Métodos de Seleção de Parâmetros para o Diagnóstico de Patologias da Laringe

LETÍCIA VEIGA CENA DA SILVA

Dissertação para obtenção do grau de Mestre em:

Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:

Prof. Dr. João Paulo Ramos Teixeira

Prof. Dr. Bruno Catarino Bispo

Bragança

Novembro de 2019

Métodos de Seleção de Parâmetros para o Diagnóstico de Patologias da Laringe

LETÍCIA VEIGA CENA DA SILVA

Dissertação para obtenção do grau de Mestre em:

Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:

Prof. Dr. João Paulo Ramos Teixeira

Prof. Dr. Bruno Catarino Bispo

Bragança

Novembro de 2019

Agradecimentos

Gostaria de agradecer primeiramente a Deus. A minha mãe, Rosana F. Veiga que sempre me deu meus melhores exemplos, sendo imagem de força e dedicação. Ao meu pai, José Marcelo Cena que me incentivou a realizar os meus maiores sonhos. As minhas irmãs, Marcela Veiga e Maria Eduarda Veiga que sempre estiveram ao meu lado e tornaram cada momento mais fácil e divertido. Aos meus avós Dirce Ferraz, Antônio Veiga, Ilca Teixeira e José Cena. Aos meus padrinhos Luciana F. Veiga e Eraldo F. de Moraes. E aos tios Adriana Veiga, Sérgio Fernandes, Andréia Cena e Fábio Alves.

Aqueles a quem considero família meu muito obrigada, nomeadamente, Rosa Letícia Mello, Luis França, Fabrício Menezes, Janaína Dias, Laísa Cardoso, Kauane Benitis, Larissa Brito, Maria Eduarda Oliveira, Camilla Shimizu, Gabriela Possebon, Nadine Martelozo, Nathália Gonçales.

Gostaria de agradecer fortemente as grandes oportunidades que o Instituto Politécnico de Bragança e a Universidade Tecnológica Federal do Paraná campus Cornélio Procópio me proporcionaram. Gostaria de estender meus cumprimentos aos professores Drs. João Paulo Teixeira e Bruno Catarino Bispo por me orientarem e disponibilizarem seu tempo. Meu muito obrigado a todos dos laboratórios LPSA-SIPECA, LCAR e UNIAG.

A todos os amigos que tive o prazer de fazer ao longo desses dos anos, desde de minha trajetória em Cornélio Procópio, Brasil em 2014/1 até a presente data em Bragança, Portugal. Quero agradecer por todos os almoços de domingo, aniversários e passeios no Cristo a Lígia Taniguchi, Lucas Miller, Matheus Presotto, Eduardo Nogueira, Paulo Yamashita, Ruhan Policarpo, Eder Iori, Afonso Sanches, Samuel Cardoso, Carolina Gasperoni, João Rafael Paulo, Karen Taniguchi, Emy Taniguchi, Patrick Oliveira, Reginaldo Junior, Wesley Gomes.

Um agradecimento especial a todos aqueles que me deram suporte, estiveram sempre junto e mesmo longe do Brasil fizeram com que me sentisse em casa. Ao Lucas Fernandes, Matheus Montanini, Luis Aguiar, Leonardo Cândido, Juliana Hermsdorf, Arthur Casarini, Luiz Miguel Vilche, Lídia França, Milena Ianela, , Kenji Matoba, Isabele Corrêa, Victor Guedes, Gustavo, Eric, Thiago, William, Ighor, Alana, Luana, Micael, Lídia, Anelise, Camila, Nágila, Edmar, João, Renato, Gustavo, John, Hiago, Gabriel, Bárbara, Ana, Miguel, Mariana.

“Na vida, não existe nada a temer, mas a entender”

Marie Curie

Resumo

Esta dissertação propõe soluções para a identificação de patologias da voz através do processamento do sinal de fala. Foram utilizados na classificação de patologias como Laringite Crónica, Disfonia e Paralisia das Cordas Vocais as redes neuronais, Multilayer Perceptron e Long-Short-Term-Memory. Os parâmetros acústicos empregados foram *jitter* relativo, *jitter* absoluto, *shimmer* relativo, *shimmer* absoluto, autocorrelação, *Harmonic to Noise Ratio*, *Noise to Harmonic Ratio* e *Mel Frequency Cepstral Coefficients*. Estes parêmetros são extraídos da base de dados *Saarbrücken Voice Database*, a partir de arquivos de áudio que contém as vogais sustentadas /a/, /i/ e /u/ nos tons baixo, normal e alto. Também empregou e testou técnicas de normalização de dados, identificação de *outliers* e seleção de parâmetros. Tais aplicações tem a finalidade de otimizar o modelo de reconhecimento, torná-lo mais eficiente e consequentemente melhorar a acurácia/exatidão do diagnóstico. Como pré-processamento utilizou-se as técnicas de normalização Z-score, Logarítmica e Raiz Quadrada para permitir uma melhor identificação dos *outliers* presente nos dados, por meio da aplicação do método do Box Plot e do Desvio Padrão. Após os experimentos, tanto o método do Desvio Padrão quanto o do Box Plot com normalização do Z-score mostraram-se muito úteis para o pré-processamento do conjunto de dados para o reconhecimento de patologias de voz. A acurácia foi melhorada entre 3 a 13 pontos em percentagem. Posteriormente, foram utilizadas as técnicas de Seleção de Parâmetros que ordenam os atributos segundo uma métrica de importância. Deste modo, os parâmetros relevantes são selecionados de acordo com o critério estabelecido pelos testes: Correlação, ReliefF, Test t de Welch, Regressão Multilinear. Ao comparar todos os algoritmos desenvolvidos, pode-se destacar que o algoritmo baseado no ReliefF teve o melhor desempenho. Com relação a acurácia teve um aumento de 9 pontos percentuais e na medida F de 8 pontos percentuais.

Palavras Chave: Classificação, Patologias da Voz, Identificação de *Outliers*, Seleção de Parâmetros.

Abstract

This thesis proposes solutions for the identification of voice pathologies through speech signal processing. Pathologies such as Chronic Laryngitis, Dysphonia and Vocal Cord Paralysis were used to classify neuronal networks, Multilayer Perceptron and Long-Short-Term-Memory. The acoustic parameters employed are relative jitter, absolute jitter, relative shimmer, absolute shimmer, autocorrelation, Harmonic to Noise Ratio, Noise to Harmonic Ratio and Mel Frequency Cepstral Coefficients. These parameters are extracted from the Saarbrücken Voice Database from audio files containing the sustained vowels /a/, /i/ and /u/ in low, normal and high tones. Data normalization, outlier identification and attribute selection techniques were used. Such applications have the purpose of optimizing the recognition model, making it more efficient and consequently improving the accuracy of the diagnosis. As preprocessing, Z-score, Logarithmic and Square Root normalization techniques were used to allow a better identification of outliers present in the data, by applying the Box Plot and Standard Deviation method. After the experiments, both the Standard Deviation method and the Z-score normalized Box Plot method proved to be very useful for data set preprocessing for speech pathology recognition. Accuracy was improved by 3 to 13 percentage points. Subsequently, we used the Parameter Selection techniques that order the attributes according to a metric of importance. Thus, the relevant parameters were selected according to the criteria established by the tests: Correlation, ReliefF, Welch T test, Multilinear Regression. When comparing all the developed algorithms, it can be highlighted that the ReliefF based algorithm had the best performance. Regarding accuracy, there was an increase of 9 percentage points and measure F of 8 percentage points.

Keywords: Classification, Vocal Pathologies, Outliers Detection, Feature Selection.

Índice de Figuras

Índice de Tabelas

Acrónimos

AM	Aprendizado de Máquina
BP	*Box Plot*
DP	Desvio Padrão
FN	Falso Negativo
FP	Falso Positivo
HNR	*Harmonic to Noise Ratio*
IA	Inteligência Artificial
IPB	Instituto Politécnico de Bragança
LSTM	*Long Short-Term Memory*
MFCC	*Mel Frequency Cepstral Coefficients*
MLP	*Multilayer Perceptron*
NHR	Noise to Harmonic Ratio
RNA	Rede Neuronal Artificial
RNN	Rede Neuronal Recorrente
SP	Seleção de Parâmetros
SVDT	Saarbrücken Voice Database
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo
VLI	Valor Limite Inferior
VLS	Valor Limite Superior

Índice

Capítulo 1: Introdução

Este capítulo é dedicado à introdução do tema, seguido pelos objetivos do trabalho, a caracterização do Estado da Arte que contém um conjunto de trabalhos relacionados e a estrutura desta dissertação.

1.1. Contextualização

A voz é o principal meio de comunicação pelo qual os homens transmitem suas ideias. E, por ser rápida e eficiente, é um instrumento de comunicação oral muito poderoso (Alcaim & Oliveira, 2011). Além de transmitir diferentes informações verbais, a voz também transmite carga emocional (Alcaim & Oliveira, 2011).

Cada indivíduo ao falar apresenta sua identidade sonora que o distingue de outro falante. O processo tradicional de detecção de patologias da laringe é extenso e invasivo (Dajer, 2010). Estes exames fundamentam-se em técnicas subjetivas e dependentes da experiência médica (Almeida, 2010).

