal como referido anteriormente [96].

Em suma, a função de perda é geralmente utilizada nestes modelos para se avaliar apenas a otimização do modelo, isto é, avaliar o quão bem o modelo aprendeu através do processo de treino. Por outro lado, a exatidão é usada para avaliar o desempenho do modelo em dados desconhecidos permitindo nomeadamente, a comparação de diferentes modelos desenvolvidos para o dado problema [96].

Técnicas de amostragem

Geralmente, no estudo de problemas de classificação clínicos os conjuntos de dados utilizados apresentam uma distribuição entre classes desequilibrada. Por exemplo, num problema de classificação binária, uma classe desequilibrada corresponde à classe que é representada por um grande número de amostras (classe maioritária) em comparação com a outra classe (classe minoritária) [97]. Os tradicionais algoritmos *deep learning*, dada a sua natureza estatística/probabilística, foram desenvolvidos considerando conjuntos de dados equilibrados, sendo este o verdadeiro problema da utilização dos dados desequilibrados [98]. Por este motivos foram desenvolvidos métodos que permitem a utilização da técnica de *deep learning* em conjuntos de dados não-balanceados, destacando as técnicas de reamostragem. A reamostragem consiste num método que pretende equilibrar a classe desequilibrada alterando o conjunto de dados através de duas técnicas distintas: sobreamostragem e subamostragem. No geral, a técnica de sobreamostragem aumenta a frequência da classe minoritária, copiando exemplos da classe, enquanto que a técnica de subamostragem remove exemplo da classe maioritária compensando assim, o desequilíbrio entre as classes [97]. Sabe-se que a aplicação destas técnicas pode ajudar na melhoria da precisão do classificador em conjuntos de dados desequilibrados [99].

Em alternativa, uma das técnicas mais populares e poderosas, trata-se da técnica de SMOTE (*Synthetic Majority Oversampling Technique*) e de um modo simples, este algoritmo gera novos dados artificiais com base nas semelhanças do espaço de recursos entre os exemplos da classe minoritária [99]. No entanto, este algoritmo ao criar o mesmo número de amostras de dados artificiais pode aumentar a probabilidade de ocorrerem sobreposições entre classes. Assim, foram desenvolvidos outros métodos de amostragem adaptativos, incluindo o algoritmo Borderline-SMOTE que corresponde a uma variação do algoritmo SMOTE [99].

Uma outra abordagem encontrada na literatura para lidar com conjuntos de dados desequilibrados reside na utilização do classificador ‘*bagging*’. Trata-se de um método de conjunto capaz de lidar com problemas de classificação, desenvolvido para melhorar a estabilidade e precisão dos algoritmos de *machine learning* [100]. No geral, este classificador opera de forma a obter uma previsão final através da combinação de classificações a partir de conjuntos de treino gerados aleatoriamente [100]. Para lidar com conjuntos de dados desequilibrados existe uma variação conhecida como ‘*balanced bagging classifier*’, que inclui uma etapa adicional para equilibrar o conjunto de dados de treino usando a técnica de subamostragem aleatória [101].

Técnicas de binarização

Por vezes é mais fácil implementar um classificador que distingue apenas duas classes, do que considerar mais do que duas num só problema de classificação, uma vez que os limites de decisão num problema de classificação binária são mais simples. Com base nisto, foram desenvolvidas técnicas de binarização com o objetivo de lidar com os problemas de classificação de várias classes, que consistem na divisão do problema original em problemas de classificação binária de resolução mais fácil [102]. Na literatura podem ser encontradas diversas estratégias de binarização, no entanto as estratégias mais comumente encontradas e utilizadas são [103]:

1. ‘**One vs One**’ (OvO): consiste em emparelhar uma determinada classe com as restantes classes do problema de classificação multiclases individualmente dividindo assim, o conjunto de dados de várias classes em vários problemas de classificação binária. O número de modelos criados depende do número de classes $n(n - 1)/2$, onde n é o número de classes. Desta forma, cada classe será emparelhada com outra classe, uma por uma, exigindo mais tempo e um maior custo computacional.
2. ‘**One vs Rest**’ (OvR): também conhecida por ‘*One vs All*’ (OvA). Através desta estratégia cada a classe é emparelhada com todas as restantes classes num problema de classificação binária, produzindo exatamente o mesmo número de modelos que o número de classes utilizada. Ou seja, se o número de classes for igual a 10, o número de modelos treinados também é igual a 10. No entanto, esta estratégia poderá ser afetada pelo

desequilíbrio do conjunto de dados principalmente se o número de classes for demasiado grande.

O sucesso da utilização destas estratégias em problemas de classificação de várias classes confirma-se pelo aumento da taxa de precisão da classificação [103].

3.4.2. Avaliação da performance

Relativamente às métricas de desempenho, estas são extremamente importantes para que seja possível analisar o comportamento dos algoritmos desenvolvidos inclusive, sobre os resultados produzidos. São excelentes ferramentas para medir e resumir a qualidade de um classificador quando este é testado em dados desconhecidos [104]. São também úteis durante o processo de treino e ajuste de parâmetros estruturais pois proporcionam feedback objetivo sobre a performance do modelo de classificação.

Como tal, as decisões tomadas por um algoritmo ou classificador podem ser representadas em problemas de classificação binária através de uma estrutura conhecida como matriz de confusão ou ainda, tabela de contingência (Tabela 3.2). Devido às suas características, é considerada como a base para muitas das métricas mais comuns [105], [106].

De uma forma geral, na matriz de confusão um valor previsto corresponde a um dado resultado em que o modelo desenvolvido considerou como verdadeiro (representado na Tabela 3.2 como ‘positivo’) ou falso (representado como ‘negativo’), enquanto que um valor real indica se um dado resultado é realmente correto ou incorreto, correspondendo respetivamente ao valor ‘positivo’ e ‘negativo’ [105]. Assim, consideram-se quatro categorias através da análise da tabela 3.2 [106]:

1. **Verdadeiros Positivos (VP)** correspondem a todos os exemplos de valor positivo corretamente classificados como positivos;
2. **Falsos Positivos (FP)** referem-se aos exemplos negativos que são incorretamente classificados como positivos;
3. **Verdadeiros Negativos (VN)** correspondem a negativos corretamente classificados como negativos;
4. **Falsos Negativos (FN)** referem-se nomeadamente aos exemplos positivos que são incorretamente classificados como negativos.

Tabela 3.2: Representação de uma matriz de confusão. Adaptado de [105].

		Valor real	
		Positivo	Negativo
Valor previsto	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo(VN)

Por exemplo, uma técnica desenvolvida e amplamente utilizada nas últimas décadas para selecionar classificadores ou algoritmos com base na avaliação do seu desempenho, trata-se do conhecido gráfico ou curva de *Receiver Operating Characteristic* (ROC). Os gráficos ROC são gráficos bidimensionais, nos quais está representada no eixo ‘x’, a taxa de falsos positivos e no eixo ‘y’, a taxa de verdadeiros positivos [105]. Desta forma, é possível observar com estes gráficos como o número de exemplos positivos classificados corretamente varia consoante o número de exemplos negativos classificados incorretamente por um determinado classificador em análise [106]. Por outras palavras, o gráfico ROC encontra-se diretamente relacionado com o custo/benefício obtido por um classificador através da análise quer da taxa de verdadeiros positivos (benefícios), quer da taxa de verdadeiros negativos (custos) [105]. No entanto, em situações que se verifica um grande desequilíbrio na distribuição da classe, as curvas ROC podem apresentar uma previsão excessivamente otimista [106].

Por outro lado, com base nas informações recolhidas pela matriz, é possível definir várias métricas ou medidas de *performance* que são comumente usadas, das quais se destacam [93], [105], [106]:

1. **Exatidão**, ou conhecida também por ‘*accuracy*’ e descrita na equação 3.7, observando-se que esta métrica se relaciona com a proporção de resultados verdadeiros, ou seja, com classificações corretas (verdadeiros positivos e negativos) e o número total de casos ou exemplos avaliados.

$$Exatidão = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.7)$$

2. **Precisão**, descrita na equação 3.8, permite medir a fração de exemplos ou casos classificados como positivos que são verdadeiramente positivos.

$$Precisão = \frac{VP}{VP + FP} \quad (3.8)$$

3. **Taxa de Verdadeiros Positivos**, também conhecida por ‘sensibilidade’ (ou *recall*), encontra-se descrita na equação 3.9. Esta relaciona-se com a proporção de verdadeiros positivos de entre todos os casos positivos da população em estudo.

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3.9)$$

8. **Taxa de Verdadeiros Negativos**, conhecida também por ‘especificidade’, encontra-se descrita na equação 3.10. Em contraste com a anterior, esta relaciona-se com a proporção de verdadeiros negativos de entre todos os casos negativos da população em estudo.

$$Especificidade = \frac{VN}{VN + FP} \quad (3.10)$$

De entre estas métricas, a exatidão foi a medida de desempenho tida em consideração na avaliação e análise dos modelos desenvolvidos. Esta métrica é especificada no momento da compilação do modelo considerando que, um valor de 100% corresponde à exatidão máxima da classificação pelo modelo. Por outras palavras, o classificador é avaliado com base na exatidão total referindo-se ao número de casos previstos corretamente [104], tal como descrita na equação 3.11 [107]:

$$\text{Exatidão da Classificação} = \frac{N^{\circ} \text{ de casos classificados corretamente}}{N^{\circ} \text{ total de casos}} \quad (3.11)$$

Existe também uma outra métrica que tem apresentado um ótimo desempenho na otimização e avaliação de um classificador para problemas de classificação binária, conhecida por '*F-measure*' e que se encontra descrita na equação 3.12. Esta medida representa a média harmônica entre os valores da métrica '*recall*' e da métrica precisão [104]:

$$F\text{-Measure} = \frac{2 * \text{precisão} * \text{recall}}{\text{precisão} + \text{recall}} \quad (3.12)$$

Deste modo, esta métrica tem particular interesse ao combinar duas importantes métricas referidas anteriormente, como a sensibilidade e a precisão, sendo considerada uma importante medida na avaliação do desempenho dos classificadores, para além da referida métrica '*accuracy*'.

3.4.3. Ferramentas

Neste estudo, para o desenvolvimento do algoritmo com a técnica DL foi indispensável a utilização de um software: *Spyder*.

O *Spyder* é um ambiente gráfico desenvolvido para programação na linguagem Python que oferece uma combinação exclusiva de recursos avançados de edição, análise, *debugging* e de visualização gráfica [108]. A integração destes diversos recursos facilita o desenvolvimento de scripts e permite uma melhor gestão dos projetos. Na figura 3.6 visualiza-se o ambiente quando este se inicia, sendo essencialmente composto por: uma área de desenvolvimento de código (figura 3.6 a.), outra que permite visualizar as variáveis existentes (figura 3.6 b.) e ainda, um terminal que permite a execução direta de comandos ou *scripts* (figura 3.6 c.). Este software é distribuído independentemente ou pode ser disponibilizado através de uma interface gráfica, instalada automaticamente após instalação do Anaconda, uma ferramenta computacional gratuita e disponível nos sistemas operativos Windows, Mac OS X e Linux.

Adicionalmente, o Anaconda é um gestor de ambientes de trabalho que permite gerir de forma fácil diferentes conjuntos de bibliotecas de função e várias versões do Python, entre outras ferramentas. Assim, antes mesmo de se iniciar o desenvolvimento do algoritmo com o software *Spyder*, criou-se exclusivamente um ambiente virtual no programa computacional Anaconda.

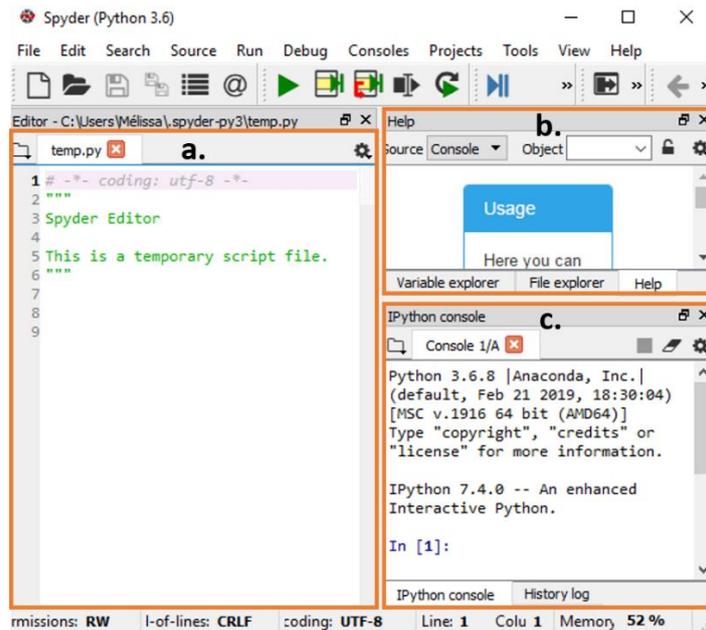


Figura 3.6: Ilustração do software Spyder.

O algoritmo foi desenvolvido em Python, na versão 3.6.8 e com o auxílio de bibliotecas adicionais ou API (*Application Programming Interface*), instaladas através da janela de comando deste software. Na tabela 3.3 estão identificadas segundo a versão utilizada, destacando-se as seguintes especificações para cada biblioteca [109]–[113]:

1. **Tensorflow:** biblioteca *open-source* para computação numérica especialmente útil no desenvolvimento de modelos de ML e DL.
2. **Keras:** possibilita o desenvolvimento e treino de modelos de DL, dando um enorme suporte para algumas das funcionalidades específicas do Tensorflow.
3. **Librosa:** biblioteca Python especialmente desenvolvida para análise de ficheiros áudio.
4. **Scikit-Learn:** módulo útil no pré-processamento de dados, bem como na comparação, validação e na definição de parâmetros em modelos de DL.
5. **Imbalanced-learn:** módulo Python que oferece várias técnicas de reamostragem serem aplicadas em conjuntos de dados com acentuado desequilíbrio entre classes.
6. **NumPy:** esta biblioteca fornece um conjunto de funções e operações que auxiliam no cálculo numérico em arrays, bastante úteis para definir os dados utilizados no processo de treino e teste em modelos de DL.
7. **Matplotlib:** possibilita a visualização dos resultados obtidos pelos modelos DL com recurso à função *plot()*.

Tabela 3.3: Listagem das bibliotecas Python instaladas para o desenvolvimento do projeto.

Módulos	Versão instalada
Tensorflow	1.13.1
Keras	2.2.4
Librosa	0.6.3
<i>Scikit-learn</i>	0.20.3
<i>Imbalanced-learn</i>	0.6.2
<i>NumPy</i>	1.16.3
<i>Matplotlib:</i>	3.0.3

Implementação

Este capítulo tem como objetivo detalhar todos os passos e etapas fundamentais na definição das estruturas de *deep learning*, bem como dos respectivos parâmetros, especialmente desenvolvidos e ajustados para o tratamento e análise dos dados recolhidos. Em 4.1 são apresentadas de uma forma geral as arquiteturas adotadas no desenvolvimento dos respectivos modelos. Na secção 4.2 são descritos todos os passos necessários na fase de pré-processamento e desenvolvimento de cada modelo. Por último, na secção 4.3 é apresentada uma breve explicação relativamente ao método utilizado para avaliação dos algoritmos desenvolvidos.

4.1. Arquitetura geral

Após a conclusão da construção da base de dados deste estudo, através dos dados recolhidos na fase anterior, foi possível definir o conjunto de dados necessário para o desenvolvimento de um classificador baseado em DL, correspondendo assim ao ponto de partida deste processo. De uma forma geral, este processo incluiu três etapas fundamentais e que se encontram esquematizadas na figura 4.1: pré-processamento, criação do modelo e ainda, avaliação e classificação dos dados.

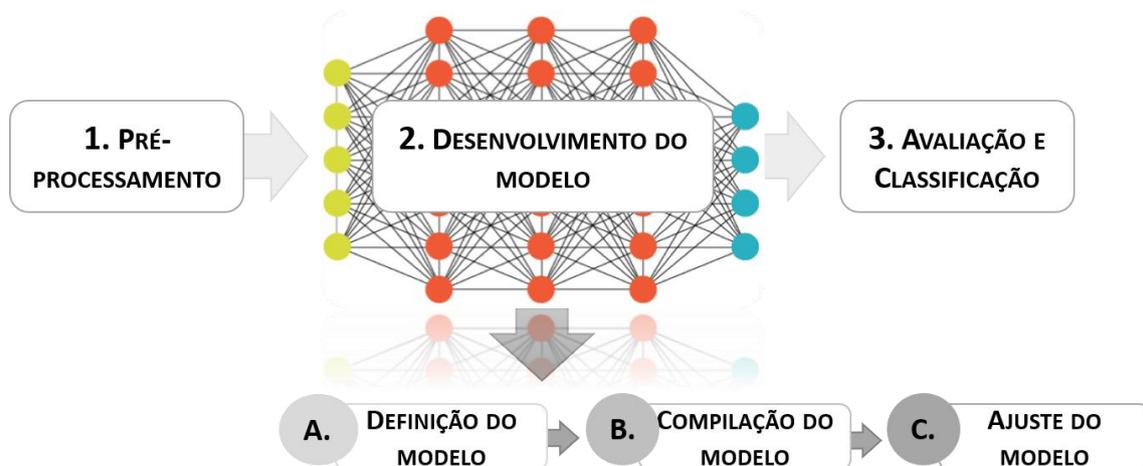


Figura 4.1: Esquematização dos passos gerais do desenvolvimento do modelo.

A primeira etapa apresentada no fluxograma da figura 4.1 corresponde ao pré-processamento do conjunto de dados, onde se realiza a sua divisão em variáveis de entrada, geralmente designada pela letra ‘X’ e nas variáveis de saída ‘Y’. Por outras palavras, nesta fase pretende-se preparar o conjunto de dados a utilizar para a etapa seguinte.

Relativamente à segunda etapa, incluem-se todos os passos fundamentais para a criação do modelo. Assim, a fase de ‘*Desenvolvimento do modelo*’ inicia-se pela etapa A como indica a figura 4.1 e relaciona-se com a definição do modelo, isto é, criação de um modelo onde são adicionadas as camadas necessárias tendo em conta a ordem no qual se pretende realizar o cálculo e o processamento dos dados. Nesta fase é necessário configurar o número de camadas que o modelo deverá conter e eleger a arquitetura mais adequada, considerando a topologia do conjunto de dados utilizado. Posto isto, é necessário otimizar a computação realizada pelo modelo através da sua compilação na etapa B, elegendo o tipo de otimizador e ainda, a definição da função de perda. Concluindo este passo, encontraram-se reunidas todas as condições para iniciar-se a última etapa (Etapa C) da fase de ‘*Desenvolvimento do modelo*’ que diz respeito ao processo de aprendizagem e treino da rede.

Por último, e após o treino da rede no passo anterior, é possível avaliar o desempenho da mesma com novos dados correspondendo à última fase esquematizada na figura 4.1, a fase de ‘*Avaliação e Classificação*’.

4.2. Pré-Processamento e desenvolvimento do modelo

Como foi explicado ao longo da seção *Análise acústica do Sinal de Voz* do Capítulo 2, esta ferramenta pode ser aplicada quer no domínio do tempo, quer no domínio das frequências. Desta forma, pretendeu-se incluir no desenvolvimento do algoritmo proposto estas duas abordagens. Para tal, foi elaborado cuidadosamente um modelo DL específico para cada abordagem, ou seja, dois modelos independentes que se distinguem na topologia do conjunto de dados usados e por sua vez, na arquitetura de rede utilizada. De seguida, são apresentados e discutidos os detalhes mais importantes do desenvolvimento de ambos os modelos de DL.

4.2.1. Modelo no domínio do tempo

Como referido anteriormente, um dos grandes objetivos da utilização da análise acústica da voz em saúde é a quantificação e caracterização não invasiva de um sinal sonoro através da obtenção de dados quantitativos por meio de um programa computacional. Por sua vez, esta técnica permite identificar diferenças significativas na fala de indivíduos que apresentam alguma patologia em comparação com indivíduos saudáveis, um dos objetivos propostos na realização do presente trabalho. De seguida, são descritas as principais etapas do desenvolvimento do modelo

DL com base na análise acústica no domínio do tempo, ou seja, por meio de parâmetros acústicos mais utilizados por esta técnica.

Pré-processamento e leitura da base de dados

Uma das principais características do desenvolvimento deste modelo é a utilização do software *Praat* na extração dos parâmetros acústicos que permitem quantificar e caracterizar o sinal de voz analisado. Desta forma, após a recolha dos dados foi utilizado um *script* específico no software *Praat* que permitiu realizar de forma automática a análise acústica de todos os sinais de voz recolhidos.

Após a realização da análise acústica, foram obtidos até cerca de 18 valores para cada sinal de voz recolhido. Estes valores relacionam-se com os parâmetros acústicos mais frequentemente utilizados e descritos anteriormente, como a frequência fundamental (ou *pitch*), as medidas de perturbação e de qualidade, estando identificados e descritos na tabela 4.1.

Tabela 4.1: Descrição dos parâmetros acústicos obtidos após análise acústica dos sinais de voz recolhidos.

Parâmetro acústico	Abreviatura utilizada	Significado
Frequência fundamental	PitchMed	Frequência fundamental vocal média
	PitchSDev	Desvio padrão da frequência fundamental vocal
	pMin	Frequência fundamental mínima
	pMax	Frequência fundamental máxima
Medida de perturbação: <i>jitter</i>	Jitt1 (local)	<i>Jitter</i> relativo, em %
	Jitt2 (local, absoluto)	<i>Jitter</i> absoluto, em segundos
	Jitt3 (rap)	Perturbação Média Relativa
	Jitt4 (ppq5)	Quociente de Perturbação do Período
	Jitt5 (ddp)	Diferença absoluta média entre diferentes ciclos de jitter
Medida de perturbação: <i>shimmer</i>	Shim1 (local)	<i>Shimmer</i> relativo, em %
	Shim2 (local, dB)	<i>Shimmer</i> absoluto, em decibel
	Shim3 (apq3)	Quociente de Perturbação de Amplitude em 3 ciclos
	Shim4 (apq5)	Quociente de Perturbação da Amplitude em 5 ciclos
	Shim5 (apq11)	Quociente de Perturbação da Amplitude em 11 ciclos
	Shim6 (dda)	Diferenças absolutas médias entre amplitudes de períodos consecutivos
Caraterísticas harmónicas	Hnr1	Autocorrelação média
	Hnr2 (NHR)	Relação média entre ruído e harmónicos
	Hnr3 (HNR)	Relação média entre harmónicos e ruído

A obtenção de todos estes valores possibilita uma análise acústica eficiente de todos os sinais de voz recolhidos. Assim, e tendo em conta o procedimento adotado no processo de recolha dos sinais de voz dos participantes, todos os valores que resultaram da análise acústica realizada automaticamente foram organizados em diferentes ficheiros CSV, sendo estes valores utilizados como variáveis de entrada para o modelo de DL em desenvolvimento.

