INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Utilização de Ferramentas de Machine Learning no Diagnóstico de Patologias da Laringe

Felipe Lage Teixeira

Dissertação apresentada à

Escola Superior de Tecnologia e Gestão Instituto Politécnico de Bragança

para obtenção do grau de Mestre em

Engenharia Industrial – Ramo Engenharia Eletrotécnica

Trabalho realizado sob a orientação de

Professor Doutor João Paulo Teixeira

Outubro de 2019

Agradecimentos

Como não poderia deixar passar em claro, tenho que agradecer aos responsáveis por ter atingido esta fase apesar das dificuldades da vida. Por isso, deixo uma palavra de carinho aos meus pais, que sempre lutaram para me ajudar a alcançar os meus objetivos, mas que infelizmente não podem assistir à conclusão de mais uma etapa da minha vida, por isso, pai e mãe, onde quer que estejam agradeço por tudo que fizeram por mim.

Agradeço, de especial maneira, ao Professor João Paulo Teixeira pela orientação, conhecimento transmitido e disponibilidade demonstrada ao longo deste percurso.

Agradeço, à minha família, por todo o apoio e motivação que me deram durante esta fase, especialmente aos que sempre me incentivaram, que todos os sacrifícios suportaram para que os meus objetivos fossem alcançados com sucesso.

Aos amigos, que durante o meu percurso académico, sempre me motivaram e se colocaram a disposição para me prestar auxílio em qualquer fase desta etapa, sem vocês seria mais difícil.

A todos, Muito Obrigado.

Resumo

Este trabalho está relacionado com o estudo e utilização de um conjunto de ferramentas de *machine learning*, nomeadamente árvores de decisão, *support vector machines* (SVM's), *Deep-learning - Deep Neural Networks*, com o prepósito de fazer a classificação entre fala patológica e fala normal, e identificar a patologia com estas ferramentas.

As patologias utilizadas neste estudo são a laringite crónica, disfonia e paralisia das cordas vocais.

Utilizou-se a base de dados Alemã *Saarbrucken Voice Database* (SVD), que se encontra disponível *online* de forma gratuita pelo Instituto de Fonética da Universidade de *Saarland*. Nesta base de dados é possível encontrar sinais de voz, entre saudáveis e patológicos de mais de 2000 sujeitos.

Foram utilizados três grupos de parâmetros, o grupo I (a), contêm parâmetros como *Jitter* relativo, *Shimmer* relativo e *Harmonic to Noise Ratio* (HNR), determinados em segmentos de fala estacionária, onde se atingiu 80.7% de exatidão para distinguir saudáveis e patológicos com SVM. O grupo I (b), contêm os parâmetros do grupo I(a), *Noise to Harmonic Ratio* (NHR) e Autocorrelação determinados em segmentos de fala estacionária, onde se atingiu 79.2% de exatidão para distinguir saudáveis e patológicos com SVM. O grupo II é baseado em *Mel Frequency Cepstral Coefficientes* (MFCC's), determinados nos segmentos de fala estacionários, onde se atingiu 83.3% de exatidão para distinguir saudáveis e laringite com SVM. O grupo III é formado por coeficientes MFCC's extraídos de fala contínua onde se atingiu 71% de exatidão para distinguir saudáveis e patológicos com Redes Neuronais.

Realizou-se uma análise estatística referente aos parâmetros do grupo I (b), com o propósito de identificar características únicas em determinados parâmetros, que permitissem diferenciar as patologias.

No decorrer deste trabalho, embora não fosse objetivo inicial, deu-se início a elaboração de um "*software"* protótipo para fazer gravação de voz, extração de parâmetros e classificação da patologia.

Palavras-chave: *Machine Learning*, *Deep Learning*, Patologias.

Abstract

This work is related to the study and use of a set of machine learning tools, namely decision trees, Support Vector Machines (SVM's), Deep learning - Deep Neural Networks (neural networks), with the purpose of classifying speech pathological and normal speech, and to identify the pathology with these tools. The pathologies used in this study are chronic laryngitis, dysphonia and vocal cord paralysis.

We use the database of the German Saarbrucken Voice Database (SVD), which is available online for free at the Institute of Phonetics at the University of Saarland. In this database it is possible to find voice signals between healthy and pathological of more than 2000 subjects.

Three groups of parameters were used, the first one is the group I (a) contains parameters such as Relative Jitter, Relative Shimmer and Harmonic to Noise Ratio (HNR), determined in stationary speech segments, where 80.7% accuracy was achieved to distinguish healthy and pathologies. The group I (b), contain parameters like Relative Jitter, Relative Shimmer, HNR, Noise to Harmonic Ratio (NHR) and Autocorrelation determined in segments of stationary speech, where it obtained 79.2% accuracy to distinguish healthy and pathological patients with SVM. Group II is based on Mel Frequency Cepstral Coefficients (MFCC's), determined in stationary speech segments, where it obtained 83.3% accuracy to distinguish the healthy and laryngitis with SVM. Group III is formed by MFCC coefficients, extracted from continuous speech, where it reached 71% of accuracy to distinguish healthy and pathologies with Neuronal Networks.

The statistical study concerning the parameters of group I (b) was performed, in which three 'a', 'i' and 'u' vowels were analyzed in three differents tones: high, low and normal. The statistical study was performed with the purpose of detecting unique characteristics in certain parameters, which allowed to distinguish the pathologies used in this dissertation.

In the course of this work, although it was not an initial objective, Started the development of prototype software to make voice recording, parameter extraction and classification of the pathology.

Keywords: Machine Learning, Deep Learning, Pathologies.

Resumen

Este trabajo está relacionado con el estudio y utilización de un conjunto de herramientas de machine learning, dígase árboles de decisión, support vector machines, Deep learning- Deep Neural Networks (redes neuronales), con el propósito de hacer la clasificación entre habla patológica y habla normal e identificar la patología con estas herramientas.

Las patologías utilizadas en este estudio son la laringitis crónica, disfonía y parálisis de las cuerdas vocales.

Se ha utilizado la base de datos alemana Saarbrucken Voice Database (SVD), que se encuentra disponible online de forma gratuita por el Instituto de Fonética de la Universidad de Saarland. En esta base de datos es posible encontrar señales de voz, entre saludables y patológicos de más de 2000 sujetos.

Se han analizado tres grupos de parámetros, el grupo I(a) contiene parámetros como Jitter relativo, Shimmer relativo, HNR, determinados en segmentos de habla estacionaria, alcanzaron una precisión del 80.7% para distinguir entre sano y patológico. El grupo I(b), contiene parámetros como Jitter relativo, Shimmer relativo, HNR, NHR y Autocorrelación, determinados en segmentos de habla estacionaria, donde se alcanzó una precisión del 79,2% para distinguir sanos y patológicos con la SVM. O grupo II está basado en coeficientes MFCC's, determinados en segmentos de habla estacionaria, donde se logró una precisión del 83.3% para distinguir los sanos y la laringitis con SVM.

El grupo III está formado por coeficientes MFCC extraídos del habla continua, que alcanzaron el 71% de precisión para distinguir los sanos y patológicas con Redes neuronales.

Se ha realizado el estudio estadístico referente a los parámetros del grupo I(b), cuyas 3 vocales "a", "i" y "u" en tres tonos disponibles alto, bajo y normal fueron analizadas. el estudio estadístico se ha realizado con el propósito de detectar características únicas en determinados parámetros, que permitieran diferenciar las patologías utilizadas en esta disertación.

En el transcurso de este trabajo, aunque no fuera el objetivo inicial, se dió inicio a la elaboración de un "*software*" prototipo para hacer grabación de voz, extracción de parámetros y clasificación de la patología.

Palabras clave: Machine Learning, Deep Learning, Patologias

Conteúdo

Lista de Tabelas

Lista de Figuras

Lista de Abreviaturas

DAG- Grafo direcionado acíclico;

DFT- Transformada Discreta de Fourier;

GPU- Unidade de processamento gráfico;

HNR- Harmonics-to-Noise Ratio;

IAPMEI- Instituto de Apoio às Pequenas e Médias Empresas e ao Investimento;

L. Crónica- Laringite crónica;

LS- Least Squares;

LS-SVM-Máquinas de vetor de suporte por Mínimos Quadrados;

MFCC- Mel-frequency cepstral coefficients;

MLP- MultiLayer Perceptron;

NHR- Noise-to-Harmonics Ratio;

P.C. Vocais- Paralisia das cordas vocais;

PCA-Principal Component Analysis;

QP- Quadratic Programming;

RBF- Radial-Basis Function;

ReLU- unidade linear rectificada;

RN MLP- Rede Neuronal MultiLayer Perceptron;

RNA- Rede Neuronal Artificial;

SMO- Sequential Minimal Optimization;

SVD- Saarbruecken Voice Database;

SVM- Support Vector Machine;

vs.- versus;

WPT- Transformada Wavelet Packet;

1. Introdução

Ao surgir o Ser Humano, nasce com ele a necessidade de se exprimir, quer seja por expressões corporais ou por meio de sons. Estes podem ser provocados por meios físicos, como bater as palmas, ou então por meio oral.

Ao surgir vontade de comunicar oralmente, o cérebro humano transmite impulsos nervosos aos músculos do sistema respiratório, que quando são contraídos comprimem o ar dos pulmões e obrigam-no a subir pela traqueia até à laringe (fazendo com que as cordas vocais ajustem o seu posicionamento e a sua vibração) e estruturas do trato vocal (fazendo ajustar a tensão na faringe, o posicionamento da língua e do palato mole).

Deste modo, a coluna de ar pulmonar é sonorizada na laringe (fonação) e modulada no trato vocal (sons da fala), figura 1.1.

Resumindo, a voz é um som (resultante de um conjunto de acontecimentos no aparelho fonador e ao longo do trato vocal) com uma determinada força, sonoridade, duração, velocidade e ritmo, regulado de forma subconsciente pela informação enviada pelo cérebro via auditiva (Guimarães, 2004).

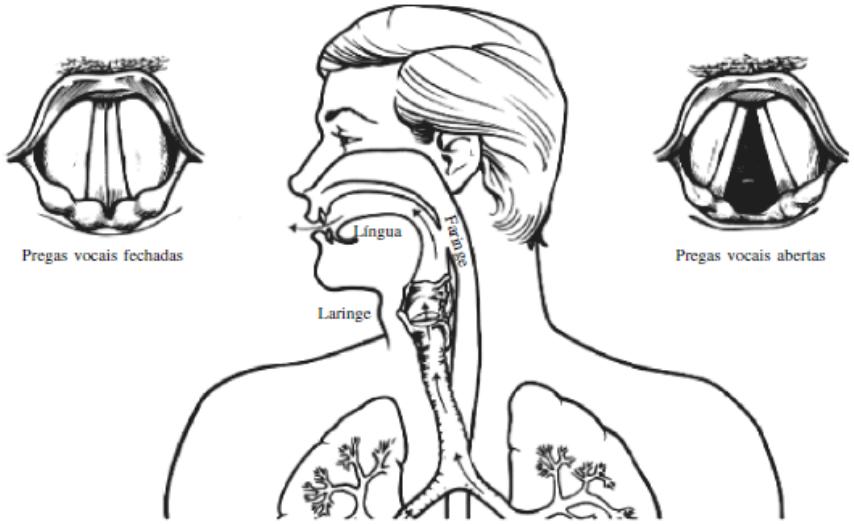

Figura 1.1- Trato vocal e aparelho fonador (Guimarães, 2004)

Cada voz, tal como cada impressão digital, é única, não existem duas vozes iguais, por mais que para o ouvido Humano não seja possível distinguir essas pequenas diferenças, cada voz é constituída por parâmetros distintos.

A voz humana tem, qualidades próprias (Matuck, 2005):

Tom – é a altura musical da voz. Segundo o tom, a voz humana classifica-se em aguda ou grave.

Timbre – é a matriz pessoal da voz. É um fenómeno complexo e está determinado pelo tom fundamental e pelos seus harmónicos. Reconhece-se pelo timbre característico a pessoa com a qual se fala. Há vozes bem timbradas e agradáveis, mas também existem roucas, agudas e chiadas;

Quantidade – é a duração do som. Segundo a quantidade, os sons podem ser longos ou curtos, com toda a gama intermediária de semi-longos, semi-curtos, etc. A quantidade depende, geralmente, das caraterísticas de cada língua, dos costumes linguísticos das regiões ou países.

Intensidade – é a maior ou menor força com que se produz a voz. Há vozes fracas e vozes fortes.

A análise acústica vocal permite quantificar e caracterizar um sinal sonoro. Utilizando a análise acústica, para o estudo da voz, consegue-se de forma não invasiva determinar e quantificar a qualidade vocal do individuo através dos diferentes parâmetros acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral. Constitui assim, um método de avaliação objetiva que permite, entre outras utilidades, detetar precocemente problemas vocais (Teixeira, Ferreira, & Carneiro, Análise Acústica Vocal-Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala, 2011).

Uma perturbação na voz traz implicações profundas na vida social e profissional de uma pessoa. Nos pacientes com patologias progressivas é importante ter acesso a um rápido diagnóstico para ser possível promover um melhor tratamento e prognóstico (Alves N. , 2016).

Na atualidade, são vários os exames que se podem realizar para fazer diagnósticos de distúrbios da voz.

Inicialmente, os diagnósticos eram feitos através da análise preceptiva auditiva, contudo, a falta de entendimento entre examinadores experientes, tornou necessária a pesquisa de uma avaliação objetiva, onde fosse possível analisar a voz com aparelhos capazes de medir parâmetros acústicos.

Existem vários exames que podem ser feitos na deteção de patologias associadas à voz, porém, estes exames são invasivos e tornam-se um pouco desconfortáveis para os pacientes, podendo, até, provocar o vómito, ou, dependem da experiência do médico que faz a avaliação (Teixeira & Fernandes, 2015) (Teixeira, Ferreira, & Carneiro, 2011).

O uso de técnicas computacionais, juntamente, com a análise acústica permite medir propriedades do sinal acústico de uma voz gravada, onde se diz de forma sustentada vogais ou um discurso (Alves N. , 2016).

A análise acústica da voz é uma técnica bastante utilizada na deteção e estudo de patologias da voz. Correlaciona-se, em geral, com o uso de técnicas computacionais que permitem medir propriedades do sinal acústico de uma voz gravada dizendo vogais de forma sustentada ou em discurso (Alves N. , 2016).

A análise acústica é capaz de fornecer o formato da onda sonora permitindo-nos avaliar determinadas características como a frequência fundamental (F0), definida como o número de vibrações por segundo produzidas pelas cordas vocais, as medidas de perturbação da frequência, *o jitter*, definido como sendo a perturbação da frequência fundamental ciclo a ciclo e ainda medidas da perturbação da amplitude, o *shimmer*, que é a variabilidade da amplitude ciclo a ciclo.

Estes parâmetros podem ser medidos através de equipamento específico próprio para o efeito, nomeadamente, *softwares* informáticos, (no caso desta dissertação utilizaram-se algoritmos desenvolvidos no software MATLAB) para extração dos valores referentes aos parâmetros da voz de cada sujeito.

Através de parâmetros existentes na voz Humana é possível distinguir vozes saudáveis, vozes patológicas, vozes femininas, vozes masculinas, entre outras.

Quando existe um distúrbio derivado de lesões nas cordas vocais, o processo de fonação é alterado, isto porque, os padrões de vibração durante a fase de abertura e fecho das cordas vocais são irregulares. Tendo isto em conta, aliando técnicas de computação baseadas em tecnologia sofisticada, que permitem a gravação da voz e a sua posterior análise, obtém-se medidas quantitativas das alterações da voz. Desta forma, torna-se possível detetar alterações da voz que auditivamente são impercetíveis.

As patologias provenientes da laringe condicionam a qualidade de vida dos pacientes. Tais patologias causam alterações na qualidade da fala, sendo que, muitas vezes, estas alterações não são detetadas ao ouvido humano.

O recurso a técnicas comuns para a deteção destas patologias, provocam desconforto ao paciente, uma vez que são técnicas invasivas e, ainda, têm a desvantagem do custo associado ser dispendioso.

No entanto, a medição de alguns parâmetros relacionados com a fala patológica, analisados independentemente, não são muito conclusivos, contudo, associados

adequadamente a uma ferramenta de Inteligência Artificial, os resultados são consideravelmente melhores, (relativamente às patologias estudadas nesta dissertação) (Teixeira, Ferreira, & Carneiro, 2011) (Teixeira, Oliveira, & Lopes, 2013) (Alves N. , 2016).

A análise de um grande número de dados, com diversas variáveis, nem sempre é possível para o ser humano, uma vez que a possibilidade de ocorrerem erros é bastante elevada.

Associando os parâmetros acústicos, contidos numa base de dados, a classificadores de inteligência artificial, pode-se reduzir a complexidade dos exames auditivos através de um método não evasivo.

As ferramentas de Inteligência Artificial são, então, uma mais-valia a ter em conta para este tipo de tarefas. Depois de realizado o treino de um sistema de inteligência artificial, este deve ter a capacidade de generalizar, isto é, encontrando uma situação nunca vista anteriormente, o sistema deve ser capaz de tomar uma decisão, sendo esta, baseada em similaridades de parâmetros vistos anteriormente (Lanc, 1992) (Alves N. , 2016).

As árvores de decisão são modelos representativos de tabelas de classificação através de uma estrutura hierárquica, onde, em cada nível da árvore é analisado um atributo e, por sua vez, é tomada uma decisão (Cordeiro H. , 2016).

Os SVM's apresentam-se como classificadores binários, isto é, apenas podem classificar os dados a analisar numa de duas classes. Assumindo que é possível separar os dados, compete ao SVM encontrar um hiperplano que as separe. A utilização desta ferramenta é uma escolha adequada, visto que, existem diversos classificadores lineares. O hiperplano do SVM é criado baseado no conceito de margem máxima, ou seja, o plano escolhido entre os dados deve maximizar a distância entre as duas opções (Cordeiro H. , 2016).

Baseadas no cérebro humano, as redes neuronais artificiais são sistemas inteligentes capazes de aprender e reconhecer padrões de acordo com os dados que são fornecidos às suas entradas, assim, generaliza este conhecimento para analisar novos dados (Haykin S. , 2001) (Mackay, 2005) (Baravieira, 2016).

Uma Deep Learning é baseada nas redes neuronais ditas normais, no entanto, apresentam um número maior de camadas ocultas na sua arquitetura, o que ajuda no processamento da informação. Cada camada escondida pode ter funções de ativação diferentes das restantes.

Nas análises realizadas nesta dissertação são utilizados três grupos de parâmetros, o grupo I parâmetros fonte (a) contêm parâmetros como *Jitter* relativo, *Shimmer* relativo, HNR, determinados em segmentos de fala estacionária (3 vogais 'a', 'i' e 'u' em três tons disponíveis alto, baixo e normal), o grupo parâmetros fonte (b) contem os parâmetros do grupo parâmetros fonte (a) em conjunto com os parâmetros NHR e Autocorrelação, determinados nos mesmos sinais de (a). O grupo II é baseado em coeficientes MFCC's, determinados nos mesmos segmentos de fala de a). O grupo III é formado por coeficientes MFCC extraídos de fala contínua ("bom dia, como estás?" traduzido do alemão).

1.1. Objetivos do trabalho

Investigações realizadas até à data, no reconhecimento de patologias da voz (identificação da patologia), são ainda precoces. Contudo, determinados estudos têm sido feitos ao longo das últimas décadas. Assim, sabe-se, como referido no estado da arte desta dissertação, que é possível fazer a identificação de patologias recorrendo a parâmetros extraídos de vozes humanas.

Trabalhos anteriores fazem referência a parâmetros como *Jitter* e *Shimmer*, o que em alguns casos não mostra resultados conclusivos para o reconhecimento de determinadas patologias.

O desenvolvimento tecnológico atual permite um acesso fácil a ferramentas de inteligência artificial. Aliar determinados parâmetros extraídos da voz a uma destas ferramentas, pode levar a que se encontre um "meio" de separação das características das diferentes patologias.

O objetivo desta dissertação consiste no estudo e utilização de um conjunto de ferramentas de *machine learning* (árvores de decisão*, Support Vector Machine, Redes neuronais - Deep Learning*) aplicando diferentes parâmetros extraídos da voz de sujeitos. Os parâmetros aplicados foram extraídos de vogais ditas de forma sustentada e de uma frase dita de forma continua.

Pretende-se, assim, distinguir a fala saudável da fala patológica, recorrendo a uma classificação binária onde os sujeitos saudáveis são distinguidos de sujeitos portadores de patologias, contudo nesta etapa a patologia não é indicada.

Por fim pretende-se fazer uma classificação quaternária, onde é indicado se o sujeito é saudável ou patológico, caso seja patológico, pretende-se fazer a distinção da patologia portadora pelo sujeito em teste.

1.2. Organização do trabalho

No presente capítulo é apresentada a contextualização da pesquisa e indicados quais os objetivos desta dissertação.

No capítulo 2 (estado da arte), estão apresentados alguns dos desenvolvimentos mais recente na área em estudo, onde é indicado, por exemplo, taxas de exatidão obtidas em estudos realizados com um prepósito semelhante ao desta dissertação.

No capítulo 3 (materiais), é indicada informação acerca da base de dados utilizada neste estudo, bem como acerca dos sinais utilizados. Também contêm informação acerca dos parâmetros extraídos do sinal acústico e, por fim, uma breve descrição das patologias em estudo e como se manifestam.

No capítulo 4 (métodos), contem abordagens teóricas acerca de ferramentas, tipos e elementos de *machine learning*. Neste capítulo, também, estão descritos os modelos utilizados nesta dissertação.

No capítulo 5 (resultados), constam os resultados obtidos no decorrer desta dissertação com aplicação dos diferentes modelos analisados.

No capítulo 6 (análise dos resultados), é feita uma análise dos resultados obtidos, comparando os diferentes grupos de parâmetros utilizados, bem como as ferramentas de *machine learning*.

Por último, no capítulo 7 (conclusão), surge a conclusão onde é feito um resumo de todo o trabalho elaborado, mencionados os melhores resultados, e comparando os resultados obtidos, com resultados de trabalhos realizados anteriormente (referidos no capítulo 2). Neste capítulo são também indicadas algumas sugestões para trabalhos futuros.

Constam ainda desta dissertação os seguintes anexos:

Anexo A – Exatidões obtidas com aplicação de SVM's para diferentes combinações.

Anexo B – Valores obtidos com as redes da primeira camada de Deep-Learning.

Anexo C – Matrizes confusão para análise detalhada.

Anexo D – Interface gráfica, desenvolvida no âmbito do programa StartUP Voucher 2018 do IAPMEI.

2.Estado da arte

O Processamento Digital de Sinal tem-se tornado mais dinâmico e em constante desenvolvimento. Os sistemas mais antigos para a realização deste processamento eram analógicos e incluíam, por exemplo, um Voder (*Voice Demonstration Recorder*), para síntese de fala por controlo manual (Rabiner & Schafer, 2011).

Por volta de 1970, a necessidade de encontrar, de forma objetiva, patologias da laringe, já era sentida.

(B.Davis, 1979) concluiu que, os métodos, envolvendo a análise acústica da voz, são mais objetivos que os métodos auditivos, na avaliação da voz. Assim, propôs que fosse utilizado um computador para calcular a filtragem inversa do sinal de voz, de forma a avaliar precocemente casos de patologia e monitorizar o progresso durante a terapia da voz.

Com o decorrer do tempo, alguns trabalhos têm vindo a ser desenvolvidos. O avanço da tecnologia tem aprofundado cada vez mais o tema em estudo nesta dissertação.

A análise acústica, na avaliação da qualidade vocal, pode ser utilizada para a determinação objetiva de alterações da função vocal, avaliações de cirurgias, tratamentos farmacológicos e de reabilitação (Godino-Llorente, Gomez-Vilda, & Blanco-Velasco, 2006).

Em 2002, prova-se que um classificador linear pode identificar as vozes patológicas e as vozes saudáveis, utilizando a Transformada *Wavelet Packet* (WPT) e o algoritmo *Best Basis* (BBA). Para as condições de trabalho na época em causa, a taxa de acerto para este classificador foi de 67,2% (Parraga, 2002).

Karthikeyan Umapathy e Sridhar Krishnan propuseram um estudo onde utilizaram frases em inglês, ditas por pacientes com diferentes patologias de origem orgânica, neurológicas e traumáticas. Efetuando o treino e teste de um classificador com o algoritmo LDB (*Local Discriminant Bases*) onde utilizaram decomposições *wavelet packet* (algoritmo *best-basis*), atingiram 96% de acerto na classificação em dois grupos (voz normal e voz patológica), e para uma classificação de quatro grupos atingiram 74% de acerto (voz normal masculina, voz normal feminina, voz patológica masculina, voz patológica feminina) (Umapathy & Krishnan, 2005).

Um algoritmo para identificar patologias laríngeas, foi apresentado por Fonseca et al. em 2007. Este algoritmo baseia-se na transformada de Wavelet discreta de *Daubchies*

(DWT-db), nos coeficientes de predição linear (LPC) e nas SVM de mínimos quadrados. Compara-se wavelet´s com tamanhos diferentes e três valores de *Kernel.* A particularidade deste método consiste num classificador adequado para as patologias da laringe, para identificar nódulos nas pregas vocais. Apresenta mais de 90% de exatidão na classificação e tem uma complexidade computacional baixa em relação ao comprimento do sinal de fala (Fonseca E. S., Guido, Scalassara, Maciel, & Pereira, 2007).

Costa (2008) avaliou a separação entre vozes do tipo Normal e vozes Patológicas e, num segundo momento, entre Normal e Edema e Outras Patologias da base de dados 'KAY Elemetrics, 1994'. Ao combinar os parâmetros utilizados (coeficientes por predição linear, *cepstral, mel-cepstral*, etc) por meio de abordagens de *Machine Learning*, obteve um desempenho superior a 92% (Costa S. , 2008).

No artigo de Henriquez et al (2009), foi estudado a utilização de medidas caóticas, baseadas na teoria dinâmica não-linear, onde se pretende fazer a distinção entre os dois níveis de qualidade de voz: patológica e saudável. As medidas utilizadas foram a 1ª e 2ª ordem da Entropia de Rényi, Entropia de Correlação e Dimensão de Correlação. Foram utilizadas duas bases de dados para este estudo, a primeira base de dados é composta por quatro níveis de qualidade de voz (voz saudável e três níveis de vozes patológicas), a outra base de dados é composta por dois níveis de qualidade de voz (saudáveis e patológicas). Utilizando um classificador baseado em redes neuronais artificiais, a taxa de sucesso para a primeira base de dados foi de 82,47% e para a segunda base de dados foi alcançada uma taxa de acerto de 99,69% (Henriquez P. Alonso J. B., 2009).

Fonseca e Pereira (2009) obtiveram a separação entre vozes normais e patológicas e entre nódulos e edemas, com ausência de validação cruzada. Com uma taxa de acerto média de cerca de 85% utilizando SVM's. A base de dados, utilizada neste estudo, era composta por 60 indivíduos (homens e mulheres, com idades compreendidas entre 4 e 72 anos) e foi usada a vogal sustentada \a\ (Fonseca & Pereira, 2009).

(Amato, 2009) apresenta uma abordagem com base no sistema web, onde é possível o utilizador enviar os sinais vocais através de uma interface. Os sinais enviados são analisados em tempo real recorrendo a técnicas de processamento de sinal, fornecendo assim informações sobre possíveis alterações na voz. Este sistema oferece várias funções que permite analisar detalhadamente casos específicos e foi testado regionalmente.

Uma nova técnica de transformação de características, para melhorar a exatidão de triagem para a deteção de vozes patológicas, é apresentada por Arias-Londoño et al. em

2010. A transformação dos dados é baseada nos modelos de Markov, obtendo a fase de transformação e classificação simultaneamente e ajustando os parâmetros do modelo que minimiza o erro de classificação. Os vetores, com os dados originais, são formados usando parâmetros com ruido a curto prazo e os MFCC's. Esta técnica demonstra uma melhoria significativa do desempenho, sem que sejam adicionados novos dados à entrada original (Arias-londono, Godino-Llorente, Sáenz-Lechón, Osma-Ruiz, & Castellanos-Domínguez, 2010).

Parâmetros como *Jitter* (perturbações de frequência), *Shimmer* (perturbações de amplitude), componentes sub-harmónicos e a distorção da envolvente do sinal de voz têm sido utilizados para a deteção de patologias na voz por (Beber, 2010) e por (Teixeira et al., 2011) (Espinola, 2014).

Kohler (2011), estudou a implementação de redes neuronais para identificar sujeitos com nódulos nas cordas vocais, sujeitos com paralisia das cordas vocais e sujeitos saudáveis, obtendo uma taxa de acerto de 96% (Kohler, 2011).

(Wang, 2011) afirma que os MFCC´s são parâmetros eficazes para a classificação de vozes patológicas ou saudáveis utilizando um Modelo de Mistura Gaussiano (GMM), mas que pode ser melhorada a eficácia do método de classificação pode ser melhorada. Então foi aplicado um SVM em simultâneo com GMM e comparados os resultados com o GMM simples para a deteção de fala patológica. Foram utilizadas gravações da base de dados da KAY para realizar os experimentos, e concluíram que é possível através de uma vogal sustentada, fazer a classificação entre normal e patológico com uma precisão de 96,1. Conclui-se também que as taxas de erro diminuem de 8% no GMM para 4,6% no GMM-SVM.

Brandt (2012) avaliou a separação entre vozes normais e vozes patológicas dos grupos – segmentada por identidade de gênero (caso I). Num segundo momento (caso II) entre "Normal" e "Edema, Nódulos e Paralisia". Esta separação foi feita por meio de árvore de decisão binária, RN MLP e Máquina de vetores de suporte (SVM). A sua eficácia, para o caso I, apresentou exatidão de 94,6% (voz feminina) e 87,3% (voz masculina). No caso II, apresentou uma taxa de acerto de 73,5% (voz feminina) e 62,5% (voz masculina) (Brandt, 2012).