A quantificação e caracterização vocal do indivíduo ocorre mediante um conjunto de parâmetros acústicos temporais e espectrais (Machado, 2013) tais como: periodicidade, amplitude, duração e composição espectral (Teixeira, Ferreira, & Carneiro, 2011). A partir destes parâmetros acústicos, pode-se diagnosticar, acompanhar e tratar desordens vocais de um modo rápido e eficiente (Almeida, 2010). Por ser uma técnica não invasiva, a análise acústica é bastante indicada na detecção e estudo de patologias da voz (Brockmann-Bauser, 2011; Godino-Llorente, Gómez-Vilda, & Blanco-Velasco, 2006).

A análise acústica, dentro da mineração de dados, extrai conhecimento aplicando técnicas estatísticas e computacionais (Berton, 2011) que visam medir propriedades do sinal sonoro de uma voz gravada, tanto em vogais de forma sustentada ou em discurso contínuo (Alves, 2016). Além disso, permite delinear o formato da onda sonora e avaliar determinadas características como a frequência fundamental (F0), índices de variação de frequência e perturbação da amplitude (*jitter* e *shimmer*), análise espectral, entre outros (Brockmann-Bauser, 2011).

A mineração de dados pretende resolver algumas das dificuldades enfrentadas pelas técnicas de análise tradicionais, como a escalabilidade, a alta dimensionalidade, a complexidade

e a distribuição dos conjuntos de dados. Algumas aplicações da mineração de dados são amplamente utilizadas, como a identificação de *outliers*, a seleção de parâmetros e classificação. A detecção de *outliers* desempenha um papel fundamental na descoberta de padrões nos dados, buscando aqueles cujas características diferem umas das outras (Campos, 2015). A seleção de parâmetros tem por objetivo reduzir a dimensionalidade dos conjuntos, de modo a selecionar o melhor e menor subconjunto de parâmetros (Fernandes, 2017). No caso de sinais de fala, a classificação é responsável pelo reconhecimento de vozes patológicas e patologias da fala que permite atribuir uma classe aos dados com base nas características de seus atributos (Cordeiro, 2016).

1.2. Motivação e Justificativa

Em pacientes com patologias progressivas, para promover um melhor tratamento é determinante que o diagnóstico realizado seja rápido. A subjetividade dos exames de diagnósticos termina quando são empregados classificadores inteligentes aliados a parâmetros acústicos (Alves, 2016). Portanto, a utilização de análise acústica vocal para o exame e diagnóstico de patologias é importante. Além de ser mais simples é também mais econômica que os exames tradicionais (Dajer, 2010).

Em um grande conjunto de dados, a classificação é uma tarefa humanamente impraticável devido à variedade de exemplos que precisam ser estudados para detectar os padrões. Deste modo, justifica-se a aplicação de um classificador inteligente, como a rede neuronal, para a seleção de parâmetros no conjunto (Alves, 2016).

A principal causa do surgimento dos *outliers* em base de dados são os erros humanos, de instrumentos, desvios em populações, comportamento fraudulento, mudanças ou falhas no comportamento de sistemas (Berton, 2011). A presença destes em conjuntos de dados pode causar distorção das análises e má interpretação dos dados. Deste modo, torna-se importante a identificação e correção das anomalias para garantir a confiabilidade dos resultados.

Além disso, grandes conjuntos de dados têm se tornado cada vez mais comuns, o que faz necessária a seleção e remoção de parâmetros pouco relevantes. A aplicação desta técnica leva a redução da dimensão dos dados e do gasto computacional (Barbosa, 2013).

1.3. Objetivos

O objetivo geral desta dissertação é utilizar técnicas de normalização de dados, identificação de *outliers* e seleção de parâmetros para pré-processar os parâmetros usados no diagnóstico de sujeitos patológicos e patologias. Tais aplicações tem a finalidade de otimizar o modelo de reconhecimento, torná-lo mais eficiente e consequentemente melhorar a acurácia/exatidão do diagnóstico.

Mais especificamente, tem-se os seguintes objetivos:

- Implementação de modelos baseados em Redes Neuronais Artificiais clássicas;
- Implementação de modelos baseados em Redes Recorrentes. Utilizando os conceitos da rede Long-Short-Term-Memory (LSTM);
- Identificação, tratamento de *outliers* e seleção de subconjunto de parâmetros representativos;
- Aplicação dos modelos em problemas binários para detecção de patologias na voz utilizando vogais;
- Estudo comparativo das metodologias implementadas e trabalhos relacionados.

1.4. Estado da Arte

Nesta seção será descrita uma revisão referente as investigações relacionadas com o uso de inteligência artificial para o problema de classificação das patologias da voz, apresentando-as em ordem cronológica.

As investigações a respeito de reconhecimento de patologias são consideravelmente recentes. Por volta da década de 60 surgiram trabalhos relacionados da distinção de vozes normais e vozes patológicas da identificação de caraterísticas presentes nos sinais de fala por meio da análise de vogais sustentadas. Estes trabalhos se dedicam a quantificar as perturbações em frequência e amplitude, isto é, quantificar os parâmetros acústicos *jitter* e *shimmer*, respectivamente (Lieberman & Affiliations, 1963; Iwata, 1972).

No trabalho de Dibazar, Narayanan, e Berger (2002) pretende-se desenvolver um sistema capaz de identificar as patologias utilizando como característica o MFCC (Coeficientes Mel-Cepstrais), sendo esta a primeira utilização dos coeficientes. A base de dados utilizada é a MEEI *(Massachusetts Eye and Ear Infirmary)* (MEEI, 1994) e a partir desta foram implementados diferentes classificadores. O HMM (*Hidden Markov Model*) utilizou MFCC

combinados com a frequência fundamental e obteve o melhor desempenho dentre os demais classificadores, um acerto percentual de 98,3%.

Fonseca, Guido, Scalassara, Maciel, e Pereira em 2007 apresentam um método que aplicou a energia das *wavelets* na classificação binária entre saudáveis e patológicos. Foram aferidas, em 4 sub-bandas do sinal considerando as frequências mais altas, valores referentes a *Root Mean Square* (RMS). Assim, comprovou-se que em vozes patológicas os valores de RMS são superiores. Ao lidar com os valores de RMS provenientes das duas bandas das *wavelets* e um classificador *Support Vector Machine* (SVM) foi atingida uma taxa de 91,6% de acerto. Portanto, concluiu-se que é viável a classificação de vozes patológicas ao analisar o ruído presente nas altas frequências dos sinais.

No trabalho de Arjmandi (2011) foram caracterizados 33 parâmetros, minuciosamente aferidos por meio do Multidimensional Voice Program (MDVP). Estes parâmetros foram extraídos a partir da vogal /a/ e pertencem a base de dados MEEI (MEEI, 1994). Os MFCC foram estimados com base em fala contínua e ao serem comparados com os MDVP, ambos alcançaram uma taxa de acerto na identificação de patologias de 97%. A taxa de acerto aumenta para 98,3% ao comparar os coeficientes Mel-Cepstrais combinados com a frequência fundamental.

Em 2013, Lee et al. efetuou uma investigação entre vozes saudáveis e patológicas (223 sujeitos saudáveis, 472 sujeitos com nódulos e 472 sujeitos com Paralisia das Cordas Vocais). Os parâmetros empregados no reconhecimento das doenças foram o *jitter* do período fundamental, *shimmer*, HNR, os valores e desvios de frequências dos formantes de primeira e segunda ordem. Na aplicação de parâmetros glotais (*jitter* do período fundamental e *shimmer*) foram alcançados os melhores resultados e ao serem combinados com os parâmetros com as caraterísticas do trato vocal (HNR) obtiveram um aumento na taxa de reconhecimento, melhorando ainda mais os resultados. E a partir destes, conclui-se que os formantes contêm informações que permitem classificar as vozes patológicas. Isto é, em sujeitos com patologia devido ao mal funcionamento da glote, existem desvios nos formantes produzidos.

Os trabalhos anteriores se preocupavam apenas em realizar o reconhecimento de vozes patológicas, sem fazer a identificação da patologia propriamente dita. Deste modo, as investigações descritas distinguem as vozes saudáveis das patológicas por meio de uma classificação binária.

Rosa, Pereira, e Grellet em 2000 propuseram o primeiro trabalho de classificação multiclasse de patologias. O qual, empregou 7 parâmetros de jitter do período fundamental

extraídos do ruído do sinal de fala calculado através dos filtros Kalman ou Wiener. Foram utilizadas 3 vogais sustentadas /a/, /e/ e /i/ de 25 sujeitos saudáveis e 48 sujeitos com uma ou mais patologias (no total 21 patologias distintas), provenientes de uma base de dados própria e criada para o estudo. Foram determinadas 8 classes, no qual 6 correspondem a patologias, 1 as vozes normais e 1 as demais patologias estudadas. O melhor modelo de reconhecimento alcançou cerca de 54,79% e foi obtido para 231 combinações de pares de teste com *jitter* do período fundamental. Apesar da baixa taxa de acerto concluiu-se que o *jitter* do perído fundamental é útil na identificação de patologias da voz.