Para o desenvolvimento do referido modelo, é de extrema importância importar as bibliotecas necessárias e que já foram anteriormente mencionadas na secção 3.4.1 do Capítulo 3. Este passo é crucial para garantir o correto funcionamento do mesmo. De seguida, importou-se a base de dados utilizada, que neste caso trata-se de um dos ficheiros CSV anteriormente referidos através da biblioteca *NumPy*. Posto isto, foi necessário dividir o conjunto de dados utilizado em variáveis de entrada e na variável de saída. Os valores resultantes da análise acústica, correspondem às variáveis de entrada do modelo, como já foi referido e os valores da variável de saída foram atribuídos considerando dois cenários:

1. **Problema de classificação binária.** Valores de saída discretos no conjunto (0,1) para distinguir entre os participantes sem doença (valor 0) e participantes com doença coronária (valor 1).
2. **Problema de classificação multiclasse.** Valores de saída discretos no conjunto (0...6) para distinguir os participantes tendo em conta as categorias do sistema de classificação CAD-RADS, incluindo a gravidade da doença coronária.

Assim, à variável de saída foram apenas associados os valores que se encontravam na última coluna do ficheiro CSV. Os valores utilizados na última coluna variaram consoante o que se pretendeu classificar na amostra recolhida, de acordo com os dois cenários apresentados e respetivas classes.

Definição do modelo

Entretanto, definiu-se o modelo de DL através da identificação e descrição da sequência de camadas necessárias para o seu desenvolvimento.

Para este modelo optou-se por utilizar uma arquitetura de rede totalmente conectada com um determinado número de camadas, também conhecida por *Deep Feed Forward* (DFF). No geral, na definição de todas as camadas do modelo foi necessário: (1) especificar o número de neurónios; (2) o método de inicialização; e por último, (3) especificar a função de ativação. Uma exceção à regra ocorre na definição da camada de entrada, sendo também necessário definir-se o número de variáveis de entrada utilizadas nesta camada. Relativamente ao método de inicialização, utilizou-se a inicialização padrão dos pesos uniforme através da biblioteca Keras. E utilizou-se a função de ativação RELU em todas as camadas exceto na camada de saída, tendo sido usada a função de ativação sigmoide para garantir que a saída obtida pela rede esteja compreendida entre 0 e 1,

quando o problema de classificação binária é o cenário considerado. Por outro lado, quando se pretende que saída pela rede esteja compreendida entre 0 e 6 no caso do problema de classificação multiclases, a função de ativação utilizada é a *'softmax'*. Na figura 4.2 encontra-se esquematizado um exemplo da arquitetura de rede DFF com integração dos referidos conceitos alusivos à rede.

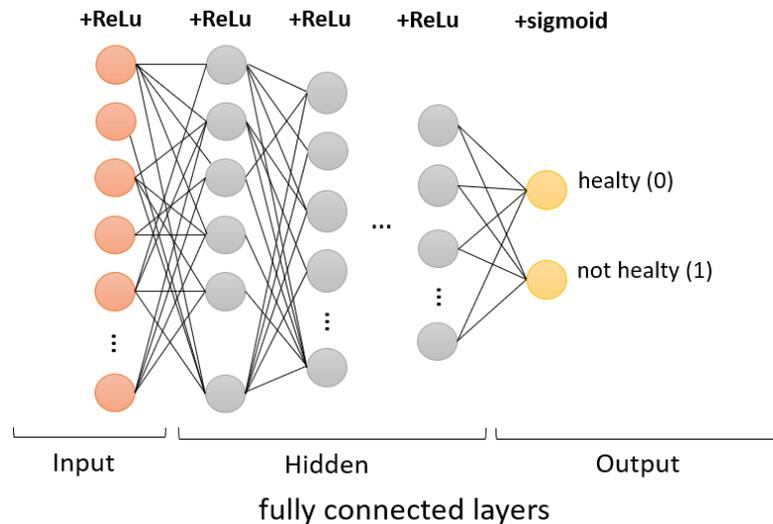


Figura 4.2: Exemplo representativo de uma arquitetura de uma *deep feed forward*.

É importante referir, que a definição final do modelo ocorreu após um intenso processo de otimização.

Compilação do modelo

Após a definição do modelo, foi necessário proceder à sua compilação. Para tal, foi necessário especificar alguns parâmetros adicionais antes de se iniciar o processo de treino e aprendizagem da rede. Destes parâmetros adicionais, destacam-se: a função de perda, o otimizador e as métricas adicionais para avaliar o desempenho do modelo. Para este modelo, a função de perda utilizada foi a perda logarítmica (perda por entropia cruzada) por se tratar de um problema de classificação. Como otimizador, utilizou-se um algoritmo de otimização de primeira ordem conhecido como *'adam'*, uma variante do gradiente estocástico descendente. Esta escolha deve-se ao facto de se tratar de um algoritmo computacionalmente eficiente e de fácil implementação [114]. Por último, pretendeu-se ainda reportar a *'accuracy'* da classificação do modelo após o processo de treino, como a principal métrica na avaliação do mesmo.

Treino do modelo

Tendo concluído as etapas anteriormente descritas, iniciou-se o treino do modelo. Para tal, foi necessário especificar o número de iterações no qual o modelo deve executar o processo de treino no conjunto de dados. Este número de interações é frequentemente designado por épocas e que foi especificado através do parâmetro *'epochs'*. Para além disto, definiu-se o número de

amostras de treino que são avaliadas antes que seja realizada uma nova atualização dos parâmetros internos da rede, com recurso ao parâmetro *'batch_size'* [115].

Na figura 4.3 observa-se o fluxograma que esquematiza todas as etapas descritas anteriormente, no que toca à fase de pré-processamento e o desenvolvimento do modelo, destacando as principais características da sua implementação.

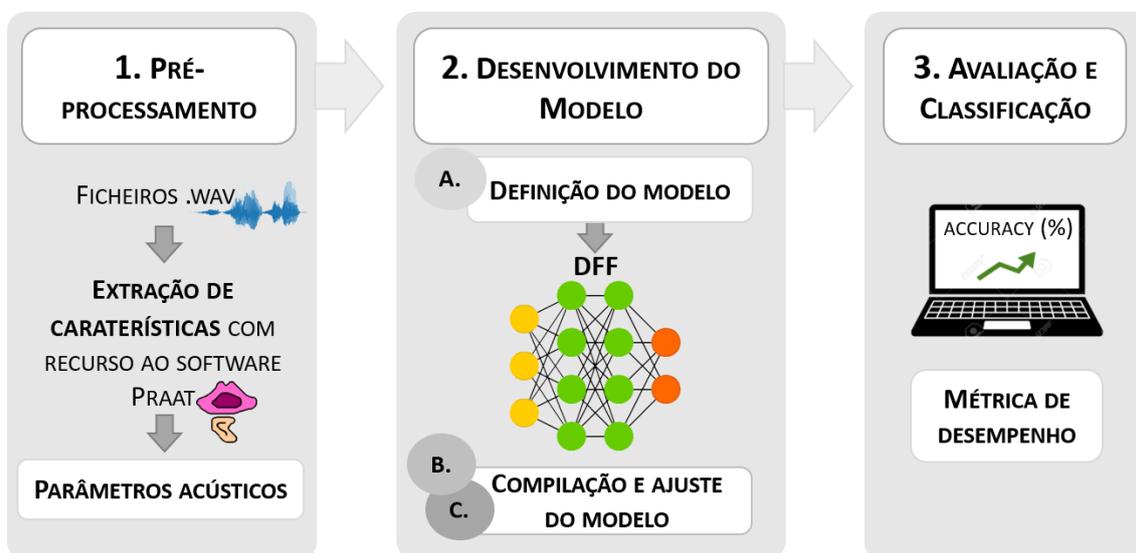


Figura 4.3: Fluxograma representativo das fases do desenvolvimento do modelo no domínio do tempo.

Após a computação eficiente do treino da rede foi possível proceder à avaliação do modelo. Porém, por se tratar de uma etapa de extrema importância será discutida detalhadamente a seguir, após apresentação e descrição do modelo no domínio das frequências.

4.2.2. Modelo no domínio das frequências

Uma alternativa ao modelo apresentado anteriormente, segue-se uma segunda abordagem da aplicação da análise acústica recorrendo ao domínio das frequências, mas mantendo o objetivo inicialmente proposto. Nesta abordagem pretende-se utilizar os coeficientes MFCC apresentados e discutidos na secção 2.2.2 do Capítulo 2. Assim, tal como no modelo anterior, são apresentados e discutidos os aspetos mais relevantes no que diz respeito às etapas de pré-processamento e do desenvolvimento do modelo DL em questão.

Pré-processamento e leitura da base de dados

Para o desenvolvimento deste modelo foi imprescindível a utilização da biblioteca Librosa, uma vez que esta disponibiliza os recursos para a extração dos MFCC de todos os ficheiros WAV recolhidos.

Inicialmente, realizou-se a importação de todas as bibliotecas necessárias para o desenvolvimento do respetivo modelo, tal como aconteceu com o modelo anteriormente apresentado.

De seguida, importaram-se todos os ficheiros WAV que se encontravam numa pasta específica através do seu diretório. Para além disto, especificaram-se os formatos das variáveis de entrada e das variáveis de saída para que fossem corretamente tratadas nos passos seguintes.

Através do método *enumerate()* procedeu-se à leitura e enumeração de todos os ficheiros WAV, para que fosse possível definir a variável de saída. Para tal, associou-se a cada ficheiro o número que corresponde ao valor da variável de saída através da sua identificação no nome de cada ficheiro utilizado. Por exemplo, quando um ficheiro de voz era recolhido de um indivíduo saudável, no final do nome deste ficheiro foi colocado o número 0, e caso se tratasse de um ficheiro recolhido de um indivíduo com diagnóstico de DAC era colocado o número 1. Este processo permitiu distinguir os diferentes ficheiros tendo em conta ao grupo ao qual pertenciam os participantes do estudo e ainda, pelas categorias do sistema CAD-RADS. Neste último cenário, o valor de cada categoria CAD-RADS foi de igual modo identificado em cada ficheiro. Por exemplo, caso se tratasse de um participante da categoria CAD-RADS 4, era colocado o número 4. Quando os ficheiros estavam a ser enumerados, foi utilizada uma técnica que permitiu identificar qual o número utilizado no nome de cada ficheiro e por sua vez, associá-lo à variável de saída. Assim, este processo permitiu que cada ficheiro de voz fosse devidamente identificado, utilizando-se esta codificação para se definir a variável de saída para o processo de treino e avaliação.

De seguida, como nem todos os ficheiros WAV recolhidos continham a mesma duração, foi necessário alterar estes ficheiros aplicando uma estratégia que permitiu selecionar a parte mais central do ficheiro áudio, utilizando a sua dimensão e a frequência de amostragem através da definição de uma fórmula. Posto isto, todos os ficheiros passaram a apresentar a mesma duração. De seguida, utilizando a biblioteca Librosa, procedeu-se à extração dos MFCC de cada ficheiro e associá-los às variáveis de entrada, recorrendo-se ao método *'feature.mfcc'*. Foi necessário especificar e identificar os ficheiros WAV, a frequência de amostragem e ainda, o número de coeficientes ceptrais pretendidos, sendo assim definida a variável de entrada X. Foram extraídos 16 coeficientes mel ceptrais, em janelas de duração 81ms com 1/3 de overlap, ao longo da duração de 2 segundos de cada ficheiro.

Por último e antes de se prosseguir com a fase de desenvolvimento do modelo, foi necessário reformular o conjunto de dados de entrada, para que estes estivessem adequados ao treino de uma rede convolucional. Uma vez que os dados de entrada são imagens, as camadas utilizadas nestas redes necessitam que os seus dados estejam configurados tendo em conta as características das imagens. Estas são expressas numa matriz tridimensional em que os primeiros eixos correspondem às dimensões das imagens utilizadas e ainda, à profundidade da cor (altura \times largura \times profundidade), sendo esta a configuração dos dados de entrada utilizados. Por último, normalizaram-se os valores para um intervalo entre 0 e 1.

Definição do modelo

Tal como aconteceu no modelo no domínio do tempo, construiu-se um modelo por camadas. No entanto, para este modelo optou-se por se utilizar uma arquitetura de rede diferente da usada no modelo anterior, uma vez que a topologia dos dados de entrada utilizada é também diferente. Utilizou-se assim uma rede CNN caracterizada por usar diferentes tipos de camadas como já foi discutido na secção *Deep Learning* do Capítulo 3. Assim, na construção deste modelo foram utilizadas camadas de convolução, de *pooling*, de *dropout* e ainda, do tipo *dense*. Para o seu desenvolvimento foi de extrema importância compreender e especificar de forma correta os argumentos necessários para cada tipo de camada, no qual destacamos os seguintes [116], [117]:

1. Nas camadas de convolução (*Conv2D*), foi necessário especificar os seguintes parâmetros: número de neurónios; tamanho do *kernel*, ou seja, o tamanho da matriz do filtro utilizado na camada de convolução; o formato dos dados de entrada, que geralmente se relaciona com a formato das imagens utilizadas; o argumento '*padding*' com o valor '*same*' que indica que podem ser adicionadas linhas e colunas de zeros, quantas forem necessárias para que a saída tenha a mesma dimensão que a entrada após o processo de convolução; e por último, a função de ativação.
2. Nas camadas de *pooling* (*MaxPooling2D*) foi apenas necessário especificar o tamanho do *pooling*.
3. As camadas de regularização (*Dropout*) foram configuradas para excluir aleatoriamente uma determinada percentagem de neurónios na camada e foram utilizadas com intuito de reduzir o *overfitting*².
4. As camadas de '*Flatten*' permitiram conectar as camadas de convolução às camadas '*dense*' também utilizadas.

Resumidamente, na figura 4.4 visualiza-se a arquitetura de uma rede CNN esquematizando todos estes conceitos através de um diagrama geral para uma melhor compreensão face à arquitetura destas redes.

Compilação do modelo

Na compilação do respetivo modelo, foram definidos de igual forma os parâmetros utilizados no modelo DL no domínio do tempo. Utilizou-se a mesma função de perda e o mesmo método de inicialização, como também se reportou a exatidão da classificação como a principal métrica para avaliação.

² Acontece quando a taxa de erro entre o treino e o teste do conjunto de dados é muito grande. Por outras palavras, o resultado obtido para o conjunto de dados de treino é muito bom, mas em novos dados o resultado obtido é bastante fraco [95].

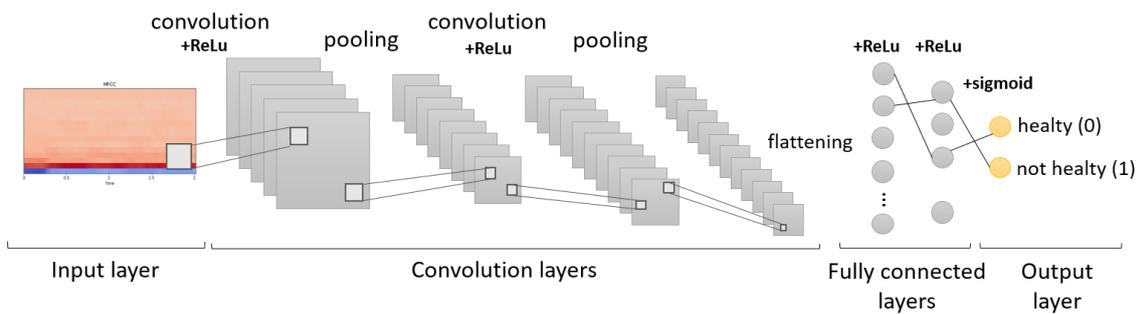


Figura 4.4: Exemplo representativo de uma arquitetura de rede neural convolucional.

Treino do modelo

Por último, iniciou-se o treino do modelo. Neste passo especificou-se o número de épocas e o *'batch_size'*.

Na figura 4.5 observa-se o fluxograma representativo de todos os passos necessários ao desenvolvimento deste modelo, destacando as características mais relevantes.

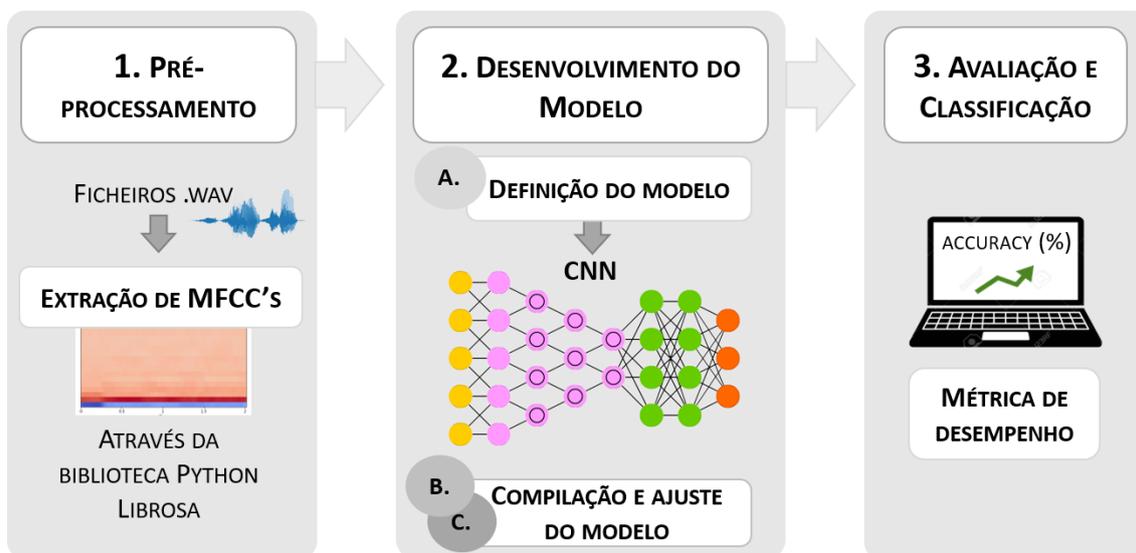


Figura 4.5: Fluxograma representativo das fases do desenvolvimento do modelo no domínio das frequências.

Neste momento, é possível comparar este fluxograma com o apresentado na figura 4.3 e identificar as principais diferenças no desenvolvimento de ambos os modelos nas duas fases anteriormente descritas e que se encontram esquematizadas em ambos os fluxogramas apresentados.

4.3. Método de Avaliação

A avaliação do desempenho obtido por ambos os modelos é considerada a parte mais crucial no desenvolvimento destes. Considera-se um bom modelo aquele que consegue prever de forma

precisa quando lhe são apresentados novos dados. Na verdade, em qualquer modelo DL só é possível calcular a sua eficácia após este ser testado com dados desconhecidos até ao momento em lhe são apresentados. Uma vez que o método de avaliação escolhido foi utilizado em ambos os modelos, optou-se por abordar em separado este último passo importante e crucial na implementação dos modelos desenvolvidos, como principal ferramenta de análise dos dados no presente trabalho.

Existem vários métodos que permitem estimar o desempenho destes modelos, no entanto por se tratar de um procedimento bastante popular, o método usado no desenvolvimento deste trabalho foi a validação cruzada *k-fold* (*k-fold cross validation*) [118]. No geral, o conceito da validação cruzada foca-se na divisão da base de dados utilizada em vários subconjuntos aproximadamente iguais. Posto isto, durante as várias iterações do processo de treino, um subconjunto é escolhido para ser o conjunto de dados utilizado para testar o modelo calculando-se uma medida de desempenho, enquanto que os restantes subconjuntos são usados para o treino. Este passo repete-se até que todos os subconjuntos tenham sido escolhidos para ser o conjunto de dados de teste, tal como se encontra esquematizado na figura 4.6. No final deste processo, a medida de desempenho resultante da validação cruzada *k-fold* trata-se da média dos valores calculados em cada iteração da medida de desempenho escolhida, sendo a exatidão geralmente a medida de desempenho mais utilizada nestes modelos. A utilização do *k-fold* faz variar os conjunto de teste e treino e as várias iterações obrigam a vários treinos. O resultado é uma rede mais robusta (os efeitos causados por mínimos locais atenuam-se ou evitam-se) e obtêm-se uma maior confiança nos resultados [119].

Recorrendo ao método da validação cruzada *k-fold* é necessário apenas especificar um único parâmetro designado por '*k*', e que representa o número de subgrupos para os quais uma base de dados ou conjunto de dados inicial deverá ser dividido. No exemplo representado e esquematizado na figura 4.6 é possível observar que o conjunto de dados utilizado foi dividido em 5 partes iguais realizando-se a validação cruzada 5 vezes, uma vez que o parâmetro '*k*' é igual a 5. Um forte argumento a favor da utilização desta ferramenta é o potencial de se usar todo o conjunto de dados em dados de teste, embora que este processo não ocorra de uma só vez [120].

Posto isto, em ambos os algoritmos desenvolvidos para este trabalho utilizou-se o '*StratifiedKFold*', uma variação do método de validação cruzada *k-fold* através da biblioteca *Scikit-Learn*, permitindo a cada subconjunto conter aproximadamente a mesma percentagem das amostras de cada classe, tal como acontece no conjunto de dados inicial [119]. Após escolhido o valor de *k*, foram criados *k* modelos sendo registados e armazenados as medidas de desempenho resultantes de cada um e no final, é obtida a média e o desvio padrão do conjunto de medidas de desempenho anteriormente registadas para se fornecer uma boa estimativa da *performance* do classificador.