Em trabalhos anteriores (Teixeira & Gonçalves, 2014) e na tese de mestrado de Gonçalves 2015, foram desenvolvidos algoritmos que efetuam a medida de alguns

parâmetros fundamentais, tais como *Jitter, Shimmer* e HNR, com elevadas precisões em vozes reais, quer sejam de controlo quer sejam patológicas (Teixeira & Gonçalves, 2014).

Sergio Espinola caracterizou um modelo de processamento digital de voz para o apoio ao diagnóstico, no contexto da construção de sistemas de identificação automatizados de patologias da fala. Separou vozes patológicas de vozes normais, bem como, patologias específicas (Paralisia, Edema de Reinke, nódulos) com uma exatidão de 100 % e cerca de 92% para nódulos contra Edema de Reinke. Utilizou métodos de *Machine Learning* (redes neuronais artificiais e SVM'S) e a sua complexidade e dimensão foi reduzida utilizando a técnica de análise de componentes principais (PCA), para aplicar na separação entre patologias. A realização de testes estatísticos, com grupos locais, confirmaram, também, limiares de indícios de anomalias, como era esperado teoricamente. A utilização de menor quantidade de parâmetros, obtida após PCA, mostrou-se também eficiente, atingindo as mesmas taxas de acerto. Foi usada uma base de dados constituída por 182 vozes e o som consistia na vogal \a\ de forma sustentada (Espinola, 2014).

(Muhammad, 2014) apresenta um método de classificação patológica baseada no áudio de baixo nível (MPEG-7), as experiências foram realizadas em gravações de uma vogal sustentada 'a' retiradas da base de dados Massachusetts Eye and Ear Infirmary (MEEI), foram utilizados SVM's para deteção patológica bem como para classificações binárias entre patologias. Atingiu-se uma exatidão de 99,994% com um desvio padrão de 0.0105% para a deteção de vozes patológicas e exatidões de até 100% para classificação binária entre patologias. Este estudo afirma que os parâmetros derivados de MPEG-7 podem ser utilizados de forma eficaz para a deteção e classificação automática das patologias da voz.

Em 2015, Panek *et al*. apresenta uma forma de deteção automática de patologias da voz, onde analisa um vetor de 28 parâmetros acústicos, utilizando a análise de componentes principais, análise de componentes principais do *Kernel* e uma rede neuronal auto-associativa, em quatro tipos de deteção de patologia (disfonia hiperfuncional, disfonia, laringite e paralisia das cordas vocais), utilizando as vogais /a/, /i/ e /u/ ditas num tom alto, médio e baixo. Concluíram que os métodos de análise dos componentes de *Kernel* e da rede neuronal auto-associativa são um avanço na deteção de patologias das pregas vocais, onde, os níveis de eficácia rondam os 100%. Este estudo abordou as técnicas mais utilizadas para o processamento do sinal de fala, onde, a

comparação dos parâmetros através de métodos de *Machine Learning* determina o estado de saúde do paciente (Panek, Skalski, Gajda, & Tadeusiewicz, 2015).

Em 2015, Forero *et al*., utilizaram os parâmetros do sinal glotal para ajudar a identificar dois tipos de patologia da voz, nódulos e paralisia unilateral. Estes parâmetros foram obtidos através de uma filtragem inversa e são utilizados como entrada para uma rede neuronal artificial, para uma SVM e também para o Modelo de Markov, para obter a classificação e comparar os resultados dos sinais de voz em três grupos diferentes. A base de dados é constituída por 248 gravações, correspondentes aos três grupos e a sua taxa de sucesso atinge 97,2% (Forero, Kohler, Vellasco, & Cataldo, 2015).

Nuno Alves (2016) desenvolveu um conjunto de Redes Neuronais Artificiais e Máquinas de Vetor de Suporte (SVM), que permitiram, com elevada exatidão (ultrapassando os 91%), identificar patologias associadas à Laringe (disfonia e paralisia das cordas vocais) (Alves N. , 2016).

Na dissertação de Hugo Cordeiro, com a utilização de árvores de decisão, conseguiu atingir 95% de acerto numa classificação binária, utilizando uma base de dados com 209 sujeitos, em que 106 sujeitos (27 saudáveis e 79 patológicos) foram utilizados para efetuar o treino, e 103 sujeitos (26 saudáveis e 77 patológicos) foram utilizados para realizar os testes (Cordeiro H. T., 2016).

Em Teixeira *et al*. (2018), pretendia-se a longo prazo desenvolver um sistema classificador baseado em redes neuronais artificiais e/ou máquina de vetor de suporte para classificar, com grande exatidão, sinais de fala entre as classes de laringite e controlo. Neste estudo encontra-se uma análise estatística de um conjunto de parâmetros sobre os grupos envolvidos (grupo de controlo e grupo com laringite crónica). A análise foi realizada com as vozes dos dois géneros (masculino e feminino). Os parâmetros utilizados neste estudo foram o *Jitter*, *Shimmer*, HNR, NHR e a Autocorrelação, retirados do som das vogais sustentadas /a/, /i/ e /u/ nos tons baixo, alto e normal. Para este estudo foi utilizada a base de dados *Saarbrucken Voice Database* (SVD). Numa primeira fase foram comparados os parâmetros por género para ambos os grupos, numa segunda análise comparou-se o grupo patológico para cada parâmetro. Verificou-se que na primeira fase só há diferenças de voz no *jitter* absoluto entre género masculino e feminino para o grupo de controlo. A comparação entre o grupo patológico e de controlo mostram conclusões semelhantes para os restantes 6 parâmetros. Estes parâmetros poderão ser importantes

para usar como ferramenta de decisão inteligente para classificar entre laringite crónica e saudável (Teixeira, Teixeira, Fernandes, & Fernandes, 2018).

(Selamtzis, 2018) investigou o efeito da vogal (sustentada e extraída de fala) através de dois parâmetros: proeminência do pico cepstral suavizado (CPPS) e entropia da amostra utilizada. Analisou-se trinta e um indivíduos portadores de disfonia (diferentes tipos de disfonia) e trinta e um indivíduos saudáveis. A vogal sustentada 'a' foi dita em tom e volume normal. A vogal 'a' em fala continua foi retirada de um texto lido através do reconhecimento automático de fala. Os parâmetros CPPS e entropia foram calculados para todas as vogais extraídas de cada sujeito, as vogais sustentadas foram analisadas em frames de 41 ms criando assim uma distribuição de valores para cada sujeito. A avaliação recorreu a análise ROC (Receiver-Operator Characteristic). Verificou-se que a entropia e CPPS estão correlacionados negativamente e que a utilização de vogais sustentadas pode ser mais eficaz para a detecção de disfonia.

J. Fernandes (2018) na dissertação de mestrado, desenvolve um algoritmo que permite determinar e extrair os parâmetros *Harmonic to Noise Ration (HNR), Noise to Harmonic Ratio (NHR)* e Autocorrelação, utilizados no âmbito desta dissertação (Fernandes, 2018).

Estudos desenvolvidos por (Fang, 2018) Basearam-se na aquisição de 60 amostras de vozes saudáveis e 402 patológicas (compreendidas entre oito patologias). Foram extraídos os parâmetros MFCC's de uma vogal sustentada e aplicados a três ferramentas de machine learning (deep neural networks, SVM's e o modelo de mistura Gaussiano). Foi concluído que as deep neural networks superam os SVM's e o modelo de mistura gaussiano atingindo 94,26 % de precisão para detectar patologias.

Guedes (2019), estudou a aplicação de mfcc's e espectrogramas de uma frase, em modelos de inteligência artificial. Afirma que dos modelos aplicados, a rede neuronal clássica, apresenta melhores resultados quando comparado com *Transfer Learning* (LSTM e Conv1D) Atingiu 99% de exatidão para a implementação de um modelo LSTM com parâmetros *Jitter, Shimmer* e Autocorrelação, na classificação binária entre laringite e saudável. Para as frases, realizou-se um estudo comparativo entre modelos de redes neuronais, convolucionais e recorrentes para os parâmetros MFCCs e Espectrogramas na escala Mel obtendo resultados de 76% de medida-F para disfonia x saudável, 68% de medida-F para laringite x saudável, 80% de medida-F para paralisia x saudável. Para

classificação multi-classe é obtido 59% e 40% de medida-F para 3 classes e 4 classes, respetivamente (Guedes, 2019).

Os melhores resultados foram encontrados por Henriquez et al (2009), onde utilizou a 1ª e 2ª ordem da Entropia de Rényi, Entropia de Correlação e Dimensão de Correlação e para uma classificação quaternária (voz saudável e três níveis de vozes patológicas) atingiu 82.47%, já para uma classificação entre saudável / patológico alcançou uma taxa de acerto de 99,69%, ambos os resultados foram alcançados com o recurso a redes neuronais artificiais.

3.Base de dados e parâmetros

Neste capítulo é feita uma abordagem às patologias utilizadas no desenvolvimento deste trabalho, bem como os parâmetros necessários para utilização nesta dissertação. É mencionado, também, uma análise estatística sobre alguns dos parâmetros.

As patologias diretamente ligadas à laringe designam-se por patologias da voz ou patologias laríngeas. Existem várias lesões que podem causar estas patologias, como lesões mínimas estruturais e/ou funcionais da laringe, lesões de massa localizada nas pregas vocais, alterações tecidulares da prega vocal, perturbações neurológicas e perturbações não orgânicas ou de tensão muscular (Cordeiro H. , 2016).

Nos estudos efetuados neste trabalho, foi utilizada uma base de dados alemã, *Saarbrucken Voice Database* (SVD), que se encontra disponível *online* de forma gratuita pelo Instituto de Fonética da Universidade de *Saarland*.

Nesta base de dados é possível encontrar sinais de voz, entre saudáveis e patológicos de mais de 2000 sujeitos. Para cada sujeito é disponibilizada a gravação de 3 fonemas (/a/, /i/ e /u/) nos tons baixo, normal e alto, ditos de forma sustentada. Existe ainda a gravação de uma frase em alemão: *"Guten Morgen, wie geht es Ihnen?"* (Bom dia, como estás?). A frequência de amostragem dos sinais de voz é de 50 kHz. O tamanho dos ficheiros situa-se entre 1 e 3 segundos com uma resolução de 16 bits.

3.1. Patologias

Há uma grande variedade de patologias que causam mudanças significativas nos padrões vibratórios, afetando a qualidade da produção vocal. Essas patologias podem derivar de infeções, tumores ou paralisias laríngeas (Teixeira, Ferreira, & Carneiro, 2011).

Aqui apresentam-se apenas as patologias usadas neste trabalho.

Disfonia

Pode ser entendido como o principal sintoma de distúrbio da comunicação oral, isto faz com que sejam apresentadas limitações ou dificuldades na transmissão da mensagem verbal de um sujeito. Por outras palavras, disfonia, representa qualquer dificuldade ou alteração na emissão natural da voz. Sendo a disfonia uma limitação vocal, pode ser classificada em quatro graus de intensidade, leve, moderado, intenso e grau extremo

(afonia). As disfonias podem ser do tipo funcionais (ou primárias), onde o uso da voz é a causa da disfonia, ou do tipo orgânica, onde o uso da voz provoca lesões nas estruturas produtoras de voz (Cantoni, 2017).

Manifesta-se através de sintomas como, esforço ao emitir a voz, rouquidão, variação da frequência da voz habitual, perda da eficiência vocal, falta de ar ao falar, dor ao emitir sons, dificuldade em manter a voz, entre outros. O tratamento para este tipo de patologia sucede aos exames médicos invasivos (exame clínico, exame de imagem). É possível reduzir a disfonia num sujeito por meio de exercícios da voz, casos mais graves podem estar sujeitos a cirurgia médica como, por exemplo, remover uma lesão (pólipo, nódulo), ou até mesmo para reconstruir pregas vocais (Romano, 2017).

Laringite crónica

A laringe encontra-se entre a garganta e a traqueia e contém as cordas vocais (Enciclopédia médica, 2001).

Uma das complicações mais comuns identificadas na laringe é a laringite. Esta inflamação pode manifestar-se na forma aguda, tendo um início abrupto e geralmente autolimitado, ou na forma crónica, persistindo os sintomas mais de três semanas (Medscape, 2017). Na figura 3.1 é possível observar a diferença entre uma laringe normal e uma laringe com laringite.

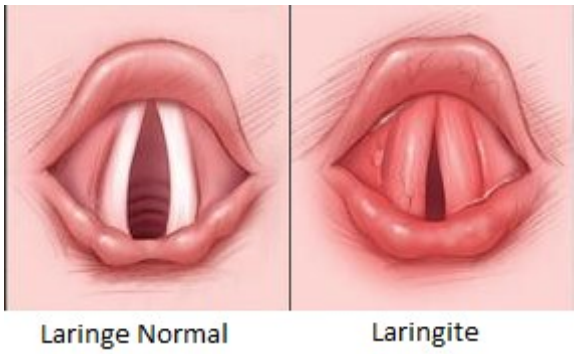

Figura 3.1- Diferença entre uma laringe normal e uma laringe portadora de laringite (Alves C. , 2018)

A etiologia da laringe aguda inclui um uso excessivo da voz, a exposição a agentes nocivos ou agentes infeciosos levando a infeções do trato respiratório superior. Os agentes infeciosos são na maioria das vezes virais, mas também podem ser bacterianos (Medscape, 2017).

Também, temos a laringite crónica que pode ser causada pela inalação de substâncias irritantes (fumo do tabaco, produtos químicos) e a longo prazo pelo uso intensivo da voz.

Os sintomas da laringite, geralmente, desenvolvem-se ao longo de um período de 12 a 24 horas e variam de acordo com as causas subjacentes. Os sintomas podem incluir perda gradual da voz, rouquidão e dores de garganta. Nas infeções mais graves pode ocorrer febre, disfagia (dificuldade na deglutição) e, raramente, edema (Enciclopédia médica, 2001) (Sasaki, Laringitis, 2017).

O diagnóstico baseia-se nos sintomas típicos e nas alterações da voz. Por vezes, o profissional de saúde faz uma laringoscopia com um tubo de visualização fino e flexível, que pode mostrar vermelhidão e pequena vasculatura dilatada nas cordas vocais (Medscape, 2017) (Sasaki, Laringitis, 2017).

Paralisia das cordas vocais

A sua causa pode ser derivada de tumores, danos nos nervos ou lesões, causados geralmente por toxinas ou infeções. Nos seus sintomas mais comuns destacam-se a mudança de voz e possivelmente dificuldades em respirar. Em média, os sujeitos de género feminino são mais afetados que os sujeitos do género masculino (Sasaki, Paralisia das cordas vocais, s.d.).

A paralisia pode afetar uma ou as duas cordas vocais. Paralisia de uma corda vocal pode ocorrer devido a doenças cerebrais (tumores, acidente vascular cerebral), doenças desmielinizantes (por exemplo esclerose múltipla) ou lesões nos nervos que ligam à laringe. As lesões nos nervos podem ser causadas de diversas formas, destacando-se entre elas, tumores (malignos e benignos), lesões e cirurgias no pescoço. Na maioria dos sujeitos a causa nunca é apurada com eficácia máxima. Um sintoma típico, neste caso, é a rouquidão, nestes tipos de caso a via respiratória não fica obstruída (Sasaki, Paralisia das cordas vocais, s.d.).

A paralisia de duas cordas vocais, é um distúrbio que pode mesmo por em causa a vida do sujeito. Pode ser causado por cirurgias na coluna, inserção de uma sonda endotraqueal ou por doenças que afetam os músculos e os nervos (por exemplo miastenia grave). Sintomas gerais passam por potência da voz reduzida, dificuldade em respirar, som rouco e agudo quando o sujeito respira (Sasaki, Paralisia das cordas vocais, s.d.).

A paralisia das cordas vocais impede o correto funcionamento das cordas vocais impedindo que abram ou fechem, isto pode afetar a fala, respiração e até mesmo a deglutição (pode fazer com que os alimentos sólidos e líquidos sejam inalados até à traqueia e aos pulmões) (Sasaki, Paralisia das cordas vocais, s.d.).

Diagnósticos para esta patologia passam por laringoscopias, estudos de diagnóstico por imagem, imagens de ressonância magnética, radiografias, entre outros. O seu tratamento baseia-se em cirurgias. No caso da paralisia de uma corda vocal, serve para mover a corda vocal paralisada de maneira a que o sujeito possa conter um falar o mais perto do normal possível, no caso das duas cordas vocais paralisadas, surge muitas vezes a necessidade de criar uma abertura na traqueia, visto que é complicado manter as vias respiratórias abertas de maneira adequada (Sasaki, Paralisia das cordas vocais, s.d.).

3.2. Parâmetros utilizados

Nesta dissertação, a separação por género não foi feita, uma vez que, em (Teixeira J. , Fernandes, Teixeira, & Odete, 2018) ficou provado que para os parâmetros usados não há diferença entre o sexo masculino e feminino.

Nesta análise utilizou-se sujeitos de controlo/saudáveis e sujeitos portadores de patologias – disfonia, laringite crónica e paralisia das cordas vocais. Na tabela 3.1 é possível observar o número de sujeitos associado a cada categoria, média de idades e desvio padrão de idades.

Tabela 3.1– Grupos e tamanho da amostra, média e desvio padrão das idades.

Grupos	Tamanho da Amostra	Média de Idades	Desvio Padrão das Idades
Controlo	194	38	14,4
Disfonia	69	47,4	16,4
L. Crónica	41	49,7	13,5
P. C. Vocais	169	57,8	13,8
Total	473	-	-

Parâmetros extraídos do sinal acústico

Jitter

O *jitter* é definido como uma medida de variação do período glotal entre ciclos de vibração das pregas vocais. A Figura 3.2 contém a representação gráfica, a fim de entender melhor a sua definição. Sujeitos que não consigam controlar a vibração das cordas vocais têm tendência a ter valores de *jitter* mais elevados. O *jitter* pode ser medido

de quatro formas diferentes, porém, neste estudo só é utilizada uma dessas formas (Alves N. , 2016).

A análise do *jitter relativo* num sinal de fala é a média da diferença absoluta entre períodos consecutivos, dividida pelo período médio e é expresso em percentagem (equação 1).

$$jitter\ (relativo) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \times 100 \quad (1)$$

Na equação 1 T_i é o tempo de duração do período glotal i e N é o número total de períodos glotais.

Shimmer

O *shimmer* foi outro parâmetro extraído e relaciona-se com a variação da amplitude a cada ciclo. Através da Figura 3.2 é possível entender melhor a definição de *shimmer*. Uma redução na resistência glotal e lesões podem causar variações da amplitude glotal correlacionadas com a soprosidade e emissão de ruído, dando lugar a um valor de *shimmer* mais elevado. Este pode ser medido de quatro formas diferentes, porém, neste estudo só será usada uma por serem diferentes formas de medir a mesma variação da amplitude glotal (Alves N. , 2016).

O *shimmer* relativo é definido como a média da diferença absoluta entre amplitudes de períodos consecutivos, dividida pela amplitude média, expresso em percentagem (equação 2).

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \times 100 \quad (2)$$

Na equação 2 A_i é a magnitude do período glotal i e N é o número total de períodos glotais.

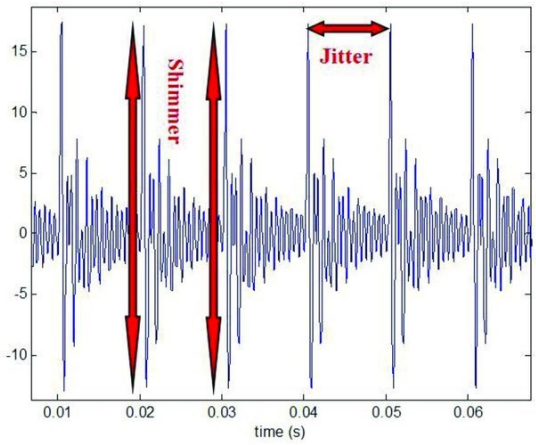

Figura 3.2- *Jitter e Shimmer* de um sinal sonoro (Teixeira & Gonçalves, Accuracy of Jitter and Shimmer Measurements, 2014)

A harmonicidade é medida com três parâmetros, HNR (*Harmonic to Noise Ratio*), NHR (*Noise to Harmonic Ratio*) e a Autocorrelação.

HNR

O HNR é um parâmetro em que a relação entre as componentes harmónicas e de ruído fornecem uma indicação de periodicidade global do sinal de voz, pela quantificação da relação entre a componente periódica (parte harmónica) e aperiódica (ruído). O valor global de HNR de um sinal varia, porque diferentes configurações do trato vocal implicam diferentes amplitudes para os harmónicos. O valor de HNR (em dB) pode ser determinado pela equação 3 (Alves N. , 2016).

$$\mathrm{HNR} = 10 \times \log_{10} \frac{\mathrm{AC_v(T)}}{\mathrm{AC_v(0)} - \mathrm{AC_v(T)}} \qquad (3)$$

Onde $AC_v(T)$ (numerador) representa a potência da componente harmónica do sinal e $AC_v(0)$ corresponde à potência total do sinal. A diferença das duas (denominador) é assumida como sendo a componente de ruído.

Autocorrelação

Na função da autocorrelação é avaliada a similaridade no mesmo sinal, ou seja, a ocorrência de trechos de amostras semelhantes entre si. Para tal é feita uma comparação em partes ao longo do sinal para encontrar as respetivas semelhanças, deste modo quanto maior o numero de semelhanças, maior será o valor da autocorrelação (Guedes, 2019) (Ortigueira, 2005) (equação 4).

De acordo com (Fernandes, 2018), matematicamente a autocorrelação pode ser determinada em três passos.

No primeiro procedimento (equação 4), tendo um sinal x(t) utiliza-se um segmento do mesmo sinal com a duração T, centrado em $t_{méd.}$ Da parte selecionada, é subtraída a média de μ_x e o resultado é multiplicado por uma função de janela w(t) de modo a obter uma janela do sinal:

$$a(t) = \left(x\left(tméd - \frac{1}{2}T + t\right) - \mu x\right)w(t) \tag{4}$$

A função de janela w(t) é simétrica em torno de t = 12 T e 0 em todos os lugares fora do intervalo de tempo [0,T]. (Boersma, 1993) menciona que a janela deve ser uma janela sinusoidal ou de Hanning, dada pela equação 5.

$$w(t) = \frac{1}{2} - \frac{1}{2}\,cos\frac{2\pi t}{T} \tag{5}$$

Em seguida é calculada a autocorrelação normalizada ra (τ) da parte do sinal selecionada. Esta é uma função simétrica ao atraso τ:

$$\text{ra}\,(\tau) = \text{ra}\,(-\tau)\frac{\int_0^{\text{T}-\tau} a(t)a(t+\,\tau)\text{dt}}{\int_0^T a^2\,(t)dt} \tag{6}$$

Por último é necessário calcular a autocorrelação normalizada rw(τ) da função de janela utilizada. Utilizando a janela de *hanning* a autocorrelação é obtida através da equação 7.

$$\text{rw}\,(\tau) = \left(1 - \frac{|\tau|}{T}\right)\left(\frac{2}{3} + \frac{1}{3}\cos\frac{2\pi\tau}{T}\right) + \frac{1}{2\pi}\,sin\frac{2\pi|\tau|}{T} \tag{7}$$

Para estimar a autocorrelação $r_x(\tau)$ do segmento de sinal original, dividimos a autocorrelação $r_a(\tau)$ da janela do sinal pela autocorrelação $r_w(\tau)$ da janela utilizada (Eq.8).

$$\text{rx}\,(\tau) = \frac{ra(\tau)}{rw(\tau)} \tag{8}$$

Onde $r_a(t)$ é a autocorrelação normalizada do sinal e $r_w(t)$ é a autocorrelação normalizada de uma janela utilizada, por exemplo, a janela de *hanning*.

Com isto, a função de autocorrelação de um sinal de voz sustentada exibe os máximos locais para valores múltiplos de τ, deste modo apenas é necessário identificar o primeiro máximo local, que será correspondente à parte harmónica (Fernandes, 2018).

NHR

O NHR quantifica a relação entre a componente aperiódica (ruído) e a componente periódica (parte harmónica). Apesar de ser o inverso do HNR, não se mede no domínio logarítmico, logo os valores não são o inverso (Boersma, 2004).

Na figura 3.3 é possível observar a determinação do NHR.

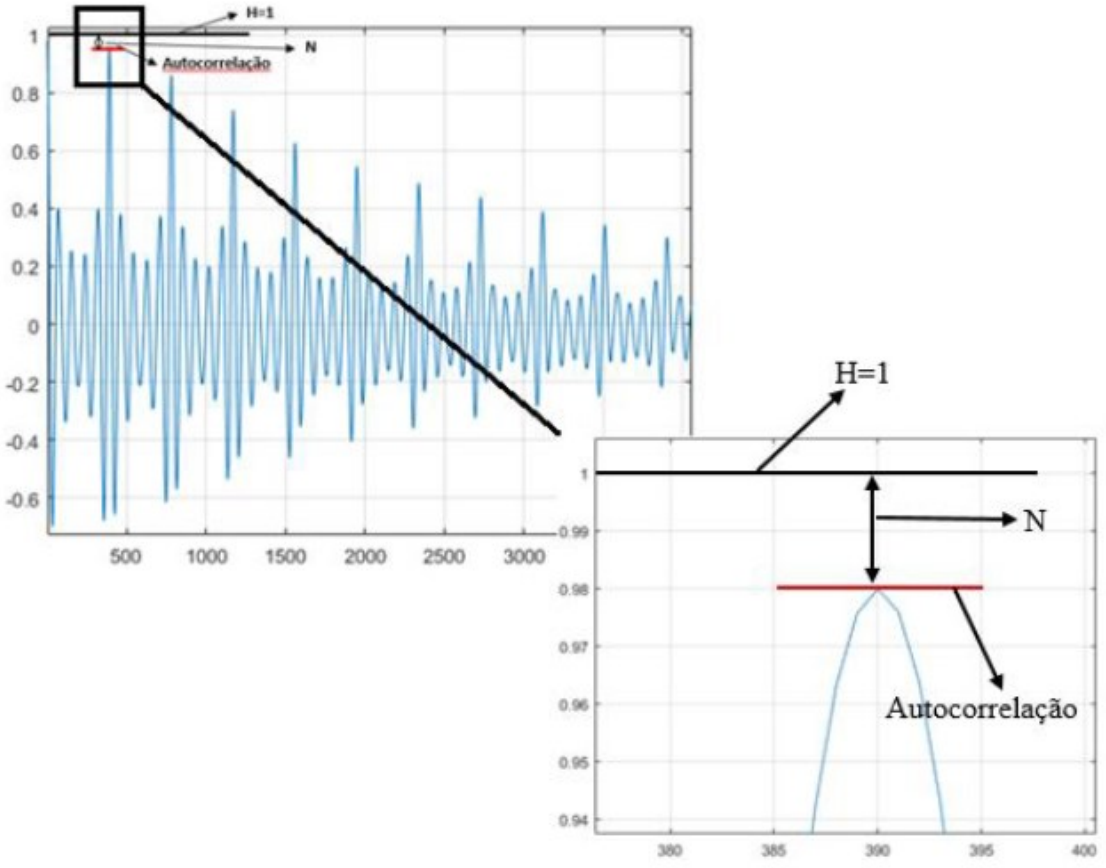

Figura 3.3- Determinação do NHR (Fernandes, 2018)

De acordo com a figura 4, o valor de ruído é mencionado como N. Este parâmetro é determinado de acordo com a equação 9.

$$NHR = \frac{N}{H} = \frac{H - Autocorr}{H} = \frac{1 - Autocorr}{1} = 1 - Autocorr \quad (9)$$

MFCC

Coeficientes Cepstrais na Frequência Mel, do inglês *Mel Frequency Cepstral Coefficients* (MFCC) surgiram com estudos feitos na área da psico-acústica (ciência que estuda a perceção auditiva humana), e indicam que a perceção humana das frequências de tons puros ou de sinais de voz não seguem uma escala linear (PUC-Rio). Estes parâmetros de curto termo são baseados no espectro.

Os MFCC's podem ser definidos como o cepstro de uma janela de análise determinada a partir de uma DFT, numa escala de frequência e magnitude que é característica da audição humana (Costa C. R., 2013) (Logan, 2000).

O cepstro consiste na representação do sinal de voz onde um sinal de fonte glótica, de variação temporal rápida, e a resposta do trato vocal, de variação lenta, são desacoplados e transformados em dois componentes aditivos. Os parâmetros do cepstro têm como particularidade conseguirem separar a excitação do trato vocal (Costa S. , 2008) (Cordeiro H. , 2016).

O processo para extração de MFCC's pode ser dividido em sete etapas: Pré-enfase, Framing, Janelamento, Transformada de Fourier discreta (DFT), Banco de filtros Mel, Transformada discreta do cosseno (The Discrete Cosine Transform (DCT)) e o cálculo da energia (Lindasalwa Muda, 2010).

A primeira etapa (Pré-Enfase) consiste em dar enfase às frequências mais altas, aumentando a energia do sinal nessas mesmas frequências. O cálculo para a Pré Enfase é dado pela equação 10 onde x[n] é o sinal de áudio (Lindasalwa Muda, 2010).

$$y[n] = x[n] - 0,95x[n-1] \tag{10}$$

A segunda etapa (Framing) tem como objetivo dividir o sinal em pequenos frames, onde é aconselhado que cada frame esteja compreendido entre 20 a 40 milisegundos.

A terceira etapa (Janelamento) tem como objetivo fazer a multiplicação de uma função de janela por cada frame do sinal.