Dibazar, Berger, e Narayanan em 2006 desenvolveram um trabalho que pretende identificar 5 patologias a partir dos parâmetros MFCC e o classificador HMM contendo sinais da vogal /a/ sustentada. A melhor taxa de identificação alcançada foi de 70%, onde cada modelo de HMM com 1 patologia é combinado com o HMM com as outras 4 patologias. O autor constata que os parâmetros que caracterizam o trato vocal, MFCC, são capazes de discriminar entre as doenças em estudo.

Em 2008, Hosseini e Almasganj propuseram a identificação das patologias Pólipos, Keratosis Leukoplakia, Disfonia Espasmódica Adutora e Nódulos. São testados 3 modelos de reconhecimento, os quais utilizam o SVM para comparar os Pólipos contra as outras patologias investigadas. Os parâmetros empregados na classificação são obtidos através da entropia das *wavelets*, o sinal foi separado em sub-bandas e calculou-se a entropia para cada uma delas. Foram utilizados 75% dos indivíduos para treino e 25 para teste (não foram especificados o número de indivíduos). O melhor modelo de identificação obteve 87,5% na taxa de acerto ao comparar sujeitos com Pólipos e sujeitos com Nódulos, foram obtidos 82,5% comparando Pólipos versus Nódulos e 81,81% para a comparação de sujeitos com Pólipos versus com Queratoses.

Em 2009, Scalassara et al. propôs a realização do reconhecimento de patologias utilizando a entropia relativa como parâmetro. Para a realização dos testes empregou-se a vogal sustentada /a/ de 48 falantes, subdividida em 3 grupos de amostras (indivíduos saudáveis, indivíduos com nódulo nas pregas vocais e indivíduos com edema de Reinke). A entropia é calculada em tramas de 50 ms, durante 1 segundo, tendo como objetivo caracterizar a incerteza do sinal. Ao analisar os sinais de sujeitos não patológicos e patológicos foi possível diferenciar ambos completamente, pois tem valores superiores de entropia relativa. Não sendo possível fazer a distinção entre as patologias diferentes. Os sinais de sujeitos patológicos sofrem

variações ao nível da frequência fundamental e em decorrência disto apresentam maiores valores de entropia. Portanto, a incerteza aumenta em sinais não saudáveis, devido aos valores altos de *jitter* do período fundamental e o *shimmer*.

Em 2011, Carvalho et al. aplicou as *wavelet* nos testes de identificação de patologias. Com a finalidade de confirmar a estabilidade do sinal foi empregado um sinal de 1 segundo decomposto em tramas de 24 ms, na qual cada trama é subdividida em 4. Como entrada de uma rede neuronal multicamadas tem-se um sinal, para o qual é aplicada uma *wavelet* e a cada sub-banda é aferida a energia das bandas da *wavelet*, individualmente. O melhor modelo alcançou uma taxa de acerto de 90% entre vozes não patológicas e vozes com Edema de Reinke e Nódulos nas cordas vocais, a partir da decomposição *wavelet* de 6ª ordem. Ao utilizar uma quarta classe que abrange indivíduos com diversas doenças referentes a patologias da voz de origem neurológica e *wavelets* até 7ª ordem, a taxa de acerto diminuiu para 87%. O primeiro trabalho no qual ocorre a identificação de várias patologias, totalizando 4 classes (incluindo vozes saudáveis).

Algumas publicações mais recentes referentes ao reconhecimento de vozes patológicas a partir de uma classificação multiclasse, tem-se:

Em 2015, Gonçalves elaborou um algoritmo que mensurou os parâmetros acústicos da fala e comparou vozes saudáveis e patológicas (Laringite, Disfonia Hiperfuncional, Disfonia Espasmódica, Pólipos das Cordas Vocais e Envelhecimento das Cordas Vocais). A base de dados utilizada é a *Saarbrucken Voice Database* (SVDT), onde estão disponibilizadas gravações das vogais /a/, /i/ e /u/ com tons baixo, normal, alto, variando entre tons e a gravação da frase *"Guten Morgen, wie geht es Ihnen*?" (“Bom dia, como estás?”), em alemão (Pützer & Barry). Extraídos da base de dados SVDT, os parâmetros utilizados no algoritmo são o *jitter, shimmer* e HNR. O algoritmo utilizou sinais sintetizados, de vozes saudáveis e vozes patológica, os valores dos parâmetros acústicos foram pré-estabelecidos e comparados com os resultados obtidos pelo *software* Praat. Para os sinais de vozes sintetizados, tem-se como resultado um erro inferior a 5 μs para o parâmetro Jitta e inferiores a 0,1% para o Shim. No entanto, em vozes reais, as diferenças dos erros obtidos pelo *software* Praat e o algoritmo foram desprezíveis. Através da comparação do comportamento dos parâmetros *jitter* e *shimmer* em vozes saudáveis e patológicas, utilizando um critério estatístico, dentre as patologias citadas apenas as três últimas apresentaram distinção significativa dos parâmetros, em relação ao grupo de sinais de voz de controlo.

Em 2016, Cordeiro propôs a identificação de patologias da voz empregando o processamento de fala em sinais de fala contínua. Os parâmetros utilizados nos classificadores são: *Pitch Jitter* local, HNR, MFCC, LSF (*Line Spectral Frequencies*) e MLSF (*Mel-Line Spectral Frequencies*). Foram usadas as bases de dados *Dutch Corpus of Pathological and Normal Speech* (Copas), da Universidade de São Paulo (USP) e a *Massachusetts Eye and Ear Infirmary* (MEEI). A base de dados Copas é composta por 319 sujeitos, dos quais 122 são saudáveis (nos sinais adquiridos constam vogais sustentadas e fala contínua). A base de dados da USP é constituída por 16 indivíduos saudáveis, 16 indivíduos com Edema de Reinke e 15 com nódulos (os sinais adquiridos são vogais /a/, /e/ e /i/). Da base de dados MEEI fazem parte 53 indivíduos saudáveis e 724 com patologia da voz (os sinais adquiridos são vogais /a/ extraídas da palavra "*rainbow*"). Foram utilizados como classificadores a máquinas de vetor de suporte (SVM), modelo de misturas Gaussianas, e dois discriminadores lineares, avaliados com validação cruzada. Os sinais de fala contínua foram sub-divididos em: saudáveis, com Edemas e Nódulos e Paralisia Unilateral das Corda Vocais. Ao utilizar o sinal de fala contínua, a taxa de acerto obtida foi de 84% nas três classes. Para a análise utilizando os formantes e o HNR, um algoritmo simples baseado em árvores de decisão foi empregado e a taxa de reconhecimento alcançou 95%.

Em 2016, Alves avaliou o desempenho dos sistemas inteligentes, na detecção de patologias da laringe utilizou diferentes conjuntos de dados. O conjunto 1 é constituído pelos parâmetros HNR e 4 medidas de *jitter* e *shimmer*, foram utilizadas apenas uma vogal e um tom ou com várias vogais e vários tons, e o conjunto 2, por 12 coeficientes cepstrais, frequência e largura de banda dos três primeiros formantes, frequência fundamental, energia, potência, momentos espectrais de ordem zero, um, dois, três e curtose. O banco de dados utilizado foi o *Saarbrucken Voice Database* (SVDT), aplicou-se a separação do gênero, de modo que os parâmetros femininos foram analisados separadamente dos masculinos. O número de indivíduos saudáveis selecionados foi o mesmo que o grupo patológico (balanceamento de classes). Utilizou um total de 334 indivíduos do sexo feminino e 196 do sexo masculino. Dos 334 indivíduos do sexo feminino, 126 tinham Paralisia das Cordas Vocais e 41 Disfonia e no sexo masculino 69 tinham Paralisia das Cordas Vocais e 29 Disfonia. Foram aplicadas técnicas de seleção de variáveis e redução da dimensão como a regressão linear passo a passo e análise dos componentes principais (PCA). Para a classificação entre saudável e patológico utilizou dois tipos de sistemas inteligentes: redes neuronais artificiais (RNA) e máquinas de vetor de

suporte (SVM). A partir da avaliação do desempenho de predição, o conjunto 1, não apresenta grande poder preditivo para apenas uma vogal e tom. Foram alcançados bons resultados devido ao uso de várias vogais e vários tons, somente para Disfonia. E o conjunto 2, ainda que apenas na vogal /a/ e tom normal, é possível obter melhores resultados na Paralisia das Cordas Vocais para várias vogais e vários tons. Contudo, o *jitter, shimmer* e HNR para várias vogais e tons continua a ser a melhor forma de classificar como patológico ou saudável quando se usa a Disfonia. O classificador com melhores resultados em média é a Rede Neuronal Artificial (RNA). Foram obtidas precisões de 100% para o primeiro conjunto de parâmetros, usando a Disfonia feminina e a masculina como grupo patológico; 78,9% usando a Paralisia das Cordas Vocais feminina como grupo patológico; 81,8% usando a Paralisia masculina como grupo patológico.