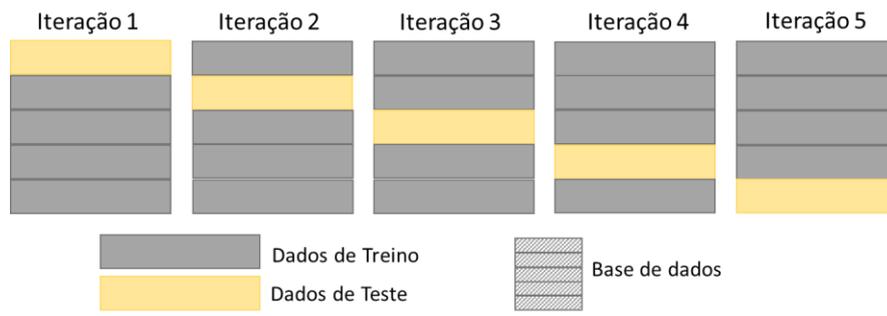


Figura 4.6: Esquematização do processo de validação cruzada k -fold estratificado com 5 interações ($k = 5$).

Resultados

Este capítulo inicia-se com uma análise descritiva da amostra final e as principais dificuldades no desenvolvimento da base de dados. Posteriormente são apresentados os resultados obtidos da classificação por ambos os algoritmos desenvolvidos e a respetiva discussão e comparação dos mesmos, tendo em conta os seguintes fatores: procedimento utilizado na recolha dos dados, a classificação CAD-RADS, o score de cálcio e ainda, o número de fatores de risco associados.

5.1. Desenvolvimento da Base de Dados

De entre os principais objetivos identificados do presente estudo, o objetivo de desenvolver uma base de dados contendo gravações de voz quer de indivíduos com patologia cardíaca já diagnosticada, quer de indivíduos saudáveis foi alcançado com sucesso. Como já referido anteriormente, a base de dados desenvolvida é constituída por todas as gravações de voz realizadas aos participantes e por um ficheiro CSV que reuniu todas as informações clínicas importantes de cada participante, nomeadamente: idade, sexo, grupo da amostra, classificação CAD-RADS, score de cálcio, número de vasos com doença obstrutiva e ainda, os principais fatores de risco associados.

Na tabela 5.1 é apresentada uma análise descritiva da amostra final face ao grupo que pertencem os participantes, sexo e a faixa etária. Em ambos os sexos, o número de indivíduos com diagnóstico de doença coronária na amostra deste estudo é significativamente mais elevado entre a faixa etária dos 50 e 59 anos, seguida da faixa etária dos 60 e 69 anos, em comparação com as restantes faixas etárias. De facto, este aspeto vai de encontro com o foi referenciado sobre o modelo SCORE no Capítulo 2, concretamente na secção 2.1.3. *Avaliação do Risco Cardiovascular*, em que suas tabelas para o cálculo do risco cardiovascular apresentam um maior destaque na faixa etária entre os 50 e os 65 anos, por se considerar este o período crítico na ocorrência de mudanças acentuadas no risco. Assim e com auxílio da figura 5.1, evidenciamos um maior número de indivíduos com diagnóstico de DAC nestas faixas etárias desta amostra, tal como seria de esperar.

Tabela 5.1: Caracterização geral dos participantes face ao grupo, faixa etária e sexo.

	Grupo em análise (patológico)			Grupo de controlo (saudável)		
	Faixa Etária (anos)	n	(%)	Faixa Etária (anos)	n	(%)
Masculino	18 – 29	0	0%	18 – 29	1	1,3%
	30 – 39	1	1,3%	30 – 39	2	2,6%
	40 – 49	4	5,3%	40 – 49	2	2,6%
	50 – 59	11	14,4%	50 – 59	2	2,6%
	60 – 69	7	9,2%	60 – 69	0	0%
	> 70	7	9,2%	> 70	0	0%
Feminino	18 – 29	0	0%	18 – 29	1	1,3%
	30 – 39	0	0%	30 – 39	1	1,3%
	40 – 49	2	2,6%	40 – 49	3	4%
	50 – 59	12	15,7%	50 – 59	4	5,3%
	60 – 69	7	9,2%	60 – 69	6	8%
	> 70	3	4%	> 70	0	0%
	Total	54	71%	Total	22	29%

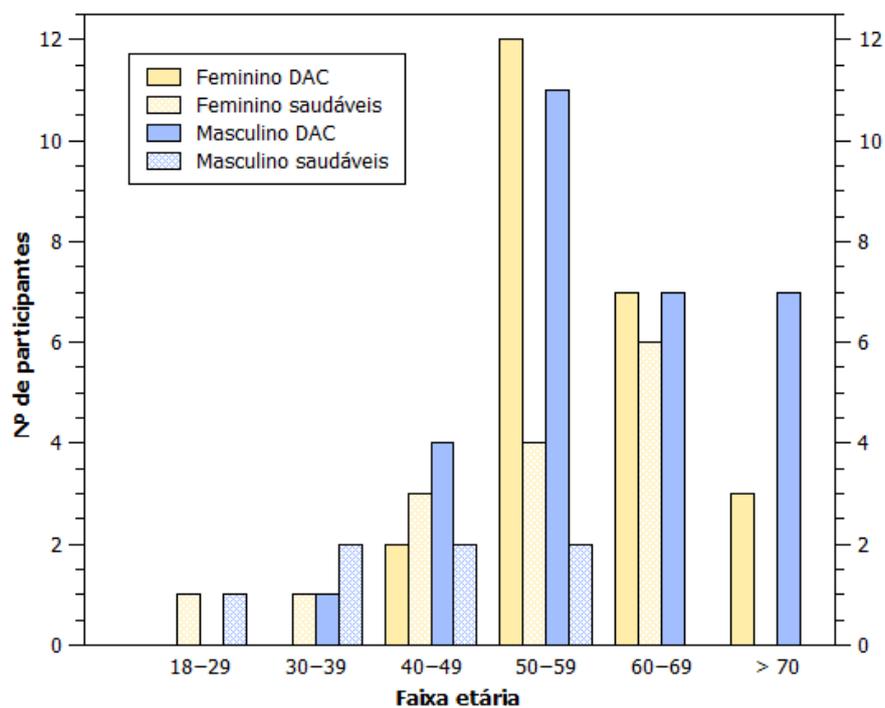


Figura 5.1: Descrição do grupo em análise e de controlo face ao sexo e faixa etária.

Relativamente ao sistema de classificação CAD-RADS, método clínico utilizado para quantificar a doença coronária aterosclerótica segundo a sua gravidade, evidencia-se na amostra um número acentuado de indivíduos na categoria 1, a menos grave, deste sistema com cerca de

30 participantes (60%). Assim sendo, mais de metade dos participantes que pertencem ao grupo com diagnóstico de doença coronária revelaram “estenose mínima não obstrutiva” no exame clínico realizado. Observando a figura 5.2 também se verifica um decréscimo exponencial no número de participantes à medida que gravidade da doença coronária se revela cada vez mais grave, especialmente no sexo masculino. É importante salientar que 4 indivíduos incluídos anteriormente no grupo com diagnóstico de doença coronária, não lhes foi atribuída nenhuma categoria do sistema de classificação CAD-RADS. Assim, optou-se por excluir estes participantes pela falta desta informação crucial, unicamente quando se analisar a amostra tendo em conta a gravidade da doença coronária com os algoritmos desenvolvidos. Não foram apresentados motivos para justificar esta ausência, podendo ter apenas ocorrido um lapso durante o preenchimento destas informações na base de dados desenvolvida.

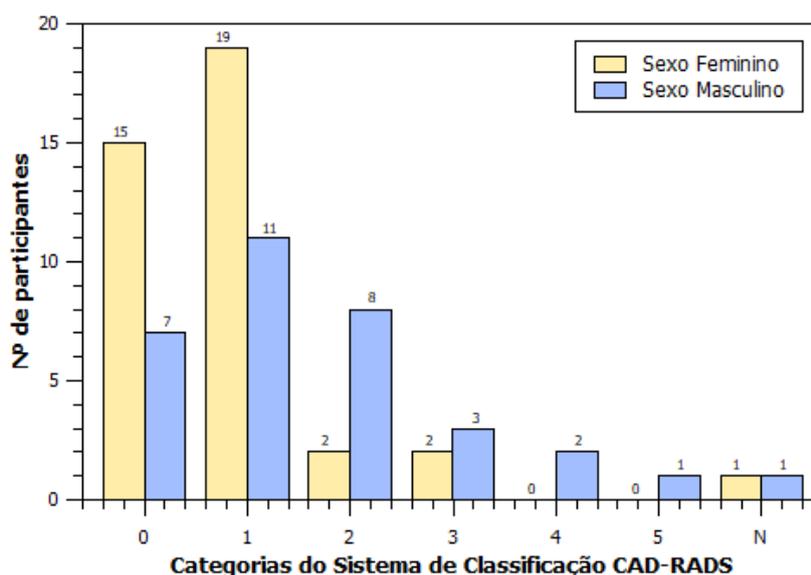


Figura 5.2: Amostra do estudo segundo o sistema de classificação CAD-RADS.

Para cada participante do estudo foram também recolhidos os valores relativamente ao cálculo do score de cálcio. De entre todos os participantes os valores variaram entre 0 e 1029, sendo que cerca de 45% dos participantes não apresentaram nenhum grau de calcificação coronária. Relembrando que a percentagem de indivíduos sem doença coronária da amostra é de 29%, evidencia-se assim que alguns dos participantes que foram diagnosticados com doença através do sistema de classificação CAD-RADS, não apresentaram qualquer sinal de calcificação coronária quando o score de cálcio foi calculado e analisado.

Para além destas informações clínicas, realizou-se também o levantamento de quais os fatores de risco associados a cada participante do estudo. Em relação aos fatores de risco, verificou-se que a hipertensão arterial foi o fator de risco mais frequente de entre todos os participantes (35,6%). Seguiu-se a dislipidemia com uma pequena diferença face ao anterior (30,69%) e o tabagismo (18,8%). Em menor frequência, como demonstra a figura 5.3, também

foram identificados participantes que apresentavam história familiar de doença coronária prematura e a diabetes *mellitus*.

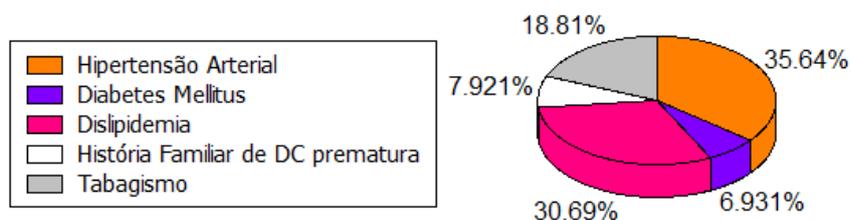


Figura 5.3: Fatores de risco associados aos participantes do estudo (em %).

No que toca à distribuição do número total de fatores de risco associados, curiosamente existe um maior número de participantes com diagnóstico de DAC que não apresentaram nenhum fator de risco em comparação com os participantes saudáveis. Ainda assim, e tal como seria de esperar, o grupo de indivíduos com diagnóstico de doença coronária apresentam mais fatores de risco associados comparativamente ao grupo de controlo.

Por último, é importante evidenciar alguns problemas imprevisíveis que surgiram no decorrer do desenvolvimento da base de dados, em particular com as gravações dos ficheiros áudio. Destacamos assim, os dois principais problemas que condicionaram a base de dados:

1. Realização das gravações de todos os participantes do estudo no mesmo local e em silêncio tornou-se uma tarefa difícil e por vezes, até impossível de se alcançar. Em alguns casos ($n = 4$), analisando posteriormente as gravações realizadas detetou-se algum ruído de fundo, impossibilitando a utilização destes ficheiros para que os resultados não fossem assim comprometidos. Somente foram afetados os ficheiros em que os referidos participantes realizavam a leitura dos dois textos pretendidos no procedimento da recolha de dados.
2. Por outro lado, em alguns dos participantes ($n = 7$) quando lhes era apresentado o primeiro texto para se iniciar a gravação da sua leitura, esta era impedida por não conseguirem ler sem o auxílio dos seus óculos. É importante referir que estes participantes iniciaram este processo porque cumpriam com os critérios de inclusão do estudo, no entanto sem ajuda deste acessório o procedimento da leitura não se realizou.

Com base nesta análise realizada à base de dados é possível identificarem-se os principais problemas e limitações quanto ao seu desenvolvimento. Adicionalmente, fornece um feedback quanto aos aspetos a melhorar no desenvolvimento da referida base de dados para trabalhos futuros.

5.2. Resultados obtidos

Posto isto, estão reunidas as condições para a aplicação dos dados recolhidos, e que constituem a base de dados analisada anteriormente, aos modelos DL. De salientar, que foram desenvolvidos dois modelos DL que realizam a análise acústica com base nas diferentes abordagens apresentadas anteriormente:

1. **Modelo 1 (M1)**, que corresponde ao modelo que realiza a análise acústica no domínio do tempo;
2. **Modelo 2 (M2)**, que corresponde ao modelo no realiza a análise acústica no domínio das frequências.

A figura 5.4 encontra-se um diagrama que esquematiza todas as variáveis consideradas por ambos os modelos e que produzem os resultados que permitem inferir sobre a sua qualidade no que toca à classificação dos participantes.

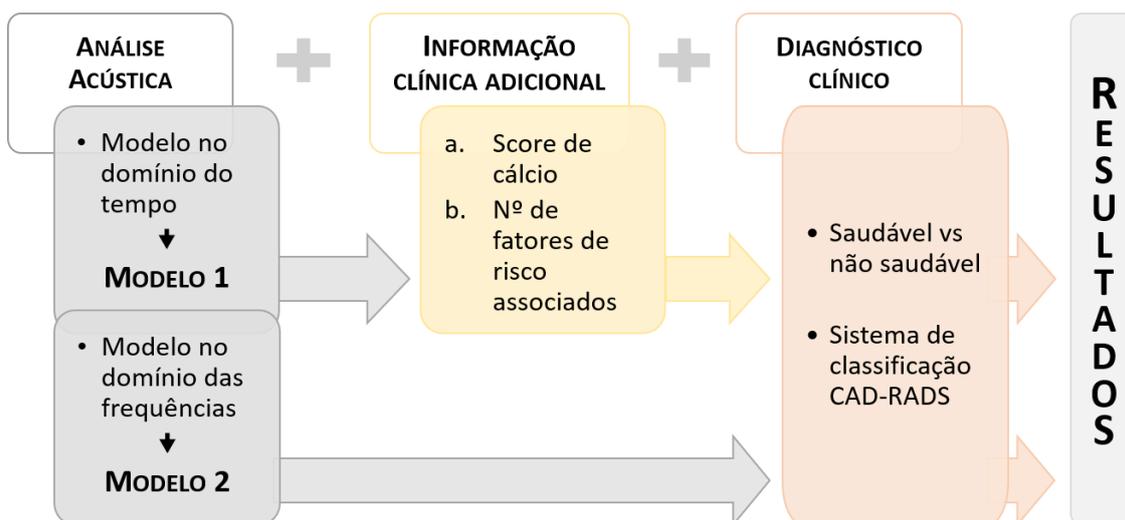


Figura 5.4: Esquema das principais variáveis utilizadas em cada modelo de *deep learning* desenvolvido.

Analisando o esquema verifica-se que com o modelo 1 é possível utilizar outras informações clínicas que se encontram presentes na base de dados, uma vez que os valores obtidos da análise acústica são tabelados num ficheiro CSV, podendo estas informações serem facilmente incorporadas através da adição de novas colunas neste ficheiro com as novas informações. Em contrapartida, não é possível adicionar estas mesmas informações no modelo 2 devido à topologia utilizada pelas variáveis de entrada deste modelo. Isto porque, a partir da extração dos coeficientes MFCC, são produzidas imagens que são expressas em matrizes tridimensionais, sendo estas utilizadas como dados de entrada para o modelo 2. Desta forma, é impossível adicionar novas informações, como por exemplo os valores obtidos pelo score de cálcio, sem comprometer a informação contida nas matrizes.

Com base nisto, e também com intuito de identificar se as informações clínicas adicionais presentes na base de dados ajudam a melhorar a exatidão da classificação dos modelos, foram criadas variações do modelo 1 que incluíram estas informações para além dos valores dos parâmetros acústicos e da classificação dos participantes. Assim, para além do modelo 1 identificado como ‘M1’, destacam-se as seguintes variações deste modelo:

- **M1_CA**: para além da análise acústica, contém os valores de score de cálcio como variável de entrada e o diagnóstico dos participantes como variável de saída.
- **M1_FR**: para além da análise acústica e do diagnóstico, contém o número de fatores de risco associados a cada participante.
- **M1_CA+FR**: contém os valores da análise acústica, o diagnóstico e todas as informações clínicas, ou seja, os valores do score de cálcio e o número de fatores de risco.

Em suma, a tabela 5.2 encontra-se um resumo das informações utilizadas nas variáveis de entrada de cada modelo apresentado, para uma melhor compreensão das diferenças entre os modelos avaliados neste estudo.

Tabela 5.2: Resumo das informações utilizadas como variáveis de entrada em cada modelo desenvolvido.

Modelo	Análise Acústica / Extração de características		Análise Clínica	
	Parâmetros Acústicos	MFCC	Score de cálcio	Fatores de risco
M1	×	-	-	-
M1_CA	×	-	×	-
M1_FR	×	-	-	×
M1_CA+FR	×	-	×	×
M2	-	×	-	-

A análise dos resultados obtidos por parte dos modelos DL foca-se principalmente no valor da métrica de desempenho ‘*accuracy*’. Como já foi mencionado, com a utilização do método de validação cruzada *k-fold* para a avaliação dos modelos, o conjunto de dados é dividido em *k* conjuntos sendo criados *k* modelos que utilizam sempre um conjunto diferente para ser o conjunto de dados de teste. Durante este processo e após a finalização do treino de cada modelo criado, os valores da ‘*accuracy*’ são recolhidos, sendo no final deste processo calculada a média desses valores. Esta média calculada corresponde ao valor percentual tido em consideração nos resultados de ambos os modelos DL desenvolvidos. Quanto maior for o seu valor, melhor será o modelo DL em análise. Neste caso quer a média, quer o desvio padrão calculados no final deste processo fornecem uma boa estimativa em relação à ‘*accuracy*’ da classificação, como também no desempenho do modelo. Adicionalmente, reportou-se a métrica ‘*F-measure*’. Esta permite-nos obter num só valor a combinação de duas outras importantes métricas de desempenho para a

avaliação de cada modelo: a precisão e o *recall*. De igual forma, pretende-se que o valor da média obtida face aos valores ‘*F-measure*’ de todos os *k* modelos criados seja próximo de 100% tanto quanto possível, que significa que o modelo avaliado obteve um ótimo desempenho quanto ao problema de classificação proposto.

De seguida, estão descritos os resultados obtidos após aplicação dos modelos DL nos dados recolhidos, tendo em conta os diferentes procedimentos adotados no momento da recolha dos dados. Para além disto, os resultados encontram-se organizados em três etapas: numa primeira fase, segundo o problema de classificação binária, em que os participantes são classificados pela ausência (0) e diagnóstico de doença coronária (1); segundo o problema de classificação multiclasse, tendo em conta o sistema de classificação CAD-RADS em que os participantes são classificados com base na gravidade da doença coronária; e por último, na avaliação individual dos diferentes graus da doença coronária vs indivíduos saudáveis. Nesta última fase, todas as categorias são avaliadas individualmente à exceção dos graus mais graves que serão analisados em conjunto, uma vez que as amostras nestas categorias são bastante reduzidas comparativamente às categorias menos graves. Isto permite averiguar de forma mais pormenorizada a qualidade dos modelos desenvolvidos tendo em conta os diversos graus da doença coronária.

5.2.1. Dados recolhidos na emissão das vogais sustentadas

Os primeiros resultados apresentados utilizam os dados recolhidos a partir das gravações das emissões sustentadas das vogais /a/, /i/ e /u/ dos participantes da amostra, sendo aplicados aos modelos desenvolvidos.

Problema de classificação binária

Após um intenso e demorado processo de otimização dos parâmetros de cada modelo na tentativa de maximizar o valor da média da ‘*accuracy*’ da classificação e minimizar o desvio padrão associado foram obtidos os resultados apresentados na tabela 5.3.

Tabela 5.3: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação binária.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	70,13%	($\pm 5,08\%$)	80,00%	($\pm 3,00\%$)
M1_CA	150 / 5	83,71%	($\pm 3,03\%$)	87,23%	($\pm 2,35\%$)
M1_FR	150 / -	75,56%	($\pm 2,73\%$)	82,98%	($\pm 1,66\%$)
M1_CA+FR	200 / 5	85,08%	($\pm 3,00\%$)	88,27%	($\pm 2,67\%$)
M2	75 / 5	70,27%	($\pm 0,74\%$)	82,54%	($\pm 0,51\%$)

Para além da métrica ‘accuracy’ também são apresentados os valores da média e do desvio padrão da métrica ‘F-measure’. No geral, quase todos os modelos alcançaram um desempenho razoavelmente bom relativamente ao problema de classificação proposto. Verifica-se que o modelo que obteve melhores resultados foi o M1_CA+FR, seguindo-se o M1_CA com uma diferença mínima entre ambos. De realçar-se que estas duas redes incorporaram os valores do score de cálculo como variável de entrada, além dos valores da análise acústica. Analisando os resultados obtidos com base apenas na realização da análise acústica, o modelo 1 (M1) e o modelo 2 (M2), o segundo obteve uma ligeira melhoria face ao valor do desvio padrão, quer com a métrica ‘accuracy’ (0,74% vs 5,08%), quer com a métrica ‘F-measure’ (0,51 vs 3,00).

Adicionalmente, é possível avaliar graficamente o comportamento ou desempenho de cada modelo durante os processos de treino quanto à exatidão da classificação através da figura 5.5.