É recomendado que seja utilizada uma janela de Hamming para a realização deste passo (Tiwari, 2010).

A equação 11 mostra a função da janela de Hamming.

$$y(n) = x(n)(0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right)) \tag{11}$$

Na quarta etapa é necessário converter as N amostras de cada frame do domínio dos tempos para o domínio das frequências, para tal é utilizada a transformada discreta de Fourier (Lindasalwa Muda, 2010).

Para o seu cálculo é utilizada a equação 12, onde X(k) são os coeficientes espectrais, x(n) o frame do sinal. É de referir que os valores de n e k devem ser maiores ou igual a zero e menor ou igual a N-1 (Guedes, 2019).

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{i2\pi nk}{N}} \quad (12)$$

A quinta etapa serve para fazer a transformação para a escala Mel.

"A perceção humana das Frequências dos sinais de áudio não segue uma escala linear, ou seja, para cada tom representado em Hz deve-se representar este mesmo tom para a escala Mel. Esta escala é uma escala de frequência linear abaixo dos 1000Hz e com um espaçamento logarítmico acima de 1000Hz. Como exemplo, um tom de 1KHz (40 decibéis) que está na perceção humana, tem o valor de 1000 mels" (Guedes, 2019).

Para fazer esta transformação é utilizada a equação 13. Neste processo são aplicados filtros triangulares no espetro para fazer a conversão.

$$Mel(f) = 2595 * log_{10}\left(\frac{f}{700} + 1\right) \quad (13)$$

Na sexta etapa surge a transformada discreta do cosseno, que consiste em transformar o espetro Mel para o domínio do tempo. Esta transformação pode ser designada como Mel Frequency Cepstrum Coefficient, onde os coeficientes de ordem mais baixa representam a forma do trato vocal e os coeficientes de ordem superior representam a periodicidade na forma de onda. (Lindasalwa Muda, 2010) (Tiwari, 2010).

Por fim, na última etapa é calculada a energia do frame de um sinal para um segmento no tempo t1 para o tempo t2 recorrendo à equação 14.

$$Energia = x2[t] \quad (14)$$

3.3. Análise Estatística aos Parâmetros

A eficiência e confiabilidade de um processo de discriminação de vozes patológicas, dependem, em grande parte, dos parâmetros ou caraterísticas que são utilizadas pelo classificador escolhido. A questão fundamental consiste em saber quais os parâmetros/caraterísticas que devem ser utilizados para identificar a patologia, e como saber qual o valor do parâmetro / caraterística que é associado à voz patológica ou à voz saudável (Espinola, 2014).

Fez-se uma análise estatística dos parâmetros *jitter* relativo, *shimmer* relativo, HNR, NHR e Autocorrelação, a fim de perceber se existe diferença entre sujeitos de controlo e patológicos. Para tal, teve-se em conta a vogal e o tom em que é dita. O objetivo principal era encontrar diferenças significativas entre os diferentes parâmetros, que permitissem fazer uma distinção entre as patologias em estudo.

Nesta análise foram utilizados diagramas de caixa ('boxplot'). A construção de diagramas de caixa necessita de 5 informações básicas, o mínimo e o máximo, que vão indicar os extremos do gráfico, e os valores do 1º quartil (Q1 ou Q0,25) e 3º quartil (Q3 ou Q0,75), e, por fim, o valor da mediana.

Neste tipo de gráficos obtêm-se, dentro da caixa, 50% dos sujeitos em estudo. Entre as linhas que definem a margem mínima e a margem máxima, estão 90% dos sujeitos em análise. Nos limites externos das margens mencionadas encontram-se os restantes 10%, que são valores discrepantes (*outliers).* O processo mencionado pode ser melhor compreendido com a análise da figura 3.4.

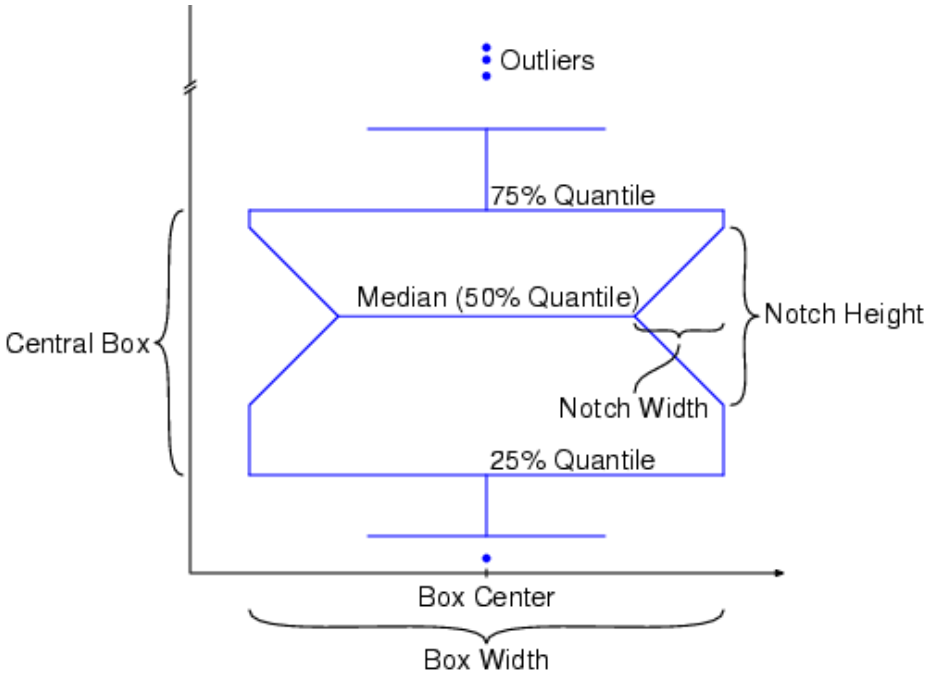

Figura 3.4- Explicação do gráfico de caixa (MathWorks, 2018)

Quando se faz uma análise deste género, devem ser utilizados uma população de, pelo menos, 30 sujeitos. Quando comparamos duas categorias, sendo por exemplo, grupo A e grupo B, podem surgir três hipóteses. Na hipótese 1 (representada na figura 3.5) onde o grupo B tem valores maiores que o grupo A, não existe sobreposição de caixas, ou o terceiro quartil do grupo A está abaixo do primeiro quartil do grupo B. Assim, afirma-se que existe uma diferença entre os grupos A e B

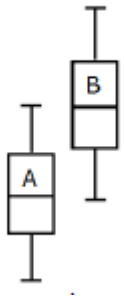

Figura 3.5- Hipótese 1- grupo B é maior que grupo A (Teixeira J. , Fernandes, Teixeira, & Odete, 2018)

Na hipótese 2 (representada na figura 3.6) onde as caixas se sobrepõem, mas as medianas são diferentes (primeiro quartil abaixo da mediana ou vice-versa), diz-se que é provável que haja uma diferença entre o grupo A e o grupo B.

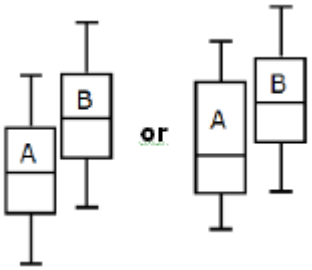

Figura 3.6- Hipótese 2- É provável que o grupo B seja maior que o grupo A (Teixeira J. , Fernandes, Teixeira, & Odete, 2018)

Na hipótese 3 (representada na figura 3.7) onde as caixas se sobrepõem, podendo ocorrer sobreposição de medianas inclusive, afirma-se que nenhuma diferença entre grupos pode ser reivindicada (Teixeira J. , Fernandes, Teixeira, & Odete, 2018).

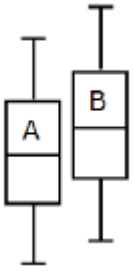

Figura 3.7- Hipótese 3- Não se pode afirmar a existência de diferenças entre os grupos (Teixeira J. , Fernandes, Teixeira, & Odete, 2018)

Após uma abordagem teórica sobre o assunto, fez-se uma análise para todos os casos, ou seja, para todas as patologias, todos os parâmetros, vogais e tons, contudo apenas serão apresentados os resultados que mais se evidenciaram.

Para o parâmetro *jitter* relativo, a patologia, paralisia das cordas vocais, por norma, apresenta valores diferentes dos restantes grupos em estudo, sendo, apenas, possível afirmar que os valores são, maiores que o grupo de controlo. A figura 3.8 demonstra a análise comparativa entre as diferentes categorias para o parâmetro *Jitter* relativo da vogal ‘a’ no tom alto.

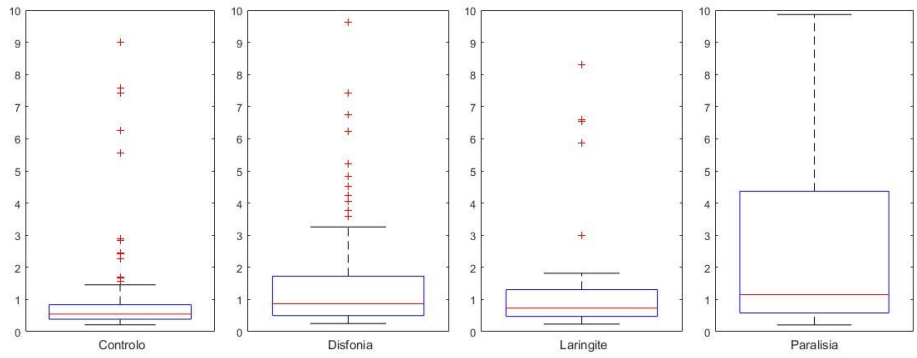

Figura 3.8- *Jitter* /a/ alto

A tabela 3.2, resume o estudo estatístico referente ao parâmetro jitter relativo.

Tabela 3.2- Estudo estatístico referente ao parâmetro: Jitter Relativo

Parâmetro:	**Jitter Relativo**	
Caso geral:	Paralisia das cordas vocais apresenta, possivelmente valores maiores em relação aos restantes grupos	
Casos em destaque		
Vogal	Tom	Descrição
\a	Alto e baixo	Paralisia apresenta valores possivelmente maiores que Controlo.
\a	baixo	Disfonia apresenta valores possivelmente maiores que Controlo.
\u	normal	Paralisia apresenta valores possivelmente maiores que Laringite e Controlo.
\i	(alto, baixo e normal)	Independentemente do tom utilizado, indica a probabilidade de existirem diferenças entre categorias.

Para o parâmetro *Shimmer* relativo, de um modo geral, os sujeitos com paralisia das cordas vocais, também, apresentam valores, possivelmente, maiores que os sujeitos de controlo. A figura 3.9 representa uma comparação entre as diferentes categorias para o parâmetro *Shimmer* relativo da vogal 'u' no tom alto.

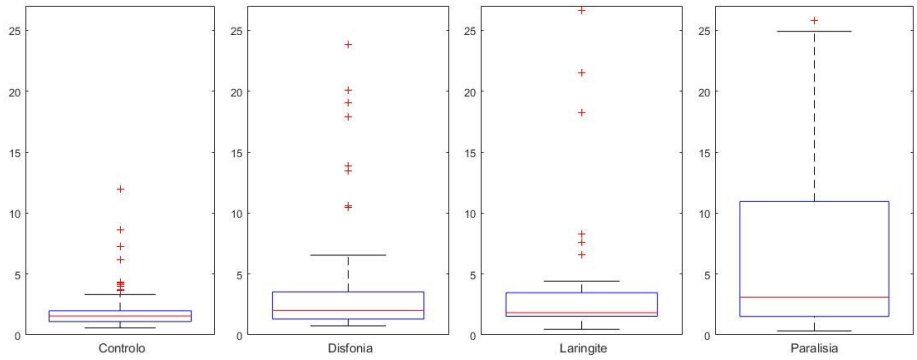

Figura 3.9- *Shimmer* relativo /u/ alto

A tabela 3.3, resume o estudo estatístico referente ao parâmetro Shimmer relativo.

Tabela 3.3- Estudo estatístico referente ao parâmetro: Shimmer Relativo

Parâmetro:	**Shimmer Relativo**	
Caso geral:	Paralisia das cordas vocais apresenta, possivelmente valores maiores em relação ao grupo de controlo.	
Casos em destaque		
Vogal	Tom	Descrição
\a\ \u	alto	Disfonia apresenta valores possivelmente maiores que Controlo.
\a	normal	Laringite apresenta valores possivelmente maiores que Controlo.

Relativamente ao parâmetro HNR, os valores dos sujeitos saudáveis são, possivelmente, maiores que os restantes sujeitos.

A análise da vogal /a/, independentemente do tom, não traz uma vantagem clara na distinção da patologia disfonia. Para a vogal /i/ tom alto, os sujeitos de controlo ou com disfonia apresentam valores, possivelmente, maiores relativamente aos sujeitos com laringite crónica, demonstrado na figura 3.10.

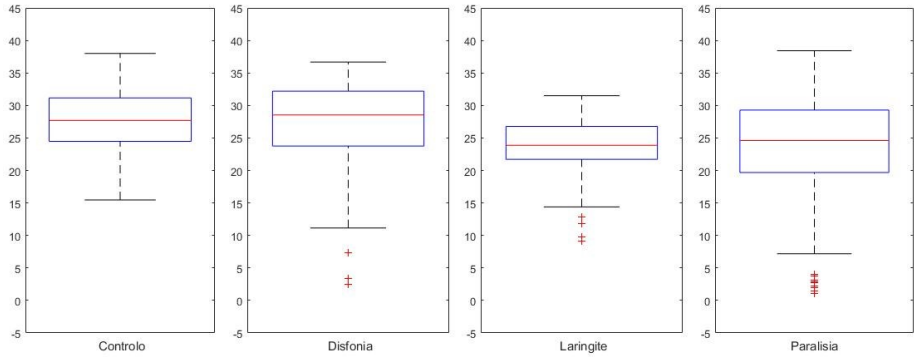

Figura 3.10- HNR /i/ alto

A tabela 3.4, resume o estudo estatístico referente ao parâmetro HNR.

Tabela 3.4- Estudo estatístico referente ao parâmetro: HNR

Parâmetro:	**HNR**	
Caso geral:	O grupo saudável apresenta, possivelmente valores maiores em relação aos restantes grupos.	
Casos em destaque		
Vogal	Tom	Descrição
\i	alto	Controlo e Disfonia apresentam valores possivelmente maiores que Laringite.
\u	baixo	Disfonia e paralisia apresentam valores possivelmente inferiores que Controlo.

No parâmetro NHR, os sujeitos patológicos apresentam valores, possivelmente, maiores caso sejam comparados com sujeitos de controlo.

Também neste parâmetro, a paralisia das cordas vocais apresenta uma diferença, na maioria dos tons.

Pode-se afirmar que a utilização do parâmetro NHR, utilizando a vogal /a/, possivelmente, permite fazer uma distinção entre sujeitos patológicos e saudáveis. A figura 3.11 representa a análise feita com o parâmetro NHR, para a vogal ‘a’ no tom

baixo, onde se pode confirmar o mencionado anteriormente. A tabela 3.5, resume o estudo estatístico referente ao parâmetro NHR.

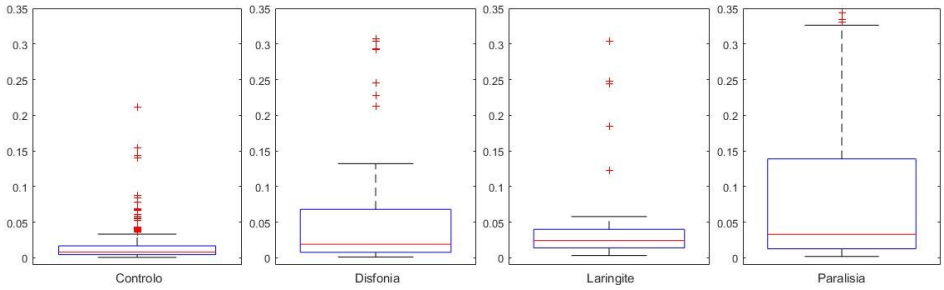

Figura 3.11- NHR /a/ baixo

Tabela 3.5- Estudo estatístico referente ao parâmetro: NHR

Parâmetro:	**NHR**	
Caso geral:	Os grupos patológicos apresentam, possivelmente valores maiores em relação ao grupo saudável.	
Casos em destaque		
Vogal	Tom	Descrição
\a	(Alto, baixo e normal)	Grupos patológicos apresentam valores possivelmente maiores que grupo saudável.
\u	baixo	Laringite apresenta valores possivelmente maiores que Controlo.

No parâmetro Autocorrelação, a vogal /a/, possivelmente, permite uma distinção entre sujeitos patológicos e saudáveis, uma vez que os valores associados a sujeitos de controlo são, provavelmente, maiores que os dos sujeitos patológicos (isto porque o parâmetro NHR está diretamente relacionado com a autocorrelação). Com recurso à figura 3.12 pode-se ter uma melhor perceção da diferença entre sujeitos saudáveis e patológicos para esta vogal. A tabela 3.6, resume o estudo estatístico referente ao parâmetro Autocorrelação.

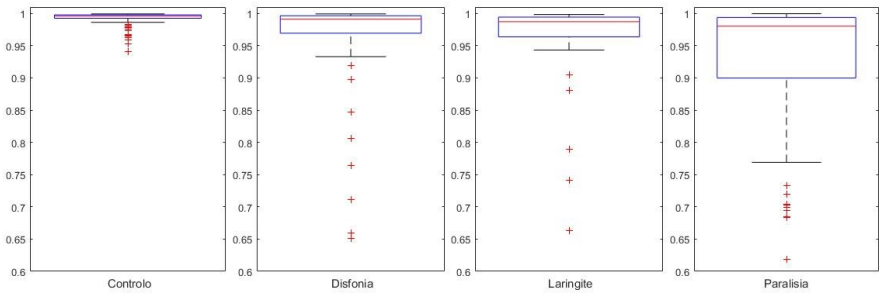

Figura 3.12- Autocorrelação /a/ alto

Tabela 3.6- Estudo estatístico referente ao parâmetro: Autocorrelação

Parâmetro:	**Autocorrelação**	
Caso geral:	O grupo de controlo apresenta, possivelmente valores maiores em relação aos grupos patológicos.	
Casos em destaque		
Vogal	**Tom**	**Descrição**
\a	alto	Grupos patológicos apresentam valores possivelmente inferiores ao grupo saudável.
\i	(alto, baixo e normal)	Paralisia e Laringite apresenta valores possivelmente inferiores a Controlo.
\u	Alto, normal	Paralisia apresenta valores possivelmente inferiores em relação a Controlo.

Perante a análise realizada, conclui-se que existem parâmetros, que de forma independente, podem levar a taxas de acerto elevadas, no que diz respeito à identificação entre sujeitos patológicos e sujeitos saudáveis.

Pode-se apurar que as vogais com informação mais relevantes são a vogal /a/ e a vogal /i/. Em relação aos tons, viu-se que todos eles podem ter informações distintas. Sujeitos portadores de paralisia das cordas vocais apresentam, na maior parte dos casos, fatores notórios que permitem a distinção relativamente a outras categorias.

A categoria de disfonia e de laringite crónica contém, na maioria dos casos, resultados semelhantes, o que se pode concluir que a análise individual dos parâmetros para a distinção destas duas classes pode não ser muito promissora.

Na sua generalidade, repara-se que não existe nenhum parâmetro, independentemente da vogal ou do tom, que permita fazer uma distinção óbvia entre as diferentes categorias.

No entanto, a utilização de todos os parâmetros em simultâneo, aliados a uma ferramenta de *machine learning* pode trazer resultados promissores, no que diz respeito à distinção de patologias.

4.Modelos em estudo

Durante esta secção é feita uma abordagem teórica referente a *machine learning* e aos diferentes tipos de ferramentas de *machine learning,* utilizadas nesta dissertação.

4.1. Ferramentas de Machine Learning

Machine Learning, (em português, aprendizagem automática) é um conceito cada vez mais implementado nos dias que correm.

Deixar que um computador aprenda de forma autónoma, por exemplo, na deteção de doenças, analisando uma base de dados com sintomas e respetivos resultados de doentes anteriores, revelou-se mais fácil do que fazer entrevistas a médicos (Domingos, 2017).

O *Machine Learning* desenvolveu-se através da evolução da inteligência artificial, no entanto, uma abordagem lógica, baseada em conhecimento, causou a separação entre a inteligência artificial e *Machine learning*.

Este surge como um campo separado por volta dos anos 90, deixando, assim, o seu objetivo de encontrar soluções para problemas com solução de natureza prática através da inteligência artificial. Passou então o seu foco para métodos e modelos da estatística e da teoria da probabilidade (Langley, 2001) .

4.2. Tipos de *Machine Learning*

Diversas técnicas de *Machine Learning* foram desenvolvidas, a fim de resolver diversos problemas em variadas áreas.

Estas técnicas podem ser classificadas em três tipos, como se pode observar na figura 4.1.

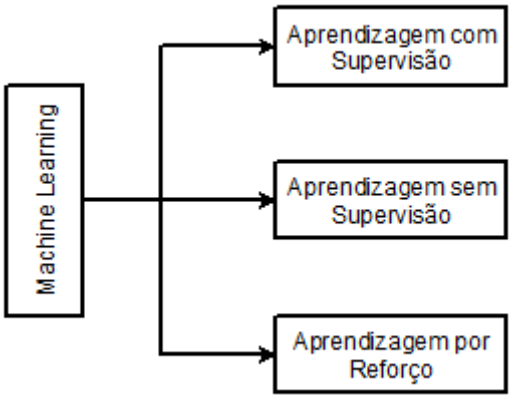

Figura 4.1– Técnicas de *Machine Learning*

4.2.1. Elementos de *Machine Learning*

Para o funcionamento de qualquer sistema de *Machine Learning* são envolvidos três elementos.

Dados

Todos os métodos de aprendizagem envolvem a utilização de dados e neste tipo de processo utiliza-se um conjunto de dados para que se possa treinar o sistema.

Os dados utilizados no decorrer desta dissertação, são os referidos no capítulo 3.

Organizaram-se os dados, em três grupos de parâmetros, no grupo I (a) os parâmetros correspondem ao *Jitter* Relativo, *Shimme*r Relativo e HNR das três vogais (/a/, /i/ e /u/) em três tons diferentes (alto, baixo e normal). O grupo I (b) contém os parâmetros do grupo I (a) em conjunto com os parâmetros NHR e Autocorrelação. No grupo II, os parâmetros são os coeficientes MFCCs das vogais, usadas no grupo I. No grupo III, os parâmetros correspondem aos coeficientes MFCCs, extraídos da fala continua (frase).

Mais informações referentes aos grupos elaborados, são explicadas no capítulo 5.

Modelos

Os modelos são formulados através de estruturas matemáticas para aprender. São baseados em observações e experiências humanas, no entanto, alguns métodos de aprendizagem automática desenvolvem os seus próprios modelos sem ser necessária intervenção humana (Michael Paluszek, 2017).

São várias as possibilidades de modelos que se podem utilizar com o *Machine Learning*. Na figura 4.2 é representado alguns dos modelos mais influentes.

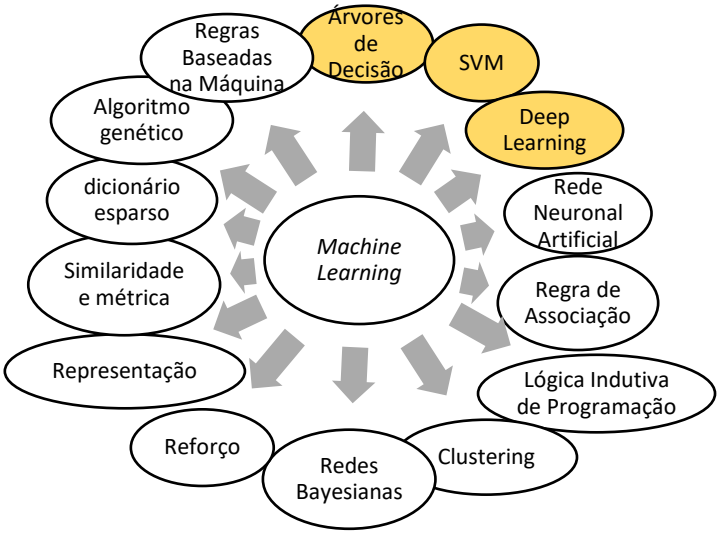

Figura 4.2- Modelos *de Machine Learning*

Treino

Qualquer sistema que através de uma entrada mapeia uma saída, necessita de ser treinado para efetuar essa tarefa de maneira útil.

O treino é realizado indicando ao sistema dados de entrada e a sua correspondente saída. Quanto maior for a diversidade de soluções, melhor será o treino efetuado. Caso o número de exemplos aplicado no treino seja suficiente, o sistema deve ser capaz de indicar saídas corretas quando novas entradas forem introduzidas (Michael Paluszek, 2017).

Existem diferentes tipos de algoritmos de treino desenvolvidos, como por exemplo Backpropagation ou Regra da Aprendizagem Delta.

Aprendizagem supervisionada

A aprendizagem supervisionada, utiliza um conjunto de características, em que cada exemplo está associado a uma categoria.

Com prévio conhecimento adquirido no processo de treino, é esperado que sejam previstas as categorias de novos dados nunca antes demonstrados ao sistema. (Ian Goodfellow, 2016)

Os problemas relacionados com aprendizagem supervisionada podem ser agrupados em dois tipos, classificação, onde o algoritmo classifica uma previsão, ou seja, quando a saída é uma categoria. Ou então, regressão, onde o algoritmo tenta prever um resultado com base nas variáveis anteriores (Brownless, 2016).

Exemplos de alguns algoritmos supervisionados de Machine Learning:

- Support Vector Machines (SVM) para problemas de classificação e regressão;

- Redes neuronais artificiais (ANN) para problemas de classificação e regressão;
- Regressão Linear, para problemas de regressão;
- K-Nearest Neighbors (kNN) para problemas de classificação e regressão.

Aprendizagem sem supervisão

Na aprendizagem sem supervisão utiliza-se um conjunto de dados com muitas características, em seguida, os algoritmos, aprendem as propriedades úteis desse mesmo conjunto de dados, geralmente, usa-se para descobrir padrões em dados para os quais não existe uma resposta conhecida. A utilização deste tipo de supervisão tem a vantagem de aprender coisas sobre os dados dos quais não há qualquer tipo de informação antecipada (Ian Goodfellow, 2016).

Os problemas relacionados com aprendizagem sem supervisão, podem ser agrupados em dois tipos, clustering, onde se pretende descobrir os agrupamentos inerentes aos dados, ou também, associação, onde na grande quantidade de dados se tenta aprender regras de associação (Brownless, 2016).

Exemplos de alguns algoritmos sem supervisão de Machine Learning:

- k-means, para problemas de clustering;
- À priori para problemas de aprendizagem de regras de associação.

Aprendizagem com semi-supervisão

Como o próprio nome indica, apenas alguns dados, em minoria, estão supervisionados. A maioria dos dados não é supervisionada, porque pode ser um processo intensivo, que requer um ser humano com qualificações específicas. A pequena quantidade de dados supervisionados serve de alavanca para os restantes dados (Michael Paluszek, 2017).

Algoritmo backpropagation

O desenvolvimento do algoritmo *backpropagation* defende que é possível "treinar" as camadas intermédias de uma ferramenta de *machine learning* (redes neuronais, por exemplo).

Cada neurónio de uma camada recebe os sinais de todos os neurónios da camada anterior e propaga os seus dados de entrada a todos os neurónios da camada posterior.

A principal vantagem de utilizar o "*backpropagation*" consiste em trabalhar em multi-camadas, que permite solucionar problemas em que a informação não é separável linearmente, e outros algoritmos são incapazes de solucionar. Outra característica

associada ao *backpropagation* é o facto de ser *feedforward*, isto significa que a conexão entre neurónios não é cíclica (Lanhellas, 2013) (Peres, 2017).

Os neurónios guardam os valores a serem calculados para definir o valor dos pesos, pois é pelo peso que a RNA consegue identificar a resposta dada (Lanhellas, 2013).

O método *backpropagation* é um algoritmo supervisionado. Os algoritmos supervisionados necessitam de um vetor de entrada e de um vetor de saída (também conhecido como vetor alvo (*target*)). Quando o vetor de entrada é aplicado, a saída é calculada e comparada com o vetor alvo correspondente. O erro encontrado é, então, reassumido pela rede neuronal e os pesos são atualizados de forma a minimizar o erro. Este processo é repetido até que o erro dos vetores seja correspondente ao valor pré-definido para cada conjunto (Peres, 2017).

Ou seja, este algoritmo tem duas fases distintas (também conhecidas como a fase *Forward* – em que é definida a saída da rede para um dado padrão de entrada, e a fase *Backward* – em que a diferença entre o resultado desejado e o resultado obtido é utilizada para atualizar os pesos das conexões). Resumindo, este algoritmo ajusta os valores dos pesos de maneira a que o erro seja minimizado (Haykin S. , 2001).

As funções de treino implementadas nos experimentos práticos são melhorias realizadas no método de *backpropagation.*

Levenberg-Marquardt backpropagation

Esta função de treino foi implementada recorrendo a função "trainlm", por definição é a função de treino padrão do Matlab, esta função atualiza os valores dos pesos e bias seguindo o conceito da otimização Levenberg-Marquardt. Esta função de treino consiste num algoritmo supervisionado, utiliza o Jacobiano para cálculos, onde é suposto que o desempenho seja uma média ou soma de erros ao quadrado (MathWorks, trainlm, 2019).

Scaled Conjugate Gradient backpropagation

Esta função de treino foi implementada recorrendo a função "trainscg", esta função atualiza os valores dos pesos e bias de acordo com o método gradiente conjugado em escala (MathWorks, trainscg, 2019).

