



UNIVERSIDADE D  
COIMBRA

Vânia Sofia Martins Dias

CLASSIFICAÇÃO DE SONS RESPIRATÓRIOS  
ADVENTÍCIOS EM CRIANÇAS

Dissertação no âmbito do Mestrado Integrado em Engenharia Física - Ramo Instrumentação,  
orientada pelo Professor Doutor Rui Pedro Pinto de Carvalho e Paiva  
e pela Professora Doutora Alda Sofia Pires de Dias Marques,  
apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia  
da Universidade de Coimbra.

Fevereiro de 2020





Vânia Sofia Martins Dias

# Classificação de sons respiratórios adventícios em crianças

Dissertação apresentada ao Departamento de Física da Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Física no ramo de Instrumentação

Orientadores:

Professor Doutor Rui Pedro Paiva (Universidade Coimbra)  
Professora Doutora Alda Marques (Universidade de Aveiro)

Júri:

Professor Doutor Rui Pedro Pinto de Carvalho e Paiva  
Professora Doutora Filipa Isabel Gouveia de Melo Borges Belo Soares  
Professor Doutor Jorge Manuel Oliveira Henriques  
Professor Doutor Fernando Manuel dos Santos Perdigão

**Coimbra, 2020**



Esta cópia da tese é fornecida na condição de que quem a consulta reconhece que os direitos de autor são pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada sem a referência apropriada.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.



# Agradecimentos

Uma tese de mestrado é um longo caminho, que inclui inúmeros desafios, tristezas, incertezas, alegrias, conquistas e muitos contratempos. É o culminar de 5 anos de trabalho intensivo, onde este esforço final levará à atribuição do grau de mestre. Trilhar este caminho só foi possível com o apoio, energia e força de várias pessoas às quais estou eternamente grata.

Em primeiro lugar, ao meu orientador Professor Doutor Rui Pedro Paiva que me deu a conhecer este projeto, pela disponibilidade e apoio que demonstrou, bem como a sua colaboração na resolução de problemas que foram surgindo ao longo do trabalho e por todas as palavras de incentivo.

À Professora Doutora Alda Marques (ESSUA), pela disponibilidade e ajuda no esclarecimento de dúvidas. E ainda, à Professora Doutora Ana Oliveira (ESSUA), pelos esclarecimentos relativamente à base de dados, à aquisição do som respiratório bem como às anotações registadas.

Ao Bruno Rocha pelos esclarecimentos e auxílios dados no âmbito dos projetos já desenvolvidos pelo grupo de investigação.

Aos meus pais, Sandra e Filipe, e à minha irmã, Liliana, pelo apoio incondicional nesta fase e em todas as outras da minha vida. Especialmente, à minha mãe, que sempre me ouviu e encorajou naqueles momentos mais difíceis, mas também pela alegria com que celebra cada vitória minha, permitindo-me chegar ao fim deste percurso.

Ao meu namorado Rodrigo, pela partilha, companheirismo e apoio incondicional, assim como a enorme motivação nesta jornada.

À minha melhor amiga Andreia, que me acompanhou nesta e em tantas outras etapas, pela sua paciência, apoio e amizade.

Aos meus avós e padrinhos, pela motivação e apoio nesta caminhada, bem como pelas palavras carinhosas, que também elas foram essenciais neste percurso.

Por último, o meu agradecimento a todos aqueles que contribuíram para a concretização desta dissertação, estimulando-me intelectual e emocionalmente.



# Resumo

As doenças respiratórias assumem-se como uma das principais causas de morte, originando problemas de saúde, sociais e económicos em todo o mundo. Nesse sentido têm sido investigados métodos para melhorar o diagnóstico, acompanhamento e prognóstico de pacientes com essas doenças. A análise computacional de sons respiratórios é uma das formas de detetar as patologias respiratórias.

Nesta dissertação pretende-se avaliar a capacidade de classificação de sons adventícios em crianças, realizando um estudo de “*feature engineering*” através das *features* mais utilizadas no estado de arte, e ainda comparar essa classificação à obtida para a população adulta. A metodologia adotada consistiu em estratificar por idades a base de dados da população infantil e utilizar 42 *features* em conjunto com 4 algoritmos de *machine learning* para classificar sibilâncias, ferveores e sons normais. Procedeu-se à seleção de *features* por meio de 3 algoritmos distintos a fim de avaliar o desempenho das *features*. A combinação das *features* mais relevantes e a otimização de parâmetros do SVM permitiu obter uma *f-measure* média de 0,794; 0,891; 0,850; 0,882; 0,933; 0,842 para as amostras 0-2 anos, 3-6 anos, todos os sons de crianças, apenas ferveores (crianças), apenas sibilâncias (crianças) e adultos, respetivamente.

Em suma, os resultados foram bastante promissores, comprovando-se a grande potencialidade dessas *features* para a classificação de sons adventícios em crianças. Ficou também demonstrado que a estratificação por idades na população infantil é significativa e melhora os resultados da classificação. Como trabalhos futuros propõe-se testar a classificação numa base de dados maior e estendê-la a outros sons adventícios, bem como averiguar a influência do local de aquisição do som respiratório e investigar as diferenças nas *features* selecionadas para a população infantil e adulta.

**PALAVRAS-CHAVE:** *features*, crianças, sons respiratórios adventícios, SVM, classificação



# Abstract

Respiratory diseases are considered one of the main causes of death, causing health, social and economic problems worldwide. In this sense, several methods to improve diagnosis, monitoring and prognostic of patients with these diseases have been investigated. The computational analysis of respiratory sounds is one of the ways to detect respiratory pathologies.

This dissertation aims to evaluate the ability to classify adventitious sounds in children, carrying out a feature engineering study using the most used features in the state of the art, and also to compare this classification to that obtained for the adult population. The adopted methodology consisted of separating the children's population database by age and using 42 features along with 4 machine learning algorithms to classify wheezes, crackles and normal sounds. The selection of features was carried out using 3 different algorithms in order to evaluate the performance of features. The combination of the most relevant features and the optimization of SVM parameters allowed to obtain an average f-measure of 0.794; 0.891; 0.850; 0.882; 0.933; 0.842 for samples 0-2 years, 3-6 years, all sounds of children, only crackles (children), only wheezes (children) and adults, respectively.

In sum, the results were very promising, proving the great potential of these features for the classification of adventitious sounds in children. It was also demonstrated that the separation by age in the child population is significant and improves the results of the classification. As future work, it is proposed to test the classification in a larger database and extend it to other adventitious sounds, as well as to investigate the influence of the place of acquisition of the respiratory sound and investigate the differences in the select features for the child and adult population.

**KEYWORDS:** features, children, adventitious respiratory sounds, SVM, classification



# Lista de Siglas

**ANN** *Artificial Neural Network*

**AR** *Autoregressive Modeling*

**ATS** *American Thoracic Society*

**CAS** *Continuous Adventitious Sounds* - Sons Adventícios Contínuos

**2CD** *Two - Cycle Duration* - Duração de 2 ciclos

**CNN** *Convolutional Neural Network*

**CORSA** *Computerized Respiratory Sound Analysis*

**DAS** *Discontinuous Adventitious Sounds* - Sons Adventícios Descontínuos

**DPOC** *Doença Pulmonar Obstrutiva Crónica*

**DWT** *Discrete Wavelet Transform* - Transformada Discreta *Wavelet*

**ESSUA** *Escola Superior de Saúde da Universidade de Aveiro*

**FD** *Fractal Dimension* - Dimensão Fractal

**FFT** *Fast Fourier Transform*

**FT** *Fourier Transform*

**GBW** *Gaussian Bandwidth* - Largura de Banda de Gauss

**GMM** *Gaussian Mixture Models*

**IDW** *Initial Deflection Width* - Largura de Deflexão Inicial

**KFD** *Katz Fractal Dimension* - Dimensão Fractal Katz

**k-NN** *k-Nearest Neighbors*

**LDW** *Largest Deflection Width* - Duração da Deflexão de Amplitude Máxima na Onda

**LPC** *Linear Predictive Coding*

**LR** *Logistic Regression* - Regressão Logística

- LVQ** *Learning Vector Quantization*
- MAR** *Multivariate Autoregressive Modelling*
- MFCCs** *Mel-Frequency Cepstral Coefficients*
- MLP** *Multi-Layer Perceptron*
- 1-NN** *One Nearest-Neighbor*
- PCA** *Principal Component Analysis* - Análise das Componentes Principais
- PF** *Peak Frequency* - Frequência de pico
- PSD** *Power Spectral Density*
- QP** *Quadratic Programming* - Programação quadrática
- RF** *Random Forest*
- SMO** *Sequential Minimal Optimization* - Otimização Mínima Sequencial
- SNR** *Signal-to-Noise Ratio* - Relação Sinal-Ruído
- SR** Som Respiratório
- STFT** *Short-time Fourier Transform*
- SVM** *Support Vector Machine*
- TD** Total Duration - Duração Total
- TDW** *Total Deflection Width* - Largura Total de Deflexão
- TF** Tempo-Frequência
- VFD** *Variance Fractal Dimension* - Variância Dimensão Fractal
- WPST-NST** *Wavelet Packet Transform-Based Stationary-Nonstationary*
- WS-SS** *Wheezes Signature in the Spectrogram Space* - Sibilâncias no espaço do espectrograma
- WT** *Wavelet Transform*

# Lista de Figuras

1.1	As 10 principais causas de morte no mundo em 2016. . . . .	1
2.1	Relação entre os termos SRs, sons adventícios, sons pulmonares e sons fisiológicos/da respiração. . . . .	8
2.2	a) Representação do SR normal, b) Sibilância, c) fervor; e os respectivos espectrogramas. . . . .	9
2.3	Formas de onda para as sibilâncias. . . . .	10
2.4	Ilustração dos parâmetros do fervor: 2CD, IDW, LDW e TD. . . . .	11
2.5	Representação de um fervor <i>coarse</i> e <i>fine</i> . . . . .	11
3.1	Algoritmo proposto por Pinho et al., para detetar ferveiros. . . . .	35
3.2	Posições para a aquisição de sons (vermelho). Para cada voluntário, selecionaram os sons adquiridos nas duas posições em que os sons adventícios/normais eram mais audíveis. . . . .	36
4.1	Visão geral das <i>features</i> disponíveis na <i>MIRtoolbox</i> . . . . .	45
4.2	Representação <i>spectral centroid</i> . . . . .	46
4.3	Representação <i>spectral brightness</i> . . . . .	46
4.4	Representação <i>spectral spread</i> . . . . .	47
4.5	Representação <i>spectral skewness</i> . . . . .	47
4.6	Representação <i>spectral kurtosis</i> . . . . .	48
4.7	Representação <i>spectral rolloff</i> 85. . . . .	48
4.8	Representação <i>spectral roughness</i> . . . . .	50
4.9	Representação <i>zero-crossing rate</i> . . . . .	51
4.10	Representação <i>key clarity</i> . . . . .	52
4.11	Representação $F_0$ <i>cepstrum</i> . . . . .	53
4.12	Representação <i>inharmonic</i> do $F_0$ <i>cepstrum</i> . . . . .	53
4.13	Sibilâncias no espaço do espectrograma (WS-SS). . . . .	55
4.14	O classificador Naive Bayes descrito como uma rede bayesiana na qual os atributos preditivos ( $X_1, X_2, \dots, X_k$ ) são condicionalmente independentes dado o atributo de classe (C). . . . .	56
4.15	Exemplo de classificação iBk: a amostra teste (ponto verde) deve ser classificada em quadrados azuis ou em triângulos vermelhos. Se $k = 3$ (círculo mais pequeno-linha a cheio), é classificado como triângulo vermelho porque existem 2 triângulos e apenas 1 quadrado dentro do círculo interno. Se $k = 5$ (círculo a tracejado) é classificado como quadrado azul (3 quadrados vs 2 triângulos dentro do círculo externo). 57	57

4.16 Painel “ <i>Preprocess</i> ” do software Weka, após o ficheiro .arff ter sido importado. . . . .	61
4.17 Painel “ <i>Classify</i> ” do Weka onde se encontra o output para o classificador SMO. . . . .	62
4.18 Painel “ <i>Select attributes</i> ” do Weka onde é possível observar o resultado para o método <i>CfsSubsetEval + Best First</i> . . . . .	65

# Lista de Tabelas

2.1	Resumo dos sons adventícios e possíveis patologias associadas. . . . .	13
4.1	Descrição da base de dados utilizada para a população infantil. . . . .	40
5.1	Classificação com 210 <i>features</i> para a amostra 0-2 anos. . . . .	70
5.2	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.1, para a amostra 0-2 anos. . . . .	71
5.3	Classificação com 210 <i>features</i> para a amostra 3-6 anos. . . . .	72
5.4	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.3, para a amostra 3-6 anos. . . . .	72
5.5	Classificação com 210 <i>features</i> para a amostra todos os sons de crianças. 73	
5.6	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.5, para a amostra todos os sons de crianças. . . . .	74
5.7	Classificação com 210 <i>features</i> para a amostra que contém apenas ferveores. . . . .	75
5.8	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.7, para a amostra que contém apenas ferveores. . . . .	76
5.9	Classificação com 210 <i>features</i> para a amostra que contém apenas sibilâncias. . . . .	77
5.10	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.9, para a amostra que contém apenas sibilâncias. . . . .	77
5.11	Diferentes métodos de seleção de <i>features</i> para a amostra 0-2 anos. . .	79
5.12	95% da classificação máxima com 42 atributos para a amostra 0-2 anos. 80	
5.13	Diferentes métodos de seleção de <i>features</i> para a amostra 3-6 anos. . .	82
5.14	95% da classificação máxima com 50 atributos para a amostra 3-6 anos. 83	
5.15	Diferentes métodos de seleção de <i>features</i> para a amostra que contém todos os sons de crianças. . . . .	85
5.16	95% da classificação máxima com 18 atributos para a amostra que contém todos os sons de crianças. . . . .	86
5.17	Diferentes métodos de seleção de <i>features</i> para a amostra apenas ferveores. . . . .	88

5.18	95% da classificação máxima com 6 atributos para a amostra que contém apenas fervores. . . . .	89
5.19	Diferentes métodos de seleção de <i>features</i> para a amostra apenas sibilâncias. . . . .	91
5.20	95% da classificação máxima com 3 atributos para a amostra que contém apenas sibilâncias. . . . .	92
5.21	Classificação após otimização de parâmetros para a amostra 0-2 anos.	93
5.22	Classificação após otimização de parâmetros para a amostra 3-6 anos.	93
5.23	Classificação após otimização de parâmetros para a amostra que contém todos os sons de crianças. . . . .	93
5.24	Classificação após otimização de parâmetros para a amostra que contém apenas fervores. . . . .	94
5.25	Classificação com 210 <i>features</i> para a população adulta. . . . .	95
5.26	Testes de significância estatística com $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.25, para a população adulta. . . . .	96
5.27	Diferentes métodos de seleção de <i>features</i> para a população adulta. . . . .	97
5.28	95% da classificação máxima com 42 atributos para a população adulta.	98
5.29	Classificação após otimização de parâmetros para a população adulta.	98
A.1	Resumo dos métodos de extração de <i>features</i> e classificação para SRs em crianças presentes na literatura. . . . .	118
A.2	Resumo dos métodos de extração de <i>features</i> e classificação para SRs em adultos presentes na literatura. . . . .	121
B.1	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.1, para a amostra 0-2 anos. . . . .	127
B.2	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.3, para a amostra 3-6 anos. . . . .	128
B.3	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.5, para a amostra todos os sons de crianças. . . . .	128
B.4	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.7, para a amostra que contém apenas fervores. . . . .	129
B.5	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na tabela 5.9, para a amostra que contém apenas sibilâncias. . . . .	129
B.6	Testes de significância estatística com $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.25, para a população adulta. . . . .	130
C.1	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval + Best First</i> para a amostra 0-2 anos. . . . .	131

C.2	<i>Features</i> selecionadas pelo algoritmo <i>WrapperSubsetEval</i> + <i>Best First</i> para a amostra 0-2 anos. . . . .	132
C.3	50 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a amostra 0-2 anos. . . . .	132
C.4	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval</i> + <i>Best First</i> para a amostra 3-6 anos. . . . .	133
C.5	<i>Features</i> selecionadas pelo algoritmo <i>WrapperSubsetEval</i> + <i>Best First</i> para a amostra 3-6 anos. . . . .	133
C.6	80 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a amostra 3-6 anos. . . . .	134
C.7	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval</i> + <i>Best First</i> para a amostra todos os sons de crianças. . . . .	135
C.8	<i>Features</i> selecionadas pelo algoritmo <i>WrapperSubsetEval</i> + <i>Best First</i> para a amostra todos os sons de crianças. . . . .	136
C.9	70 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a amostra todos os sons de crianças. . . . .	136
C.10	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval</i> + <i>Best First</i> para a amostra que contém apenas ferveores. . . . .	137
C.11	<i>Features</i> selecionadas pelo algoritmo <i>WrapperSubsetEval</i> + <i>Best First</i> para a amostra que contém apenas ferveores. . . . .	138
C.12	50 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a amostra que contém apenas ferveores. . . . .	138
C.13	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval</i> + <i>Best First</i> para a amostra que contém apenas sibilâncias. . . . .	139
C.14	<i>Features</i> selecionadas pelo algoritmo <i>WrapperSubsetEval</i> + <i>Best First</i> para a amostra que contém apenas sibilâncias. . . . .	139
C.15	60 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a amostra que contém apenas sibilâncias. . . . .	140
C.16	<i>Features</i> selecionadas pelo algoritmo <i>CfsSubsetEval</i> + <i>Best First</i> para a população adulta. . . . .	141
C.17	185 primeiros atributos selecionados pelo algoritmo <i>ReliefFAttributeEval</i> + <i>Ranker</i> para a população adulta. . . . .	141
A.1	Recomendações CORSA para a aquisição de sons respiratórios. . . . .	147
B.1	<i>Features</i> propostas por Mendes et al. e Rocha et al. . . . .	149
C.1	Resultados de desempenho do método proposto por Rocha et al. . . . .	151



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização . . . . .	1
1.2	Objetivos e abordagens . . . . .	3
1.3	Organização da tese . . . . .	4
<b>2</b>	<b>Análise de sons respiratórios: contexto e definições</b>	<b>7</b>
2.1	Definições de sons respiratórios . . . . .	7
2.1.1	Sons adventícios . . . . .	8
2.2	Importância do diagnóstico através de sons adventícios . . . . .	12
2.3	Auscultação dos sons respiratórios em crianças vs adultos . . . . .	15
2.4	Análise computacional do som respiratório . . . . .	15
2.5	Definição de <i>feature</i> . . . . .	16
<b>3</b>	<b>Revisão da Literatura</b>	<b>17</b>
3.1	Métodos de extração de <i>features</i> e classificação . . . . .	17
3.1.1	Crianças . . . . .	17
3.1.2	Adultos . . . . .	25
3.2	Limitações do estado de arte . . . . .	38
<b>4</b>	<b>Métodos</b>	<b>39</b>
4.1	Base de dados . . . . .	39
4.1.1	População infantil . . . . .	39
4.1.2	População adulta . . . . .	43
4.2	<i>Features</i> extraídas . . . . .	44
4.2.1	<i>Features</i> da <i>MIRtoolbox</i> . . . . .	44
4.2.2	Outras <i>features</i> . . . . .	54
4.3	Classificadores . . . . .	55
4.3.1	SMO . . . . .	55
4.3.2	Naive Bayes . . . . .	56
4.3.3	iBk . . . . .	56
4.3.4	J48 . . . . .	57
4.4	Pré-processamento e extração de <i>features</i> no Matlab . . . . .	58
4.5	Processamento no Weka . . . . .	60
4.6	Testes de significância estatística . . . . .	64
4.7	Seleção de <i>features</i> . . . . .	64
4.8	Otimização de parâmetros . . . . .	66

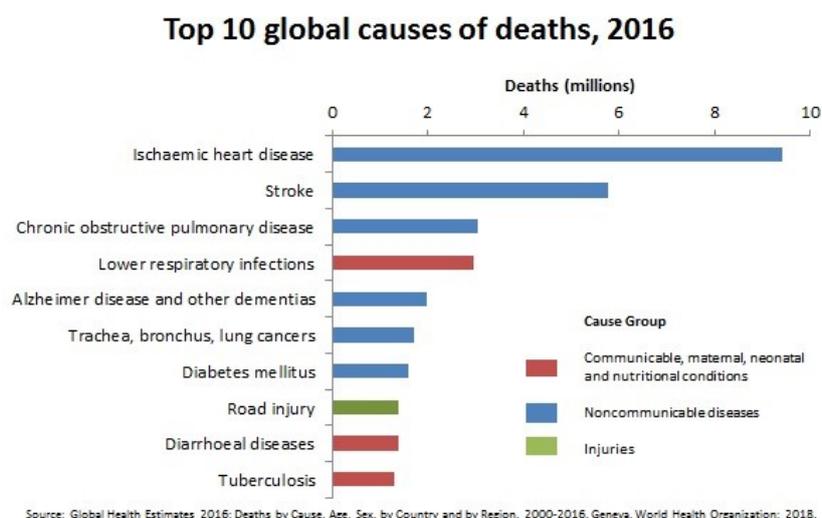
<b>5</b>	<b>Resultados</b>	<b>69</b>
5.1	População infantil . . . . .	69
5.1.1	Classificação com 210 <i>features</i> . . . . .	69
5.1.2	Seleção de <i>features</i> . . . . .	78
5.1.3	Otimização de parâmetros . . . . .	92
5.2	População adulta . . . . .	94
5.2.1	Classificação com 210 <i>features</i> . . . . .	94
5.2.2	Seleção de <i>features</i> . . . . .	96
5.2.3	Otimização de parâmetros . . . . .	98
<b>6</b>	<b>Discussão de Resultados</b>	<b>99</b>
<b>7</b>	<b>Conclusão e trabalho futuro</b>	<b>101</b>
	<b>Referências</b>	<b>103</b>
	<b>Apêndices</b>	<b>114</b>
<b>A</b>	<b>Tabela de resumo dos métodos de extração de <i>features</i> e classificação dos sons respiratórios para crianças e adultos.</b>	<b>117</b>
A.1	Crianças . . . . .	118
A.2	Adultos . . . . .	121
<b>B</b>	<b>Tabelas dos testes de significância estatística para <math>\alpha=0,05</math></b>	<b>127</b>
B.1	Amostra 0-2 anos . . . . .	127
B.2	Amostra 3-6 anos . . . . .	128
B.3	Amostra todos os sons de crianças . . . . .	128
B.4	Amostra apenas ferveores (crianças) . . . . .	129
B.5	Amostra apenas sibilâncias (crianças) . . . . .	129
B.6	Adultos . . . . .	130
<b>C</b>	<b>Tabelas seleção de <i>Features</i></b>	<b>131</b>
C.1	Amostra 0-2 anos . . . . .	131
C.2	Amostra 3-6 anos . . . . .	133
C.3	Amostra todos os sons crianças . . . . .	135
C.4	Amostra apenas ferveores (crianças) . . . . .	137
C.5	Amostra apenas sibilâncias (crianças) . . . . .	139
C.6	Adultos . . . . .	141
	<b>Anexos</b>	<b>147</b>
<b>A</b>	<b>Recomendações CORSA para a aquisição de sons respiratórios</b>	<b>147</b>
<b>B</b>	<b><i>Features</i> propostas por Mendes et al. e Rocha et al.</b>	<b>149</b>
<b>C</b>	<b>Resultados de desempenho do método proposto por Rocha et al.</b>	<b>151</b>

# Introdução

Este capítulo introduz o problema a que esta dissertação pretende dar resposta e a visão global do mesmo. Para além disto, inclui também os objetivos do projeto, bem como os resultados obtidos e uma descrição sumária da organização do documento.

## 1.1 Contextualização

As patologias respiratórias têm vindo a apresentar taxas de mortalidade e morbilidade cada vez maiores. A Doença Pulmonar Obstrutiva Crónica (DPOC) e as infeções do trato respiratório inferior apresentam-se, respetivamente, como a terceira e quarta causas de morte no mundo [World Health Organization, 2018] e representam claramente uma carga significativa a nível económico, social e de saúde [Marques et al., 2014]. A Figura 1.1 mostra as 10 principais causas de morte em todo o mundo em 2016, cujos dados provêm de um estudo estatístico concluído em 2018.



**Figura 1.1:** As 10 principais causas de morte no mundo em 2016 [World Health Organization, 2018].

As estatísticas mostram que Portugal segue a tendência dos países desenvolvidos, aumentando a esperança média de vida devido ao acesso a uma medicina evoluída e a medidas terapêuticas inovadoras e eficazes. Este aumento na longevidade tem como consequência um aumento dos doentes crónicos. As doenças respiratórias em Portugal, com destaque nas pneumonias, são desde 2015 a 3<sup>a</sup> causa de morte, logo após o cancro e as doenças cardiovasculares. As 3 principais causas de internamento em 2016 por doenças respiratórias foram as pneumonias (40 345), a DPOC (7 864) e os cancros (5 541) e tendem a manter-se futuramente. Os hábitos tabágicos e a reduzida atividade física influenciam negativamente a evolução das doenças respiratórias crónicas [Observatório Nacional das Doenças Respiratórias, 2018].

Relativamente às crianças estima-se que cerca de 1 milhão com idades inferiores a 5 anos morrem anualmente devido a infeções agudas do trato respiratório inferior. Apenas a pneumonia é responsável pela morte de 1,6 milhões de crianças a nível mundial. Ou seja, uma taxa de mortalidade maior comparativamente ao HIV, malária e tuberculose [Emmanouilidou et al., 2018, Gross et al., 2000]. Outras doenças respiratórias como constipação, gripe, pneumonia e bronquite assumem-se como as principais causas de mortalidade infantil no mundo [Song, 2015]. Sabe-se ainda que a incidência de infeções do trato respiratório inferior em crianças pode potencializar o desenvolvimento de doenças respiratórias crónicas na idade adulta [Merkus, 2003].

O Observatório Nacional das Doenças Respiratórias estima que em 2020 as doenças respiratórias sejam responsáveis por cerca de 12 milhões de mortes no mundo [Observatório Nacional das Doenças Respiratórias, 2018].

O diagnóstico das doenças respiratórias, bem como a sua progressão são geralmente avaliadas pela auscultação, provas de função respiratória e técnicas de imagiologia médica. O estetoscópio, inventado no início do século XIX, é a ferramenta mais utilizada para auscultação, bem como para a avaliação e diagnóstico de anormalidades respiratórias ou infeções [Emmanouilidou et al., 2018]. Isto porque o estetoscópio é simples, não invasivo, económico, prático e aplicável a todas as populações e ambientes. Contudo, apesar da sua universalidade na prática clínica, é limitado pela subjetividade na auscultação, ou seja, a interpretação do som respiratório (SR) depende das propriedades do estetoscópio, da capacidade auditiva e da experiência clínica dos profissionais de saúde, tal como da sua capacidade em memorizar os padrões sonoros [Pinho et al., 2015, Puder et al., 2016]. Para além disso, a presença de ruído de fundo afeta a qualidade das auscultações, podendo ocultar a presença de anormalidades no sinal. As limitações referidas são ainda mais evidenciadas na pediatria devido à movimentação das crianças, ao choro, ao desen-

volvimento do seu sistema respiratório e à falta de cooperação, e ainda em países em desenvolvimento, nos quais os contextos clínicos sofrem de várias restrições, em particular quanto ao número de profissionais de saúde e às técnicas disponíveis [Ellington et al., 2012, Ellington et al., 2014].

A radiografia ao tórax é uma técnica bastante usada para detetar e monitorizar as infeções respiratórias. No entanto, o raio-x apresenta várias limitações uma vez que nem sempre está disponível em todos os contextos clínicos, não é portátil, é caro, apresenta altos níveis de subjetividade entre os observadores e envolve doses consideráveis de radiação. Todos esses fatores impedem a monitorização dos pacientes com a frequência necessária [Cherian et al., 2005, Neuman et al., 2012]. Outros meios para avaliar os SRs, como a espirometria, dependem da motivação e da cooperação do paciente, ou outros, como as análises de gases sanguíneos e imagens não se encontram disponíveis em todos os locais clínicos [Marques et al., 2014].

Presentemente, não existe um acompanhamento continuado das patologias respiratórias, levando a episódios frequentes de descompensação com hospitalização, os quais têm impacto significativo na saúde e qualidade de vida dos pacientes, na baixa de produtividade por perda de dias de trabalho e em elevados custos de tratamento ambulatorio e internamento.

É sabido que cada episódio de descompensação é precedido por uma fase de incremento gradual que varia entre várias horas e vários dias antes de ocorrer o pico da descompensação. Assim, a deteção precoce dessa tendência pode prevenir a sua ocorrência e levar a um quadro clínico significativamente mais leve. Nesse sentido, são requeridos sistemas de diagnóstico precoce de descompensações. Neste contexto, um sintoma bastante importante são os SRs adventícios, os quais estão frequentemente associados a desordens respiratórias.

Para colmatar as limitações referidas anteriormente na análise de SRs, na última década muita pesquisa foi realizada a fim de aperfeiçoar o diagnóstico precoce e a monitorização de pacientes com doenças respiratórias. A análise computadorizada do SR consiste em gravar SRs com um dispositivo eletrónico e analisá-los/classificá-los objetivamente, com base em técnicas avançadas de processamento de sinal, revelando-se uma alternativa bastante promissora para o diagnóstico e monitorização de doenças respiratórias [Gross et al., 2000, Pinho et al., 2015].

## 1.2 Objetivos e abordagens

A presente dissertação tem como objetivo geral avaliar a capacidade de classificação de sons adventícios em crianças, mais especificamente:

- Realizar um estudo de “*feature engineering*” versando as *features* mais utilizadas no estado de arte;
- Classificar usando diferentes técnicas de análise de sinais áudio e aprendizagem computacional (*machine learning*/reconhecimento de padrões);
- Explorar diferentes abordagens para organizar os dados da população infantil;
- Comparar a classificação de sons adventícios em crianças vs adultos;
- Analisar criticamente os resultados.

A fim de cumprir estes objetivos, utilizou-se a base de dados do Laboratório de Investigação e Reabilitação Respiratória - Lab3R da Escola Superior de Saúde da Universidade de Aveiro (ESSUA), e as *features* do estado de arte, em particular da MIRtoolbox, e testaram-se classificadores de diversos tipos, e.g., probabilísticos (Naive Bayes), árvores de decisão (J48<sup>1</sup>), funcionais (SMO<sup>2</sup>) e métodos baseados em instâncias (iBk<sup>3</sup>). Obtiveram-se resultados bastante promissores, nomeadamente valores da *f-measure* de 0,794; 0,891; 0,850; 0,882; 0,933; 0,842 para as amostras 0-2 anos, 3-6 anos, todos os sons de crianças, apenas ferveiros (crianças), apenas sibilâncias (crianças) e adultos, respetivamente.

Salienta-se ainda que numa fase inicial um dos objetivos era atualizar um sistema de aquisição de SRs - a ferramenta CLASS<sup>4</sup> para que conseguisse adquirir sons multicanal. Este objetivo foi concretizado através de uma solução simples, nomeadamente um software de gravação de som - Reaper.

### 1.3 Organização da tese

Esta dissertação encontra-se organizada em 7 capítulos da seguinte forma:

- Capítulo 1: Contém a introdução do documento;
- Capítulo 2: Contém definições e contexto dos sons respiratórios;
- Capítulo 3: Revisão de literatura sobre os diferentes métodos de extração de *features* e classificadores usados tanto para crianças como adultos;

---

<sup>1</sup>Usou-se a designação do software Weka, que corresponde à implementação específica desta plataforma. Equivalente a C4.5.

<sup>2</sup>Usou-se a designação do software Weka, que corresponde à implementação específica desta plataforma. Equivalente a SVM.

<sup>3</sup>Usou-se a designação do software Weka, que corresponde à implementação específica desta plataforma. Equivalente a k-NN.

<sup>4</sup>Software de gravação e anotação de SRs [Semedo et al., 2015].

- Capítulo 4: Metodologia usada para alcançar os objetivos desta dissertação;
- Capítulo 5: Apresenta os resultados obtidos;
- Capítulo 6: Contém a discussão dos resultados;
- Capítulo 7: Conclusão do trabalho desenvolvido e proposta de trabalho futuro.



# Análise de sons respiratórios: contexto e definições

Este capítulo contém a definição de determinados conceitos relacionados com os sons respiratórios (SRs), bem como a sua análise, que serão mencionados ao longo desta dissertação.

## 2.1 Definições de sons respiratórios

A distinção entre SRs fisiológicos (normais) e patológicos/adventícios (como fer-vores<sup>1</sup>, sibilâncias,<sup>2</sup> ...) reveste-se de grande importância para a obtenção de um diagnóstico médico preciso. Os SRs permitem então perceber a integridade da via aérea.

Os SRs definem-se como todos os sons relacionados com a respiração, incluindo sons fisiológicos, adventícios, a tosse, entre outros (Figura 2.1). Os sons da fala que surgem durante a respiração não se incluem nos SRs [Sovijarvi et al., 2000, Reichert et al., 2008].

Os sons pulmonares são todos os SRs ouvidos ao longo da via aérea, incluindo sons fisiológicos e adventícios. Todos os sons são gerados pelo fluxo de ar no trato respiratório, durante a inspiração e expiração, e podem ser gravados no tórax, na traqueia ou na boca. Caracterizam-se por um amplo espectro acústico com uma faixa de frequências dependente do local onde se adquire o som. Esta variação ocorre devido à filtração natural por parte do parênquima pulmonar que, consoante a sua densidade, faz variar a frequência dos diferentes sons [Pasterkamp et al., 2016].

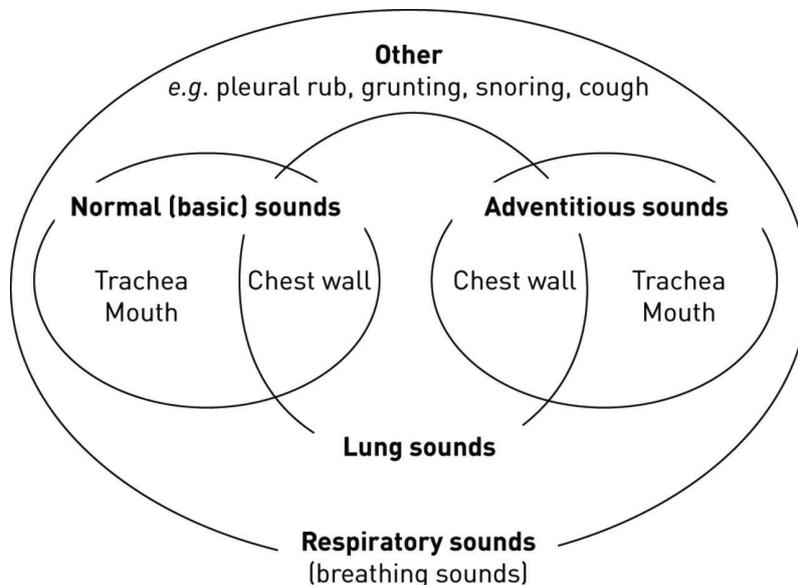
Os sons fisiológicos ocorrem na ausência de qualquer processo patológico e como tal excluem os sons adventícios [Sovijarvi et al., 2000, Reichert et al., 2008]. Como explicado acima, o som varia consoante a constituição do parênquima pulmonar. Desde o nascimento até à idade adulta, as estruturas respiratórias crescem em

---

<sup>1</sup>Corresponde ao termo *crackles* em inglês.

<sup>2</sup>Corresponde ao termo *wheezes* em inglês.

número e tamanho. Como tal é esperado que a auscultação a um bebé seja diferente da realizada a um adulto. Esta questão será abordada na secção 2.3.



**Figura 2.1:** Relação entre os termos SRs, sons adventícios, sons pulmonares e sons fisiológicos/da respiração [Sovijarvi et al., 2000].

Na traqueia, o SR normal ou ruído traqueal é caracterizado por um amplo espectro acústico, contendo uma frequência mais alta, audível tanto durante a fase inspiratória quanto na expiratória [Sovijarvi et al., 2000]. Já os SRs ouvidos no peito sofrem atenuação pelos pulmões e pela caixa torácica. O parênquima pulmonar e a parede torácica atuam como um filtro passa-baixo, não permitindo a passagem de sons de alta frequência. Portanto, o som auscultado no peito é constituído principalmente por baixas frequências [Sarkar et al., 2015].

Os sons adventícios caracterizam-se por serem SRs adicionais que se sobrepõem aos sons da respiração normais. Podem ser contínuos (como sibilâncias) ou descontínuos (fervores). Alguns pacientes podem manifestar ambos em simultâneo. Geralmente, a presença de sons adventícios indica a existência de patologia pulmonar [Sovijarvi et al., 2000, Reichert et al., 2008].

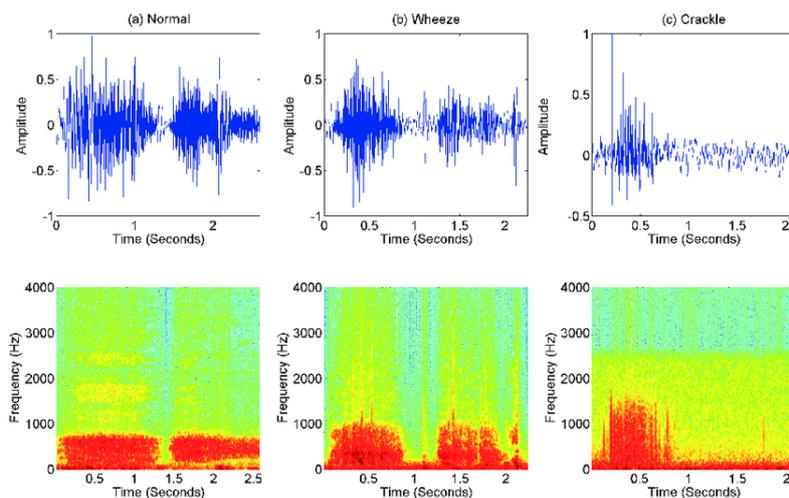
A Figura 2.1 resume a relação entre as definições aqui apresentadas.