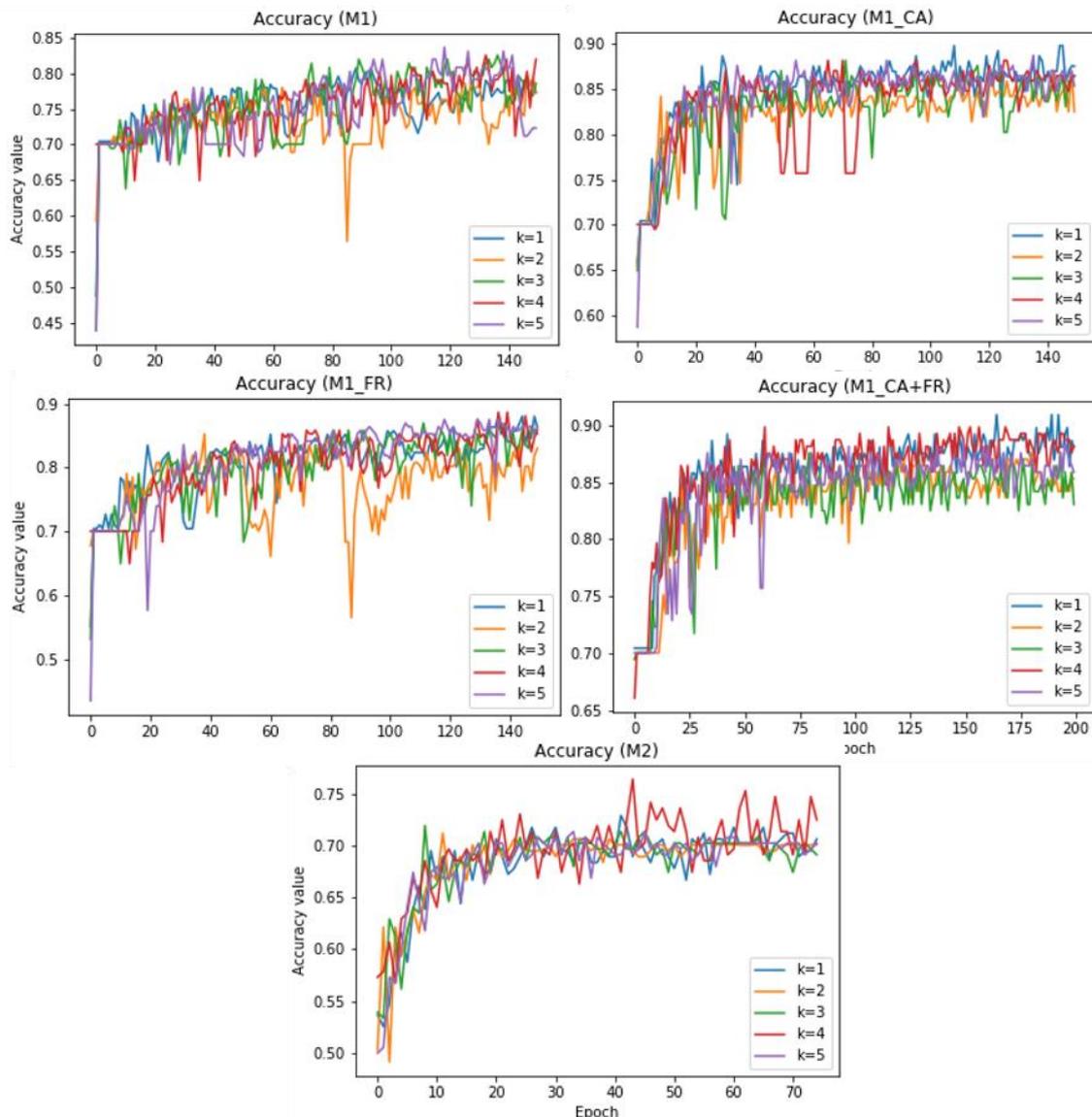


Figura 5.5: Desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação binária.

Na figura 5.5 encontram-se os gráficos com os processos de treino de cada modelo, apresentando cada um deles cinco curvas diferentes, uma vez que o conjunto de dados inicial foi dividido em cinco conjuntos com base no método de validação cruzada *k-fold* estratificado (onde $k=5$). Visualiza-se que todos os modelos conseguiram aprender o problema durante as primeiras 30 épocas. Inclusivamente, verifica-se que a ‘*accuracy*’ é mais estável nos modelos M1_CA, M_CA+FR e em particular, com o modelo M2 em que as curvas representadas dos processos de treino de cada k modelo criado são definitivamente mais suaves.

Ainda antes de se prosseguir com a análise dos resultados obtidos tendo em conta novos cenários para a classificação, tal como mencionado, é também importante investigar qual a influência do fator tabagismo nas cordas vocais na amostra avaliada através dos resultados apresentados na tabela 5.3. Sabe-se que estas estruturas são fortemente afetadas pelo estilo de vida, especialmente pelo tabaco [47]. Mas o tabagismo é um importante e prevalente fator de risco na nossa população e portanto, a exclusão de participantes fumadores no estudo levaria à exclusão de uma significativa parte da população, quer com diagnóstico, quer com elevado risco de desenvolver a doença coronária. Ainda assim, a influência do tabaco nas cordas vocais não deve ser totalmente ignorada e por este motivo, optou-se pela realização de um teste adicional com intuito de confirmar se os resultados apresentados na tabela 5.3, resultam da influência do tabaco nas cordas vocais ou, se de facto estes comprovam a existência de uma associação entre a voz e a doença coronária. Para tal, uma nova amostra foi constituída exclusivamente por indivíduos não fumadores de ambas as classes em estudo. Após exclusão dos participantes fumadores, o novo conjunto de dados inicial foi testado nos dois modelos DL que utilizam apenas a informação extraída pela análise acústica nas variáveis de entrada. Os resultados obtidos, mais precisamente pelos modelos M1 e M2, encontram-se descritos na tabela 5.4.

Tabela 5.4: Resultados obtidos pelos modelos com os dados recolhidos na emissão das vogais sustentadas e sem os participantes fumadores, no cenário de problema de classificação binária.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	69,09%	($\pm 3,53\%$)	79,29%	($\pm 3,15\%$)
M2	75 / 5	69,82%	($\pm 1,10\%$)	82,22%	($\pm 0,76\%$)

Através da comparação dos resultados entre as duas tabelas, identifica-se uma diferença de 1% no valor da média em ambos os modelos, quer da métrica ‘*accuracy*’, quer da métrica ‘*F-Measure*’. Para além disto, verifica-se que o desvio-padrão de ambos os modelos também não se alterou significativamente. Com base nisto, podemos afirmar que a influência do tabaco nas cordas vocais é mínima e que os resultados satisfatórios obtidos pelos classificadores se relacionam com as características da DAC.

Problema de classificação multiclases

De igual forma, após um longo processo de otimização são apresentados os resultados obtidos pelos modelos, no entanto pretende-se que classificação seja realizada com base no sistema de classificação CAD-RADS. Desta forma, é possível avaliar o desempenho dos mesmos modelos anteriormente analisados tendo em conta os diferentes graus da gravidade da doença coronária. Na tabela 5.5 encontram-se os resultados obtidos por parte dos modelos para o problema de classificação multiclasse proposto.

Tabela 5.5: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclases.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	250 / 10	38,84%	($\pm 2,70\%$)	11,44%	($\pm 4,64\%$)
M1_CA	250 / 10	62,62%	($\pm 4,20\%$)	55,51%	($\pm 4,84\%$)
M1_FR	300 / 10	49,07%	($\pm 5,29\%$)	20,24%	($\pm 7,43\%$)
M1_CA+FR	300 / 10	64,07%	($\pm 6,85\%$)	53,47%	($\pm 10,89\%$)
M2	75 / 5	42,02%	($\pm 3,45\%$)	0%	($\pm 0\%$)

No geral, os resultados obtidos foram insatisfatórios em todos os modelos comparativamente aos apresentados anteriormente. Inclusivamente, os piores valores foram registados pelo modelo M1 e M2, com base nos valores obtidos por ambas as métricas. Revela-se assim ser impossível obter uma classificação viável com os modelos DL, quando a amostra enquadrava os diferentes graus de gravidade da doença coronária. Ainda assim, verifica-se alguma concordância nos resultados obtidos com os anteriores. Isto porque os modelos onde se verificaram os melhores valores na média da exatidão na classificação descritos na tabela 5.5 (64% e 63% aproximadamente) continham os valores do score de cálculo para além da análise acústica. Observando as curvas de aprendizagem na figura 5.6 para se avaliar de uma forma melhor o desempenho de todos os modelos, identifica-se uma maior dificuldade no processo de aprendizagem sendo utilizado um maior número de épocas em quase todos os modelos.

É importante reforçar que neste problema de classificação multiclasse identifica-se um desequilíbrio bastante mais acentuado face ao problema de classificação binária anteriormente analisado. Os resultados apresentados na tabela 5.5 foram obtidos pelos modelos com base neste desequilíbrio, uma vez que os modelos criados através do método de validação cruzada *k-fold* contêm aproximadamente a mesma percentagem de exemplos das amostras de cada classe do conjunto de dados inicial.

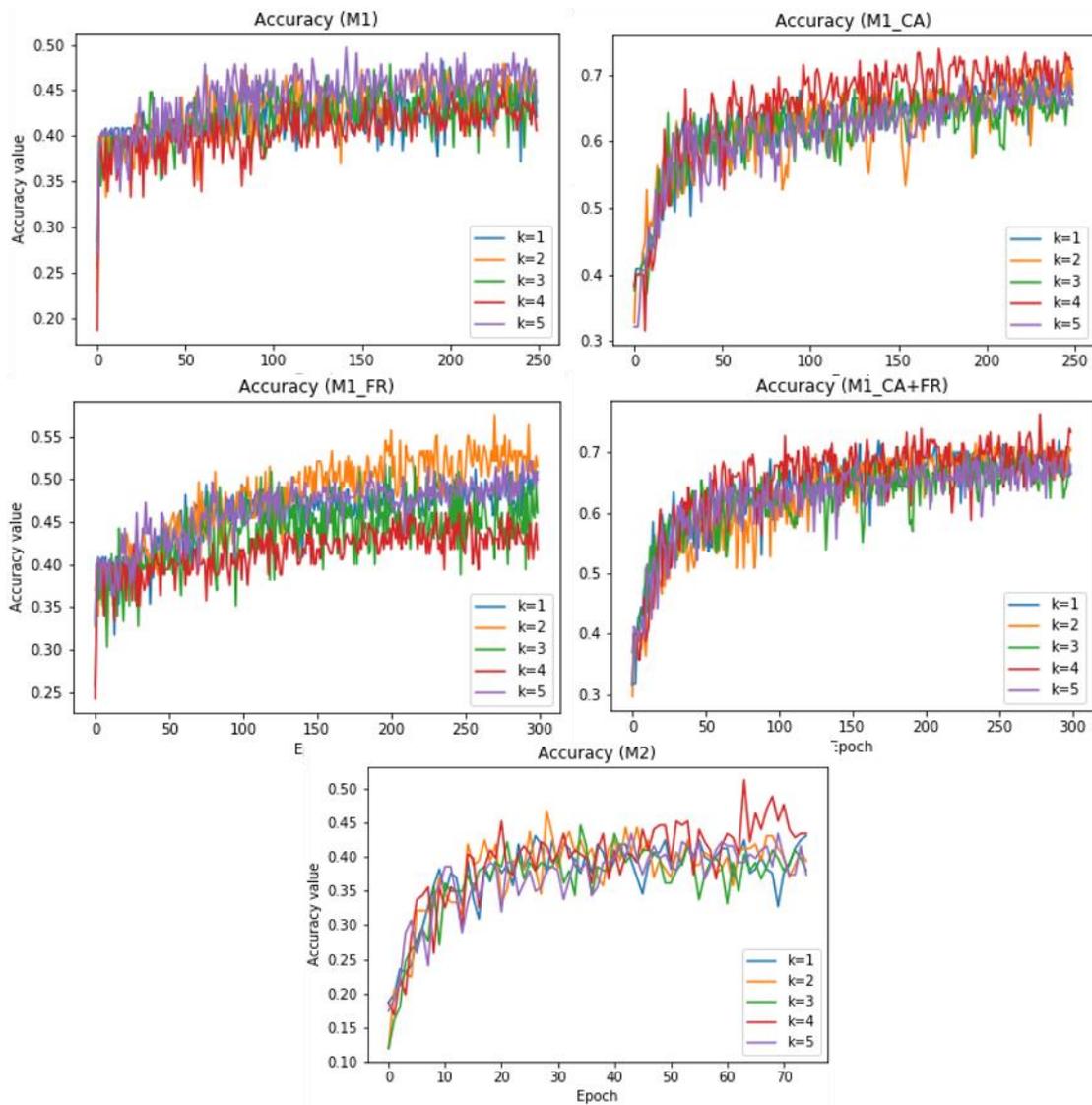


Figura 5.6: Desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclases.

No entanto, verifica-se um severo desequilíbrio principalmente com as classes 4 e 5: classe 0 apresenta uma percentagem de 32%; classe 1 de 40%; classe 2 de 15%; classe 3 de 7%; e a classe 4 e 5 ambas com uma percentagem de 3% dos dados. Foram assim aplicadas técnicas de reamostragem para compensar este desequilíbrio e comparar os seus resultados relativamente aos apresentados na tabela 5.5. Relativamente à métrica ‘accuracy’, o modelo M1 apresentou uma média de 30,63% ($\pm 6,79\%$), M1_CA cerca de 56,31% ($\pm 2,70$), M1_FR apresentou 40,30% ($\pm 2,62\%$) e o modelo M1_CA+FR cerca de 65,02% ($\pm 8,63\%$) com a aplicação da técnica de reamostragem BorderlineSMOTE. Para o modelo M2 utilizou-se uma outra abordagem, no qual foi realizado o cálculo dos pesos de cada classe obtendo-se uma média de 38,66% ($\pm 2,41$) face à exatidão da classificação. No geral, visualiza-se um pequeno decréscimo em quase todos os valores.

Uma vez que os resultados obtidos após aplicação de técnicas de reamostragem também revelaram ser insatisfatórios, e por se tratar de um problema de classificação multiclases, recorreu-se a uma técnica de binarização conhecida por ‘*OneVsRest*’ na tentativa de se alcançarem melhores resultados. A tabela 5.6 encontram-se os resultados obtidos por parte dos modelos DL através da utilização de um classificador que aplica a referida estratégia OvR

Tabela 5.6: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na emissão das vogais sustentadas, no cenário de problema de classificação multiclases.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	40,78%	($\pm 4,19\%$)	34,17%	($\pm 3,17\%$)
M1_CA	150 / 5	67,48%	($\pm 5,44\%$)	64,65%	($\pm 6,49\%$)
M1_FR	150 / -	46,62%	($\pm 3,77\%$)	41,56%	($\pm 5,05\%$)
M1_CA+FR	200 / 5	69,41%	($\pm 4,58\%$)	67,97%	($\pm 4,95\%$)
M2	75 / 5	37,22%	($\pm 3,51\%$)	25,66%	($\pm 3,36\%$)

Comparando os resultados obtidos com os da tabela 5.5, verifica-se uma melhoria nos valores de ambas as métricas avaliadas, exceto no valor da ‘*accuracy*’ do modelo M2. Ainda assim, os resultados são insatisfatórios, à exceção dos modelos M1_CA e M1_CA+FR em que os valores da média e do desvio padrão já demonstram um desempenho razoável na classificação. O elevado número de classes usado pode justificar os resultados obtidos, uma vez que a utilização de um número grande de classes pode condicionar a utilização da estratégia OvR

Categoria CAD-RADS 0 vs CAD-RADS 1

Na tabela 5.7 estão descritos os resultados do modelo 1, a variação do modelo 1 (M1_CA+FR) que incorpora todas informações clínicas adicionais e o modelo 2 numa amostra que inclui apenas os indivíduos sem doença e os indivíduos da categoria CAD-RADS 1.

Tabela 5.7: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 1.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	200 / 5	56,32%	($\pm 5,22\%$)	66,22%	($\pm 2,60\%$)
M1_CA+FR	200 / 10	83,22%	($\pm 2,12\%$)	82,67%	($\pm 2,67\%$)
M2	75 / 5	56%	($\pm 3,27\%$)	66,82%	($\pm 6,03\%$)

Uma vez mais, os resultados demonstram que o modelo M1_CA+FR obteve o melhor resultado com uma média de 83% quer na exatidão da classificação, quer na métrica ‘*F-measure*’. Os restantes apresentaram resultados menos satisfatórios.

Categoria CAD-RADS 0 vs CAD-RADS 2

Por sua vez, na tabela 5.8 estão descritos os resultados da média e desvio padrão obtidos por cada modelo numa amostra representada por indivíduos pertencentes as categorias CAD-RADS 0 e 2. Ou seja, pretende-se avaliar a capacidade de os modelos distinguirem na amostra os indivíduos sem diagnóstico de doença coronária e aos indivíduos diagnosticados com DAC não obstrutiva leve.

No que toca à exatidão da classificação, visualiza-se uma ligeira melhoria no valor da média em todos os modelos comparativamente aos resultados obtidos com amostra anterior (CAD-RADS 0 vs CAD-RADS 1). Destaca-se em particular, uma melhoria significativa com o modelo M1 que obteve um resultado razoável na classificação de indivíduos com DAC não obstrutiva leve com uma ‘*accuracy*’ de aproximadamente 74%, em contraste com o que se verificou na classificação de indivíduos com DAC não obstrutiva mínima. No entanto, o valor da métrica ‘*F-measure*’ obteve resultados pouco satisfatórios nos modelos M1 e M2.

Tabela 5.8: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 2.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	73,89%	($\pm 4,96\%$)	49,31%	($\pm 13,08\%$)
M1_CA+FR	200 / 5	88,53%	($\pm 3,97\%$)	76,85%	($\pm 9,22\%$)
M2	75 / 5	66,63%	($\pm 4,41\%$)	17,36%	($\pm 21,33\%$)

Categoria CAD-RADS 0 vs CAD-RADS +3

Por fim, são apresentados os resultados obtidos (tabela 5.9) por parte dos modelos com base numa amostra constituída por participantes sem diagnóstico de DAC e participantes que foram diagnosticados com os graus mais elevados da doença coronária. Estes resultados indicam uma melhoria no valor da média em quase todos os modelos. O modelo M1_CA+FR que engloba também as informações do score de cálcio e o número de fatores de risco associados, regista novamente o melhor resultado de entre todos os modelos comparativamente aos resultados obtidos com as amostras anteriores. Face aos restantes, existe uma melhoria quer na média, quer no desvio padrão de ambas as métricas avaliadas, mas esta diferença é menos significativa.

Tabela 5.9: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS +3.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	70,99%	($\pm 2,17\%$)	57,19%	($\pm 5,97\%$)
M1_CA+FR	150 / 15	93,57%	($\pm 3,91\%$)	91,99%	($\pm 4,34\%$)
M2	75 / 5	70,99%	($\pm 2,17\%$)	59,71%	($\pm 2,30\%$)

No entanto, este problema de classificação apresenta um ligeiro desequilíbrio: classe 0 apresenta uma percentagem de 71%; classe 3 de 16%; e as classes 4 e N de 6,5%. De igual forma, para confirmar se os dados apresentados na tabela 5.9 são ou não demasiado otimistas, aplicaram-se técnicas de reamostragem aos conjuntos de dados. Com o classificador *Balanced Bagging*, o modelo M1 apresentou uma média de 62,28% ($\pm 7,28\%$) e M1_CA+FR apresentou 86,08% ($\pm 6,20\%$) quanto à ‘accuracy’. Também o modelo M1_CA+FR apresentou cerca de 86,14% ($\pm 9,09\%$) com a utilização da técnica BorderlineSMOTE, e o modelo M2 obteve uma média de 68,77% ($\pm 7,28\%$). Comparando os valores que se encontram na tabela 5.9, visualiza-se um pequeno decréscimo nos valores da média e por outro lado, um aumento no valor do desvio padrão associado à métrica exatidão.

Por se tratar de um problema de classificação multiclases, recorreu-se à estratégia OvR sendo registados os valores obtidos pelos modelos DL na tabela 5.10. Verifica-se que os desempenhos obtidos pelos modelos foram bastante próximos aos obtidos sem aplicação da estratégia OvR apresentados na tabela 5.9.

Tabela 5.10: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na emissão das vogais sustentadas, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS +3.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	150 / -	70,99%	($\pm 2,17\%$)	58,97%	($\pm 2,85\%$)
M1_CA+FR	150 / 15	92,51%	($\pm 2,49\%$)	90,49%	($\pm 1,95\%$)
M2	75 / 5	70,99 %	($\pm 2,17\%$)	58,97%	($\pm 2,85\%$)

Por último, é importante analisar graficamente e em conjunto os desempenhos dos processos de treino de todos os modelos para cada amostra que analisa individualmente as categorias do sistema de classificação CAD-RADS através da Figura D.1 em anexo. Verifica-se que à medida que a gravidade da doença coronária na amostra avança, a exatidão da classificação melhora com

curvas de aprendizagem de treino que demonstram sinais de convergência também cada vez melhores. Para além deste aspeto, é visível a diferença quanto à irregularidade das curvas quando as amostra incluem indivíduos com DAC não obstrutiva mínima comparativamente às amostra que incluem indivíduos que apresentam estenoses ou até mesmo, oclusões coronárias totais.

5.2.2. Dados recolhidos na leitura dos textos foneticamente balanceados

Nesta fase são apresentados os resultados obtidos pelos modelos DL utilizando os dados recolhidos quer na leitura do texto foneticamente balanceado que exprime emoções positivas, quer a leitura do texto que exprime emoções negativas. De igual forma, pretende-se estudar as mesmas variáveis e modelos DL que foram analisados com os dados anteriores.

Problema de classificação binária

Na tabela 5.11 encontram-se descritos os resultados obtidos após aplicação dos modelos para o problema de classificação binária proposto com os dados recolhidos na leitura do texto que expressa emoções positivas e negativas.

Tabela 5.11: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, no cenário de problema de classificação binária.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	75 / -	71,05%	($\pm 3,48\%$)	79,39%	($\pm 5,94\%$)
M1_CA	95 / 5	84,40%	($\pm 5,38\%$)	88,24%	($\pm 4,03\%$)
M1_FR	115 / -	71,14%	($\pm 4,46\%$)	82,25%	($\pm 1,77\%$)
M1_CA+FR	105 / 5	88,34%	($\pm 4,77\%$)	90,67%	($\pm 4,23\%$)

Uma vez mais, todos os modelos DL alcançaram um desempenho relativamente bom face ao problema de classificação binária proposto. Evidencia-se o modelo M1_CA+FR com uma média de aproximadamente 88,34% quanto à ‘accuracy’ da classificação, sendo que este modelo contém todas as informações clínicas, além dos valores da análise acústica. Destaca-se também o resultado bastante positivo obtido pelo modelo M1, quer no valor da média, quer no valor do desvio padrão de ambas as métricas avaliadas, indicando bons resultados na classificação da doença coronária com base apenas na análise acústica realizada. Analisando o desempenho obtido por cada modelo no que diz respeito à ‘accuracy’ da classificação em relação aos seus processos de treino, visualiza-se na figura 5.7 que todos os modelos aprenderam razoavelmente bem o problema, mostrando um bom comportamento e sinais de convergência.