4.3. Árvores de Decisão

As árvores de decisão têm uma diversa aplicação em diferentes campos do *Machine Learning*.

Uma árvore de decisão consiste num mapa dos possíveis resultados de uma série de escolhas relacionadas. É possível, através de uma árvore de decisão, comparar ações com base nos custos, probabilidades e benefícios. Podem ser utilizadas desde conduzir diálogos informais, até algo mais complexo como, por exemplo, mapear um algoritmo, que se supõe que seja a melhor escolha, matematicamente (Lucidchart, 2018).

As árvores de decisão (figura 4.3) podem ser interpretadas como sendo uma resposta à pergunta de como se deve atuar quando as regras de mais do que um conceito se aplica a uma instância.

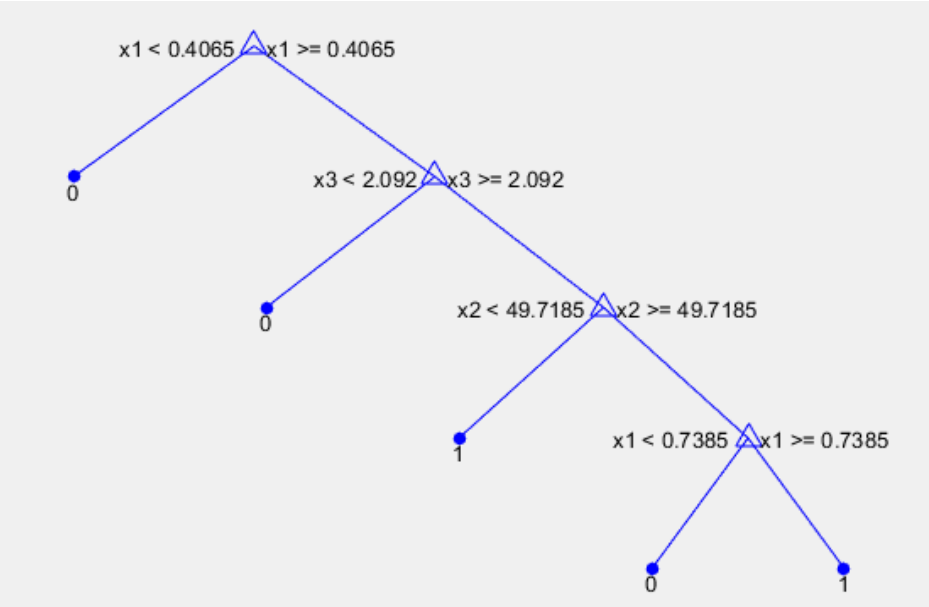

Figura 4.3- Árvore de decisão para identificação de sujeitos saudáveis ou patológicos

As árvores de decisão garantem, à partida, que para cada instância se aplicará exatamente uma regra. Um conceito define duas classes, o conceito em si e a sua negação (Domingos, 2017).

Geralmente, uma árvore de decisão inicia-se com um único nó, que se divide em resultados possíveis. Cada um desses resultados conduz a nós adicionais, que se ramificam em outras possibilidades, criando, assim, a forma de uma árvore (Lucidchart, 2018).

Cada nó interno da árvore corresponde a um teste do valor de uma propriedade. Os ramos dos nós são rotulados com os possíveis resultados do teste (Lima, 2012).

As árvores de decisão podem ser classificadas em dois tipos:

- Árvores de classificação – produzem saídas por categorias;
- Árvores de regressão – produzem saídas numéricas.

4.4. Support Vector Machine

Entende-se por SVM's (do inglês, *Support Vector Machine*) como sendo um conceito utilizado para a análise de grandes quantidades de números e reconhecimento de padrões. São aplicados para a classificação e análise de regressão, que maximiza a exatidão preditiva de um modelo, sem causar ajuste dos dados de treino (Sobre o SVM, 2015).

Os SVM's têm características que os tornam mais atrativos, tais como, boa capacidade de generalização; robustez em grandes dimensões; convexidade da função objetivo e uma teoria bem definida dentro da matemática e estatística (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

O SVM opera quando mapeia dados para um espaço variável altamente dimensional. Na figura 4.4 é possível observar o conjunto de dados original, na qual os pontos de dados se situam em duas categorias diferentes.

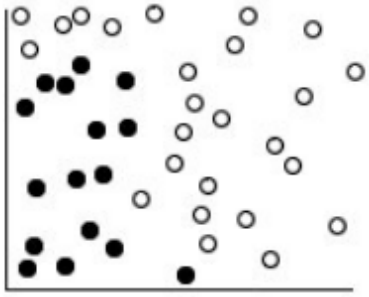

Figura 4.4- Conjunto de dados original (Como o SVM Funciona, 2015)

Os pontos de dados podem ser atribuídos a uma categoria, mesmo quando os dados não são linearmente separáveis. Na figura 4.5 pode-se observar as duas categorias separadas por uma curva.

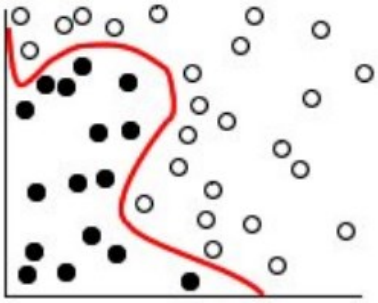

Figura 4.5- Dados com separador incluído (Como o SVM Funciona, 2015)

Assim que um separador entre as categorias é encontrado, os dados são transformados para que o separador possa ser desenhado como um hiperplano. Na figura 4.6 observar-se o limite entre as duas categorias definido por um hiperplano.

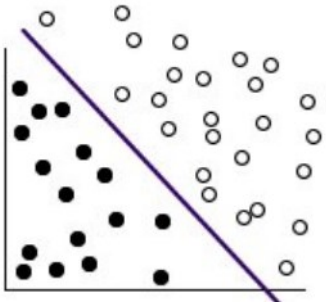

Figura 4.6- Dados separados por hiperplano (Como o SVM Funciona, 2015)

Sempre que é adicionado um novo dado, as suas características são utilizadas para prever a qual grupo deve pertencer.

O primeiro modelo SVM a ser introduzido é chamado de Classificador de Margem Máxima e trabalha com dados linearmente separáveis, o que limita as suas aplicações práticas. Contudo, este modelo, apresenta propriedades importantes e é fundamental para a formulação de SVM's mais sofisticados (Meloni, 2009). Este classificador é geralmente usado quando num espaço de ordem *N*, um classificador linear não tem uma generalização razoável, ou seja, classifica bem o conjunto de treino, mas depois não consegue generalizar eficazmente para o conjunto de teste (Cordeiro H. T., 2016).

Os SVM pertencem à categoria de classificadores lineares, pois induzem separadores lineares ou hiperplanos, quer no espaço original dos exemplos de entrada, se forem separáveis ou quase separáveis (ruído), ou em um espaço transformado (espaço de recursos), se os exemplos não forem linearmente separáveis no espaço original.

A procura do hiperplano de separação, através de um dos espaços transformados, é feita de forma implícita, através das funções de *Kernel* (Suárez, 2014).

SVM para classificação binária de exemplos separados linearmente

Dado um conjunto separável de exemplos $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, onde $x_i \in \mathbb{R}^d$ e $y_i \in \{+1, -1\}$, pode-se definir um hiperplano de separação, figura 4.7, como uma função linear, que é capaz de separar este conjunto sem erro, equação 15 (Suárez, 2014).

$$D(x) = (w_1 x_1, \dots, w_d x_d) + b =< w, x > +b \tag{15}$$

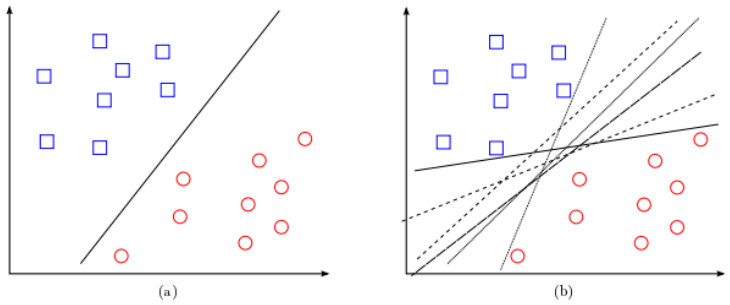

Figura 4.7- Hiperplanos de separação num espaço bidimensional de um conjunto de exemplos separáveis em duas classes: (a) exemplo de hiperplano de separação (b) outros exemplos de hiperplanos de separação entre os vários possíveis (Suárez, 2014)

Onde w e b são coeficientes reais. O hiperplano de separação cumprirá com as seguintes restrições para todo o x_i do conjunto de exemplos:

$$\begin{aligned} (w, x_i) + b \geq 0 \quad & se \quad y_i = +1 \\ (w, x_i) + b \leq 0 \quad & se \quad y_i = -1, \;\; i = 1, \dots, n \end{aligned} \tag{16}$$

ou também,

$$y_i((w, x_i) + b) \geq 0, \quad i = 1, \dots, n \tag{17}$$

ou de forma mais compacta

$$y_i D(x_i) \geq 0, \quad i = 1, \dots, n \tag{18}$$

Como se pode deduzir através da figura 4.7, o hiperplano que permite separar os exemplos não é único, ou seja, existem infinitos hiperplanos separáveis, representados por todos aqueles hiperplanos que são capazes de cumprir as restrições impostas por

qualquer uma das expressões equivalentes (16-18). Assim sendo, surge a necessidade de estabelecer um critério adicional, que permita definir um hiperplano de separação ótima (Suárez, 2014).

Hiperplanos de Separação Ótima

Designa-se por hiperplano de margem máxima (ou de separação ótima) quando um conjunto de vetores é separado sem erros e a distância entre os vetores (das classes opostas) mais próximos ao hiperplano é máxima. Na figura 4.8 (a) é possível observar um hiperplano com margem pequena, enquanto que, na figura 4.8 (b) observa-se um hiperplano de separação ótima, para um conjunto de treino bidimensional (Meloni, 2009).

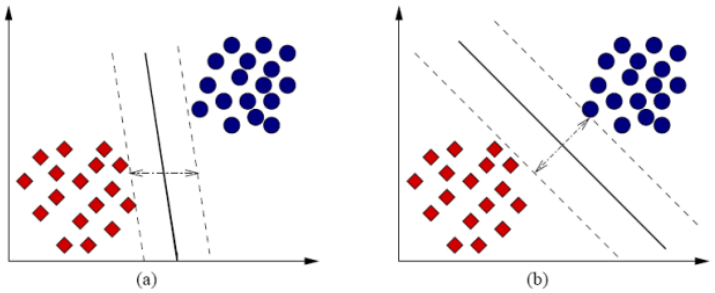

Figura 4.8- (a) Um hiperplano de separação com margem pequena. (b) Um hiperplano de separação ótima (Meloni, 2009)

Para haver um hiperplano de separação ótima é necessário que se defina o conceito de margem de um hiperplano de separação, designado por τ, como a distância mínima entre o referido hiperplano e os exemplos mais próximos de qualquer uma das classes (ver figura 4.8 (a)). Partindo desta definição, designa-se de hiperplano ótimo, se a sua margem for de tamanho máximo (ver figura 4.8 (b)). Existem propriedades que levam, imediatamente, a definir um hiperplano de separação ideal, ou seja, tem de ser equidistante do exemplo mais próximo de cada classe. Supondo que a distância do hiperplano ideal, para o exemplo mais próximo da classe +1 é menor que a correspondente ao exemplo mais próximo da classe -1, significa que o hiperplano da classe +1 pode ser movido, de modo, a que fique maior do que estava anteriormente na classe +1 e permaneça menor do que a distância ao exemplo mais próximo da classe -1 (Suárez, 2014).

A distância de qualquer exemplo x_i ao hiperplano de separação ótima é dada por $\frac{|D(x_i)|}{\|w\|}$. Se esse exemplo pertence ao conjunto de vectores de suporte (identificado por cor sólida), a distância ao hiperplano é sempre $\frac{1}{\|w\|}$.

Além disso, os vetores de suporte aplicados à função de decisão cumprem sempre $|D(x)| = 1$. Na figura 4.9 é possível observar um hiperplano de separação ótima.

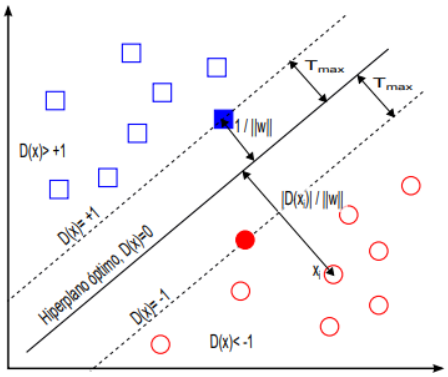

Figura 4.9- Hiperplano de separação ótima (Suárez, 2014)

De acordo com a definição, um hiperplano ótimo é aquele que tem uma margem máxima, ou seja, um valor mínimo de $\|w\|$. Para tal utiliza-se a equação 19 (Suárez, 2014).

$$y_i\big((w, x_i) + b\big) \geq 1, \quad i = 1, \ldots, n \tag{19}$$

O conceito de margem máxima está diretamente relacionado com a capacidade de generalização do hiperplano de separação, de tal forma que, quanto maior for a margem, maior distância de separação existirá entre as duas classes (Suárez, 2014).

SVM para classificação binária de exemplos quase separáveis linearmente

A maior parte dos problemas reais caracterizam-se por terem exemplos ruidosos e não serem perfeita e linearmente separáveis.

Para este tipo de problemas reais, a estratégia é "relaxar" o grau de separação do conjunto de exemplos, permitindo erros de classificação em alguns dos exemplos do conjunto de treino. Porém, o objetivo continua a ser encontrar um hiperplano ótimo para os restantes exemplos separáveis (Suárez, 2014).

Um exemplo é não separável se não cumprir a condição imposta na equação 19. Neste caso, podem acontecer duas situações, uma, em que o exemplo está dentro da margem associada à classe correta, tendo em conta o limite de decisão que define o hiperplano de separação. Na outra situação, o exemplo fica do outro lado do hiperplano. Nestas duas situações define-se o exemplo como não separável, porém, na primeira situação o exemplo foi classificado corretamente, enquanto que, na segunda situação não. Na figura 4.10 é possível observar este exemplo.

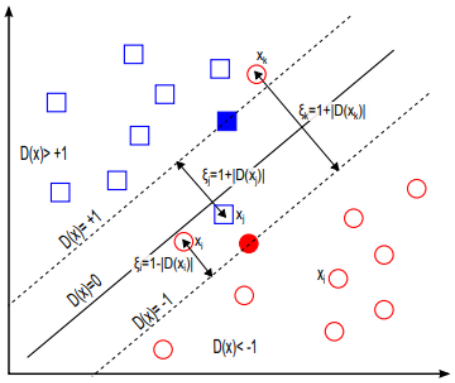

Figura 4.10- Caso de exemplos não separáveis

No caso de exemplos não separáveis, as variáveis de folga medem o desvio da margem da classe respetiva. Assim, os exemplos x_i, x_j e x_k são, cada um deles, não selecionáveis ($\xi_i, \xi_j, \xi_k > 0$). Porém, x_i está corretamente classificado, enquanto x_j e x_k estão no lado errado do limite de decisão e, portanto, mal classificada (Suárez, 2014).

Para abordar este novo problema, é necessário que na equação 19 seja introduzido um conjunto de variáveis reais positivas, denominadas de variáveis de folga, ξ_i, $i=1,\ldots,n$, que permitirá quantificar o número de exemplos não separáveis, que se está disposto a admitir (equação 20) (Suárez, 2014).

$$y_i\big((w, x_i) + b\big) \geq 1 - \xi_i, \qquad \xi_i \geq 0 i = 1, \ldots, n \tag{20}$$

Para um exemplo (x_i, y_i), a variável de folga, ξ_i, representa o desvio do caso separável, medido desde a borda da margem que corresponde à classe y_i. Tendo em conta esta definição, as variáveis de folga de valor zero, correspondem a exemplos separáveis, enquanto que maiores que zero correspondem a exemplos não separáveis, e valores maiores que um correspondem a exemplos não separáveis mal classificados. Ou seja, a soma de todas as variáveis de folga, $\sum_{i=1}^{n} \xi_i$, permite, de alguma maneira, medir a quantidade de exemplos associados não-separáveis (Suárez, 2014).

A equação 20 não é suficiente para criar o objetivo de maximizar a margem, pois, esta classifica erradamente muitos exemplos. Para que haja uma classificação mais correta, a função de otimização deve incluir os erros de classificação que o hiperplano de separação está a cometer (equação 21) (Suárez, 2014).

$$f(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \qquad (21)$$

Onde *C* é uma constante, suficientemente grande, definida pelo utilizador, que permite controlar em que grau influência a quantificação de exemplos não separáveis. Então, um valor *C* elevado, permitiria valores de ξi muito pequenos. No limite ($C \to \infty$), seriam considerados exemplos perfeitamente separáveis ($\xi i \to 0$). Por outro lado, um valor *C* muito pequeno permitiria valores de ξi elevados, ou seja, admitia-se muitos exemplos mal classificados. No caso limite ($C \to 0$), todos os exemplos permitidos são mal classificados ($\xi i \to \infty$) (Suárez, 2014).

Por isso, o novo problema de otimização consiste em encontrar um hiperplano, definido por *w* e *b* que minimize a função (eq.21), sujeito às restrições dadas pela equação 10. Este hiperplano designa-se por hiperplano de separação de margens suaves (Suárez, 2014).

SVM para classificação binária de exemplos não separáveis linearmente

Os SVM's descritos anteriormente são eficazes na classificação de dados linearmente separáveis ou quase separáveis, isto faz com que sejam limitados, uma vez que, existem muitos casos em que não é possível dividir linearmente os dados. Na Figura 4.11 é apresentado um exemplo em que a utilização de uma fronteira curva seria mais adequada para separar as duas classes. Essa fronteira é definida pela Eq.22 (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

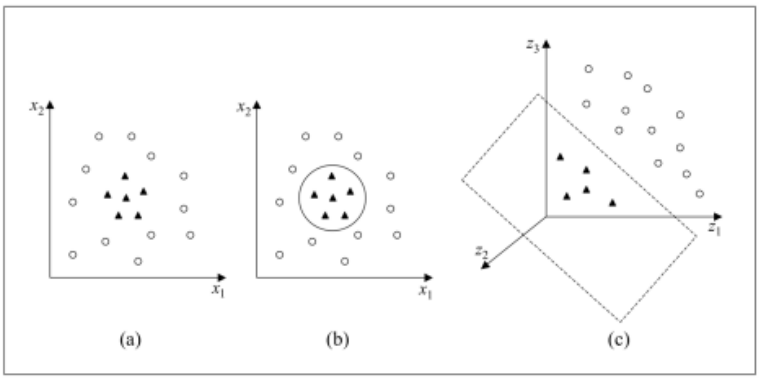

Figura 4.11- (a) Conjunto de dados não lineares; (b) Fronteira não Linear; (c) Fronteira linear no espaço de características (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007)

$$x_1^2 + x_2^2 < r^2 \tag{22}$$

Os SVM's enfrentam problemas não lineares ao mapearem o conjunto de treino do espaço original, definido como espaço de entradas, para um novo espaço chamado de espaço de características. Se $\Phi : X \rightarrow \mathfrak{I}$ for um mapeamento, onde X é o espaço de entradas e $\mathfrak{I}$ corresponde ao espaço de características (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

O teorema de Cover motiva este procedimento, uma vez que dado um conjunto de dados não linear no espaço de entradas X, este pode ser transformado em um espaço de características $\mathfrak{I}$, no qual, com uma probabilidade elevada, os dados são linearmente separáveis. Porém, é necessário que duas condições sejam satisfeitas. É necessário que a transformação seja não linear e que a dimensão do espaço de características seja suficientemente alta (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

Considerando o conjunto de dados da Figura 4.11 (a), e transformando os dados de $\mathbb{R}^2$ para $\mathbb{R}^3$ através do mapeamento representado na Equação 23, o conjunto de dados não linear em $\mathbb{R}^2$passa a ser linearmente separável em $\mathbb{R}^3$(Figura 4.11(c)). Deste modo, é possível encontrar um hiperplano capaz de separar os dados através da equação 24. Apesar de esta equação ser linear em $\mathbb{R}^3$, ela corresponde a uma fronteira não linear em $\mathbb{R}^2$ (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

$$\Phi(x) = \Phi(x_1, x_2) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right) \tag{23}$$

$$f(x) = w.\Phi(x) + b = w_1x_1^2 + w_2\sqrt{2}x_1x_2 + w_3x_2^2 + b = 0 \tag{24}$$

Inicialmente mapeiam-se os dados para um espaço de maior dimensão através de Φ e a este espaço aplica-se a SVM linear. Assim, é encontrado o hiperplano com maior margem de separação, o que garante uma boa generalização. O SVM a utilizar é o SVM linear com margens suaves, uma vez que permite lidar com ruídos e *outliers* presentes nos dados (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

A única informação necessária sobre o mapeamento Φ é uma definição de como o produto interno $\Phi(x_i).\Phi(x_j)$ pode ser realizado, para quaisquer x_i e x_j pertencentes ao espaço de entradas. Isto obtém-se com a introdução do conceito de *Kernels* (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

Funções de *Kernel*

Um *Kernel K* é uma função que recebe dois pontos x_i e x_j do espaço de entrada e calcula o produto escalar desses dados no espaço de características, como descrito na Equação 25 (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

$$K(x_i, x_j) = \Phi(x_i).\Phi(x_j) \tag{25}$$

Para o mapeamento apresentado na Equação 25 e dois dados $x_i = (x_{1i}, x_{2i})$ e $x_j = (x_{1j}, x_{2j})$ em $\mathbb{R}^2$, por exemplo, o *Kernel* é dado pela Equação 26 (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

$$\mathrm{K}(x_i, x_j) = (x_{1i}^2, \sqrt{2}x_{1i}x_{2i}, x_{2i}^2).(x_{1j}^2, \sqrt{2}x_{1j}x_{2j}, x_{2j}^2) = (x_i.x_j)^2 \tag{26}$$

Geralmente a função de *Kernel* é usada sem se conhecer o mapeamento Φ, que é gerado implicitamente. Os *Kernels* são úteis, uma vez que, o seu cálculo é bastante simples e têm capacidade de representar espaços abstratos (Lorena & Carvalho, Uma Introdução às Support Vector Machines, 2007).

Alguns dos *Kernels* mais utilizados são os Gaussianos ou RBF (*Radial- Basis Function*) e os Sigmoidais, apresentados na tabela 4.1.

Tabela 4.1- Principais Kernels nos SVM´s (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003)

Tipo de Kernel	Função K(xi,xj) correspondente	Comentários
Polinomial	$(Xi^T.Xj+1)^P$	Onde P deve ser especificada pelo utilizador
Gaussiano	$\exp\left(-\frac{1}{2\sigma^2}\|Xi-Xj\|^2\right)$	A amplitude σ^2 é especificada pelo utilizador
Sigmoidal	$\tanh(\beta 0 Xi.Xj+\beta 1)$	Utilizado apenas para alguns valores de $\beta 0$ *e* $\beta 1$

A escolha de um classificador através do uso de SVM's envolve a escolha de uma função *Kernel*, além de parâmetros desta função e do algoritmo para determinação do hiperplano ótimo (como o valor da constante *C*, por exemplo). A escolha do *Kernel* e dos parâmetros considerados tem efeito no desempenho do classificador obtido, pois eles definem a fronteira de decisão induzida. O modelo pode ser selecionado através de várias técnicas, que resultam de meios para determinação da função *Kernel* e parâmetros do algoritmo (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

4.4.1. SVM's para várias classes

Para além das classes referidas anteriormente, onde se definem em duas classes, sendo uma positiva e outra negativa, existem muitas aplicações que envolvem o agrupamento em mais que duas classes. Este fator não torna as SVM's inutilizáveis, uma vez que são propostas diversas técnicas para este tipo de problemas (multiclasses).

Num problema de multiclasses, o conjunto de treino é composto por pares (x_i, y_i), tal que, $y_i \in 1, \dots, k$, sendo $k > 2$ (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

Qualquer método para gerar classificadores multiclasses, partindo de preditores binários, pode utilizar SVM's como base (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003). De seguida são apresentadas duas

abordagens usadas geralmente para realizar a tarefa de classificação de multiclasses. São denominadas, respetivamente, decomposição "um-contra-todos" e "todos-contra-todos".

4.4.1.1. Decomposição "um-contra-todos"

Uma abordagem usual em problemas de multiclasse consiste na geração de *k* SVM's, onde *k* é o número de classes. Neste tipo de problemas, uma classe é fixada como positiva e as restantes como negativas. Dá-se o nome de decomposição "um-contra-todos" (1-c-t) a esta metodologia, e é independente do algoritmo de aprendizagem utilizado no treino dos classificadores. Na predição da classe de um padrão *x*, só é necessário escolher a saída com valor máximo entre as *k* SVM's, conforme a Equação 27 (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

$$f(x) = arg \max_{1 \le i \le k} (w_i . \Phi(x) + b_i) \tag{27}$$

Este método tem a desvantagem de não ser possível prever limites no erro de generalização através do seu uso. Além disso, o tempo de treino é normalmente longo (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

4.4.1.2. Decomposição "todos-contra-todos"

Para a solução de multiclasses existe uma outra abordagem, que a partir de classificadores binários envolve a construção de $k(k-1)/2$ SVM's, onde cada classe é separada uma da outra. A este método dá-se o nome de "todos-contra-todos" (t-c-t). De modo a unir estes classificadores, *Friedman* (1996) propôs usar um esquema de votação por maioria, onde cada um dos classificadores fornece uma classe como resultado, sendo a solução final a classe que recebeu mais indicações. Esta metodologia, contudo, também não prevê limites no erro de generalização e o tamanho dos classificadores gerados é em geral grande e a avaliação dos resultados pode ser lenta (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

Para a resolução destes problemas, Platt el al. (2000) sugerem a utilização de um grafo direcionado acíclico (DAG). Deste modo, o problema de agrupamento em várias classes é decomposto em diversas classificações binárias em cada nó do grafo. Este processo pode ser observado na Figura 4.12 (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

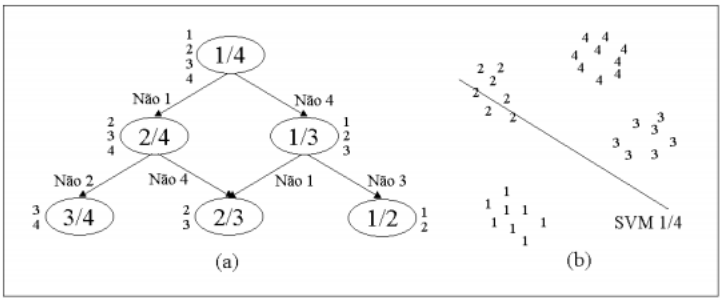

Figura 4.12- (a) Exemplo de grafo direcionado utilizado para classificar quatro classes a partir de SVM's binárias, as quais estão representadas pelos nós da árvore. (b) diagrama do espaço de características do problema 1/4 (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003)

A Figura 4.12 apresenta um exemplo com quatro classes. O processo é feito pelo DAG e equivale à operação de uma lista. No início, essa lista é composta por todas as classes. Em cada nó gera-se uma SVM que separa os elementos da primeira e última classe da lista. Após ser escolhida uma das classes, a outra é eliminada da lista e o processamento contínua da mesma forma, até que um nó seja atingido. A classificação de um exemplo de um problema de k classes requer a avaliação de k-1 nós, ou seja, k-1 SVM's. Deste modo, consegue-se reduzir o tempo de geração de resultados (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

Platt et al. (2000) demonstraram que esta técnica tem erros de generalização, limitado pela margem máxima obtida em cada nó. Como as SVM's utilizam o princípio de maximização de margens, o seu uso como indutor base é bastante adequado (Lorena & Carvalho, Introdução às Máquinas de Vetores Suporte (Support Vector Machines), 2003).

4.5. Rede Neuronal Artificial

Neurónio Biológico

O cérebro humano é constituído por estruturas e neurónios que interagem entre si e tomam decisões adequadas a partir de informações recebidas. Um neurónio é constituído por corpo celular, axónios e dendrites. As dendrites (zonas recetivas) formam uma malha de finos filamentos à volta do neurónio. O axónio (linhas de transmissão) consiste num tubo longo e fino que no seu final se divide em ramos que terminam em pequenos bolbos, que quase tocam as dendrites dos outros neurónios. O espaço entre o fim do bolbo e a dendrite é conhecido como sinapse, que tem como papel memorizar a informação (Matsunaga, 2012). Na figura 4.13 é possível observar o exemplo de um neurónio.

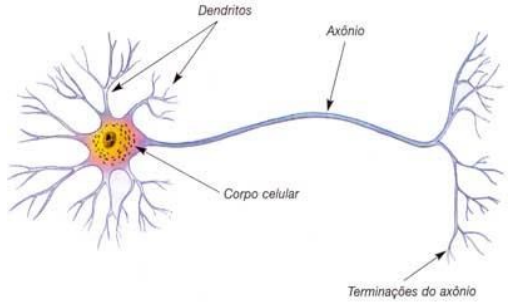

Figura 4.13- Modelo de Neurónio Biológico (Remes, 2013)

Neurónio Artificial

De um modo geral, todo o tipo de RNA apresentam a mesma unidade de processamento, um neurónio artificial, que simula o comportamento de um neurónio biológico.

Os neurónios artificiais possuem várias entradas, que correspondem às conexões sinápticas com outras unidades similares a ele, e uma saída, em que o seu valor depende diretamente do somatório ponderado de todas as outras saídas dos neurónios que estão conectados a ele. Um modelo de um neurónio artificial pode ser observado na figura 4.14.