### 2.1.1 Sons adventícios

#### Sons adventícios contínuos

Os sons adventícios contínuos (*Continuous Adventitious Sounds* - Sons Adventícios Contínuos (CAS)) consistem em sons pulmonares musicais que geralmente duram

mais de 250 milisegundos e são maioritariamente expiratórios, contudo, podem ser também inspiratórios ou ambos [Sarkar et al., 2015]. Estes sons são tipicamente divididos em sibilâncias, roncos<sup>3</sup>, estridor<sup>4</sup> e *squawks*<sup>5</sup>. A Figura 2.2 ilustra a representação e espectrograma de um som normal, sibilância e fervor.



**Figura 2.2:** a) Representação do SR normal, b) Sibilância, c) fervor; e os respetivos espectrogramas [Sengupta et al., 2016].

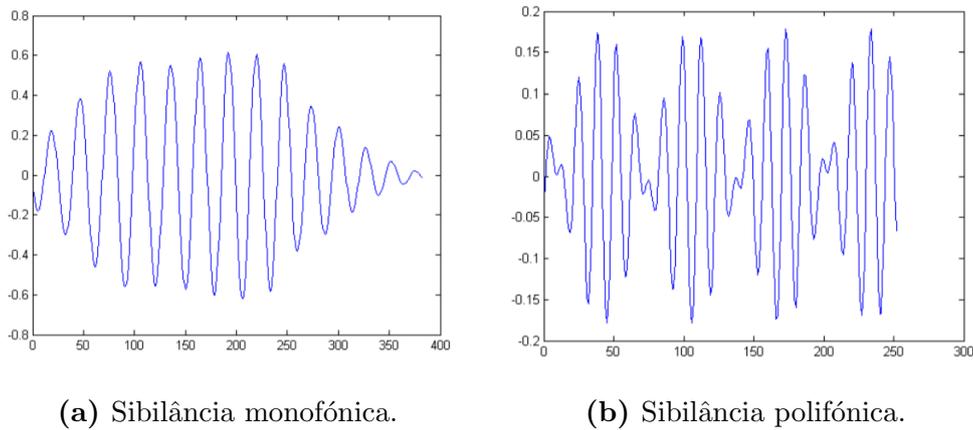
O Comité *American Thoracic Society* (ATS), relativamente à nomenclatura pulmonar, define sibilâncias como sons agudos de alta frequência com uma frequência dominante de 400 Hz ou mais, e roncos como sons contínuos com frequências mais baixas, de cerca de 150-200 Hz ou menos [Bunin and Loudon, 1979]. Tanto as sibilâncias como os roncos são sinais sinusoides, com frequências entre 100 e 1000 Hz [Pramono et al., 2017]. Geralmente as sibilâncias são sons mais agudos comparativamente aos outros SRs e são possíveis de ouvir no paciente com a boca aberta ou através da auscultação na traqueia e no tórax [Pasterkamp et al., 1997]. Os roncos, como têm frequências mais baixas, são mais perceptíveis através da auscultação no tórax [Bohadana et al., 2014, Sarkar et al., 2015].

As sibilâncias denominam-se de monofónicas ou polifónicas, uma vez que apresentam duração variável e frequências fundamentais distintas. Assim, uma sibilância é designada monofónica se incluir apenas uma frequência fundamental, e é designada de polifónica se incluir várias frequências fundamentais, tal como é possível observar na Figura 2.3 [Ulukaya et al., 2015]. Sabe-se ainda que as sibilâncias polifónicas e monofónicas estão associadas respetivamente a doenças das vias aéreas inferiores e superiores [Hashemi et al., 2011].

<sup>3</sup>Corresponde ao termo *rhonchi* em inglês.

<sup>4</sup>Corresponde ao termo *stridor* em inglês.

<sup>5</sup>Este termo não possui tradução para português.



**Figura 2.3:** Formas de onda para as sibilâncias [Ulukaya et al., 2015].

Os *squawks* são sibilâncias que ocorrem na inspiração, de curta duração (inferior a 200 milissegundos) [Sarkar et al., 2015]. A frequência fundamental varia entre 200 e 300 Hz [Bohadana et al., 2014]. Os *squawks* geralmente ocorrem no final da inspiração e são frequentemente precedidos por fervores [Sarkar et al., 2015].

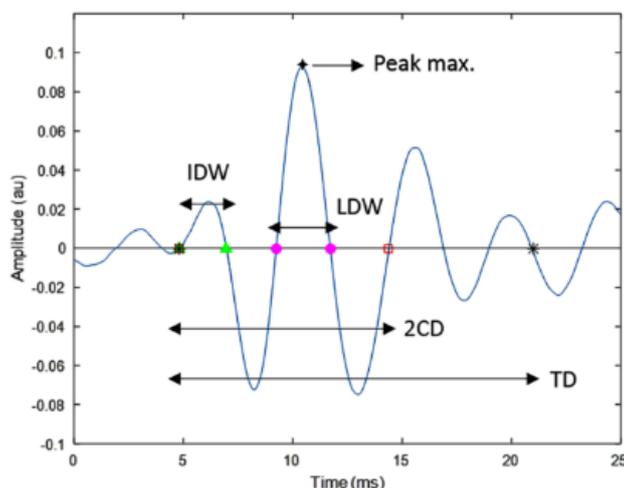
O estridor é um som musical agudo produzido pela obstrução do trato respiratório superior [Bohadana et al., 2014]. O estridor surge principalmente na inspiração e caracteriza-se por um pico a cerca de 1000 Hz no espectro de frequência [Sovijarvi et al., 2000].

### Sons adventícios descontínuos

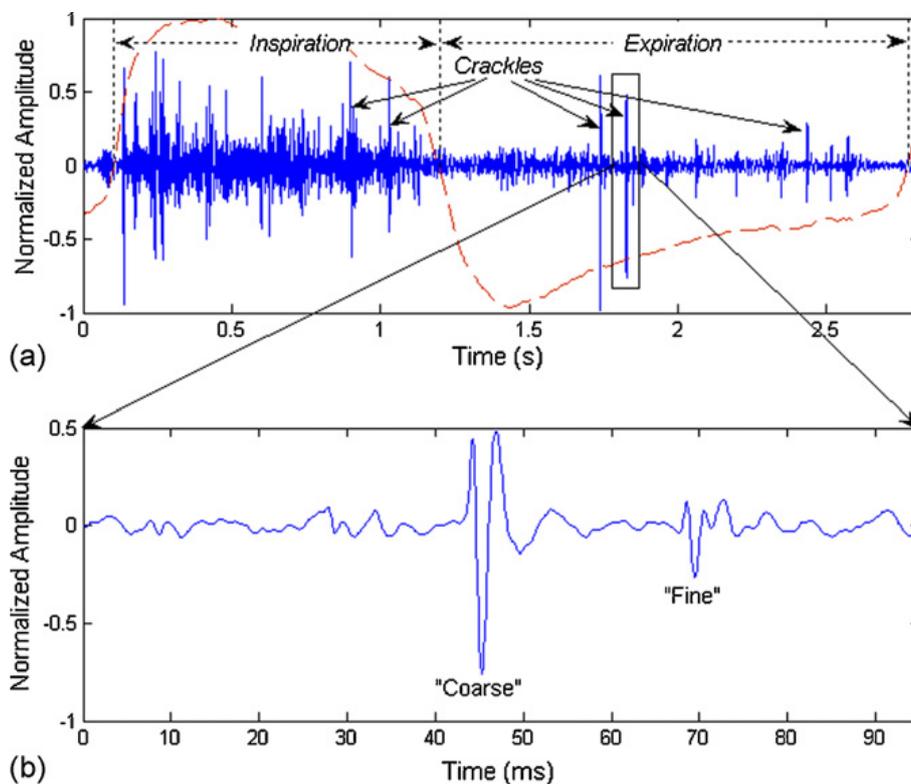
Os sons adventícios descontínuos (*Discontinuous Adventitious Sounds* - Sons Adventícios Descontínuos (DAS)) são sons pulmonares explosivos e não musicais. Os fervores são o principal tipo de som adventício descontínuo. Classificam-se em fervores de baixa frequência (*coarse*) ou de alta frequência (*fine*) com base na sua duração, altura do som e tempo do ciclo respiratório [Sarkar et al., 2015]. Três parâmetros são geralmente usados para avaliar o tipo de fervores: i) *Initial Deflection Width* - Largura de Deflexão Inicial (IDW), que é o tempo de duração entre o início e a primeira deflexão do fervor acima ou abaixo da linha de base; ii) *Two - Cycle Duration* - Duração de 2 ciclos (2CD) que é, tal como o nome sugere, a duração dos dois primeiros ciclos do fervor; iii) *Largest Deflection Width* - Duração da Deflexão de Amplitude Máxima na Onda (LDW) e iv) *Total Duration* - Duração Total (TD), que corresponde à duração total do fervor [Lozano et al., 2016]. A Figura 2.4 representa um fervor, bem como os parâmetros que o caracterizam.

Os fervores de baixa frequência (*coarse*), aproximadamente 350 Hz, são sons graves, de alta amplitude e com 2CD maior que 10 ms, enquanto os fervores de

alta frequência (*fine*), aproximadamente 650 Hz, são sons agudos, baixa amplitude e com 2CD menor que 10 ms (Figura 2.5) [Sovijarvi et al., 2000]. De acordo com os critérios do ATS, os ferveiros *coarse* têm durações médias de IDW e 2CD de 1,5 e 10 ms, e os dos ferveiros *fine* de 0,7 e 5 ms, respectivamente [Sarkar et al., 2015].



**Figura 2.4:** Ilustração dos parâmetros do fervero: 2CD, IDW, LDW e TD [Management Association, 2019].



**Figura 2.5:** Representação de um fervero *coarse* e *fine* [Yeginer and Kahya, 2008].

## 2.2 Importância do diagnóstico através de sons adventícios

Os SRs evidenciam a sua importância ao fornecerem informações sobre a integridade do sistema respiratório [Abbas and Fahim, 2010]. O diagnóstico de distúrbios pulmonares através da auscultação é dependente da presença de sons adventícios [Yeginer et al., 2004]. Os sons da respiração podem ser anormais em caso de patologias nas vias aéreas. Os SRs com frequências e intensidades irregularmente altas são característicos de muitas doenças com obstrução das vias aéreas, como asma e bronquite crônica [Sovijarvi et al., 2000].

Em [Gavriely et al., 1995] mostrou-se que alterações nos espectros de sons pulmonares poderiam ajudar na detecção precoce de doenças nas vias aéreas. O estudo em [Plante et al., 1998] sugeriu analisar os sons da traqueia a fim de identificar alterações estruturais e dinâmicas na via aérea superior.

As sibilâncias são muito comuns na asma e em pacientes com doenças obstrutivas nas vias aéreas. Em [Marini et al., 1979] demonstrou-se que o grau de obstrução dos brônquios relaciona-se com a presença e características das sibilâncias.

O estridor é uma sibilância de alta intensidade que indica uma obstrução morfológica ou dinâmica na laringe ou na traqueia [Sovijarvi et al., 2000]. Este manifesta-se geralmente durante a inspiração, quando fora da via aérea, e durante a expiração, quando dentro da via aérea, a menos que a obstrução seja fixa e, nesse caso, o estridor pode aparecer em ambas as fases da respiração [Bingham, 1987]. A complexidade do espectro do estridor é dependente da doença, do local da obstrução, do fluxo de ar e do volume [Sovijarvi et al., 2000].

Em pacientes com fibrose pulmonar, os ferveores são frequentemente *fine*, repetitivos e ocorrem no final da inspiração. Por sua vez, os ferveores associados à obstrução crônica das vias aéreas (por exemplo, DPOC) são *coarse*, menos repetíveis e surgem precocemente na inspiração [Nath and Capel, 1974]. O aparecimento de ferveores pode ser considerado um sinal precoce de uma doença respiratória [Sovijarvi et al., 2000]. O número de ferveores por respiração está associado à gravidade da doença [Epler et al., 1978]. Para além disso, a forma de onda e os episódios temporais dos ferveores revelam-se importantes para o diagnóstico de distúrbios cardiorrespiratórios, onde se verificam frequentemente [Piirila and Sovijarvi, 1995].

A Tabela 2.1 resume os sons adventícios e as suas características como tipo, gama de frequência, duração, fase do ciclo respiratório, origem, e ainda as doenças respiratórias mais comuns associadas a cada um deles.

**Tabela 2.1:** Resumo dos sons adventícios e possíveis patologias associadas [Kandaswamy et al., 2004, Bohadana et al., 2014, Sarkar et al., 2015, Pramono et al., 2017]

Som adventício	Tipo	Gama de frequência (Hz)	Duração (ms)	Fase do ciclo	Origem	Doença associada	Outros
					estreitamento das vias aéreas,	asma, fibrose cística, DPOC	patologias nas vias aéreas superiores
Sibilância	CAS	alta, >400	>80	inspiração ou expiração	limitação do fluxo aéreo		patologias nas vias aéreas inferiores
Roncos	–	baixa, 150-200 ou menos	>80	geralmente na expiração, inspiração e ambas	secreção bronquial, mucosa mais espessa	bronquite crônica, tumores, pneumonia, DPOC	pode desaparecer após a tosse
Estridor	CAS	alta, >500 (1000)	>250	geralmente na inspiração, expiração e ambas	fluxo de ar oscilante na laringe ou nos brônquios inferiores (obstrução das vias aéreas superiores)	laringite, paralisia das cordas vocais, inflamação das vias aéreas após extubação, tumores, estenose traqueal	mais perceptível no pescoço do que no peito; frequente em crianças

## 2. Análise de sons respiratórios: contexto e definições

Squawks	-	CAS	baixa, 200-300	<200	inspiração	oscilação das vias aéreas periféricas	alveolite alérgica, fibrose pulmonar, fibrose intersticial	precedidos por fervores
	<i>fine</i>		alta, 650	<10 ( $\pm 5$ )	inspiração (no fim)	abertura explosiva das vias aéreas pequenas	pneumonia, insuficiência cardíaca congestiva, fibrose pulmonar	
Fervores		DAS			geralmente inspiração (início), expiração, ambas	bolha de ar nos grandes brônquios	bronquite crónica, bronquiectasia, DPOC	
	<i>coarse</i>		baixa, 350	>10 ( $\pm 15$ )				

## 2.3 Auscultação dos sons respiratórios em crianças vs adultos

No bebé até aos 2 anos, devido à menor quantidade de alvéolos e à maior densidade da caixa torácica e do parênquima pulmonar, é audível ruído respiratório brônquico, ou seja, ambos os tempos respiratórios (inspiração e expiração) são audíveis com clareza ao longo de todo o tórax. Porém, como os bebés apresentam uma frequência respiratória (entre 30 a 60 ciclos por minuto) e cardíaca (entre 90 a 190 batimentos por minuto) muito elevada e o parênquima pulmonar tem um elevado poder de propagação sonora, a auscultação torna-se difícil, pois o som propaga-se muito rapidamente e muitas vezes o som gerado numa área propaga-se ao longo de toda a via aérea, sendo difícil a correta distinção [Pasterkamp et al., 1996].

No adolescente e no adulto o ruído respiratório brônquico é considerado patológico e o expectável é o ruído respiratório normal, ou seja, ruído de baixa frequência que vai diminuindo dos ápices pulmonares para as bases, sendo que nas bases a expiração é quase inaudível [Pasterkamp et al., 1996].

A aquisição de SRs em crianças é, para além das razões já mencionadas, dificultada pelo seu choro, movimentação e falta de cooperação, o que leva frequentemente os profissionais de saúde a diagnosticar com base nos relatos dos pais [Bokov et al., 2016, Emmanouilidou et al., 2018].

## 2.4 Análise computacional do som respiratório

A análise computacional promove uma aprendizagem autónoma, competências técnicas contemporâneas e de resolução de problemas e pode integrar-se nas várias formas de ensino, estudos de caso e resolver grande parte da subjetividade associada à auscultação.

A *Computerized Respiratory Sound Analysis* (CORSA) é um método simples, objetivo e não invasivo para detetar e caracterizar sons adventícios, que regista os SRs dos pacientes com um dispositivo eletrónico e analisa-os com base em características específicas do sinal [Marques et al., 2014]. A CORSA tem várias vantagens comparativamente à auscultação realizada pelo especialista de saúde, nomeadamente uma melhor correlação com os critérios que definem SRs, redução da variabilidade entre ouvintes e diagnóstico de distúrbios respiratórios mais assertivo e objetivo [Gurung et al., 2011, Fischer et al., 2016].

Em 2000, a *European Respiratory Society* publicou normas para a CORSA com o objetivo de padronizar tanto o registo quanto o processamento de sinais dos SRs

[Sovijärvi et al., 1997]. A Tabela A.1 que se encontra na secção A dos anexos, resume as recomendações relativas à CORSA para a aquisição de SRs, nomeadamente as características dos sensores e a localização dos microfones [Vannuccini et al., 2000].

### 2.5 Definição de *feature*

Em *machine learning* e no reconhecimento de padrões, um atributo<sup>6</sup>/*feature* é uma propriedade mensurável ou uma característica de um fenómeno que é analisado. A escolha de *features* informativas, discriminatórias e independentes é uma etapa crucial para algoritmos eficazes no reconhecimento e classificação de padrões [Liu and Motoda, 2012]. Um conjunto de *features* numéricas pode ser convenientemente descrito por um vetor de *features*, que é precisamente um vetor de atributos numéricos que representam um objeto. Em *machine learning* existem muitos algoritmos que requerem uma representação numérica de objetos, pois essas representações facilitam o processamento e a análise estatística [Liu and Motoda, 2012].

O conjunto inicial de *features* pode ser redundante e demasiado grande para ser gerido. Portanto, uma etapa preliminar bastante usada em aprendizagem computacional é a seleção de um subconjunto de *features* ou a construção de um novo e reduzido conjunto para facilitar a aprendizagem [Piramuthu and Sikora, 2009].

---

<sup>6</sup>Usou-se o termo “atributo” como tradução de *feature*, embora ambos os termos sejam usados nesta dissertação.

## Revisão da Literatura

Os sistemas de análise automatizada de sons respiratórios (SRs) geralmente incluem um método para aquisição de sinal, análise e classificação. Este capítulo apresenta uma revisão de literatura sobre os métodos de aquisição de SRs, extração de *features* e algoritmos usados para a classificação de SRs em crianças e adultos.

### 3.1 Métodos de extração de *features* e classificação

#### 3.1.1 Crianças

Em [Fenton et al., 1985], 5 indivíduos asmáticos com idades entre os 10 e 16 anos foram estudados durante ataques de asma espontâneos ou induzidos, ao longo de 20 minutos. Os sons foram registrados por dois acelerômetros de contacto (*EMT-25B, ElemaSchonander Siemens, Iselin, NJ*) que gravavam simultaneamente sons no peito (pulmonares) e pescoço (traqueais). Os espectros de potência<sup>1</sup> foram calculados usando a *Fast Fourier Transform* (FFT). Dois parâmetros foram utilizados para descrever os espectros: i) o limite superior de frequência  $f_u$  e a ii) frequência da potência máxima  $f_p$ . O objetivo era detectar sibilâncias e para isso, analisaram os picos grandes e estreitos no espectro de potência. O limite superior  $f_u$  indicava a largura de banda do sinal e foi definida como a frequência além da qual a energia foi sempre menor que 10% do pico de potência. O pico de potência foi definido numa largura de banda média especificada ( $f_1-f_2$ ), fora da faixa de energia de baixa frequência dos sons cardíacos e do ruído muscular. Os sons normais do pulmão em adultos encontram-se abaixo de  $f_1$ , mas em crianças asmáticas e sobre a traqueia são de menor importância. Um valor acima de 200 Hz pode ser considerado anormal, permitindo identificar sibilâncias presentes nos sons pulmonares. O método para a detecção de sibilâncias nos sons traqueais verificava se o pico de potência era um fator  $A_f$  vezes a potência média  $P_{avg}$  na largura de banda  $f_1-f_2$ . Assim, a sibilância

---

<sup>1</sup>Espectro de potência: a magnitude ao quadrado do espectro de frequência de um sinal [Waitman et al., 2000].

foi definida como sendo um pico no espectro de potência acima de 200 Hz e pelo menos uma vez maior que  $P_{avg}$ . A potência abaixo de 200 Hz é considerada normal e pode conter contribuições de sons cardíacos. As sibilâncias detetadas apresentavam alta amplitude - geralmente superior a 15 vezes a potência média. No caso das sibilâncias polifônicas notou-se que todos os picos de potência começaram juntos e continuaram até o final da sibilância. O estudo mostrou ainda que, os sons traqueais normais são mais altos na expiração e a frequência do pico de potência é um meio confiável para a identificação de sibilâncias. Verificou-se também que a traqueia era um local preferível relativamente ao pulmão para analisar sibilâncias.

Na abordagem proposta em [Gnitecki et al., 2004], os SRs de crianças dos 9 aos 15 anos de idade com broncoconstrição foram registados através da inspiração de metacolina, no lobo pulmonar posterior inferior direito enquanto os indivíduos respiravam normalmente durante 60 s (a gravação durou 70 s, incluindo 10 s a suster a respiração). A *Signal-to-Noise Ratio* - Relação Sinal-Ruído (SNR) foi determinada a partir do RMS dos SRs. Aplicaram-se técnicas de *Fractal Dimension* - Dimensão Fractal (FD) nomeadamente *Variance Fractal Dimension* - Variância Dimensão Fractal (VFD) e *Katz Fractal Dimension* - Dimensão Fractal Katz (KFD). O KFD calcula-se através da equação 3.1,

$$KFD = \frac{\log_{10}(n)}{\log_{10}(n) + \log_{10}(\frac{d}{L})} \quad (3.1)$$

onde  $n$  é o número de incrementos entre amostras da forma de onda sobre a qual o KFD é calculado,  $L$  é a soma de todas as distâncias entre incrementos sucessivos e  $d$  é o valor da distância máxima medida desde o início do primeiro incremento. A VFD é determinada pelo expoente de Hurst,  $H$ , que se baseia na relação da *law power* entre a variação dos incrementos de amplitude de um sinal,  $B(t)$  (neste caso sons pulmonares), ao longo de um incremento de tempo  $\Delta t = |t_2 - t_1|$ , com  $B(t_2) - B(t_1)$  indicado como  $(\Delta B)_{\Delta t}$ . A *law power* é  $\text{Var}[(\Delta B)_{\Delta t}] \sim \Delta t^{2H}$ , e o expoente Hurst define-se pela equação 3.2,

$$H = \lim_{\Delta t \rightarrow 0} \left[ \frac{1}{2} \frac{\log_b \text{Var}(\Delta B)_{\Delta t}}{\log_b(\Delta t)} \right] \quad (3.2)$$

com  $b = 2$  para este estudo. Usando a equação 3.2, a VFD para um caso de dimensão euclidiana  $E$  (igual a 1 para formas de onda), é determinado pela equação 3.3.

$$VFD = E + 1 - H \quad (3.3)$$

Neste estudo, os sinais foram filtrados através de um filtro digital *Butterworth* de 5ª ordem e a classificação foi feita através do algoritmo *One Nearest-Neighbor* (1-NN). Os resultados obtidos para os verdadeiros positivos (falsos positivos) foi de 90,3%(23,4%), 63,5% (24,9%) e 58,3% (26,1%) para RMS-SNR + KFD, RMS-SNR +VFD e RMS-SNR, respetivamente. Em suma, pode-se concluir que o RMS-SNR e KFD apresentam o melhor desempenho na classificação 1-NN, dentro de 75-600 Hz, em relação ao RMS-SNR + VFD e RMS-SNR apenas.

No estudo em [Jin et al., 2011], o conjunto de dados 1 é constituído por SRs registados na traqueia de 7 indivíduos saudáveis e 14 patológicos com diferentes graus de obstrução das vias aéreas (8 sexo masculino e 13 sexo feminino,  $15 \pm 9$  anos). E no conjunto de dados 2, os SRs foram capturados em várias posições no tórax posterior direito/esquerdo superior/inferior usando um estetoscópio eletrónico e um microfone condensador (ECM-77B, Sony, Inc., Japão), de 3 indivíduos saudáveis e 12 patológicos (9 sexo masculino e 6 sexo feminino,  $11 \pm 7$  anos). Os sujeitos teste apresentavam disfunções pulmonares como: asma, pneumonia e bronquiolite. O tempo de gravação foi de 600 s para cada sujeito. E cada segmento sonoro foi classificado em 4 categorias: i) SR normal, ii) estridor, iii) sibilância monofónica e iv) polifónica. A derivada espectral define-se como a soma absoluta das derivadas de primeira ordem do espectro de magnitude. Propõem-se 3 *features* baseadas na derivada espectral: i) derivada espectral média: distribuição dos valores da derivada espectral para cada segmento de SR. Trata-se de um parâmetro significativo, pois o espectro da sibilância polifónica tem vários picos distintos que fornecem um valor da derivada espectral alto, enquanto a sibilância monofónica e o estridor apresentam apenas um pico, fornecendo valores mais baixos da derivada espectral. Por outro lado, o SR normal com ruído possui um valor de derivada espectral ainda mais baixo; ii) *spread* temporal: a variação da derivada espectral ao longo do tempo analisa as durações dos picos espectrais. Como a sibilância e o estridor podem ter variações espectrais contínuas, altos valores de  $T_{spread}$  inferem a sua presença. Por outro lado, os SRs com ruído fornecem baixos valores de  $T_{spread}$ ; iii) posição espectral: indica a posição espectral do desvio espectral máximo em  $T_{max}$ . É útil para diferenciar a sibilância monofónica do estridor que tem uma frequência fundamental diferente. Os atributos apresentados com base na derivada espectral resumem com exatidão a duração e a intensidade do espectro de magnitude, no eixo da frequência para cada intervalo de tempo. O classificador *k-Nearest Neighbors* (k-NN) alcançou uma exatidão<sup>2</sup> de  $92,4 \pm 2,9\%$ .

No artigo em [Emmanouilidou et al., 2012] pretende-se analisar os sons pulmona-

---

<sup>2</sup>Utilizou-se o termo exatidão como tradução de *accuracy*, nesta dissertação.

res anômalos na ausculta pediátrica, mais precisamente em bebês do Nepal contendo choro, conversas de fundo e ruído ambiental. Os sons foram adquiridos com um estetoscópio de gravação digital *ThinkLabs Inc.* ligado a um MP3 *player*. Os sujeitos eram crianças saudáveis ou com doenças respiratórias. Foram selecionadas um total de 28 gravações de 15 s: 10 casos normais, 10 com sibilância e 8 com ferveores. Foi aplicado o método de extração de *features* proposto em [Waitman et al., 2000] (este estudo será apresentado na secção dos adultos 3.1.2): o espectro de potência de cada segmento foi obtido e somado ao longo do eixo de frequência que varia de 0-800 Hz para formar um vetor de *features*. O algoritmo *Support Vector Machine* (SVM) foi capaz de discriminar casos normais e anormais (sibilâncias e ferveores) com sensibilidade de 89,44% e especificidade de 80,50%.

Na abordagem em [Xie et al., 2012], recolheram SRs de 7 indivíduos saudáveis e 14 patológicos com diferentes graus de obstrução das vias aéreas (8 sexo masculino e 13 sexo feminino,  $15 \pm 9$  anos). Outro conjunto de dados foi adquirido através de um estetoscópio eletrônico com um microfone (*ECM-77B, Sony Inc.*, Japão) em várias posições sobre o tórax posterior direito/esquerdo superior e inferior, a partir de 3 indivíduos saudáveis e 12 patológicos (9 sexo masculino e 6 do sexo feminino,  $11 \pm 7$  anos). Os sujeitos teste encontravam-se com disfunção pulmonar, nomeadamente: asma, pneumonia e bronquiolite. É proposto um método de extração de *features* que combina a *Discrete Wavelet Transform* - Transformada Discreta *Wavelet* (DWT) com a *Principal Component Analysis* - Análise das Componentes Principais (PCA), ou seja, um PCA multi-escala. A FFT foi aplicada aos SRs para obter os espectros de potência. Alcançaram uma exatidão de  $97,3 \pm 2,7\%$ .

Em [Emmanouilidou and Elhilal, 2013], usaram um estetoscópio digital *ThinkLabs Inc.* para recolher SRs de 53 crianças saudáveis em ambientes não ideais. As contaminações de sinal consideradas neste estudo envolvem ruído ambiente, conversas em segundo plano, choro, interferência eletrónica e ruídos produzidos por deslocamentos intencionais ou não intencionais do estetoscópio. Do espectro de amplitude, vários atributos foram extraídos:

- Largura do pico;
- Inclinação do espectro. O espectro produzido pelas gravações de som pulmonar decai exponencialmente com frequências de 100 Hz;
- *Power ratio*;
- Taxa de frequência baixa a alta.

Para classificar diferentes tipos de ruído usou-se o SVM. 80% dos dados foram utilizados para treino e 20% dos dados para teste.

A detecção de sibilâncias também foi estudada em [Yu et al., 2013], com sons obtidos através de um estetoscópio que foi construído usando um microfone. Os sons foram gravados no pescoço durante 5-30 s de 8 crianças, 6 do sexo masculino e 2 duas do sexo feminino (1-7 anos, com uma média de  $2,5 \pm 1,87$ ) e foram utilizadas 59 gravações, 25 com sibilância e 34 normais. Sendo que para cada paciente com sibilância, os pediatras confirmaram que o som adventício foi causado por asma. A *feature* utilizada foi o coeficiente de correlação do espectro respiratório adaptativo (*Adaptive Respiratory Spectrum Correlation Coefficient*), enquanto o classificador foi empiricamente determinado. Vários coeficientes de correlação consecutivos foram considerados um evento de sibilância. Assim, quando uma sibilância surge, cada espectro *Short-time Fourier Transform* (STFT) exibe componentes harmônicos contínuos ao longo do eixo do tempo, de modo que os espectros são bastante semelhantes entre si como indicado pelos seus valores de correlação cruzada. Finalmente, cada registo é classificado como contendo sibilância ou som normal. O desempenho alcançado foi de 88% de sensibilidade com 94% de especificidade na detecção de sibilâncias. A limitação deste sistema é que as crianças podem chorar quando a asma ataca e o choro causa baixa exatidão na determinação da sibilância.

No artigo em [Ellington et al., 2014], através de um estetoscópio digital (*ThinkLabs ds32a*, Centennial Colorado, EUA) e um gravador de MP3 registaram sons pulmonares de 186 crianças saudáveis, com uma média de  $2,2 \pm 1,4$  anos, onde 47% eram do sexo masculino, 27% eram crianças com 2 a 12 meses, 44% tinham 1 a 2 anos e 30% tinham 3 a 5 anos, em cada um dos 8 locais torácicos durante 10 s: esquerdo e direito anterior superior (AS), direito e esquerdo anterior inferior (AI), posterior superior direito e esquerdo (PS) e posterior inferior esquerdo e direito (PI). Foram utilizadas apenas 151 gravações sonoras (71 sexo feminino e 80 sexo masculino). Os sinais adquiridos foram filtrados com um filtro passa-baixo *Butterworth* de 4<sup>a</sup> ordem com uma frequência de corte de 1 kHz. Isso porque 91% da energia total do sinal foi encontrada em frequências abaixo de 1 kHz. Os sons foram processados em segmentos de curta duração com cerca de 2 segundos. Excluíram-se segmentos contaminados por ruído caracterizados por irregularidades de alta frequência na faixa de 200-500 Hz, o que corresponde ao choro das crianças. O espectro de potência foi obtido usando a FFT e foi posteriormente aplicado um filtro *Butterworth* de 5<sup>a</sup> ordem a 20 Hz. Usaram 10 atributos de representações espectrais e espectro-temporais das gravações sonoras. A análise espectral informa sobre o conteúdo da frequência do sinal gravado, como variações lentas ou rápidas

no sinal, o que inclui parâmetros como largura de pico, inclinação do espectro, *power of regression line* e *power ratio*. Estes atributos avaliam a concentração de energia ao longo da frequência. Para além disso, extraíram também *Mel-Frequency Cepstral Coefficients* (MFCCs). Utilizaram-se ainda parâmetros de tempo-frequência, incluindo forma espectral e modulações temporais. A largura do pico representa, tal como nome indica a largura do espectro de potência. O pico do espectro foi identificado na faixa de 0-200 Hz. A inclinação do espectro é a inclinação da linha de regressão linear ajustada ao espectro em eixos logarítmicos, que diminui exponencialmente para frequências superiores a 75 Hz. *Power of regression line* consiste no total de “*power*” sob a linha de regressão. *Power ratio* é a relação entre a potência total calculada e a potência da linha de regressão. Um valor de *power ratio* próximo de 1 significa que o espectro logarítmico segue a linha de regressão. Para este estudo, três coeficientes (MFCC1, MFCC2, MFCC3) foram usados para cada indivíduo, calculando a média de todos os MFCCs extraídos num curto espaço de tempo, correspondendo a filtros centrados nas frequências 56, 116, 181 Hz respetivamente. A forma espectral (escalas) estima o quão amplo ou estreito o perfil espectral é, indicando as variações na frequência. Modulações (taxas) temporais referem a velocidade com que o conteúdo da frequência se altera com o tempo e em qual fase (direção), positiva ou negativa. As *features* como largura de pico espectral, MFCCs e modulações espectro-temporais também mostraram variações com o local de gravação.

Em [Jin et al., 2014], registaram sons traqueais de 7 indivíduos saudáveis e 14 patológicos com diferentes graus de obstrução das vias aéreas (8 sexo masculino e 13 sexo feminino, com  $15 \pm 9$  anos), através de um microfone condensador (*ECM-77B*, *Sony Inc.*, Japão). Pediu-se aos sujeitos de teste que prendessem a respiração por 15 s e depois respirassem normalmente sem fluxo direcionado. A gravação de 600 s foi guardada para cada sujeito em monocal. Neste estudo, usaram também sons pulmonares de gravações recolhidas nos pulmões, tórax superior e inferior de 5 indivíduos saudáveis e 19 patológicos (14 do sexo masculino e 10 do sexo feminino, com idades compreendidas entre os  $11 \pm 7$  anos). Os sons registados incluem sibilância, estridor, roncos, e mistura destes. Propõem-se 3 métodos de extração de *features* para classificar SRs: média auto-regressiva (DP), *kurtosis* instantâneo (*kurtosis-KP*) medido recursivamente e distorção dos histogramas pela entropia da amostra (SEP). Para o conjunto de dados, o desempenho é superior a 90% para inspirações e expirações quando o conjunto de *features* individuais de DP ou KP é usado. Pelo contrário, o conjunto de atributos do SEP fornece uma exatidão ligeiramente menor. O desempenho melhora significativamente quando o SEP é

combinado com os outros 2 conjuntos de *features*. Combinando os 3 conjuntos de *features*, é possível obter exatidões de classificação de 100% e 97,1% para expiração e inspiração, respetivamente.

No estudo em [Mazić et al., 2015], recolheram um total de 45 sons, com uma duração de 10 s, de 16 crianças com idades entre 1-6 anos de idade. Portanto, de um total de 45 fonopneumogramas<sup>3</sup>, 21 deles foram utilizados para treino e validação, enquanto os restantes 24 fonopneumogramas foram utilizados para o teste. Em 21 fonopneumogramas, 5 deles continham sibilância, 9 continham outros sons (inspiração, expiração, estridor de inspiração e ronco), e os 12 fonopneumogramas restantes continham sons não classificados. Como etapa de pré-processamento do algoritmo de aprendizagem computacional, cada sinal dos fonopneumogramas foi primeiro filtrado com o filtro passa-alto de 50<sup>a</sup> ordem *Yule-Walker* com a frequência de corte de 100 Hz para reduzir o impacto de sons cardíacos e ruído muscular. Posteriormente, é calculada a STFT. Pela análise dos espectrogramas, um médico classifica visualmente e manualmente como SRs normais, sibilâncias, estridor ou roncós. Para o caso de sibilância, é importante determinar o início e o fim dessa sequência. O artigo propõe uma arquitetura de reconhecimento de padrões de 2 camadas para deteção de sibilância em crianças com asma. A primeira camada consiste em dois classificadores SVM. A segunda camada usa um limiar de deteção digital, que aprimora ainda mais a estrutura proposta com o objetivo de melhorar o processo de deteção de sibilâncias. O primeiro nível do algoritmo de classificação foi assegurado com uma cascata SVM que consiste em 2 classificadores paralelos que apresentam melhor desempenho comparativamente a um único classificador SVM e ajuda a discriminar sons não asmáticos de sibilâncias de asma. Como extração de *features* foram usados os MFCCs, *kurtosis* e entropia. Além disso, foi verificado experimentalmente que uma condição necessária para a deteção bem sucedida de sibilâncias, com base nos recursos do MFCCs, pressupõe que a diferença entre as frequências centrais da sibilância testada e o seu vizinho mais próximo do conjunto de treino, deve ser limitada a 40 Hz. Para diferenças acima de 60 Hz, a sibilância será classificada incorretamente. Concluiu-se que, usando os MFCCs como *features* e *Gaussian Mixture Models* (GMM) ou SVM como classificadores, a deteção de sibilâncias pode atingir uma exatidão superior a 95%.