Adicionalmente, e tal como aconteceu com os dados recolhidos na emissão das vogais sustentadas, também se realizou um teste adicional neste conjunto de dados para avaliar a

influência do tabaco nas cordas vocais. Uma nova amostra foi constituída após exclusão de todos os participantes fumadores, sendo testada pelo modelo M1 que utiliza apenas as informações extraídas da análise acústica nas variáveis de entrada. Os resultados demonstraram uma diferença no valor da média de cerca de 10% quanto à métrica ‘accuracy’ ($60,32 \pm 15,30\%$) e de 20% quanto à métrica ‘F-measure’ ($59,07 \pm 31,08\%$). Adicionalmente, os valores do desvio-padrão também foram significativamente mais elevados. Este aumento significativo do desvio-padrão revela que durante o processo de validação cruzada *k-fold* estratificado, registaram-se *k* modelos com uma ótima classificação, mas em contrapartida *k* modelos obtiveram uma péssima classificação e consequentemente, os valores da média quer da ‘accuracy’, quer do ‘F-measure’ baixaram. Com base nisto, é possível que estes resultados tenham sido afetados pelo reduzido número total de dados inicial da amostra comparativamente aos dados recolhidos no procedimento anteriormente analisado, em que o número de dados analisados é bastante superior e onde apenas se registou uma diferença de 1%.

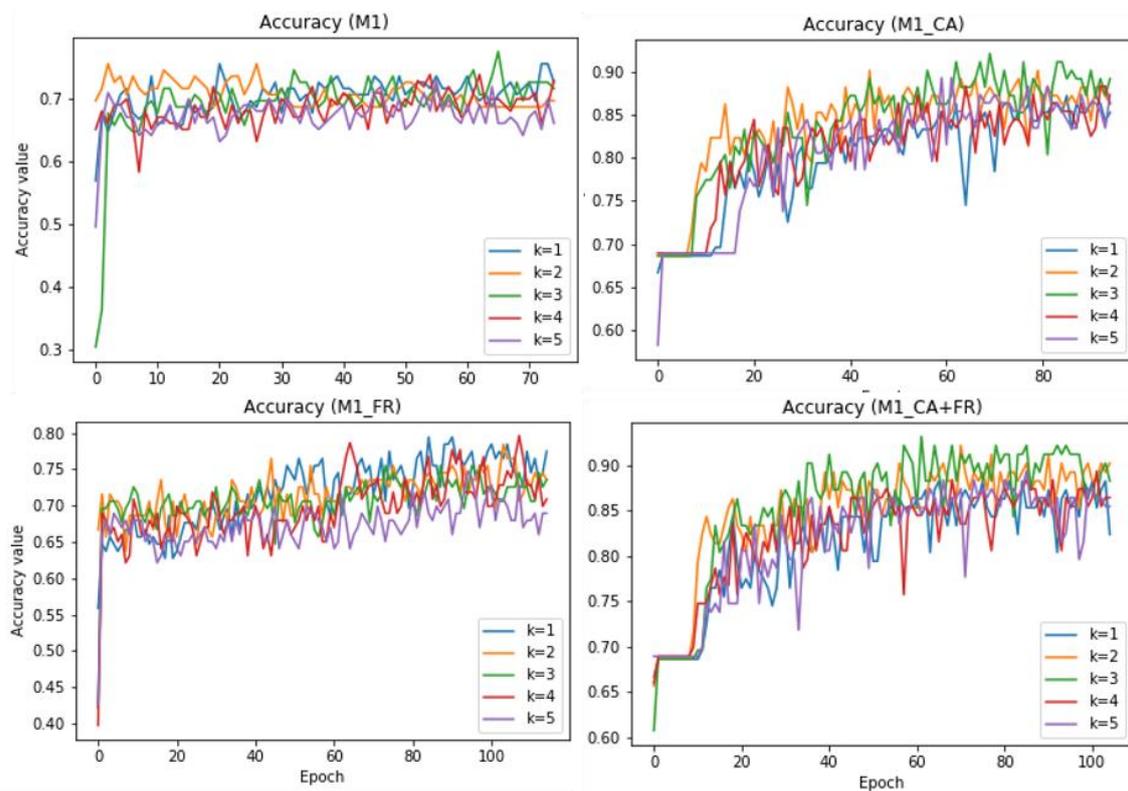


Figura 5.7: Desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na leitura dos textos, no cenário de problema de classificação binária.

Problema de classificação multiclases

Os melhores resultados de cada modelo encontram-se registados na tabela 5.12, tendo em conta o problema de classificação multiclases, ou seja, com base na classificação dos participantes face à gravidade da patologia através do sistema de classificação CAD-RADS. No

geral, estes resultados são insatisfatórios comparativamente aos apresentados anteriormente no problema de classificação binária.

Tabela 5.12: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclases.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	120 / 5	40,32%	($\pm 7,28\%$)	14,30%	($\pm 7,66\%$)
M1_CA	75 / 5	64,03%	($\pm 3,23\%$)	55,31%	($\pm 11,56\%$)
M1_FR	75 / 5	42,06%	($\pm 4,82\%$)	19,92%	($\pm 9,37\%$)
M1_CA+FR	95 / 5	64,11%	($\pm 6,64\%$)	57,53%	($\pm 4,43\%$)

Verifica-se que os modelos com os melhores resultados relativamente ao valor da média, quer da métrica ‘accuracy’, quer da métrica ‘F-measure’ são os modelos M1_CA e M1_CA+FR.

Analisando graficamente o desempenho de cada modelo quanto à métrica ‘accuracy’ nos seus processos de treino através da figura 5.8, confirma-se que o modelo M1_CA e M1_CA+FR apresentam melhores. Por outro lado, todos os modelos apresentam sinais de convergência relativamente aos processos de treino.

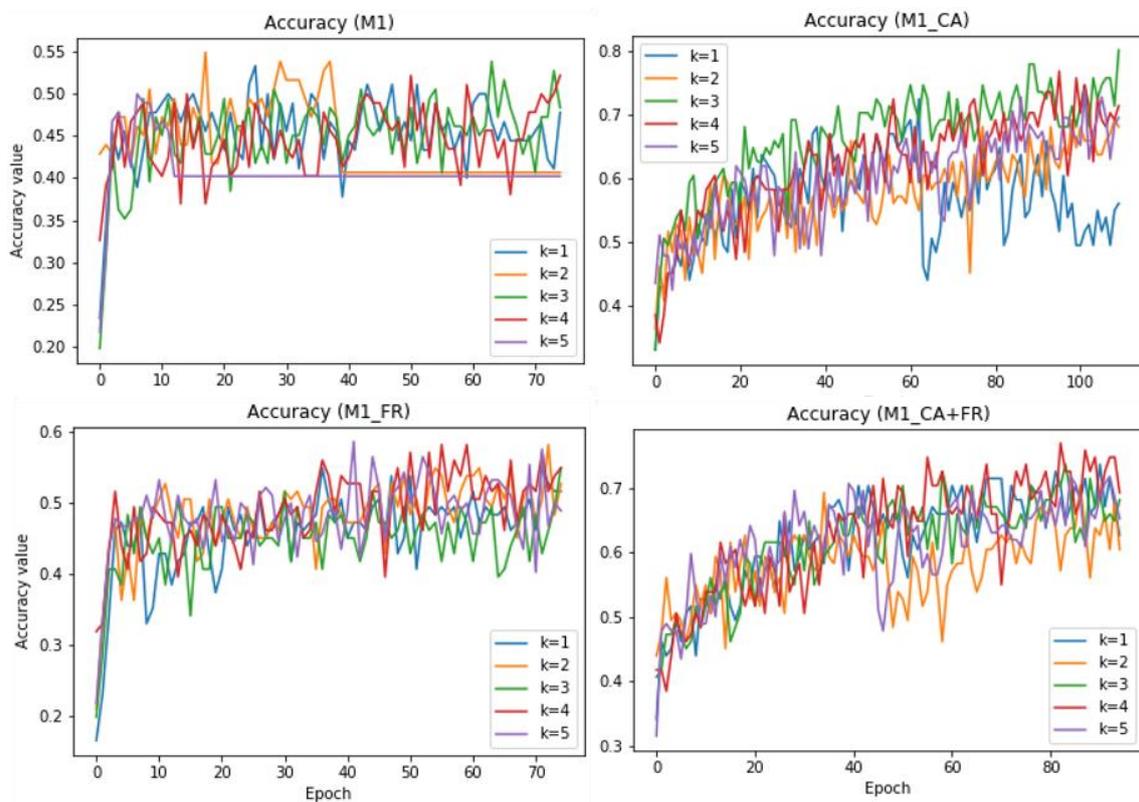


Figura 5.8: Desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclases.

Verifica-se neste problema de classificação multiclasse um desequilíbrio entre classes: classe 0: 35%; classe 1: 40%; classe 3: 18%; e classe 4: 7%. Este desequilíbrio é bastante mais acentuado face ao anterior problema de classificação binária, podendo o método de avaliação e de amostragem aplicado não ser suficiente para este problema de classificação devido aos resultados insatisfatórios. Foram aplicadas algumas técnicas de reamostragens aos conjuntos de dados utilizados no treino dos modelos DL anteriormente analisados que, após um processo de otimização, se registaram os seguintes melhores resultados: o modelo M1 com uma média na ‘accuracy’ de 35,06% ($\pm 5,31$); M1_CA com 56,09% ($\pm 4,43$); M1_FR com 42,02 ($\pm 6,63\%$); e por último, o modelo M1_CA+FR com 64,03% ($\pm 5,06\%$). Em todos os modelos os resultados correspondem à aplicação do método de sobreamostragem ‘BorderlineSMOTE’ nos dados utilizados para o processo de treino à exceção do modelo M1 em que foi aplicado o método de sobreamostragem SMOTE. Comparativamente aos resultados apresentados na tabela 5.12, visualiza-se uma diferença muito pouco significativa (em alguns casos, nula) entre estes últimos resultados apresentados em que utilizaram métodos de reamostragem.

Tendo em conta os resultados obtidos após aplicação de técnicas de reamostragem e por se tratar de um problema de classificação multiclasse, aplicou-se a estratégia OvR aos modelos obtendo-se os resultados apresentados na tabela 5.13.

Tabela 5.13: Resultados obtidos pelos modelos desenvolvidos aplicando a técnica de binarização, com os dados recolhidos na leitura dos textos, no cenário de problema de classificação multiclasse.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	Epochs /mini-batches	Accuracy		F-Measure	
M1	75 / 5	45,49%	($\pm 13,35\%$)	38,74%	($\pm 12,28\%$)
M1_CA	75 / 5	65,77%	($\pm 1,99\%$)	62,32%	($\pm 2,22\%$)
M1_FR	75 / 5	45,49%	($\pm 12,77\%$)	38,91%	($\pm 11,51\%$)
M1_CA+FR	75 / 5	70,16%	($\pm 5,15\%$)	67,3%	($\pm 6,57\%$)

Identificam-se melhorias nos valores obtidos quer da média, quer do desvio padrão de ambas as métricas, sendo o melhor desempenho obtido pelo modelo M1_CA+FR. Ainda assim, os modelos M1 e M1_FR apresentam resultados bastante insatisfatórios quanto à classificação.

Categoria CAD-RADS 0 vs CAD-RADS 1

Com base apenas numa amostra constituída por indivíduos na categoria CAD-RADS 0 e CAD-RADS 1, o modelo M1 e M1_CA+FR foram testados sendo registados os melhores resultados obtidos na tabela 5.14. Identifica-se uma diferença significativa nos resultados obtidos por ambos os modelos. O modelo M1_CA+FR conseguiu alcançar um desempenho bastante positivo com cerca de 79% em contraste com o modelo M1. Relativamente ao último, não é

possivelmente mencionar que tenha conseguido realizar uma boa classificação, realçando neste problema a vantagem de se utilizar dados clínicos adicionais na identificação entre indivíduos sem doença diagnosticada e com DAC não obstrutiva mínima.

Tabela 5.14: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 1.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	95 / 5	52,22%	($\pm 5,16\%$)	51,31%	($\pm 16,78\%$)
M1_CA+FR	105 / 10	79,15%	($\pm 6,70\%$)	80%	($\pm 3,44\%$)

Categoria CAD-RADS 0 vs CAD-RADS 2

Analisando a tabela 5.15 identifica-se uma diferença significativa dos resultados obtidos por ambos os modelos, sendo estes alcançaram um desempenho bastante bom na classificação entre indivíduos sem doença e com doença não obstrutiva leve. O modelo M1_CA+FR obteve o melhor resultado com uma média aproximadamente de 92% relativamente à exatidão na classificação e cerca de 87,4% no valor da média da métrica '*F-measure*'. De salientar que com o modelo M1 verificou-se um valor bastante insatisfatório quer com a média, quer com o desvio padrão em relação à métrica '*F-measure*', não sendo um bom indicador.

Tabela 5.15: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e 2.

Modelo	Parâmetros da rede	Medidas de desempenho ($\mu (\pm\sigma)$)			
	<i>Epochs /mini-batches</i>	<i>Accuracy</i>		<i>F-Measure</i>	
M1	95 / 5	70,00%	($\pm 4,08\%$)	29,43%	($\pm 24,64\%$)
M1_CA+FR	115 / 5	91,67%	($\pm 5,27\%$)	87,43%	($\pm 6,66\%$)

Categoria CAD-RADS 0 vs CAD-RADS 3

Por último, os modelos DL são testado numa amostra com apenas indivíduos sem doença e com doença grave (presença de estenose), sendo os melhores resultados identificados na tabela 5.16. Quanto às restantes categorias mais graves do sistema de classificação CAD-RADS não foram tidas em consideração, porque não havia número suficiente de amostras para garantir a sua representação em todos os k modelos criados pelo método de validação cruzada k -fold estratificado. Verifica-se novamente uma melhoria no valor da média nos resultados obtidos quanto à '*accuracy*' da classificação, quer com o modelo M1, quer com o M1_CA+FR. No entanto, estes valores apresentados podem ser demasiado otimistas, uma vez que este problema

de classificação apresenta um ligeiro desequilíbrio, com aproximadamente 83% para a classe maioritária (Classe 0) e 17% para a classe minoritária (Classe 1).

Tabela 5.16: Resultados obtidos pelos modelos desenvolvidos com os dados recolhidos na leitura dos textos, numa amostra constituída por indivíduos da categoria CAD-RADS 0 e CAD-RADS 3.

Modelo	Parâmetros da rede <i>Epochs /mini-batches</i>	Medidas de desempenho ($\mu (\pm\sigma)$)			
		<i>Accuracy</i>		<i>F-Measure</i>	
M1	75 / 5	83,56%	($\pm 4,35\%$)	33,33%	($\pm 27,89\%$)
M1_CA+FR	75 / 5	100%	($\pm 0,00\%$)	100%	($\pm 0,00\%$)

De igual modo, ambos os modelos foram novamente testados aplicando métodos de reamostragem aos conjuntos de dados, tal como aconteceu anteriormente com o problema de classificação multiclases. O modelo M1_CA+FR registou uma média da ‘accuracy’ 100% ($\pm 0,00$) quer com a técnica SMOTE e BorderlineSMOTE, enquanto que o modelo M1 obteve uma média de 83,33% ($\pm 10,28$) e 81,78% ($\pm 13,14$), com o classificador *Balanced Bagging* e com a técnica BoderlineSMOTE respetivamente.

Para concluir esta análise, é relevante observar a evolução no que diz respeito ao desempenho dos processos de treino dos modelos graficamente que analisaram individualmente as categorias do sistema de classificação CAD-RADS e que originaram os resultados anteriormente apresentados. Observa-se através da figura D.2 em anexo, que as curvas de aprendizagem do treino dos modelos mostraram melhores desempenhos à medida que a gravidade da doença coronária aumenta. Uma vez que estas curvas são mais regulares e apresentam um melhor comportamento de convergência.

5.3. Comparação e discussão dos resultados

Os resultados demonstraram que a análise acústica poderá ser de facto uma ferramenta eficaz na classificação, devido aos resultados satisfatórios obtidos pelos modelos DL quando avaliaram indivíduos saudáveis e com doença coronária através de um problema de classificação binária. Estes resultados foram consistentes quer com a análise acústica realizada no domínio do tempo, utilizando até cerca de 18 parâmetros acústicos, quer no domínio das frequências através dos coeficientes MFCC. Deste modo, não se identificaram diferenças significativas quanto às duas abordagens utilizadas para a realização da análise acústica. Este resultado indica que as diferentes abordagens adotadas são igualmente eficazes, sugerindo ambas a existência de uma possível associação entre a voz e a doença coronária.

Sabe-se que o sistema nervoso autónomo é um sistema crucial nas respostas fisiológicas e patológicas do sistema cardiovascular, dividindo-se ainda em dois sistemas distintos: o sistema

nervoso simpático e o sistema nervoso parassimpático [121]. Ambos trabalham em conjunto para manter a homeostase corporal [122]. Com base nisto, pressupõe-se que o sistema nervoso autónomo funcione normalmente na ausência de qualquer doença cardíaca. Nestas condições, por exemplo, a frequência cardíaca deve refletir adequadamente as respostas face às condições ambientais, bem como uma pressão arterial adequada e outras respostas vaso regulatórias face aos diversos estímulos [121], [123]. Mas quando uma doença cardíaca altera as condições hemodinâmicas, a situação altera-se [123]. Inclusivamente, muitos dos mecanismos compensatórios que podem levar até à modificação da fisiologia cardíaca envolvem nomeadamente, o sistema nervoso autónomo [123].

Adicionalmente, considera-se que o sistema nervoso autónomo esteja relacionado com a voz devido em grande parte à função do nervo vago, um dos nervos cranianos cruciais na produção do sinal de voz [122]. O processo que envolve quer a produção, quer o processamento do sinal de voz estão diretamente dependentes da cooperação de aproximadamente 100 músculos, inervados por uma rede diversificada de nervos cranianos e espinhais, incluindo também algumas partes subcorticais e corticais do cérebro e ainda, processos cardiorrespiratórios. Deste modo, quando ocorre um desequilíbrio ao nível fisiológico ou até emocional, é provável que este sinal seja comprometido. O sinal de voz trata-se assim de um processo psicofisiológico, influenciado por estímulos ambientais e/ou internos [124].

No trabalho desenvolvido por Alvear, Barón-López, Alguacil e Dawid-Milner [125] comprovaram que todas as variáveis cardiovasculares estão significativamente associadas à função vocal da laringe quer em condições de repouso, quer condições de stress, sendo as alterações mediadas pelo sistema nervoso autónomo. Inclusivamente à mais de três décadas, Orlikoff e Baken [126] já tinham identificado que a frequência cardíaca está associada à modulação da frequência fundamental do sinal de voz.

Por outro lado, existe um vasto número de estudos epidemiológicos, fisiopatológicos e de ensaios clínicos que comprovam a importância da frequência cardíaca em doenças cardiovasculares [127]. Nomeadamente, Fox e Ferrari [127] apresentam os principais mecanismos pelos quais a frequência cardíaca pode promover o desenvolvimento e progressão da DAC, podendo até provocar a rutura das placas ateroscleróticas.

Assim, sugere-se a existência de uma associação entre as características do sinal de voz, o sistema nervoso autónomo e a DAC. Por outras palavras, podem ocorrer alterações nas características do sinal de voz devido aos fatores fisiopatológicos da DAC, sendo estas alterações mediadas pelo sistema nervoso autónomo.

O procedimento para a recolha dos dados centrou-se em 3 etapas: (1) emissão de vogais sustentadas, (2) leitura de um texto que expressa emoções positivas e (3) que expressa emoções negativas. No entanto, no momento do tratamento dos dados do presente estudo optou-se por

analisar em conjunto os dados recolhidos na leitura de ambos os textos com intuito de se avaliar o impacto das emoções, independentemente do tipo de emoção expressa ser positiva ou negativa. Analisar estes dados individualmente com base no tipo de emoção provocada, não era viável pelo reduzido número de dados recolhidos. Mas através da análise conjunta destes dados, observou-se inclusivamente que os resultados foram ligeiramente superiores comparativamente aos obtidos com os dados recolhidos pela emissão de vogais sustentadas. Estes resultados vão de encontro com os resultados apresentados por Maor *et al.* [7]. No seu trabalho desenvolvido sugeriram uma associação entre a voz e a aterosclerose, mediada pela hipersensibilidade do sistema adrenérgico ao stress emocional [7], uma vez que existem diversos estudos que demonstraram a relação entre o stress, o sistema adrenérgico e a aterosclerose [128], [129]. Este mecanismo é também suportado por outros estudos que evidenciam que as alterações emocionais como o stress, a depressão ou ansiedade, são um importante fator de risco para a DAC [130], [131] e ainda, pelas fortes evidências que o stress emocional afeta o sinal de voz, sendo estas alterações mediadas pelo sistema adrenérgico, ou seja, pelo sistema nervoso autónomo [132], [133].

5.3.1. Impacto das informações clínicas adicionais

Durante o desenvolvimento da base de dados, foram recolhidos dados clínicos adicionais sobre os participantes do estudo, inclusive o número de fatores de risco associados e os valores do score de cálcio. Incorporando estes dados adicionais aos modelos DL como variáveis de entrada para além da análise acústica, observou-se um melhor desempenho da classificação por parte dos modelos DL. Por outro lado, analisando estas duas variáveis isoladamente em conjunto com a análise acústica, verificou-se que os valores do score de cálcio proporcionaram melhores resultados. Este resultado deve-se ao facto de que o número de fatores de risco associados e recolhidos dos participantes não seguir uma relação linear com a doença coronária, tal como acontece com os valores do score de cálcio. Por exemplo, de entre os 76 participantes deste estudo, apenas um participante apresentou um total máximo de 4 fatores de risco e no entanto, apresentou ausência de DAC. No entanto, na prática clínica é bastante mais simples e rápido realizar um levantamento dos fatores de risco associados, em contraste com o cálculo do score de cálcio. Em relação a este, parece existir uma forte relação entre o cálculo do cálcio presente nas artérias coronárias e a carga aterosclerótica, evidenciando-se ser um importante marcador da doença coronária. Porém, apresenta a desvantagem de ser necessário adquirir-se imagens imagiológicas para se realizar a quantificação do cálcio presente nas artérias coronárias.