Em 2018, Fernandes implementou um algoritmo que determinava os parâmetros HNR, NHR e autocorrelação, e os utilizou como entradas de um sistema inteligente para diagnóstico de patologias da fala. Comparou os valores do algoritmo e do *software* Praat, para identificar a melhor janela e o seu comprimento, em número de períodos glotais. A janela de Hanning com um comprimento correspondente a 6 períodos glotais foi a escolhida. O algoritmo permite extrair os parâmetros HNR, NHR e autocorrelação com valores suficientemente próximos dos valores de referência. Foi ainda desenvolvido um algoritmo que seleciona apenas a parte do sinal onde ocorre fala, eliminando as zonas de silêncio iniciais e finais, para, posteriormente, se extrair os MFCCs, os Linear *Prediction Coefficientes* (LPC) e os *Line Spectral Frequency* (LSF). Os parâmetros foram extraídos de 9 locuções correspondentes a 3 vogais em 3 tons e a uma frase, para sujeitos com 19 patologias, mais os sujeitos de controlo. A base de dados curada iniciada numa investigação anteriormente realizada pelo autor foi complementada. Esta base de dados passou a possuir novos parâmetros e patologias sendo estes 13 coeficientes MFCC, HNR, NHR, autocorrelação, *jitter* absoluto, *jitter* relativo, *shimmer* absoluto, *shimmer* relativo. Esta base de dados curada disponibiliza um conjunto de parâmetros sobre estes sinais de fala para a investigação sobre as 19 patologias.

Em 2019, Guedes classificou as patologias relacionadas a voz utilizando conceitos de *Deep Learning.* Foi proposto a implementação dos principais modelos de *Deep Learning* para a classificação de patologias da voz em fala contínua, utilizando a frase alemã "*Guten Morgen, wie geht es Ihnen*?" da base de dados *Saarbruecken Voice Database*. Para a classificação em fala sustentada com vogais foram obtidos bons resultados, porém não se tem muitos trabalhos relacionadas a classificação utilizando fala contínua. Para as análises multiclasse e binária

foram utilizadas as patologias de Disfonia, Laringite e Paralisia das Cordas Vocais, além da classe dos saudáveis. Foi realizado um estudo para a classificação com vogais nas mesmas patologias. O melhor resultado para as vogais é de 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, é realizado 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 multiclasse é obtido 59% e 40% de medida-F para 3 classes e 4 classes, respectivamente.

Em 2019, Teixeira propôs a utilização de técnicas *Deep-Learning* e SVM em vogais sustentadas para a classificação multiclasse. Os parâmetros aplicados foram extraídos da base SVDT. Este trabalho se caracteriza com uma investigação em andamento.

1.5. Organização do Trabalho

No Capítulo 1 é realizada uma contextualização do assunto. São enunciados a motivação, os objetivos e o Estado da Arte. No Capítulo 2 é apresentado o objeto de trabalho. São enunciados a descrições de patologias na Laringe e caracterização dos parâmetros acústicos e a base de dados. O Capítulo 3 contém a fundamentação teórica das técnicas de aprendizagem computacional abordados na dissertação. O Capítulo 4 contém a fundamentação teórica acerca dos principais métodos de detecção e correção de *outliers.* O Capítulo 5 contém os resultados, análises e conclusões referentes as técnicas de tratamento de *outliers* utilizadas. No Capítulo 6 é abordada a fundamentação teórica dos principais métodos de seleção de parâmetros abordados na dissertação. No Capítulo 7 é apresentado o desenvolvimento experimental e resultados a respeito dos parâmetros selecionados. Por fim, no Capítulo 8 é apresentada uma conclusão global que compara as conclusões obtidas e introduz as possibilidades de trabalhos futuros.

Capítulo 2: Objeto de Trabalho

Este capítulo aborda os conceitos de geração da voz, a caracterização das patologias, parâmetros acústicos e a base de dados utilizada nesta dissertação.

2.1. Contextualização

O modelo fonte-filtro é fundamentado nas características do mecanismo vocal humano. Deste modo, o modelo discreto no tempo proposto é um modelo linear de produção de fala, no qual, a fonte de excitação e o aparelho vocal são considerados como dois sistemas separados (Alcaim & Oliveira, 2011).

O sinal de voz é a resposta dos sistemas de filtragem do aparelho vocal a uma ou mais fontes de excitação, e suas propriedades são especificadas ao longo do tempo em termos das características individuais da fonte e do filtro (Rabiner & Schafer, 2009).

Considerando o tempo discreto, a modelagem do processo de produção do sinal de voz $s(n)$ pode ser obtida pela convolução entre o sinal de excitação $e(n)$ correspondente ao pulso glotal e a resposta impulsiva $h(n)$ correspondente à configuração do trato vocal (Alcaim & Oliveira, 2011; Rabiner & Schafer, 2009). O modelo descrito é ilustrado na Figura 1.

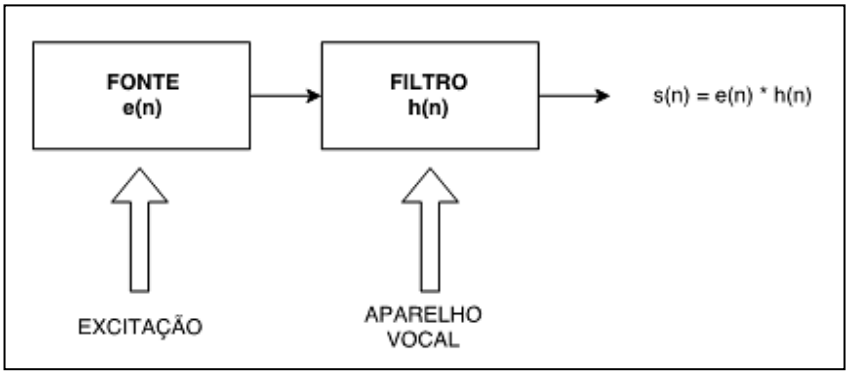

Figura 1 – Diagrama de blocos de geração da voz (Adaptado de Machado, 2013).

A Equação (2.1) representa matematicamente o modelo discreto de geração da voz.

$$s(n) = e(n) * h(n) \tag{2.1}$$

Sendo assim, pela análise do sinal de fala torna-se possível a identificação de anomalias na fonte, referentes a patologias associadas à laringe, e anomalias no trato vocal que estão associadas a patologias de ordem neurológica, ou mesmo fisiológica que limitam o normal funcionamento dos músculos do trato vocal.

2.2. Patologias da Voz

Nesta dissertação será abordada a classificação das seguintes patologias relacionadas com a voz: Disfonia, Laringite Crónica, Paralisia das Cordas Vocais.

2.2.1. Laringite Crónica

As inflamações da Laringe e das áreas próximas são geralmente disfunções denominadas de laringites (Kumar et. al, 2010; Behlau, Azevedo, & Madazio, 2010). A inflamação da mucosa laríngea, localiza nas cordas vocais, precisa persistir por pelo menos 3 semanas, para ser considerada crónica. Em alguns casos pode apresentar muitos anos de evolução. A inflamação, frequentemente causada por irritação prolongada das cordas vocais, provoca uma tumefação ou inchaço da mucosa laríngea e uma abundante produção de secreções. Os sintomas comuns são: dor de garganta, tosse, problemas de deglutição e rouquidão, embora variável provoca principalmente rouquidão devido à inflamação na garganta (Huche & Allali, 2005).

As causas da Laringite podem ser resultado do uso extensivo da voz, de uma manifestação viral, bacteriana ou química. No entanto, é mais comum partir de uma infecção generalizada do trato respiratório superior ou o resultado de exposição maciça ao fumo de tabaco e álcool (Kumar et al., 2010). Trata-se de uma patologia associada à fonte.

2.2.2. Disfonia

Disfonia é uma designação médica para uma desordem vocal, um transtorno na comunicação que dificulta a produção vocal. Existem diversos causadores como por exemplo, uma disfunção, abuso vocal ou uso indevido da voz (Teixeira & Fernandes, 2015). Esta patologia pode ter como sintomas: rouquidão, dor de garganta ou garganta seca, dificuldade em

manter a voz, fadiga vocal, variações na frequência usual, falta de volume e projeção, perda de eficiência vocal e pouca resistência ao falar (Teixeira & Fernandes, 2015).

A Disfonia é comumente encontrada em indivíduos que usam sua voz abundantemente e de maneira incorreta. É uma patologia que pode se manifestar como sintoma secundário ou como principal (Teixeira & Fernandes, 2015). Segundo Teixeira e Fernandes (2015), a Disfonia pode ser classificada como orgânica ou funcional. A orgânica é uma alteração anatômica nas cordas vocais, como nódulos ou tumores benignos, e na funcional não existem alterações anatômicas. Assim, esta patologia pode ser associada a problemas referentes a fonte, no caso de Disfonia Orgânica, ou a fonte e/ou trato vocal no caso de Disfonia Funcional.

2.2.3. Paralisia das Cordas Vocais

A Paralisia das Cordas Vocais, ocorre quando as cordas vocais não abrem e fecham, de forma apropriada, podendo ser unilateral ou bilateral. A paralisia unilateral é uma patologia considerada comum, enquanto a bilateral é rara e pode implicar risco de morte. Como resultado de uma paralisia, as cordas vocais podem permanecer abertas deixando as vias respiratórias e pulmões desprotegidos (Silva et al., 2017).

As principais causas desta patologia podem ser atribuídas a um trauma na cabeça, pescoço ou peito, a manifestações neurológicas (como esclerose múltipla, doença de Parkinson) ou a um acidente vascular cerebral AVC (Silva et al., 2017).