Em [Song, 2015], numa amostra de 376 crianças (212 do sexo masculino e 164 do sexo feminino), 153 tiveram pneumonia. O conjunto de dados foi agrupado nas seguintes faixas etárias: 3-12 meses, 13-24 meses, 25-36 meses, 37-48 meses e 49-60

---

<sup>3</sup>Fonopneumograma (*Phonopneumogram*): é uma representação simultânea e sobreposta do sinal sonoro e do fluxo de ar no domínio do tempo durante a respiração [Piiirila and Sovijarvi, 1995].

meses. O som foi gravado colocando o microfone próximo do nariz ou da boca (se o nariz estiver bloqueado) por pelo menos 1 minuto. Usaram 18 tipos de *features* acústicos agrupados em *features prosodic*, espectrais, *cepstrais*, operadores de energia Teager e informações temporais. *Prosodic features* são *features* geralmente usadas para a análise de SRs, isso inclui frequência fundamental, *formants*, *zero-crossing rate*, volume, *crest factor*, *log-variance*, *skewness* e *kurtosis*. *Formants* são os picos espectrais de um espectro sonoro. As *features* espectrais são calculadas transformando o sinal para o domínio da frequência pela *Fourier Transform* (FT). As seguintes *features* espectrais foram usadas: *spectral centroid*, *spectral flux*, *spectral entropy*, *spectral spread* e *power spectral density* (PSD). Utilizou-se o método *Welch* para calcular a média das estimativas do PSD. Os atributos *cepstrais* incluem as *features cepstrais* e os 13 MFCCs. Para classificação, utilizaram dois classificadores: k-NN e SVM usando o Weka. Cada classificador foi aplicado em cada faixa etária separadamente. Para avaliar o desempenho da classificação, foi utilizado o método de validação cruzada *leave-out* (LOOCV), uma vez que em cada faixa etária o número de dados não era grande (por exemplo, a faixa etária de 25 a 36 meses contém 41 dados). A classificação revela pneumonia vs não-pneumonia, e a exatidão média para o k-NN foi de 90,68% e para SVM foi de 91,98%. O sistema alcançou especificidade de 89,88% e 90,68%, sensibilidade de 90,22% e 92,06%, para k-NN e SVM, respetivamente.

No estudo em [Bokov et al., 2016], as crianças foram selecionadas caso tivessem entre 1 dia e 12 anos de idade e apresentassem SRs anormais na avaliação clínica inicial (pelo menos um som: sibilâncias, ferveores, roncós; uma das doenças: laringite aguda, rinite). A duração da gravação foi de 30 s. No total, foram obtidas 186 gravações de crianças com idade média de 20 meses. Desenvolveram um algoritmo de reconhecimento de sibilâncias a partir de SRs registados com um smartphone colocado próximo da boca. Optaram por usar 14 *features* com base na definição acústica de sibilância (equações 3.4-3.11).

$$F_{1,2} = \text{mean}(f_i^{1..N_i}) \quad (3.4)$$

com  $i$  a ordem harmónica (fundamental ou segunda harmónica),  $f$ , a frequência média de um determinado segmento sibilância e  $N_i$  o número de sinais compatíveis com sibilância detetados em uma determinada gravação para cada harmónica.

$$F_{3,4} = [\max(f_i^{1..N_i}) - \min(f_i^{1..N_i})] / \text{mean}(f_i^{1..N_i}) \quad (3.5)$$

$$F_{5,6} = \text{mean}(PSD_i^{1..N_i}) \quad (3.6)$$

Define-se  $PSD_i^m$  como o PSD médio ao longo dos intervalos de tempo incluídos numa determinada sibilância  $m$  para o harmónico  $i$ , onde  $f_l=200$  Hz e  $f_L=2500$  Hz são a frequência limite para o cálculo do PSD médio, enquanto  $F$  é o conjunto de frequências correspondentes à definição de um pico de sibilância (equação 3.7).

$$PSD_i^m = \text{mean} \left[ \sum_{f \in F} PSD_i(f) \ / \ \sum_{f_i}^{f_L} PSD(f) \right] \quad (3.7)$$

$$F_{7,8} = [\max(PSD_i^{1..N_i}) - \min(PSD_i^{1..N_i})] / \text{mean}(PSD_i^{1..N_i}) \quad (3.8)$$

$$F_{9,10} = \text{mean}(\tau_i^{1..N_i}) \quad (3.9)$$

$$F_{11,12} = [\max(\tau_i^{1..N_i}) - \min(\tau_i^{1..N_i})] / \text{mean}(\tau_i^{1..N_i}) \quad (3.10)$$

$$F_{13,14} = \sum_{k=1}^{N_i} \tau_i^k \ / \ T \quad (3.11)$$

onde  $\tau$  é a duração do  $i$ -ésimo harmónico de uma determinada sibilância. Um sinal compatível com sibilância é definido como tendo 3 ou menos harmónicos detetáveis em cada segmento de tempo e duração de pelo menos 100 ms. Neste estudo, uma das condições para um pico no espectro de potência ser considerado como sibilância era que a largura do pico em 95% do pico de potência fosse menor que 40 Hz. O melhor desempenho (sensibilidade de 71,4% e especificidade de 88,9%) foi observado com o algoritmo SVM.

### 3.1.2 Adultos

Em [Pesu et al., 1996], os SRs adquiridos no tórax e a taxa de fluxo respiratório da boca foram registados em 9 pacientes com diferentes doenças pulmonares e em 4 indivíduos saudáveis, com um tempo de gravação a variar de 30-40 s. Os segmentos temporais foram então classificados em 4 classes: sibilâncias, ferveores *fine*, ferveores *coarse* ou som normal. O método de extração de *features* usado aqui é o seguinte: dividir o sinal em segmentos; calcular a decomposição do *wavelet packet* para cada segmento; procurar a melhor base; usar os resultados da melhor pesquisa da base para construir o vetor de *features*. A classificação dos vetores de *features* é realizada

usando o *Learning Vector Quantization* (LVQ). Os ferveores *fine* foram claramente melhor detetados que os ferveores *coarse*. Isto porque, o fervor *coarse* tem uma gama de frequência mais baixa e, portanto, a diferença para o som normal é menor. As sibilâncias apresentaram melhor desempenho na detecção comparativamente aos ferveores *coarse*, mas o número de falsos positivos foi bastante alto.

Em [Kahya et al., 1997], foram analisados registros de 28 pacientes com DPOC, 23 pacientes com doença pulmonar restritiva e 18 indivíduos saudáveis. Os SRs foram classificados em normal e anormal (ferveores), para tal usaram *Autoregressive Modeling* (AR)<sup>4</sup> como método de extração de *features* e o k-NN como classificador, alcançando uma exatidão de 69,59%.

No artigo em [Waitman et al., 2000], gravaram os SRs dos pacientes com os microfones (*LS-60 Adult Precordial Sensors by Novasonics Inc., Rye, New York*) colocados no tórax anterior e utilizaram dois sensores (um para cada pulmão) para permitir a gravação simultânea dos SRs de cada pulmão. Selecionaram 17 sujeitos, 6 do sexo feminino e 11 do sexo masculino com idades entre 19 e os 75 anos para obter uma variedade de sons pulmonares normais e patológicos. Os sons incluíram ferveores, sibilâncias, estridor, SRs de menor intensidade, turbulência do ventilador, pacientes em ventilação e sons normais da respiração. Quanto ao método de extração de *features*: os SRs foram representados pelos seus espectros de potência, somados em vetores de *features* no espectro de frequências de 0 a 800 Hz. Depois de calcular a FFT do SR, o espectro de potência foi calculado. De notar que as frequências de interesse devem ficar abaixo de 800 Hz, para que essas *features* sejam espaçadas de 0 a 800 Hz. As *features* separadas foram calculadas para o ciclo inspiratório e expiratório. Como etapa final, o logaritmo do valor de cada *feature* é utilizado para reduzir o intervalo. Alcançaram uma exatidão de 73% e uma sensibilidade e especificidade de 87% e 95% respetivamente, através de diferentes redes neuronais de retropropagação.

Em [Baydar et al., 2003], foram estudados 9 indivíduos saudáveis e 11 indivíduos patológicos com diferentes tipos de doenças respiratórias. A duração de cada gravação foi de cerca de 13 s, tendo cada uma incluído 3 a 4 ciclos respiratórios. Neste trabalho, propõe-se a aplicação de um novo método de extração de *features*, denominado método de coerência de sinais, para a análise dos SRs. Coerência de sinal é um tipo de medida que indica a quantidade de variabilidade aleatória em cada componente do sinal de *Fourier*. O sistema de classificação proposto para os SRs consiste em três etapas: i) extração de *features*, na qual se calcula os valores de coerência do sinal

---

<sup>4</sup>AR: Métodos que estimam o espectro de potência de um sinal resolvendo equações lineares. O valor do sinal num determinado momento é assumido como uma combinação linear dos valores anteriores dos sinais multiplicados pelos coeficientes [Waitman et al., 2000].

para representar os SRs; ii) redução da dimensão das *features* extraídas para um espaço dimensional menor; iii) detecção, por meio de um classificador de distância euclidiana, *K-Means*.

No estudo em [Bahoura and Pelletier, 2003] propõe-se o uso dos MFCCs como *features*. As *features* extraídas com base na análise *cepstral* (MFCCs e parâmetros *cepstrais*) apresentaram melhores resultados na classificação comparativamente à *Wavelet Transform* (WT), *Linear Predictive Coding* (LPC) e FFT. Usando o algoritmo LVQ classificaram os segmentos sonoros como contendo sibilâncias ou SRs normais, atingindo-se uma exatidão de 75,80% e 77,50% para sibilâncias e não sibilâncias, respetivamente.

A abordagem em [Kandaswamy et al., 2004], apresenta um método de análise tempo-frequência para sinais sonoros pulmonares usando coeficientes da WT e classificação numa de 6 categorias: normal, sibilância, fervor, *squawk*, estridor ou roncocal usando 4 algoritmos de uma *Artificial Neural Network* (ANN) de retropropagação: *Adaptive Learning Rate BP* (GDA), *Resilient BP* (RP), *Scaled Conjugate Gradient* (SCG), *Levenberg Marquardt* (LM). Os sons pulmonares foram gravados a partir de vários sujeitos, deitados em posição supina e em estado relaxado, e assim terão diferentes níveis de intensidade. Os sinais foram decompostos em subcamadas de frequência usando a DWT. Um conjunto de *features* estatísticas foi extraído das subcamadas para representar a distribuição de coeficientes das *wavelets*. Estes coeficientes extraídos fornecem uma representação compacta que mostra a distribuição de energia do sinal no tempo e na frequência. As seguintes *features* estatísticas foram usados para representar a distribuição tempo-frequência dos sinais pulmonares: i) média dos valores absolutos dos coeficientes em cada subcamada; ii) potência média dos coeficientes e *wavelets* em cada subcamada; iii) desvio padrão dos coeficientes em cada subcamada; iv) rácio dos valores médios absolutos das subcamadas adjacentes. As *features* i) e ii) representam a distribuição da frequência do sinal e as *features* iii) e iv) a quantidade de alterações na distribuição da frequência. A exatidão do método é de 81%.

Em [Bahoura and Pelletier, 2004], propõem o algoritmo GMM para classificar os SRs em duas categorias: normal e sibilância. São seguidas duas abordagens para extração de *features*: a primeira utiliza 24 MFCCs e a segunda utiliza 24 coeficientes de *subband based cepstral*. O método proposto é também comparado com outras redes neuronais: LVQ e *Multi-Layer Perceptron* (MLP). O melhor desempenho é obtido com o GMM usando os MFCCs, com uma sensibilidade de 70% e especificidade de 60%.

Em [Güler et al., 2005], os sons pulmonares foram obtidos de 96 indivíduos, 56

deles sofriam de doenças pulmonares e os restantes eram indivíduos saudáveis. De cada intervalo sonoro com duração de 15 a 20 s foram selecionados ciclos completos da respiração. De cada ciclo respiratório selecionado, calculou-se a PSD da potência de *Fourier*. Para prever sons normais, sibilâncias e ferveiros, interpretando os sons pulmonares, esses mesmos sons pulmonares são processados por meio do método *Welch* e os cálculos de estimativa da PSD (métodos de extração de *features*) são aplicados ao algoritmo genético<sup>5</sup> híbrido e às redes neurais (GANN). Neste estudo, usando uma rede supervisionada com aprendizagem de retropropagação e arquitetura MLP, obtiveram-se exatidões de classificação de 81-91% e 83-93% usando (só rede neuronal) e ANN e GA (rede neuronal com algoritmo genético), respetivamente.

Na investigação em [Martinez-Hernandez et al., 2006], um conjunto de 10 indivíduos formaram o grupo saudável enquanto o grupo de pacientes continha 19 indivíduos. Um conjunto de microfones gravou os SRs nas costas dos indivíduos. A extração de *features* compreendia um *Multivariate Autoregressive Modelling* (MAR), a redução da dimensionalidade dos vetores de *features* foi realizada por *singular value decomposition* e PCA, e a classificação ficou assegurada por uma ANN supervisionada de retropropagação baseada na regra de *Levenberg-Marquardt*. O MAR em combinação com PCA forneceu o melhor resultado, sendo possível diferenciar sons normais de anormais com uma exatidão de classificação de 87,68%, sensibilidade e especificidade de 81,36% e 83,64% respetivamente.

O objetivo do estudo em [Chien et al., 2007] foi classificar os SRs em normais e anormais (sibilância) usando a análise *cepstral* com base nos GMM. O sinal sonoro foi dividido em segmentos sobrepostos, os quais são caracterizados por vetores de *features* de dimensões reduzidas usando MFCCs. Durante a fase de teste, um som desconhecido foi comparado a todos outros através dos modelos GMM e a classificação é baseada no critério *Maximum Likelihood*<sup>6</sup>. Este critério baseia-se no valor limite: se o limite for maior que zero o som é normal, caso contrário é sibilância. O som foi gravado usando um microfone condensador (*ECM, KEC-2738*) com um estetoscópio (*3M Littmann Classic S.E.*). Nesta situação, a largura de banda desejada é de 60-4000 Hz para análise do pulmão e da traqueia. Utilizou-se um filtro passa-baixo *Butterworth* de 4<sup>a</sup> ordem. A exatidão na identificação de sibilâncias foi até 90%.

---

<sup>5</sup>Um algoritmo genético típico mantém uma população de soluções e implementa uma estratégia de “sobrevivência do mais apto” na busca por melhores soluções. O algoritmo genético é usado para obter uma estrutura de rede neuronal quase ótima [Güler et al., 2005].

<sup>6</sup>Teste estatístico que compara a qualidade do ajuste de dois modelos estatísticos: um modelo nulo e um modelo alternativo. O teste é baseado na razão de probabilidade, que expressa a probabilidade dos dados seguirem um modelo e não outro. Essa razão de probabilidade pode ser usada para calcular o *p-value*.

No presente estudo em [Taplidou and Hadjileontiadis, 2007], é introduzido um detetor de sibilâncias por tempo-frequência (TF-WD) que localiza e identifica automaticamente episódios de sibilâncias durante gravações de SRs com base na análise tempo-frequência (TF). O TF-WD foi avaliado com base nos SRs de 13 pacientes adultos com asma, DPOC e pneumonia, perfazendo um total de 337 sibilâncias testadas. As sibilâncias adquiridas possuíam amplitudes distintas e ocorriam em diferentes instantes do ciclo respiratório. Os resultados do detetor foram comparados aos da auscultação clínica realizada por dois especialistas. Os SRs foram registados em 5 microfones (*ECM-77B, Sony, Inc.*, Tóquio, Japão) aplicados sobre a traqueia, axilas direita e esquerda e posterior direita e esquerda dos pulmões. A duração total das gravações por sujeito foi de 5 minutos. A representação TF foi obtida através da STFT. A técnica proposta é realizada através das seguintes etapas: i) cálculo da representação TF do som gravado; ii) subtração do som da respiração “básica” ao som total da respiração; iii) deteção do pico na representação TF; iv) classificação dos picos detetados como sibilância e não sibilância. Um conjunto de critérios é aplicado para avaliar os picos detetados na etapa iii), em termos de presença de sibilância ou não. Esses critérios incluem:

- Os máximos locais, ou seja, os picos devem ter a magnitude máxima numa janela de frequência de 70 Hz;
- Coexistência de pico, ou seja, o número de picos que coexistem em cada intervalo de tempo não deve ser maior que 4, levando em consideração que as sibilâncias geralmente não têm mais que 3 harmónicos;
- Continuidade no tempo, isto é, os picos devem ter uma duração maior que 150 ms;
- O agrupamento, isto é, os picos pertencem à mesma sibilância quando a proximidade da frequência dos picos referentes a posições sucessivas da janela do tempo não excede 50 Hz.

Os picos detetados que satisfazem os critérios acima mencionados correspondem a sibilâncias, caso contrário, são sons sem sibilâncias e são descartados. Em termos de desempenho, obteve-se uma sensibilidade de  $95,5 \pm 4,8\%$  e especificidade de  $93,7 \pm 9,3\%$ . De notar que o TF-WD tem um desempenho superior para as sibilâncias adquiridas na traqueia do que no peito. No entanto, o desempenho geral é bastante promissor, indicando que o TF-WD pode ser utilizado tanto na sibilância traqueal quanto torácica, inspiratória ou expiratória, com intensidade variável.

O artigo em [Lu and Bahoura, 2008], apresenta um sistema automatizado integrado para reconhecimento de ferveiros *fine/coarse*. Este sistema compreende 3 módulos com as seguintes funções: i) separação dos ferveiros dos SRs normais usando um filtro *Wavelet Packet Transform-Based Stationary-Nonstationary* (WPST-NST); ii) detecção de ferveiros por FD; iii) classificação de ferveiros com base nos GMM. Relativamente à extração de *features*, as características espectrais e a forma de onda são a forma mais direta de analisar as características dos ferveiros. De acordo com sua definição, os ferveiros são classificados em *fine* e *coarse* pela duração do IDW e 2CD. Além disso, existem várias outras definições para a descrição de fervor, tais como *Total Deflection Width* - Largura Total de Deflexão (TDW), LDW, *Peak Frequency* - Frequência de pico (PF) e *Gaussian Bandwidth* - Largura de Banda de Gauss (GBW). Neste caso, 4 parâmetros são testados: LDW, PF, TDW e GBW. O cálculo do TDW e do LDW é proposto em [Vannuccini et al., 1998]. O LDW é calculado por 2 pontos cruzados no eixo do tempo e tangente da deflexão máxima. Os parâmetros PF e GBW são derivados da FT do sinal. PF é a frequência na qual o espectro dos ferveiros atinge o valor máximo. GBW é aproximadamente igual à gama de frequência entre o ponto onde a amplitude do espectro decai 3 dB do seu valor mais alto. A separação de ferveiros dos sons pulmonares é baseada no facto dos ferveiros terem um carácter explosivo, com um gráfico de energia excedendo o nível de ruído. O detetor do pico dos ferveiros proposto é capaz de encontrar com alta precisão as posições (limites) dos ferveiros, mesmo que estejam disfarçadas. O filtro *denoising* usado remove a maioria dos ruídos no fervor e melhora significativamente a confiança do resultado na deteção. O melhor desempenho/exatidão (91,5%) é obtido pela análise espectral e de forma de onda (GBW, PF e LDW) e pelos GMM, com uma sensibilidade e especificidade de 95,6% e 63,3%, respetivamente.

No trabalho desenvolvido em [Alsmadi and Kahya, 2008], um sistema de diagnóstico em tempo real baseado no *DSP 56311* da Motorola foi projetado e implementado, sendo capaz de classificar os sons pulmonares em duas classes: saudável e patológica. Para tal, foi usada uma amostra testada em offline: 20 pacientes com fibrose pulmonar, pneumonia e edema pulmonar e doenças pulmonares obstrutivas como asma, bronquite e enfisema. A média de idades foi de  $36,7 \pm 11,9$  anos. E ainda, 20 indivíduos do sexo masculino não fumadores cuja idade média era  $34,4 \pm 7,4$  anos. E em tempo real testaram-se 25 indivíduos diferentes: 13 com sons patológicos e 12 com SRs saudáveis. Os sujeitos com sons patológicos eram 6 do sexo feminino e 7 do sexo masculino e sofriam de várias doenças respiratórias restritivas ou obstrutivas. A média de idades foi de  $40,1 \pm 16,4$  anos. Por outro lado, um total de 12 indivíduos saudáveis não fumadores (10 do sexo masculino e 2 do sexo feminino) foram utiliza-

dos para testar o desempenho online da classe saudável, com uma média de idades de  $35,8 \pm 11,3$  anos. Os microfones usados para captar o SR foram colocados na base posterior direita e esquerda dos pulmões. Relativamente ao processamento de sinal, o filtro passa-banda é construído por um filtro *Bessel* passa-alto de 6<sup>a</sup> ordem com um filtro *Butterworth* e um passa-baixo de 8<sup>a</sup> ordem. O filtro passa-alto é usado para reduzir o efeito do ruído de baixa frequência resultante de sons cardiovasculares, sons musculares e ruído de fricção causados pelo movimento do paciente ou do microfone. O filtro de *Bessel* é usado na configuração do filtro de alta frequência para obter uma resposta em fase aproximadamente linear, de modo que as formas de onda que contenham ferveres ou outros sons sejam minimamente afetadas. O filtro passa-baixo é um filtro *antialiasing* para o sinal que é digitalizado com uma frequência de amostragem de 8 kHz. Relativamente à extração de *features*, o método AR é usado para representar SRs segmentados em diferentes classificadores. Cada amostra do sinal sonoro respiratório,  $s(n)$ , é descrita pela equação 3.12:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + e(n) \quad (3.12)$$

onde  $p$  é a ordem do modelo,  $a_1, a_2, \dots, a_p$  são os coeficientes AR,  $e(n)$  é o erro de predição que é assumido como gaussiano. O algoritmo de *Levinson-Durbin* é usado para estimar os parâmetros AR. O sistema oferece a opção de escolher um dos dois classificadores para realizar o processo de classificação: um classificador baseado na distância mínima com a medida de distância *mahalanobis* ou um classificador k-NN (com  $k=5$ ) com a opção de escolher entre diferentes medidas de distância nomeadamente as distâncias *itakura*, euclidiana e *city-block*. O desempenho em tempo real do sistema desenvolvido na classificação de sons pulmonares em ambiente clínico foi muito promissor, tendo o melhor resultado uma exatidão de 96%, sensibilidade de 92% e especificidade de 100% para o classificador de k-NN usando a distância métrica *city-block*. O desempenho do classificador com as outras duas distâncias métricas, euclidiana e *itakura*, também resultou em altas taxas de exatidão, ambas com 88%. O alto valor de sensibilidade, mais precisamente 100% e 92% obtido com diferentes classificadores em operação online, indica que o sistema é especialmente confiável para detetar sons patológicos. Os resultados de especificidade para os classificadores com as distâncias métricas euclidiana e *itakura* também foram altos, com 83% e 75%, respetivamente.

Em [Charleston-Villalobos et al., 2011], através de 25 sensores acústicos constituídos por microfones, adquiriram-se sons pulmonares multicanais com duração de 15 s em toda a superfície torácica posterior. Uma amostra constituída por 8

indivíduos não fumadores (6 do sexo feminino e 2 do sexo masculino) formaram o grupo saudável enquanto o grupo de pacientes era composto por 19 indivíduos (13 do sexo feminino e 6 do sexo masculino) com diagnóstico clínico de doenças pulmonares intersticiais. O objetivo foi classificar os sons em normais e anormais, para tal além da técnica convencional da PSD, os valores próprios da matriz de covariância e os modelos AR univariados e multivariados foram aplicados para a construção de vetores de *features* como entrada para uma rede neuronal supervisionada. Os resultados mostraram a eficácia dos modelos AR univariados com uma exatidão de classificação de 75% e 93% para indivíduos saudáveis e doentes, respetivamente, e valores de sensibilidade de 80-100% e de especificidade de 73-100%.

Em [Serbes et al., 2011], recolheram SRs de 26 indivíduos (13 saudáveis e 13 patológicos) e propuseram um método de deteção de ferveiros no qual usaram a *dual-tree complex wavelet transform* como uma etapa de pré-processamento para remover as bandas de frequência que não continham nenhuma informação, e posteriormente extraíram vários conjuntos de *features* usando análise TF e tempo-escala. A classificação foi feita pelo algoritmo SVM. Adicionalmente, para minimizar o ruído de atrito e a interferência do som cardíaco usaram filtros passa-baixo *Butterworth* de 8<sup>a</sup> ordem com frequência de corte de 4 kHz, filtros passa-alto *Bessel* de 6<sup>a</sup> ordem com frequência de corte de 80 Hz e ainda um filtro *antialiasing*. O melhor resultado de exatidão foi de 97,20%.

O objetivo da investigação em [Hashemi et al., 2011] foi analisar as sibilâncias e classificá-las como monofónicas e polifónicas. Os sons pertenciam a 140 indivíduos diferentes com DPOC e asma, resumindo-se em 77 sibilâncias polifónicas e 63 monofónicas. Seis *features* estatísticas foram extraídas dos coeficientes *wavelet*:

- Média dos valores absolutos em cada subcamada ( $\mu_{di}$ );
- Potência média em cada subcamada ( $P_{di}$ );
- Desvio padrão em cada subcamada ( $\sigma_{di}$ );
- Rácio dos valores médios absolutos de subcamadas adjacentes ( $\frac{\mu_{di}}{\mu_{di+1}}$ );
- Assimetria em cada subcamada ( $sk_{di}$ );
- *Kurtosis* em cada subcamada ( $ku_{di}$ ).

As primeiras duas *features* enumeradas representam a distribuição da frequência do sinal, o desvio padrão e o rácio dos valores médios absolutos das subcamadas adjacentes mostram as alterações na distribuição da frequência. A assimetria mede a assimetria da distribuição e *kurtosis* mede o pico da distribuição. Ao comparar

a contribuição das *features*, concluímos que a média, o desvio padrão e a energia aparecem com mais frequência nos primeiros 100 atributos. Utilizaram uma MLP como classificador. Os resultados mostram que, usando um conjunto de 15 *features* selecionadas e uma rede MLP, as sibilâncias podem ser classificadas com uma exatidão de 89,28%.

Um trabalho bastante semelhante, em [Abbasi et al., 2013], propõe a extração de *features* de sons pulmonares utilizando coeficientes de *wavelet* e classificação por redes neuronais probabilísticas e SVM em 6 categorias: sons normais, sibilâncias, roncos, ferveores, estridor e *squawks*. As seguintes *features* foram extraídas: i) a média dos coeficientes em qualquer subcamada; ii) a média da potência dos coeficientes de *wavelet* em qualquer subcamada; iii) o desvio padrão dos coeficientes em qualquer subcamada; iv) a proporção dos valores médios nas subcamadas. Os atributos 1 e 2 demonstram a distribuição de frequência do sinal e os atributos 3 e 4 indicam as mudanças na distribuição de frequência. Os resultados revelam que o algoritmo SVM obteve um melhor desempenho na classificação de sons pulmonares normais e anormais com uma exatidão de 93,51-100%.

Neste estudo em [Morillo et al., 2013], recorreram a SRs de 53 pacientes hospitalizados devido à exacerbação aguda da DPOC. Pelo menos 3 ciclos respiratórios completos foram registados após indicar aos pacientes para respirarem o mais profundamente possível. O primeiro grupo de pacientes incluiu pacientes com ausência de sons anormais e o segundo continha pacientes com sons adventícios quando davam entrada na unidade hospitalar (principalmente sibilâncias e roncos), que poderiam diminuir significativamente ou até desaparecer no momento da alta médica. A análise de sinais não estacionários foi realizada usando a STFT e WT, que analisam os sinais simultaneamente no domínio do tempo e da frequência. Testaram 13 *features*: parâmetros da frequência (média e mediana), parâmetros espectrais (*spectral crest factor*), entropias (*Shannon*, *Rényi* e *Tsallis*), *relative power* para 8 gamas de frequência (50-200Hz, 200-400Hz, 400-800Hz e 800-2000Hz), momento de 2<sup>a</sup> ordem, *skewness* e *spectral kurtosis*. Para a redução de dimensionalidade usaram PCA. Alcançaram uma exatidão de 77,6%.

Na investigação em [Mondal et al., 2014], recolheram 120 ciclos de 30 gravações de 10 indivíduos saudáveis e 20 patológicos com disfunções pulmonares: DPOC, doenças pulmonares intersticiais e asma, a partir do sulco anterior supraesternal. Os sons anormais dos pulmões incluem sibilâncias, ferveores e *squawks*. Foram testadas 4 *features*: em termos de informações de textura (lacunaridade), índice de irregularidade (entropia da amostra), momento de 3<sup>a</sup> ordem (assimetria) e momento de 4<sup>a</sup> ordem (*kurtosis*). Salienta-se que o parâmetro lacunaridade mede informações so-

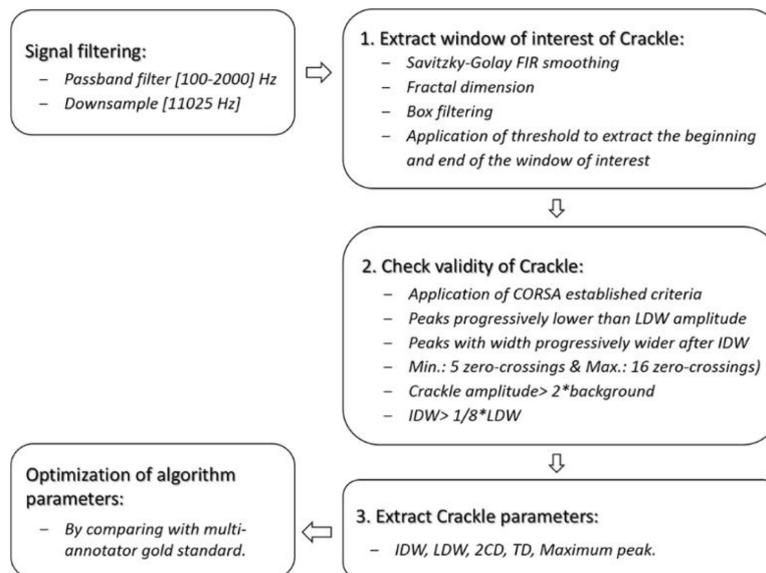
bre textura ou heterogeneidade de quaisquer objetos que possam ser fractais ou não fractais. A informação da textura dos SRs normais e patológicos pode ser medida através do índice de lacunaridade. O som pulmonar anormal é mais heterogêneo e por isso, fornece um valor de lacunaridade mais alto que o som pulmonar normal. Estes atributos são analisados usando dois classificadores diferentes: *extreme learning machine* e SVM. Os resultados são obtidos usando a validação cruzada com 5 *folds*. A *extreme learning machine* oferece um desempenho relativamente superior ao SVM. Os resultados ótimos fornecem uma exatidão de 92,86%, sensibilidade de 86,30% e especificidade de 86,90%.

No estudo em [İçer and Gengeç, 2014], analisaram os sons pulmonares de 20 indivíduos saudáveis e 40 indivíduos com diversas patologias respiratórias como DPOC, asma, fibrose pulmonar, pneumonia, tuberculose, doença pulmonar intersticial, bronquiectasias, cancro do pulmão e fungo pulmonar, com um tempo de gravação de 11 s definido automaticamente. Utilizando um estetoscópio eletrônico recolheram os sons de 6 zonas pulmonares: costas dos pacientes, incluindo a parte direita e esquerda superior, direita e esquerda média e direita e esquerda inferiores. O processo de extração de *features* compreende 3 módulos de processamento de sinais com as seguintes funções: i)  $f_{min} / f_{max}$  a partir da PSD baseada no método de *Welch*; ii) frequência instantânea média e o tempo de troca da frequência instantânea calculados pela transformada de *Hilbert Huang*; iii) os valores próprios obtidos a partir da análise do espectro. No processo de classificação, foi utilizado o algoritmo SVM para distinguir os ferveiros, roncos e sons pulmonares normais. O rácio da frequência da PSD e os valores próprios demonstram maior exatidão de classificação (90-100%) relativamente aos cálculos da média e ao tempo de troca da frequência instantânea.

Em [Chen et al., 2015], os MFCCs foram usados como *features* de um conjunto de 20 sons pulmonares e, em seguida o algoritmo *K-means* foi usado para agrupar as *features* e reduzir a quantidade de dados. Os sons pulmonares foram classificados em: normais, sibilâncias, ferveiros e roncos. A taxa média de identificação do som pulmonar normal foi de 95% enquanto os sons pulmonares anormais foi de 91,3%.

Em [Pinho et al., 2015], desenvolveram um algoritmo para deteção e caracterização automática de ferveiros e avaliaram o seu desempenho e exatidão comparando com a análise de especialistas clínicos. O algoritmo é baseado em 3 procedimentos principais: i) extração de um segmento de interesse de um potencial fervor (com base nas FDs e nas técnicas de filtragem *box*); ii) verificação da validade do potencial fervor considerando os critérios estabelecidos pela CORSA e iii) caracterização e extração de parâmetros do fervor. A partir de 10 pacientes foram selecionados 24

gravações (15 ficheiros áudio de 6 pacientes com pneumonia; com  $46 \pm 14,6$  anos, adquiridos com um estetoscópio digital *Thinklabs®* (*Thinklabs® Rhythm: ds32a*, EUA) e 9 ficheiros áudio de 4 pacientes com fibrose cística; com  $32,3 \pm 18,1$  anos, adquiridos com um estetoscópio digital *WelchAllyn* (*WelchAllyn Meditron*, 5079-402)), adquiridos durante 10 s em contextos clínicos. A filtragem de sinal consistiu na aplicação de um filtro passa-banda de 100-2000 Hz (resposta impulsional finita, projetada com uma janela *Blackman* de 83 ms), para eliminar o ruído de alta frequência, garantindo que os principais atributos dos ferveores fossem preservados. A extração de uma janela de interesse implica diferentes etapas do processamento do sinal: i) suavização da resposta impulsional finita de Savitzky-Golay (polinómio) para separar os picos de ruído das restantes altas frequência; ii) estimativa da FD; iii) filtragem *box*; e iv) aplicação de um limite para extrair o início e o fim de uma janela de interesse. O diagrama da Figura 3.1 representa o algoritmo proposto para detetar ferveores. O desempenho do algoritmo foi avaliado por comparação com as anotações de três especialistas clínicos. Foram alcançados altos níveis de sensibilidade: 89%, valor preditivo positivo: 95% e desempenho geral: 92%.



**Figura 3.1:** Algoritmo proposto por [Pinho et al., 2015] para detetar ferveores.

No artigo em [Mendes et al., 2015], através de um estetoscópio eletrónico *3M Littman* (modelo 3200) recolheram dados de 12 indivíduos ( 9 patológicos e 3 saudáveis), sendo que os SRs de 6 pacientes continham sibilâncias ou sibilâncias e ferveores, enquanto os indivíduos saudáveis apresentavam apenas SRs normais, e o restante conjunto de 3 pacientes apresentava apenas ferveores. As posições de auscultação foram selecionadas entre as 6 posições possíveis ilustradas na Figura 3.2.

Para cada sujeito, selecionaram os dados adquiridos nas duas posições em que os SRs eram mais audíveis. Assim sendo, foram utilizadas 24 aquisições de aproximadamente 30 s.



**Figura 3.2:** Posições para a aquisição de sons (vermelho). Para cada voluntário, selecionaram os sons adquiridos nas duas posições em que os sons adventícios/normais eram mais audíveis [Mendes et al., 2015].

Neste trabalho, 30 *features* foram testadas com o objetivo de identificar o melhor conjunto de *features* para a detecção de sibilâncias. As 30 *features* incluem a detecção de *Wheezes Signature in the Spectrogram Space* - Sibilâncias no espaço do espectrograma (WS-SS) e 29 *features* musicais bastante usadas no contexto da recuperação de informações musicais. O método proposto para a detecção de WS-SS possui as seguintes etapas: i) filtrar o sinal; ii) calcular o espectrograma; iii) subtração do fundo; iv) detecção de pico; v) redução de falsos positivos; vi) cálculo da matriz de pesos ( $\omega$ ). As 29 *features* musicais, calculadas usando a *MIRtoolbox*, incluem: *Root-mean square energy*, *Brightness*, *Centroid*, *Flatness*, *Irregularity*, *Keyclarity*, *Kurtosis*, 13 MFCCs, *Rolloff 85*, *Rolloff 95*, *Roughness*, *Spread*, *Skewness*, *Zerocross*, *Chromagram centroid*, *Chromagram peak* e *Mode* (estas *features* encontram-se na Tabela B.1 da secção B dos anexos e serão detalhadamente explicadas na secção 4.2 do capítulo dos métodos). Para avaliar o desempenho das diferentes *features* usaram o classificador *Logistic Regression* - Regressão Logística (LR) e o classificador *Random Forest* (RF). Quatro algoritmos de seleção de *features* foram usados para classificar a importância das *features*: *the least absolute shrinkage and selection operator* (LASSO), a importância da variável estimada usando um RF (RF-VI), a seleção sequencial de *features* na direção para frente (*SFS-Forward*) e na direção para trás (*SFS-Backward*). A função objetivo usada na seleção sequencial de *features* foi a maximização do coeficiente de correlação de *Matthews* (MCC<sup>7</sup>) medida após a classificação dos dados usando o classificador LR ou RF. Todos os algoritmos de seleção classificaram a *feature* WS-SS como a mais importante. Por outro lado, o uso de

---

<sup>7</sup>MCC: é uma medida de desempenho balanceada, especialmente adequada quando o conjunto de dados está desequilibrado.

mais de 10 *features* permitiu apenas um pequeno aumento do valor do MCC, ou seja, a adição de mais *features* conduz a uma melhoria pouco significativa. Como esperado, os critérios de desempenho (MCC, especificidade, sensibilidade e exatidão) foram maiores com o classificador RF. Utilizando 30 *features* combinadas com esse classificador, observa-se uma MCC, sensibilidade, especificidade e exatidão iguais a  $92,7 \pm 1\%$ ,  $90,9 \pm 2\%$ ,  $99,4 \pm 1\%$  e  $97,9 \pm 1\%$  respetivamente.