Este modelo inclui um sinal adicional, *bias (b),* que favorece ou limita a possibilidade de ativação do neurónio. O processo sináptico é representado pelos *pesos (w)* que amplificam cada um dos sinais recebidos. A *função de ativação (f)* modela a forma como o neurónio responde ao nível da excitação, limitando e definindo a saída da rede neural.

A função de ativação pode ser representada de diferentes formas, sendo que, os três tipos básicos são: limiar, linear e sigmoide. A escolha do tipo varia conforme o objetivo do projeto.

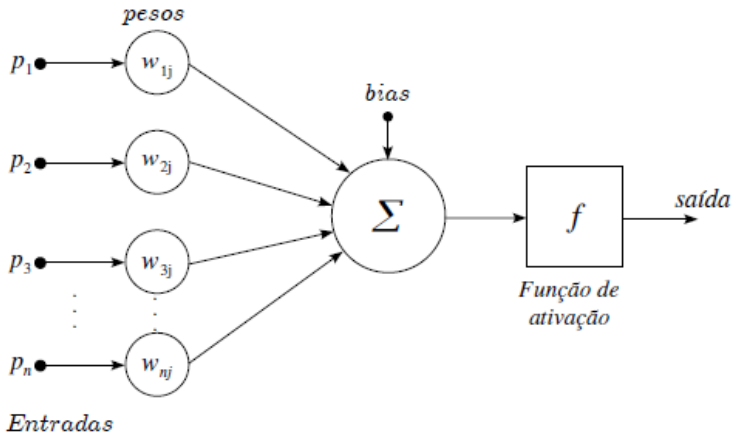

Figura 4.14- Modelo de neurónio artificial

4.5.1. Rede Neural Artificial (RNA)

Do inglês *Artificial Neural Network* (ANN), as redes neuronais baseiam-se em sistemas de computação adaptados, inspirados no processamento de informação existentes nos neurónios humanos e nas características das suas interconexões. O cérebro humano possui uma enorme capacidade de processar informação, deste modo, o estudo das redes neuronais é baseado no comportamento das redes neuronais biológicas (Matsunaga, 2012).

As RNA são constituídas por um conjunto de neurónios, conectados por canais de comunicação, associados a um determinado peso, que interagem entre si. Isto leva a formar-se um sistema de computação paralelo, em que as suas implementações podem ser em *hardware* ou em *software*. O seu comportamento inteligente deve-se às interações entre as diferentes unidades de processamento (Matsunaga, 2012) (Carvalho, 2009).

A construção das redes neuronais copia a arquitetura do cérebro, sabendo que o cérebro é formado por uma gigante rede de neurónios ligados entre si, as redes neuronais são

formadas por nós (o correspondente aos neurónios no cérebro). Na figura 4.15 é possível observar a comparação entre os elementos do cérebro e de uma rede neuronal.

Cérebro		**Rede Neuronal**
Neurónio	→	Nó
Ligação do neurónio	→	Peso da conexão

Figura 4.15- Comparação entre cérebro e rede neuronal

Na figura 4.16 é possível entender a estrutura de funcionamento do mecanismo (nó) de uma rede neuronal, onde X1, X2 e X3 são os sinais de entrada, W1, W2 e W3 corresponde ao peso correspondente a cada sinal de entrada, b corresponde ao erro sistemático e por último y é a saída pretendida.

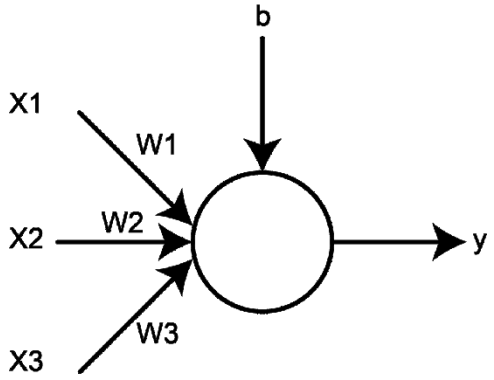

Figura 4.16- Nó a receber informação em três entradas

As redes neuronais artificiais têm a capacidade de adquirir, utilizar e armazenar informação baseada em acontecimentos (treino). Vão adquirindo conhecimento através da experiência, e são capazes de detetar padrões existentes num conjunto de dados. Tal como o ser humano utiliza o cérebro para reter informação conforme aprende, o computador usa também a sua memória para armazenar a informação (Kim, 2017).

A sua arquitetura é tipicamente organizada em camadas, em que algumas unidades podem estar conectadas às unidades da camada seguinte (figura 4.17).

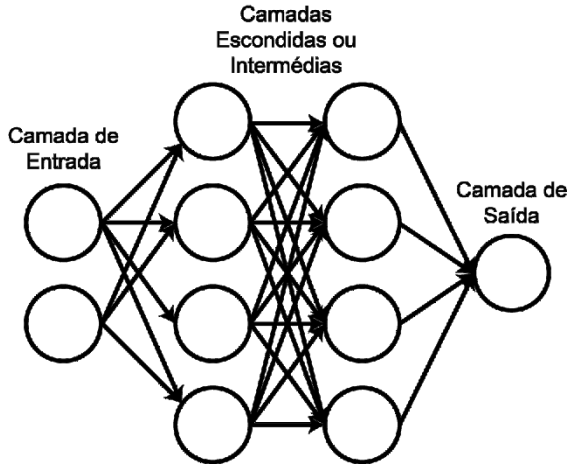

Figura 4.17- Organização em camadas

As ligações da rede neuronal dependem, essencialmente, do número de nós a interligar-se. Na camada de entrada os padrões são apresentados à rede, de seguida, nas camadas intermediárias ou escondidas realiza-se a maioria do processamento, por último, na camada de saída o resultado final é concluído e apresentado.

As redes neuronais inicialmente foram desenvolvidas com uma arquitetura simples, tornando-se cada vez mais complexa.

Single-layer neural networks

Este tipo de arquitetura é o mais simples, baseia-se apenas nas entradas para obter de forma imediata as saídas, como o próprio nome indica consiste em ligações simples de redes neuronais. A figura 4.18 serve de auxílio para uma melhor interpretação deste tipo de RNA (Kim, 2017).

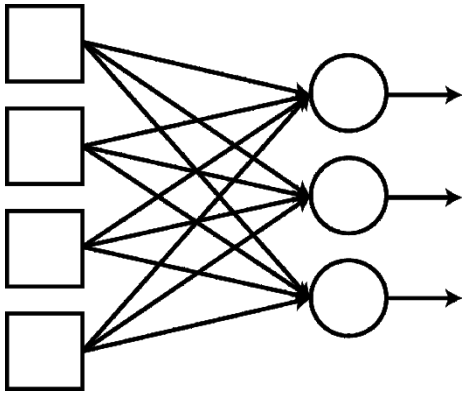

Figura 4.18- *Single-layer Neural Network*

Multi-layer neural networks

Consiste numa arquitetura mais elaborada, entre as entradas e as saídas, existem camadas escondidas (figura 4.19).

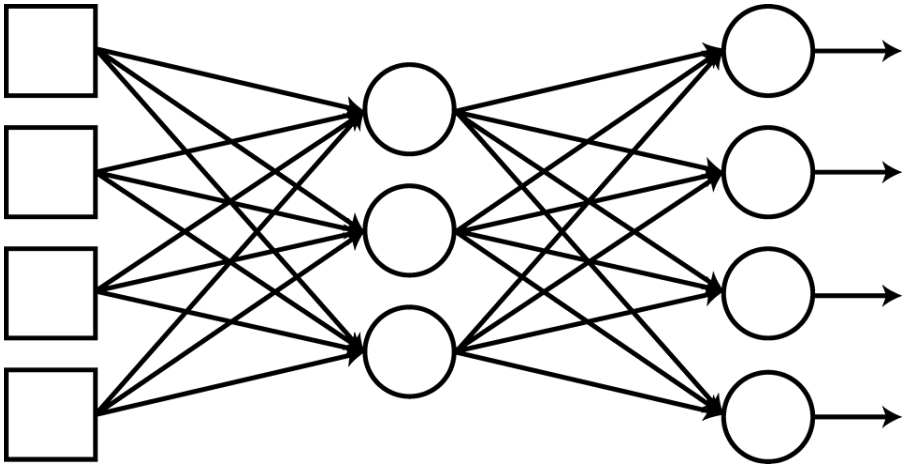

Figura 4.19- *(Shallow) Multi-layer Neural Network*

Este tipo de RNA possui diferentes tipos de constituição, sendo que, a sua diferença está no número de camadas escondidas.

Shallow neural networks

Este género de RNA aplica-se quando a arquitetura *multi-layer* é formada apenas por uma camada escondida. A figura 4.19 pode ser interpretada como uma rede neuronal "rasa" (Kim, 2017).

Deep neural network

Este género de RNA aplica-se quando a arquitetura *multi-layer* possui duas ou mais camadas escondidas. A figura 4.20 pode ser interpretada como uma rede neuronal "profunda" (Kim, 2017).

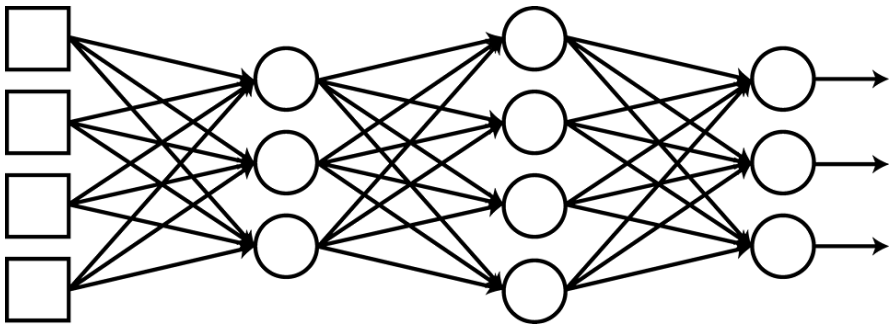

Figura 4.20- *Deep Neural Networks*

Algumas características importantes numa RNA:

- Flexibilidade – esta característica indica que a RNA pode ser ajustada a novos ambientes por meio de um processo de treino, fazendo com que aprenda novas ações, com base na informação contida nos dados utilizados para treino;
- Processamento de informação incerta – correndo o risco de que a informação fornecida para o treino possa estar incorreta, afetada por ruído, é possível obter um raciocínio correto;
- Paralelismo – elevado número de neurónios podem estar ativos ao mesmo tempo. Não existe nenhum tipo de restrição que obrigue a trabalhar uma instrução após outra;
- Robustez e tolerâncias a falhas – a eliminação de alguns neurónios faz com que o desempenho global não seja substancialmente afetado.

4.5.2. Deep Neural Networks (Deep Learning)

Desde a criação das redes neuronais artificiais, não demorou muito tempo até que fossem reveladas algumas limitações, porque a regra de aprendizagem para *multi-layer neural network* não era encontrada. Visto que, uma rede neuronal só armazena informação com o seu próprio treino, ter uma rede "intransitável" é inútil (Kim, 2017).

O treino de *multi-layer neural network* foi resolvido com a introdução do algoritmo "*back-propagation*". Mais tarde, aparece o conceito *Deep Learning*, que consiste numa vertente de redes neuronais, em que existem duas ou mais camadas escondidas no seu modelo.

Com o desenvolvimento tecnológico, o *Deep Learning* supera as restantes técnicas de *Machine Learning*. Resumindo, o *Deep Learning* foi a solução para otimizar o algoritmo "*back-propagation*" aplicado às *multi-layer neural networks.*

O seu objetivo é a aprendizagem aprofundada, baseada num conjunto de algoritmos que tentam moldar um conjunto de dados usando um grafo (representação abstrata de um conjunto de objetos e das relações entre eles) com várias camadas de processamento, compostas por diversas transformações lineares e não lineares.

A utilização do *Deep Learning* é cada vez mais uma realidade, porque, cada vez mais, se consegue obter melhores resultados. Estes vão melhorando com o avançar de trabalhos e pesquisas desenvolvidas, assim, a sua exatidão de reconhecimento, atinge um nível nunca antes alcançado, tornando o *Deep Learning* crucial para diversas aplicações.

A grande parte dos métodos de *Deep Learning* utiliza arquiteturas de redes neuronais, devido a isto, os modelos de *Deep Learning* são muitas vezes designados *como Deep Neural networks*. O termo "*Deep*" refere-se normalmente ao número de camadas escondidas na rede neuronal, geralmente, redes neuronais "normais" contêm 2-3 camadas escondidas, já nas *Deep Learning* o número de camadas escondidas pode atingir as 150 (Figura 4.21).

Os modelos treinam através de elevados conjuntos de dados e arquiteturas de redes neuronais, que aprendem diretamente dos dados, sem haver necessidade de intervenção manual.

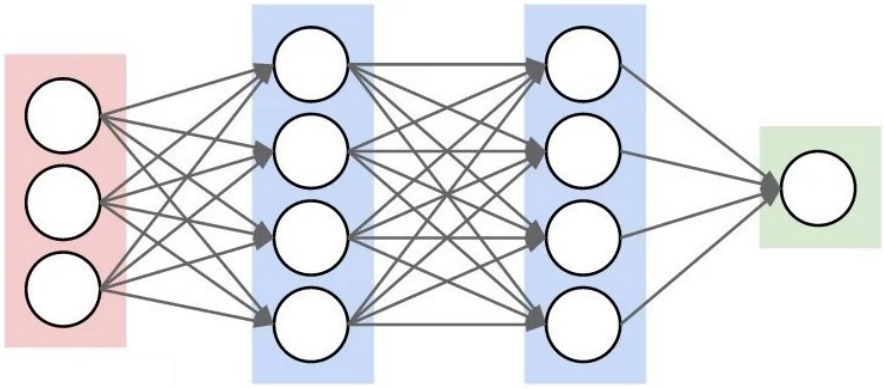

Figura 4.21- *Deep Neural Networks* - rede com duas camadas escondidas (networks, 2015)

Uma vez que, o *Deep Learning* é uma vertente mais aprofundada de *Machine Learning*, os seus algoritmos são, também, mais complexos, visto que, em níveis mais superficiais os dados tendem a convergir (MathWorks, Deep Learning, s.d.).

Sempre que se adiciona dados de treino à rede, noutros métodos de *Machine Learning* o desempenho não passa de um determinado nível, já na sua vertente *Deep Learning*, na maioria das vezes, quantos mais dados são adicionados para o treino, melhor será o seu nível de desempenho (MathWorks, Deep Learning, s.d.).

As *deep-neural networks* podem ser treinadas utilizando derivações dos algoritmos de *backpropagation* de uma função de custo, que mede a discrepância entre as saídas da rede e os valores alvo reais, produzidas para cada caso de treino (Hinton, et al., 2012).

De forma a reduzir a sobreposição, os pesos grandes são ajustados em proporção à sua magnitude quadrática, ou então, a aprendizagem pode ser encerrada quando o desempenho de um conjunto de validação começa a piorar. Nas *deep-neural networks* (à semelhança de outros tipos de rede), com conectividade completa entre camadas adjacentes, os pesos iniciais recebem pequenos valores aleatórios para evitar que todas as

unidades ocultas de uma camada obtenham exatamente o mesmo gradiente (Hinton, et al., 2012).

As *deep-neural networks* com muitas camadas ocultas são difíceis de otimizar. Para além dos problemas de otimização, as *deep-neural networks* com muitas camadas ocultas e muitas unidades por camada, são modelos muito flexíveis com um número muito grande de parâmetros. Isso torna-as capazes de modelar relações muito complexas e altamente não-lineares entre entradas e saídas. Esta característica é importante para a modelagem acústica de alta qualidade, mas também permite modelar as regularidades falsas, que são uma propriedade acidental dos exemplos específicos no conjunto de treino (Hinton, et al., 2012).

Melhoria das *Deep Neural Networks*

Apesar das melhorias constantes, o *Deep Learning* não tem uma tecnologia que possa ser apresentada. A sua inovação deve-se ao facto de pequenas tecnologias irem sendo melhoradas com o decorrer do tempo.

O algoritmo *backpropagation*, durante o processo de treino, testa os itens descritos de seguida, que são considerados obstáculos para as *deep learning*

- *Vanishing gradient;*
- *Overfitting;*
- *Carga computacional.*

Vanishing gradient

Em português, gradiente de fuga, é o tipo de obstáculo que surge quando se tenta treinar uma rede neuronal, aplicando técnicas de otimização baseadas em gradiente. À medida que camadas escondidas são adicionadas ao modelo, a capacidade de aprendizagem melhora, tornando-se cada vez mais eficaz. De uma maneira geral, ao adicionar camadas escondidas, tende a tornar a rede capaz de aprender cada vez mais funções e funções mais complexas, fazendo assim um trabalho de previsão cada vez mais eficaz (Walia, 2017).

O problema surge quando é aplicado o método de *backpropagation*, em que têm que ser calculados os erros em relação aos pesos, num sentido retrógrado. Assim, os gradientes têm tendência a ficar menores à medida que se recua na rede. Isto leva a que os neurónios das camadas anteriores aprendam de forma mais lenta, comparando-os com os neurónios de camadas posteriores (Walia, 2017) (Kim, 2017).

Camadas anteriores são importantes porque são elas as responsáveis por identificar os padrões mais básicos, e com isto tornam-se as camadas fundamentais na hierarquia da rede. Caso as camadas anteriores indiquem resultados errados, fará com que o tempo de treino seja maior e a exatidão de previsão do modelo diminui. (Walia, 2017)

Uma solução possível para o gradiente de fuga consiste no uso da função Unidade Linear Retificada (ReLU) como a função de ativação. Baseado em Kim (2017) e Walia (2017) sabe-se que a utilização da função *Sigmoid* ou *Tanh* como função de ativação não produz resultados muito favoráveis.

A função ReLU atribui a entradas negativas o valor zero e é definida através da Equação 28.

$$\varphi(x) = \begin{cases} x, x > 0 \\ 0, x \leq 0 \end{cases} = \max(0, x) \tag{28}$$

Funções implementadas nos experimentos práticos

Elliot symmetric sigmoid transfer

Esta função de transferência foi implementada recorrendo a função "elliotsig", que converte valores compreendidos entre [-∞ ∞] para [-1 1]. Com a implementação desta função pode ser necessário a utilização de um maior número de neurónios, ou precisar de mais iterações no processo de treino para obter os mesmo resultados (MathWorks, elliotsig, 2019).

Elliot 2 symmetric sigmoid transfer

Esta função de transferência foi implementada recorrendo a função "elliot2sig", esta função é uma variação da função original de Elliot, tem uma inclinação mais ingreme (semelhante á função "tansig"), no entanto o centro da função não é tão suavizado.

A função "elliot2sig" converte valores compreendidos entre [-∞ ∞] para [-1 1].

Vantagens associadas a esta função passam por, não necessitar de funções exponenciais ou trigonométricas e permite cálculos rápidos em hardware simples (MathWorks, elliot2sig, 2019).

Overfitting

Entende-se como o processo que ocorre quando está numa fase de aprendizagem, mas utiliza, também, algum tipo de ruído existente nos dados de treino, assim, afeta o

desempenho do modelo para a previsão com novos dados (Gupta, 2017) (Waseem Rawat, 2017).

Pode-se ver a existência de *overfitting*, caso se obtenha a exatidão da classificação dos dados utilizados para treino e esta, mesma exatidão, seja muito distante da exatidão do conjunto de teste, deste modo, conclui-se que a rede está a aprender com ruído. A ocorrência de *overfitting* é muito comum no *deep learning* devido à grande quantidade de pesos e bias (parâmetros) (Kim, 2017) (Gupta, 2017).

Para o treino ser eficaz, é necessário encontrar uma maneira de detetar se ocorre ou não *overfitting,* e caso ocorra o treino deve ser interrompido.

O *overfitting* pode ser evitado usando dados para efetuar uma validação, ou *dropout* (desistência) em que treina apenas alguns dos nós e pesos selecionados aleatoriamente, pode ser entendido na figura 4.22. As percentagens recomendadas em Kim (2017) são de 50% para camadas ocultas e 25% para camadas de entrada (Kim, 2017) (Gupta, 2017) (Waseem Rawat, 2017).

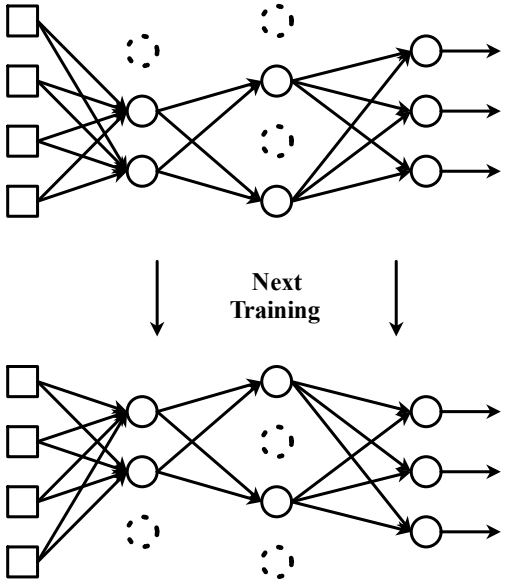

Figura 4.22- Processo de treino para *Overfitting*

Carga computacional

Um poder computacional elevado faz com que o tempo de treino seja menor. O poder computacional é importante, visto que, à medida que existem cada vez mais dados, mais pesos e *bias* vão existir, que são calculados automaticamente de maneira a obter o melhor desempenho. Este problema suavizou com a introdução de *hardwares* de alto desempenho (GPU's) e a criação de algoritmos criados para o efeito (Kim, 2017).

5. Desenvolvimento e Resultados

5.1 Parâmetros

Neste capítulo é descrito o desenvolvimento técnico no decorrer da dissertação.

Tal como em trabalhos realizados anteriormente (Alves N. , 2016), foram analisados vários parâmetros e a sua capacidade de distinção entre sujeito saudável e sujeito patológico. Elaboraram-se três grupos de parâmetros, no grupo I (a) os parâmetros correspondem ao *Jitter* Relativo, *Shimme*r Relativo e HNR para três vogais (/a/, /i/ e /u/) em três tons diferentes (alto, baixo e normal), o que na sua totalidade corresponde a 27 parâmetros por sujeito. Estes parâmetros foram extraídos utilizando o algoritmo desenvolvido por Teixeira e Gonçalves (2016) (Teixeira & Gonçalves, Algorithm for jitter and shimmer measurement in pathologic voices, 2016) e por Fernandes (2018) (Fernandes, 2018).

O grupo I (b) contém os parâmetros do grupo I (a) em conjunto com os parâmetros NHR e Autocorrelação, num total de 45 parâmetros por sujeito, extraídos utilizando o algoritmo desenvolvido por Fernandes (2018) (Fernandes, 2018).

No grupo II, os parâmetros são os coeficientes MFCCs das vogais, usadas no grupo I. A extração destes parâmetros teve por base o trabalho realizado por (Fernandes, 2018), onde, originalmente se tinham 9 (para as nove vogais/tons) conjuntos de 13 coeficientes. Assim, o grupo II apresenta um vetor com dimensão 117 para cada sujeito.

No grupo III, os parâmetros correspondem aos coeficientes MFCCs, extraídos da fala continua (frase). A extração destes parâmetros, também, teve por base o trabalho realizado por (Fernandes, 2018). Foram usados segmentos de 35 ms com uma sobreposição variável de forma a que a frase de cada sujeito corresponda a 50 segmentos. Para cada sujeito obteve-se uma matriz (50 * 13). Semelhante ao sucedido no grupo II, foi necessário converter para um único vetor, então, para cada sujeito existe um vetor com dimensão 650.

Mais detalhes relacionados com os parâmetros, são referidos no capítulo 3 desta dissertação.

5.2 Medidas utilizadas para avaliar o desempenho

Para avaliar os modelos de machine learning implementados, foram utilizados parâmetros para medir desempenhos.

As medidas obtidas dependem diretamente de umas medidas obtidas previamente, Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falso Positivo (FP), Falso Negativo (FN) (Hossin, 2015).

É demonstrado um pequeno exemplo com o prepósito de entender como são adquiridos os valores a utilizar, imagine-se a distinção entre sujeitos patológicos e sujeitos saudáveis (Guedes, 2019):

Verdadeiro Positivo (VP): os verdadeiros positivos são as pessoas que realmente tem patologia e que foram classificadas corretamente como doentes;

Verdadeiros Negativos (VN): os verdadeiros negativos é o oposto, são as pessoas saudáveis e que são classificadas corretamente como saudáveis;

Falso Positivo (FP): os falsos positivos são aquelas pessoas que são saudáveis e foram classificadas como doentes;

Falso Negativo (FN): os falsos negativos são aquelas pessoas que são doentes é que foram classificadas equivocadamente como saudáveis.

Estes conceitos podem ser melhor compreendidos observando a tabela 5.1 onde é mostrada uma Matriz de Confusão. Na matriz de confusão está a quantidade de sujeitos pertencentes a cada categoria de acordo com a classificação feita, e é possível entender também se a classificação foi bem realizada ou não.

Tabela 5.1- Matriz confusão aplicada para análise de resultados (Alves N. , 2016)

		Resultado da classificação	
		Saudável	Patológico
Diagnóstico	Saudável	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Patológico	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Tendo isto, é possível apurar alguns valores de medida geralmente utilizados para avaliar desempenhos, destacando-se assim a Exatidão, Sensibilidade, Especificidade e Medida F.

Exatidão

É calculada de acordo com a equação 29 e corresponde à soma dos Verdadeiros Positivos e Verdadeiros Negativos dividido pelo total de casos existentes (sejam Verdadeiros ou Falsos, Positivos ou Negativos). Esta medida apesar de ser muito utilizado, não é aconselhável para testes com dados desbalanceados.

$$Exatidão\ (\%) = \frac{VP + VN}{VP + VN + FP + FN} * 100 \tag{29}$$

Precisão

É calculada de acordo com a equação 30 e corresponde a todas as classificações feitas como positivas e que são realmente positivas, nesta medida, os diagnósticos negativos não são tidos em conta.

$$Precisão\ (\%) = \frac{VP}{VP + FP} * 100 \tag{30}$$

Sensibilidade e Especificidade

A Sensibilidade é calculada de acordo com a equação 31 e tem como objetivo identificar corretamente as previsões positivas.

$$Sensibilidade\ (\%) = \frac{VP}{VP + FN} * 100 \tag{31}$$

A Especificidade é calculada de acordo com a equação 32 e tem como objetivo identificar corretamente as previsões negativas.

$$Especificidade\ (\%) = \frac{VN}{VN + FP} * 100 \tag{32}$$

Medida F

A Medida F é definida pela medida harmónica existente da relação da Precisão e Sensibilidade, é calculada de acordo com a equação 33 e quanto mais próximo de 1 for o resultado, mais fiável é o modelo. Este parâmetro apresenta uma avaliação mais confiável no que diz respeito a utilização de categorias desbalanceadas.

$$Medida\ F = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade} \tag{33}$$

5.3 Validação cruzada "Leave-one-out"

É comum recorrer a técnicas de validação cruzada quando os dados existentes são poucos, ou quando as diferentes classes estão desbalanceadas. A falta de dados pode levar a dificuldades na generalização do modelo, isto porque, podem não existir dados suficientes para criar conjuntos de treino, validação e teste fixos.

Este tipo de problema pode ser ultrapassado aplicando técnicas de validação cruzada, uma dessas possíveis técnicas é validação cruzada em k-fold.

A utilização de k-fold, tem como objetivo dividir aleatoriamente a base de dados em N grupos (k-folds), onde cada instância é treinada N-1 vezes e testada uma única vez (Guedes, 2019).

O método "Leave-one-Out" é um caso especifico do k-fold, onde é retirado da base de dados apenas uma amostra para teste e as restantes amostras dividem-se entre treino e validação, este processo é repetido tantas vezes quanto número de amostras existirem na base de dados, por outras palavras, vai treinar e testar tantas vezes quanto sujeitos existam na base de dados (Webb, 2011). A figura 5.1 demonstra o método.

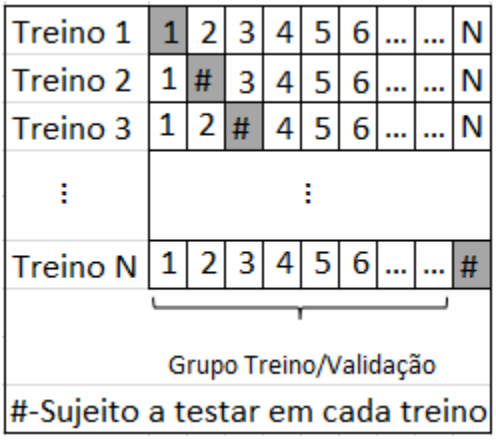

Figura 5.1- Método "Leave-one-out"

5.4. Implementação da árvore de decisão

A implementação das árvores de decisão, no Matlab, pode ser por meio de *script's* realizados pelo utilizador, onde são implementadas as condições padrão para que as decisões sejam tomadas, ou então pode-se recorrer a funções já existentes que devolvem a árvore de decisão de maneira "automática".

Para a aplicação de árvores de decisão, foi utilizada uma função já existente na toolbox do *software*, designada *'fitctree'*. Esta função devolve uma árvore de decisão binária, ajustada com base nos atributos de entrada e de saída. Os nós das ramificações são divididos com base nos valores dos atributos de entrada (MathWorks, fitctree, 2018).

Outra função de extrema importância foi a *'classregtree'*, que também está disponível no *software*. Esta função permite entender as condições, em que a árvore de decisão faz as suas escolhas (MathWorks, classregtree, s.d.).

Uma árvore de decisão apresenta maior número de nós e ramificações, quantos mais parâmetros forem utilizados. Isto implica a existência de mais ou menos níveis. Os níveis referidos são a fase em que a árvore de decisão se encontra estável num nó e tem que tomar a decisão do próximo passo.