Uma Paralisia das Cordas Vocais ocorre quando os músculos laríngeos não conseguem controlar a tensão e posição das cordas vocais. Como o sistema nervoso é responsável por controlar a ação dos músculos presentes na laringe, se apenas uma prega vocal é afetada pela paralisia, as frequências de vibração das cordas serão divergentes e o indivíduo produzirá um som bitonal, isto é, o paciente perde o poder de amplificação vocal e não consegue falar alto. Se as duas cordas são afetadas pela paralisia, há o risco da glote não se abrir totalmente, provocando problemas respiratórios e ruídos devido a passagem de ar por alguma abertura das cordas vocais (Cordeiro, 2016).

Rouquidão, soprosidade, dificuldades durante a respiração, respiração ruidosa e problemas de deglutição são alguns dos sintomas relacionados a paralisia das cordas vocais. Podem ainda ocorrer alterações na qualidade de voz como a perda de volume ou da frequência fundamental (Silva et al., 2017). Esta é claramente uma patologia associada à fonte.

2.3. Parâmetros do Sinal Acústicos

Serão descritos a seguir alguns dos principais parâmetros acústicos utilizados na caracterização do trato vocal. Os parâmetros *jitter* e *shimmer* que foram extraídos a partir do algoritmo desenvolvido por Teixeira & Gonçalves (2016). Os parâmetros que contêm as características harmônicas da voz, HNR (*Harmonic to Noise Ratio*), NHR (*Noise to Harmonic Ratio*) e Autocorrelação foram extraídos por meio do algoritmo desenvolvido por Fernandes (2018), e disponíveis na base de dados curada (Fernandes et al., 2019). Por fim, os coeficientes mel-cepstrais (MFCC).

2.3.1. Perturbações em Frequência

O conceito do parâmetro *jitter* é estabelecido como uma medida de variação da duração do período glotal, entre ciclos de vibração das cordas vocais (Teixeira & Fernandes, 2015). Os sujeitos com vozes patológicas não conseguem controlar a vibração das cordas vocais e em consequência, os valores das perturbações em frequência medidos, alcançam valores mais elevados. O conceito de *jitter* apresentado é ilustrado na Figura 2. Este parâmetro pode ser medido por diferentes maneiras conhecidas na literatura como *jitter* absoluto, *jitter* relativo, perturbação média relativa (*relative average perturbation-rap*) e o quociente de perturbação (*five-points period perturbation quotient-ppq5*) (Alves, 2016). Nas equações listadas os índices T_i e N correspondem ao tamanho do período glotal i e o número total de períodos glotais, respectivamente.

Jitter* absoluto *(jitta) é a variação absoluta do período glotal entre ciclos consecutivos, expressos pela Equação (2.2).

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i-1}\right| \tag{2.2}$$

***Jitter* relativo (jitt)** é a diferença absoluta média entre os períodos glotais consecutivos divididos pelo período médio e expresso em percentagem, definida na Equação (2.3).

$$jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i-1}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \times 100 \quad (2.3)$$

***Jitter* (rap)** ou perturbação média relativa (*relative average perturbation*) é a diferença absoluta média entre um período e a média desse e os seus dois vizinhos, dividida pelo período médio. É expresso em percentagem na Equação (2.4).

$$rap = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1}T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \times 100 \quad (2.4)$$

***Jitter* (ppq5)** ou quociente de perturbação do período num intervalo de cinco pontos (*fivepoints period perturbation quotient*-ppq5) é a diferença absoluta média entre um período e a média desse e os seus quatro vizinhos dividida pelo período médio. É expresso em percentagem na Equação (2.5).

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|T_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2}T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \times 100 \quad (2.5)$$

2.3.2. Pertubações em Amplitude

As variações de magnitude por toda a extensão dos períodos glotais, em cada ciclo fonatório, dá origem ao parâmetro acústico denominado de *shimmer*, ilustrado na Figura 2. Algumas variações na magnitude glotal de pacientes patológicos são derivadas de lesões ou redução na resistência glotal devido a respiração e a emissão de ruído (Teixeira & Fernandes, 2015).

Este parâmetro pode ser medido de diversas maneiras como *shimmer* absoluto em dB, shimmer relativo em percentagem, quociente de perturbação da amplitude em três ciclos *(three point amplitude perturbation quotient-apq3)* e quociente de perturbação da amplitude em cinco pontos *(five point amplitude perturbation quotient-apq5)* (Teixeira & Gonçalves, 2014). Nas

equações listadas abaixo, os índices A_i e N correspondem à amplitude e ao número total de períodos glotais, respectivamente.

***Shimmer* absoluto (ShdB)** é a variação da amplitude pico a pico em decibel (dB), sendo definido pela Equação (2.6).

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\times\log\left(\frac{A_{i+1}}{A_i}\right)\right| \tag{2.6}$$

***Shimmer* relativo (Shim)** é diferença média absoluta das amplitudes de dois períodos consecutivos, normalizada pela amplitude média percentual expresso na Equação (2.7).

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_{i+1}-A_i\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i}\times 100 \tag{2.7}$$

***Shimmer* (apq3)** é o quociente de perturbação da amplitude em três pontos (*three point amplitude perturbation quotient-apq3*) que equivale a diferença absoluta média entre a amplitude de um período e a média das amplitudes dos seus vizinhos, dividida pela amplitude média. É expresso em percentagem na Equação (2.8).

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i-\left(\frac{1}{3}\sum_{n=i-1}^{i+1}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i}\times 100 \tag{2.8}$$

***Shimmer* (apq5)** é o quociente de perturbação da amplitude em cinco pontos (*five point amplitude perturbation quotient*-apq5) que representa a diferença absoluta média entre a amplitude de um período e a média das amplitudes dos seus quatro vizinhos, dividida pela amplitude média. É também expresso em percentagem na Equação (2.9).

$$apq5 = \frac{\frac{1}{N-1}\sum_{i=3}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \quad (2.9)$$

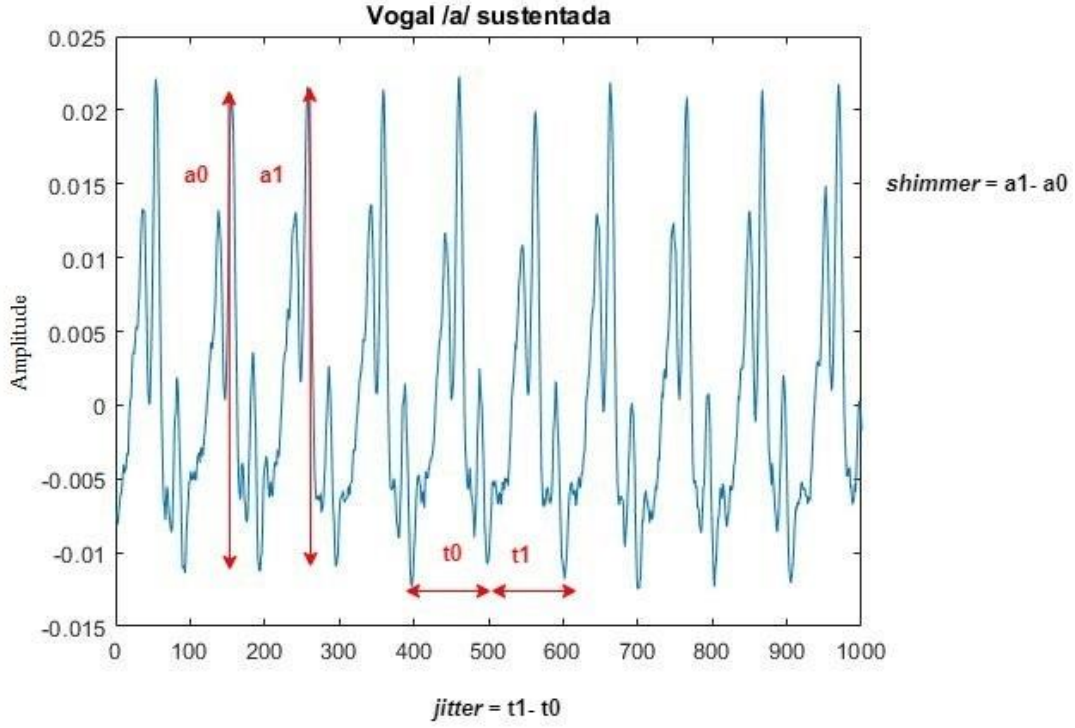

Figura 2 – Representação do jitter e shimmer em um sinal de fala.

2.3.3. Autocorrelação

A autocorrelação descreve a distribuição da magnitude espectral de um sinal no domínio temporal. É a transformada inversa de Fourier de um espectro de energia distribuída (Teorema de Wiener–Khinchin) (Ferreira, 2012). É uma ferramenta matemática que permite evidenciar padrões repetidos, como a presença de um sinal periódico camuflado por ruído. Quanto maior o valor da autocorrelação, maior é a repetição de eventos no sinal. (Boersma, 1993). Trabalhos indicam que, quando comparadas com vozes saudáveis, vozes patológicas apresentam menores valores de autocorrelação (Fernandes, 2018).