No estudo em [Mendes et al., 2016], é proposta uma abordagem para a deteção de segmentos que contenham um ou mais ferveores, para tal foi testado o desempenho de 35 *features*. Essas *features* incluem 31 atributos normalmente usados no contexto de recuperação de informações musicais (Tabela B.1 da secção B dos anexos e serão detalhadamente explicadas na secção 4.2 do capítulo dos métodos, com exceção da  $F_0$ , *Harmonic Change Detection Function* e *Spectral Rolloff* de 75 em vez de 95), um atributo baseado em *wavelets* proposto por [Bahoura, 2009] para detetar ferveores, bem como a energia de Teager. Um atributo especialmente projetado para detetar sibilâncias no espaço do espectrograma (WS-SS), também foi incluído no conjunto de atributos testados [Mendes et al., 2015]. A classificação foi realizada utilizando um classificador de LR. O conjunto de dados foi recolhido de 20 voluntários, 17 pacientes com presença de sons adventícios e 3 indivíduos saudáveis. Estes dados sonoros incluíam ferveores, sibilâncias e sons normais do pulmão. Dentro da amostra, 15 pacientes manifestavam presença ferveores ou ferveores juntamente com outros sons adventícios e dois pacientes continham sons adventícios que não eram ferveores. As aquisições foram realizadas usando um estetoscópio eletrónico *3M Littman* (modelo 3200) e as posições de auscultação são as mesmas que as representadas na Figura 3.2. Deste modo, um total de 40 gravações de som com aproximadamente 30 s cada foram utilizadas neste estudo. Para as condições testadas, os resultados mais promissores foram obtidos para o tamanho da frame igual a 128 ms e 27 *features*, com um desempenho de 76% de sensibilidade e 77% de exatidão.

Num estudo em [Aykanat et al., 2017], recolheram-se 11 gravações de áudio de cada um dos 1630 indivíduos saudáveis e doentes, totalizando 17 930 ficheiros de áudio. Aplicaram dois tipos de algoritmos de aprendizagem: MFCCs num algoritmo SVM e imagens do espectrograma usando STFT numa *Convolutional Neural Network* (CNN). Para classificar o som, avaliaram-se 4 grupos para cada algoritmo CNN e SVM: i) classificação saudável vs patológica; ii) roncos e classificação sonora normal; iii) classificação singular do tipo de SR; e iv) classificação do tipo de áudio. Exatidão dos resultados experimentais foram: i) CNN 86%, SVM 86%, ii) CNN 76%, SVM 75%, iii) CNN 80%, SVM 80% e iv) CNN 62%, SVM 62%, respetivamente. Como resultado, observou-se que a classificação da imagem por espectrograma com

CNN funciona tão bem quanto o SVM, e dada a grande quantidade de dados, CNN e SVM podem classificar e pré-diagnosticar SR.

O principal objetivo em [Rocha et al., 2018], foi projetar um método para a detecção automática de tosse e SRs adventícios recorrendo a gravações áudio de 18 pacientes, adultos e idosos, que apresentaram DPOC com comorbidades (insuficiência cardíaca, diabetes, hipertensão). Os sons foram recolhidos em 6 locais do tórax, de acordo com a Figura 3.2. Na fase de pré-processamento, o sinal de áudio foi filtrado, usando um filtro passa-alto de resposta impulsional infinita de 2ª ordem (IIR) a 80 Hz (abaixo do limite inferior típico da voz humana adulta), e normalizado. A *MIRtoolbox* foi usada para extrair 35 *features* musicais relacionadas com a dinâmica, timbre, afinação e conteúdo harmónico (Tabela B.1 da secção B dos anexos e serão detalhadamente explicadas na secção 4.2 do capítulo dos métodos). Outras 7 *features* foram extraídas: i) o atraso do grupo *chirp*, ii) a razão harmónica / ruído (HNR) foi calculada para as faixas de frequência (0-500 e 0-1500 Hz), iii) o máximo da entropia e iv) o máximo da energia de Teager em cada frame, v) o máximo da KFD do filtro WPST-NST e vi) WS-SS. Os sons foram classificados pelo algoritmo RF. Para avaliar o modelo, 10 versões do conjunto de dados foram usadas: i) completo, com 5 classes (tosse, sibilância, ferveores, fala, outras); ii) intercalado, com 4 classes, onde sibilâncias e ferveores foram fundidos (tosse, sons adventícios, fala, outros); iii) completo 50, ou seja, completo com os melhores 50 atributos; iv) intercalados 50, ou seja, intercalados com os 50 melhores atributos; v) completo 20; vi) intercalados 20; vii) completo 10; viii) intercalados 10; ix) completo 5; x) intercalados 5. Os melhores resultados obtiveram-se com o ii) (intercalado com 4 classes) e com o iv) (intercalo com os melhores 50 atributos), ambos com uma especificidade de 90%, sensibilidade de 69%, *precision* de 70% e *f-measure* de 69%. A Tabela C.1 presente na secção C dos anexos mostra os resultados para todos os conjuntos testados.

## 3.2 Limitações do estado de arte

Tanto no caso da população infantil como na população adulta existem várias limitações aos estudos apresentados anteriormente, tais como o i) número de indivíduos analisados ser pequeno; ii) não realizarem estratificação de idades, no caso das crianças; iii) análise de sons adventícios isolados e não em conjunto, ou simplesmente não diferenciá-los (isto é, fundi-los); iv) pouca variedade de *features* utilizadas e v) pouca variabilidade nos locais de aquisição dos SRs.

## 4

# Métodos

Após a revisão de literatura dos trabalhos neste âmbito, delimitou-se a metodologia a utilizar nesta dissertação.

Numa fase primordial, estudou-se a base de dados dos sons respiratórios (SRs) das crianças. Seguidamente, extraíram-se *features* de acordo com os estudos em [Mendes et al., 2015, Mendes et al., 2016, Rocha et al., 2018], classificaram-se os SRs, selecionaram-se as *features* mais relevantes consoante a sua performance, e por fim otimizaram-se parâmetros. O procedimento foi repetido para a população adulta.

## 4.1 Base de dados

Os ficheiros de sons utilizados foram fornecidos pelo Laboratório de Investigação e Reabilitação Respiratória - Lab3R da Escola Superior de Saúde da Universidade de Aveiro (ESSUA) e por uma base de dados online, criada por outros elementos do grupo de investigação, que contém anotações e informações acerca dos indivíduos auscultados, disponível em [ICBHI Challenge, 2017].

### 4.1.1 População infantil

Numa primeira fase, a população alvo são crianças, pelo que se excluíram para esta amostra todos os sons de sujeitos com idade igual ou superior a 18 anos. Assim, utilizou-se um total de 76 sons (44 sons normais, 32 sons anormais que incluem sibilâncias e/ou ferveores) de 49 indivíduos distintos (saudáveis e com diferentes patologias respiratórias) que se encontram caracterizados pormenorizadamente na Tabela 4.1. Por cada ficheiro de som .wav, existem ficheiros .txt de anotações com o instante de início e fim de cada som adventício (sibilâncias ou/e ferveores). Todas as amostras de som são monocanal, têm uma duração de 20 segundos e são adquiridas na traqueia e noutras 6 regiões do corpo que incluem costas e tórax: anterior esquerdo e direito, posterior esquerdo e direito, lateral esquerdo e direito, tal como se encontra esquematizado na Figura 3.2. A maioria dos SRs utilizados neste estudo fo-

#### 4. Métodos

ram recolhidos através do estetoscópio eletrónico *WelchAllyn Meditron Master Elite Electronic Stethoscope* (Meditron), à exceção de um som recolhido através do *3M Littmann Classic II SE Stethoscope* (LittC2SE).

**Tabela 4.1:** Descrição da base de dados utilizada para a população infantil.

Pac. <sup>1</sup>	Id. <sup>2</sup>	Sexo	Saud. <sup>3</sup>	Doença	N <sup>o</sup> grav. <sup>4</sup>	Local da aquisição	Som adventício	N <sup>o</sup> anotações de sons adventícios
101	3	F		IRTS <sup>5</sup>	2	Al <sup>6</sup> Pr <sup>7</sup>	–	–
102	0,75	F	X		1	Ar <sup>8</sup>	–	–
105	7	F		IRTS	1	Tc <sup>9</sup>	–	–
108	3	M		IRTI <sup>10</sup>	1	Al	–	–
115	0,58	M		IRTI	1	Ar	sibilâncias	1
119	2	F		IRTS	1	Ar	–	–
121	13	F	X		2	Tc	–	–
123	5	M	X		1	Al	–	–
125	14	M	X		1	Tc	fervores	2
126	1	F	X		1	Al	–	–
127	2	M	X		1	Ar	fervores	5
129	6	M		IRTS	1	Ar	fervores	15
131	3	M		IRTS	1	Al	sibilâncias	2
136	5	M	X		1	Ar	–	–
137	4	M		IRTS	2	Ar Ll <sup>11</sup>	–	–
143	0,25	F	X		1	Al	–	–
144	3	M	X		2	Al Tc	–	–
148	4	M		IRTS	1	Al	–	–
						Al	sibilâncias	19
							fervores	3
149	0,67	M		bronquiolite	3	Lr <sup>12</sup>	fervores	3
						Pl	sibilâncias	6
							fervores	3

150	0,67	F		IRTS	1	Al	sibilâncias	1
							fervores	13
152	16	M	X		1	Al	–	–
153	3	M	X		1	Al	–	–
						Al	sibilâncias	1
159	0,83	F	X		4	Ar	–	–
						Ll	–	–
						Pr	–	–
						Al	fervores	8
161	2	F		bronquiolite	2	Al	sibilâncias	5
						Pl <sup>13</sup>	sibilâncias	7
164	1	M		IRTS	1	Ll	fervores	1
							sibilâncias	3
						Ar	fervores	8
165	2	F		IRTS	3	Pl	–	–
						Pr	sibilâncias	1
							fervores	1
167	1	F		bronquiolite	2	Al	–	–
						Pr	sibilâncias	10
171	9	M	X		1	Al	–	–
173	3	M		bronquiolite	1	Al	fervores	6
						Al	sibilâncias	1
179	10	F	X		2		fervores	5
						Tc	fervores	4
							sibilâncias	1
182	11	M	X		1	Tc	fervores	9
183	14	F	X		2	Pl	fervores	5
						Tc	–	–
184	2	F	X		1	Ar	–	–
187	0,5	F	X		1	Ll	–	–
						Al		
188	3	M		IRTS	4	Ar	–	–
						Pl		
						Tc		
190	3	F		IRTS	1	Tc	–	–

#### 4. Métodos

194	2	M	X		2	Lr Pr	–	–
197	16	F		IRTS	2	Al Tc	– sibilâncias fervores	– 1 2
202	2	M	X		1	Ar	–	–
206	3	M		bronquiolite	3	Ar Lr Pl	sibilâncias fervores sibilâncias	8 1 12 11
208	5	F	X		1	Pr	–	–
209	14	F	X		1	Tc	–	–
210	1	F		IRTS	2	Al Ar	fervores fervores	1 1
214	5	M	X		1	Ar	–	–
216	1	M		bronquiolite	2	Al Pl	fervores sibilâncias	6 9
217	12	F	X		1	Tc	–	–
224	10	F	X		2	Tc Al	–	–
225	0,83	M	X		1	Pl	–	–
226	4	M		pneumonia	3	Al Ll Pl	– fervores	2 9 9

<sup>1</sup> Paciente; <sup>2</sup> Idade; <sup>3</sup> Saudável; <sup>4</sup> N<sup>o</sup> Gravações; <sup>5</sup> **IRTS**: Infecção Respiratória do Trato Superior; <sup>6</sup> **Al**: *Anterior left*; <sup>7</sup> **Pr**: *Posterior right*; <sup>8</sup> **Ar**: *Anterior right*; <sup>9</sup> **Tc**: *Trachea*; <sup>10</sup>**IRTI**: Infecção Respiratória do Trato Inferior; <sup>11</sup> **Ll**: *Lateral left*; <sup>12</sup> **Lr**: *Lateral right*; <sup>13</sup>**Pl**: *Posterior left*;

Em seguida, sintetiza-se a informação presente na Tabela 4.1 referente aos dados utilizados para a população infantil.

- **N<sup>o</sup> pacientes:** 49;
- **Média de idades:** 4,8 anos;
- **N<sup>o</sup> gravações:** 76;
- **N<sup>o</sup> pacientes com fervores:** 17;

- N<sup>o</sup> pacientes com sibilâncias: 13;
- N<sup>o</sup> pacientes sem sons adventícios: 28;
- N<sup>o</sup> de sibilâncias anotadas: 99;
- N<sup>o</sup> de fervores anotados: 121;
- N<sup>o</sup> pacientes doentes: 23;
- N<sup>o</sup> pacientes saudáveis: 26;
- N<sup>o</sup> pacientes sexo feminino: 23;
- N<sup>o</sup> pacientes sexo masculino: 26.

Com o intuito de obter uma análise detalhada, os testes realizados partiram de diferentes amostras: i) todos os sons de crianças com menos de 18 anos; estratificação por idades: ii) 0-2 anos, iii) 3-6 anos, iv) mais de 7 anos, v) apenas sibilâncias (crianças) e vi) apenas fervores (crianças).

Realizar experiências com os sons adventícios separados (sibilâncias e fervores) é mais simples e é o estudo apresentado em quase todos os trabalhos da literatura. As amostras com vários tipos de sons adventícios tornam o problema mais realista, embora mais complexo, e por norma, os resultados são inferiores.

Nas crianças a estratificação por idades faz sentido visto que estas passam por diferentes fases de desenvolvimento do sistema respiratório. Assim sendo, os SRs das crianças foram agrupados de acordo com a maturidade do seu sistema respiratório, que se divide em três estágios conforme a idade: i) crescimento exponencial do número de alvéolos, de 20 para aproximadamente 300 milhões (0-2 anos) ii) desenvolvimento de ventilação colateral através da formação de poros do canal de Kohn e Lambert (3-6 anos) e iii) deposição de elastina no pulmão que diminui a resistência das vias aéreas respiratórias (7-8 anos) [Oliveira, 2014]. Os SRs estão diretamente relacionados com as alterações do tecido pulmonar e, conseqüentemente, com a idade das crianças.

Salienta-se que a amostra que contém sons de crianças com mais de 7 anos não pôde ser analisada devido à desigualdade entre as classes (é composta por apenas 3 sibilâncias). Por conseguinte, esta amostra foi descartada neste estudo.

#### 4.1.2 População adulta

Para a população adulta consideraram-se todos os sons de sujeitos com idade igual ou superior a 18 anos. Assim, utilizou-se um total de 847 sons que incluem 1799 sibilâncias e 8776 fervores de 76 pacientes distintos com idade média de 67,6 (saudáveis e com diferentes patologias respiratórias: asma, DPOC, infecção respiratória do trato superior, infecção respiratória do trato inferior, pneumonia e bronquiectasia). Tal como para a população infantil, por cada ficheiro de som .wav, exis-

tem ficheiros .txt de anotações com o instante de início e fim de cada som adventício (sibilâncias ou/e ferveores). Existem gravações de som monocanal e multicanal, com durações de 10 a 90 segundos, que foram adquiridas nos mesmos locais da população infantil, ou seja, na traqueia e noutras 6 regiões do corpo que incluem costas e tórax: anterior esquerdo e direito, posterior esquerdo e direito, lateral esquerdo e direito, tal como se encontra esquematizado na Figura 3.2. Os SRs foram recolhidos através do microfone AKG C417L (AKGC417L), estetoscópio 3M *Littmann Classic II SE* (LittC2SE), estetoscópio eletrónico 3M *Littmann 3200* (Litt3200) e estetoscópio eletrónico *WelchAllyn Meditron Master Elite* (Meditron).

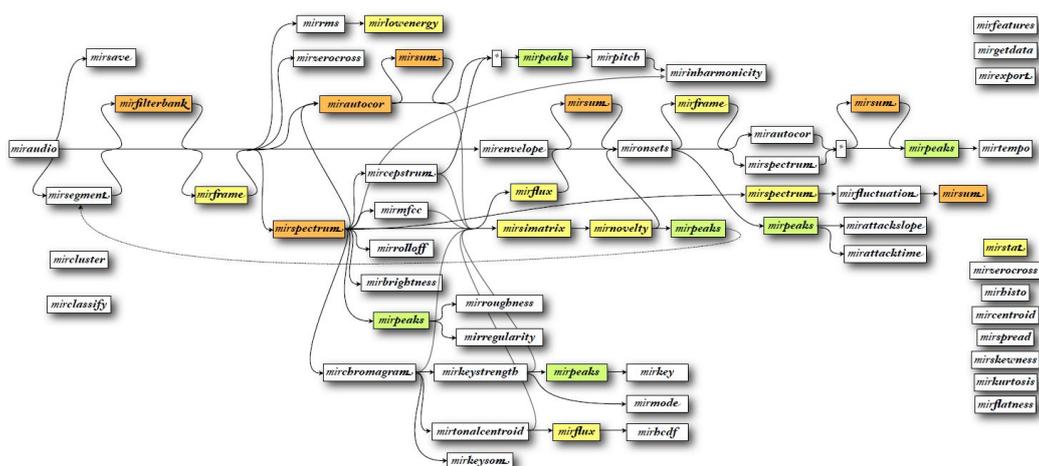
Atendendo à desigualdade das 3 classes para a amostra dos adultos, utilizou-se o filtro *SpreadSubSample* no Weka para produzir uma subamostra aleatória do conjunto de dados, na qual algumas instâncias foram eliminadas com o intuito de criar um conjunto de dados balanceado com um máximo de 2000 eventos para cada classe. Escolheu-se um máximo de 2000 eventos por classe devido à distribuição desigual das classes (8791 ferveores, 3331 sons normais e 1819 sibilâncias).

## 4.2 *Features* extraídas

Neste estudo usaram-se 42 *features*, das quais 35 são *features* implementadas através da *MIRtoolbox* e estão relacionadas com a dinâmica, timbre, afinação e conteúdo harmónico. Estas apresentam-se de seguida [Olivier Lartillot et al., 2007, Mendes et al., 2015, Mendes et al., 2016, Rocha et al., 2018].

### 4.2.1 *Features* da *MIRtoolbox*

A *MIRtoolbox* oferece um conjunto de funções escritas em Matlab que inclui cerca de 50 *features* de áudio e música [Olivier Lartillot et al., 2007]. Uma visão geral da maioria das *features* disponíveis nesta *toolbox* pode ser vista na Figura 4.1.



**Figura 4.1:** Visão geral das *features* disponíveis na *MIRtoolbox* [Olivier Lartillot et al., 2007].

### Categoria: Dinâmica

#### 1. *Feature*: RMS

**Descrição:** A energia global de um sinal,  $x$ , pode ser calculada pela equação 4.1. A raiz quadrada da média do quadrado da amplitude também chamada RMS, é descrita pela equação 4.2.

$$\sum_n |x(n)|^2 \quad (4.1)$$

$$\sqrt{\frac{1}{N} \sum_n |x(n)|^2} \quad (4.2)$$

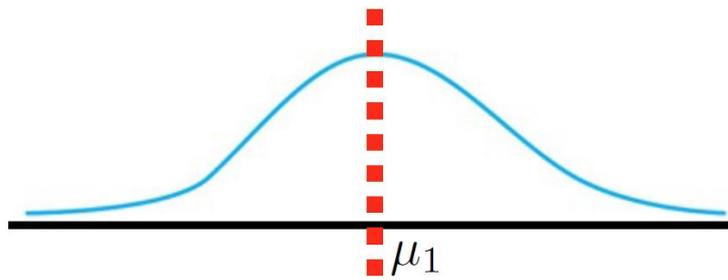
### Categoria: Timbre

#### 1. *Feature*: *Spectral centroid*<sup>1</sup>

**Descrição:** Uma descrição importante e útil da forma de uma distribuição pode ser obtida através do uso dos seus momentos. O 1º momento, designado de média, é o centro geométrico (centróide) da distribuição e mede a tendência central da variável aleatória (Figura 4.2). Calcula-se pela equação 4.3.

$$\mu_1 = \int x f(x) dx \quad (4.3)$$

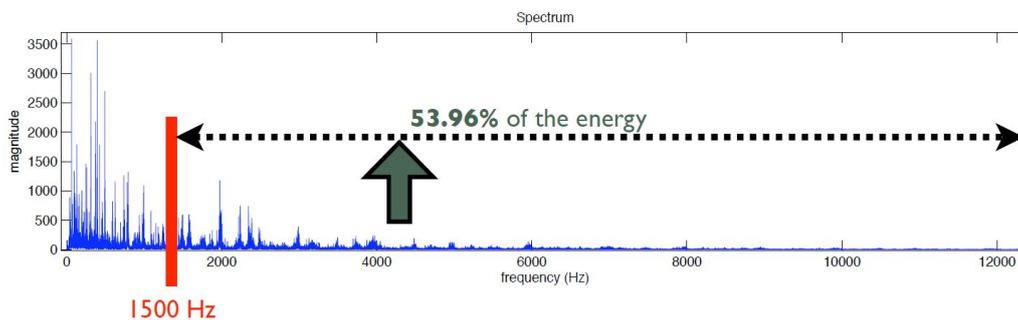
<sup>1</sup>Usou-se os termos em inglês por uma questão de uniformidade com o estado de arte.



**Figura 4.2:** Representação *spectral centroid* [Olivier Lartillot et al., 2007].

### 2. *Feature: Spectral brightness*

**Descrição:** Consiste em fixar uma frequência de corte e medir a quantidade de energia acima dessa frequência (1500Hz), tal como se exemplifica na Figura 4.3.

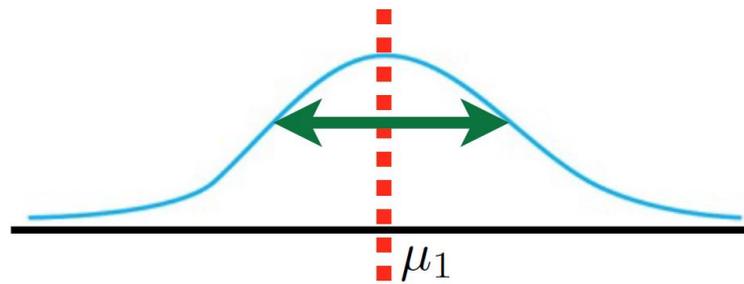


**Figura 4.3:** Representação *spectral brightness* [Olivier Lartillot et al., 2007].

### 3. *Feature: Spectral spread*

**Descrição:** O 2º momento central, designado de variância, representa-se por  $\sigma^2$  e é definido como sendo o desvio da variável aleatória em relação ao seu valor médio ao quadrado. A variância é sempre positiva e é uma medida da dispersão da variável aleatória (Figura 4.4 e equação 4.4).

$$\sigma^2 = \mu_2 = \int (x - x_1)^2 f(x) \, dx \quad (4.4)$$



**Figura 4.4:** Representação *spectral spread* [Olivier Lartillot et al., 2007].

#### 4. **Feature:** *Spectral skewness*

**Descrição:** O 3º momento central é designado de assimetria e avalia a simetria da distribuição (equação 4.5). A assimetria pode ter um valor positivo, nesse caso diz-se que a distribuição é inclinada positivamente com alguns valores muito maiores que a média e, portanto, prolonga-se mais para a direita (Figura 4.5). Uma distribuição inclinada negativamente prolonga-se mais para a esquerda. Uma distribuição simétrica tem uma assimetria zero. O coeficiente de assimetria é a razão entre a assimetria e o desvio padrão ao cubo (equação 4.6). O coeficiente de assimetria possui unidades mais convenientes do que a assimetria e geralmente varia de -3,0 a 3,0. Analogamente, uma distribuição simétrica tem um coeficiente de assimetria igual a zero e um coeficiente positivo de assimetria indica que a distribuição exibe uma concentração à esquerda e prolonga-se para a direita, enquanto um valor negativo indica o contrário. É o coeficiente de assimetria que se usa neste trabalho.

$$\mu_3 = \int (x - \mu_1)^3 f(x) \, dx \quad (4.5)$$

$$\frac{\mu_3}{\sigma_3} \quad (4.6)$$



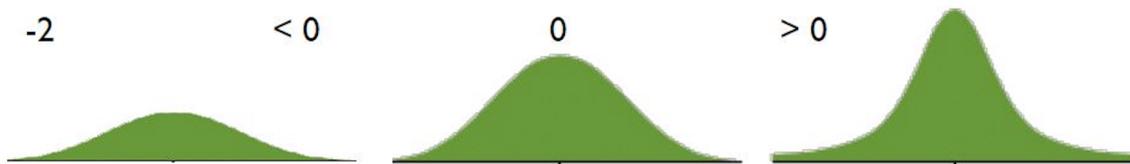
**Figura 4.5:** Representação *spectral skewness* [Olivier Lartillot et al., 2007].

5. **Feature:** *Spectral kurtosis*

**Descrição:** O 4º momento padronizado é definido pela equação 4.7 e o excesso de *kurtosis* é definido pela equação 4.8. O “menos 3” no final desta fórmula representa uma correção para tornar o achatamento da distribuição normal igual a zero (Figura 4.6).

$$\frac{\mu_4}{\sigma_4} \tag{4.7}$$

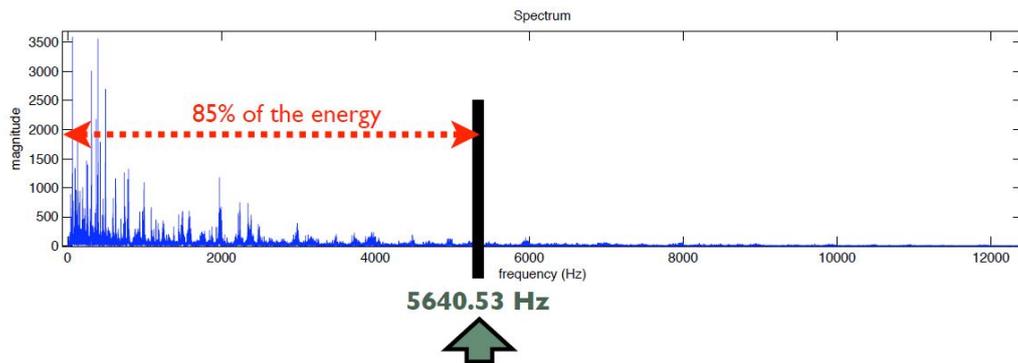
$$\frac{\mu_4}{\sigma_4} - 3 \tag{4.8}$$



**Figura 4.6:** Representação *spectral kurtosis* [Olivier Lartillot et al., 2007].

6. **Feature:** *Spectral rolloff 85*

**Descrição:** Frequência tal que 85% da energia total está contida abaixo dessa frequência, tal como se representa na Figura 4.7.



**Figura 4.7:** Representação *spectral rolloff 85* [Olivier Lartillot et al., 2007].

7. **Feature:** *Spectral rolloff 95*

**Descrição:** Frequência tal que 95% da energia total está contida abaixo dessa frequência.

8. **Feature:** *Spectral entropy*

**Descrição:** Refere-se à entropia relativa de *Shannon*. A entropia de *Shannon* é definida pela equação 4.9. Para obter uma medida de entropia independente do comprimento da sequência usa-se a entropia relativa (equação 4.10). A entropia de *Shannon* descreve a curva de entrada  $p$  e indica se contém picos predominantes ou não. Caso a curva seja extremamente plana, correspondendo a uma situação de máxima incerteza relativamente à saída da variável aleatória  $x$  da função de massa de probabilidade  $P(x_i)$ , então a entropia é máxima. Contrariamente, se a curva contiver apenas um pico nítido a entropia será mínima, indicando uma situação de incerteza mínima, pois a saída é determinada por esse pico.

$$H(X) := - \sum_{i=1}^n p(x_i) \log_b p(x_i) \quad (4.9)$$

$$H(p) = - \frac{\text{sum}(p. \times \log(p))}{\log(\text{length}(p))} \quad (4.10)$$

9. **Feature:** *Spectral flatness*

**Descrição:** Indica se a distribuição é suave ou pontiaguda. Define-se pela razão entre a média geométrica e a média aritmética (equação 4.11).

$$\frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum \prod_{n=0}^{N-1} x(n)}{N}\right)} \quad (4.11)$$

10. **Feature:** *Spectral roughness*

**Descrição:** Calcula os picos do espectro e a média da dissonância entre todos os pares possíveis de picos (Figura 4.8).

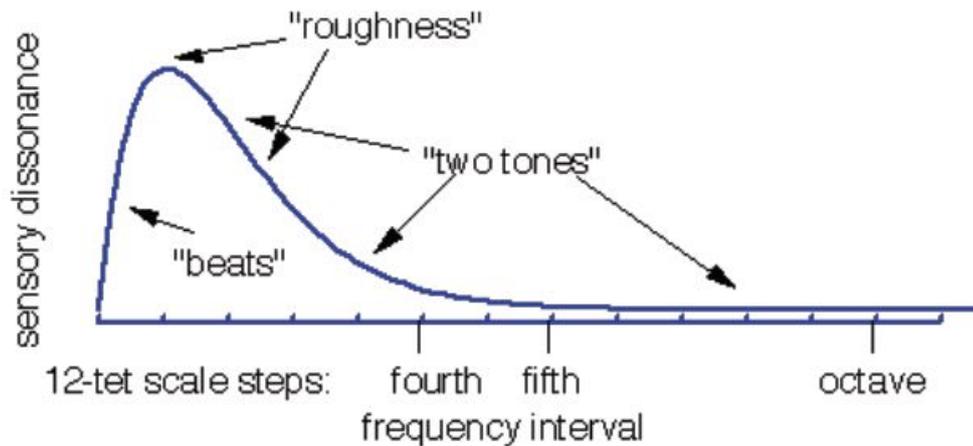


Figura 4.8: Representação *spectral roughness* [Olivier Lartillot et al., 2007].

11. **Feature:** *Spectral irregularity*

**Descrição:** Consiste no somatório do quadrado da diferença da amplitude entre as partes adjacentes [Jensen and Arnspang, 1999] (equação 4.12).

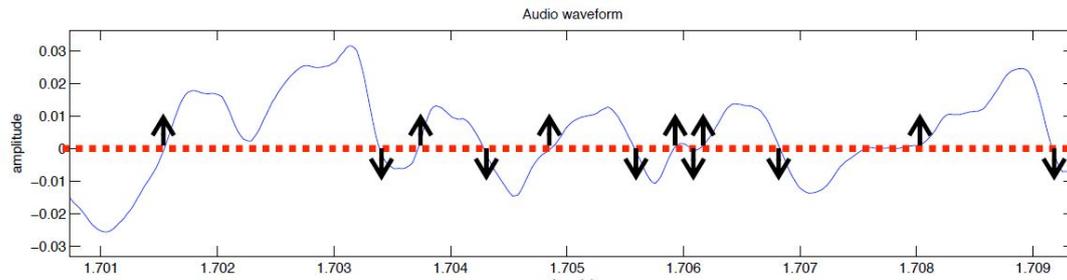
$$\frac{\sum_{k=1}^N (a_k - a_{k+1})^2}{\sum_{k=1}^N a_k^2} \quad (4.12)$$

12. **Feature:** MFCC

**Descrição:** MFCC descreve a forma espectral do som. As bandas de frequência são posicionadas logaritmicamente (na escala Mel), que faz uma melhor aproximação da resposta do sistema auditivo humano comparativamente às bandas de frequência espaçadas linearmente. Recorre à transformada discreta de cosseno que é semelhante à transformada discreta de *Fourier*, só que apenas usa números reais. Possui uma forte propriedade de “compactação de energia”: a maioria das informações do sinal tende a concentrar-se nos componentes de baixa frequência da transformada discreta de cosseno. É por isso que, por padrão, apenas os 13 primeiros componentes são retornados.

13. **Feature:** *Zero-crossing rate*

**Descrição:** Contabiliza o número de vezes que o sinal cruza o eixo  $x$ , ou seja, as mudanças de sinal/monotonia (Figura 4.9).



**Figura 4.9:** Representação *zero-crossing rate* [Olivier Lartillot et al., 2007].

### Categoria : Tonalidade

#### 1. **Feature:** *Chromagram centroid*

**Descrição:** Calcula o vetor centróide tonal de 6<sup>a</sup> dimensão do *chromagram*<sup>2</sup>. Corresponde a uma projeção dos acordes ao longo dos círculos dos 5<sup>os</sup> e dos 3<sup>os</sup> menores, e dos 3<sup>os</sup> maiores.

#### 2. **Feature:** *Chromagram peak*

**Descrição:** Pico do centróide tonal.

#### 3. **Feature:** *Key clarity*

**Descrição:** Fornece uma estimativa ampla das posições centrais tonais e da sua respectiva clareza (Figura 4.10).

<sup>2</sup>Mostra a distribuição de energia ao longo da frequência ou classes de frequência; semelhante a um histograma.

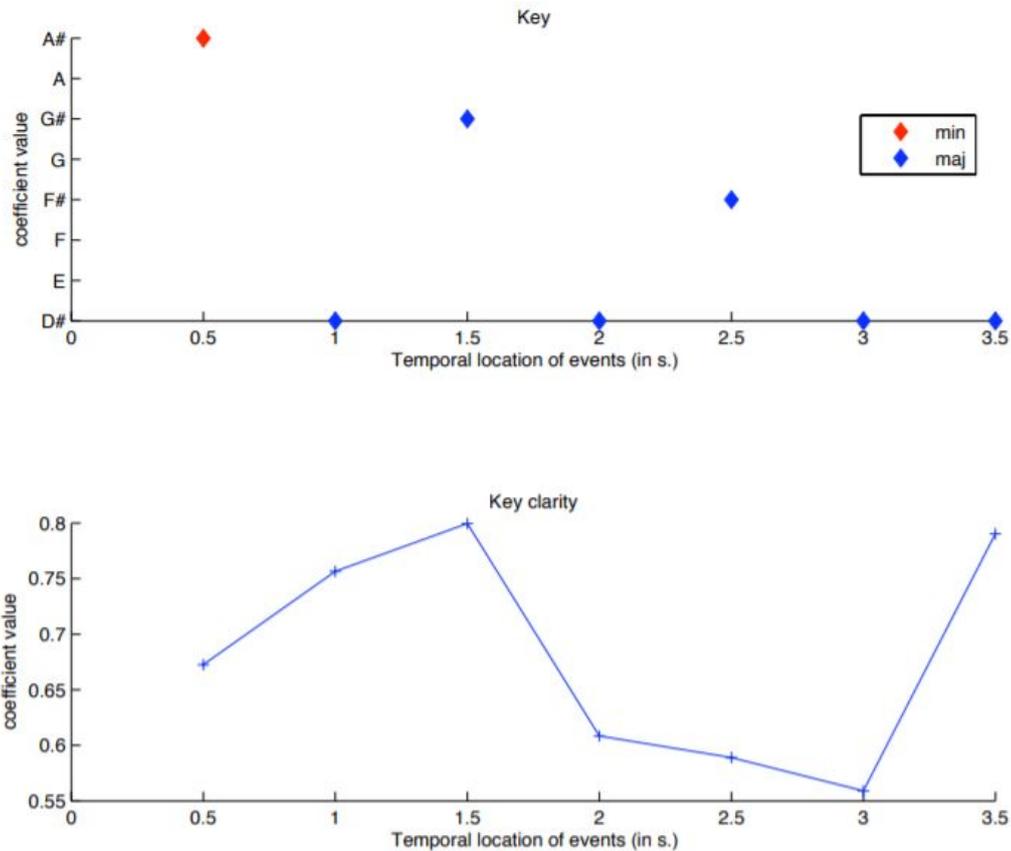


Figura 4.10: Representação *key clarity* [Olivier Lartillot et al., 2007].

#### 4. *Feature: Mode*

**Descrição:** Maior vs menor, retornando um valor numérico entre -1 e +1: quanto mais próximo de +1 maior o número de fragmentos previstos, e quanto mais próximo do valor -1 mais pequeno o fragmento pode ser.

#### 5. *Feature: Harmonic Change Detection Function (HCDF)*

**Descrição:** Fluxo do centróide tonal.

**Categoria :** Altura

##### 1. *Feature: $F_0$ cepstrum*

**Descrição:** Corresponde ao maior pico do *cepstrum* em cada frame (Figura 4.11).

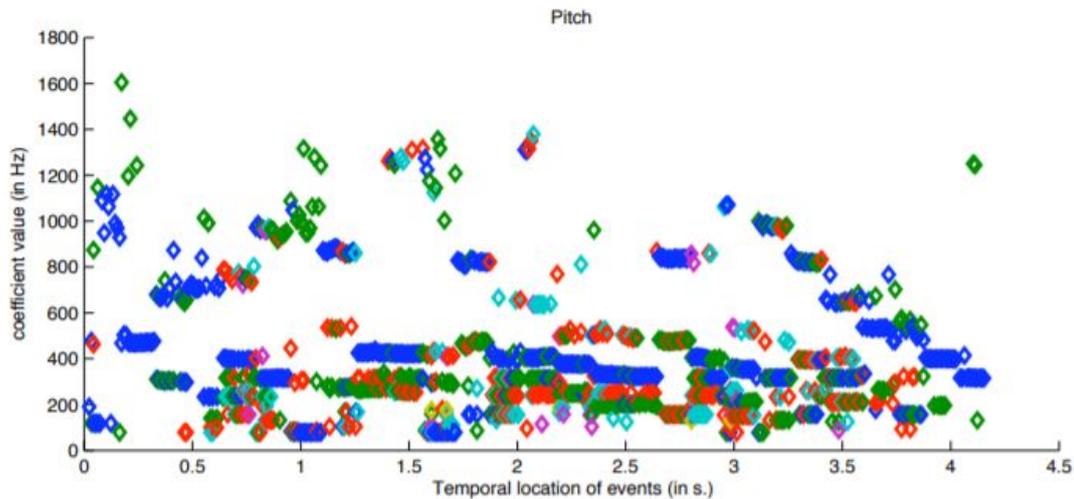


Figura 4.11: Representação  $F_0$  cepstrum [Olivier Lartillot et al., 2007].