Para a aplicação das árvores de decisão, independentemente do grupo de parâmetros, utilizou-se 75 % dos sujeitos para treino e os restantes 25% para teste.

Para o caso da classificação binária (saudável / patológico) a quantidade de sujeitos é mencionada na tabela 9.

Para o grupo de parâmetros I (a), obteve-se 58% para classificar sujeitos patológicos e saudáveis. Para a discriminação entre as patologias, a taxa de exatidão atingida foi de 38,5%.

Para o grupo de parâmetros I (b), obteve-se 65,8% para classificar sujeitos patológicos e saudáveis. Para a discriminação entre duas patologias, a taxa de exatidão atingida foi de 41,9%.

Para o grupo de parâmetros II, obteve-se 59,8% para discriminar sujeitos patológicos e saudáveis. Para a discriminação entre duas patologias, a taxa de exatidão atingida foi de 33,3%.

Para o grupo de parâmetros III, obteve-se 59,8% para discriminar sujeitos patológicos e saudáveis. Para a discriminação entre duas patologias, a taxa de exatidão atingida foi de 41,9%.

O gráfico 1 demonstra a exatidão alcançada, com a aplicação dos diferentes grupos de parâmetros, para fazer a distinção entre sujeitos saudáveis e sujeitos patológicos.

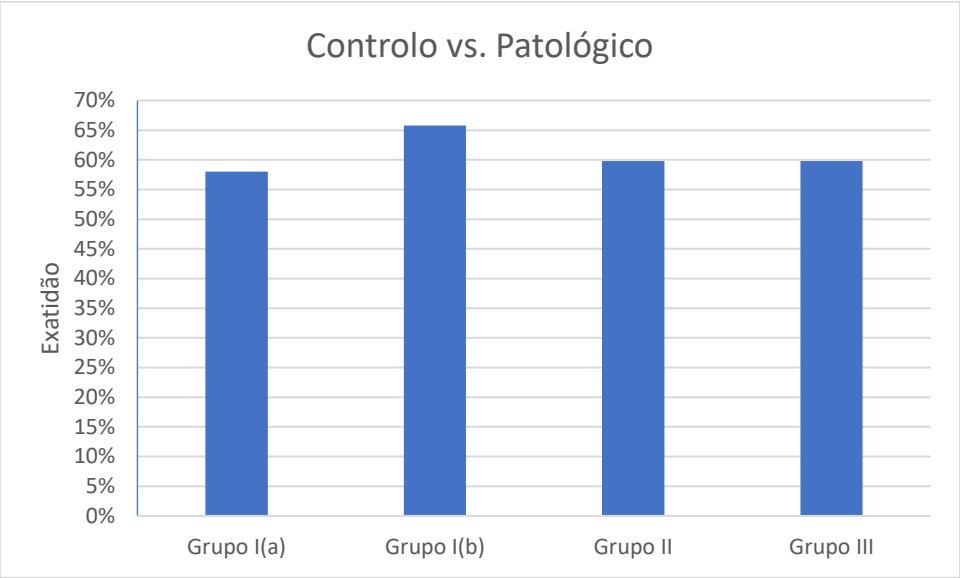

Gráfico 1- Exatidão alcançada com árvores de Decisão para a distinção entre sujeitos saudáveis e patológicos

O gráfico 2 demonstra a exatidão máxima alcançada, com a aplicação dos diferentes grupos de parâmetros, para fazer a distinção entre duas das diferentes categorias em estudo.

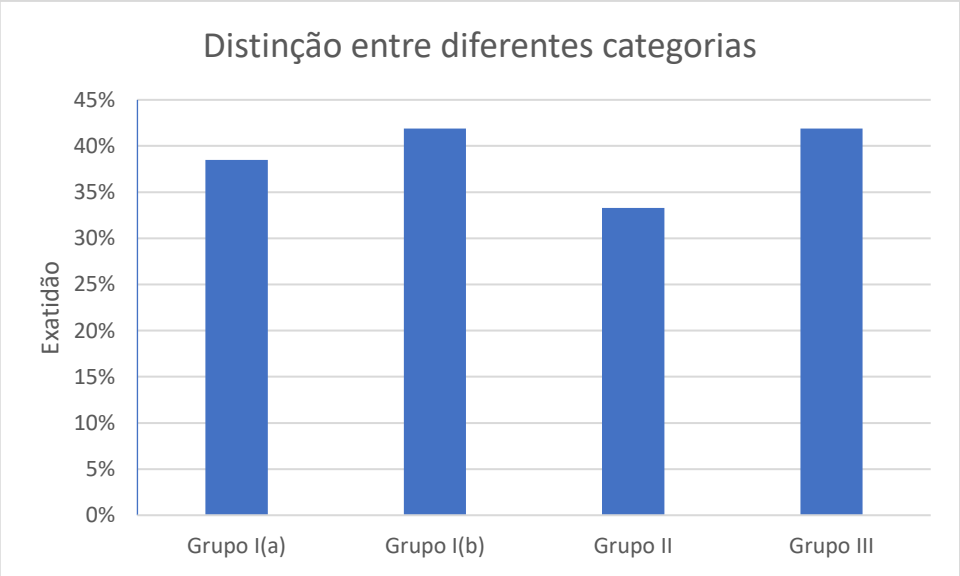

Gráfico 2- Exatidão alcançada com árvores de Decisão para a distinção entre sujeitos de duas diferentes categorias (patologias)

A otimização deste modelo não foi realizada por as espectativas de se obterem bons resultados serem reduzidas em face dos resultados apresentados.

5.5. Implementação de SVM

A implementação dos SVM's foi feita com *software* Matlab, onde existem funções especificas para criar, treinar e classificar os dados. A função '*svmtrain*' permite treinar o SVM, e a função *'svmclassify*' permite apurar o poder de predição do SVM após o treino realizado.

Descartou-se a hipótese de discriminar os sujeitos por género, porque em trabalhos anteriores Teixeira *et al* (2018) concluiu-se que para os parâmetros usados não há diferença entre sujeitos de diferentes géneros.

Inicialmente, foram testados os três grupos de parâmetros, fazendo apenas a classificação binária entre sujeitos de controlo e patológicos.

A percentagem utilizada para treino foi de 75% e os restantes 25% para teste (tabela 5.2), onde (146 / 209) e (48 / 70) representam a quantidade de sujeitos utilizados em treino e teste, pertencentes a cada classe correspondentemente.

Tabela 5.2- Sujeitos utilizados para o grupo de treino e teste

Distinção	Sujeitos Totais	Sujeitos em Treino	Sujeitos em Teste
Controlo vs. Patológicos	473	(146/209)	(48/70)

Testou-se os diferentes tipos da função de *Kernel*, bem como os diferentes métodos disponíveis para encontrar o hiperplano de separação. As funções de *Kernel* disponíveis são: *'linear', 'polynomial', 'gaussian' ou 'rbf'* e *'quadratic'*. Em relação aos métodos de otimização disponíveis são três: *Least Squares (LS), Quadratic Programming (QP)* e *Sequential Minimal Optimization (SMO)*.

A utilização de todos os sujeitos para a distinção de saudáveis e patológicos, permite apurar que os parâmetros da fonte (parâmetros I(a)) (*Jitter, Shimmer* e HNR) contêm informação mais relevante, permitindo fazer uma classificação com maior exatidão (tabela 5.3).

Fez-se o estudo, acrescentando mais parâmetros (NHR e Autocorrelação) (parâmetros I(b)), para tentar obter melhores resultados. No entanto, com quatro ou até mesmo cinco parâmetros (parâmetros I(b)), a exatidão alcançada foi praticamente a mesma (tabela 5.3).

Atingiu-se assim, um total de 80,7% de sujeitos classificados corretamente, no conjunto de teste (tabela 5.3).

A utilização de parâmetros MFCC's, quer sejam das vogais ou da frase, apresentam resultados inferiores, com precisões de 61,1 % e 52,6% correspondentemente. Prevê-se que os MFCC's não contenham tanta capacidade discriminante como os parâmetros da fonte (tabela 5.3).

Tabela 5.3- Distinção entre sujeitos saudáveis/ patológicos nos vários grupos de parâmetros

Designação	Grupo de parâmetros	Exatidão (%)	Método	Função
Saudáveis vs. Patológicos	Parâmetros da fonte (I(a))	80,7	SMO	MLP
	Parâmetros da fonte (I(b))	79.2	SMO	MLP
	MFCC's das vogais (II)	61,1	SMO	MLP
	MFCC's da frase (III)	52,6	LS	MLP

A aplicação do método *'SMO'* permite, em conjunto com a função '*MLP*', atingir uma exatidão máxima. Reparou-se, no decorrer de diversas experiências, que o método '*QP*' em conjunto com a função '*MLP*', não apresentava nenhum resultado, porque, devido à elevada quantidade de dados, o problema não sofre uma otimização, tornando-se assim um problema não convexo.

Segue-se um processo de classificação binária separando as diferentes categorias, sendo consideradas todas as combinações. A tabela 5.4 mostra a quantidade de sujeitos utilizados para treino e teste, para as diferentes combinações.

Tabela 5.4- Sujeitos utilizados discriminando as categorias

Descrição	Total de sujeitos	Sujeitos de Treino	Sujeitos de Teste
Controlo vs. Disfonia	263	146/52	48/17
Controlo vs. P.C. Vocais	363	146/127	48/42
Controlo vs. L. Crónica	235	146/32	48/9
Disfonia vs P.C. Vocais	237	52/127	17/42
Disfonia vs. L. Crónica	110	52/32	17/9
Paralisia vs. L. Crónica	210	127/32	42/9

As tabelas 5.5, 5.6 e 5.7 demonstram para cada grupo de parâmetros, as exatidões alcançadas, bem como o método utilizado e a função aplicada. Tal como no processo anterior, aqui, também, foram testados diferentes métodos e funções, contudo, apenas os melhores resultados são mencionados (Teixeira F. , et al., 2018).

Para a classificação entre as diferentes categorias, apenas foram utilizados sujeitos das categorias em causa.

Na tabela 5.5, foram utilizados os parâmetros da fonte (grupo I(a)), em que a maior exatidão é encontrada quando se faz uma discriminação entre sujeitos saudáveis (ou controlo) e sujeitos portadores de laringite, atingindo 80%, no entanto é de salientar que o grupo de teste pode conter as classes desbalanceadas, portanto os resultados atingidos devem ser considerados com alguma prudência.

Conclui-se, que sujeitos portadores de disfonia, quando comparados com outros sujeitos patológicos, apresentam taxas de exatidão relativamente baixas. O que pode indicar que entre sujeitos patológicos, os parâmetros da fonte não tenham informação discriminatória para fazer uma distinção patológica, quando aplicados num SVM.

Tabela 5.5- Classificação com SVM, aplicando os parâmetros do grupo I(a)

Grupo de parâmetros	Designação	Exatidão (%)	Método	Função
Parâmetros da fonte	Controlo vs. Disfonia	76,1	*SMO*	*LINEAR*
	Controlo vs. P.C. Vocais	73,9	*QP*	*POLYNOMIAL*
	Controlo vs. L. Crónica	**80**	***QP** **	***RBF***
	Disfonia vs P.C. Vocais	67,2	*LS*	*RBF*
	Disfonia vs. L. Crónica	62	*QP* *	*QUADRATIC* *
	Paralisia vs. L. Crónica	72,2	*SMO*	*RBF*

*- a exatidão indicada pode ser atingida utilizando diferentes métodos ou funções.

Na tabela 5.6, utiliza-se os MFCC's das vogais, que por sua vez, apresentam uma exatidão um pouco mais elevada no que diz respeito à classificação entre sujeitos patológicos. Distinguir sujeitos com laringite crónica de sujeitos saudáveis também melhorou cerca de 3%.

Tabela 5.6- Classificação com SVM, aplicando os parâmetros do grupo II

Grupo de parâmetros	Designação	Exatidão (%)	Método	Função
MFCC's das vogais	Controlo vs. Disfonia	74,6	*SMO**	*RBF*
	Controlo vs. P.C. Vocais	65,2	*QP**	*QUADRATIC*
	Controlo vs. L. Crónica	**83,3**	***LS****	***RBF***
	Disfonia vs P.C. Vocais	72,1	*QP **	*RBF*
	Disfonia vs. L. Crónica	69	*SMO **	*POLYNOMIAL*
	Paralisia vs. L. Crónica	**83,3**	***LS***	***POLYNOMIAL***

[i]*- a exatidão indicada pode ser atingida utilizando diferentes métodos ou funções

Na tabela 5.7, utiliza-se os MFCC's da frase, e comparando sujeitos saudáveis com sujeitos portadores de disfonia, ocorre uma melhoria de aproximadamente 5%, quando comparado com os grupos de parâmetros restantes.

Tabela 5.7- Classificação com SVM, aplicando os parâmetros do grupo III

Grupo de parâmetros	Designação	Exatidão (%)	Método	Função
MFCC'S da frase	Controlo vs. Disfonia	80,6	*LS*	*LINEAR*
	Controlo vs. P.C. Vocais	67,4	*SMO**	*LINEAR*
	Controlo vs. L. Crónica	**83,3**	***QP****	**** RBF***
	Disfonia vs. P.C. Vocais	72,1	*QP**	**POLYNOMIAL*
	Disfonia vs. L. Crónica	65,5	*QP **	*RBF*
	Paralisia vs. L. Crónica	81,5	*QP **	*RBF*

*- a exatidão indicada pode ser atingida utilizando diferentes métodos ou funções

Os estudos realizados envolvendo este capítulo foram publicados na conferência HCIST, com o título "Classification of Control/Pathologic Subjects with Support Vector Machines".

5.6. Implementação de Rede Neuronal

Como referido na parte teórica desta dissertação, o Matlab disponibiliza também vários tipos de redes neuronais, tais como diferentes funções de treino, funções de aprendizagem, performance e funções de transferência.

Inicialmente, à semelhança do sucedido na aplicação de SVM's, aqui nas redes neuronais também foi feita uma classificação entre duas categorias, criando um grupo de treino, de validação e de teste. A Dimensão do conjunto de treino é 70%, validação e teste 15%. Nesta fase apenas foram utilizados sujeitos das duas categorias envolventes.

A tabela 5.8 mostra os valores de exatidão no conjunto de teste, quando aplicados os parâmetros do grupo I(a), para classificações binárias com redes neuronais.

Tabela 5.8 Exatidão obtida no conjunto de teste, com redes neuronais

Descrição	Exatidão (%)
Controlo/Patológico	91,4
Controlo/Disfonia	94,9
Controlo/Paralisia	92,6
Controlo/L. Crónica	100
Disfonia/Paralisia	94,3
Disfonia/L. Crónica	100
Paralisia/L. Crónica	100

Reparou-se que os valores de exatidão alcançados poderiam ser bons, no entanto reparou-se que as classes estavam desbalanceadas o que poderia induzir em erro a performance atingida pelas diferentes redes. Optou-se assim por realizar novos experimentos com redes neuronais que consistem em utilizar o método "*leave-one-out*", assim todos os sujeitos foram analisados e permite analisar com melhor veracidade os resultados obtidos.

O método do "*leave-one-out*" aplicado no conceito de deep-learning utilizado nesta dissertação foi implementado de acordo com o algoritmo representado na figura 5.2.

```
i=1
Enquanto i<Total de Sujeitos
    Teste = sujeito (i)
    Treino e Validação = Todos os sujeitos – sujeito(i)
    J=1
    Enquanto J<=7
        Cria Target
        Cria Rede(J)
        Treina Rede(J)
        Simula Rede(J) com Teste
        Guarda Resultado
        Simula Rede(J) com Treino e Validação
        Guarda Saída
        J=J+1
    Fim do Enquanto
    Cria Target para Rede Final
    Cria Entrada da Rede Final com as Saídas das 7 Redes anteriores
    Cria Rede Final
    Treina Rede Final
    Simula Rede Final com a saída das 7 redes anteriores para o Teste
    Guarda Saída Final para o Teste
    i=i+1
Fim do Enquanto
Compara Saída Final Para todos os sujeitos com o Target para todos os
sujeitos
```

Figura 5.2- Algoritmo do método "Leave-one_out" aplicado de modo a que o Sujeito seja testados em todas as redes neuronais

As redes neuronais utilizadas apenas variam os nós nas camadas intermédias (camadas escondidas), na camada de entrada o número de nós é igual ao número de parâmetros

utilizados, assim quando aplicados os parâmetros do grupo I(a) a camada de entrada é composta por 27 nós, com os parâmetros do grupo I(b) é composta por 45 nós, com os parâmetros do grupo II é composta por 117 nós, e aplicando os parâmetros do grupo III a camada inicial é composta por 650 nós.

A camada de saída das redes é composta exclusivamente com um único nó.

Nesta etapa foi decidido aplicar três classes em cada rede neuronal, por exemplo o sujeito em análise pode pertencer a controlo, ou a disfonia, caso não pertença a nenhuma das duas opções é inserido na classe de "outras", onde, neste caso, a classe "outras" corresponderá a laringite e a paralisia.

Para avaliar o desempenho das redes neuronais, fez-se três avaliações em cada uma delas, considerando apenas duas categorias.

No exemplo demonstrado na tabela 5.9, o primeiro momento considerou-se controlo como sendo uma categoria e as restantes classes correspondem a uma outra classe, no segundo momento considerou-se disfonia como uma classe e controlo, laringite e paralisia como sendo a outra classe, no terceiro momento de avaliação laringite e paralisia foram considerados uma classe e controlo e disfonia como sendo outra classe. A tabela 16 mostra valores de exatidão e da Medida F, utilizando os parâmetros do grupo I(a), para a rede que faz a classificação entre controlo, disfonia ou outras. Os restantes resultados obtidos para as diferentes redes são demonstrados no anexo B.

Tabela 5.9- Parâmetros de avaliação na classificação de Controlo/Disfonia/Outras

	Exatidão (%)	**Medida F (%)**
Controlo / Outras	59,55	49,54
Disfonia / Outras	74,07	2,33
Outras / (controlo + disfonia)	52,29	59,22

Como complemento, a tabela 5.10 demonstra os valores medidos para a distinção entre sujeitos saudáveis e patológicos, utilizando os diferentes grupos de parâmetros. Na tabela 5.11 é possível observar as configurações utilizadas nas diferentes redes para diagnosticar sujeitos saudáveis e sujeitos patológicos onde F. treino corresponde à função de treino, C. E a camada escondida, F. Ati a função de ativação.

Tabela 5.10 – Valores medidos para classificar sujeitos saudáveis/patológicos

Parâmetros	Exatidão (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)	Medida F
Grupo I(a)	**72.7**	***69.6***	*65.9*	***78***	***67.7***
Grupo I(b)	72.3	*68*	*65.7*	*77.2*	*66.8*
Grupo II	67.2	53	61.7	70.3	57
Grupo III	71	57.2	**67.3**	73	61.8

Tabela 5.11- Caraterísticas das redes para classificar saudáveis/patológicos.

Parâmetros	Tipo de rede	F. treino	Nós C. E	F. Ati. C.E	F. Ati. C.S
Grupo I(a)	Feed-Forward	Levenberg-Marquardt	*50*	*Elliot2sig*	*Elliot2sig*
Grupo I(b)		Scaled Conjugate Gradient	*60*	*Elliot2sig*	*Tansig*
Grupo II			65		Elliot2sig
Grupo III			75		Tansig

As sete redes neuronais (sete, devido ao "confronto" entre as diferentes categorias) mencionada na tabela 5.12 formam o primeiro nível de deep-learning implementado.

Tabela 5.12 redes que compõe o primeiro nível implementado de Deep-Learning

R.N 1	**Controlo / Patológico**
R.N 2	Controlo / Disfonia / Outras
R.N 3	Controlo / Laringite / Outras
R.N 4	Controlo / Paralisia / Outras
R.N 5	Disfonia / Paralisia / Outras
R.N 6	Disfonia / Laringite / Outras
R.N 7	Laringite / Paralisia / Outras

Para o segundo nível do deep-learning foram feitos alguns testes e guardou-se a melhor configuração encontrada, assim é composta por uma rede feed-forward com 7 nós na primeira camada (respostas das sete redes de primeiro nível) e 75 nós na camada escondida, utiliza uma função de ativação "elliot2sig" em simultâneo com o treino Levenberg-Marquardt. A configuração mencionada é demonstrada na figura 5.3.

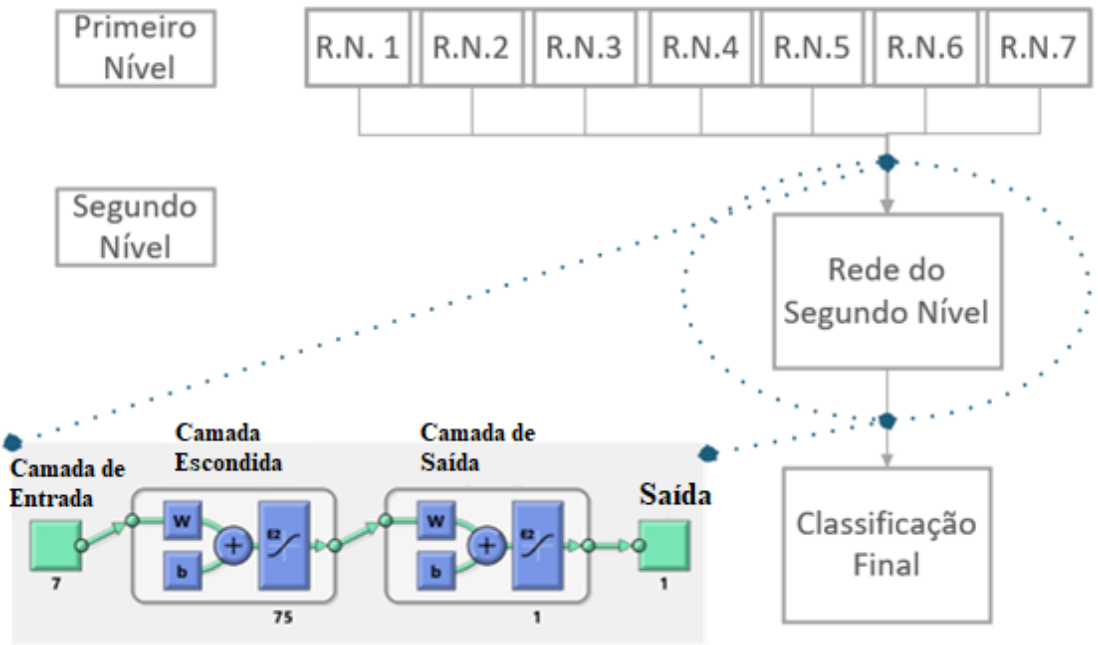

Figura 5.3 Arquitetura do Deep-Learning

O segundo nível implementado tem como principal objetivo a identificação da patologia portadora pelo individuo, na tabela 5.13 mostra a exatidão alcançada com diferentes números de nós na camada escondida da rede implementada no nível 2, mostra também o total do tempo gasto para a realização de diagnóstico dos sujeitos totais expresso em minutos. Este processo consistiu em treinar as 8 redes (7 na primeira camada e 1 na segunda camada) 473 vezes (quantidade de sujeitos) usando o método '*leave-one-out*'. os valores mencionados foram obtidos com recurso a um processador Intel ® Core ™ i5-3337u CPU @ 1.80GHz.

Tabela 5.13- Exatidão obtida e tempo gasto de acordo com o número de nós.

Nº de nós na Camada Escondida	**Exatidão obtida (%)**	**Tempo total (minutos)**
50	30.4	141.1
75	**39.5**	**168.5**
80	36.4	116.4

Obteve-se assim como exatidão máxima um valor de 39,5%, correspondendo a 187 sujeitos classificados corretamente de acordo com a categoria a que pertencem. A matriz confusão (figura 5.4) permite entender as classificações realizadas pelo algoritmo implementado.

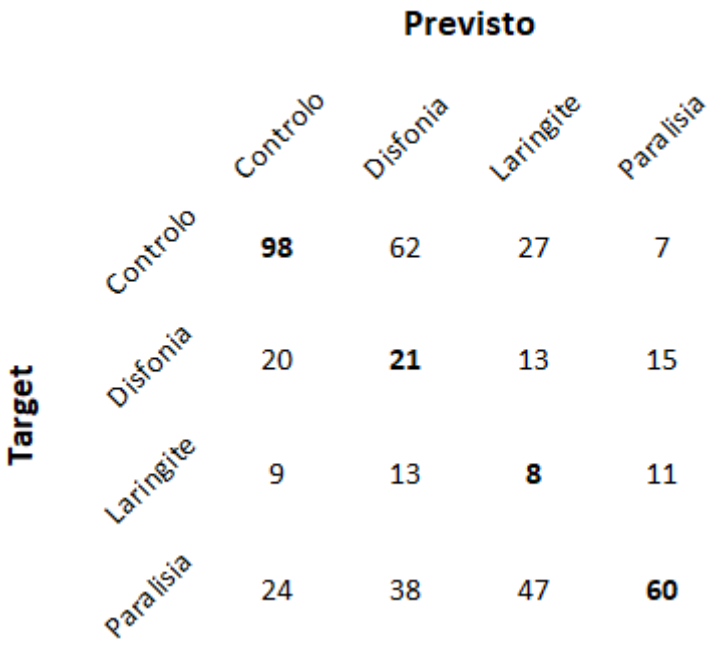

Figura 5.4 Matriz Confusão Final

Fazendo uma análise detalhada, considerando apenas duas categorias, para a classificação entre controlo / outras atinge-se uma exatidão de 68.5%, uma precisão de 50.5%, sensibilidade de 64.9%, especificidade de 70.2% e uma Medida F de 56.8%. Estes parâmetros de avaliação foram calculados tendo em conta a matriz confusão representada na tabela 5.5.

As classificações Disfonia / Outras, Laringite / Outras e Paralisia / Outras são demonstradas no anexo C desta dissertação.

Tabela 5.5 Matriz confusão. Controlo / Outras

Controlo / outras

		Previsto	
		Controlo	outras
Target	Controlo	98	96
	outras	53	226

6.Análise dos resultados

Neste capítulo é feita uma análise dos resultados obtidos no decorrer do desenvolvimento desta dissertação.

É de referir que a normalização de valores foi feita no momento de extração de parâmetros, o que leva a não serem implementadas técnicas com esse prepósito, durante a realização dos experimentos mencionados.

Inicialmente, fez-se a classificação entre sujeitos saudáveis de sujeitos patológicos, aplicando as três ferramentas de *machine learning* estudadas nesta dissertação. Para tal utilizaram-se os grupos de parâmetros I(a), I(b), II e grupo III.

As árvores de decisão são a ferramenta que apresenta valores de exatidão mais baixos, atingindo o seu valor máximo de exatidão de 65,8% quando aplicados os parâmetros do grupo I(b) para diagnóstico de sujeitos saudáveis e patológicos, já para a discriminação de duas patologias o mesmo grupo de parâmetros à semelhança do grupo III apresentam uma exatidão de 41,9%.

A utilização de SVM's para este tipo de distinção (controlo / patológico) atinge uma exatidão máxima de 80,7% quando utilizados os parâmetros do grupo I(a). No entanto, a exatidão alcançada com os parâmetros do grupo I(b) tem uma diferença insignificante, onde o grupo I(b) tem cerca de 1,5% menos.

Para discriminação entre duas categorias, os parâmetros do grupo II atingem uma exatidão máxima de 83,3% para diagnóstico de sujeitos de paralisia / laringite e sujeitos de controlo / laringite.

Com os experimentos realizados comprovou-se que os parâmetros dos grupos II e III não apresentam fatores tão discriminantes como os parâmetros do grupo I.

Para a aplicação de redes neuronais apenas foram tidos em conta os resultados obtidos com as redes que discriminam entre três categorias (com exceção da rede controlo/ patológico), atingiu-se uma exatidão média de 79% para identificação de disfonia / laringite / outras utilizando os parâmetros do grupo I (a ou b). A sensibilidade média atingiu o valor máximo de 67.3% com a rede controlo / patológico com os parâmetros do grupo III. A especificidade média mais alta foi de 84.5% utilizando os parâmetros do grupo I(b) com a rede disfonia/ paralisia / outras. A medida F com valor médio mais alto atingido foi 67.7% com a utilização da rede controlo / patológico recorrendo aos parâmetros do grupo I (a).

Os resultados com as árvores de decisão foram os piores e inferiores a resultados obtidos com os SVM's e Redes Neuronais.

A tabela 6.1 mostra os valores médios obtidos com a aplicação das redes neuronais da primeira camada.

Tabela 6.1- Valores de medida médios nas redes neuronais da primeira camada

Rede Neuronal (grupo de parâmetros)	*Medida*	*Valor (%)*
Controlo / Patológico (I(a))	Precisão	69.6
Controlo / Patológico (III)	Sensibilidade	67.3
Controlo / Patológico (I(a))	Medida-F	67.7
Dis / Lar/Outras (I(b))	Exatidão	79.65
Dis / Par/Outras (I(b))	Especificidade	84.5

Tendo em consideração os resultados atingidos com os parâmetros da fonte (b) serem idênticos aos obtidos com os parâmetros da fonte (a), foram usados apenas os (a) por serem em menor número e também por uma questão de uniformização (visto que a maioria das discriminações apresentaram melhores resultados com este grupo de parâmetros) para utilização em trabalho futuro.

A exatidão atingida com aplicação de SVM's tem valores mais altos quando comparados com as redes neuronais, mas é de salientar que os experimentos com SVM's apenas testam determinados sujeitos (Conjunto de teste) e aplicando apenas parâmetros de sujeitos pertencentes às categorias discriminadas, podendo levar a que as classes no conjunto de teste estejam desbalanceadas, já as redes neuronais, por inferior que seja o valor de exatidão, todos os sujeitos entram em análise tendo em conta o método "*leave-one-out*", outro fator a ter em conta é que fazem discriminação em três categorias (categoria 1, categoria 2 ou outra categoria).