Dado um sinal de fala $x(t)$ é selecionada uma seção com duração T que é centrada em t_{mid}. A partir desta, subtrai-se a média μ_x e multiplica-se por uma janela $w(t)$ (Boersma, 1993). O sinal resultante $a(t)$ do sinal dada na Equação (2.10).

$$a(\mathrm{t}) = \left(x\left(t_{mid} - \frac{1}{2}T + t \right) - \mu_x \right) w(\mathrm{t}) \tag{2.10}$$

Sugere-se o uso da função de janela $w(\mathrm{t})$ sinusoidal ou janela de Hanning expressa pela Equação (2.11) (Boersma, 1993).

$$w(\mathrm{t}) = \frac{1}{2} - \frac{1}{2}\cos\frac{2\pi t}{T} \tag{2.11}$$

Na etapa seguinte é calculada a autocorrelação normalizada da parte do sinal selecionada $r_a(\tau)$ que é expressa pela Equação (2.12). Assume-se que τ é uma variável de atraso (Boersma, 1993).

$$r_a(\tau) = \mathrm{r}_a(-\tau) = \frac{\int_0^{T-\tau} \mathrm{a(t)\,a(t+\tau)\,dt}}{\int_0^{T} \mathrm{a^2(t)\,dt}} \tag{2.12}$$

O cálculo da autocorrelação normalizada da função de janela é expresso por meio da Equação (2.13). A janela empregada é a janela de Hanning (Boersma, 1993).

$$r_w(\tau) = \left(1 - \frac{|\tau|}{T}\right)\left(\frac{2}{3} + \frac{1}{3}\cos\frac{2\pi\tau}{T}\right) + \frac{1}{2\pi}\sin\frac{2\pi|\tau|}{T} \tag{2.13}$$

Ao dividir $r_a(\tau)$ pela $r_w(\tau)$, pode-se obter uma estimativa da autocorrelação $r_x(\tau)$ do segmento de sinal original (Boersma, 1993). A Equação (2.14) é a representação da $r_x(\tau)$. Através da Figura 3 pode-se ilustrar a aplicação da função de autocorrelação.

$$r_x(\tau) = \frac{r_a(\mathrm{t})}{r_w(\mathrm{t})} \tag{2.14}$$

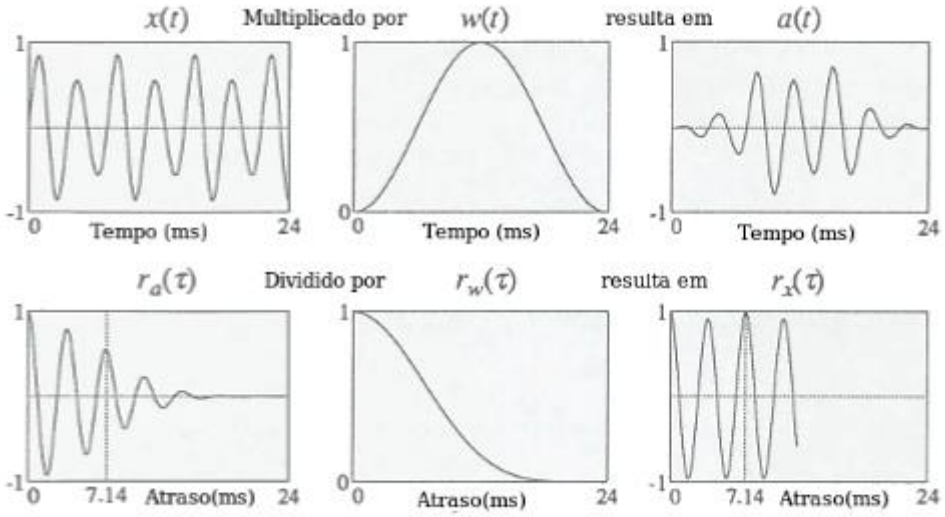

Figura 3 – Representação da autocorrelação do sinal (Adaptado de Boersma, 1993).

2.3.4. HNR

O *Harmonic to Noise Ratio* (HNR) é a razão entre as componentes harmônicas e de ruído em um sinal de fala sustentada e vozeada. Consiste na relação entre a componente periódica resultado da vibração das cordas vocais e a componente não periódica resultado do ruído glótico (Guimarães, 2007). A quantificação de HNR permite indicar integralmente se o sinal tem um comportamento periódico (Alves, 2016). O desempenho do processo de fonação pode ser avaliado pela HNR, onde um valor elevado significa uma voz saudável e está associado a impressão de voz sonora e harmônica, enquanto que um valor pequeno indica uma voz astênica ou disfônica (Lopes, 2008).

A autocorrelação $r_x(\tau)$ em função do atraso τ, é definida na Equação (2.15), para um sinal de tempo x(t) (Boersma, 1993). A autocorrelação é usada para obter a separação em componente harmônica H e de ruído N.

$$r_x(\tau) = \int x(\mathrm{t})\, \mathrm{x}(\mathrm{t}+\tau)\, \mathrm{dt} \qquad (2.15)$$

A função $r_x(\tau)$ tem seu máximo em $\tau = 0$. Se existirem máximos globais fora de 0 e os outros máximos da função estiverem situados em nT_0, como $r_x(\mathrm{nT}_0) = \mathrm{r}_x(0)$, indica periodicidade do sinal (Boersma, 1993). Ainda pode haver máximos locais mesmo que não existam máximos globais fora de 0. Se o maior máximo local ocorre em $\tau_{\max}$ e a sua altura

$r_x(\tau_{\max})$ for suficiente, o sinal terá uma componente periódica e a sua força harmônica R (*e*ntre 0 e 1) é igual ao máximo local $r'_x(\tau)$ (Boersma, 1993), o qual é definido na Equação (2.16).

$$r'_x(\tau) = \frac{r_x(\tau)}{r_x(0)} \tag{2.16}$$

A Equação (2.17) representa a autocorrelação total, a qual é equivalente ao somatório das componentes harmônicas e componentes de ruído (Boersma, 1993).

$$r_x(0) = \mathrm{r}_H(0) + \mathrm{r}_N(0) \tag{2.17}$$

Caso o ruído seja branco, ou seja, tenha densidade espectral plana (não se correlacionará com ele mesmo). O máximo local é obtido em $\tau_{\max} = T_0$ com a altura $r_x(\tau_{\max}) = \mathrm{r}_H(\mathrm{T}_0) = \mathrm{r}_H(0)$ (Boersma, 1993). Pode-se medir o HNR, em dB, através da Equação (2.18).

$$HNR = 10 \times \log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \tag{2.18}$$

Por fim, a função autocorrelação de um sinal de fala sustentada possui máximos locais para valores múltiplos de τ e T_0 . Logo, faz-se necessário calcular a função autocorrelação do sinal e identificar o primeiro pico (correspondente a componente harmônica do sinal e considerar a energia restante como de ruído), para determinar o HNR. A componente de ruído $N = 1 - H$ equivale a diferença entre 1 e parte harmônica H (Lopes, 2008).

2.3.5. NHR

O *Noise-to-Harmonic-Ratio* (NHR) é definido como a relação entre as componentes aperiódicas e as componentes periódicas, presentes num sinal vocal. As componentes aperiódicas são resultantes do ruído glótico. Enquanto que as componentes harmônicas têm origem da vibração das pregas vocais.

Por meio da Equação (2.19) apresentam-se os cálculos do NHR (Fernandes, 2018). No qual, N equivale a componente do ruído, H a componente harmônica e A corresponde ao valor

da autocorrelação (Fernandes, 2018). Sendo assim, o valor da componente de ruído, HNR é expresso em função da função de autocorrelação por meio da Equação (2.20).

$$NHR = \frac{N}{H} = \frac{H - A}{H} = \frac{1 - A}{1} \tag{2.19}$$

$$HNR = 1 - A \tag{2.20}$$

2.3.6. Coeficientes Mel-Cepstrais

Os Coeficientes Cepstrais na Frequência Mel, derivado do inglês, *Mel Frequency Cepstral Coefficients* (MFCC), são características acústicas de curto termo que se fundamentam no espectro (Oppenheim, Schafer, & Buck, 1999). Os Coeficientes Mel-Cepstrais são altamente eficazes no reconhecimento de voz de falantes (Xu et al., 2004; Ittichaichareon, Suksri, & Yingthawornsuk, 2012).

Foram desenvolvidos os MFCC's de modo a respeitar as percepções de frequências interpretadas pelo aparelho auditivo humano, isto é, respeitar uma relação de escala não-linear (Alves, 2016). Surgem com a introdução de informação perceptiva, através da filtragem do espectro do sinal com um banco de filtros de escala Mel (Cordeiro, 2016). Os coeficientes mel-cepstrais originam-se a partir dos coeficientes cepstrais (Davis & Mermelstein, 1980; Cordeiro, 2016). O processo de obtenção dos parâmetros acústicos MFCC são esquematizados na Figura *4*.

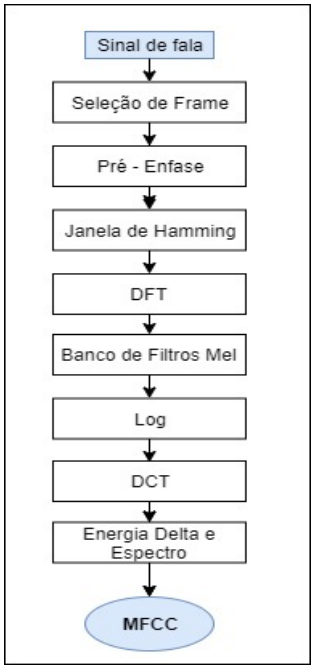

Figura 4 – Processo de criação de um coeficiente mel-cepstral (Adaptado de Logan, 2000).