2. **Feature:**  $F_0$  spectrum

**Descrição:** Corresponde ao maior pico espectral em cada frame.

3. **Feature:** *Inharmonicity* do  $F_0$  cepstrum

**Descrição:** Estima a inarmonicidade, isto é, a quantidade de parciais que não são múltiplos da frequência fundamental, com um valor entre 0 e 1. Mais precisamente, a inarmonicidade considera a quantidade de energia fora da série harmônica ideal. Para isso, usa-se uma função simples que estima a inarmonicidade de cada frequência, dada a frequência fundamental  $F_0$  (Figura 4.12).

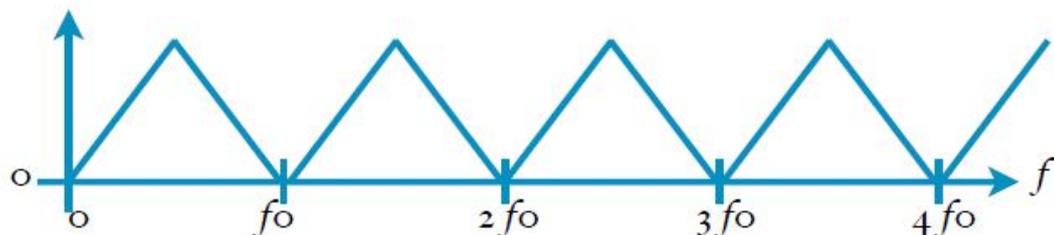


Figura 4.12: Representação *inharmonicity* do  $F_0$  cepstrum [Olivier Lartillot et al., 2007].

**Categoria:** Não aplicável

1. **Feature:** *Spectral flux*

**Descrição:** Distância euclidiana entre frames sucessivas no espectro.

### 4.2.2 Outras *features*

De acordo com [Rocha et al., 2018], extraíram-se outras 7 *features*:

- o **atraso do grupo *chirp***, que é uma medida baseada na fase proposta em [Drugman et al., 2011] para destacar turbulências durante a produção glótica, computacionalmente calcula-se os espectros de atraso do grupo de *chirp* (de sinal com fase zero) num círculo definido pelo raio R para todas as frames de fala extraídas [Bozkurt et al., 2007];
- a **razão harmónica / ruído (HNR)** foi calculada para as gamas de frequência compreendidas nos intervalos **0-500 Hz** e de **0-1500 Hz**, usando a *toolbox Voice Sauce* [Krom, 1993, Shue et al., 2009];
- o **máximo da entropia** em cada frame, sendo a entropia da informação uma medida da desordem de um sistema;
- o **máximo da energia de Teager** em cada frame;
- o **máximo da dimensão fractal de Katz (KFD) do filtro WPST-NST**;
- ***WS-SS-wheeze signature in the spectrogram space*** [Mendes et al., 2015, Mendes et al., 2016].

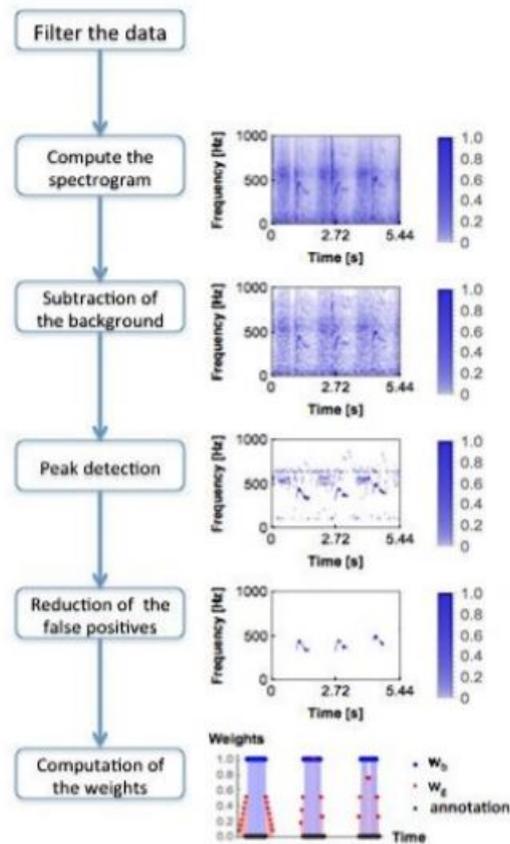
O máximo da entropia foi calculado com recurso a uma função do Matlab “*entropyfilt*”, que executa a filtragem da entropia numa imagem definida no input, em escala de cinza [Gonzalez et al., 2004].

Para cada frame foi calculado o máximo da energia de Teager do sinal normalizado. Para sinais discretos  $x[n]$ , o operador de energia Teager é dado pela equação 4.13.

$$\psi(x[n]) = x^2[n] - x[n-1]x[n+1] \quad n \in \mathbb{Z} \quad (4.13)$$

Para a deteção de ferveores, [Bahoura and Lu, 2006] propuseram um método baseado em *wavelet* (WPST-NST) usando a *wavelet Daubechies* 8 com 5 níveis de decomposição. Aplicou-se um filtro à parte não estacionária do sinal e a KFD foi calculada.

O método proposto para detetar sibilâncias no espaço do espectrograma (WS-SS) [Mendes et al., 2015, Mendes et al., 2016], já explicado na secção 3.1.2, encontra-se representado no diagrama da Figura 4.13.



**Figura 4.13:** Sibilâncias no espaço do espectrograma (WS-SS) [Mendes et al., 2015].

## 4.3 Classificadores

Nesta secção encontram-se explicados os algoritmos que posteriormente serão usados neste trabalho para a classificação dos SRs. Estes algoritmos foram escolhidos com base em diferentes paradigmas: funcionais (e.g., SMO), probabilísticos (e.g., Naive Bayes), árvores de decisão (e.g., J48) e métodos baseados em instâncias (e.g., iBk).

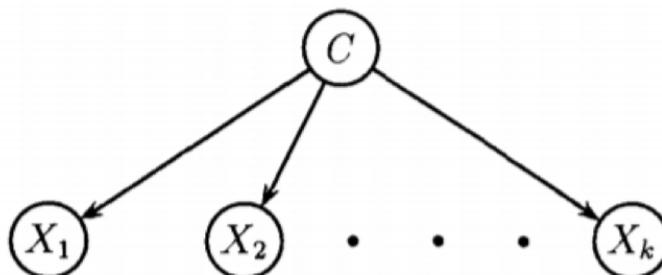
### 4.3.1 SMO

A *Sequential Minimal Optimization* - Otimização Mínima Sequencial (SMO) é um algoritmo para treino do SVM. O treino de um SVM requer a solução de um problema de otimização de *Quadratic Programming* - Programação quadrática (QP) muito grande. O SMO divide esse grande problema de QP numa série de problemas menores de QP que são resolvidos analiticamente, o que evita uma otimização numérica de QP demorada como um *loop* interno. Como a quantidade de memória

necessária para o SMO é linear para o conjunto de treino, o SMO pode lidar com conjuntos de treino muito grandes. A cada passo, o SMO escolhe dois multiplicadores *Lagrange* para otimizar o conjunto, encontra os valores ideais para esses multiplicadores e atualiza o SVM para retornar os novos valores ótimos [Platt, 1999].

### 4.3.2 Naive Bayes

O classificador Naive Bayes é simples e usa uma semântica clara para representar e aprender o conhecimento probabilístico. O classificador é visto como uma rede bayesiana, denominada “*naive*” porque baseia-se em duas suposições simplistas importantes. Em particular, assume que os atributos preditivos são condicionalmente independentes dada a classe, e considera que nenhum atributo oculto influencia o processo de previsão. Assim, um classificador Naive Bayes representa-se conforme está ilustrado na Figura 4.14, na qual todas as setas são direcionadas do atributo classe para os atributos preditivos observados [John and Langley, 1995]. Os valores numéricos de precisão do estimador são escolhidos com base na análise dos dados de treino [Source Forge, 2009b].

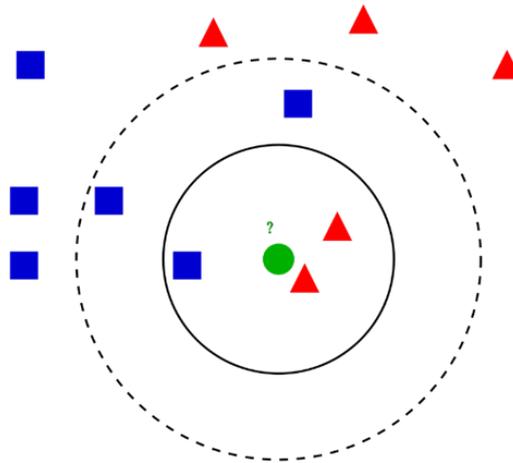


**Figura 4.14:** O classificador Naive Bayes descrito como uma rede bayesiana na qual os atributos preditivos ( $X_1, X_2, \dots, X_k$ ) são condicionalmente independentes dado o atributo de classe ( $C$ ) [John and Langley, 1995].

### 4.3.3 iBk

No reconhecimento de padrões, o algoritmo k-NN é um método não paramétrico usado para classificação e regressão. Em ambos os casos, a entrada consiste nos k exemplos de treino mais próximos no espaço das *features*. Na classificação k-NN, a saída resulta de uma associação da classe. Um objeto é classificado pela maioria dos seus vizinhos, sendo o objeto atribuído à classe o mais comum entre os k vizinhos mais próximos (k é um número inteiro positivo, geralmente pequeno), tal como é possível observar na Figura 4.15. Se  $k = 1$ , o objeto é simplesmente atribuído à classe

do único vizinho mais próximo [Altman, 1992]. A distância euclidiana é normalmente usada para variáveis contínuas, enquanto a distância de *Hamming* usa-se para variáveis discretas como a classificação de texto [Jaskowiak and Campello, 2011]. Uma desvantagem da classificação considerando a maioria dos vizinhos ocorre no caso da distribuição da classe ser distorcida. Ou seja, exemplos de uma classe mais frequente tendem a dominar a previsão do novo exemplo, porque tendem a ser frequentes entre os  $k$  vizinhos mais próximos. Uma técnica capaz de superar esse problema consiste em atribuir pesos às contribuições dos vizinhos, para que os vizinhos mais próximos contribuam mais para a média comparativamente aos mais distantes. Por exemplo, é frequente atribuir a cada vizinho um peso de  $1/d$ , onde  $d$  é a distância do vizinho. Os vizinhos são obtidos de um conjunto de objetos para os quais a classe é conhecida [Coomans and Massart, 1982].



**Figura 4.15:** Exemplo de classificação  $k$ -NN: a amostra teste (ponto verde) deve ser classificada em quadrados azuis ou em triângulos vermelhos. Se  $k = 3$  (círculo mais pequeno- linha a cheio), é classificado como triângulo vermelho porque existem 2 triângulos e apenas 1 quadrado dentro do círculo interno. Se  $k = 5$  (círculo a tracejado) é classificado como quadrado azul (3 quadrados vs 2 triângulos dentro do círculo externo) [Analytics Vidhya, 2018].

#### 4.3.4 J48

O J48 é uma implementação *open source* em Java no Weka do algoritmo C4.5 [Source Forge, 2009a]. C4.5 consiste num algoritmo que gera uma árvore de decisão desenvolvida por Ross Quinlan [Quinlan, 1993]. C4.5 é uma extensão do algoritmo ID3, anterior a Quinlan. As árvores de decisão geradas pelo C4.5 podem ser usadas para classificação e, por isso o C4.5 é designado de classificador estatístico [Witten et al., 2005]. Segundo [Quinlan, 1993] os algoritmos de árvore de decisão partem de um conjunto de casos ou exemplos e criam uma estrutura de dados em

árvore que pode ser usada para classificar os novos casos. Cada caso é descrito por um conjunto de atributos que podem ter valores numéricos ou simbólicos. Associado a cada caso de treino, há um rótulo que indica o nome da classe. Cada nó interno de uma árvore de decisão contém um teste, cujo resultado é usado para decidir qual a ramificação a seguir a partir desse nó. Por exemplo, um teste pode conter a pergunta “ $x > 4$  para o atributo  $x$ ?”, se o teste for verdadeiro segue-se pelo ramo da esquerda, caso contrário segue-se pelo ramo da direita. Os nós folha contêm rótulos de classe em vez de testes. Na classificação, quando um caso de teste (sem rótulo) atinge um nó folha, o C4.5 classifica-o usando o rótulo armazenado nele. Resumidamente, o C4.5 constrói árvores de decisão a partir de um conjunto de dados de treino da mesma maneira que o ID3, usando o conceito de entropia das informações. Os dados de treino são um conjunto  $S = s_1, s_2, \dots$  de amostras já classificadas. Cada amostra  $s_i$  consiste num vetor  $p$ -dimensional  $x_{1,i}, x_{2,i}, \dots, x_{p,i}$ , onde o  $x_j$  representa os valores do atributo da amostra, bem como a classe na qual  $s_i$  se enquadra. Em cada nó da árvore, o C4.5 escolhe o atributo dos dados que divide o conjunto de amostras em subconjuntos. O critério de divisão é o ganho de informação normalizado (diferença na entropia). O atributo com o maior ganho de informações normalizadas é escolhido para tomar a decisão. Para as diferentes aplicações do C4.5, as árvores produzidas são pequenas e precisas, resultando em classificadores rápidos e confiáveis.

### 4.4 Pré-processamento e extração de *features* no Matlab

Numa primeira fase e de acordo com os trabalhos desenvolvidos neste âmbito por [Mendes et al., 2015, Mendes et al., 2016, Rocha et al., 2018], definiu-se o *windowSize* e o *hopSize* (linhas 2-3), dividindo o sinal sonoro em frames de 50 ms e com uma sobreposição de 80% (*hop size* de 20%), resultando em 1996 frames (linhas 16-20). Usou-se um filtro passa-alto de resposta impulsional infinita de 8<sup>a</sup> ordem a 80 Hz (abaixo do limite inferior típico da voz humana adulta [Baken and Orlikoff, 2000]) (linhas 4 e 13). Posteriormente, procedeu-se à verificação, e caso seja necessário à reamostragem dos dados para uma taxa de amostragem de 4000 Hz (linhas 6-10) e normalizou-se o sinal sonoro através do valor máximo absoluto do áudio (linha 14). Na linha 24 extraem-se as 42 *features* explicadas na secção 4.2 para cada uma das 1996 frames.

Noutro *script*, geram-se eventos aleatórios de sons normais para que se possam classificar sons adventícios e normais, e não apenas os sons adventícios existentes

nos ficheiros de anotações. Criam-se mais ou menos eventos aleatórios se existirem poucas ou muitas anotações, respetivamente. Os eventos aleatórios gerados têm uma duração de 50 ms e 150 ms em proporções iguais. Estes valores da duração dos eventos são baseados na duração das sibilâncias e ferveores.

Por último, agregam-se as frames que correspondem aos eventos aleatórios gerados (sons normais) e aos eventos anotados (sibilâncias/ferveores) e calcula-se a média, mediana, mínimo, máximo e desvio padrão para cada uma das 42 *features*, perfazendo um total de 210 *features* para cada evento (linhas 26-36). Toda a informação acerca das 210 *features* extraídas para os eventos criados e anotados, as respetivas classes e os pacientes, é guardada num ficheiro .arff para posteriormente ser processado no Weka.

```

1 % initialize constants
2     windowSize = .05;
3     hopSize = .2;
4     hpf = designfilt('highpassiir', 'FilterOrder', 8, ...
5         'HalfPowerFrequency', 80, 'SampleRate', 4000);
6
7 % resample to 4000 Hz;
8     if fs≠4000
9         audio = resample(audio, 4000, fs);
10        fs = 4000;
11    end
12 % filter and normalize audio
13    audio = filtfilt(hpf, audio); % filter audio
14    audio = audio/max(abs(audio)); % normalize audio
15
16    mir = miraudio(audio, fs);
17    fr = mirframe(mir, windowSize, hopSize);
18    frpos = get(fr, 'FramePos');
19    frpos = frpos{1,1}{1,1};
20    frpos = mean(frpos);
21
22 % extract file events and features
23    fileEvents{file} = cell2mat(annotation(:,1:2));
24    features = extractFeatures(audio, fs, windowSize, hopSize);
25
26    % loop over events
27    for eve = 1:size(fileEvents{file},1)
28        [~, idStart] = min(abs(fileEvents{file}(eve,1) - ...
29            frpos)); % event first frame

```

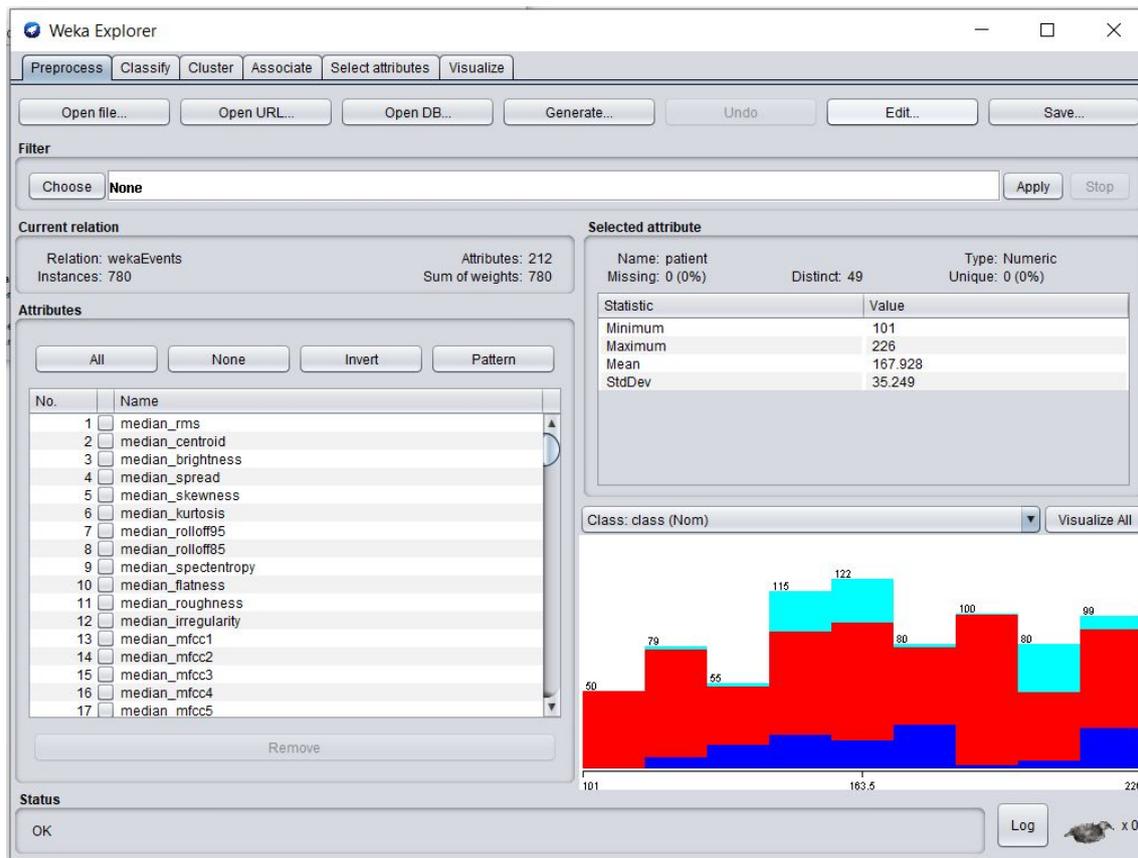
```
29         [idStart, idEnd] = min(abs(fileEvents{file}(eve,2) - ...  
30             frpos)); % event last frame  
31     if idEnd - idStart > ceil(1/windowSize)  
32         idEnd = idStart + ceil(1/windowSize);  
33     end  
34     [medianFeatures(eve,:), maxFeatures(eve,:), ...  
35         stdFeatures(eve,:), ...  
36         minFeatures(eve,:), meanFeatures(eve,:)] = ...  
37         ICBHI_sumfeatures(features(idStart:idEnd,:));  
38     end
```

## 4.5 Processamento no Weka

*Waikato Environment for Knowledge Analysis (Weka)* é um software *open source* desenvolvido na Universidade de Waikato, Nova Zelândia. O Weka é composto por um conjunto de algoritmos de aprendizagem computacional capazes de realizar processos de *data mining*, mais especificamente pré-processamento de dados, classificação, regressão, *clustering*, visualização e seleção de *features* [Hall et al., 2009].

Após selecionar a interface do *Explorer* do Weka, acede-se ao painel do “*Preprocess*” para importar o ficheiro *.arff* gerado pelo *script* do Matlab, descrito na secção 4.4 (Figura 4.16). O painel “*Classify*” permite aplicar algoritmos de classificação e regressão ao conjunto de dados, devolvendo estimativas de exatidão do modelo resultante, previsões de erros, curvas ROC, valor da *f-measure*, matriz de confusão, entre outros parâmetros, tal como se pode visualizar na Figura 4.17.

Usou-se a predefinição do Weka para a validação cruzada, ou seja, 10  *folds*. Assim, o conjunto de dados é ordenado aleatoriamente e em seguida dividido em 10 partições ( *folds*) de tamanho igual. Em cada iteração, 1  *fold* é usada para teste e as outras 9  *folds* são usadas para treino do classificador. Os resultados fornecidos consistem na média das 10  *folds*.



**Figura 4.16:** Painel “Preprocess” do software Weka, após o ficheiro .arff ter sido importado.

Experimentaram-se 4 classificadores (SMO, iBk, J48 e Naive Bayes), já explicados na secção 4.3. Para cada amostra, os resultados de cada classificador são apresentados com base nos seguintes parâmetros: *recall*, *precision*, *f-measure* e matriz de confusão, sendo que para as 3 primeiras apresenta-se a média para as 3 classes e os resultados por classe.

**Classifier**

Choose: SMO - C 1.0 - L 0.001 - P 1.0E-12 - N 0 - V 1 - W 1 - K "weka.classifiers.functions.supportVector.PolyKernel-E 1.0 - C 250007 - calibrator "weka.classifiers.functions.Logistic-R 1.0E-8 - M 1 - num-decimal-places 4"

**Test options**

Use training set  
 Supplied test set  
 Cross-validation Folds: 10  
 Percentage split: % 50  
 More options...

(Nom) class

Start Stop

Result list (right-click for options)

14.91.49 - funcoes.SMO

**Classifier output**

```

+ 0.0032 * (normalized) mean_inharmonicity
- 0.0113 * (normalized) mean_F0
+ 0.0033 * (normalized) mean_egg
- 0.5294 * (normalized) mean_hnr500
+ 0.0447 * (normalized) mean_hnr1500
+ 0.0085 * (normalized) mean_wsss
+ 0.0027 * (normalized) mean_frachim
+ -0.0036 * (normalized) mean_beager
- 0.0041 * (normalized) mean_entropy
- 0.8577

Number of kernel evaluations: 73378 (85.38% cached)

Time taken to build model: 0.68 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      669      85.7652 %
Incorrectly Classified Instances    111      14.2308 %
Kappa statistic                    0.659
Mean absolute error                 0.2591
Root mean squared error             0.3313
Relative absolute error             85.8141 %
Root relative squared error         85.5234 %
Total Number of Instances          780

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.602   0.035   0.763    0.602   0.673    0.426   0.872    0.578   crackle
          0.848   0.339   0.674   0.948   0.809   0.657   0.803   0.866   normal
          0.673   0.018   0.850   0.673   0.751   0.725   0.895   0.646   wherez
Weighted Avg.   0.858   0.250   0.853   0.858   0.852   0.661   0.826   0.752

=== Confusion Matrix ===
  a  b  c  <-- Classified as
 74 45 4  | a = crackle
 21 527 8  | b = normal
 2  31 68  | c = wherez
  
```

Figura 4.17: Painel “Classify” do Weka onde se encontra o output para o classificador SMO.

### Matriz de confusão

No caso ilustrado na Figura 4.17, que corresponde a um output deste estudo no painel “*Classify*”, tem-se 3 classes (a=*crackle*; b=*normal*; c= *wheeze*), ou seja, uma matriz de confusão 3x3. O número de instâncias classificadas corretamente é a soma das diagonais da matriz. Todos os outros elementos da matriz foram classificados incorretamente (a classe “*crackle*” é classificada erradamente como “*normal*” em 3 instâncias e nenhuma instância foi classificada erradamente como “*wheeze*”; a classe “*normal*” é classificada erradamente como “*crackle*” em 8 instâncias e erradamente classificada como “*wheeze*” em 2 instâncias; a classe “*wheeze*” foi erradamente classificada como “*crackle*” em apenas 1 instância e erradamente classificada em 2 instâncias como “*normal*”) [Hall et al., 2009].

### *Precision*

Corresponde à proporção dos casos que realmente pertencem à classe  $x$  entre todos aqueles que foram classificados como classe  $x$ . É também conhecido como valor preditivo positivo. Na matriz, corresponde ao elemento da diagonal dividido pela soma da coluna em questão (as colunas na Figura 4.17: *crackle*, *normal*, *wheeze*), ou seja, obtém-se pela equação 4.14 onde TP representa *True Positive* - verdadeiros positivos e FP *False Positive* - falsos positivos [Hall et al., 2009].

$$Precision = \frac{TP}{TP + FP} \quad (4.14)$$

### *Recall*

Corresponde à taxa de TP que por sua vez consiste na proporção de casos que foram classificados corretamente como classe  $x$ , entre todos os casos que realmente têm classe  $x$  (equação 4.15). É equivalente à sensibilidade [Hall et al., 2009].

$$Recall = \frac{TP}{TP + FN} \quad (4.15)$$

### *F-measure*

É uma medida que combina a *precision* e a *recall* (equação 4.16) [Hall et al., 2009].

$$F-measure = 2 \times Precision \times \frac{Recall}{Precision + Recall} \quad (4.16)$$

## 4.6 Testes de significância estatística

Para analisar cada instância pormenorizadamente, isto é, a classe real e a classe prevista pelo classificador foi necessário atribuir um ID a cada instância. Isso foi feito no painel “*Preprocess*” do Weka, escolhendo um filtro de atributos não supervisionado, nomeadamente o “*AddId*”. Seguidamente, no separador “*Classify*” escolheu-se o filtro “*meta-FilteredClassifier*” e selecionou-se nos filtros de atributos não supervisionados o *Remove* e definiram-se o atributo de índice 1 que corresponde ao ID, e o índice do atributo “*patient*” que corresponde ao número do paciente. Depois, na janela “*More options*” no “*Test options*” selecionou-se “*Output predictions*” e inseriu-se 1 e o índice do “*patient*” no “*Output additional attributes*” para que o atributo ID e “*patient*” fossem visíveis no output da classificação, permitindo identificar o ID e o paciente para cada instância.

Posteriormente, foram realizados os testes de significância estatística para avaliar quais os resultados de classificação são estatisticamente significativos. Isto porque, embora se pudesse alcançar um valor da *f-measure* superior num determinado algoritmo não implica que o mesmo seja um classificador preferível, pois poderá ter previsto a classe ao acaso. Os testes de significância estatística foram efetuados através de um *script* do Matlab, que tem como *input* da função 4 parâmetros: i)  $x_1$  e ii)  $x_2$  que são dois vetores de tamanhos iguais que contêm a previsão do classificador 1 e 2, respetivamente; iii) *alpha* que representa o limite de significância estatística tipicamente 0,05 (5%) ou 0,01 (1%); iv) *type* que indica o tipo de teste (emparelhado ou não). Como *output* temos 3 parâmetros: *h* que pode tomar o valor de 1 ou 0 caso as diferenças sejam estaticamente significativas ou não, respetivamente; *p-value*, que deve ser menor que *alpha* para o teste ser estatisticamente significativo; *g* que pode tomar o valor de 1 se ambas as variáveis seguirem uma distribuição gaussiana ou 0 se pelo menos uma das distribuições não é gaussiana (o parâmetro *g* não é relevante para este estudo). Usou-se um valor de *alpha*=0,01 e *alpha*=0,05, e *type*=*paired*=1 (emparelhado).

## 4.7 Seleção de *features*

Após a realização dos testes de significância estatística entre os 4 classificadores que permitiram averiguar qual o classificador mais adequado, prosseguiu-se para a seleção de atributos. O painel “*Select attributes*” do Weka fornece várias técnicas para identificar quais os atributos mais relevantes para a classificação do conjunto de dados (Figura 4.18).

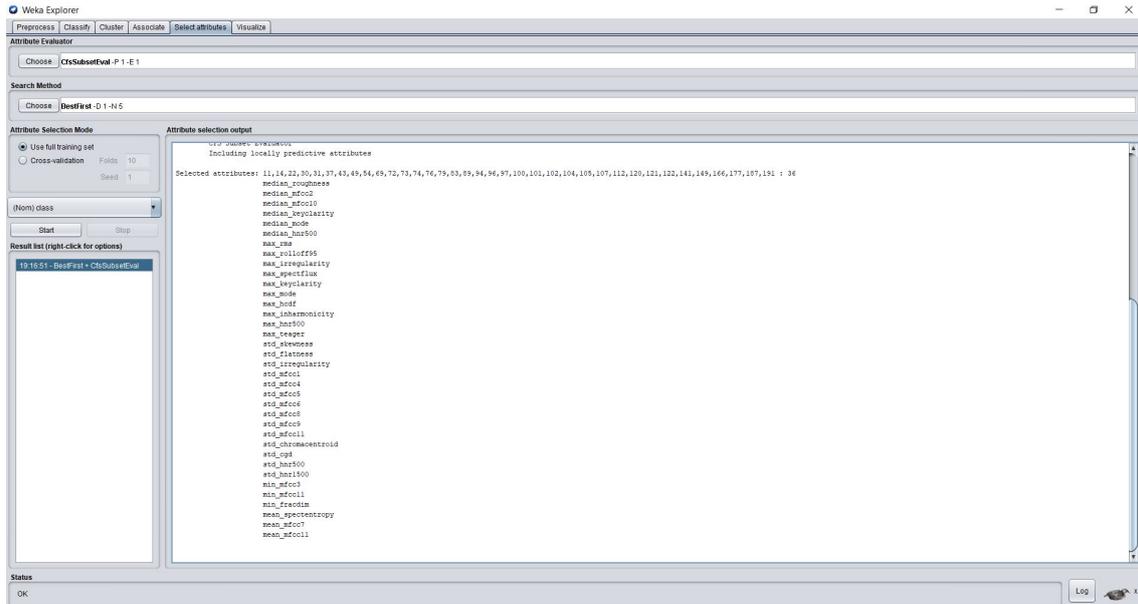


Figura 4.18: Painel “*Select attributes*” do Weka onde é possível observar o resultado para o método *CfsSubsetEval* + *Best First*.

Neste trabalho escolheram-se 3 técnicas de seleção de atributos: *CfsSubsetEval* + *Best First*, *ReliefFAttributeEval* + *Ranker* e *WrapperSubsetEval* + *Best First* (com o algoritmo SMO) que corresponde a “*Attribute Evaluator*” + “*Search Method*”.

- ***CfsSubsetEval***: O algoritmo *CfsSubsetEval* avalia o valor de um subconjunto de atributos considerando a capacidade individual preditiva de cada atributo, juntamente com o grau de redundância entre eles. Os subconjuntos de atributos que são altamente correlacionados com a classe enquanto apresentam baixa correlação são os preferidos [Hall, 1998, Hall et al., 2009].
- ***ReliefFAttributeEval***: Avalia o valor de um atributo amostrando repetidamente uma instância e considerando o valor do atributo fornecido para a instância mais próxima da mesma e de outra classe [Kira and Rendell, 1992, Hall et al., 2009].
- ***WrapperSubsetEval***: Avalia conjuntos de atributos recorrendo a um esquema de aprendizagem, sendo que é usada a validação cruzada para estimar a exatidão desse esquema [Kohavi and John, 1997, Hall et al., 2009].
- ***Best First***: Seleciona o nó mais promissor que se gerou até ao momento e que ainda não foi expandido. Um nó promissor é aquele que tem uma estimativa de exatidão pelo menos maior que a melhor encontrada até ao momento. A configuração do número de nós consecutivos sem melhoria permite controlar

o nível de retorno feito. O *Best First* pode começar com o conjunto vazio de atributos e pesquisar para a frente, ou começar com o conjunto completo de atributos e pesquisar para trás, ou iniciar em qualquer ponto e pesquisar nas duas direções (considerando todas as adições possíveis e exclusões de atributos únicos num determinado ponto) [Hall et al., 2009].

- **Ranker:** Classifica os atributos de acordo com as suas avaliações individuais [Hall et al., 2009].

Após se conhecerem os atributos selecionados procede-se à classificação com esses mesmos atributos. Para tal, regressa-se ao painel “*Preprocess*” do Weka e escolhe-se o filtro de atributos não supervisionado “*Reorder*” que permite ordenar os atributos de acordo com a seriação do método de seleção de atributos selecionado. De notar que o atributo da classe deve ser o último. Depois, procede-se à classificação novamente e avalia-se se os resultados melhoraram ou não. No caso do *Relief* como ordena as 210 *features* pela ordem de importância usa-se o filtro “*remove*” para seleccionar apenas os  $n$  primeiros atributos (mais o atributo da classe) e simulam-se os resultados.

Depois de se obter o número de *features* ótimo, realizaram-se novamente os testes de significância estatística entre o melhor modelo encontrado com o número de *features* ótimo e o melhor modelo obtido primeiramente com 210 *features*. Verifica-se ainda, qual o número de *features* com que se atinge 95% dos resultados da melhor classificação obtida.

## 4.8 Otimização de parâmetros

Uma maneira tradicional de executar a otimização de parâmetros é através do *grid search*. O *grid search* é um processo que ajusta parâmetros a fim de determinar os valores ideais para um determinado modelo [Chicco, 2017]. Isto é importante, pois o desempenho do modelo é baseado nos valores dos parâmetros especificados. O *grid search* realiza uma pesquisa exaustiva através de um subconjunto especificado manualmente de parâmetros de um algoritmo de aprendizagem. Por exemplo, um classificador SVM possui pelo menos dois parâmetros que podem ser ajustados para obter um desempenho melhor: parâmetro de complexidade  $C$  e o expoente do kernel polinomial [Hsu et al., 2003]. Ambos os parâmetros são contínuos, por isso seleccionou-se um conjunto de valores “razoáveis” para cada um, ou seja, variou-se  $C$  de 0,1 a 2 e o expoente do PolyKernel de 0,1 a 3, ambos em intervalos de 0,1. O *grid search* treina um SVM para cada par  $(C, e)$  no produto cartesiano desses dois

conjuntos e avalia o seu desempenho. Por fim, identificam-se os pares que atingiram os valores mais altos para a classificação.



# Resultados

Neste capítulo são apresentados os resultados da classificação dos SRs, os testes de significância estatística, seleção de *features* e otimização de parâmetros para as diferentes amostras selecionadas.

## 5.1 População infantil

Nesta secção apresentam-se os resultados dos diferentes testes realizados para a população infantil.

### 5.1.1 Classificação com 210 *features*

Os resultados da classificação com as 210 *features* para as 5 amostras selecionadas, nomeadamente 0-2 anos, 3-6 anos, todos os sons, apenas ferveores e apenas sibilâncias, para os 4 classificadores (Naive Bayes, iBk, j48 e SMO) encontram-se nas Tabelas 5.1, 5.3, 5.5, 5.7 e 5.9, respetivamente. As tabelas indicam para além dos valores da *f-measure*, a *recall*, *precision* e matriz de confusão. Nos testes de significância estatística comparam-se os resultados obtidos na classificação para os 4 classificadores com o intuito de verificar se as diferenças entre eles são estatisticamente significativas ou não. Estes resultados podem ser observados nas Tabelas 5.2, 5.4, 5.6, 5.8 e 5.10, correspondentes às amostras referidas anteriormente.

#### Amostra 0-2 anos

Visualizando a Tabela 5.1, o classificador SMO foi o que atingiu maiores valores para a *f-measure* nas 3 classes. Tal como se pode observar pela matriz de confusão classificou corretamente 38 dos 55 eventos de ferveores, 33 dos 50 eventos normais e 57 dos 65 eventos de sibilâncias. Ou seja, entre as 3 classes, as sibilâncias foram as que apresentaram o maior número de instâncias corretamente classificadas. A *recall* e a *precision* refletem obviamente os valores da *f-measure*. Os classificadores Naive Bayes e iBk, e logo a seguir o j48, apresentaram os piores resultados para os

fervores. Em relação à classe dos eventos normais e sibilâncias, o j48 obteve o pior resultado comparativamente ao Naive Bayes e iBk.

**Tabela 5.1:** Classificação com 210 *features* para a amostra 0-2 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	fervor	0,571	0,073	0,129	a b c <- classified as
	normal	0,506	0,860	0,637	37 8 10   a = fervor
	sibilância	0,603	0,723	0,657	9 34 7   b = normal
	weighted avg.	0,564	0,553	0,480	9 6 50   c = sibilância
j48	fervor	0,446	0,455	0,450	a b c <- classified as
	normal	0,510	0,500	0,505	25 14 16   a = fervor
	sibilância	0,631	0,631	0,631	17 25 8   b = normal
	weighted avg.	0,536	0,535	0,535	14 10 41   c = sibilância
Naive Bayes	fervor	0,571	0,073	0,129	a b c <- classified as
	normal	0,506	0,860	0,637	4 26 25   a = fervor
	sibilância	0,603	0,723	0,657	1 43 6   b = normal
	weighted avg.	0,564	0,553	0,480	2 16 47   c = sibilância
SMO	fervor	0,717	0,691	0,704	a b c <- classified as
	normal	0,717	0,660	0,688	38 11 6   a = fervor
	sibilância	0,803	0,877	0,838	9 33 8   b = normal
	weighted avg.	0,750	0,753	0,750	6 2 57   c = sibilância

Relativamente aos testes de significância estatística com  $alpha=0,01$  (Tabela 5.2) verificaram-se diferenças estatisticamente significativas entre o classificador Naive Bayes e os restantes (j48, iBk, SMO) com um  $p-value$  bastante inferior a 0,01, o que confere confiabilidade ao resultado. Deste modo, as diferenças nos resultados da classificação do Naive Bayes para os restantes classificadores são relevantes, logo este classificador não é o mais adequado aos dados. Por outro lado, os restantes classificadores (iBk, j48 e SMO) ainda que apresentem diferenças entre eles na classificação, estas não são estatisticamente significativas e estão longe de o serem (o valor de  $p-value$  é elevado). Ainda assim, considera-se o SMO o classificador mais adequado à amostra 0-2 anos, uma vez que foi o que apresentou melhores resultados.