Tendo-se apurado que as redes neuronais são mais "poderosas" para o tipo de previsões que se pretende, foram utilizadas para a continuidade desta dissertação.

Na segunda camada de Deep-learning aplicada, foi feito o experimento de retirar a informação dada pela rede neuronal que identifica entre sujeito saudável e patológico (colocou-se o primeiro nível com seis redes neuronais, retirando a R.N-1), contudo a exatidão atingida foi de 30% para discriminação entre as quatro categorias.

Os sete parâmetros (sete redes neuronais do primeiro nivel) utilizados como entrada na segunda camada, permitem atingir uma exatidão de 39,5% para análise de todos os sujeitos, o que significa que 187 indivíduos foram corretamente categorizados de acordo com a sua patologia.

7.Conclusão

Este capítulo aborda as conclusões obtidas com a realização desta dissertação.

Na realização desta dissertação foi indispensável o estudo prévio sobre ferramentas de *machine learning*.

Analisou-se estatisticamente os parâmetros *Jitter* relativo, *Shimmer* relativo, HNR, NHR e Autocorrelação para as vogais /a/, /i/ e /u/, com o intuito de descobrir diferenças significativas entre as diferentes patologias. Apurou-se que nenhum parâmetro utilizado de forma independente permite classificar corretamente. Contudo, utilizar parâmetros em simultâneo, nomeadamente *Jitter* relativo, *Shimmer* relativo, HNR, permite classificações mais assertivas.

Como principal objetivo tinha-se a aplicação de ferramentas de *machine learning* para a identificação de patologias laríngeas, onde se analisaram três tipos de ferramenta (árvores de decisão, SVM's e redes neuronais) e concluiu-se que as redes neuronais prestam um serviço mais vantajoso quando aplicadas no âmbito dos objetivos desta dissertação. Isto fez com que árvores de decisão e SVM's fossem descartados na construção do modelo final.

Todas as ferramentas testadas, foram implementadas utilizando três grupos de parâmetros, o grupo I, referente aos parâmetros fonte (a) que contêm parâmetros como *Jitter* relativo, *Shimmer* relativo e HNR, determinados em segmentos de fala estacionária (três vogais em três tons). O grupo parâmetros fonte (b) contêm os parâmetros do grupo parâmetros fonte (a) em conjunto com os parâmetros NHR e Autocorrelação, determinados de igual modo aos parâmetros fonte (a). O grupo II é baseado em coeficientes MFCC's, determinados em segmentos de fala estacionária (3 vogais em três tons). O grupo III é formado por coeficientes MFCC extraídos de fala contínua.

Conclui-se que os parâmetros do grupo I (a) são os mais promissores para a identificação das patologias aqui analisadas.

Construíram-se redes neuronais para classificar entre três hipóteses possíveis, por exemplo, um sujeito poderá pertencer à patologia 1, à patologia 2 ou a outra patologia não definida na rede. Este conceito é mais prático porque se aproxima mais da realidade, embora se soubesse à partida que a taxa de exatidão não era muito elevada.

A utilização de sete redes, faz com que todas as combinações entre categorias, sejam analisadas. Sabendo que nem todos os sujeitos tem um diagnostico correto no primeiro nível do deep-learning, não implica que a sua classificação final esteja incorreta.

Atingiu-se uma exatidão de 39,5% de sujeitos diagnosticados corretamente recorrendo a redes neuronais.

Tendo em conta os resultados mencionados no estado desta dissertação, conclui-se que apesar de os parâmetros aqui utilizados, apresentarem algumas características que permitem classificar corretamente alguns sujeitos, percebe-se que há outros parâmetros que permitem melhores resultados. onde (Henriquez P. Alonso J. B., 2009) conseguiu 82.47% para classificação de quatro categorias, nesta dissertação atingiu-se 39,5%. já para uma classificação entre saudável / patológico alcançou uma taxa de acerto de 99,69%, enquanto que neste estudo atingiu-se 72.7%.

Durante a realização desta dissertação submeteram-se artigos ao HCIST como autor "Classification of Control/Pathologic Subjects with Support Vector Machines", e como co-autor "Long Short Term Memory on Chronic Laryngitis Classification", "Harmonic to Noise Ratio Measurement - Selection of Window and Length", "Parameters for Vocal Acoustic Analysis - Cured Database", "Learning with AudioSet to Voice Pathologies Identification in Continuous Speech", "Outliers Treatment to Improve the Recognition of Voice Pathologies".

7.1. Trabalhos Futuros

Como trabalhos a desenvolver no futuro, gostaria de sugerir um estudo para descobrir o porquê de só se alcançar 39,5% de exatidão para a classificação entre as 4 classes, fazer a mesma análise utilizando outro tipo de ferramentas de *Machine Learning,* assim como, outras bases de dados. Seria, por exemplo, interessante aplicar máquinas de discriminação paraconsistente (do inglês, Discriminative Paraconsistent Machine), esta ferramenta baseia-se num modelo discriminativo, com treino supervisionado, que aplica critérios de paraconsistência e permite fazer um tratamento inteligente de incertezas e contradições.

Sugiro também que seja feita uma análise com o mesmo propósito desta dissertação, aplicando os parâmetros Line Spectral Frequencies (LSF), Mel Line Spectral Frequencies (MFSF) e Linear Predictive Cepstral Coefficients (LPCC).

Uma das bases de dados a utilizar poderia ser "VOice ICar fEDerico II" onde existem 208 vozes gravadas entre sujeitos saudáveis e patológicos, a utilização desta base de

dados poderá ser útil visto que para cada sujeito refere também informação sobre o estilo de vida, um diagnóstico médico e contêm também os resultados de dois questionários médicos específicos. Esta base de dados foi elaborada com o apoio médico da SIFEL (Società Italiana di Foniatria e Logopedia) e SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 Statement.

Outra base de dados sugerida é "Massachusetts Eye and Ear Infirmary database", desenvolvida pelo MEEI Voice and Speech Lab, onde constam mais de 1400 exemplos de vozes gravadas (vogal 'a'). É uma base de dados comercializada por Kay Elemetrics e tem uma vasta utilização em estudos realizados com o mesmo prepósito desta dissertação.

Sabe-se que alguns diagnósticos atuais são baseados em imagens obtidas por meio de laringoscopias, propõe-se que futuramente se realize o estudo, onde seja possível ter imagens (das cordas vocais, por exemplo) e parâmetros extraídos da voz, de maneira a que possam ser implementados em redes neuronais, a fim de entender se o acréscimo de características das imagens aos parâmetros podem trazer informações relevantes para a deteção de patologias.

Seria também interessante que fossem utilizadas outras patologias e realizar as mesmas análises.

Durante o desenvolvimento deste trabalho surgiu a oportunidade de candidatura ao programa *StartUp Voucher* 2018 do IAPMEI. Esta candidatura foi aceite e com isso iniciou-se o desenvolvimento de uma interface gráfica que tem como objetivo determinar patologias da fala. Para isso, aplica-se o algoritmo desenvolvido por Fernandes, 2018 (Fernandes, 2018) para extrair os parâmetros necessários que foram aplicados a redes neuronais desenvolvidas no âmbito desta dissertação (Anexo D). Propõe-se assim que este projeto possa também ser continuado com melhorias constantes e acréscimo de novas patologias.

Referências

Alves, C. (17 de Março de 2018). *Laringite: o que é, Remédios, Tratamento, Sintomas e Causas*. (OPAS) Obtido em Agosto de 2019, de https://opas.org.br/laringite-o-que-e-remedios-tratamento-sintomas-e-causas/

Alves, N. (2016). *Diagnóstico Inteligente de Patologias da Laringe.* Dissertação de mestrado em Tecnologia Biomédica , Bragança.

Amato, F. C. (2009). Early detection of voice diseases via a web-based system. *Biomedical Signal Processing and Control, ScienceDirect*, 206-211.

Arias-londono, J., Godino-Llorente, J., Sáenz-Lechón, N., Osma-Ruiz, V., & Castellanos-Domínguez, G. (2010). An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recognition (ELSEVIER).*

B.Davis, A. l. (1979). Acoustic Characteristics of Normal and Pathological Voices. *Speech and Language*, 271-335.

Baravieira, P. B. (2016). *Aplicação de uma rede neural artificial para avaliação da rugosidade e soprosidade vocal.* Tese de doutoramento, Universidade de São Paulo.

Brandt, R. V. (2012). *Modelagem Acústica para Classificação de Vozes Patológicas Utilizando Análise Paramétrica e Não Paramétrica.* Tese de Doutoramento em Engenharia Elétrica, Universidade Federal de campina Grande, Campina Grande - Brasil.

Brownless, J. (16 de Março de 2016). *Supervised and Unsupervised Machine Learning Algorithms*. Obtido em Julho de 2019, de machinelearningmastery: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Cantoni, L. A. (2017). *Disfonia-Conceito.* Obtido em 17 de Maio de 2018, de http://luizcantoni.com.br/wp-content/uploads/2013/12/Disfonias_e_Laringopatias_Dr_Luiz_Cantoni.pdf

Carvalho, A. P. (2009). *Redes Neurais Artificiais*. Obtido em novembro de 2017, de http://conteudo.icmc.usp.br/pessoas/andre/research/neural/

Como o SVM Funciona. (2015). Obtido em 12 de Dezembro de 2017, de IBM: https://www.ibm.com/support/knowledgecenter/pt-br/SS3RA7_17.1.0/modeler_mainhelp_client_ddita/clementine/SVM_howwork.html

Cordeiro, H. (2016). *Reconhecimento De Patologias da Voz usando Técnicas de Processamento da Fala.* Dissertação de Doutoramento, Universidade Nova de Lisboa.

Cordeiro, H. T. (2016). *Reconhecimento de Patologias de Voz usando Técnicas de Processamento de Fala.* Tese de Mestrado, Universidade Nova de Lisboa.

Costa, C. R. (2013). *Reconhecimento Robusto de Vogais Isoladas.* Porto.

Costa, S. (2008). *Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para discriminação de vozes patológicas.* Tese Doutoramento, Campina Grande - Brasil.

Domingos, P. (2017). *A REVOLUÇÃO DO ALGORITMO MESTRE* (2ª EDIÇÃO ed.). (F. S. Pereira, Trad.) MANUSCRITO.

Enciclopédia médica, d. f. (2001). *Enciclopédia médica da família.* Livraria Civilização Editora.

Espinola, S. d. (2014). *Análise Acústica para Classificação de Patologias da Voz empregando Análise de Componentes Principais, Redes Neurais Artificiais e Máquina de Vetores de Suporte.* Campina Grande - Paraíba - Brasil.

Fang, S.-H. T.-J.-Y.-H.-C. (19 de Março de 2018). Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. *Journal of Voice*, 1-8.

features, P. v.-7. (2014). Muhammad, G., Melhem, M. *Biomedical Signal Processing and Control*, 1-9.

Fernandes, J. (2018). *Determinação da Autocorrelação, HNR e NHR para Análise Acústica Vocal.* Instituto Politécnico de Bragança: Tese de Mestrado.

Fonseca, E. S., & Pereira, J. C. (2009). Normal versus pathological voice signals. Em *Engineering in Medicine and Biology Magazine* (pp. 44-48).

Fonseca, E. S., Guido, R. C., Scalassara, P. R., Maciel, C. D., & Pereira, J. C. (2007). Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Computers in Biology and Medicine (ELSEVIER).*

Fonseca, E. S., Guido, R. C., Scalassara, P., Maciel, C. D., & Pereira, J. C. (2007). Wavelet time-frequency analysis and least squares support vector machine for the identification of voice disorders. *Elsevier, Computers in Biology and Medicine.*

Forero, L., Kohler, M., Vellasco, M. M., & Cataldo, E. (2015). Analysis and Classification of Voice Pathologies Using Glottal Signal Parameters. Rio de Janeiro - Brasil.

Godino-Llorente, J. I., Gomez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessement system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Transations on Biomedical Engineering*, *53*, pp. 1943-1953.

Guedes, V. d. (2019). *Deep Learning aplicado a classificação de patologias da voz.* Escola Superior de Tecnologia e Gestão de Bragança. Bragança: Instituto Politécnico de Bragança. Obtido em Junho de 2019

Guimarães, I. (2004). Os problemas de voz nos professores: prevalência, causas, efeitos e formas de prevenção. *REVISTA PORTUGUESA DE SAÚDE PÚBLICA, 22*(como se produz a voz?), 33/34.

Gupta, T. (12 de Fevereiro de 2017). *Deep Learning: Overfitting*. (Towards Data Science) Obtido em Maio de 2018, de https://towardsdatascience.com/deep-learning-overfitting-846bf5b35e24

Haykin, S. (2001). *Redes Neurais - Princípios e práticas.* Porto Alegre.

Haykin, S. (2001). *REDES NEURAIS Principios e prática* (2ª Edição ed.). (P. M. Engel, Trad.) Hamilton, Ontário, Canadá: Bookman.

Henriquez P. Alonso J. B., F. M.-L.-D.-M. (2009). Characterization oh Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics. *IEEE Transactions on Audio, Speech, and Language Processing.*

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., . . . Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*.

Hossin, M. &. (March de 2015). A Review On Evaluation Metrics For Data Classification Evaluation. *International Journal of Data Mining & Knowledge Management Process (IJDKP), 5, No.2*, 1-4.

Ian Goodfellow, Y. B. (2016). *Machine Learning Basics.* (MIT, Ed.) Obtido de deeplearningbook: http://www.deeplearningbook.org

Kim, P. (2017). *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence.* Seoul, Korea (Republic of): APRESS.

Kingsbury, B., Sainath, T., & Soltau, H. (s.d.). Scalable minimum bayes risk training of deep neural network acoustic models using distributed hessian-free optimization. NY-USA.

Kohler, M. R. (2011). *Redes Neurais Artificiais para classificação de patologias vocais.* Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica. Rio de Janeiro, Brasil: PUC Rio. Obtido em Dezembro de 2018

Lanc, T. (1992). *The Importance of Input Variables to a Neural Network Fault-diagnostic System for Nuclear Power Plants.* Iowa.

Langley, P. (2001). *"The changing science of machine learning". Machine Learning.*

Lanhellas, R. (2013). *Redes Neurais Artificiais : Algoritmo Backpropagation*. (DEVMEDIA) Obtido em 2017, de Devmedia: https://www.devmedia.com.br/redes-neurais-artificiais-algoritmo-backpropagation/28559

Lieberman, P. (1961). Pertubations in Vocal Pitch. *The Journal of the Acoustic Society of America, nº5*, 597-603.

Lima, E. S. (2012). INF - Inteligência Artifical Aula15- Árvores de Decisão. Brasil. Obtido em Junho de 2018, de http://edirlei.3dgb.com.br/aulas/ia_2012_1/IA_Aula_15_Arvores_de_Decisao.pdf

Lindasalwa Muda, M. B. (3 de March de 2010). Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. (J. O. COMPUTING, Ed.) *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, 2*. Obtido em Maio de 2019, de HTTPS://SITES.GOOGLE.COM/SITE/JOURNALOFCOMPUTING/

Logan, B. (2000). Mel Frequency Cepstral Coefficients for Music Modeling. *Cambrige Research Laboratory*.

Lorena, A., & Carvalho, A. (2007). Uma Introdução às Support Vector Machines. *RITA, XIV*(2).

Lorena, A., & Carvalho, A. d. (2003). *Introdução às Máquinas de Vetores Suporte (Support Vector Machines).* São Carlos.

Lucidchart. (2018). *O que é um diagrama de árvore de decisão?* Obtido em novembro de 2017, de Lucidchart: https://www.lucidchart.com/pages/pt/tudo-sobre-%C3%A1rvores-de-decis%C3%A3o

Mackay, D. (2005). *Information Theory, Inference; and Learning Algorithms.* Cambridge University Press, Cambridge.

Manfredi, C. (2000). Adaptive Noise Energy Estimation in Pathological Speech Signals. *IEEE Transaction on Biomedical Engineering.*

MathWorks. (2018). *fitctree*. Obtido em Junho de 2018, de https://www.mathworks.com/help/stats/fitctree.html?searchHighlight=fitctree&s_tid=doc_srchtitle

MathWorks. (2018). *plot::Boxplot*. Obtido em junho de 2018, de https://www.mathworks.com/help/symbolic/mupad_ref/plot-boxplot.html?s_tid=srchtitle

MathWorks. (2019). *elliot2sig*. Obtido de Mathworks.com: https://www.mathworks.com/help/deeplearning/ref/elliot2sig.html?s_tid=srchtitle

MathWorks. (2019). *elliotsig*. Obtido em Agosto de 2019, de MathWorks.com: https://www.mathworks.com/help/deeplearning/ref/elliotsig.html

MathWorks. (2019). *leargdm*. Obtido de mathworks.com: https://www.mathworks.com/help/deeplearning/ref/learngdm.html?searchHighlight=learngdm&s_tid=doc_srchtitle

MathWorks. (2019). *tansig*. Obtido de mathworks.com: https://www.mathworks.com/help/deeplearning/ref/tansig.html

MathWorks. (2019). *trainlm*. Obtido de mathworks.com: https://www.mathworks.com/help/deeplearning/ref/trainlm.html?searchHighlight=trainlm&s_tid=doc_srchtitle

MathWorks. (2019). *trainscg*. Obtido de mathworks.com: https://www.mathworks.com/help/deeplearning/ref/trainscg.html?searchHighlight=trainscg&s_tid=doc_srchtitle

MathWorks. (s.d.). *classregtree*. Obtido em Junho de 2018, de https://www.mathworks.com/help/stats/classregtree.html?searchHighlight=classregtree&s_tid=doc_srchtitle

MathWorks. (s.d.). *Deep Learning*. Obtido em dezembro de 2017, de https://www.mathworks.com/discovery/deep-learning.html

Matsunaga, V. Y. (2012). *Curso de Redes Neurais utilizando o MATLAB.* Belém-Pará-Brasil.

Matuck, G. (2005). *Processamento de Sinais de Voz Padrões Comportamentais por Redes Neurais Artificiais.* São José dos Campos.

Measurements, A. o. (2014). Teixeira, J.P.; Gonçalves, A. *CENTERIS 2014 / ProjMAN 2014 / HCIST 2014.*

Medscape. (2017). *Acute Laryngitis*. (Medscape) Obtido em 2 de Novembro de 2017, de https://emedicine.medscape.com/article/864671-overview

Meloni, R. (2009). *Classificação Supervisionada.*

Michael Paluszek, S. T. (2017). *MATLAB Machine Learning.* New Jersey USA: APRESS.

Mitchell, T. (1997). *Machine Learning. McGraw Hill.*

networks, D. A. (2015). *Deep Art: creative neural networks*. (altitude) Obtido em 22 de Março de 2019, de http://cxblog.altitude.com/creative-neural-networks

Nielson, M. (2015). *Neural Networks and Deep Learning.*

Ortigueira, M. (2005). *Processamento Digital de Sinais.* Lisboa: Fundação Gulbenkian.

Panek, D., Skalski, A., Gajda, J., & Tadeusiewicz, R. (2015). Accoustic Analysis Assessement In Speech Pathology Detetion. *AMCS.*

Parraga, A. (2002). *Aplicação de Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas.* Rio Grande do Sul - Brasil.

Peres, R. (Julho de 2017). ALGORITMO BACKPROPAGATION . *PROGRAMAR*, pp. 16-18.

PUC-Rio (Ed.). (s.d.). Atributos para Reconhecimento de Voz Distribuído.

Rabiner, L., & Schafer, R. (2011). *Theory and Application of Digital Speech Processing.* Prentice Hall.

Remes, C. L. (2013). *Caracterização por simulação numérica de painéis fotovoltaicos e método de rastreamento do máximo pontode potência baseado em redes neuronais artificiais.* UNIVERSIDADE DO ESTADO DE SANTA CATARINA , CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT. Joinville, SC: UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC . Obtido de https://www.researchgate.net/profile/Chrystian_Remes

Romano, C. (Agosto de 2017). *O que é a disfonia, como evitar e como tratar?* Obtido em 17 de Maio de 2018, de https://blog.cristianeromano.com.br/disfonia-o-que-e-como-evitar/

Sasaki, C. T. (2017). *Laringitis*. (MANUAL MERCK) Obtido em 2 de Novembro de 2017, de http://www.merckmanuals.com/es-us/professional/trastornos-otorrinolaringol%C3%B3gico/trastornos-de-la-laringe/laringitis

Sasaki, C. T. (s.d.). *Paralisia das cordas vocais*. (Versão Saúde para a Familia) Obtido em 20 de Junho de 2018, de msdmanuals.com: https://www.msdmanuals.com/pt-

pt/casa/dist%C3%BArbios-do-ouvido,-nariz-e-garganta/doen%C3%A7as-da-boca-e-da-garganta/paralisia-das-cordas-vocais

Selamtzis, A. C. (2018). Effect of vowel context in cepstral and entropy analysis of pathological voices. *Biomedical Signal Processing and Control*.

Sobre o SVM. (2015). Obtido em 12 de Dezembro de 2017, de IBM: https://www.ibm.com/support/knowledgecenter/pt-br/SS3RA7_17.1.0/modeler_mainhelp_client_ddita/clementine/SVM_about.html

Suárez, E. (2014). Tutorial sobre Máquinas de Vectores Soporte (SVM). Madrid: Dpto.de Inteligencia Artificial, ETS de Ingeniería Informática, Uniersidade Nacional de Educación a Distancia (UNED).

Teixeira, F., Fernandes, J., Guedes, V., Júnior, A., Teixeira, & J.P. (2018). Classification of Control/Pathologic Subjects with Support Vector Machines. *Classification of Control/Pathologic Subjects with Support Vector Machines, 138*, pp. 272-279. Obtido de https://www.sciencedirect.com/science/article/pii/S1877050918316727

Teixeira, J. P., & Fernandes, P. (Dezembro de 2015). Acoustic Analysis of Vocal Dysphonia. *Procedia Computer Science - ELSEVIER*.

Teixeira, J. P., & Gonçalves, A. (2016). Algorithm for jitter and shimmer measurement in pathologic voices. Em C. 2016 (Ed.), *HCist 2016.* Procedia Computer Science.

Teixeira, J. P., Ferreira, D. B., & Carneiro, S. M. (2011). Análise Acústica Vocal - Determinação do Jitter e Shimmer para diagnóstico de patologias da fala. Obtido de https://bibliotecadigital.ipb.pt/bitstream/10198/7282/1/artigo_publicado.pdf

Teixeira, J. P., Ferreira, D. B., & Carneiro, S. M. (2011). Análise Acústica Vocal - Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala.

Teixeira, J. P., Teixeira, F., Fernandes, J., & Fernandes, P. O. (2018). Acoustic Analysis of Chronic Laryngitis - satistical Analysis of Sustained Speech Parameters.

Teixeira, J., & Gonçalves, A. (2014). Accuracy of Jitter and Shimmer Measurements. *Procedia Technology, 16,*, 1190-1199.

Teixeira, J., Fernandes, J., Teixeira, F., & Odete, P. (2018). Acoustic Analysis of Chronic Laryngitis.

Teixeira, J., Ferreira, D., & Carneiro, S. (2011). Análise Acústica Vocal-Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala. *CLME.*

Teixeira, J., Oliveira, C., & Lopes, C. (2013). Vocal Acoustic Analysis- Jitter, Shimmer and HNR Parameters. *CENTERIS 2013.*

Tiwari, V. (10 de Feb de 2010). MFCC and its applications in speaker recognition. *MFCC and its applications in speaker recognition, 1*. Obtido em 2019, de https://pdfs.semanticscholar.org/b4e9/c14c67b8aa431a40041cce0a3564144e1a2a.pdf

Umapathy, K., & Krishnan, S. (2005). Feature Analysis of Pathological Speech Signals Using Local Discriminant Bases Technique. *IEEE Med. Biol. Eng. Compu.*, *43*.

Walia, A. S. (1 de Junho de 2017). *The Vanishing Gradient Problem*. (Medium) Obtido em Maio de 2018, de https://medium.com/@anishsingh20/the-vanishing-gradient-problem-48ae7f501257

Wang, X. Z. (2011). Discrimination Between Pathological and Normal Voices Using GMM-SVM. *Journal of Voice*.

Waseem Rawat, Z. W. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Florida 1710, South Africa: mitpressjournals. Obtido de https://www.mitpressjournals.org/doi/pdfplus/10.1162/neco_a_00990

Webb, G. I. (2011). Leave-One-Out Cross-Validation. Em G. I. Claude Sammut (Ed.), *Encyclopedia of Machine Learning*, (pp. 600-602). Boston. Obtido em june de 2019, de SpringerLink: https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_469

Anexos

Anexo A –Exatidões obtidas com aplicação de SVM's para diferentes combinações.

As precisões referidas em anexo foram calculadas de acordo com a equação 20.

As precisões referidas em anexo são expressas em percentagem (%).

Repare-se que a aplicação do método 'QP' em simultâneo com a função 'MLP', representado neste anexo por '**', não apresenta nenhuma taxa de exatidão associada, porque devido à quantidade de dados o problema não sofre uma otimização, que faz com que seja um problema não convexo.

A.1- Taxas de exatidão obtidas com aplicação de SVM's para diferentes combinações, para discriminação entre sujeitos saudáveis e sujeitos patológicos.

Neste anexo, é mencionada a taxa de exatidão obtida, com a aplicação de SVM's para a discriminação entre sujeitos saudáveis e sujeitos portadores de patologia.

Como mencionado no decorrer desta dissertação, foram aplicados diferentes métodos e diferentes funções de *Kernel*.

As tabelas A, B, C e D mostram as taxas de exatidão para distinção entre sujeitos saudáveis/controlo e sujeitos portadores de patologias. Utilizando o método SMO com a função MLP com os parâmetros do grupo I (a) (tabela A) atingiu-se a taxa de exatidão mais elevada, 80,7%, no entanto, para o tipo de distinção em causa, considera-se que este valor não pode ser considerado um valor de importância relevante. Verifica-se, também, que a utilização dos parâmetros do grupo I (b), em condições de utilização idênticas, apresenta uma taxa de exatidão muito semelhante.

Tab. A - Discriminação entre sujeitos de controlo e patológicos utilizando os parâmetros do grupo I (a)

Método	Função				
	Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	23,5	32,7	15,3	20,49	**
SMO	25,4	33	15,3	20,8	80,7
LS	31,2	43,1	15,3	27,5	44,95

Tab. B - Discriminação entre sujeitos de controlo e patológicos utilizando os parâmetros do grupo I (b)

Método	Função				
	Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	19,9	23,5	15,3	35,5	**
SMO	20,8	26,6	15,3	34,6	79,2
LS	31,8	29	15,3	32,4	50,8

Tab. C - Discriminação entre sujeitos de controlo e patológicos utilizando os parâmetros do grupo II

Método	Função				
	Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	46,23	22,59	18,52	34,8	**
SMO	43,3	24,4	18,52	35,56	61,1
LS	54,4	24,81	18,52	35,19	48,5

Tab. D - Discriminação entre sujeitos de controlo e patológicos utilizando os parâmetros do grupo III

Método	Função				
	Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	30,24	17,18	17,18	17,18	**
SMO	30,24	16,8	17,18	17,18	28,9
LS	28,5	17,18	17,18	17,18	52,58

Como não se obteve uma taxa de exatidão relativamente alta utilizando a totalidade dos sujeitos patológicos, resolveu-se discriminar os sujeitos patológicos de acordo com a sua patologia, assim a ideia passa a ser criar todas as possibilidades entre as quatro categorias existentes, referido em A.2.

A.2- Taxas de exatidão obtidas com aplicação de SVM's para diferentes combinações, para discriminação entre as diferentes categorias.

Este anexo menciona as precisões obtidas com aplicação de SVM's para diferentes combinações, onde se pretende distinguir as diferentes categorias.

Para as diferentes opções foram, também, testados os diferentes métodos com as diferentes funções. A exatidão correspondente é demonstrada nas tabelas E a P.

Para os parâmetros do grupo I (a), concluiu-se que a aplicação dos diferentes métodos demonstra poucas diferenças significativas. No entanto, destaca-se uma exatidão de 80% aplicando qualquer um dos métodos com a função 'RBF' na discriminação Controlo vs. L. Crónica.

Pela análise das tabelas, depara-se que a função 'Quadratic' apresenta sempre uma exatidão inferior, por isto, pode ser descartada em testes futuros quando realizados com este grupo de parâmetros.

Tab. E - Discriminação das diferentes categorias aplicando o método QP com parâmetros do grupo I (a)

Método	**Discriminação**	**Função**				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	Controlo vs. Disfonia	73,1	70,1	73,1	65,7	**
	Controlo vs. P.C. Vocais	67,4	73,9	65,2	69,6	
	Controlo vs. L. Crónica	58,3	66,7	80	63,3	
	Disfonia vs P.C. Vocais	49,2	49,2	65,6	54	
	Disfonia vs. L. Crónica	51,7	51,7	58,6	62	
	P.C. Vocais vs. L. Crónica	55,56	59,3	70,4	68,5	

Tab. F - Discriminação das diferentes categorias aplicando o método SMO com parâmetros do grupo I (a)

Método	**Discriminação**	**Função**				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
SMO	Controlo vs. Disfonia	76,1	70,1	73,1	68,7	73,1
	Controlo vs. P.C. Vocais	67,4	71,7	67,4	69,6	60,9
	Controlo vs. L. Crónica	50	66,7	80	68,3	50
	Disfonia vs P.C. Vocais	47,5	47,5	63,9	57,4	50,8
	Disfonia vs. L. Crónica	51,7	51,7	62	55,2	55,2
	P.C. Vocais vs. L. Crónica	55,56	59,3	72,2	68,5	61,1

Tab. G - Discriminação das diferentes categorias aplicando o método LS com parâmetros do grupo I (a)

Método	Discriminação	Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
LS	Controlo vs. Disfonia	64,2	59,7	73,1	65,67	52,2
	Controlo vs. P.C. Vocais	67,4	70,65	65,2	67,4	41,3
	Controlo vs. L. Crónica	58,3	65	80	60	36,67
	Disfonia vs P.C. Vocais	52,5	50,8	67,2	52,5	60,66
	Disfonia vs. L. Crónica	51,7	55,2	58,6	62	51,7
	P.C. Vocais vs. L. Crónica	55,56	57,4	70,4	64,8	55,56

O método SMO para discriminar controlo de disfonia e controlo de P.C.vocais, é o que apresenta precisões maiores, contudo, as precisões alcançadas com estes parâmetros ficam um bocado longe do desejado. Para discriminar disfonia de P.C. vocais e de L.Crónica o método LS apresenta melhores precisões, quando utilizado com a função '*polynomial'*.