Como primeira etapa do processo tem-se a pré-ênfase, que aumenta a energia do sinal $x(n)$ nas frequências mais altas. Esta etapa é expressa por meio da Equação (2.21), a qual, $a = 0.95$ (Muda et al., 2010).

$$y(\mathrm{n}) = \mathrm{x}(\mathrm{n}) - a\mathrm{x}(\mathrm{n}-1) \tag{2.21}$$

Na segunda etapa tem-se a seleção de *frames*, onde o sinal é dividido em N *frames*, com durações a partir de 20 ms até 40 ms. A terceira etapa consiste no janelamento (multiplicação da função janela, *frame* a *frame*). Para isto é utilizada a janela de Hamming $w(n)$, expressa por meio na Equação (2.22) (Oppenheim et al., 1999).

$$w(\mathrm{n}) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \ge N-1 \tag{2.22}$$

A quarta etapa utiliza a Transformada Discreta de Fourier (DFT) para transformar N amostras de cada *frame*, do domínio dos tempos para o domínio das frequências, por meio da Equação (2.23). No qual, $X(\mathrm{k})$ são coeficientes espectrais, $x(\mathrm{n})$ é o sinal de áudio e N é o comprimento da DFT, a que corresponde o comprimento de $x(\mathrm{n})$ e $X(\mathrm{k})$ (Oppenheim et al., 1999).

$$X(\mathrm{k}) = \sum_{n-0}^{N-1} x(\mathrm{n})\mathrm{e}^{-\frac{i2\pi nk}{N}} \tag{2.23}$$

Na quinta etapa utiliza-se um filtro de acordo com a escala Mel. A frequência pode ser convertida da escala Hz para a escala Mel através da aplicação da Equação (2.24) (Alves, 2016; Oppenheim et al., 1999).

$$F(\mathrm{Mel}) = 2595 \times \log_{10}\left[\frac{1+f}{700}\right] \tag{2.24}$$

Na sexta etapa obtém-se os *MFCC's* ao realizar a conversão do espectro na base log Mel para o domínio temporal, por meio da Transformada Discreta do Cosseno (DCT) (Oppenheim et al., 1999).

Na etapa final é calculada a energia do sinal $x(n)$ pela equação (2.25). Onde são medidas variações de amplitude do sinal de fala (Alves, 2016; Oppenheim et al., 1999).

$$Energia = \sum x^2(\mathrm{n}) \tag{2.25}$$

2.4.Base de Dados

O banco de vozes chamado *Saarbrücken Voice Database* (SVDT) é constituído por uma coleção de gravações de vozes formada por mais de 2000 pessoas com nacionalidade alemã, algumas diagnosticadas com alguma patologia e outras saudáveis (Barry & Pützer, 2018). É disponibilizado pelo Instituto de Fonética da Universidade de Saarland de modo gratuito e *online*, sendo facilmente encontrado por toda a comunidade (Barry & Pützer, 2018).

Cada indivíduo tem um arquivo de gravação individual que contém: as vogais sustentadas /a/, /i/ e /u/ nos tons baixo, normal e alto; as vogais /a/, /i/ e /u/ com a frequência fundamental crescente; e a frase em alemão *"Guten Morgen, wie geht es Ihnen?"* ("Bom dia, como estás?") (Barry & Pützer, 2018). O tamanho do arquivo de som situa-se entre 1 e 3 segundos e têm uma resolução de 16 bits e uma frequência de amostragem de 50 kHz (Teixeira & Gonçalves, 2014).

A partir da base de dados curada parâmetros de interesse foram extraídos da SVDT pelos algoritmos desenvolvidos por (Fernandes et al., 2019) e (Teixeira & Gonçalves, 2014). A Tabela 1 apresenta uma caracterização da base utilizada, composta por sujeitos patológicos com 19 doenças diferentes e também sujeitos saudáveis, num total de 901 indivíduos.

Tabela 1 – Caracterização da base de dados (Adaptado de Fernandes et al., 2019).

Grupo de Testes	Tamanho das amostras		Média de Idades	Desvio Padrão de Idades
	Feminino	Masculino		
Controle	123	71	38,06	14,36
Disfonia	40	29	47,38	16,27
Laringite Crónica	16	25	49,69	13,47
Paralisia das Cordas Vocais	102	67	57,75	13,77
Cisto	2	1	47,5	15,56
Pólipos de Cordas Vocais	10	17	52,28	13,41
Carcinoma de Cordas Vocais	1	18	57,00	6,60
Tumor Laríngeo	1	3	53,50	8,17
Granuloma	1	1	44,50	4,50
Granuloma Intubação	-	3	53,00	11,22
Tumor Hipofaríngeo	-	5	59,50	9,29
Fibroma	1	-	46,00	0
Displasia Laríngea	-	1	69,00	0
Edema de Reinke	29	5	56,10	11,37
Disfonia Funcional	51	24	47,12	14,54
Disfonia Hipofuncional	4	8	41,63	15,07
Disfonia Hiperfuncional	95	32	42,32	13,62
Disfonia Hipotônica	-	2	49,50	12,50
Disfonia Psicogênica	38	13	51,40	9,40
Disfonia Espasmódica	40	22	57,15	15,75
Total	554	347	-	-

Capítulo 3: Aprendizagem Computacional

Nesta seção serão apresentadas as descrições das principais arquiteturas de redes e os critérios de avaliação utilizados para determinar a performance do sistema inteligente.

3.1.Redes Neuronais Artificiais

As Redes Neuronais Artificiais (RNA) são determinadas como coleções de pequenas unidades de processamento individualmente interconectadas, no qual a informação é transmitida entre unidades de interconexão. A RNA tem como característica principal a capacidade de generalização, a partir da sua estrutura e a habilidade de aprender, permitindo a solução de problemas complexos (Haykin, 1999).

Um modelo computacional de neurónio artificial foi estabelecido, fundamentado nas atividades de responsabilidade do neurónio biológico (Zhang & Zhang, 1999). Um grupo de unidades chamadas "neurónios artificias" estruturam as redes neuronais artificiais, interconectados por meio de pesos sinápticos (Haykin, 1999).

No modelo matemático do neurónio ocorre a união aditiva representada pela soma das entradas de sinais e as sinápses. Esta soma ponderada pode ou não receber a ação das bias, levando a um crescimento ou diminuição da entrada da função de ativação. Posteriormente, o sinal da saída é limitado, devido à restrição da amplitude imposta pela função de ativação (Ribeiro, 2007).

A Figura 5, é um esquema do funcionamento de uma rede neuronal artificial, contendo os sinais de entrada, os pesos, função de ativação e sinais de saída.

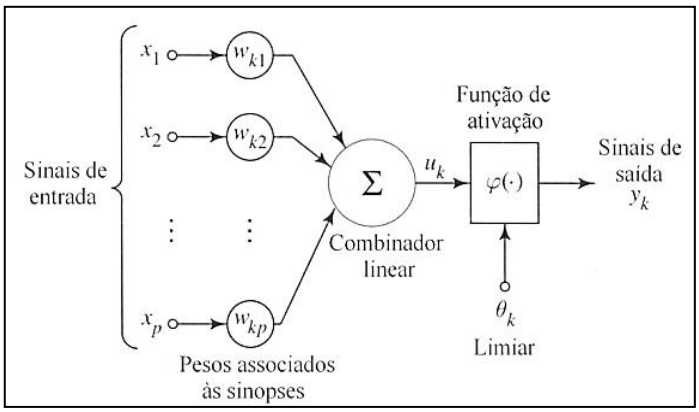

Figura 5 – Modelo não linear de um neurônio artificial (Adaptado de Haykin, 2001).

Matematicamente pode-se representar o valor de saída do neurónio pelo y_k na Equação (3.1). Na qual: $x1, x2, ...xn$ são sinais de entrada com n terminais; $W_{k1}, W_{k2}...W_{kp}$ são os pesos sinápticos do neurónio k ; b_k , é a representação do bias; φ , representa a função de ativação (Moraes, Valiati, & Neto, 2013; Haykin, 2001).

$$y_k = \varphi\left(\sum_{j=1}^{m} w_{kj} x_j + b_k\right) \tag{3.1}$$

Em geral as redes neuronais podem ser classificadas de acordo com sua arquitetura. O modo como os neurónios se distribuem e dispõem estruturalmente está muito relacionado com os seus algoritmos de treinamento. Possuem três principais tipologias: Perceptrão, *Multilayer Perceptron* e as rede recorrentes (Haykin, 2001).

3.1.1. Perceptrão

O Perceptrão é uma rede neuronal que se propaga progressivamente e resolve problemas separáveis de forma linear. É adequado para solucionar problemas simples de classificação de padrões. Pode ser organizada por camada e como as conexões são sempre unidirecionais, não existem ciclos (Cortez & Neves, 2000).