Note que, a fim de avaliar os resultados para uma correlação diferente (95%), procedeu-se à realização dos testes de significância estatística para um  $alpha=0,05$ . Os resultados foram exatamente os mesmos, e podem ser consultados na Tabela B.1 na secção B dos apêndices.

**Tabela 5.2:** Testes de significância estatística com  $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.1, para a amostra 0-2 anos.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=0,9691	h=1; p=7,0682 $\times 10^{-7}$	h=0; p= 0,6401
j48	–	–	h=1; p=3,7741 $\times 10^{-7}$	h=0; p= 0,5667
Naive Bayes	–	–	–	h=1; p=6,6343 $\times 10^{-6}$
SMO	–	–	–	–

### Amostra 3-6 anos

Visualizando a Tabela 5.3, o classificador SMO foi o que atingiu maiores valores para a *f-measure* nas 3 classes. Tal como se pode observar pela matriz de confusão classificou corretamente 39 dos 42 eventos de fervores, 26 dos 36 eventos normais e 30 dos 33 eventos de sibilâncias. A *recall* e a *precision* refletem obviamente os valores da *f-measure*. O classificador j48 foi o que apresentou os piores resultados para as 3 classes. Já os classificadores iBk e Naive Bayes atingiram valores da *f-measure* bastante semelhantes, mas não superiores aos conseguidos pelo SMO. Para os 4 classificadores, as sibilâncias foram as que apresentaram o maior número de instâncias corretamente classificadas, seguindo-se os fervores e por fim os eventos normais.

Nos testes de significância estatística para  $\alpha=0,01$  e  $\alpha=0,05$  (Tabela 5.4 e Tabela B.2 da secção B dos apêndices) ambos verificaram diferenças estatisticamente significativas entre o classificador j48 e SMO com um *p-value* inferior a 0,01, o que confere confiabilidade ao resultado. Este teste não surpreende, visto que o SMO obteve de facto melhores resultados que o j48. Contrariamente, entre os classificadores j48 e iBk as diferenças não são estatisticamente significativas e estão longe de o serem (*p-value*=0,3671), tal como entre o classificador Naive Bayes e SMO com um *p-value*=1 estão muito longe de serem estatisticamente significativas (tanto para  $\alpha=0,01$  como para  $\alpha=0,05$ ). Por outro lado, para  $\alpha=0,01$  entre os conjuntos de classificadores j48 e Naive Bayes, Naive Bayes e iBk, e iBk e SMO apesar de os testes indicarem que as diferenças na classificação não são estatisticamente sig-

## 5. Resultados

nificativas, como os valores de  $p$ -value são baixos ( $p$ -value=0,0129; 0,0323; 0,0148, respetivamente) afere-se que estão próximas de serem estatisticamente significativas. Aliás, para  $\alpha=0,05$  as diferenças já se revelaram estatisticamente significativas. Assim sendo, considera-se o SMO o classificador mais adequado à amostra 3-6 anos, uma vez que foi o que apresentou melhores resultados.

**Tabela 5.3:** Classificação com 210 *features* para a amostra 3-6 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	fervor	0,805	0,786	0,795	a b c <- classified as
	normal	0,733	0,611	0,667	33 8 1   a = fervor
	sibilância	0,825	1,000	0,904	8 22 6   b = normal
	weighted avg.	0,788	0,793	0,786	0 0 33   c = sibilância
j48	fervor	0,765	0,619	0,684	a b c <- classified as
	normal	0,500	0,528	0,514	26 11 5   a = fervor
	sibilância	0,641	0,758	0,694	8 19 9   b = normal
	weighted avg.	0,642	0,631	0,632	0 8 25   c = sibilância
Naive Bayes	fervor	0,735	0,857	0,791	a b c <- classified as
	normal	0,759	0,611	0,677	36 6 0   a = fervor
	sibilância	0,909	0,909	0,909	11 22 3   b = normal
	weighted avg.	0,794	0,793	0,789	2 1 30   c = sibilância
SMO	fervor	0,813	0,923	0,867	a b c <- classified as
	normal	0,839	0,722	0,776	39 3 0   a = fervor
	sibilância	0,938	0,909	0,923	8 26 2   b = normal
	weighted avg.	0,858	0,856	0,854	1 2 30   c = sibilância

**Tabela 5.4:** Testes de significância estatística com  $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.3, para a amostra 3-6 anos.

	iBk	j48	Naive Bayes	SMO
iBk	-	h=0; p=0,3671	h=0; p= 0,0323	h=0; p= 0,0148
j48	-	-	h=0; p=0,0129	h=1; p= 0,0063
Naive Bayes	-	-	-	h=0; p=1
SMO	-	-	-	-

### Amostra todos os sons de crianças

Visualizando a Tabela 5.5, o classificador SMO foi o que atingiu maiores valores para a *f-measure* nas 3 classes. Tal como se pode observar pela matriz de confusão classificou corretamente 101 dos 123 eventos de fervores, 106 dos 135 eventos normais e 83 dos 101 eventos de sibilâncias. A *recall* e a *precision* refletem obviamente os valores da *f-measure*. O classificador Naive Bayes foi o que apresentou os piores resultados para as 3 classes. Já os classificadores iBk e j48 atingiram valores da *f-measure* semelhantes, embora o j48 tenha alcançado valores ligeiramente superiores ao iBk, mas não superiores aos conseguidos pelo SMO. Para os 3 classificadores (iBk, j48 e SMO), as sibilâncias foram as que apresentaram o maior número de instâncias corretamente classificadas, seguindo-se os fervores e por fim os eventos normais.

**Tabela 5.5:** Classificação com 210 *features* para a amostra todos os sons de crianças.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	fervor	0,706	0,683	0,694	a b c ← classified as
	normal	0,662	0,696	0,679	84 27 12   a = fervor
	sibilância	0,714	0,693	0,704	25 94 16   b = normal
	weighted avg.	0,692	0,691	0,691	10 21 70   c = sibilância
j48	fervor	0,759	0,691	0,723	a b c ← classified as
	normal	0,704	0,704	0,704	85 26 12   a = fervor
	sibilância	0,696	0,772	0,732	18 95 22   b = normal
	weighted avg.	0,721	0,719	0,719	9 14 78   c = sibilância
Naive Bayes	fervor	1,000	0,024	0,048	a b c ← classified as
	normal	0,591	0,674	0,630	3 51 69   a = fervor
	sibilância	0,441	0,881	0,587	0 91 44   b = normal
	weighted avg.	0,689	0,510	0,418	0 12 89   c = sibilância
SMO	fervor	0,821	0,821	0,821	a b c ← classified as
	normal	0,757	0,785	0,771	101 18 4   a = fervor
	sibilância	0,865	0,822	0,843	20 106 9   b = normal
	weighted avg.	0,809	0,808	0,808	2 16 83   c = sibilância

Relativamente aos testes de significância estatística com  $\alpha=0,01$  (Tabela 5.6) verificaram-se diferenças estatisticamente significativas entre o classificador Naive Bayes e os restantes (j48, iBk, SMO) com um *p-value* bastante inferior a 0,01, o que confere confiabilidade ao resultado. Deste modo, as diferenças nos resultados da classificação do Naive Bayes para os restantes classificadores são relevantes, logo este classificador não é o mais adequado aos dados. Contrariamente, entre os clas-

sificadores SMO e iBk as diferenças não são estatisticamente significativas e estão longe de o serem, pois o valor de  $p$ -value é elevado ( $p$ -value=0,5907), tal como entre o classificador iBk e j48, embora neste caso o valor de  $p$ -value seja menor ( $p$ -value=0,1669), logo as diferenças estarão mais próximas de serem estatisticamente significativas. Estes resultados são surpreendentes, uma vez que existem diferenças entre os valores da  $f$ -measure encontrados para os classificadores iBk, j48 e SMO. Quanto aos classificadores SMO e j48, embora o teste indique que não existem diferenças estatisticamente significativas o baixo valor do  $p$ -value permite concluir que as diferenças estão próximas de serem estatisticamente significativas. Assim sendo, considera-se o SMO o classificador mais adequado à amostra que contém todos os sons de crianças, uma vez que foi o que apresentou melhores resultados.

Note que, a fim de avaliar os resultados para uma correlação diferente (95%), procedeu-se à realização dos testes de significância estatística para um  $alpha=0,05$ . Os resultados foram exatamente os mesmos, e podem ser consultados na Tabela B.3 na secção B dos apêndices.

**Tabela 5.6:** Testes de significância estatística com  $alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.5, para a amostra todos os sons de crianças.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=0,1669	h=1; p=3,8271 $\times 10^{-27}$	h=0; p=0,5907
j48	–	–	h=1; p=2,8552 $\times 10^{-23}$	h=0; p=0,0640
Naive Bayes	–	–	–	h=1; p=4,4281 $\times 10^{-27}$
SMO	–	–	–	–

### Amostra apenas ferveores

Visualizando a Tabela 5.7, os classificadores SMO e j48 foram os que atingiram os maiores valores para a  $f$ -measure nas 2 classes. Tal como se pode observar pela matriz de confusão o SMO (j48) classificou corretamente 53 (54) dos 62 eventos de ferveores e 48 (48) dos 57 eventos normais. A  $recall$  e a  $precision$  refletem obviamente os valores da  $f$ -measure. O classificador Naive Bayes foi o que apresentou os piores resultados para as 2 classes. Já o classificador iBk ainda que tenha atingido melhores resultados que o Naive Bayes, estes não foram superiores aos dos j48 e SMO. Para

os 3 classificadores (iBk, j48 e SMO), os fervores apresentaram maior número de instâncias corretamente classificadas em relação aos eventos normais.

**Tabela 5.7:** Classificação com 210 *features* para a amostra que contém apenas fervores.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	fervor	0,746	0,758	0,752	a b <- classified as
	normal	0,732	0,719	0,726	47 15   a = fervor
	weighted avg.	0,739	0,739	0,739	16 41   b = normal
j48	fervor	0,857	0,871	0,864	a b <- classified as
	normal	0,857	0,842	0,850	54 8   a = fervor
	weighted avg.	0,857	0,857	0,857	9 48   b = normal
Naive Bayes	fervor	0,750	0,097	0,171	a b <- classified as
	normal	0,495	0,965	0,655	6 56   a = fervor
	weighted avg.	0,628	0,513	0,403	2 55   b = normal
SMO	fervor	0,855	0,855	0,855	a b <- classified as
	normal	0,842	0,842	0,842	53 9   a = fervor
	weighted avg.	0,849	0,849	0,849	9 48   b = normal

Relativamente aos testes de significância estatística com  $\alpha=0,01$  (Tabela 5.8), verificaram-se diferenças estatisticamente significativas entre o classificador Naive Bayes e os restantes (j48, iBk, SMO) com um *p-value* bastante inferior a 0,01, o que confere confiabilidade ao resultado. Deste modo, as diferenças nos resultados da classificação do Naive Bayes para os restantes classificadores são relevantes, logo este classificador não é o mais adequado aos dados. Por outro lado, os restantes classificadores (iBk, j48 e SMO) ainda que apresentem diferenças entre eles na classificação, estas não são estatisticamente significativas e estão longe de o serem (o valor de *p-value* é elevado).

Note que, a fim de avaliar os resultados para uma correlação diferente (95%), procedeu-se à realização dos testes de significância estatística para um  $\alpha=0,05$ . Os resultados foram exatamente os mesmos, e podem ser consultados na Tabela B.4 na secção B dos apêndices.

Assim sendo, considera-se o SMO o classificador mais adequado à amostra dos fervores. Atente que, embora o j48 tenha resultados bastante semelhantes ao SMO, ou até valores da *f-measure* ligeiramente superiores, o SMO apresenta os melhores resultados em quase todos os cenários, assim por uma questão de simplicidade optou-se por usar sempre o SMO. Além disso, os resultados na literatura suportam

## 5. Resultados

a utilização do SMO (SVM) (Tabela A dos apêndices). E tal como se observará na secção 5.2 referente à população adulta, o SMO teve o melhor desempenho (estatisticamente significativo).

**Tabela 5.8:** Testes de significância estatística com  $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.7, para a amostra que contém apenas ferveores.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=1	h=1; p=8,0448 $\times 10^{-13}$	h=0; p=0,8415
j48	–	–	h=1; p=1,2053 $\times 10^{-13}$	h=0; p=0,8348
Naive Bayes	–	–	–	h=1; p=1,3358 $\times 10^{-12}$
SMO	–	–	–	–

### Amostra apenas sibilâncias

Visualizando a Tabela 5.9, os classificadores j48, iBk e SMO foram os que atingiram os maiores valores para a *f-measure* nas 2 classes. Os classificadores iBk e j48 alcançaram exatamente os mesmos resultados (matriz de confusão, *f-measure*, *precision* e *recall*). Tal como se pode observar pela matriz de confusão o SMO (j48 e iBk) classificou corretamente 47 (48) dos 55 eventos de sibilâncias e 38 (40) dos 49 eventos normais. A *recall* e a *precision* refletem obviamente os valores da *f-measure*. O classificador Naive Bayes foi o que apresentou os valores mais baixos da *f-measure* para as 2 classes. Para os 3 classificadores (iBk, j48 e SMO), as sibilâncias apresentaram maior número de instâncias corretamente classificadas em relação aos eventos normais.

Relativamente aos testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$  (Tabela 5.10 e Tabela B.5 na secção B dos apêndices, respetivamente) ambos revelaram para os conjuntos iBk e j48, iBk e SMO, e, j48 e SMO, valores do *p-value* bastante elevados (*p-value*=1; 0,8084; 0,8084, respetivamente) o que implica que as diferenças estão longes de serem estatisticamente significativas, uma vez que os resultados da classificação desses 3 algoritmos foram bastante semelhantes entre si. Por outro lado, embora o teste com  $\alpha=0,01$  entre o Naive Bayes e os 3 restantes classificadores indique que as diferenças não são estatisticamente significativas, os valores do *p-value* são bastante mais baixos, permitindo aferir que estão próximos

de serem estatisticamente significativos, aliás para um  $alpha=0,05$  as diferenças já se revelaram estatisticamente significativas, o que não surpreende dado que o classificador Naive Bayes apresentou os valores mais baixos da  $f$ -measure. Assim sendo, considera-se o SMO o classificador mais adequado à amostra dos sibilâncias. Note que, embora o j48 tenha resultados bastante semelhantes ao SMO, ou até valores da  $f$ -measure ligeiramente superiores, o SMO apresenta os melhores resultados em quase todos os cenários, assim por uma questão de simplicidade optou-se por usar sempre o SMO. Além disso, os resultados na literatura suportam a utilização do SMO (SVM) (Tabela apêndices A). E tal como se observará na secção 5.2 referente à população adulta, o SMO teve o melhor desempenho (estatisticamente significativo).

**Tabela 5.9:** Classificação com 210 *features* para a amostra que contém apenas sibilâncias.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	normal	0,851	0,816	0,833	a b <- classified as
	sibilância	0,842	0,873	0,857	40 9   a = normal
	weighted avg.	0,846	0,846	0,846	7 48   b = sibilância
j48	normal	0,851	0,816	0,833	a b <- classified as
	sibilância	0,842	0,873	0,857	40 9   a = normal
	weighted avg.	0,846	0,846	0,846	7 48   b = sibilância
Naive Bayes	normal	0,719	0,837	0,774	a b <- classified as
	sibilância	0,830	0,709	0,765	41 8   a = normal
	weighted avg.	0,778	0,769	0,769	16 39   b = sibilância
SMO	normal	0,826	0,776	0,800	a b <- classified as
	sibilância	0,810	0,855	0,832	38 11   a = normal
	weighted avg.	0,818	0,817	0,817	8 47   b = sibilância

**Tabela 5.10:** Testes de significância estatística com  $alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.9, para a amostra que contém apenas sibilâncias.

	iBk	j48	Naive Bayes	SMO
iBk	-	h=0; p=1	h=0; p= 0,0499	h=0; p= 0,8084
j48	-	-	h=0; p=0,0330	h=0; p=0,8084
Naive Bayes	-	-	-	h=0; p=0,0116
SMO	-	-	-	-

### 5.1.2 Seleção de *features*

Nesta fase é importante encontrar o número ótimo de *features* com que se atinge o melhor resultado de classificação, mas também o número *features* com que se atinge 95% dessa classificação.

#### Amostra 0-2 anos

Na amostra 0-2 anos, o algoritmo *CfsSubsetEval + Best First* selecionou as 36 *features* indicadas na Tabela C.1, que se encontra na secção C.1 dos apêndices, como as mais relevantes.

Contudo, a classificação com estes 36 atributos não melhorou comparativamente à classificação com 210 atributos e com o algoritmo SMO (Tabela 5.1), observando-se decréscimos dos valores da *f-measure* para as 3 classes (Tabela 5.11).

O algoritmo *WrapperSubsetEval + Best First* com SMO selecionou 13 *features* indicadas na Tabela C.2, que se encontra na secção C.1 dos apêndices, como as mais relevantes.

Mais uma vez, não se verificaram melhorias relativamente à classificação com 210 atributos e com o algoritmo SMO (Tabela 5.1), embora os valores da *f-measure* sejam superiores aos obtidos pelo método *CfsSubsetEval + Best First* para todas as classes (Tabela 5.11).

Relativamente ao algoritmo *ReliefFAttributeEval + Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado com o algoritmo, cujos valores se encontram na Tabela 5.11, obteve-se com os primeiros 50 atributos descritos na Tabela C.3, que se encontra na secção C.1 dos apêndices.

É possível observar que os valores da *f-measure* apresentam aumentos pouco significativos nas 3 classes comparativamente à classificação com 210 atributos e com o algoritmo SMO.

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features* *ReliefFAttributeEval + Ranker* (50 atributos).

Tabela 5.11: Diferentes métodos de seleção de *features* para a amostra 0-2 anos.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	fervor	0,781	0,455	0,575	a b c <- classified as
		normal	0,596	0,620	0,608	25 13 17   a = fervor
		sibilância	0,605	0,800	0,689	2 31 17   b = normal
		weighted avg.	0,659	0,635	0,628	5 8 52   c = sibilância
WrapperSubsetEval + Best First; com SMO	SMO	fervor	0,729	0,636	0,680	a b c <- classified as
		normal	0,721	0,620	0,667	35 6 14   a = fervor
		sibilância	0,671	0,815	0,736	7 31 12   b = normal
		weighted avg.	0,704	0,700	0,697	6 6 53   c = sibilância
ReliefAttributeEval + Ranker	SMO	fervor	0,714	0,727	0,721	a b c <- classified as
		normal	0,821	0,640	0,719	40 6 9   a = fervor
		sibilância	0,773	0,892	0,829	10 32 8   b = normal
		weighted avg.	0,768	0,765	0,761	6 1 58   c = sibilância

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.1) e o melhor modelo obtido com 50 atributos (Tabela 5.11), revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value bastante elevado de 0,6547. Estes resultados não surpreendem, visto que não existem diferenças muito significativas nas medidas da  $f$ -measure para os dois modelos (1,1% de diferença entre os dois resultados).

Com os primeiros 42 atributos da Tabela C.3, que se encontra na secção C.1 dos apêndices, atinge-se 95% da classificação máxima (50 atributos presentes na Tabela C.3 e com o SMO: Tabela 5.11) obtida para a amostra 0-2 anos. Os resultados da classificação com os 42 atributos estão expressos na Tabela 5.12.

**Tabela 5.12:** 95% da classificação máxima com 42 atributos para a amostra 0-2 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,645	0,727	0,684	a b c ← classified as
	normal	0,763	0,580	0,659	40 8 7   a = fervor
	sibilância	0,786	0,846	0,815	13 29 8   b = normal
	weighted avg.	0,734	0,729	0,727	9 1 55   c = sibilância

### Amostra 3-6 anos

Na amostra 3-6 anos, o algoritmo *CfsSubsetEval + Best First* selecionou 35 *features*, indicadas na Tabela C.4, que se encontra na secção C.2 dos apêndices, como as mais relevantes.

A classificação com estes 35 atributos levou a um aumento pouco significativo na classe sibilância e manteve-se praticamente constante nas classes normal e fervor, comparativamente à classificação com 210 atributos e com o algoritmo SMO (Tabelas 5.3, 5.13).

O algoritmo *WrapperSubsetEval + Best First* com SMO selecionou 12 *features*, indicadas na Tabela C.5, que se encontra na secção C.2 dos apêndices, como as mais relevantes.

Verificaram-se melhorias pouco significativas na classificação relativamente à classificação com 210 atributos e com o algoritmo SMO (Tabela 5.3). Observa-se que os valores da  $f$ -measure aumentaram ligeiramente para as classes normal e sibilância, e desceram para a classe fervor (Tabela 5.13).

Relativamente ao algoritmo *ReliefFAttributeEval + Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado, cujos valores se encontram na Tabela 5.13, foi obtido através dos primeiros 80 atributos descritos na

Tabela C.6, que se encontra na secção C.2 dos apêndices. É possível observar que os valores da *f-measure* apresentam aumentos sobretudo nas classes fervor e sibilância comparativamente aos outros modelos de seleção de *features* e à classificação com 210 atributos e com o algoritmo SMO.

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features* *ReliefFAttributeEval* + *Ranker* (80 atributos).

**Tabela 5.13:** Diferentes métodos de seleção de *features* para a amostra 3-6 anos.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	fervor	0,826	0,905	0,864	a b c <- classified as
		normal	0,818	0,750	0,783	38 4 0   a = fervor
		sibilância	0,969	0,939	0,954	8 27 1   b = normal
		weighted avg.	0,866	0,865	0,864	0 2 31   c = sibilância
WrapperSubsetEval + Best First; com SMO	SMO	fervor	0,796	0,929	0,857	a b c <- classified as
		normal	0,929	0,722	0,813	39 2 1   a = fervor
		sibilância	0,912	0,939	0,925	8 26 2   b = normal
		weighted avg.	0,873	0,865	0,863	2 0 31   c = sibilância
ReliefFAttributeEval + Ranker	SMO	fervor	0,886	0,929	0,907	a b c <- classified as
		normal	0,829	0,806	0,817	39 3 0   a = fervor
		sibilância	0,938	0,909	0,923	5 29 2   b = normal
		weighted avg.	0,883	0,883	0,883	0 3 30   c = sibilância

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.3) e o melhor modelo obtido com 80 atributos (Tabela 5.13), revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value elevado de 0,3438. Estes resultados não surpreendem, pois embora existam diferenças nas medidas da  $f$ -measure entre os dois modelos, estas são pouco significativas (2,9% de diferença entre os dois resultados).

Com os primeiros 50 atributos da Tabela C.6, que se encontra na secção C.2 dos apêndices, atinge-se 95% da classificação máxima (80 atributos presentes na Tabela C.6 e com o SMO: Tabela 5.13) obtida para a amostra 3-6 anos. Os resultados da classificação com os 50 atributos estão expressos na Tabela 5.14.

**Tabela 5.14:** 95% da classificação máxima com 50 atributos para a amostra 3-6 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,760	0,905	0,826	a b c ← classified as
	normal	0,828	0,667	0,738	38 4 0   a = fervor
	sibilância	0,969	0,939	0,954	11 24 1   b = normal
	weighted avg.	0,844	0,838	0,836	1 1 31   c = sibilância

### Amostra todos os sons

Na amostra que contém todos os sons de crianças com idade inferior a 18 anos, o algoritmo *CfsSubsetEval + Best First* selecionou 45 *features*, indicadas na Tabela C.7, que se encontra na secção C.3 dos apêndices, como as mais relevantes.

Para estes 45 atributos, os valores da  $f$ -measure diminuíram ligeiramente para a classe fervor e sibilância, e aumentaram na classe normal, comparativamente à classificação com 210 atributos e com o algoritmo SMO (Tabelas 5.5 e 5.15).

O algoritmo *WrapperSubsetEval + Best First* com SMO proporcionou o melhor resultado, selecionando 22 *features* indicadas na Tabela C.8, que se encontra na secção C.3 dos apêndices, como as mais relevantes.

A classificação com estes 22 atributos apresentou valores da  $f$ -measure ligeiramente superiores para as 3 classes em relação à classificação com 210 atributos e com o algoritmo SMO e ao método *CfsSubsetEval + Best First* (Tabelas 5.5 e 5.15).

Relativamente ao algoritmo *ReliefFAttributeEval + Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado para este método, cujos valores se encontram na Tabela 5.15, obteve-se a partir dos 70 primeiros atributos descritos na Tabela C.9, que se encontra na secção C.3 dos apêndices. Verifica-se que os valores da  $f$ -measure são bastante semelhantes aos alcançados com

os 210 atributos e com o classificador SMO.

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features WrapperSubsetEval + Best First* (22 atributos).

**Tabela 5.15:** Diferentes métodos de seleção de *features* para a amostra que contém todos os sons de crianças.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	fervor	0,890	0,724	0,798	a b c <- classified as
		normal	0,694	0,889	0,779	89 26 8   a = fervor
		sibilância	0,826	0,703	0,759	8 120 7   b = normal
		weighted avg.	0,798	0,780	0,780	3 27 71   c = sibilância
WrapperSubsetEval + Best First; com SMO	SMO	fervor	0,907	0,789	0,843	a b c <- classified as
		normal	0,736	0,867	0,796	97 23 3   a = fervor
		sibilância	0,860	0,792	0,825	8 117 10   b = normal
		weighted avg.	0,829	0,819	0,820	2 19 80   c = sibilância
ReliefFAttributeEval + Ranker	SMO	fervor	0,825	0,805	0,815	a b c <- classified as
		normal	0,753	0,815	0,783	99 20 4   a = fervor
		sibilância	0,882	0,812	0,845	18 110 7   b = normal
		weighted avg.	0,814	0,811	0,811	3 16 82   c = sibilância

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.5) e o melhor modelo obtido com 22 atributos (Tabela 5.15), revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value de 0,1917. Estes resultados não surpreendem, pois embora existam diferenças nas medidas da  $f$ -measure entre os dois modelos, estas são pouco significativas (1,2% de diferença entre os dois resultados).

Com os primeiros 18 atributos da Tabela C.8, que se encontra na secção C.3 dos apêndices, atinge-se 95% da classificação máxima (22 atributos presentes na Tabela C.8 e com o SMO: Tabela 5.15) para a amostra que contém todos os sons de crianças. Os resultados da classificação com os 18 atributos estão expressos na Tabela 5.16.

**Tabela 5.16:** 95% da classificação máxima com 18 atributos para a amostra que contém todos os sons de crianças.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,876	0,748	0,807	a b c ← classified as
	normal	0,719	0,852	0,780	92 26 5   a = fervor
	sibilância	0,819	0,762	0,790	8 115 12   b = normal
	weighted avg.	0,801	0,791	0,792	5 19 77   c = sibilância

### Amostra apenas ferveores

Na amostra de sons que contém apenas ferveores, o algoritmo *CfsSubsetEval* + *Best First* selecionou 23 *features* indicadas na Tabela C.10, que se encontra na secção C.4 dos apêndices, como as mais relevantes.

Contudo, os valores da  $f$ -measure na classificação com estes 23 atributos não melhoraram comparativamente à classificação com 210 atributos e com o classificador SMO, verificando-se decréscimos nas 2 classes (Tabelas 5.7 e 5.17).

O algoritmo *WrapperSubsetEval* + *Best First* com SMO selecionou 8 *features* indicadas na Tabela C.11, que se encontra na secção C.4 dos apêndices, como as mais relevantes.

A classificação com estes 8 atributos apresentou valores da  $f$ -measure superiores nas 2 classes relativamente à classificação com 210 atributos e com o classificador SMO (Tabelas 5.7 e 5.17). Estes valores foram também melhores em relação aos alcançados pelo algoritmo *CfsSubsetEval* + *Best First*.

Relativamente ao algoritmo *ReliefFAttributeEval* + *Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado para este método, cujos valores se encontram na Tabela 5.17, foi obtido através dos 50 primeiros atributos descritos na Tabela C.12, que se encontra na secção C.4 dos apêndices.

Apesar destes resultados serem superiores aos alcançados com os 210 atributos e com o classificador SMO, não superam os alcançados com o método *WrapperSubsetEval* + *Best First*.

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features* *WrapperSubsetEval* + *Best First* (8 atributos).

**Tabela 5.17:** Diferentes métodos de seleção de *features* para a amostra apenas fervores.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	fervor	0,768	0,855	0,809	a b <- classified as
		normal	0,820	0,719	0,766	53 9   a = fervor
		weighted avg.	0,793	0,790	0,789	16 41   b = normal
WrapperSubsetEval + Best First; com SMO	SMO	fervor	0,862	0,903	0,882	a b <- classified as
		normal	0,889	0,842	0,865	56 6   a = fervor
		weighted avg.	0,875	0,874	0,874	9 48   b = normal
ReliefAttributeEval + Ranker	SMO	fervor	0,848	0,903	0,875	a b <- classified as
		normal	0,887	0,825	0,855	56 6   a = fervor
		weighted avg.	0,867	0,866	0,865	10 47   b = normal

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.7) e o melhor modelo obtido com 8 atributos (Tabela 5.17), revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value de 0,4669. Estes resultados não surpreendem, pois embora existam diferenças nas medidas da  $f$ -measure entre os dois modelos, estas são pouco significativas (2,5% de diferença entre os dois resultados).

Com os primeiros 6 atributos da Tabela C.11, que se encontra na secção C.4 dos apêndices, atinge-se 95% da classificação máxima (8 atributos presentes na Tabela C.11 e com o SMO: Tabela 5.17) para a amostra que contém apenas ferveores. Os resultados da classificação com os 6 atributos estão expressos na Tabela 5.18.

**Tabela 5.18:** 95% da classificação máxima com 6 atributos para a amostra que contém apenas ferveores.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,791	0,855	0,822	a b <- classified as
	normal	0,827	0,754	0,789	53 9   a = fervor
	weighted avg.	0,808	0,807	0,806	14 43   b = normal

### Amostra apenas sibilâncias

Na amostra de sons que contém apenas sibilâncias, o algoritmo *CfsSubsetEval* + *Best First* selecionou 25 *features*, indicadas na Tabela C.13, que se encontra na secção C.5 dos apêndices, como as mais relevantes.

Os valores da  $f$ -measure para a classificação com estes 25 atributos foram superiores nas 2 classes comparativamente aos 210 atributos com o classificador SMO (Tabelas 5.9 e 5.19).

O algoritmo *WrapperSubsetEval* + *Best First* com SMO selecionou apenas 5 *features* indicadas na Tabela C.14, que se encontra na secção C.5 dos apêndices, como as mais relevantes.

A classificação com estes 5 atributos apresentou valores da  $f$ -measure superiores comparativamente aos 210 atributos para o classificador SMO e ainda ao algoritmo *CfsSubsetEval* + *Best First*, para as 2 *classes* (Tabelas 5.9 e 5.19).

Relativamente ao algoritmo *ReliefFAttributeEval* + *Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado para este método, cujos valores se encontram na Tabela 5.19, foi obtido através dos 60 primeiros atributos descritos na Tabela C.15, que se encontra na secção C.5 dos apêndices. Apesar destes valores da  $f$ -measure serem superiores aos alcançados com 210 atributos e com o classificador SMO, não superaram os encontrados pelo método *Wrap-*

*perSubsetEval + Best First.*

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features WrapperSubsetEval + Best First* (5 atributos).

**Tabela 5.19:** Diferentes métodos de seleção de *features* para a amostra apenas sibilâncias.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	normal	0,878	0,878	0,878	a b <- classified as
		sibilância	0,891	0,891	0,891	43 6   a = normal
		weighted avg.	0,885	0,885	0,885	6 49   b = sibilância
WrapperSubsetEval + Best First; com SMO	SMO	normal	0,957	0,898	0,926	a b <- classified as
		sibilância	0,914	0,964	0,938	44 5   a = normal
		weighted avg.	0,934	0,933	0,933	2 53   b = sibilância
ReliefFAttributeEval + Ranker	SMO	normal	0,875	0,857	0,866	a b <- classified as
		sibilância	0,875	0,891	0,883	42 7   a = normal
		weighted avg.	0,875	0,875	0,875	6 49   b = sibilância

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.9) e o melhor modelo obtido com 5 atributos (Tabela 5.19), revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value de 1, o que implica que estão muito longe de serem estatisticamente significativas. Estes resultados são surpreendentes, pois existe uma diferença de 11,6% entre as medidas da  $f$ -measure nos dois modelos.

Com os primeiros 3 atributos da Tabela C.14, que se encontra na secção C.5 dos apêndices, atinge-se 95% da classificação máxima (5 atributos presentes na Tabela C.14 e com o SMO: Tabela 5.19) para a amostra que contém apenas sibilâncias. Os resultados da classificação com os 3 atributos estão expressos na Tabela 5.20.

**Tabela 5.20:** 95% da classificação máxima com 3 atributos para a amostra que contém apenas sibilâncias.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	normal	0,911	0,837	0,872	a b <- classified as
	sibilância	0,864	0,927	0,895	41 8   a = normal
	weighted avg.	0,886	0,885	0,884	4 51   b = sibilância

### 5.1.3 Otimização de parâmetros

Note-se que em todos os modelos referidos anteriormente os parâmetros utilizados foram os estabelecidos por defeito ( $C=1$ ,  $e=1$ ), com  $C$  o parâmetro de complexidade e,  $e$  o expoente do kernel polinomial.

#### Amostra 0-2 anos

Para a amostra 0-2 anos os parâmetros ótimos encontrados são de  $C=0,3$  e  $e=2$ . Os resultados para a classificação estão expressos na Tabela 5.21 e representam o melhor modelo alcançado para esta amostra. Verificou-se um aumento pouco significativo na média da  $f$ -measure para as 3 classes, sendo que todas as classes aumentaram comparativamente ao melhor resultado encontrado para a classificação após seleção das *features* (Tabela 5.11).

**Tabela 5.21:** Classificação após otimização de parâmetros para a amostra 0-2 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,734	0,855	0,790	a b c <- classified as
	normal	0,778	0,700	0,737	47 5 3   a = fervor
	sibilância	0,869	0,815	0,841	10 35 5   b = normal
	weighted avg.	0,799	0,794	0,794	7 5 53   c = sibilância

**Amostra 3-6 anos**

Para a amostra 3-6 anos os parâmetros ótimos encontrados são de  $C=0,8$  e  $e=1$ . Os resultados para a classificação estão expressos na Tabela 5.22 e representam o melhor modelo alcançado para esta amostra. Verificou-se um aumento pouco significativo na média da *f-measure* para as 3 classes, embora apenas a classe das sibilâncias e dos sons normais tenham aumentado comparativamente ao melhor resultado encontrado para a classificação após seleção das *features* (Tabela 5.13).

**Tabela 5.22:** Classificação após otimização de parâmetros para a amostra 3-6 anos.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,886	0,929	0,907	a b c <- classified as
	normal	0,853	0,806	0,829	39 3 0   a = fervor
	sibilância	0,939	0,939	0,939	5 29 2   b = normal
	weighted avg.	0,891	0,892	0,891	0 2 31   c = sibilância

**Amostra todos os sons de crianças**

Para a amostra que contém todos os sons de crianças os parâmetros ótimos encontrados são de  $C=1,7$  e  $e=3$ . Os resultados para a classificação estão expressos na Tabela 5.23 e representam o melhor modelo alcançado para esta amostra. Verificou-se um aumento da *f-measure* nas 3 classes comparativamente ao melhor resultado encontrado para a classificação após seleção das *features* (Tabela 5.15).

**Tabela 5.23:** Classificação após otimização de parâmetros para a amostra que contém todos os sons de crianças.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,920	0,837	0,877	a b c <- classified as
	normal	0,788	0,852	0,819	103 17 3   a = fervor
	sibilância	0,861	0,861	0,861	9 115 11   b = normal
	weighted avg.	0,854	0,850	0,850	0 14 87   c = sibilância

### Amostra apenas fervores

Para a amostra que contém apenas fervores os parâmetros ótimos encontrados são de  $C=1,5$  e  $e=2$ . Os resultados para a classificação estão expressos na Tabela 5.24 e representam o melhor modelo alcançado para esta amostra. Verificou-se um aumento pouco significativo na média da *f-measure* para as 2 classes, sendo que na classe dos eventos normais registou-se a maior diferença comparativamente ao melhor resultado encontrado para a classificação após seleção das *features* (Tabela 5.17).

**Tabela 5.24:** Classificação após otimização de parâmetros para a amostra que contém apenas fervores.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,887	0,887	0,887	a b <- classified as
	normal	0,877	0,877	0,877	55 7   a = fervor
	weighted avg.	0,882	0,882	0,882	7 50   b = normal

### Amostra apenas sibilâncias

Para amostra que contém apenas sibilâncias não se conseguiu obter uma otimização dos parâmetros que melhorasse os resultados obtidos na classificação. Assim, diferentes combinações verificaram exatamente os melhores resultados (*precision*, *recall*, *f-measure* e matriz de confusão) já encontrados para a classificação com 5 atributos após seleção das *features* (Tabela 5.19):  $C=1$ ,  $e=1$ ;  $C=1,9$ ,  $e=0,6$ ;  $C=2$ ,  $e=0,6$ ;  $C=2$ ,  $e=0,7$ ;  $C=2$ ,  $e=0,8$ ;  $C=2$ ,  $e=0,9$ ;  $C=2$ ,  $e=1,1$ ;  $C=2$ ,  $e=1,2$ ;  $C=2$ ,  $e=1,3$ ;  $C=2$ ,  $e=1,4$ ;  $C=2$ ,  $e=1,5$ .