Ainda assim, estes resultados demonstram a importância e o potencial de se utilizar a análise acústica em conjunto com outras ferramentas, nomeadamente os fatores de risco, no qual já são utilizados para estimar o risco cardiovascular em modelos como o SCORE. O cálculo do risco cardiovascular pode ser melhorado dado que as ferramentas mais frequentemente utilizadas carecem de algumas limitações. Através de pesquisas realizadas em alguns países europeus,

diversos profissionais afirmaram preferir usar a sua própria experiência na prática clínica por acreditar que o controlo através dos fatores de risco é insuficiente [12].

Sugere-se assim que análise acústica em conjunto com as *guidelines* utilizadas e recomendadas na prática clínica, possa ajudar a melhorar a precisão dos modelos para estimativa do risco cardiovascular, por se tratar de um método não invasivo e de fácil aplicação.

5.3.2. Principais técnicas e parâmetros utilizados nos modelos DL

O maior desafio durante o processo de treino e na avaliação dos modelos centrou-se no desequilíbrio entre classes das amostras em estudo. Por exemplo, este desequilíbrio variou entre 47/53% a 83/17% quando as categorias CAD-RADS eram avaliadas isoladamente com a categoria dos indivíduos saudáveis (CAD-RADS 0). Ainda no âmbito dos problemas de classificação binária avaliados, quando a amostra apenas tinha como base a ausência ou presença de doença coronária independentemente do grau de gravidade, o desequilíbrio variou entre 31/69% a 30/70% para os diferentes dados recolhidos nos procedimentos do processo de recolha de dados.

Tendo em conta que todas as amostras apresentavam um ligeiro desequilíbrio nas suas classes, optou-se por utilizar a estratégia da estratificação no método de avaliação do desempenho dos modelos e através do método de validação cruzada *k-fold*, como já discutido anteriormente. Por outras palavras, o que acontece é que quando são treinados os *k* modelos criados tendo em conta a validação cruzada *k-fold*, o conjunto de dados de cada modelo mantém a percentagem de amostras de cada classe do conjunto de dados inicial. Esta é uma prática recomendada para conjuntos de dados desequilibrados [134]. Inclusive, Raschaka [134] defende que o método de estratificação é benéfico em modelos DL e de fácil implementação, demonstrando que este tem um efeito positivo na variação e no viés da estimativa na validação cruzada *k-fold*.

No entanto, onde se verificou o maior desequilíbrio foi nos problemas de classificação multiclases. Uma vez que os resultados da ‘*accuracy*’ da classificação foram superiores quando amostra incluía somente os indivíduos CAD-RADS 0 e 3 em comparação com a amostra que incluía os indivíduos CAD-RADS 0 e 1, sugere-se que os maus resultados obtidos por parte dos modelos nos problemas de classificação multiclases resultem da reduzida representação nas amostras das categorias mais severas do sistema de classificação CAD-RADS, nomeadamente as categorias 4, 5 e N.

Em todos os problemas de classificação foi utilizada o método de validação cruzada *k-fold* estratificada devido aos desequilíbrios entre classes. No entanto, este método pode revelar-se insuficiente e apresentar resultados demasiados otimistas quando o desequilíbrio é bastante acentuado. Por este motivo, nas amostras em que a classe maioritária apresentava uma percentagem igual ou superior a 80% ou quando as classes minoritárias apresentavam uma percentagem igual ou inferior a 7%, foram aplicadas técnicas de reamostragem aos conjuntos de

dados. Na comparação dos dados obtidos, não se verificaram diferenças significativas no valor da média em relação à *'accuracy'*, mas o desvio padrão associado foi bastante superior na maioria dos casos. Na verdade, as técnicas de reamostragem podem ser eficazes em melhorar o desempenho dos modelos, mas também apresentam alguns problemas associados. Por exemplo, no caso da subamostragem aleatória, podem ser removidos exemplos importantes da classe maioritária, prejudicando o desempenho obtido por parte do classificador. Para além disto, He e Garcia [99] referem alguns estudos que demonstraram que certos algoritmos treinados a partir de conjuntos de dados desequilibrados, os seus resultados são comparáveis com os que foram treinados a partir do mesmo conjunto de dados equilibrado através de técnicas de amostragem. Ainda na procura contínua de alcançar os melhores resultados utilizou-se a estratégia *'OnevsRest'*, uma técnica de binarização aos problemas de classificação multiclases. Optou-se por se utilizar esta estratégia devido ao menor gasto computacional necessário, em comparação com a outra técnica de binarização (OvO) também apresentada neste trabalho. No geral, verificou uma pequena melhoria nos resultados obtidos pelos modelos em relação às duas métricas avaliadas, identificando até modelos com desempenhos razoáveis na classificação. Ainda assim, considera-se que estes resultados também poderão ter sido condicionados, uma vez que esta estratégia pode ser afetada pela utilização de um grande número de classes.

Um outro aspeto, crucial na avaliação da qualidade dos modelos trata-se da escolha adequada das métricas de desempenho. Em todos os problemas de classificação apresentados neste trabalho recorreu-se principalmente à métrica *'accuracy'*. Esta seleção baseia-se no facto de se pretender prever o valor das classes, sendo elas igualmente importantes entre si em todos os problemas de classificação [135].

Por outro lado, o desequilíbrio entre classes que se verificou nas diversas amostras em estudo não foi a única grande limitação encontrada na aplicação dos modelos DL desenvolvidos. Este trabalho apresenta uma outra limitação no que toca nomeadamente, à aplicação do modelo M2 aos dados recolhidos durante a leitura dos textos que transmitem emoções positivas e negativas aos participantes deste estudo. Como é possível observar nos resultados apresentados anteriormente, este modelo que realiza a análise acústica no domínio das frequências não foi utilizado, nem tido em consideração pelos dados recolhidos neste procedimento de recolha. Isto deve ao facto de o modelo necessitar que os ficheiros WAV utilizados tivessem exatamente a mesma dimensão, para assim ser possível a extração dos coeficientes MFCC para o treino do modelo. No entanto, é compreensível que em relação a este aspeto, nem todos os participantes conseguiram ler os textos propostos com a mesma rapidez. Este processo funcionou para os ficheiros recolhidos durante a emissão das vogais sustentadas, até porque na literatura é recomendável analisar a porção mais estável da vogal sustentada correspondendo à porção mais central [49]. No entanto, não seria viável aplicar este mesmo método de limitar ou escolher apenas

uma porção dos ficheiros recolhidos durante a leitura dos textos, podendo essa situação comprometer os resultados e até introduzir algum viés.

Por último, em relação às diferentes arquitetura de rede utilizadas, é possível identificar algumas diferenças face alguns aspetos nomeadamente, na rapidez do processo de treino. Por exemplo, o modelo M2 que utiliza a arquitetura CNN exigiu de um intervalo de tempo superior para o processo de treino comparativamente ao modelo 1 e as suas variações. Por outro lado, o modelo M2 necessitou de menos épocas para aprender o problema comparando o número de épocas utilizado pelos restantes modelos.

5.3.3. Teste Adicional: exclusão de participantes fumadores

As estruturas vocais responsáveis pela produção do sinal de voz são fortemente afetadas pelo estilo de vida, destacando inclusive a influência do tabagismo [47]. Por outro lado, o tabagismo é um dos mais importantes e prevalente fatores de risco na população face ao desenvolvimento e progressão das doenças cardíacas, logo a exclusão de participantes fumadores neste estudo provocaria a ausência de uma significativa parte da população em análise. No entanto, a influência do tabaco nas cordas vocais não deve ser ignorada e por este motivo, durante a análise dos resultados obtidos por parte dos classificadores desenvolvidos, foi pertinente investigar esta influência nas amostras avaliadas. De facto, foi extremamente importante realizar este teste adicional para a avaliação da fiabilidade dos resultados obtidos surgirem devido às características da doença coronária.

Como tal, nos problemas de classificação binária de cada procedimento de recolha de dados analisados, realizou-se o referido teste para se avaliar a influência do tabaco nas cordas vocais. Constituíram-se novas amostras em que os participantes fumadores foram excluídos e de seguida, foram comparados os resultados obtidos dos modelos DL que utilizaram somente as informações extraídas da análise acústica com as amostra sem fumadores e com fumadores. No geral, os resultados demonstraram um diferença mínima quanto ao valor da média de ambas as métricas de desempenho avaliadas. Inclusivamente, é possível visualizar-se na figura 5.9 a comparação dos resultados apresentados nas tabela 5.3 e 5.4 relativamente aos dados recolhidos durante a emissão das vogais sustentadas, identificando-se uma diferença inferior a 1% nos valores obtidos pela métrica de desempenho ‘*accuracy*’ avaliada. Isto comprova que os resultados satisfatórios por parte dos classificadores surgem devido às características da doença coronária e não, pela influência direta do tabagismo nas cordas vocais. De referir que este teste adicional apenas foi realizado no cenário de classificação binária devido aos problemas revelados nos problemas de classificação multiclases nomeadamente, por causa do desequilíbrio entre classes identificado com a utilização do sistema de CAD-RADS.

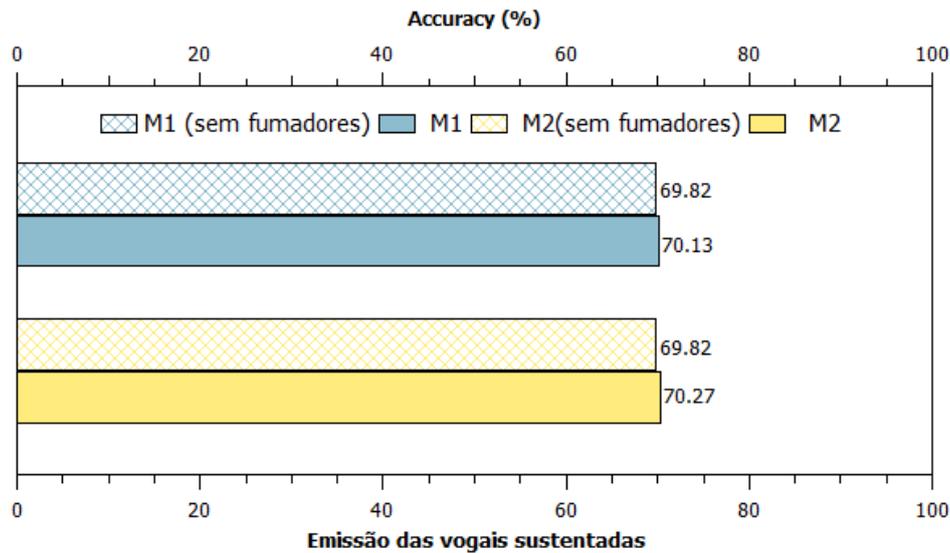


Figura 5.9: Comparação dos resultados obtidos pelos modelos desenvolvidos entre amostras com e sem participantes fumadores, no cenário de problema de classificação binária.

Assim, optou-se por realizar as comparações dos resultados obtidos entre amostras de dados sem fumadores e com fumadores apenas nos problemas de classificação binária, uma vez que os resultados obtidos nestes foram os mais robustos e confiáveis e consequentemente, alcançar-se uma conclusão confiável tanto quanto possível.

5.4. Síntese dos resultados

Após o desenvolvimento dos modelos para análise dos dados recolhidos com recurso às técnicas de *deep learning*, estes foram aplicados aos diferentes conjuntos de dados recolhidos, sendo analisados em diversos cenários de classificação e no final, avaliados através das métricas de desempenho principalmente pela métrica ‘*accuracy*’.

No geral, em ambos os conjuntos de dados avaliados, os problemas de classificação binária alcançaram os melhores resultados após análise dos valores obtidos pelas métricas de desempenho ‘*accuracy*’ e ‘*F-measure*’. Nomeadamente, através da visualização da figura 5.10, os modelos obtiveram desempenhos na ordem dos 70,13% a 83,71% nos dados recolhidos durante a emissão das vogais sustentadas, e nos dados recolhidos durante a leitura dos textos, obtiveram desempenhos na ordem dos 71,05% a 88,34% na exatidão da classificação entre indivíduos saudáveis e com doença coronária.

Os problemas de classificação multiclases obtiveram resultados menos satisfatórios, até mesmo com a aplicação de técnicas de reamostragem ou de binarização (tal como é possível observar na figura 5.11), sendo estes resultados influenciados pelo desequilíbrio entre classes (ou das diferentes categorias do sistema de classificação CAD-RADS).

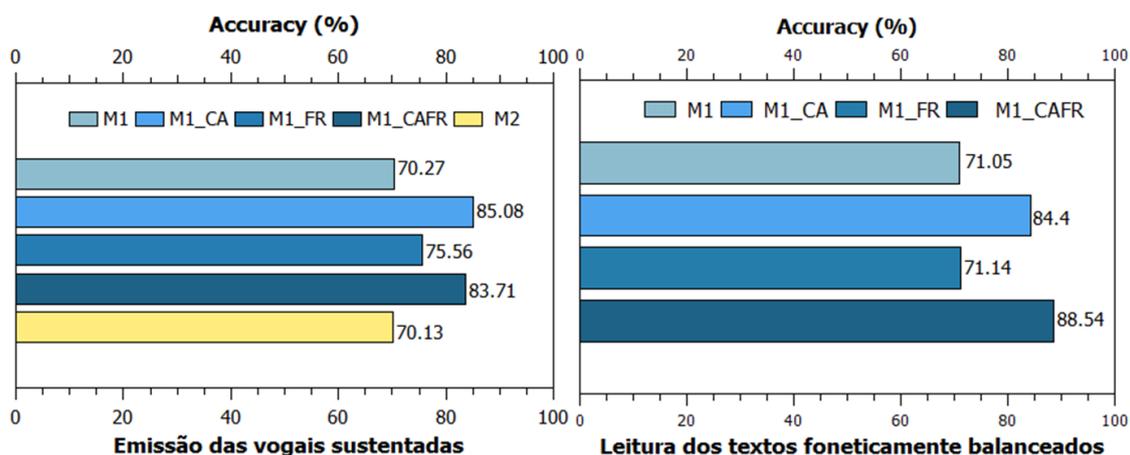


Figura 5.10: Comparação dos resultados obtidos pelos modelos desenvolvidos quanto à ‘accuracy’ entre os dois conjuntos de dados recolhidos, no cenário de problema de classificação binária.

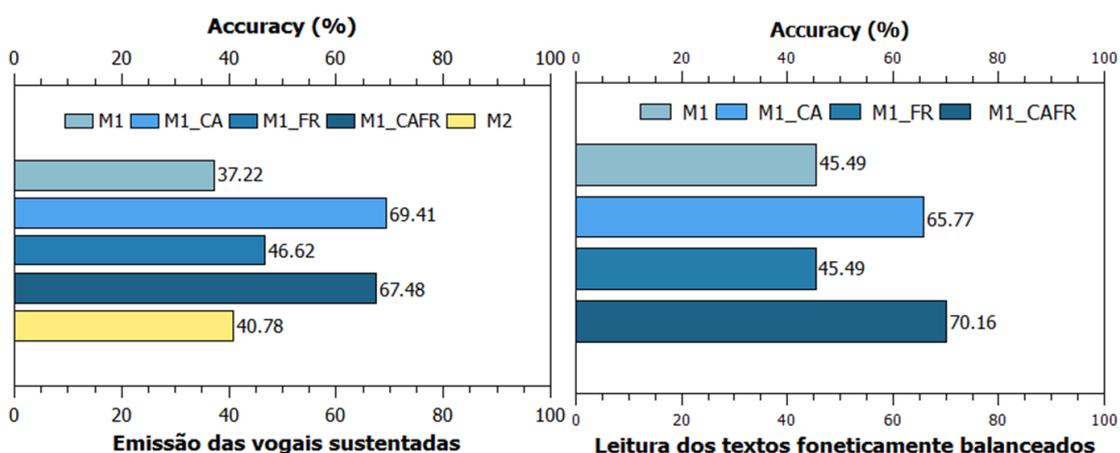


Figura 5.11: Comparação dos resultados obtidos pelos modelos desenvolvidos quanto à ‘accuracy’ entre os dois conjuntos de dados recolhidos, no cenário de problema de classificação multiclass.

Relativamente à avaliação individual dos diferentes graus da doença coronária vs indivíduos saudáveis, os modelos obtiveram melhores resultados quando as categorias mais graves do sistema de classificação CAD-RADS eram avaliadas (por exemplo, CAD-RADS 0 vs CAD-RADS 3) com os dados recolhidos quer na emissão das vogais sustentadas, quer na leitura dos textos foneticamente balanceados.

Por outro lado, em relação aos desempenhos dos diferentes modelos DL desenvolvidos e analisados, verificou-se que o modelo que incorporou todas as informações clínicas adicionais (M1_CA+FR) obteve sempre os melhores desempenhos na classificação. Adicionalmente, verificou-se que as diferentes abordagens utilizadas para a extração de características utilizando a análise acústica são igualmente eficazes, uma vez que não se registaram diferenças significativas nos resultados obtidos entre os modelos M1 e M2.

Conclusão

Neste último capítulo são apresentadas as conclusões retiradas com o trabalho desenvolvido, referindo os objetivos alcançados e algumas limitações encontradas. Através destas limitações, novas perspectivas podem ser estabelecidas para novos desenvolvimento ou trabalhos futuros sobre esta temática.

6.1. Considerações finais

Em todo o mundo, as doenças cardiovasculares têm demonstrado serem devastadoras pelas elevadas taxas de mortalidade associadas, destacando inclusivamente a DAC. Pelo facto de esta patologia estar fortemente associada a diversos fatores de risco, foram desenvolvidas diversas ferramentas para avaliação do risco cardiovascular nas últimas décadas com base na análise destes fatores de risco. Ainda assim, a avaliação do risco cardiovascular na prática clínica parece carecer de algumas limitações com a utilização das referidas ferramentas, revelando-se a necessidade de melhorar o uso destas ou ainda, procurar outro tipo de técnicas adicionais para ajudar na prevenção de eventos coronários fatais.

Recentemente, têm surgido na literatura estudos que sugerem uma possível associação entre os parâmetros acústicos do sinal de voz e as doenças cardiovasculares, através do uso de uma técnica de avaliação vocal: a análise acústica. Esta trata-se de uma ferramenta bastante usada na prática clínica por ser eficaz, não invasiva e de fácil aplicação para avaliações objetivas da função vocal.

De facto, têm sido realizados inúmeros estudos em que utilizaram esta técnica para identificar se as características vocais se encontram relacionadas com determinadas condições patológicas como por exemplo, as doenças neurodegenerativas, sendo que esta relação já é bem conhecida e estudada nas últimas décadas. Com base nisto, nas evidências recentes da existência de uma associação entre a voz e as doenças cardiovasculares e ainda, tendo em consideração as dificuldades de um cálculo eficaz do risco cardiovascular na prática clínica, pretendeu-se com o

desenvolvimento deste trabalho identificar o potencial da análise acústica relativamente à possibilidade de se distinguirem padrões acústicos associados à doença coronária aterosclerótica em indivíduos falantes do Português Europeu.

Para tal, foi necessário ainda desenvolver uma base de dados de voz contendo registo de indivíduos com diagnóstico e ausência de doença coronária, um dos objetivos imprescindíveis para a realização deste projeto. Esta base de dados é constituída pelas gravações de voz e registo de informações clínicas indispensáveis recolhidas de um total de 76 indivíduos que concordaram e aceitaram voluntariamente em participar neste estudo.

Para o processamento e análise dos dados recolhidos utilizou-se uma das técnicas da IA, (*deep learning*) com base no sucesso dos algoritmos desenvolvidos através desta técnica nomeadamente, pela capacidade de identificar padrões e discriminar dados. De facto, diversos estudos têm sido desenvolvidos com a aplicação da técnica *deep learning* na área da medicina. Cada vez mais, a evolução tecnológica das últimas décadas tem permitido grandes avanços nesta área, devido ao seu poder em transformar e melhorar a prevenção, tratamento e prestação de cuidados de saúde. Deste modo, foram desenvolvidos dois modelos com a técnica de *deep learning* com base nas diferentes abordagens para realização da análise acústica às gravações de voz incluídas na base de dados criada.

Os resultados obtidos pelos modelos desenvolvidos foram apresentados segundo os diferentes procedimentos adotados no processo de recolha dos dados e dividiram-se em problemas de classificação binária e multiclases, sendo ainda testadas as diferentes categorias do sistema CAD-RADS individualmente vs indivíduos saudáveis. Adicionalmente, estes focaram-se principalmente no valor da métrica de desempenho '*accuracy*'. Para os diferentes dados analisados, foram obtidos resultados satisfatórios com desempenhos na ordem dos 70,13% a 88,34% de exatidão por parte dos modelos avaliados nos problemas de classificação binária. De realçar que os resultados registados foram consistentes pelas diferentes abordagens utilizadas para a realização da análise acústica, por não terem sido registadas diferenças significativas nos resultados entre os dois modelos que incluíram estas duas abordagens, sugerindo que ambas são igualmente eficazes. No entanto, os resultados obtidos foram insatisfatórios para os diferentes conjuntos de dados quando analisados num contexto de problema de classificação multiclases, ou seja, segundo o sistema de classificação CAD-RADS. Diversas estratégias foram aplicadas para se alcançarem melhores resultados, tendo-se verificado em alguns modelos uma ligeira melhoria dos resultados, mas o acentuado desequilíbrio entre as classes e elevado número de classes utilizadas poderão ter sido fatores que condicionaram as referidas estratégias adotadas. Por outro lado, de entre todos os modelos analisado, verificou-se que os que incorporaram informações clínicas adicionais nomeadamente, o número de fatores de risco e os valores de score de cálcio, obtiveram melhores desempenhos na classificação. Isto sugere que análise acústica em conjunto com as *guidelines* recomendadas, pode ajudar a ultrapassar algumas das limitações

associadas à utilização destas na prática clínica e a melhorar a precisão dos modelos para o cálculo do risco cardiovascular.

Com base nos resultados obtidos neste estudo, também se identificou uma possível associação entre os parâmetros do sinal de voz e a DAC. Deste modo, confirmou-se o potencial da análise acústica em distinguir indivíduos com diagnóstico de doença cardíaca e indivíduos saudáveis, mas estudos adicionais são necessários para uma maior confiança na utilização futura desta ferramenta como por exemplo, no auxílio na estimativa do risco cardiovascular.