P.C. Vocais e L.Crónica atinge um valor máximo de exatidão utilizando a função RBF.

Tab. H - Discriminação das diferentes categorias aplicando o método QP com parâmetros do grupo I (b)

Método	Discriminação	Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	Controlo vs. Disfonia	71,6	79,1	74,6	68,7	**
	Controlo vs. P.C. Vocais	69,6	73,9	66,3	68,5	
	Controlo vs. L. Crónica	58,3	63,3	81,7	66,7	
	Disfonia vs P.C. Vocais	57,4	52,45	67,2	45,9	
	Disfonia vs. L. Crónica	48,3	41,4	62	51,7	
	P.C. Vocais vs. L. Crónica	57,4	63	70,4	68,5	

Tab. I - Discriminação das diferentes categorias aplicando o método SMO com parâmetros do grupo I (b)

Método	Discriminação	Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
SMO	Controlo vs. Disfonia	71,6	79,1	74,6	65,7	55,2
	Controlo vs. P.C. Vocais	71,7	73,9	65,2	67,4	66,3
	Controlo vs. L. Crónica	55	63,3	83,3	68,3	46,7
	Disfonia vs P.C. Vocais	57,4	50,8	63,9	49,2	47,5
	Disfonia vs. L. Crónica	51,7	41,4	62	51,7	44,8
	P.C. Vocais vs. L. Crónica	55,6	61,1	70,4	68,5	44,4

Tab. J Discriminação das diferentes categorias aplicando o método LS com parâmetros do grupo I (b)

Método	Discriminação	Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
LS	Controlo vs. Disfonia	70,1	71,6	74,6	62,7	46,3
	Controlo vs. P.C. Vocais	67,4	69,6	67,4	66,3	46,7
	Controlo vs. L. Crónica	60	66,7	81,7	63,3	53,3
	Disfonia vs P.C. Vocais	55,7	49,1	67,2	50,8	49,2
	Disfonia vs. L. Crónica	48,3	44,8	62	58,6	44,8
	P.C. Vocais vs. L. Crónica	53,7	53,7	70,4	63	55,6

À semelhança do mencionado anteriormente, a distinção entre controlo e L. Crónica, apresenta melhores resultados, atingindo 83,3% de exatidão aplicando o método 'SMO' com a função 'RBF'.

Uma diferença significativa encontra-se na distinção entre controlo e disfonia utilizando o método 'QP' e 'SMO' com a função '*Polynomial*'.

Pela análise feita, percebe-se que o facto de existirem parâmetros novos para ajudar na classificação, nem sempre se traduz em resultados práticos. (Sendo que, a diferença do grupo de parâmetros I (a) para o grupo de parâmetros I (b) consiste em ser formado por mais dois parâmetros (NHR e Autocorrelação)).

As diferenças significativas para este grupo de parâmetros encontram-se na função '*Polynomial'* para distinção entre sujeitos saudáveis e portadores de disfonia, e na função linear para distinção entre disfonia e P.C.Vocais (principalmente no método 'QP' e 'SMO'). No método 'LS' com função 'MLP' houve uma melhoria de 16,6 %, no entanto esta melhoria não traz vantagens, visto que, a exatidão aqui alcançada já foi ultrapassada utilizando outras funções.

Apesar de algumas melhorias, também o acréscimo de parâmetros fez com que algumas taxas de exatidão fossem reduzidas significativamente, nota-se em especial na distinção entre disfonia e L.Crónica com a função '*Polynomial*' independentemente do método. No geral a aplicação da função MLP piorou o desempenho analisando o grupo de parâmetros em causa.

Tab. K- Discriminação das diferentes categorias aplicando o método QP com parâmetros do grupo II

Método	Discriminação	Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	Controlo vs. Disfonia	49,2	68,7	74,6	65,7	**
	Controlo vs. P.C. Vocais	53,3	62	54,4	65,2	
	Controlo vs. L. Crónica	66,7	80	83,3	71,7	
	Disfonia vs. P.C. Vocais	55,7	65,6	72,1	67,2	
	Disfonia vs. L. Crónica	48,3	65,5	65,5	58,6	
	P.C. Vocais vs. L. Crónica	75,9	81,5	81,5	79,6	

Tab. L - Discriminação das diferentes categorias aplicando o método SMO com parâmetros do grupo II

Método		Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
SMO	Controlo vs. Disfonia	47,8	64,2	74,6	65,7	32,8
	Controlo vs. P.C. Vocais	***	63	54,3	65,2	56,5
	Controlo vs. L. Crónica	68,3	78,3	83,3	70	81, 7
	Disfonia vs. P.C. Vocais	52,5	67,2	72,1	65,6	65,6
	Disfonia vs. L. Crónica	51,7	69	65,5	58,6	55,2
	P.C. Vocais vs. L. Crónica	76	85,2	81,5	79,6	79,6

***- Não converge dentro do número máximo de iterações.

Tab. M - Discriminação das diferentes categorias aplicando o método LS com parâmetros do grupo II

Método		Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
LS	Controlo vs. Disfonia	65,7	65,7	74,6	64,2	53,7
	Controlo vs. P.C. Vocais	62	63	54,3	64,1	57,6
	Controlo vs. L. Crónica	55	78,8	83,3	71,67	65
	Disfonia vs. P.C. Vocais	59	67,2	72,1	67,2	70,5
	Disfonia vs. L. Crónica	41,4	69	65,5	58,6	44,8
	P.C. Vocais vs. L. Crónica	64,8	83,3	81,5	79,63	46,3

Utilizando os parâmetros do grupo III, a distinção com exatidão mais alta entre sujeitos de controlo e sujeitos portadores de L.Crónica, independentemente do método aplicado, atinge-se uma exatidão de 83,3%.

A exatidão é na sua maioria semelhante, independentemente do método escolhido.

Tab. N - Discriminação das diferentes categorias aplicando o método QP com parâmetros do grupo III

Método		Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
QP	Controlo vs. Disfonia	77,6	74,6	74,6	74,6	**
	Controlo vs. P.C. Vocais	67,4	54,3	54,3	60,9	
	Controlo vs. L. Crónica	75	83,3	83,3	81,7	
	Disfonia vs. P.C. Vocais	60,7	72,1	72,1	72,1	
	Disfonia vs. L. Crónica	58,6	65,5	65,5	62	
	P.C. Vocais vs. L. Crónica	72,2	81,5	81,5	79,6	

Tab. O - Discriminação das diferentes categorias aplicando o método SMO com parâmetros do grupo III

Método		Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
SMO	Controlo vs. Disfonia	77,6	73,1	74,6	74,6	76,1
	Controlo vs. P.C. Vocais	67,4	53,3	54,3	57,6	59,8
	Controlo vs. L. Crónica	75	81,7	83,3	81,7	81,7
	Disfonia vs. P.C. Vocais	60,7	70,5	72,1	72,1	54
	Disfonia vs. L. Crónica	58,6	62	65,5	65,5	48,3
	P.C. Vocais vs. L. Crónica	72,2	79,6	81,5	79,6	79,6

Tab. P - Discriminação das diferentes categorias aplicando o método LS com parâmetros do grupo III

Método		Função				
		Linear	*Polynomial*	*RBF*	*Quadratic*	*MLP*
LS	Controlo vs. Disfonia	80,6	74,6	74,6	74,6	49,3
	Controlo vs. P.C. Vocais	66,3	63	54,3	57,6	54,4
	Controlo vs. L. Crónica	73,3	83,3	83,3	81,7	45
	Disfonia vs. P.C. Vocais	67,2	72,1	72,1	72,1	54,1
	Disfonia vs. L. Crónica	58,6	65,5	65,5	65,5	48,3
	P.C. Vocais vs. L. Crónica	72,2	81,49	81,5	79,6	55,6

Anexo B – Valores obtidos com as redes da primeira camada no Deep-learning

Neste anexo são demonstrados os resultados obtidos nas experiências realizadas com a aplicação de redes neuronais artificiais. Os melhores valores são referenciados a **negrito**.

Na tabela 1, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre saudável/patológico.

Tabela 1- Valores obtidos para classificação entre saudável / patológico com os diferentes grupos de parâmetros

	Controlo / Patológico			
Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Exatidão (%)	**72,73**	72,3	67,23	71,04
Precisão (%)	**69,59**	68,04	53,09	57,22
Sensibilidade (%)	65,85	65,67	61,68	**67,27**
Especificidade (%)	**77,99**	77,21	70,26	73,05
Medida F	**67,67**	66,84	57,06	61,84
Tipo de Rede	Feed-Forward			
nº de nós	50	60	65	75
funcao c.escondida	elliot2sig			
função c. de saída	elliot2sig	tansig	elliot2sig	tansig
função de treino	trainlm	trainscg		
função de aprendizagem	Padrão do matlab*			

Na tabela 2, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Controlo/Disfonia/Outras.

Tabela 2- Valores obtidos para classificação entre Controlo/Disfonia/Outras com os diferentes grupos de parâmetros

		Controlo / Disfonia / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Controlo vs Outras	Exatidão (%)	59,55	**65,52**	57,21	62,44
	Precisão (%)	41,24	41,24	46,91	**47,42**
	Sensibilidade (%)	62,02	**75,47**	53,85	65,25
	Especificidade (%)	58,39	**62**	59,45	60,92
	Medida F	49,54	53,33	50,14	**54,93**
Disfonia vs Outras	Exatidão (%)	74,07	**77,78**	76,34	73,18
	Precisão (%)	5,88	18,18	20	**20,51**
	Sensibilidade (%)	1,45	2,9	2,9	**11,59**
	Especificidade (%)	93,73	96,7	**96,77**	88,69
	Medida F	2,33	5	5,06	**14,81**
Outras vs (Controlo+ disfonia)	Exatidão (%)	52,29	**57,33**	54,02	55,53
	Precisão (%)	75,71	**87,62**	70,95	71,9
	Sensibilidade (%)	48,62	**51,69**	50,68	51,54
	Especificidade (%)	61,36	**75,93**	60,39	62,89
	Medida F	59,22	**65,02**	59,13	60,04
	Tipo de Rede	Feed-Forward			
	nº de nós	45	50	65	60
	funcao c.escondida	elliot2sig			
	função c. de saída	tansig		elliot2sig	tansig
	função de treino	trainlm		trainscg	
	função de aprendizagem	Padrão do matlab*			

Na tabela 3, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Controlo/Laringite/Outras.

Tabela 3- Valores obtidos para classificação entre Controlo/Laringite/Outras com os diferentes grupos de parâmetros

		Controlo / Laringite / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Controlo vs Outras	Exatidão (%)	64,43	62,73	61,12	**64,73**
	Precisão (%)	45,88	37,63	52,06	**58,76**
	Sensibilidade (%)	64,49	**64,6**	55,8	61,29
	Especificidade (%)	64,41	62,07	64,77	**67,35**
	Medida F	53,61	47,56	53,87	**60**
Laringite vs Outras	Exatidão (%)	85,32	84,95	**85,8**	82,54
	Precisão (%)	11,11	**18,18**	16,67	9,01
	Sensibilidade (%)	2,44	4,88	2,44	4,88
	Especificidade (%)	97,2	96,76	**98,19**	93,27
	Medida F	4	7,69	4,26	**63,35**
Outras vs (Controlo+ Laringite)	Exatidão (%)	60	58,15	59,65	**61,18**
	Precisão (%)	79,41	**82,35**	71,43	68,49
	Sensibilidade (%)	57,98	56,16	59,44	**61,51**
	Especificidade (%)	**64,75**	64,1	60	60,73
	Medida F	**67,02**	66,78	64,89	64,81
	Tipo de Rede	Feed-Forward			
	nº de nós	10	25	40	50
	funcao c.escondida	elliot2sig	tansig	elliotsig	elliot2sig
	função c. de saída	elliot2sig	tansig	elliot2sig	elliot2sig
	função de treino	trainscg			
	função de aprendizagem	Padrão do matlab*			

Na tabela 4, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Controlo/Paralisia/Outras.

Tabela 4- Valores obtidos para classificação entre Controlo/Paralisia/Outras com os diferentes grupos de parâmetros

		Controlo / Paralisia / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Controlo vs Outras	Exatidão (%)	**61,26**	58,17	54,64	58,97
	Precisão (%)	41,24	32,47	36,08	**46,91**
	Sensibilidade (%)	74,77	**80,77**	62,5	65,47
	Especificidade (%)	**55,64**	51,66	51,18	55,02
	Medida F	53,16	46,32	45,75	**54,65**
Paralisia vs Outras	Exatidão (%)	**63,71**	60,78	59,17	60,45
	Precisão (%)	60,09	**71,11**	67,42	63,11
	Sensibilidade (%)	**44,97**	37,87	35,5	38,46
	Especificidade (%)	81,22	**84,24**	82,84	80
	Medida F	**54,48**	49,42	46,51	47,79
Outras vs (Controlo+ Paralisia)	Exatidão (%)	49,01	43,56	45,25	**49,77**
	Precisão (%)	60,91	**69,09**	63,64	55,45
	Sensibilidade (%)	26,17	24,92	25,74	**26,41**
	Especificidade (%)	78,39	**78,88**	76,47	76,1
	Medida F	36,61	36,63	**36,65**	35,78
	Tipo de Rede	Feed-Forward	newcf	Feed-Forward	
	nº de nós	20	35	80	98
	funcao c.escondida	elliot2sig	radbas	elliot2sig	elliot2sig
	função c. de saída	elliotsig	tansig		
	função de treino	trainlm		trainscg	
	função de aprendizagem	Padrão do matlab*			

Na tabela 5, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Disfonia/Paralisia/Outras.

Tabela 5- Valores obtidos para classificação entre Disfonia/Paralisia/Outras com os diferentes grupos de parâmetros

		Disfonia / Paralisia / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Disfonia vs Outras	Exatidão (%)	76,44	**79,23**	72,38	72,84
	Precisão (%)	4,35	4,35	2,9	**5,8**
	Sensibilidade (%)	15,79	**42,86**	6,67	13,33
	Especificidade (%)	79,94	**80**	78,66	78,69
	Medida F	6,82	7,89	4,04	**8,05**
Paralisia vs Outras	Exatidão (%)	**65,52**	64,96	63,2	60,1
	Precisão (%)	**76,36**	72,73	66,67	52,8
	Sensibilidade (%)	24,85	23,67	28,4	**39,05**
	Especificidade (%)	**94,51**	93,8	89,33	75,11
	Medida F	37,5	35,71	39,83	**44,9**
Outras vs (Disfonia+ Paralisia)	Exatidão (%)	**58,08**	57,42	54,49	54,34
	Precisão (%)	94,04	**95,32**	84,68	74,04
	Sensibilidade (%)	**55,39**	54,5	53,64	54,72
	Especificidade (%)	76,27	**79,63**	58,14	53,44
	Medida F	**69,72**	69,35	65,68	62,93
	Tipo de Rede	newpr	Feed-Forward		
	nº de nós	20	25	60	75
	funcao c.escondida	elliot2sig	elliot2sig		elliot2sig
	função c. de saída	tansig	tansig		
	função de treino		trainscg		
	função de aprendizagem	learngdm	Padrão do matlab*		

Na tabela 6, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Disfonia/Laringite/Outras.

Tabela 6- Valores obtidos para classificação entre Disfonia/Laringite/Outras com os diferentes grupos de parâmetros

		Disfonia / Laringite / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Disfonia vs Outras	Exatidão (%)	78,09	**80,95**	61,8	53,75
	Precisão (%)	2,9	8,7	31,88	**40,58**
	Sensibilidade (%)	6,9	26,09	16,67	16,28
	Especificidade (%)	83,25	**84,13**	83,15	82,02
	Medida F	4,08	13,04	21,89	23,24
Laringite vs Outras	Exatidão (%)	**87,47**	85,21	76,97	68,69
	Precisão (%)	11,11	0	**16,98**	10,96
	Sensibilidade (%)	2,44	0	**21,95**	19,51
	Especificidade (%)	**97,66**	94,97	84,78	76,1
	Medida F	4	0	**19,15**	14,04
Outras vs (Disfonia+ Laringite)	Exatidão (%)	71,43	**72,81**	55,34	47,99
	Precisão (%)	91,46	**92,01**	61,43	49,31
	Sensibilidade (%)	76,32	77,31	77,43	**78,51**
	Especificidade (%)	8,82	17,14	**18,13**	16,36
	Medida F	83,21	**84,03**	68,51	60,58
	Tipo de Rede		Feed-Forward		
	nº de nós	30	50	80	140
	funcao c.escondida		elliotsig		
	função c. de saída				
	função de treino		trainscg		
	função de aprendizagem	learngd	learngdm		learngd

Na tabela 7, demonstra-se os valores máximos atingidos para os critérios de avaliação, quando aplicados os diferentes grupos de parâmetros, para a classificação entre Laringite/Paralisia/Outras.

Tabela 7- Valores obtidos para classificação entre Laringite/Paralisia/Outras com os diferentes grupos de parâmetros

		Laringite / Paralisia / Outras			
	Parâmetros	Grupo I (a)	Grupo I (b)	Grupo II	Grupo III
Laringite vs Outras	Exatidão (%)	84,66	**86,88**	81,9	81,68
	Precisão (%)	7,32	2,44	**9,76**	2,44
	Sensibilidade (%)	15,79	**16,67**	15,38	5
	Especificidade (%)	**88,59**	88,13	87,67	86,75
	Medida F	10	4,26	**11,94**	3,28
Paralisia vs Outras	Exatidão (%)	**68,82**	68,19	61,95	61,31
	Precisão (%)	**70,73**	68,29	52,48	50,93
	Sensibilidade (%)	34,32	33,14	31,36	**48,52**
	Especificidade (%)	**90,91**	90,3	81,68	69,62
	Medida F	46,22	44,62	39,26	**49,7**
Outras vs (Laringite+ Paralisia)	Exatidão (%)	**64,92**	64,22	58,55	57,42
	Precisão (%)	90,11	**91,63**	79,85	68,44
	Sensibilidade (%)	**63,71**	62,6	60,69	61,64
	Especificidade (%)	70,11	**72,15**	51,82	50
	Medida F	**74,65**	74,38	68,97	64,86
	Tipo de Rede	Feed-Forward			
	nº de nós	20	35	60	100
	funcao c.escondida	elliot2sig	elliotsig	elliot2sig	
	função c. de saída	elliot2sig			
	função de treino	trainscg			
	função de aprendizagem	learngd		learngdm	

Anexo C – Matrizes Confusão para análise detalhada.

Para uma análise entre Disfonia / Outras, a tabela 1 deste anexo demonstra a quantidade de sujeitos classificados de acordo com a classificação feita e o target correspondente. Os parâmetros de avaliação obtidos são representados na tabela 2.

Tabela 1 Classificação Disfonia / Outras

Disfonia / outras

		Previsto	
		Disfonia	outras
Target	Disfonia	21	48
	outras	113	291

Tabela 2 Parâmetros de avaliação obtidos para classificação Disfonia/Outras

Exatidão (%)	66,0
Precisão (%)	30,4
Sensibilidade (%)	15,7
Especificidade (%)	85,8
Medida F (%)	20,7

Para uma análise entre Laringite / Outras, a tabela 3 deste anexo demonstra a quantidade de sujeitos classificados de acordo com a classificação feita e o target correspondente. Os parâmetros de avaliação obtidos são representados na tabela 4.

Tabela 3 Classificação Laringite / Outras

Laringite / outras

		Previsto	
		Laringite	outras
Target	Laringite	8	33
	outras	87	345

Tabela 4 Parâmetros de avaliação obtidos para classificação Laringite/Outras

Exatidão (%)	74,6
Precisão (%)	19,5
Sensibilidade (%)	8,4
Especificidade (%)	91,3
Medida F (%)	11,8

Para uma análise entre Paralisia / Outras, a tabela 5 deste anexo demonstra a quantidade de sujeitos classificados de acordo com a classificação feita e o target correspondente. Os parâmetros de avaliação obtidos são representados na tabela 6.

Tabela 5 Classificação Paralisia / Outras

Paralisia / outras

		Previsto	
		Paralisia	outras
Target	Paralisia	60	109
	outras	33	271

Tabela 6 Parâmetros de avaliação obtidos para classificação Paralisia/Outras

Exatidão (%)	70,0
Precisão (%)	35,5
Sensibilidade (%)	64,5
Especificidade (%)	71,3
Medida F (%)	45,8

Anexo D – Interface gráfica, desenvolvida no âmbito do programa StartUP Voucher 2018 do IAPMEI.

Este anexo demonstra uma aplicação em desenvolvimento, no âmbito do Programa StartUP Voucher 2018 do IAPMEI.

No decorrer desta dissertação surgiu a oportunidade de concorrer ao StartUP Voucher 2018, conciliando o objetivo desta dissertação em conjunto com o trabalho desenvolvido por Fernandes (2019).

Assim, encontra-se em desenvolvimento uma interface que permite: gravar sons ou carregar ficheiros de som, que sejam extraídos e organizados, determinados parâmetros, para que aplicados a ferramentas de *"machine learning"*, nomeadamente *"deep-learning",* de modo, a auxiliar especialistas da área médica na realização diagnósticos mais eficazes.

A imagem 1 demonstra a página inicial da aplicação desenvolvida até ao momento, onde existe a opção de escolher abrir ficheiros já existentes (imagem 2) ou gravar novos sons (imagem 3).

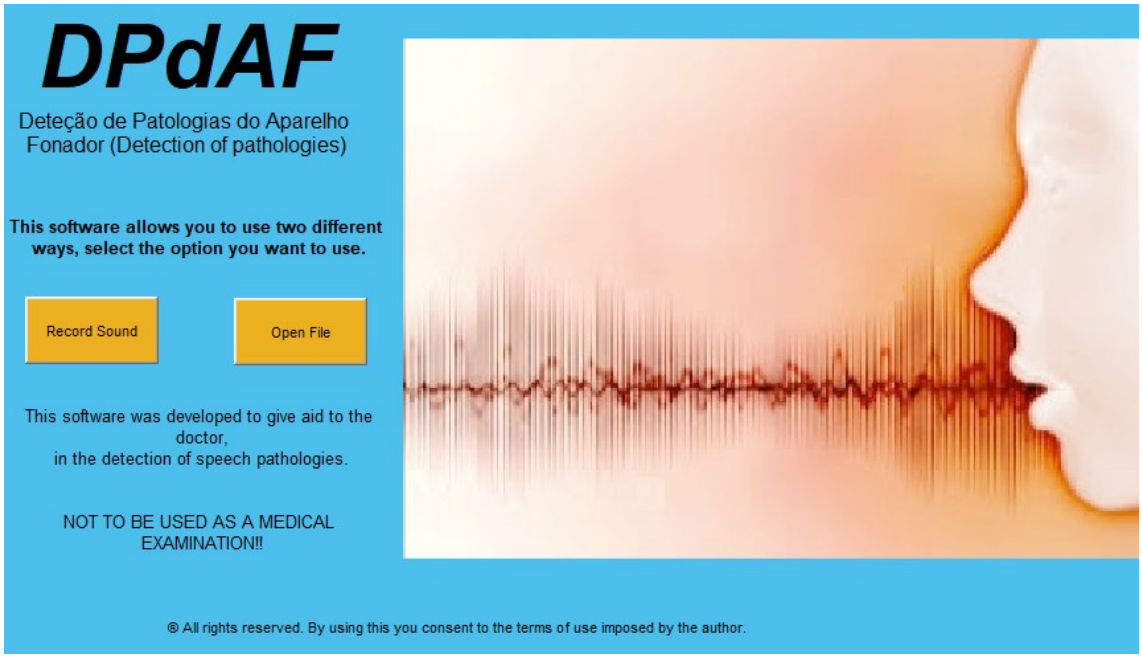

Imagem 1 – Máscara inicial da interface desenvolvida

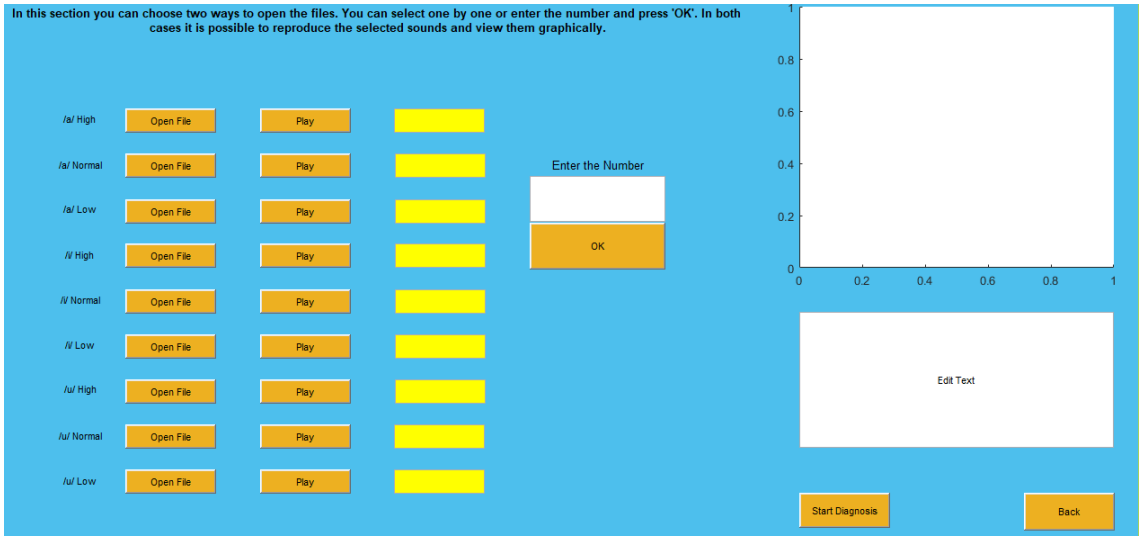

Imagem 2 - Opção de carregar ficheiros existentes

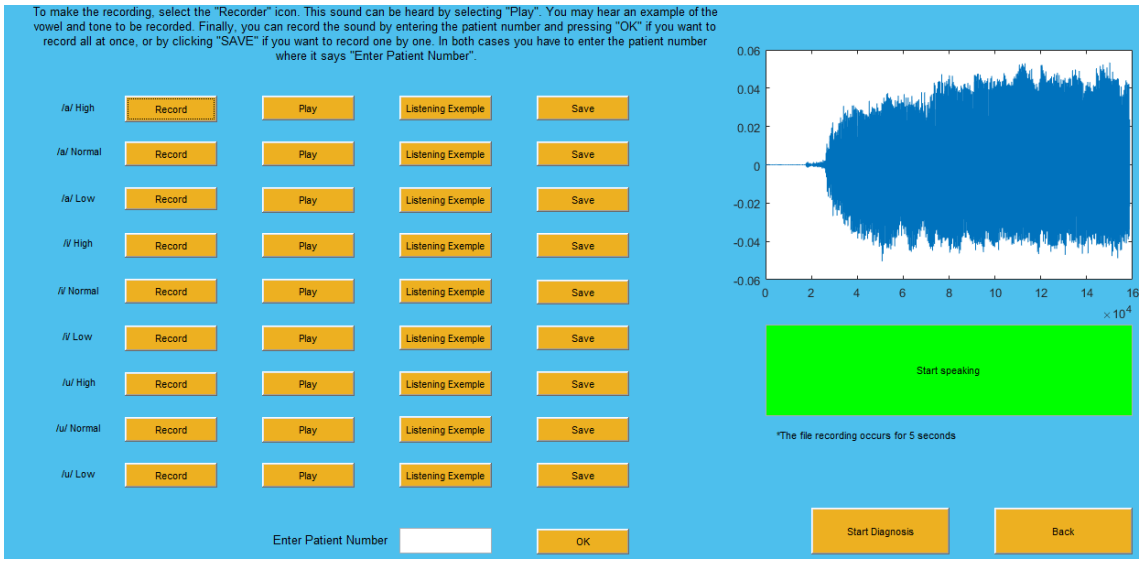

Imagem 3 - Opção de gravar ficheiros sonoros em tempo "real".

Na imagem 2, pode-se carregar ficheiros existentes de maneira independente, e ouvi-los, sendo mostrada também a sua representação gráfica. Caso se pretenda carregar os dados relativos a um determinado paciente, basta introduzir a identificação do sujeito. Caso o objetivo seja iniciar um diagnóstico com os ficheiros carregados, pressiona-se "*Start Diagnosis*".

A imagem 3, faz referência à opção em que o utilizador, pode gravar os ficheiros sonoros em tempo real. Existe a opção de reproduzir um som de exemplo, semelhante ao

pretendido. Caso o som gravado não seja pretendido pelo utilizador, pode ser novamente gravado.

Cada gravação deve ocorrer durante 5 segundos, sendo que o tempo de gravação ocorre com a janela verde, caso não se encontre no tempo de gravação a janela fica de cor vermelha.

Com os sons todos gravados é possível iniciar o diagnóstico.