## 5.2 População adulta

Nesta secção apresentam-se os resultados dos diferentes testes realizados para a população adulta.

### 5.2.1 Classificação com 210 *features*

Visualizando a Tabela 5.25, o classificador SMO foi o que atingiu maiores valores para a *f-measure* nas 3 classes. Tal como se pode observar pela matriz de confusão classificou corretamente 1730 dos 2000 eventos de fervores, 1476 dos 2000 eventos normais e 1524 dos 1819 eventos de sibilâncias. Ou seja, o maior número

de instâncias corretamente classificadas verificou-se na classe dos ferveores, seguidamente nas sibilâncias e por fim na classe dos sons normais. A *recall* e a *precision* refletem obviamente os valores da *f-measure*. Os classificadores j48 e iBk obtiveram desempenhos idênticos e superiores ao Naive Bayes, mas não ao SMO.

**Tabela 5.25:** Classificação com 210 *features* para a população adulta.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
iBk	fevor	0,781	0,673	0,723	a b c <- classified as
	normal	0,592	0,626	0,609	1346 455 199   a = fevor
	sibilância	0,665	0,724	0,693	284 1252 464   b = normal
	weighted avg.	0,680	0,673	0,674	94 408 1317   c = sibilância
j48	fevor	0,758	0,751	0,754	a b c <- classified as
	normal	0,594	0,603	0,598	1502 402 96   a = fevor
	sibilância	0,719	0,714	0,716	383 1205 412   b = normal
	weighted avg.	0,689	0,688	0,689	97 423 1299   c = sibilância
Naive Bayes	fevor	0,649	0,086	0,152	a b c <- classified as
	normal	0,450	0,781	0,571	172 1362 466   a = fevor
	sibilância	0,593	0,678	0,633	56 1562 382   b = normal
	weighted avg.	0,563	0,510	0,446	37 548 1234   c = sibilância
SMO	fevor	0,887	0,865	0,876	a b c <- classified as
	normal	0,745	0,738	0,742	1730 218 52   a = fevor
	sibilância	0,808	0,838	0,822	213 1476 311   b = normal
	weighted avg.	0,813	0,813	0,813	8 287 1524   c = sibilância

Relativamente aos resultados dos testes de significância estatística (Tabela 5.26) com um  $\alpha=0,01$ , verificaram-se diferenças estatisticamente significativas entre os 4 classificadores, com exceção do teste SMO vs J48 com um  $p\text{-value}=0,0149$ , o que implica que está próximo de ser estatisticamente significativo, aliás para um  $\alpha=0,05$  este teste já se revelou estatisticamente significativo (Tabela B.6 na secção B dos apêndices). Este resultado não surpreende dado que o classificador j48 apresentou valores mais baixos da *f-measure* em relação ao SMO. Assim sendo, considera-se o SMO o classificador mais adequado à amostra dos adultos, uma vez que foi o que apresentou melhores resultados.

**Tabela 5.26:** Testes de significância estatística com  $\alpha=0,01$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.25, para a população adulta.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=1; p= $1,4731 \times 10^{-13}$	h=1; p= $7,4789 \times 10^{-126}$	h=1; p= $4,2125 \times 10^{-10}$
j48	–	–	h=1; p= $2,6640 \times 10^{-191}$	h=0; p=0,0149
Naive Bayes	–	–	–	h=1; p= $9,3894 \times 10^{-187}$
SMO	–	–	–	–

### 5.2.2 Seleção de *features*

O algoritmo *CfsSubsetEval + Best First* selecionou as 45 *features* indicadas na Tabela C.16, que se encontra na secção C.6 dos apêndices, como as mais relevantes. Contudo, a classificação com estes 45 atributos não melhorou comparativamente à classificação com 210 atributos e com o algoritmo SMO (Tabela 5.25), observando-se decréscimos dos valores da *f-measure* para as 3 classes (Tabelas 5.25 e 5.27).

Relativamente ao algoritmo *ReliefFAttributeEval + Ranker*, os 210 atributos foram ordenados de acordo com a sua relevância. O melhor resultado com o algoritmo, cujos valores se encontram na Tabela 5.27, obteve-se com os primeiros 185 atributos descritos na Tabela C.17, que se encontra na secção C.6 dos apêndices. É possível observar aumentos praticamente insignificantes nos valores da *f-measure* para as classes sibilâncias e sons normais comparativamente à classificação com 210 atributos e com o algoritmo SMO. Os resultados do método *ReliefFAttributeEval + Ranker* superam os obtidos pelo *CfsSubsetEval + Best First*, embora se mantenham praticamente iguais aos obtidos pela classificação com todas as *features*.

Portanto, para esta amostra o melhor resultado obteve-se através do método de seleção de *features* *ReliefFAttributeEval + Ranker* (185 atributos).

De salientar também que, atendendo ao grande conjunto de dados da população adulta não foi possível usar o método *WrapperSubsetEval + Best First* usado na população infantil.

**Tabela 5.27:** Diferentes métodos de seleção de *features* para a população adulta.

Attribute Evaluator + Search Method	Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
CfsSubsetEval + Best First	SMO	fervor	0,804	0,793	0,798	a b c <- classified as
		normal	0,637	0,602	0,619	1585 314 101   a = fervor
		sibilância	0,715	0,769	0,741	339 1204 457   b = normal
		weighted avg.	0,719	0,720	0,719	47 373 1399   c = sibilância
ReliefAttributeEval + Ranker	SMO	fervor	0,886	0,867	0,876	a b c <- classified as
		normal	0,747	0,738	0,743	1733 215 52   a = fervor
		sibilância	0,808	0,839	0,823	213 1476 311   b = normal
		weighted avg.	0,814	0,814	0,814	9 284 1526   c = sibilância

Os testes de significância estatística com  $\alpha=0,01$  e  $\alpha=0,05$ , entre a classificação obtida com os 210 atributos (Tabela 5.25) e o melhor modelo obtido com 185 atributos (Tabela 5.27) revelaram que as diferenças não são estatisticamente significativas com um  $p$ -value elevado de 0,7316, o que implica que as diferenças estão longe de serem estatisticamente significativas. Estes resultados não surpreendem, visto que as diferenças nas medidas da  $f$ -measure para os dois modelos são praticamente insignificantes (0,1% de diferença entre os dois resultados).

Com os primeiros 42 atributos da Tabela C.17, que se encontra na secção C.6 dos apêndices, atinge-se 95% da classificação máxima (185 atributos presentes na Tabela C.17 e com o SMO: Tabela 5.27) obtida para a população adulta. Os resultados da classificação com os 42 atributos estão expressos na Tabela 5.28.

**Tabela 5.28:** 95% da classificação máxima com 42 atributos para a população adulta.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,844	0,841	0,842	a b c ← classified as
	normal	0,691	0,692	0,691	1681 254 65   a = fervor
	sibilância	0,778	0,781	0,779	276 1383 341   b = normal
	weighted avg.	0,771	0,771	0,771	34 364 1421   c = sibilância

### 5.2.3 Otimização de parâmetros

Para a população adulta os parâmetros ótimos encontrados são de  $C=2$  e  $e=1,5$ . Os resultados para a classificação estão expressos na Tabela 5.29 e representam o melhor modelo alcançado para esta amostra. Verificou-se um aumento da  $f$ -measure para as 3 classes comparativamente ao melhor resultado encontrado para a classificação após seleção das *features* (Tabela 5.27).

**Tabela 5.29:** Classificação após otimização de parâmetros para a população adulta.

Classifier	Class	Precision	Recall	F-measure	Confusion Matrix
SMO	fervor	0,908	0,884	0,896	a b c ← classified as
	normal	0,786	0,780	0,783	1768 188 44   a = fervor
	sibilância	0,831	0,862	0,846	166 1560 274   b = normal
	weighted avg.	0,842	0,841	0,842	14 237 1568   c = sibilância

## Discussão de Resultados

Relativamente à classificação com as 210 *features*, concluiu-se que os resultados foram bastante satisfatórios com uma *f-measure* média de 0,750; 0,854; 0,808; 0,849; 0,817; 0,813 para as amostras 0-2 anos, 3-6 anos, todos os sons de crianças, apenas ferveores (crianças), apenas sibilâncias (crianças) e adultos, respetivamente.

Em relação aos testes de significância estatística verificou-se que existiam diferenças mais significativas entre os 4 classificadores na população adulta do que na população infantil. Isso deve-se provavelmente à dimensão da base de dados que contém os sons de adultos, pois como existem mais gravações, os testes de significância estatística são mais conclusivos. Na população infantil, sendo o número de gravações mais reduzido, há uma maior incerteza relativamente à significância estatística.

Após a obtenção do número ótimo de *features*, os resultados melhoraram, obtendo-se uma *f-measure* média de 0,761; 0,883; 0,820; 0,874; 0,933; 0,814 para as amostras 0-2 anos, 3-6 anos, todos os sons de crianças, apenas ferveores (crianças), apenas sibilâncias (crianças) e adultos, respetivamente.

Em relação à seleção de *features*, verificou-se que as *features* MFCCs, *spectral rolloff* 95, *spectral rolloff* 85,  $F_0$  *spectrum*,  $F_0$  *cepstrum*, *spectral irregularity*, *chromagram centroid* (exceto para a amostra 0-2 anos), *zero-crossing rate* (exceto para a amostra apenas sibilâncias da população infantil) e *spectral flux* (não tão frequente como as referidas anteriormente) foram as mais relevantes tanto em crianças como em adultos. Outras, como o máximo da entropia, WS-SS e HCDF são muito mais relevantes em adultos do que em crianças. Já a *feature spectral entropy* apresentou maior importância na população infantil do que na população adulta. Por outro lado, outras *features* são praticamente insignificantes, como o máximo da energia de Teager e máximo KFD do filtro WPST-NST, para todas as amostras. Outros casos, como a *feature spectral brightness* que se revelou com pouca importância em quase todas as amostras à exceção da amostra 3-6 anos, ou o *spectral kurtosis* que embora tenha alguma importância na amostra adultos e apenas sibilâncias, é pouco relevante na população infantil. Nota-se uma tendência semelhante nas *features* menos

frequentes na amostra 0-2 anos e na amostra apenas sibilâncias, isto porque a amostra 0-2 anos é constituída pelo dobro de eventos de sibilâncias comparativamente à amostra 3-6 anos, por exemplo. Observa-se ainda que a *feature mode* não se demonstrou frequente na amostra adultos, sibilâncias, ferveores e todos os sons de crianças, mas nas amostras individuais 0-2 anos e 3-6 anos já apresentou maior importância. Ainda assim, é importante considerar que não houve grande consistência nas *features* selecionadas pelos diferentes métodos nem nas diferentes amostras testadas, além de que os resultados do *ranking* foram bastante semelhantes.

A otimização de resultados permitiu obter os modelos ótimos para cada uma das amostras, com uma *f-measure* média de 0,794; 0,891; 0,850; 0,882; 0,933; 0,842 para as amostras 0-2 anos, 3-6 anos, todos os sons de crianças, apenas ferveores (crianças), apenas sibilâncias (crianças) e adultos, respetivamente. Analisando as instâncias classificadas incorretamente para os modelos ótimos da população infantil, observou-se que o maior número de falhas na classificação ocorreu para os pacientes com os seguintes IDs: 149, 159, 161, 165, 167, 179 e 188 (Tabela 4.1). Ou seja, a maioria das instâncias classificadas erradamente correspondem a crianças inseridas na amostra 0-2 anos com patologias respiratórias e com sons adventícios anotados. Estes resultados justificam o facto da amostra 0-2 anos ser a que apresenta os valores mais baixos para a classificação. Isto pode dever-se ao facto das crianças entre os 0-2 anos serem as mais difíceis de auscultar, devido à falta de cooperação com os profissionais de saúde, às movimentações inesperadas e ao choro.

Ao separar sibilâncias de ferveores os resultados melhoraram, pois classificou-se isoladamente ferveores vs sons normais e sibilâncias vs sons normais.

Por último, é importante notar que contrariamente ao que se esperava, os resultados da população infantil foram semelhantes aos da população adulta. Contudo, o som é filtrado de forma diferente pelos pulmões e pela caixa torácica das crianças, uma vez que estas apresentam um menor número de alvéolos, menor tamanho de pulmões e caixa torácica, e menos massa muscular o que vai afetar a forma como o som se transmite, logo as *features* selecionadas para a população adulta e infantil foram diferentes.

## Conclusão e trabalho futuro

Esta dissertação propôs usar *features* do estado de arte, anteriormente testadas em adultos, na população infantil. Os resultados foram bastante promissores, comprovando-se a grande potencialidade dessas *features* para a classificação de sons adventícios em crianças. Contudo, embora os resultados da população adulta e infantil tenham sido semelhantes, devido às diferenças fisiológicas entre essas populações, as *features* selecionadas foram diferentes. Ficou também demonstrado que a estratificação por idades na população infantil é significativa e melhora os resultados da classificação.

Seguem-se algumas propostas de possíveis trabalhos futuros:

- Aumentar a base de dados para avaliar uma maior variabilidade de sibilâncias e ferveores, permitindo que a estratificação de idades contenha um número equilibrado de sons adventícios;
- Desenvolver um software *open source* que inclua aquisição, detecção e classificação de sons adventícios para a população adulta e infantil, integrando o trabalho desenvolvido nesta dissertação;
- Alargar a detecção e classificação de sons adventícios a roncos, *squawks*, estridor, ferveores *fine* e *coarse*, sibilâncias monofónicas e polifónicas;
- Analisar as instâncias classificadas incorretamente com base no local de aquisição da gravação sonora (traqueia/tórax), para além do paciente, a fim de averiguar a influência do local de aquisição;
- Investigar as diferenças nas *features* selecionadas para a população infantil e adulta.



# Referências

- [Abbas and Fahim, 2010] Abbas, A. and Fahim, A. (2010). An automated computerized auscultation and diagnostic system for pulmonary diseases. *Journal of medical systems*, 34(6):1149–1155.
- [Abbasi et al., 2013] Abbasi, S., Derakhshanfar, R., Abbasi, A., and Sarbaz, Y. (2013). Classification of normal and abnormal lung sounds using neural network and support vector machines. In *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, pages 1–4. IEEE.
- [Alsmadi and Kahya, 2008] Alsmadi, S. and Kahya, Y. P. (2008). Design of a dsp-based instrument for real-time classification of pulmonary sounds. *Computers in biology and medicine*, 38(1):53–61.
- [Altman, 1992] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.
- [Analytics Vidhya, 2018] Analytics Vidhya (2018). Introduction to k-nearest neighbors: A powerful machine learning algorithm (with implementation in python r). <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>. [Online; Acedido: 4 dezembro 2019].
- [Aykanat et al., 2017] Aykanat, M., Kılıç, Ö., Kurt, B., and Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017(1):65.
- [Bahoura, 2009] Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in biology and medicine*, 39(9):824–843.
- [Bahoura and Lu, 2006] Bahoura, M. and Lu, X. (2006). Separation of crackles from vesicular sounds using wavelet packet transform. In *2006 IEEE International*

- Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II. IEEE.
- [Bahoura and Pelletier, 2003] Bahoura, M. and Pelletier, C. (2003). New parameters for respiratory sound classification. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, volume 3, pages 1457–1460. IEEE.
- [Bahoura and Pelletier, 2004] Bahoura, M. and Pelletier, C. (2004). Respiratory sounds classification using gaussian mixture models. In *Electrical and Computer Engineering, 2004. Canadian Conference on*, volume 3, pages 1309–1312. IEEE.
- [Baken and Orlikoff, 2000] Baken, R. J. and Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.
- [Baydar et al., 2003] Baydar, K. S., Ertuzun, A., and Kahya, Y. P. (2003). Analysis and classification of respiratory sounds by signal coherence method. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, volume 3, pages 2950–2953. IEEE.
- [Bingham, 1987] Bingham, W. (1987). The neonatal airway: Problems and management. *Canadian Family Physician*, 33:1467.
- [Bohadana et al., 2014] Bohadana, A., Izbicki, G., and Kraman, S. S. (2014). Fundamentals of lung auscultation. *New England Journal of Medicine*, 370(8):744–751.
- [Bokov et al., 2016] Bokov, P., Mahut, B., Flaud, P., and Delclaux, C. (2016). Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Computers in biology and medicine*, 70:40–50.
- [Bozkurt et al., 2007] Bozkurt, B., Couvreur, L., and Dutoit, T. (2007). Chirp group delay analysis of speech signals. *Speech communication*, 49(3):159–176.
- [Bunin and Loudon, 1979] Bunin, N. J. and Loudon, R. G. (1979). Lung sound terminology in case reports. *Chest*, 76(6):690–692.
- [Charleston-Villalobos et al., 2011] Charleston-Villalobos, S., Martinez-Hernandez, G., Gonzalez-Camarena, R., Chi-Lem, G., Carrillo, J. G., and Aljama-Corrales, T. (2011). Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. *Computers in biology and medicine*, 41(7):473–482.

- [Chen et al., 2015] Chen, C.-H., Huang, W.-T., Tan, T.-H., Chang, C.-C., and Chang, Y.-J. (2015). Using k-nearest neighbor classification to diagnose abnormal lung sounds. *Sensors*, 15(6):13132–13158.
- [Cherian et al., 2005] Cherian, T., Mulholland, E. K., Carlin, J. B., Ostensen, H., Amin, R., Campo, M. d., Greenberg, D., Lagos, R., Lucero, M., Madhi, S. A., et al. (2005). Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bulletin of the World Health Organization*, 83:353–359.
- [Chicco, 2017] Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1):35.
- [Chien et al., 2007] Chien, J.-C., Wu, H.-D., Chong, F.-C., and Li, C.-I. (2007). Wheeze detection using cepstral analysis in gaussian mixture models. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3168–3171. IEEE.
- [Coomans and Massart, 1982] Coomans, D. and Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, 136:15–27.
- [Drugman et al., 2011] Drugman, T., Dubuisson, T., and Dutoit, T. (2011). Phase-based information for voice pathology detection. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4612–4615. IEEE.
- [Ellington et al., 2014] Ellington, L. E., Emmanouilidou, D., Elhilali, M., Gilman, R. H., Tielsch, J. M., Chavez, M. A., Marin-Concha, J., Figueroa, D., West, J., and Checkley, W. (2014). Developing a reference of normal lung sounds in healthy peruvian children. *Lung*, 192(5):765–773.
- [Ellington et al., 2012] Ellington, L. E., Gilman, R. H., Tielsch, J. M., Steinhoff, M., Figueroa, D., Rodriguez, S., Caffo, B., Tracey, B., Elhilali, M., West, J., et al. (2012). Computerised lung sound analysis to improve the specificity of paediatric pneumonia diagnosis in resource-poor settings: protocol and methods for an observational study. *BMJ open*, 2(1):e000506.
- [Emmanouilidou and Elhilal, 2013] Emmanouilidou, D. and Elhilal, M. (2013). Characterization of noise contaminations in lung sound recordings. In *2013 35th*

- Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2551–2554. IEEE.
- [Emmanouilidou et al., 2018] Emmanouilidou, D., McCollum, E. D., Park, D. E., and Elhilali, M. (2018). Computerized lung sound screening for pediatric auscultation in noisy field environments. *IEEE Transactions on Biomedical Engineering*, 65(7):1564–1574.
- [Emmanouilidou et al., 2012] Emmanouilidou, D., Patil, K., West, J., and Elhilali, M. (2012). A multiresolution analysis for detection of abnormal lung sounds. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3139–3142. IEEE.
- [Epler et al., 1978] Epler, G. R., Carrington, C. B., and Gaensler, E. A. (1978). Crackles (rales) in the interstitial pulmonary diseases. *Chest*, 73(3):333–339.
- [Fenton et al., 1985] Fenton, T. R., Pasterkamp, H., Tal, A., and Chernick, V. (1985). Automated spectral characterization of wheezing in asthmatic children. *IEEE transactions on biomedical engineering*, (1):50–55.
- [Fischer et al., 2016] Fischer, H. S., Puder, L. C., Wilitzki, S., Usemann, J., Bühner, C., Godfrey, S., and Schmalisch, G. (2016). Relationship between computerized wheeze detection and lung function parameters in young infants. *Pediatric pulmonology*, 51(4):402–410.
- [Gavriely et al., 1995] Gavriely, N., Nissan, M., Rubin, A., and Cugell, D. W. (1995). Spectral characteristics of chest wall breath sounds in normal subjects. *Thorax*, 50(12):1292–1300.
- [Gnitecki et al., 2004] Gnitecki, J., Moussavi, Z., and Pasterkamp, H. (2004). Classification of lung sounds during bronchial provocation using waveform fractal dimensions. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 2, pages 3844–3847. IEEE.
- [Gonzalez et al., 2004] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). *Digital image processing using MATLAB*. Pearson Education India.
- [Gross et al., 2000] Gross, V., Dittmar, A., Penzel, T., Schuttler, F., and Von Wichert, P. (2000). The relationship between normal lung sounds, age, and gender. *American journal of respiratory and critical care medicine*, 162(3):905–909.

- 
- [Güler et al., 2005] Güler, İ., Polat, H., and Ergün, U. (2005). Combining neural network and genetic algorithm for prediction of lung sounds. *Journal of Medical Systems*, 29(3):217–231.
- [Gurung et al., 2011] Gurung, A., Scrafford, C. G., Tielsch, J. M., Levine, O. S., and Checkley, W. (2011). Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respiratory medicine*, 105(9):1396–1403.
- [Hall et al., 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- [Hall, 1998] Hall, M. A. (1998). *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand.
- [Hashemi et al., 2011] Hashemi, A., Arabalibiek, H., and Agin, K. (2011). Classification of wheeze sounds using wavelets and neural networks. In *International Conference on Biomedical Engineering and Technology*, volume 11, pages 127–131. IACSIT Press.
- [Hsu et al., 2003] Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.
- [ICBHI Challenge, 2017] ICBHI Challenge (2017). Respiratory sound database. <https://bhichallenge.med.auth.gr/>. [Online; Acedido: 1 setembro 2019].
- [İçer and Gengeç, 2014] İçer, S. and Gengeç, Ş. (2014). Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing*, 28:18–27.
- [Jaskowiak and Campello, 2011] Jaskowiak, P. A. and Campello, R. (2011). Comparing correlation coefficients as dissimilarity measures for cancer classification in gene expression data. In *Proceedings of the Brazilian symposium on bioinformatics*, pages 1–8. Brasília.
- [Jensen and Arnsfang, 1999] Jensen, K. and Arnsfang, J. (1999). Binary decision tree classification of musical sounds. In *ICMC*.
- [Jin et al., 2011] Jin, F., Krishnan, S., and Sattar, F. (2011). Adventitious sounds identification and extraction using temporal–spectral dominance-based features. *IEEE Transactions on Biomedical Engineering*, 58(11):3078–3087.

- [Jin et al., 2014] Jin, F., Sattar, F., and Goh, D. Y. (2014). New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. *Neurocomputing*, 123:362–371.
- [John and Langley, 1995] John, G. H. and Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc.
- [Kahya et al., 1997] Kahya, Y. P., Guler, E. C., and Sahin, S. (1997). Respiratory disease diagnosis using lung sounds. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No. 97CH36136)*, volume 5, pages 2051–2053. IEEE.
- [Kandaswamy et al., 2004] Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., and Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in biology and medicine*, 34(6):523–537.
- [Kira and Rendell, 1992] Kira, K. and Rendell, L. A. (1992). A practical approach to feature selection. In Sleeman, D. H. and Edwards, P., editors, *Ninth International Workshop on Machine Learning*, pages 249–256. Morgan Kaufmann.
- [Kohavi and John, 1997] Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324. Special issue on relevance.
- [Krom, 1993] Krom, G. d. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language, and Hearing Research*, 36(2):254–266.
- [Liu and Motoda, 2012] Liu, H. and Motoda, H. (2012). *Feature selection for knowledge discovery and data mining*, volume 454. Springer Science & Business Media.
- [Lozano et al., 2016] Lozano, M., Fiz, J. A., and Jané, R. (2016). Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency. *IEEE journal of biomedical and health informatics*, 20(2):486–497.

- [Lu and Bahoura, 2008] Lu, X. and Bahoura, M. (2008). An integrated automated system for crackles extraction and classification. *Biomedical Signal Processing and Control*, 3(3):244–254.
- [Management Association, 2019] Management Association, I. (2019). *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*. IGI Global.
- [Marini et al., 1979] Marini, J. J., Pierson, D. J., Hudson, L. D., and Lakshminarayan, S. (1979). The significance of wheezing in chronic airflow obstruction. *American Review of Respiratory Disease*, 120(5):1069–1072.
- [Marques et al., 2014] Marques, A., Oliveira, A., and Jácome, C. (2014). Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: a systematic review. *Respiratory care*, 59(5):765–776.
- [Martinez-Hernandez et al., 2006] Martinez-Hernandez, H., Aljama-Corrales, C., Gonzalez-Camarena, R., Charleston-Villalobos, V., and Chi-Lem, G. (2006). Computerized classification of normal and abnormal lung sounds by multivariate linear autoregressive model. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 5999–6002. IEEE.
- [Mazić et al., 2015] Mazić, I., Bonković, M., and Džaja, B. (2015). Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children’s respiratory sounds. *Biomedical Signal Processing and Control*, 21:105–118.
- [Mendes et al., 2015] Mendes, L., Vogiatzis, I., Perantoni, E., Kaimakamis, E., Chouvarda, I., Maglaveras, N., Tsara, V., Teixeira, C., Carvalho, P., Henriques, J., et al. (2015). Detection of wheezes using their signature in the spectrogram space and musical features. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5581–5584. IEEE.
- [Mendes et al., 2016] Mendes, L., Vogiatzis, I. M., Perantoni, E., Kaimakamis, E., Chouvarda, I., Maglaveras, N., Henriques, J., Carvalho, P., and Paiva, R. P. (2016). Detection of crackle events using a multi-feature approach. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3679–3683. IEEE.
- [Merkus, 2003] Merkus, P. J. (2003). Effects of childhood respiratory diseases on the anatomical and functional development of the respiratory system. *Paediatric respiratory reviews*, 4(1):28–39.

- [Mondal et al., 2014] Mondal, A., Bhattacharya, P., and Saha, G. (2014). Detection of lungs status using morphological complexities of respiratory sounds. *The Scientific World Journal*, 2014.
- [Morillo et al., 2013] Morillo, D. S., Moreno, S. A., Granero, M. Á. F., and Jiménez, A. L. (2013). Computerized analysis of respiratory sounds during copd exacerbations. *Computers in biology and medicine*, 43(7):914–921.
- [Nath and Capel, 1974] Nath, A. and Capel, L. (1974). Inspiratory crackles—early and late. *Thorax*, 29(2):223–227.
- [Neuman et al., 2012] Neuman, M. I., Lee, E. Y., Bixby, S., Diperna, S., Hellinger, J., Markowitz, R., Servaes, S., Monuteaux, M. C., and Shah, S. S. (2012). Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. *Journal of hospital medicine*, 7(4):294–298.
- [Observatório Nacional das Doenças Respiratórias, 2018] Observatório Nacional das Doenças Respiratórias (2018). Relatório completo ondr 2018. <https://www.ondr.pt/index.html>. [Online; Acedido: 27 dezembro 2019].
- [Oliveira, 2014] Oliveira, A. L. A. (2014). *Sons respiratórios adventícios em crianças com infeção respiratória*. Tese de Mestrado em Fisioterapia, Escola Superior de Saúde, Universidade de Aveiro.
- [Olivier Lartillot et al., 2007] Olivier Lartillot et al. (2007). Mirtoolbox documentation. <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox>. [Online; Acedido: 8 dezembro 2019].
- [Pasterkamp et al., 2016] Pasterkamp, H., Brand, P. L., Everard, M., Garcia-Marcos, L., Melbye, H., and Priftis, K. N. (2016). Towards the standardisation of lung sound nomenclature. *European Respiratory Journal*, 47(3):724–732.
- [Pasterkamp et al., 1997] Pasterkamp, H., Kraman, S. S., and Wodicka, G. R. (1997). Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 156(3):974–987.
- [Pasterkamp et al., 1996] Pasterkamp, H., Powell, R. E., and Sanchez, I. (1996). Lung sound spectra at standardized air flow in normal infants, children, and adults. *American journal of respiratory and critical care medicine*, 154(2):424–430.

- [Pesu et al., 1996] Pesu, L., Ademovic, E., Pesquet, J.-C., and Helisto, P. (1996). Wavelet packet based respiratory sound classification. In *Proceedings of Third International Symposium on Time-Frequency and Time-Scale Analysis (TFTS-96)*, pages 377–380. IEEE.
- [Piiirila and Sovijjarvi, 1995] Piiirila, P. and Sovijjarvi, A. (1995). Crackles: recording, analysis and clinical significance. *European Respiratory Journal*, 8(12):2139–2148.
- [Pinho et al., 2015] Pinho, C., Oliveira, A., Jácome, C., Rodrigues, J., and Marques, A. (2015). Automatic crackle detection algorithm based on fractal dimension and box filtering. *Procedia Computer Science*, 64:705–712.
- [Piramuthu and Sikora, 2009] Piramuthu, S. and Sikora, R. T. (2009). Iterative feature construction for improving inductive learning algorithms. *Expert Systems with Applications*, 36(2):3401–3406.
- [Plante et al., 1998] Plante, F., Kessler, H., Sun, X., Cheetham, B., and Earis, J. (1998). Inverse filtering applied to upper airway sounds. *Technology and Health Care*, 6(1):23–32.
- [Platt, 1999] Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization, advances in kernel methods. *Support Vector Learning*, pages 185–208.
- [Pramono et al., 2017] Pramono, R. X. A., Bowyer, S., and Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PloS one*, 12(5):e0177926.
- [Puder et al., 2016] Puder, L. C., Wilitzki, S., Bühner, C., Fischer, H. S., and Schmalisch, G. (2016). Computerized wheeze detection in young infants: comparison of signals from tracheal and chest wall sensors. *Physiological measurement*, 37(12):2170.
- [Quinlan, 1993] Quinlan, J. (1993). The morgan kaufmann series in machine learning. *San Mateo*.
- [Reichert et al., 2008] Reichert, S., Gass, R., Brandt, C., and Andrés, E. (2008). Analysis of respiratory sounds: state of the art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2:CCRPM–S530.

- [Rocha et al., 2018] Rocha, B., Mendes, L., Chouvarda, I., Carvalho, P., and Paiva, R. (2018). Detection of cough and adventitious respiratory sounds in audio recordings by internal sound analysis. In *Precision Medicine Powered by pHealth and Connected Health*, pages 51–55. Springer.
- [Sarkar et al., 2015] Sarkar, M., Madabhavi, I., Niranjana, N., and Dogra, M. (2015). Auscultation of the respiratory system. *Annals of thoracic medicine*, 10(3):158.
- [Semedo et al., 2015] Semedo, J., Oliveira, A., Machado, A., Moreira, J., Rodrigues, J., Aparício, J., Pasterkamp, H., Jesus, L. M., and Marques, A. (2015). Computerised lung auscultation—sound software (class). *Procedia Computer Science*, 64:697–704.
- [Sengupta et al., 2016] Sengupta, N., Sahidullah, M., and Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Computers in biology and medicine*, 75:118–129.
- [Serbes et al., 2011] Serbes, G., Sakar, C. O., Kahya, Y. P., and Aydin, N. (2011). Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3314–3317. IEEE.
- [Shue et al., 2009] Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K. (2009). Voice-sauce. *p*. Program available online at <http://www.seas.ucla.edu/spapl/voicesauce/>. UCLA.
- [Song, 2015] Song, I. (2015). Diagnosis of pneumonia from sounds collected using low cost cell phones. In *2015 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE.
- [Source Forge, 2009a] Source Forge (2009a). Class j48. <https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/J48.html>. [Online; Acedido: 2 dezembro 2019].
- [Source Forge, 2009b] Source Forge (2009b). Class naivebayes. <https://weka.sourceforge.io/doc.dev/weka/classifiers/bayes/NaiveBayes.html>. [Online; Acedido: 2 dezembro 2019].
- [Sovijarvi et al., 2000] Sovijarvi, A., Dalmasso, F., Vanderschoot, J., Malmberg, L., Righini, G., and Stoneman, S. (2000). Definition of terms for applications of respiratory sounds. *European Respiratory Review*, 10(77):597–610.

- [Sovijärvi et al., 1997] Sovijärvi, A., Vanderschoot, J., and Earis, J. (1997). Standardization of computerized respiratory sound analysis. *Crit Care Med*, 156:974–987.
- [Taplidou and Hadjileontiadis, 2007] Taplidou, S. A. and Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. *Computers in biology and medicine*, 37(8):1073–1083.
- [Ulukaya et al., 2015] Ulukaya, S., Sen, I., and Kahya, Y. P. (2015). Feature extraction using time-frequency analysis for monophonic-polyphonic wheeze discrimination. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5412–5415. IEEE.
- [Vannuccini et al., 2000] Vannuccini, L., Earis, J., Helisto, P., Cheetham, B., Rossi, M., Sovijarvi, A., and Vanderschoot, J. (2000). Capturing and preprocessing of respiratory sounds. *European Respiratory Review*, 10(77):616–620.
- [Vannuccini et al., 1998] Vannuccini, L., Rossi, M., and Pasquali, G. (1998). A new method to detect crackles in respiratory sounds. *Technology and Health Care*, 6(1):75–79.
- [Waitman et al., 2000] Waitman, L. R., Clarkson, K. P., Barwise, J. A., and King, P. H. (2000). Representation and classification of breath sounds recorded in an intensive care setting using neural networks. *Journal of clinical monitoring and computing*, 16(2):95–105.
- [Witten et al., 2005] Witten, I. H., Frank, E., and Hall, M. A. (2005). Practical machine learning tools and techniques. *Morgan Kaufmann*, page 578.
- [World Health Organization, 2018] World Health Organization (2018). The top 10 causes of death 2016. <https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death/>. [Online; Acedido: 10 fevereiro 2019].
- [Xie et al., 2012] Xie, S., Jin, F., Krishnan, S., and Sattar, F. (2012). Signal feature extraction by multi-scale pca and its application to respiratory sound classification. *Medical & biological engineering & computing*, 50(7):759–768.
- [Yeginer et al., 2004] Yeginer, M., Ciftci, K., Cini, U., Sen, I., Kilinc, G., and Kahya, Y. (2004). Using lung sounds in classification of pulmonary diseases according to respiratory subphases. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 482–485. IEEE.

- [Yeginer and Kahya, 2008] Yeginer, M. and Kahya, Y. P. (2008). Elimination of vesicular sounds from pulmonary crackle waveforms. *Computer methods and programs in biomedicine*, 89(1):1–13.
- [Yu et al., 2013] Yu, C., Tsai, T.-H., Huang, S.-I., and Lin, C.-W. (2013). Soft stethoscope for detecting asthma wheeze in young children. *Sensors*, 13(6):7399–7413.

# Apêndices



A

Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

## A.1 Crianças

**Tabela A.1:** Resumo dos métodos de extração de *features* e classificação para SRs em crianças presentes na literatura.

Referências	População	Local de aquisição	Duração das gravações	Sons respiratórios	Features	Algoritmos de machine learning	Resultados
[Fenton et al., 1985]	5 crianças asmáticas, com 10 - 16 anos	Peito e traqueia	20 minutos	1 categoria: sibilâncias	Parâmetros do espectro de potência: limite superior de frequência $f_u$ e a frequência da potência máxima $f_p$		
[Gnitecki et al., 2004]	Crianças com 9 a 15 anos	Lobo inferior direito posterior pulmonar	70 s	1 categoria: ferveores	Análises no domínio do tempo, RMS e fractais: a SNR, a VFD e KFD	1-NN	Verdadeiros positivos (falsos positivos) : 90,3%(23,4%), 63,5%(24,9%) e 58,3%(26,1%) para RMS-SNR + KFD, RMS-SNR +VFD e RMS-SNR, respetivamente
[Jin et al., 2011]	Grupo 1: 7 indivíduos saudáveis e 14 patológicos (8 sexo masculino/13 sexo feminino, 15±9 anos) Grupo 2: 3 indivíduos saudáveis e 12 patológicos (9 sexo masculino/6 sexo feminino, 11±7 anos) Patologias: asma, pneumonia e bronquiolite	Tórax posterior direito/esquerdo superior/inferior	600 s	4 categorias: sons normais, estridor, sibilâncias monofónicas e polifónicas	3 <i>features</i> baseadas em derivada espectral: derivada espectral média, <i>spread</i> temporal, posição espectral	k-NN	Exatidão=92,4±2,9%
[Emmanouilidou et al., 2012]	Crianças saudáveis ou com doenças respiratórias: 10 casos normais, 10 com sibilâncias e 8 com ferveores.		15 s	2 categorias: sibilâncias e ferveores	Parâmetros do espectro de potência	SVM	Sensibilidade = 89,44% e especificidade = 80,50%

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Xie et al., 2012]	Treino: 7 indivíduos saudáveis e 14 patológicos (8 sexo masculino/13 sexo feminino, 15±9 anos) Teste: 3 indivíduos saudáveis e 12 patológicos (9 sexo masculino/6 sexo feminino, 11±7 anos) Patologias: asma, pneumonia e bronquiolite	Tórax posterior direito/esquerdo superior/inferior			Combinação da DWT com a análise de PCA	Classificador empírico	Exatidão=97,3±2,7%
[Emmanouilidou and Elhilal, 2013]	53 crianças				Largura do pico, inclinação do espectro, <i>power ratio</i> , taxa de frequência	SVM	
[Yu et al., 2013]	8 crianças: (6 sexo masculino/2 sexo feminino (1-7 anos, com uma média de 2,5± 1,87)	Pescoço	5 e 30 s	1 categoria: sibilâncias	Coefficiente de correlação do espectro respiratório		Sensibilidade=88% e especificidade=94%
[Ellington et al., 2014]	151 gravações (71 sexo feminino/80 sexo masculino): 186 crianças saudáveis, com uma média de 2,2±1,4 anos, onde 47% eram do sexo masculino, 27% eram crianças (2-12 meses), 44% tinham 1-2 anos e 30% tinham 3-5 anos	8 locais torácicos: esquerdo e direito anterior superior (AS), direito e esquerdo anterior inferior (AI), posterior superior direito e esquerdo (PS) e posterior inferior esquerdo e direito (PI)	10 s		10 atributos espectrais e espectro-temporais: largura de pico, inclinação do espectro, <i>power of regression line</i> e <i>power ratio</i> ; MFCCs; parâmetros de TF (forma espectral e modulações temporais)		

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Jin et al., 2014]	Sons traqueais: 7 indivíduos saudáveis e 14 patológicos (8 sexo masculino/13 sexo feminino, com $15\pm 9$ anos); Sons pulmonares: 5 indivíduos saudáveis e 19 patológicos (14 sexo masculino/10 sexo feminino, com $11\pm 7$ anos)	Traqueia, pulmões/tórax superior/inferior	600 s	4 categorias: sibilâncias, estridor, roncos, e mistura destes	DP, KP, SEP		Exatidões para expiração e inspiração, respetivamente 100% e 97,1%
[Mazić et al., 2015]	45 sons de 16 crianças com idades entre 1-6 anos		10 s	2 categorias: sibilâncias e sons não asmáticos	MFCCs, <i>kurtosis</i> e entropia	GMM ou SVM	Exatidão=95%
[Song, 2015]	376 crianças (0,3-5 anos, 212 do sexo masculino/164 do sexo feminino), 153 com pneumonia		1 m	2 categorias: pneumonia e não-pneumonia	18 features acústicos agrupados em <i>prosodic</i> , espectrais, <i>cepstrais</i> , operadores de energia Teager e informações temporais	k-NN, SVM	Exatidão, especificidade e sensibilidade para k-NN e SVM, respetivamente: 90,68% e 91,98%; 89,88% e 90,68%; 90,22% e 92,06%
[Bokov et al., 2016]	186 gravações de crianças de 1 dia-12 anos com idade média de 20 meses	boca	30 s	1 categoria: sibilâncias	Espectro de potência, frequência fundamental e harmónicos	SVM	Sensibilidade=71,4% e especificidade=88,9%

## A.2 Adultos

**Tabela A.2:** Resumo dos métodos de extração de *features* e classificação para SRs em adultos presentes na literatura.