6.2. Principais limitações e perspetivas futuras

Relativamente aos objetivos e metas delineadas, inclusivamente na secção *Objetivos* do Capítulo 1, é possível concluir que todos foram alcançados com sucesso:

1. Desenvolvimento de uma base de dados de voz contendo registos de indivíduos saudáveis e com patologia cardíaca já diagnosticada, bem como outras informações clínicas relevantes. Esta base de dados em Português Europeu trata-se de um importante instrumento de estudo por ser única, ter sido recolhida em ambiente hospital e anotada por clínicos experientes;
2. Criação de modelos eficazes para a análise dos dados de voz recolhidos e que demonstraram ser poderosas ferramentas de classificação para o apoio à decisão;
3. Confirmação do potencial da análise acústica quanto à possibilidade na identificação de padrões acústicos associados à patologia cardíaca.

Também é possível dar resposta às hipóteses de investigação formuladas no Capítulo 1 e responder afirmativamente à primeira hipótese formulada (H1), que é possível identificar padrões de patologia cardíaca utilizando a análise acústica do sinal de voz. Adicionalmente, na perspetiva clínica identificou-se a capacidade de avaliar a gravidade da doença coronária aterosclerótica segundo o sistema de classificação CAD-RADS, podendo ser realizada uma análise diferencial. Quanto às hipóteses formuladas na perspetiva linguística e de engenharia, concluiu-se que os modelos desenvolvidos e utilizados para a análise foram bastante eficazes na identificação de padrões, permitindo a identificação dos indivíduos saudáveis e com patologia cardíaca. Para além disto, identificou-se que as diferentes abordagens adotadas na extração das características do sinal de voz demonstram ser igualmente eficazes para o reconhecimento da patologia.

No entanto, é importante realçar que existem alguns aspetos que poderiam ser melhorados, com intuito de superar algumas limitações que surgiram no decorrer do desenvolvimento deste estudo. De seguida, são enumeradas algumas sugestões para inspirar o desenvolvimento de trabalhos futuros:

- 1. Adicionar novos dados à base de dados.** Introduzir novos registos e de gravações áudio quer de indivíduos com diagnóstico de doença coronária, quer de indivíduos saudáveis permite aumentar os recursos para o processo de aprendizagem dos modelos desenvolvidos e fornecer maior robustez aos resultados obtidos. Inclusive, é desejável obter mais dados de indivíduos com doença coronária nas categorias mais graves, de modo a combater o desequilíbrio entre as classes identificado no desenvolvimento do presente estudo.
- 2. Aumentar o quantidade e diversidade de parâmetros acústicos.** Apesar de em geral os resultados serem satisfatórios, aumentar o número de parâmetros acústicos extraídos e consequentemente, aumentar as variáveis de entrada dos modelos avaliados, poderá permitir melhorar os recursos para o processo de aprendizagem, bem como aumentar a exatidão da classificação dos modelos. A utilização de modelos de seleção de características, diminuindo a redundância dos dados e contribuindo para modelos de decisão mais simples e eficientes, será também considerada.
- 3. Ampliar a extração dos coeficientes MFCC a todos os dados recolhidos.** Adotar estratégias eficientes que permitem a utilização e a extração dos coeficientes MFCC às gravações realizadas com a leitura dos textos foneticamente balanceados que transmitem emoções.

Bibliografia

- [1] N. Cummins, A. Baird, and B. W. Schuller, “Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning,” *Methods*, 2018.
- [2] J. I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, “Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, 2006.
- [3] Ö. Eskidere and A. Gürhanl, “Voice Disorder Classification Based on Multitaper Mel Frequency Cepstral Coefficients Features,” *Comput. Math. Methods Med.*, 2015.
- [4] Y. S. Bonne, Y. Levanon, O. Dean-Pardo, L. Lossos, and Y. Adini, “Abnormal Speech Spectrum and Increased Pitch Variability in Young Autistic Children,” *Front. Hum. Neurosci.*, vol. 4, no. January, pp. 1–7, 2011.
- [5] M. J. V. García, I. Cobeta, G. Martín, H. Alonso-Navarro, and F. J. Jimenez-Jimenez, “Acoustic analysis of voice in Huntington’s disease patients,” *J. Voice*, vol. 25, no. 2, pp. 208–217, 2011.
- [6] P. Gomez *et al.*, “Voice quality analysis to detect neurological diseases,” *Model. Anal. Vocal Emiss. Biomed. Appl. - 7th Int. Work. MAVIBA 2011*, no. January, 2011.
- [7] E. Maor, J. D. Sara, D. M. Orbelo, L. O. Lerman, Y. Levanon, and A. Lerman, “Voice Signal Characteristics Are Independently Associated With Coronary Artery Disease,” *Mayo Clin. Proc.*, vol. 93, no. 7, pp. 840–847, 2018.
- [8] D. M. Shavelle, “Almanac 2015: Coronary artery disease,” *Heart*, vol. 11, no. 5–6, pp. 188–200, 2016.
- [9] M. Alcaíno and D. Lama, “Coronary Artery Disease in the Elderly,” *Ann. Behav. Med.*, vol. 3, no. 3, pp. 11–14, 2017.
- [10] J. F. Beltrame, R. Dreyer, and R. Tavella, “Epidemiology of Coronary Artery Disease,” in *Coronary artery disease-current concepts in epidemiology, pathophysiology, diagnostics and treatment*, 2012.
- [11] T. Tomasik, J. Krzysztóń, K. Dubas-Jakóbczyk, V. Kijowska, and A. Windak, “The systematic coronary risk evaluation (SCORE) for the prevention of cardiovascular diseases. Does evidence exist for its effectiveness? A systematic review,” *Acta Cardiol.*, vol. 72, no. 4, pp. 370–379, 2017.
- [12] M. F. Piepoli *et al.*, “2016 European Guidelines on cardiovascular disease prevention in clinical practice,” *Eur. Heart J.*, vol. 37, no. 29, pp. 2315–2381, 2016.
- [13] E. R. Dorsey and E. J. Topol, “State of Telehealth,” *N. Engl. J. Med.*, vol. 375, no. 2, pp. 154–161, 2016.
- [14] B. M. Kuehn, “Telemedicine Helps Cardiologists Extend Their Reach,” *Circulation*, vol. 134, no. 16, pp. 1189–1191, 2016.

- [15] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-Based Models for Speech Recognition," *Adv. Neural Inf. Process. Syst.*, no. June, 2015.
- [16] G. Hinton *et al.*, "Deep Neural Networks for Acoustic Modeling in Speech Recognition," *IEEE Signal Process. Mag.*, vol. 29, pp. 1–27, 2012.
- [17] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief. Bioinform.*, vol. 18, no. 5, pp. 851–869, 2017.
- [18] R. Wason, "Deep Learning: Evolution and Expansion," *Cogn. Syst. Res.*, vol. 52, pp. 701–708, 2018.
- [19] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–119, 2017.
- [20] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [21] S. Hahm and J. Wang, "Parkinson's Condition Estimation using Speech Acoustic and Inversely Mapped Articulatory Data Callier Center for Communication Disorders," *Sixt. Annu. Conf. Int. Speech Commun. Assoc.*, pp. 513–517, 2015.
- [22] World Health Organization, "Cardiovascular disease," 2019. [Online]. Available: https://www.who.int/cardiovascular_diseases/en/. [Accessed: 16-Apr-2019].
- [23] M. Aggarwal, B. Aggarwal, and J. Rao, "Integrative Medicine for Cardiovascular Disease and Prevention," *Med. Clin.*, vol. 101, no. 5, pp. 895–923, 2017.
- [24] Serviço Nacional de Saúde, "Doenças cardiovasculares," 2017. [Online]. Available: <https://www.sns.gov.pt/noticias/2017/10/04/doencas-cardiovasculares/>. [Accessed: 26-Mar-2019].
- [25] OECD Publishing, *Health at a Glance 2017: OECD Indicators*. Paris, 2017.
- [26] G. K. Hansson, "Inflammation, Atherosclerosis, and Coronary Artery Disease," *N. Engl. J. Med.*, vol. 352, no. 16, pp. 1685–1695, 2005.
- [27] D. R. Labarthe, *Epidemiology and Prevention of Cardiovascular Diseases: a Global Challenge*, 2nd ed. Massachusetts: Jones and Bartlett Publishers, 2010.
- [28] P. Dehkordi, E. P. Bauer, K. Tavakolian, V. Zakeri, A. P. Blaber, and F. Khosrow-Khavar, "Identifying Patients With Coronary Artery Disease Using Rest and Exercise Seismocardiography," *Front. Physiol.*, vol. 10, no. 1211, pp. 1–10, 2019.
- [29] R. C. Cury *et al.*, "Journal of Cardiovascular Computed Tomography CAD-RADS™ Coronary Artery Disease Reporting and Data System. An expert consensus document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Radiology (ACR) and the N," *J. Cardiovasc. Comput. Tomogr.*, vol. 10, no. 4, pp. 269–281, 2016.
- [30] A. V. Khera and S. Kathiresan, "Genetics of coronary artery disease: discovery, biology and clinical translation," *Nat Rev Genet*, vol. 18, no. 6, pp. 331–344, 2017.
- [31] J. Plutzky, "The Vascular Biology of Atherosclerosis," *Am. J. Med.*, vol. 115, no. 8, pp. 55–61, 2003.
- [32] P. Libby, "Vascular Biology of Atherosclerosis: Overview and State of the Art," *Am. J. Cardiol.*, vol. 91, no. 3, pp. 3–6, 2003.
- [33] MAYO CLINIC, "Arteriosclerosis / atherosclerosis," *Symptoms & causes*, 2018. [Online].

Available: <https://www.mayoclinic.org/diseases-conditions/arteriosclerosis-atherosclerosis/symptoms-causes/syc-20350569>. [Accessed: 28-Jun-2019].

- [34] P. Greenland *et al.*, “Major Risk Factors as Antecedents of Fatal and Nonfatal Coronary Heart Disease Events,” *JAMA*, vol. 290, no. 7, pp. 891–897, 2003.
- [35] I. Fronteira, “Estudos Observacionais na Era da Medicina Baseada na Evidência: Breve Revisão Sobre a Sua Relevância, Taxonomia e Desenhos,” *Acta Med. Port.*, vol. 26, no. 2, pp. 161–170, 2013.
- [36] U. N. Khot *et al.*, “Prevalence of Conventional Risk Factors in Patients With Coronary Heart Disease,” *JAMA*, vol. 290, no. 7, pp. 898–904, 2003.
- [37] P. L. Luz and D. Favarato, “Doença Coronária Crônica,” *Arq Bras Cardiol*, vol. 72, no. 1, pp. 5–21, 1999.
- [38] R. A. Payne, “Cardiovascular risk,” *Br. J. Clin. Pharmacol.*, vol. 74, no. 3, pp. 396–410, 2012.
- [39] G. I. Gorodeski, “Update on cardiovascular disease in post-menopausal women,” *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 16, no. 3, pp. 329–355, 2002.
- [40] F. Mafra and H. Oliveira, “Avaliação do risco cardiovascular – metodologias e suas implicações na prática clínica,” *Rev. Port. Med. Geral e Fam.*, vol. 24, no. 3, pp. 391–400, 2008.
- [41] C. F. Azevedo, C. E. Rochitte, and J. A. C. Lima, “Escore de Cálcio e Angiotomografia Coronariana na Estratificação do Risco Cardiovascular,” *Arq. Bras. Cardiol.*, vol. 98, no. 6, pp. 559–568, 2012.
- [42] P. O. Neves, J. Andrade, and H. Monção, “Escore de cálcio coronariano: estado atual,” *Radiol. Bras.*, vol. 50, no. 3, pp. 182–189, 2017.
- [43] C. Brotons *et al.*, “Implementation of Spanish adaptation of the European guidelines on cardiovascular disease prevention in primary care,” *BMC Fam. Pract.*, vol. 14, no. 1, 2013.
- [44] Z. Reiner, Z. Sonicki, and E. Tedeschi-reiner, “Physicians’ perception, knowledge and awareness of cardiovascular risk factors and adherence to prevention guidelines: The PERCRO-DOC survey,” *Atherosclerosis*, vol. 213, no. 2, pp. 598–603, 2010.
- [45] J. Dallongeville *et al.*, “Survey of physicians’ practices in the control of cardiovascular risk factors: the EURIKA study,” *Eur. J. Prev. Cardiol.*, vol. 19, no. 3, pp. 541–550, 2011.
- [46] E. Demirhan, E. M. Unsal, C. Yilmaz, and E. Ertan, “Acoustic Voice Analysis of Young Turkish Speakers,” *J. Voice*, vol. 30, no. 3, 2016.
- [47] S. Freitas, “Avaliação Acústica e Áudio Percetiva na Caracterização da Voz Humana,” Faculdade de Engenharia Universidade do Porto, 2012.
- [48] L. W. Lopes *et al.*, “Accuracy of Acoustic Analysis Measurements in the Evaluation of Patients With Different Laryngeal Diagnoses,” *J. Voice*, vol. 31, no. 3, 2016.
- [49] M. Shu, J. J. Jiang, and M. Willey, “The Effect of Moving Window on Acoustic Analysis,” *J. Voice*, vol. 30, no. 1, pp. 5–10, 2016.
- [50] C. Carson, J. Ryalls, K. Hardin-hollingsworth, M. Le Normand, and B. Ruddy, “Acoustic Analyses of Prolonged Vowels in Young Adults With Friedreich Ataxia,” *J. Voice*, vol. 30, no. 3, pp. 272–280, 2016.
- [51] K. R. Moon, S. M. Chung, H. S. Park, and H. S. Kim, “Materials of Acoustic Analysis: Sustained Vowel Versus Sentence,” *J. Voice*, vol. 26, no. 5, pp. 563–565, 2012.

- [52] A. P. Mendes, L. J. L. Ferreira, and E. Castro, “Softwares e hardwares de análise acústica da voz e da fala Softwares and hardwares of acoustic,” *Distúrbios da Comun.*, vol. 24, no. 3, pp. 421–430, 2012.
- [53] S. Vaz-freitas, P. Melo, V. Almeida, and A. Ferreira, “Biomedical Signal Processing and Control Acoustic analysis of voice signal: Comparison of four applications software,” *Biomed. Signal Process. Control*, vol. 40, pp. 318–323, 2018.
- [54] H. F. Wertzner, S. Schreiber, and L. Amaro, “Analysis of the fundamental frequency, jitter, shimmer and vocal intensity in phonological disordered children,” vol. 71, no. 5, pp. 582–588, 2005.
- [55] J. P. Teixeira, D. B. Ferreira, and S. M. Carneiro, “Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patologias da fala,” *6º Congr. Luso-Moçambicano Eng. Maputo, Moçambique*, 2011.
- [56] J. Teixeira and A. Gonçalves, “Accuracy of Jitter and Shimmer Measurements,” *Procedia Technol.*, vol. 16, pp. 1190–1199, 2014.
- [57] R. Villafuerte-gonzalez, V. M. Valadez-Jimenez, J. A. Sierra-ramirez, P. A. Ysunza, K. Chavarria-Villafuerte, and X. Hernandez-Lopez, “Acoustic Analysis and Electroglottography in Elite Vocal Performers,” *J. Voice*, vol. 31, no. 3, 2016.
- [58] L. S. Finger, C. A. Cielo, and K. Schwarz, “Medidas vocais acústicas de mulheres sem queixas de voz e com laringe normal,” *Braz. J. Otorhinolaryngol.*, vol. 75, no. 3, pp. 432–440, 2009.
- [59] E. T. Stathopoulos, J. E. Huber, and J. Sussman, “Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4–93 Years of Age,” *J. Speech, Lang. Hear. Res.*, vol. 54, pp. 1011–1021, 2011.
- [60] V. Tiwari, “MFCC and its applications in speaker recognition,” *Int. J. Emerg. Technol.*, vol. 1, no. 1, pp. 19–22, 2010.
- [61] D. P. Cardoso, “Identificação de locutor usando modelos de mistura de gaussianas,” Escola Politécnica da Universidade de São Paulo, 2009.
- [62] E. Thiago, “Reconhecimento de Voz utilizando extração de Coeficientes Mel-Cepstrais e Redes Neurais Artificiais,” Instituto Federal de Santa Catarina – IFSC, 2017.
- [63] J. K. Siqueira and A. Alcaim, “Comparação dos Atributos MFCC , SSCH e PNCC para Reconhecimento Robusto de Voz Contínua,” *XXIX Simpósio Bras. Telecomunicações*, 2011.
- [64] F. Fernandes, “Determinação da Autocorrelação , HNR e NHR para Análise Acústica Vocal,” Escola Superior de Tecnologia e Gestão - Instituto Politécnico de Bragança, 2018.
- [65] D. A. M. Ramírez, V. M. V. Jiménez, X. H. López, and P. A. Ysunza, “Acoustic Analysis of Voice and Electroglottography in Patients With Laryngopharyngeal Reflux,” *J. Voice*, vol. 32, no. 3, pp. 281–284, 2017.
- [66] K. Szklanny, R. Gubrynowicz, J. Ratyńska, and D. Chojnacka-wądołowska, “Electroglottographic and acoustic analysis of voice in children with vocal nodules,” *Int. J. Pediatr. Otorhinolaryngol.*, vol. 122, 2019.
- [67] J. P. Teixeira, P. O. Fernandes, and N. Alves, “Vocal Acoustic Analysis – Classification Dysphonic Voices with Artificial Neural Networks,” *Procedia Comput. Sci.*, vol. 121, pp. 19–26, 2017.
- [68] J. P. Teixeira and P. O. Fernandes, “Acoustic Analysis of Vocal Dysphonia,” *Procedia Comput. Sci.*, vol. 64, pp. 466–473, 2015.

- [69] D. Lu *et al.*, “Changes After Voice Therapy in Acoustic Voice Analysis of Chinese Patients With Voice Disorders,” *J. Voice*, vol. 32, no. 3, 2017.
- [70] A. I. Gillespie, C. Dastolfo, N. Magid, and J. Gartner-schmidt, “Acoustic Analysis of Four Common Voice Diagnoses: Moving Toward Disorder-Specific Assessment,” *J. Voice*, vol. 28, no. 5, 2014.
- [71] P. Zhuge, H. You, H. Wang, Y. Zhang, and H. Du, “An Analysis of the Effects of Voice Therapy on Patients With Early Vocal Fold Polyps,” *J. Voice*, vol. 30, no. 6, pp. 698–704, 2015.
- [72] K. Guranski and R. Podemski, “Emotional prosody expression in acoustic analysis in patients with right hemisphere ischemic stroke,” *Neurol. Neurochir. Pol.*, vol. 49, no. 2, pp. 113–120, 2015.
- [73] Y. Ikui *et al.*, “An Aerodynamic Study of Phonations in Patients With Parkinson Disease (PD),” *J. Voice*, vol. 29, no. 3, pp. 273–280, 2015.
- [74] F. Martínez-sánchez *et al.*, “Speech rhythm alterations in Spanish-speaking individuals with Alzheimer ’ s disease,” *Aging, Neuropsychol. Cogn.*, vol. 0, no. 0, pp. 1–17, 2016.
- [75] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, D. E. López, L. Millian-Morell, and J. M. Arana, “Speech in alzheimer’s disease: Can temporal and acoustic parameters discriminate dementia?,” *Dement. Geriatr. Cogn. Disord.*, vol. 37, no. 5–6, pp. 327–334, 2014.
- [76] A. König *et al.*, “Automatic speech analysis for the assessment of patients with predementia and Alzheimer ’ s disease,” *Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit.*, vol. 1, no. 1, pp. 112–124, 2015.
- [77] P. Vizza *et al.*, “Methodologies of speech analysis for neurodegenerative diseases evaluation,” *Int. J. Med. Inform.*, vol. 122, pp. 45–54, 2018.
- [78] I. Lopez-moreno, J. Gonzalez-dominguez, D. Martinez, O. Plchot, J. Gonzalez-Rodriguez, and P. Moreno, “On the use of deep feedforward neural networks for automatic language identification,” *Comput. Speech Lang.*, vol. 40, pp. 46–59, 2016.
- [79] I. Lopez-moreno, J. Gonzalez-Dominguez, O. Plchot, D. Martinez, J. Gonzalez-Rodriguez, and P. Moreno, “Automatic language identification using deep neural networks,” *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 5337–5341, 2014.
- [80] L. Deng *et al.*, “Recent Advances In Deep Learning for speech research at microsoft,” *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 8604–8608, 2013.
- [81] O. M. Murton *et al.*, “Acoustic speech analysis of patients with decompensated heart failure: A pilot study,” *J. Acoust. Soc. Am.*, vol. 142, no. 4, 2017.
- [82] V. Pareek and R. K. Sharma, “Coronary Heart Disease Detection from Voice Analysis,” in *2016 IEEE Students’ Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2016, pp. 1–6.
- [83] V. van Heuven and P. Boersma, “Speak and unSpeak with PRAAT,” *Glott Int.*, vol. 5, no. 9–10, pp. 341–347, 2001.
- [84] A. P. Mendes, A. N. Costa, A. D. Martins, A. F. O. Fernandes, S. M. D. do R. Vicente, and T. C. S. Freitas, “Contributos para a construção de um texto foneticamente equilibrado para o Português-Europeu,” *Rev. CEFAC*, vol. 14, no. 10, 2012.
- [85] H. Kabir and A. M. Saleem, “Speech assessment methods phonetic alphabet (SAMPA): analysis of urdu,” *CRULP Annu. Student Rep. Publ. Akhbar-e-Urdu*, pp. 6–11, 2002.
- [86] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp.

- 436–444, 2015.
- [87] G. Chartrand *et al.*, “Deep Learning: A Primer for Radiologists,” *Radiographics*, vol. 37, no. 7, pp. 2113–2131, 2017.
- [88] C. Cao *et al.*, “Deep Learning and Its Applications in Biomedicine,” *Genomics. Proteomics Bioinformatics*, vol. 16, no. 1, pp. 17–32, 2018.
- [89] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, “Survey on deep learning for radiotherapy,” *Comput. Biol. Med.*, vol. 98, pp. 126–146, 2018.
- [90] S. A. Puente, “Single and Multi-Label Environmental Sound Classification Using Convolutional Neural Networks,” Chalmers University of Technology, 2018.
- [91] P. Jain, “Complete Guide of Activation Functions,” *Towards Data Science*, 2019. [Online]. Available: <https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044>. [Accessed: 19-May-2020].
- [92] T. Ioanna, “Artificial Neural Networks: an overview and applications,” University of Piraeus, 2019.
- [93] J. M. F. dos Santos, “Data Classification with Neural Networks and Entropic Criteria,” Faculdade de Engenharia da Universidade do Porto, 2007.
- [94] F. Van Veen, “The Neural Network Zoo,” *The Asimov Institute for Artificial Creativity & Constraint*, 2016. [Online]. Available: <http://www.asimovinstitute.org/neural-network-zoo/>. [Accessed: 28-Jul-2019].
- [95] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [96] J. Brownlee, “Loss and Loss Functions for Training Deep Learning Neural Networks,” *Better Deep Learning*, 2019. [Online]. Available: <https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/>. [Accessed: 27-Jul-2019].
- [97] M. Al Khaldy and C. Kambhampati, “Resampling imbalanced class and the effectiveness of feature selection methods for heart failure dataset,” *Int. Robot. Autom. J.*, vol. 4, no. 1, pp. 37–45, 2018.
- [98] A. Fernández, V. López, M. Galar, M. José, and F. Herrera, “Knowledge-Based Systems Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches,” *Knowledge-based Syst.*, vol. 42, pp. 97–110, 2013.
- [99] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [100] M. Zareapoor and P. Shamsolmoali, “Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier,” *Procedia Comput. Sci.*, vol. 48, no. 2015, pp. 679–686, 2015.
- [101] M. Lango and J. Stefanowski, “Multi-class and feature selection extensions of Roughly Balanced Bagging for imbalanced data,” *J Intell Inf Syst*, vol. 50, pp. 97–127, 2018.
- [102] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” vol. 44, pp. 1761–1776, 2011.
- [103] A. R. A. Raziff, N. Sulaiman, N. Mustapha, and T. Perumal, “Single classifier, OvO, OvA and RCC multiclass classification method in handheld based smartphone gait identification Single Classifier, OvO, OvA and RCC Multiclass Classification Method in Handheld Based Smartphone Gait Identification,” *AIP Conf. Proc.*, no. 1, 2017.

- [104] M. Hossin and M. N. Sulaiman, “A review on evaluation metrics for data classification evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–11, 2015.
- [105] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, pp. 861–874, 2006.
- [106] J. Davis and M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves,” *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 233–240, 2006.
- [107] H. J. V L and S. Duraisamy, “Novel pre-processing framework to improve classification accuracy in opinion mining,” *Int. J. Comput.*, vol. 17, no. 4, pp. 234–242, 2018.
- [108] The Spyder Website Contributors, “Overview,” *SPYDER: The Scientific Python Development Environment*, 2018. [Online]. Available: <https://www.spyder-ide.org/>. [Accessed: 29-Jun-2019].
- [109] Anaconda Inc., “Anaconda Distribution,” *Anaconda*, 2019. [Online]. Available: <https://www.anaconda.com/distribution/>. [Accessed: 29-Jun-2019].
- [110] Librosa Development Team, “Tutorial,” *Librosa*, 2018. [Online]. Available: <https://librosa.github.io/librosa/tutorial.html>. [Accessed: 29-Jun-2019].
- [111] Scikit-learn Developers, “Scikit-learn: Machine Learning in Python,” 2019. [Online]. Available: <https://scikit-learn.org/stable/>. [Accessed: 29-Jun-2017].
- [112] NumPy Developers, “NumPy,” 2019. [Online]. Available: <http://www.numpy.org/>. [Accessed: 29-Jun-2019].
- [113] P. S. Foundation, “Imbalanced-learn 0.6.2,” *Project description*, 2020. [Online]. Available: <https://pypi.org/project/imbalanced-learn/>. [Accessed: 22-Apr-2020].
- [114] D. P. Kingma and J. L. Ba, “ADAM: a method for stochastic optimization,” *ICLR 2015*, pp. 1–15, 2014.
- [115] J. Brownlee, “Machine Learning Mastery,” *What is the Difference Between a Batch and an Epoch in a Neural Network?*, 2018. [Online]. Available: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>. [Accessed: 30-Aug-2019].
- [116] J. Torres, “Convolutional Neural Networks for Beginners: Practical Guide with Python and Keras,” 2018. [Online]. Available: <https://towardsdatascience.com/convolutional-neural-networks-for-beginners-practical-guide-with-python-and-keras-dc688ea90dca>. [Accessed: 08-Oct-2019].
- [117] J. Brownlee, “Dropout Regularization in Deep Learning Models With Keras,” 2019. [Online]. Available: <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/>. [Accessed: 08-Oct-2019].
- [118] T. Wong, “Performance evaluation of classification algorithms by k -fold and leave-one-out cross validation,” *Pattern Recognit.*, pp. 1–8, 2015.
- [119] Scikit-learn Developers, “Scikit-learn,” *3.1. Cross-validation: evaluating estimator performance*, 2019. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html. [Accessed: 14-Jul-2019].
- [120] R. B. Rao, G. Fung, and R. Rosales, “On the Dangers of Cross-Validation. An Experimental Evaluation,” *Proc. 2008 SIAM Int. Conf. data mining. Soc. Ind. Appl. Math.*, pp. 588–596, 2008.
- [121] C. N. B. Merz, O. Elboudwarej, and P. Mehta, “The Autonomic Nervous System and Cardiovascular Health and Disease: a Complex Balancing Act,” *JACC Hear. Fail.*, vol. 3,

no. 5, pp. 383–385, 2015.

- [122] R. Cardoso, J. Lumini-Oliveira, and R. F. Meneses, “Associations Between Autonomic Nervous System Function, Voice, and Dysphonia: A Systematic Review,” *J. Voice*, 2019.
- [123] P. Coumel, “Cardiac Arrhythmias and the Autonomic Nervous System,” *J. Cardiovasc. Electrophysiol.*, vol. 4, no. 3, pp. 338–355, 1993.
- [124] M. Van Puyvelde, X. Neyt, F. Mcglone, and N. Pattyn, “Voice Stress Analysis : A New Framework for Voice and Effort in Human Performance,” *Front. Psychol.*, vol. 9, no. 1994, pp. 1–25, 2018.
- [125] R. M. B. Alvear, F. J. Barón-López, M. D. Alguacil, and M. S. Dawid-Milner, “Interactions between voice fundamental frequency and cardiovascular parameters. Preliminary results and physiological mechanisms,” *Logop. Phoniatr. Vocology*, vol. 38, pp. 52–58, 2013.
- [126] R. F. Orlikoff and R. J. Baken, “The effect of the heartbeat on vocal fundamental frequency perturbation,” *J. Speech Hear. Res.*, vol. 32, no. 3, pp. 576–582, 1989.
- [127] K. M. Fox and R. Ferrari, “Heart rate: a forgotten link in coronary artery disease?,” *Nat. Publ. Gr.*, vol. 8, no. 7, pp. 369–379, 2011.
- [128] J. E. Dimsdale, “Psychological Stress and Cardiovascular Disease,” *J. Am. Coll. Cardiol.*, vol. 51, no. 13, pp. 1237–1246, 2008.
- [129] R. M. Carney, K. E. Freedland, and R. C. Veith, “Depression, the Autonomic Nervous System, and Coronary Heart Disease,” *Psychosom. Med.*, vol. 33, pp. S29–S33, 2005.
- [130] C. Tennant and L. Mclean, “The impact of emotions on coronary heart disease risk,” *J. Cardiovasc Risk*, vol. 8, pp. 175–183, 2001.
- [131] P. H. Wirtz and R. Von Känel, “Psychological Stress, Inflammation, and Coronary Heart Disease,” *Curr. Cardiol. Rep.*, vol. 19, p. 111, 2017.
- [132] C. L. Giddens, K. W. Barron, J. Byrd-Craven, K. F. Clark, and A. S. Winter, “Vocal Indices of Stress: A Review,” *J. Voice*, vol. 27, no. 3, p. 390.e21-390.e29, 2013.
- [133] B. Johannes *et al.*, “Non-linear function model of voice pitch dependency on physical and mental load,” *Eur. J. Appl. Physiol.*, vol. 101, pp. 267–276, 2007.
- [134] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” 2018.
- [135] J. Brownlee, “Tour of evaluation metrics for imbalanced classification,” *Imbalanced Classification*, 2020. [Online]. Available: <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>. [Accessed: 21-Apr-2020].

Anexo **A**

Parecer da Comissão de Ética do Centro
Hospitalar de Vila Nova de Gaia/Espinho

PARECER

174/2019-1

“Padroes acústicos de voz na detecção de doença coronária aterosclerótica ”

INSTITUIÇÃO /SERVIÇO – Cardiologia

INVESTIGADOR: Melissa Patricio

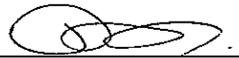
PARECER DA CES - emitido na reunião plenária de 29 / 08 / 2019

Nada a opor do ponto visto ético

Documentos analisados:

- resposta ao solicitado na reunião de 18-07-2019

 O Presidente da CES


(Enf. Ana Saraiva)

Autorizada
comunicar ao investigador e
ao departamento de parecer
ético e investigar
Diana Mota
Diretora Clínica 25/4/19/11

Remetido ao Secretariado da Comissão de Ética em 29 / 08 / 2019

Nada a opor
 25/9/19
CHVNG/E, E.P.E.
D. JOSÉ RIBEIRO
UGI Tórax e Circulação
Nº Mecanográfico: 1576

CHVNG/E, EPE
Secretariado UGI Tórax e Circulação
Doc. nº 1711/19
1ª Entrada 12/09/2019
2ª Entrada / /
3ª Entrada / /

Anexo **B**

Consentimento Informado Livre e
Esclarecido

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM INVESTIGAÇÃO

de acordo com a Declaração de Helsínquia³ e a Convenção de Oviedo⁴

Por favor, leia com atenção a seguinte informação. Se achar que algo está incorrecto ou que não está claro, não hesite em solicitar mais informações. Se concorda com a proposta que lhe foi feita, queira assinar este documento.

Título do estudo: Padrões Acústicos de Voz na deteção de Doença Coronária Aterosclerótica

Enquadramento: Recentemente na literatura científica, tem sido abordada a possibilidade de diagnosticar patologias cardíacas através da voz. Deste modo, o presente estudo tem como objetivo investigar esta possibilidade com o Português Europeu através da recolha de sinais de voz de pacientes diagnosticados com patologia cardíaca e de indivíduos saudáveis. Assim, no âmbito académico do Mestrado de Engenharia de Computação e Instrumentação Médica do Instituto Superior de Engenharia do Porto, sob orientação do Prof. Luís Coelho e do Dr. Nuno Ferreira, solicitamos a sua colaboração no presente estudo. Este estudo será realizado na Unidade de Diagnóstico e Intervenção Cardiovascular – Serviço de Cardiologia do Centro Hospitalar Vila Nova de Gaia/Espinho.

Explicação do estudo: A participação neste estudo envolverá a gravação de uma amostra de voz que posteriormente irá ser analisada tendo em conta diversos parâmetros acústicos. Adicionalmente, irão ser realizadas algumas questões antes do procedimento para despistar problemas de linguagem do participante. O método para a recolha dos dados é a entrevista sendo solicitado ao participante numa fase inicial, a realização da emissão sustentada das vogais /a/, /i/ e /u/ durante 5 segundos e posteriormente, a leitura de dois textos pré-definidos. Pretende-se apenas que esta recolha ocorra em apenas uma única sessão com uma duração média de aproximadamente 5 minutos. Adicionalmente, a recolha dos dados deverá ser realizada numa sala do serviço com pouco ruído ambiental, com os participantes sentados confortavelmente.

Condições e financiamento: A participação neste estudo é voluntária e desta forma, poderá a qualquer momento optar por desistir de participar, sem causar qualquer prejuízo ou consequência. Este projeto de investigação não oferece riscos ou prejuízos físicos, emocionais e/ou financeiros ao participante, não envolvendo custos acrescidos ao participante.

Confidencialidade e anonimato: Será garantida a confidencialidade e o anonimato das informações recolhidas ao longo de todo o processo. Nomeadamente, o sinal de voz recolhido apenas irá ser utilizado no cumprimento dos objetivos propostos no âmbito desta investigação.

Pedir consentimento para publicação do estudo: Após a conclusão do projeto de investigação, os resultados obtidos poderão ser divulgados no meio científico, sem que haja qualquer quebra de confidencialidade.

Agradecimentos: O investigador responsável está ao dispor do participante para responder a qualquer dúvida que considere que não foi totalmente esclarecida. Caso pretender contactar o investigador em caso de dúvidas: Méliça Patrício – melissapatricio.m@gmail.com

Identificação Investigador

Nome: Méliça Patrício

N.º Cédula Profissional:

Assinatura: _____



³ http://portal.arsnorte.min-saude.pt/portal/page/portal/ARSNorte/Comiss%C3%A3o%20de%20C%C3%89tica/Ficheiros/Declaracao_Helsinquia_2008.pdf

⁴ <http://dre.pt/pdf1sdip/2001/01/002A00/00140036.pdf>

Identificação da pessoa que pede o consentimento (preencher se for diferente do investigador)

Nome:

N.º Cédula Profissional:

Assinatura: _____

Declaro ter lido e compreendido este documento, bem como as informações verbais que me foram fornecidas pela/s pessoas/s que acima assina/m. Foi-me garantida a possibilidade de, em qualquer altura, recusar participar neste estudo sem qualquer tipo de consequências. Desta forma, aceito participar neste estudo e permito a utilização dos dados que de forma voluntária forneço, confiando em que apenas serão utilizados para esta investigação e nas garantias de confidencialidade e anonimato que me são dadas pelo/a investigador/a.

Nome: _____ Processo Clínico n.º _____

Assinatura: _____ Data: __ / __ / ____

**ESTE DOCUMENTO, COMPOSTO DE 2 PÁGINA/S, É FEITO EM DUPLICADO:
UMA VIA PARA O/A INVESTIGADOR/A, OUTRA PARA A PESSOA QUE CONSENTE**

Anexo **C**

Guião como instrumento para a recolha
de dados

INSTRUMENTO DE COLHEITA DE DADOS

Guião da Entrevista

O presente instrumento de colheita de dados contém o procedimento que será aplicado para a recolha de dados no âmbito desta investigação. O investigador estará ao dispor para esclarecer qualquer dúvida durante este procedimento. Desta forma, é solicitado ao participante:

1. Emissão sustentada da vogal /a/ durante aproximadamente de 5 segundos;
2. Emissão sustentada da vogal /i/ durante aproximadamente de 5 segundos;
3. Emissão sustentada da vogal /u/ durante aproximadamente de 5 segundos;
4. Leitura do seguinte texto foneticamente balanceado (45 seg. aproximadamente):

O Sol pode não nascer durante semanas ou até meses em alguns países nórdicos. Ali os Invernos são longos e rigorosos com noites muito compridas e dias curtos, pálidos e tristonhos. Só os pinheiros continuam verdes no meio das florestas geladas e despidas. Só eles parecem vivos no meio do grande silêncio imóvel e branco. Tudo parece exausto e sufocado pelo olhar da escuridão porque não há como escapar dela, nem da depressão de Inverno que surge quando o Sol não nasce.

5. Leitura do seguinte texto foneticamente balanceado (35 seg. aproximadamente):

Ao ver as maravilhas da Primavera através da minha janela, como as ruas ricas em flores de cores vibrantes, sinto-me a pessoa mais feliz do mundo. Pelas janelas abertas entra a luz brilhante desta estação. Agora, os dias começam a ficar cada vez mais longos e quentes. É uma das épocas ideais para passear, conhecer e até aproveitar para se fazer um grande e divertido piquenique com a família e os amigos mais próximos.

Anexo **D**

Evolução da ‘*accuracy*’ em cada categoria CAD-RADS

D.1 Desempenho dos processos de treino no que diz respeito à métrica ‘accuracy’ nos dados recolhidos pela emissão das vogais sustentadas

Na figura D.1 visualiza-se a evolução dos desempenhos obtidos na avaliação individual de cada categoria CAD-RADS vs indivíduos saudáveis: na primeira coluna da figura, as amostras incluíam apenas indivíduos das categorias CAD-RADS 0 e 1; na segunda coluna apenas indivíduos das categorias CAD-RADS 0 e 2; e na última coluna, apenas as categorias CAD-RADS 0 e superiores ou iguais à categoria CAD-RADS 3.

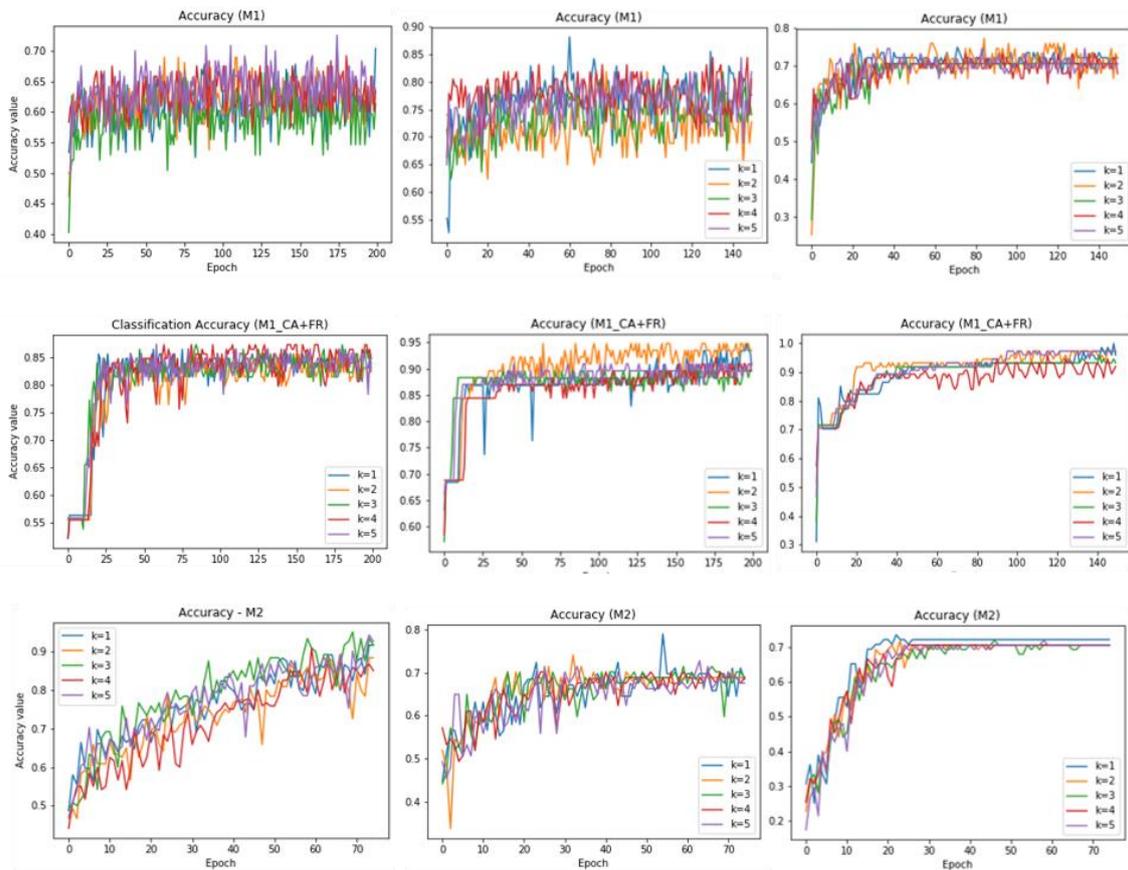


Figura D.1: Evolução do desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na emissão das vogais sustentadas, quanto à avaliação individual das categorias CAD-RADS.

D.2 Desempenho dos processos de treino no que diz respeito à métrica ‘accuracy’ nos dados recolhidos pela leitura dos textos foneticamente balanceados

Na figura D.2 visualiza-se a evolução dos desempenhos obtidos na avaliação individual de cada categoria CAD-RADS vs indivíduos saudáveis: na primeira coluna da figura, as amostra incluíam apenas indivíduos das categorias CAD-RADS 0 e 1; na segunda coluna apenas indivíduos das categorias CAD-RADS 0 e 2; e na última coluna apenas das categorias CAD-RADS 0 e 3.

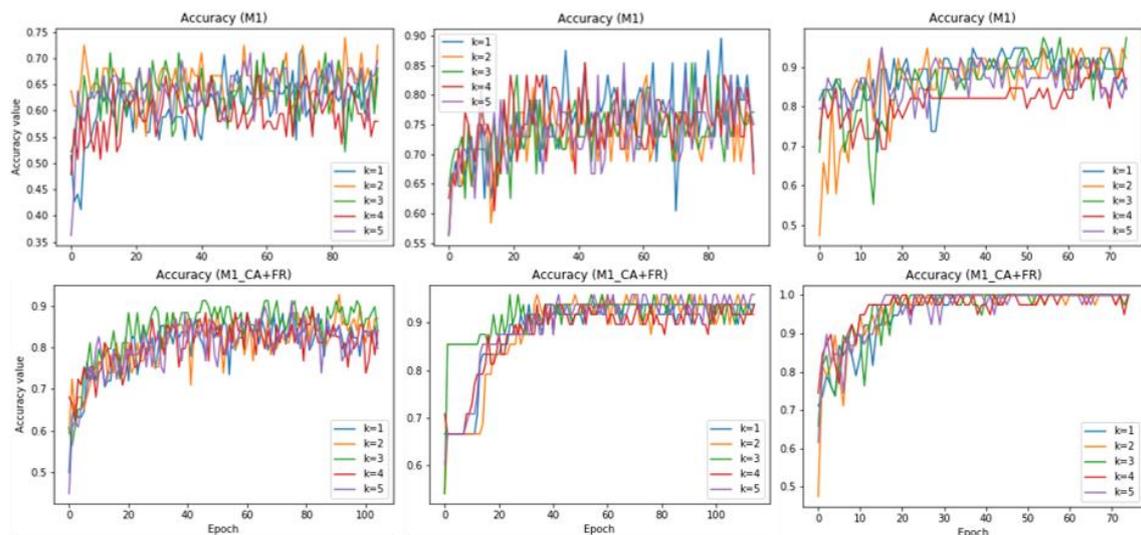


Figura D.2: Evolução do desempenho da ‘accuracy’ durante os processos de treino da validação cruzada *k-fold* com os dados recolhidos na leitura dos textos, quanto à avaliação individual das categorias CAD-RADS.