Referências	População	Local de aquisição	Duração das gravações	Sons respiratórios	Features	Algoritmos de machine learning	Resultados
[Pesu et al., 1996]	9 pacientes com doenças pulmonares e 4 indivíduos saudáveis	Tórax	30 a 40 s	4 categorias: sibilâncias, fervores <i>fine</i> , fervores <i>coarse</i> , sons normais	Pacote <i>wavelet</i>	LVQ	Os fervores <i>fine</i> e as sibilâncias apresentaram melhor desempenho relativamente aos fervores <i>coarse</i>
[Kahya et al., 1997]	28 pacientes DPOC, 23 pacientes com doença pulmonar restritiva e 18 indivíduos saudáveis			2 categorias: sons normais e fervores	Modelo AR	k-NN	Exatidão=69,59 %
[Waitman et al., 2000]	17 sujeitos (6 do sexo feminino/11 sexo masculino) com idades entre 19-75 anos	Tórax anterior		2 categorias: sons normais e anormais	Espectros de potência	Diferentes redes neuronais de retropropagação	Exatidão de 73% e uma sensibilidade e especificidade de 87% e 95% respetivamente
[Baydar et al., 2003]	9 indivíduos saudáveis e 11 patológicos		13 s		Método de coerência de sinais		
[Bahoura and Pelletier, 2003]				2 categorias: sibilâncias e sons normais	MFCC	LVQ	Exatidão de 75,80% e 77,50% para sibilâncias e sons normais, respetivamente

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Kandaswamy et al., 2004]	-	Sujeitos em estado relaxado e deitados em posição supina	-	6 categorias: sons normais, sibilâncias, ferveores, <i>squawks</i> , estridor e roncos	Coeficientes da WT: i) Média dos valores absolutos dos coeficientes em cada sub-camada, ii) Potência média dos coeficientes e <i>wavelets</i> em cada sub-camada, iii) Desvio padrão dos coeficientes em cada sub-camada, iv) Razão dos valores médios absolutos das sub-camadas adjacentes	4 algoritmos ANN de retropropagação: GDA, RP, SCG, LM	Exatidão=81%
[Bahoura and Pelletier, 2004]	-	-	-	2 categorias: sons normais e sibilâncias	Duas abordagens: 24 MFCCs e 24 coeficientes de parâmetros <i>subband based cepstral</i>	GMM	Sensibilidade=70% e especificidade=60%
[Güler et al., 2005]	96 indivíduos: 56 doentes pulmonares e os 40 saudáveis	-	15 a 20 s	3 categorias: sons normais, sibilâncias e ferveores	Método <i>Welch</i> e cálculos de estimativa da PSD	Algoritmo genético híbrido e MLP	Exatidões de 81-91% e 83-93% usando só MLP e MLP com algoritmo genético, respetivamente
[Martinez-Hernandez et al., 2006]	10 indivíduos saudáveis e 19 pacientes	Costas dos indivíduos	-	2 categorias: sons normais e anormais	Modelo multivariado AR, redução da dimensionalidade do vetor de <i>features</i> por <i>singular value decomposition</i> e PCA	Rede neuronal supervisionada com retropropagação baseado na regra de <i>Levenberg-Marquardt</i>	Exatidão de classificação de 87,68%, sensibilidade e especificidade de 81,36% e 83,64% respetivamente
[Chien et al., 2007]	-	-	-	2 categorias: sons normais e sibilâncias	MFCCs	GMM, critério <i>Maximum Likelihood</i>	Exatidão=90%

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Taplidou and Hadjileontiadis, 2007]	13 pacientes com diagnóstico de asma, DPOC e pneumonia	Traqueia, axilas direita e esquerda e posterior direita e esquerda bases dos pulmões	5 m	1 categoria: sibilâncias	Parâmetros da STFT		Sensibilidade=95,5± 4,8% e especificidade=93,7± 9,3%
[Lu and Bahoura, 2008]				2 categorias: ferveores <i>fine/coarse</i>	TDW, LDW, PF e GBW	GMM	Exatidão=91,5% usando GBW, PF e LDW, com uma sensibilidade e especificidade de 95,6% e 63,3%, respetivamente
[Alsmadi and Kahya, 2008]	Offline: 20 pacientes (fibrose pulmonar, pneumonia, edema pulmonar, asma, bronquite e enfisema), 36,7±11,9 anos e 20 indivíduos do sexo masculino não fumadores: 34,4±7,4 anos; Tempo real: 25 indivíduos diferentes: 13 com sons patológicos (6 sexo feminino/7 sexo masculino) com doenças respiratórias restritivas ou obstrutivas e 12 com SRs saudáveis, 40,1 ± 16,4 anos; Online: 12 indivíduos saudáveis não fumadores (10 sexo masculino/2 sexo feminino) com 35,8 ± 11,3.	Base posterior direita e esquerda dos pulmões		2 categorias: saudável e patológica	AR	Distância <i>mahalanobis</i> ou k-NN (com k = 5) com as medidas de distância <i>itakura</i> , euclidiana e <i>city-block</i>	Melhor online: exatidão de 96%, sensibilidade de 92% e especificidade de 100% para o classificador de k-NN usando a distância métrica <i>city-block</i> ; Euclidiana e <i>itakura</i> ambas com 88% de exatidão; 100% e 92% de sensibilidade e 83% e 75% de especificidade, respetivamente
[Charleston-Villalobos et al., 2011]	8 indivíduos saudáveis não fumadores (6 sexo feminino/2 sexo masculino) e 19 indivíduos patológicos (13 sexo feminino/6 sexo masculino) com doenças pulmonares intersticiais	Superfície torácica posterior	15 s	2 categorias: sons em normais e anormais	PSD, valores próprios da matriz de covariância e modelos AR univariados e AR multivariados	Rede neuronal supervisionada	Exatidão de 75% e 93% para indivíduos saudáveis e doentes, e valores de sensibilidade 80-100%, especificidade 73-100%, respetivamente

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Serbes et al., 2011]	26 indivíduos (13 saudáveis, 13 patológicos)		1 categoria: ferveores	TF e tempo-escala	SVM	Exatidão=97,20%
[Hashemi et al., 2011]	140 indivíduos diferentes com DPOC e asma		2 categorias: sibilâncias polifônicas e monofônicas	6 <i>features</i> dos coeficientes <i>wavelet</i> : $\mu_{di}$ , $P_{di}$ , $\sigma_{di}$ , $\frac{\mu_{di}}{\mu_{di+1}}$ , $sk_{di}$ e $ku_{di}$	MLP	Exatidão=89,28%
[Abbasi et al., 2013]			6 categorias: sons normais, sibilâncias, roncos, ferveores, estridor e <i>squawks</i>	i) Média dos coeficientes em qualquer sub-camada, ii) Média da potência dos coeficientes de <i>wavelet</i> em qualquer sub-camada, iii) Desvio padrão dos coeficientes em qualquer sub-camada, iv) Proporção dos valores médios nas sub-camadas	Redes neuronais probabilísticas e algoritmo SVM	Exatidão=93,51-100% (melhor resultado com SVM)
[Morillo et al., 2013]	53 pacientes DPCO		3 categorias: sons normais, sibilâncias e roncos	Parâmetros da frequência (média e mediana), parâmetros espectrais ( <i>spectral crest factor</i> ), entropias ( <i>Shannon</i> , <i>Rényi</i> e <i>Tsallis</i> ), <i>relative power</i> para 8 gamas de frequência (50-200Hz, 200-400Hz, 400-800Hz e 800-2000Hz), momento de 2ª ordem, assimetria e <i>spectral kurtosis</i>		Exatidão=77,6%
[Mondal et al., 2014]	10 indivíduos saudáveis e 20 pacientes (DPOC, doenças pulmonares intersticiais e asma)	Sulco anterior supraesternal	3 categorias: sibilância, ferveores e <i>squawks</i>	Lacunaridade, entropia da amostra, assimetria, <i>kurtosis</i>	Rede ELM e SVM	Exatidão de 92,86% e sensibilidade de 86,30% e especificidade de 86,90% (rede ELM tem melhor resultado do que SVM)

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[İçer and Gengeç, 2014]	20 indivíduos saudáveis e 40 indivíduos como DPOC, asma, fibrose pulmonar, pneumonia, tuberculose, doença pulmonar intersticial, bronquiectasias, cancro do pulmão e fungo pulmonar	6 zonas pulmonares: na parte de trás dos pacientes, incluindo a parte direita e esquerda superior, direita e esquerda média e direita e esquerda inferiores.	11 s	2 categorias: fervores e roncós	i) Rácio da frequência PSD baseada no método de <i>Welch</i> ; ii) Frequência instantânea média e o tempo de troca da frequência instantânea iii) Valores próprios da análise de espectro singular	Exatidão entre 90% e 100%	
[Chen et al., 2015]	20 sons pulmonares			3 categorias: sibilâncias, fervores, roncós	MFCCs	k-NN	Identificação do som pulmonar normal : 95%, sons pulmonares anormais: 91,3%.
[Pinho et al., 2015]	10 pacientes de 24 gravações (15 ficheiros áudio de 6 pacientes com pneumonia, $46 \pm 14,6$ anos, 9 ficheiros áudio de 4 pacientes com fibrose cística, $32,3 \pm 18,1$ anos)		10 s	1 categoria: fervores	FD e técnicas de filtragem de <i>boxes</i>		Sensibilidade = 89%, valor preditivo positivo=95% e desempenho geral (índice F = 92%)
[Mendes et al., 2015]	12 indivíduos (9 pacientes e 3 saudáveis)	Anterior direito e esquerdo, posterior direito e esquerdo, lateral direito e esquerdo	30 s	2 categorias: sibilâncias e fervores	30 <i>features</i> : 29 <i>features</i> musicais e 1 <i>feature</i> WS-SS	LR e RF	Melhores resultados com RF: sensibilidade, especificidade e exatidão: $90,9 \pm 2\%$ , $99,4 \pm 1\%$ e $97,9 \pm 1\%$ respetivamente

A. Tabela de resumo dos métodos de extração de *features* e classificação dos sons respiratórios para crianças e adultos.

[Mendes et al., 2016]	17 pacientes e 3 indivíduos saudáveis	Anterior direito e esquerdo, posterior direito e esquerdo, lateral direito e esquerdo	30 s	2 categorias: fervores e sibilâncias	35 <i>features</i> : 31 <i>features</i> musicais, 1 <i>feature</i> baseada em <i>wavelets</i> , energia de Teager, entropia e WS-SS	LR	Exatidão=77% e sensibilidade=76%
[Aykanat et al., 2017]	11 gravações áudio de 1630 indivíduos saudáveis e doentes				MFCCs, imagens de espectrograma usando STFT	SVM e CNN	Exatidão: CNN 86%, SVM 86%
[Rocha et al., 2018]	18 pacientes DPOC	Anterior direito e esquerdo, posterior direito e esquerdo, lateral direito e esquerdo		4 categorias: tosse, sibilância, fervores, fala	35 <i>features</i> musicais e outras 7 <i>features</i> : o atraso do grupo <i>chirp</i> , HNR, o máximo da entropia e o máximo da energia de Teager, KFD do filtro WPST-NST e WS-SS	RF	Especificidade=90%, sensibilidade=69%, <i>precision</i> =70% e <i>f-measure</i> =69%

# B

## Tabelas dos testes de significância estatística para $\alpha=0,05$

### B.1 Amostra 0-2 anos

**Tabela B.1:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.1, para a amostra 0-2 anos.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=0,9691	h=1; p=7,0682 $\times 10^{-7}$	h=0; p= 0,6401
j48	–	–	h=1; p=3,7741 $\times 10^{-7}$	h=0; p= 0,5667
Naive Bayes	–	–	–	h=1; p=6,6343 $\times 10^{-6}$
SMO	–	–	–	–

## B.2 Amostra 3-6 anos

**Tabela B.2:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.3, para a amostra 3-6 anos.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=0,3671	h=1; p= 0,0323	h=1; p= 0,0148
j48	–	–	h=1; p=0,0129	h=1; p= 0,0063
Naive Bayes	–	–	–	h=0; p=1
SMO	–	–	–	–

## B.3 Amostra todos os sons de crianças

**Tabela B.3:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.5, para a amostra todos os sons de crianças.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=0,1669	h=1; p=3,8271 $\times 10^{-27}$	h=0; p=0,5907
j48	–	–	h=1; p=2,8552 $\times 10^{-23}$	h=0; p=0,0640
Naive Bayes	–	–	–	h=1; p=4,4281 $\times 10^{-27}$
SMO	–	–	–	–

## B.4 Amostra apenas fervores (crianças)

**Tabela B.4:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.7, para a amostra que contém apenas fervores.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=1	h=1; p=8,0448 $\times 10^{-13}$	h=0; p=0,8415
j48	–	–	h=1; p=1,2053 $\times 10^{-13}$	h=0; p=0,8348
Naive Bayes	–	–	–	h=1; p=1,3358 $\times 10^{-12}$
SMO	–	–	–	–

## B.5 Amostra apenas sibilâncias (crianças)

**Tabela B.5:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na tabela 5.9, para a amostra que contém apenas sibilâncias.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=0; p=1	h=1; p= 0,0499	h=0; p= 0,8084
j48	–	–	h=1; p=0,0330	h=0; p=0,8084
Naive Bayes	–	–	–	h=1; p=0,0116
SMO	–	–	–	–

## B.6 Adultos

**Tabela B.6:** Testes de significância estatística com  $\alpha=0,05$ , entre os 4 classificadores para os resultados obtidos na Tabela 5.25, para a população adulta.

	iBk	j48	Naive Bayes	SMO
iBk	–	h=1; p= $1,4731 \times 10^{-13}$	h=1; p= $7,4789 \times 10^{-126}$	h=1; p= $4,2125 \times 10^{-10}$
j48	–	–	h=1; p= $2,6640 \times 10^{-191}$	h=1; p=0,0149
Naive Bayes	–	–	–	h=1; p= $9,3894 \times 10^{-187}$
SMO	–	–	–	–

# C

## Tabelas seleção de *Features*

### C.1 Amostra 0-2 anos

**Tabela C.1:** *Features* selecionadas pelo algoritmo *CfsSubsetEval + Best First* para a amostra 0-2 anos.

Features		
median_roughness	max_hcdf	std_mfcc9
median_mfcc2	max_inharmonicity	std_mfcc11
median_mfcc10	max_hnr500	std_chromacentroid
median_keyclarity	max_teager	std_cgd
median_mode	std_skewness	std_hnr500
median_hnr500	std_flatness	std_hnr1500
max_rms	std_irregularity	min_mfcc3
max_rolloff95	std_mfcc1	min_mfcc11
max_irregularity	std_mfcc4	min_fracdim
max_spectflux	std_mfcc5	mean_spectentropy
max_keyclarity	std_mfcc6	mean_mfcc7
max_mode	std_mfcc8	mean_mfcc11

**Tabela C.2:** *Features* selecionadas pelo algoritmo *WrapperSubsetEval* + *Best First* para a amostra 0-2 anos.

Features		
median_mfcc10	max_F0_cepstrum	min_zerocross
median_mfcc12	min_spectentropy	min_fracdim
median_mode	min_roughness	mean_F0_spectrum
max_rms	min_irregularity	
max_mfcc8	min_mfcc10	

1

**Tabela C.3:** 50 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval* + *Ranker* para a amostra 0-2 anos.

Features		
median_F0_spectrum_spectrum	min_brightness	std_rolloff85
mean_F0_spectrum_spectrum	min_zerocross	mean_centroid
min_flatness	mean_rolloff95	mean_zerocross
min_spread	median_irregularity	median_spectentropy
min_rolloff95	min_keyclarity	std_zerocross
median_skewness	median_spread	std_F0_cepstrum
max_F0_spectrum	median_entropy	median_zerocross
min_F0_spectrum	min_chromacentroid	min_mfcc10
min_centroid	median_rolloff85	min_mfcc2
min_rolloff85	mean_spectflux	median_spectflux
min_spectentropy	mean_spread	min_mfcc7
std_F0_spectrum	min_mode	std_spread
median_kurtosis	min_mfcc5	min_F0_cepstrum

<sup>1</sup>Note que a leitura das tabelas resultantes do método *ReliefFAttributeEval* + *Ranker* deve ser feita por coluna, uma vez que a ordem segue o *ranking* das *features*.

min_hcdf	mean_rolloff85	min_mfcc11
median_rolloff95	median_centroid	min_inharmonicity
min_roughness	median_flatness	min_mfcc3
median_mfcc5	std_centroid	

## C.2 Amostra 3-6 anos

**Tabela C.4:** *Features* selecionadas pelo algoritmo *CfsSubsetEval + Best First* para a amostra 3-6 anos.

Features		
median_brightness	std_spectentropy	min_mfcc5
median_rolloff95	std_mfcc2	min_mfcc6
median_irregularity	std_mfcc7	min_mfcc13
median_zerocross	std_mfcc9	min_zerocross
median_chromacentroid	std_mfcc12	min_hcdf
max_rms	std_chromacentroid,	min_fracdim
max_irregularity	std_keyclarity	min_teager
max_mfcc7	std_mode	mean_mfcc5
max_mfcc9	std_F0_cepstrum	mean_mfcc9
max_mode	std_hnr1500	mean_zerocross
max_hcdf	std_wsss	mean_F0_cepstrum
max_inharmonicity	std_entropy	

**Tabela C.5:** *Features* selecionadas pelo algoritmo *WrapperSubsetEval + Best First* para a amostra 3-6 anos.

Features		
median_rolloff95	median_chromapeak	min_spectentropy

median_mfcc5	max_rolloff95	min_hnr500
median_mfcc8	max_mfcc7	mean_kurtosis
median_zerocross	std_hnr500	mean_hnr1500

**Tabela C.6:** 80 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval* + *Ranker* para a amostra 3-6 anos.

Features		
median_F0_cepstrum	max_mfcc12	std_spread
mean_F0_cepstrum	std_F0_cepstrum	min_mfcc1
max_mfcc9	min_rms	median_rolloff95
min_mfcc6	min_inharmonicity	std_zerocross
median_inharmonicity	min_F0_cepstrum	median_rolloff85
max_mfcc7	min_mfcc8	max_inharmonicity
median_mfcc6	mean_mfcc5	min_brightness
mean_mfcc9	mean_rms	std_flatness
mean_mfcc6	median_rms	median_brightness
mean_mfcc7	median_zerocross	mean_rolloff95
median_mfcc9	max_hcdf	std_mfcc7
median_mfcc7	min_mfcc9	std_mfcc1
mean_inharmonicity	min_mfcc11	max_mfcc5
max_irregularity	min_mfcc7	median_flatness
min_flatness	max_mfcc1	mean_flatness
min_mfcc13	std_inharmonicity	median_centroid
max_chromacentroid	min_mfcc5	std_keyclarity
max_chromapeak	max_zerocross	mean_brightness
median_mfcc1	mean_zerocross	median_spectentropy
min_spread	max_mfcc6	min_F0_spectrum

std_rms	max_wsss	std_spectflux
max_mode	std_chromacentroid	std_spectentropy
max_rms	median_spectflux	mean_centroid
mean_mfcc1	max_cgd	median_mfcc3
std_irregularity	mean_spectentropy	max_spectflux
min_zerocross	mean_spectflux	mean_chromacentroid
median_mfcc5	max_mfcc10	

### C.3 Amostra todos os sons crianças

**Tabela C.7:** *Features* selecionadas pelo algoritmo *CfsSubsetEval + Best First* para a amostra todos os sons de crianças.

Features		
median_rms	max_mode	std_spectflux
median_spectentropy	max_hcdf	std_chromacentroid
median_mfcc7	max_F0_spectrum	std_hnr500
median_zerocross	max_hnr500	std_hnr1500
median_spectflux	max_wsss	min_mfcc8
median_F0_spectrum	std_rms	min_mfcc11
median_cgd	std_roughness	min_mfcc13
max_rolloff95	std_irregularity	min_spectflux
max_irregularity	std_mfcc1	min_chromapeak
max_mfcc5	std_mfcc3	min_F0_cepstrum
max_mfcc10	std_mfcc5	min_fracdim
max_mfcc12	std_mfcc6	min_teager
max_spectflux	std_mfcc8	mean_mfcc12
max_chromacentroid	std_mfcc10	mean_F0_spectrum

max_keyclarity	std_mfcc13	mean_cgd
----------------	------------	----------

**Tabela C.8:** *Features* selecionadas pelo algoritmo *WrapperSubsetEval + Best First* para a amostra todos os sons de crianças.

Features		
median_skewness	median_F0_spectrum	std_spectflux
median_roughness	median_teager	min_flatness
median_mfcc1	max_roughness	min_roughness
median_mfcc6	max_mfcc2	min_mfcc13
median_mfcc9	max_zerocross	mean_spread
median_zerocross	max_chromacentroid	mean_mfcc10
median_spectflux	max_F0_spectrum	
median_mode	std_mfcc1	

**Tabela C.9:** 70 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval + Ranker* para a amostra todos os sons de crianças.

Features		
median_wsss	max_F0_cepstrum	median_F0_spectrum
min_wsss	min_keyclarity	median_spectentropy
median_entropy	mean_chromacentroid	min_centroid
min_chromacentroid	median_skewness	min_mfcc7
median_F0_cepstrum	min_rolloff85	max_rolloff95
median_rolloff95	mean_zerocross	min_mfcc2
median_irregularity	median_rms	std_rolloff95
median_keyclarity	mean_rolloff85	median_cgd
median_inharmonicity	mean_spread	std_roughness
median_spectflux	min_cgd	min_zerocross

median_chromacentroid	min_inharmonicity	median_mfcc5
median_flatness	max_rolloff85	min_brightness
mean_rolloff95	median_zerocross	max_F0_spectrum
min_rms	min_F0_cepstrum	mean_F0_spectrum
min_flatness	min_hcdf	min_mfcc3
mean_spectflux	min_spectentropy	std_spread
min_rolloff95	std_F0_cepstrum	max_roughness
min_entropy	median_kurtosis	min_chromapeak
median_spread	median_centroid	min_fracdim
min_spread	mean_centroid	min_mfcc5
mean_F0_cepstrum	median_mfcc7	min_mfcc11
median_hcdf	max_zerocross	min_skewness
median_rolloff85	std_rolloff85	
min_spectflux	min_mfcc10	

## C.4 Amostra apenas fervores (crianças)

**Tabela C.10:** *Features* selecionadas pelo algoritmo *CfsSubsetEval + Best First* para a amostra que contém apenas fervores.

Features		
median_rms	max_keyclarity	min_mfcc8
median_centroid	max_hcdf	min_mfcc10
median_F0_spectrum	std_mfcc8	min_spectflux
median_hnr500	std_mfcc13	mean_rolloff85
max_spread	std_zerocross	mean_mfcc5
max_rolloff95	std_spectflux	mean_spectflux
max_mfcc12	std_keyclarity	mean_F0_spectrum

max\_chromacentroid    min\_mfcc5

---

**Tabela C.11:** *Features* selecionadas pelo algoritmo *WrapperSubsetEval* + *Best First* para a amostra que contém apenas ferveores.

Features		
median_mode	min_skewness	min_mfcc10
max_mfcc1	min_mfcc5	mean_hcdf
max_hnr1500	min_mfcc7	

**Tabela C.12:** 50 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval* + *Ranker* para a amostra que contém apenas ferveores.

Features		
median_chromacentroid	min_spectflux	std_F0_spectrum
median_keyclarity	std_rolloff85	mean_rolloff95
mean_chromacentroid	max_centroid	mean_chromapeak
min_rms	min_F0_cepstrum	min_chromacentroid
max_rolloff85	median_spectflux	mean_F0_spectrum
min_entropy	median_spread	max_zerocross
min_inharmonicity	mean_spread	mean_zerocross
median_entropy	max_rolloff95	min_mfcc10
median_hcdf	mean_spectflux	median_zerocross
median_rolloff85	min_skewness	min_rolloff95
min_cgd	median_chromapeak	median_mfcc5
mean_rolloff85	mean_centroid	min_mfcc5
median_spectentropy	min_F0_spectrum	min_kurtosis
median_F0_spectrum	median_centroid	min_rolloff85
median_flatness	median_kurtosis	max_spread

median_skewness	median_rms	min_flatness
median_rolloff95	min_chromapeak	

## C.5 Amostra apenas sibilâncias (crianças)

**Tabela C.13:** *Features* selecionadas pelo algoritmo *CfsSubsetEval+ Best First* para a amostra que contém apenas sibilâncias.

Features		
median_spread	max_mfcc12	std_inharmonicity
median_mfcc7	max_spectflux	std_hnr500
median_spectflux	max_inharmonicity	std_wsss
median_mode	std_rms	min_hcdf
max_kurtosis	std_spectentropy	mean_irregularity
max_roughness	std_irregularity	mean_mfcc2
max_irregularity	std_mfcc4	mean_mfcc11
max_mfcc9	std_mode	
max_mfcc11	std_F0_cepstrum	

**Tabela C.14:** *Features* selecionadas pelo algoritmo *WrapperSubsetEval + Best First* para a amostra que contém apenas sibilâncias.

Features		
median_mfcc5,	max_irregularity	min_mfcc13
median_mfcc6	max_inharmonicity	

**Tabela C.15:** 60 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval* + *Ranker* para a amostra que contém apenas sibilâncias.

Features		
min_F0_cepstrum	min_rolloff95	mean_skewness
min_inharmonicity	median_mfcc1	median_rolloff95
max_irregularity	mean_rolloff85	median_skewness
std_F0_cepstrum	median_cgd	min_spread
std_inharmonicity	median_F0_cepstrum	mean_kurtosis
max_F0_spectrum	mean_irregularity	median_mfcc11
std_irregularity	mean_rolloff95	mean_mfcc3
max_mfcc9	min_flatness	median_mfcc3
mean_F0_cepstrum	max_cgd	max_mfcc1
std_F0_spectrum	mean_mfcc6	max_inharmonicity
mean_F0_spectrum	min_keyclarity	max_mfcc11
min_chromacentroid	min_mfcc3	median_flatness
median_F0_spectrum	std_chromacentroid	median_kurtosis
mean_inharmonicity	mean_cgd	min_mfcc1
mean_mfcc7	max_skewness	max_hcdf
median_mfcc7	max_kurtosis	median_inharmonicity
mean_mfcc9	median_mfcc6	max_rms
median_mfcc9	min_rolloff85	min_mfcc6
max_mfcc7	median_irregularity	median_brightness
mean_mfcc1	median_rolloff85	mean_flatness

## C.6 Adultos

**Tabela C.16:** *Features* selecionadas pelo algoritmo *CfsSubsetEval + Best First* para a população adulta.

Features		
median_spectflux	std_mfcc6	std_hnr500
max_irregularity	std_mfcc7	std_hnr1500
max_mfcc10	std_mfcc8	std_entropy
max_mfcc12	std_mfcc9	min_irregularity
max_chromacentroid	std_mfcc10	min_mfcc7
max_inharmonicity	std_mfcc11	min_mfcc8
max_wsss	std_mfcc12	min_mfcc10
std_rms	std_mfcc13	min_mfcc11
std_rolloff85	std_zerocross	min_mfcc12
std_flatness	std_chromacentroid	min_mfcc13
std_mfcc1	std_chromapeak	min_spectflux
std_mfcc2	std_mode	min_hcdf
std_mfcc3	std_hcdf	mean_chromapeak
std_mfcc4	std_inharmonicity	mean_F0_spectrum
std_mfcc5	std_F0_spectrum	mean_wsss

**Tabela C.17:** 185 primeiros atributos selecionados pelo algoritmo *ReliefFAttributeEval + Ranker* para a população adulta.

Features		
max_F0_cepstrum	min_centroid	mean_mfcc3
median_inharmonicity	median_kurtosis	mean_hcdf
median_wsss	min_mfcc3	max_chromapeak
min_inharmonicity	mean_kurtosis	mean_wsss

---

min_mfcc10	mean_rolloff85	mean_entropy
std_F0_cepstrum	mean_zerocross	mean_mfcc7
min_hcdf	mean_spectflux	std_chromapeak
min_rms	std_centroid	mean_irregularity
min_entropy	min_mfcc4	mean_mfcc6
mean_F0_cepstrum	median_spectentropy	mean_inharmonicity
median_F0_cepstrum	median_zerocross	mean_mfcc5
min_mfcc5	median_centroid	mean_mfcc8
mean_chromacentroid	min_F0_spectrum	mean_mfcc9
min_mfcc7	mean_centroid	std_roughness
min_mfcc12	min_mfcc6	mean_mfcc13
min_keyclarity	median_mfcc8	mean_roughness
median_entropy	max_spread	mean_mfcc11
min_skewness	mean_F0_spectrum	mean_mfcc12
min_flatness	median_mfcc3	mean_keyclarity
min_spread	median_mfcc4	std_hcdf
min_rolloff95	median_mfcc2	std_mfcc1
min_spectentropy	min_zerocross	max_mfcc1
median_irregularity	median_mfcc10	mean_rms
median_chromacentroid	median_mfcc13	max_hcdf
min_cgd	median_F0_spectrum	mean_mfcc10
min_F0_cepstrum	min_brightness	mean_flatness
max_rolloff85	median_mfcc12	std_skewness
median_hcdf	max_spectflux	std_inharmonicity
min_irregularity	median_mfcc6	mean_spectentropy
max_zerocross	min_mfcc1	max_skewness
max_rolloff95	median_teager	mean_fracdim
median_keyclarity	median_mfcc1	std_keyclarity

---

min_wsss	mean_chromapeak	std_entropy
min_mode	median_cgd	max_irregularity
median_spread	min_teager	std_mfcc2
std_zerocross	min_chromapeak	mean_teager
median_skewness	median_chromapeak	std_mfcc4
min_spectflux	median_mfcc11	max_wsss
min_mfcc13	max_hnr1500	std_irregularity
median_rolloff95	min_hnr1500	mean_cgd
median_flatness	median_mfcc9	std_wsss
min_chromacentroid	max_chromacentroid	max_mfcc2
std_rolloff85	max_hnr500	max_mfcc3
mean_spread	std_hnr500	max_entropy
min_mfcc8	min_hnr500	max_mfcc4
max_F0_spectrum	mean_mfcc1	std_mfcc3
max_centroid	std_hnr1500	max_mfcc5
median_rms	median_brightness	std_mfcc7
mean_rolloff95	std_spectflux	max_mfcc6
median_mfcc5	mean_hnr1500	max_keyclarity
min_rolloff85	median_hnr1500	std_mfcc6
median_mfcc7	min_fracdim	std_mfcc11
std_rolloff95	median_fracdim	mean_brightness
min_mfcc2	median_hnr500	max_mfcc10
min_mfcc9	mean_hnr500	mean_mode
min_kurtosis	mean_skewness	max_inharmonicity
std_F0_spectrum	max_kurtosis	max_mfcc7
std_chromacentroid	std_kurtosis	std_mfcc10
median_spectflux	median_mode	std_mfcc9
min_mfcc11	mean_mfcc2	std_mfcc12

### C. Tabelas seleção de *Features*

---

std_spread	mean_mfcc4	max_mfcc13
median_rolloff85	max_roughness	

---

---

# Anexos



# A

## Recomendações CORSA para a aquisição de sons respiratórios

**Tabela A.1:** Recomendações CORSA para a aquisição de sons respiratórios [Vannuccini et al., 2000].

Parâmetros	Valores
Intervalo da frequência	100–5 kHz
Função de resposta em frequência	plano na banda de áudio das frequências em estudo; irregularidade <6 dB
Intervalo dinâmico	> 60 dB
Sensibilidade	Não deve depender da frequência dos sinais, direção dos sons ou pressão aplicada no sensor
Relação sinal-ruído	> 60 dB
Direção	Omnidirecional
Sensor	microfone condensador; microfone piezoelétrico, acelerômetro
Proteção contra ruído e interferência	Sons acústicos: proteção contra vibração, revestindo com preenchimento absorvente do som; Interferência eletromagnética: cabo torcido ou cabo de microfone blindado
Localização dos microfones	Traqueia ou tórax (posterior, anterior, lateral: esquerda e direita)



## B

# *Features* propostas por Mendes et al. e Rocha et al.

**Tabela B.1:** *Features* propostas por [Mendes et al., 2015, Mendes et al., 2016, Rocha et al., 2018].

<i>Feature</i>	Descrição
RMS	Raíz quadrada da média da energia quadrática da frame
<i>Spectral Centroid</i>	Centro de massa (centro geométrico) da distribuição espectral
<i>Spectral Brightness</i>	Quantidade de energia acima de 1500 Hz
<i>Spectral Spread</i>	Variação da distribuição espectral
<i>Spectral Skewness</i>	Assimetria da distribuição espectral
<i>Spectral Kurtosis</i>	Excesso de <i>kurtosis</i> da distribuição espectral
<i>Spectral Rolloff</i> 95	Frequência tal que 95% da energia total está contida abaixo dessa frequência
<i>Spectral Rolloff</i> 85	Frequência tal que 85% da energia total está contida abaixo dessa frequência
<i>Spectral Entropy</i>	Complexidade do espectro
<i>Spectral Flatness</i>	Ruído do espectro
<i>Spectral Roughness</i>	Estimativa da dissonância sensorial
<i>Spectral Irregularity</i>	Grau de variação dos picos sucessivos do espectro
MFCC	13 coeficientes cepstrais da frequência de mel

B. *Features* propostas por Mendes et al. e Rocha et al.

---

<i>Zero-crossing Rate</i>	Taxa de alteração da forma de onda do sinal (monotonia)
<i>Spectral Flux</i>	Distância euclidiana entre frames sucessivos no espectro
<i>Chromagram Centroid</i>	Centróide tonal
<i>Chromagram Peak</i>	Pico do centróide tonal
<i>Key Clarity</i>	Probabilidade dos principais candidatos
<i>Mode</i>	Estimativa de modalidade
<i>Harmonic Change Detection Function</i> (HCDF)	Fluxo do centróide tonal
<i>Pitch</i>	Estimativa da frequência fundamental do som ( <i>cepstrum</i> )
<i>Pitch Inharmonicity</i>	Proporção de parciais que não são múltiplos da frequência fundamental, considerando a quantidade de energia fora da série harmónica ideal
F0	Estimativa da frequência fundamental ( <i>spectrum</i> )

---

# C

## Resultados de desempenho do método proposto por Rocha et al.

**Tabela C.1:** Resultados de desempenho do método proposto por [Rocha et al., 2018].

Data set	Specificity	Sensitivity	Precision	F-measure
Complete	0.90±0.01	0.67±0.03	0.67±0.03	0.67±0.03
Merged	0.90±0.01	0.69±0.03	0.70±0.03	0.69±0.03
Complete 50	0.90±0.01	0.67±0.03	0.67±0.04	0.67±0.03
Merged 50	0.90±0.01	0.69±0.03	0.70±0.03	0.69±0.03
Complete 20	0.90±0.01	0.65±0.04	0.65±0.04	0.65±0.04
Merged 20	0.89±0.01	0.67±0.03	0.68±0.03	0.67±0.03
Complete 10	0.90±0.01	0.62±0.03	0.63±0.03	0.62±0.03
Merged 10	0.88±0.01	0.65±0.03	0.65±0.03	0.65±0.03
Complete 5	0.88±0.01	0.57±0.03	0.58±0.03	0.57±0.03
Merged 5	0.87±0.01	0.61±0.03	0.62±0.03	0.61±0.03