Contém a camada de entrada de nós que se projeta sobre a camada de saída (nós computacionais). Os neurónios estão organizados na forma de camadas, a qual possui uma camada única de saída, não são considerados os nós da camada de entrada (Haykin, 2001). Pode ser observado na Figura 6, um exemplo de um perceptrão, neste caso de 4 nós, tanto na camada de entrada (quadrados azuis) como na de saída (círculos verdes).

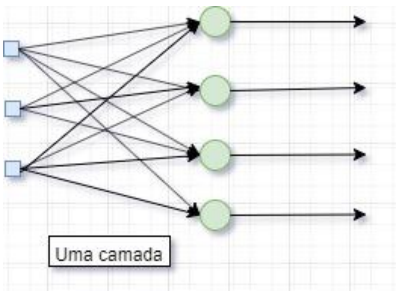

Figura 6 – Rede alimentação progressiva com uma única camada de neurónios (Adaptado de Haykin, 2001).

3.1.2. Multilayer Perceptron

O *Multilayer Percepton* (MLP) é uma Rede Neural Artificial de arquitetura feed-forward (os neurónios estão conectados em apenas uma direção), formado por uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Usam funções de ativação não-lineares que permitem resolver problemas não-lineares (Bishop, 1995).

A função da camada escondida é de intervir de forma útil entre a entrada e a saída da rede. Ao acrescentar camadas intermediárias aumenta-se a capacidade da rede em modelar as funções de maior complexidade, como um número elevado de nós na camada oculta. Uma consequência do aumento da complexidade do modelo é o aumento exponencial do tempo de aprendizagem (Cortez & Neves, 2000). A Figura 10 ilustra a arquitetura de um MLP, com duas camadas além da entrada.

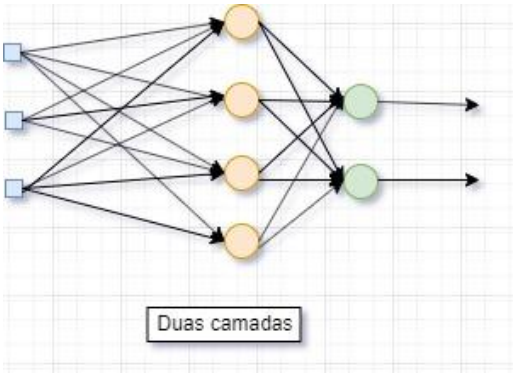

Figura 7 – Rede de alimentação progressiva totalmente conectada, com uma camada oculta e uma camada de saída (Adaptado de Haykin, 2001).

3.1.3. Redes Recorrentes

A Rede Artificial Recorrente (RNN) apresenta realimentação e também auto alimentação, de modo que a saída de um neurónio fornece parâmetros de entrada para a camada anterior. Permitem gerar e reconhecer padrões temporais e padrões espaciais (Haykin, 1999).

A Figura 8 ilustra conexões de realimentação, em uma rede recorrente, onde estas se originam dos neurónios ocultos bem como dos neurónios de saída. A presença de laços de realimentação afeta a capacidade de aprendizagem da rede e o seu desempenho. Além disso, os laços de realimentação envolvem o uso de ramos particulares compostos de elementos de atraso unitário, o que resulta em comportamento dinâmico não-linear, admitindo-se que a rede neural contenha unidades não-lineares (Haykin, 2001).

O algoritmo utilizado é o *Backpropagation Through Time*. Na fase *forward* a propagação de valores entre as camadas também utiliza pesos, entretanto agora é importante incluir os pesos que fazem ciclos de propagação para o neurónio analisado. As Equações (3.2) e (3.3) definem a saída do neurónio (Bodén, 2001).

$$y_i(\mathrm{t}) = \varphi(v_i) \tag{3.2}$$

$$v_i = \sum_i w_i x_i(\mathrm{t}) + \sum_j u_j y_j(\mathrm{t}-1) + \theta \tag{3.3}$$

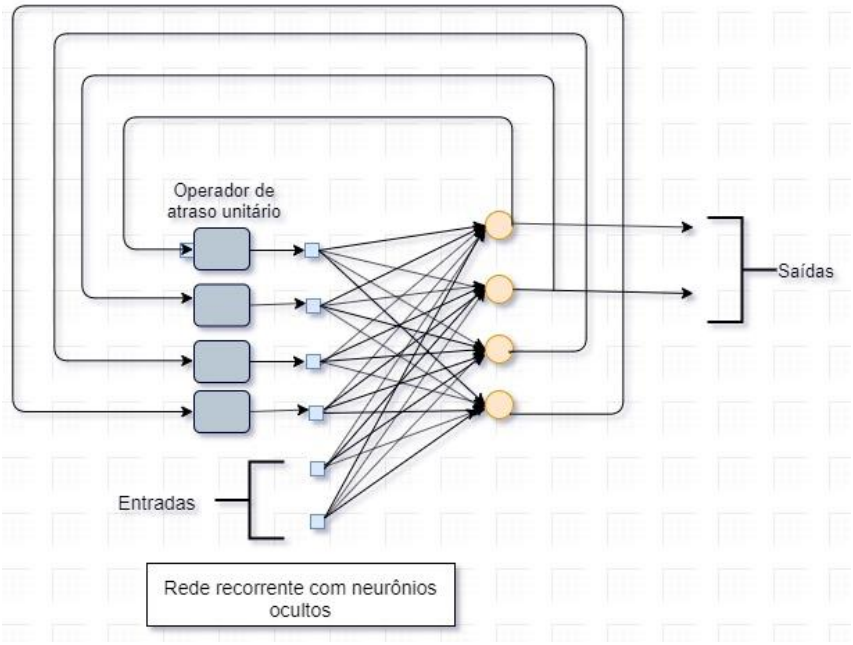

Figura 8 – Rede Recorrente sem laços de auto alimentação e com neurónios ocultos (Adaptado de Haykin, 2001).

3.1.4. Long Short-Term Memory

Redes de Memória Longa de Curto Prazo são um tipo especial de RNN composta por unidades LSTM, capaz de aprender dependências de longo prazo. A LSTM foi introduzida por Hochreiter & Schmidhuber (1997). Essas unidades contêm células de memórias com auto conexões capazes de armazenar o estado temporal da rede, além de unidades especiais chamadas portas que são responsáveis por controlar o fluxo de informações (Graves, 2002).

Também existem três unidades, conhecidas como porta de entrada, de saída e o de esquecimento com as respectivas funções de escrita, leitura e *reset* da célula. As entradas das

portas estão associadas com o vetor de entrada atual, saída do neurônio (bloco) anterior e o seu respectivo *bias* (viés) (Hochreiter & Schmidhuber, 1997).

A porta de entrada atribui informação na célula. A Equação (3.4) determina o valor a adicionar, onde xt é o valor de entrada no tempo t, h_{t-1} é a saída do neurónio oculto anterior, c_{t-1} é o valor anterior da ativação da célula e ϕ é a função de ativação. Normalmente é aplicada a esta porta a função Sigmoide Logística (Yu & Deng, 2015).

$$pe_t = \phi(\mathrm{w}^{xpe}\mathrm{xt} + \mathrm{w}^{hpe}\mathrm{h}_{t-1} + \mathrm{w}^{cpe}\mathrm{c}_{t-1} + \theta^{pe}) \tag{3.4}$$

A porta de esquecimento calcula a quantidade de informação lembrada ou esquecida na célula. Frequentemente é aplicada a função Sigmóide Logística (Yu & Deng, 2015). É expressa pela Equação (3.5).

$$pes_t = \phi(\mathrm{w}^{xpes}\mathrm{xt} + \mathrm{w}^{hpes}\mathrm{h}_{t-1} + \mathrm{w}^{cpes}\mathrm{c}_{t-1} + \theta^{pes}) \tag{3.5}$$

O cálculo referente à ativação da célula de memória é expresso pela Equação (3.6). Normalmente tem como função a Tangente Hiperbólica (Graves, 2002).

$$c_t = pes_t * c_{t-1} + pe_t * \phi(\mathrm{w}^{xc}\mathrm{xt} + \mathrm{w}^{hc}\mathrm{h}_{t-1} + \theta^{c}) \tag{3.6}$$

A porta de saída lê a informação da célula e envia de volta à RNN. De acordo com o valor dado é armazenado o quanto de saída será desejado. Para a entrada na porta utiliza-se a Equação (3.7), e para efetivar a saída é utilizado a Equação (3.8). A função de ativação empregada é a Sigmóide Logística (Graves, 2002).

$$s_t = \phi(\mathrm{w}^{xs}\mathrm{xt} + \mathrm{w}^{hs}\mathrm{h}_{t-1} + \mathrm{w}^{cs}\mathrm{c}_t + \theta^{s}) \tag{3.7}$$

$$h_t = s_t * \phi(\mathrm{c}_t) \tag{3.8}$$

3.2. Avaliação do Desempenho do Modelo

A avaliação da performance de um modelo de aprendizagem de redes neuronais fundamenta-se na matriz confusão (*confusion matrix*), a qual é composta por linhas e colunas, de acordo com o sistema de decisão. As linhas ponderam o número de observações de cada um

dos valores das classes no conjunto de teste enquanto as colunas, ponderam o número de observações previstas para cada classe (Apolónia, 2018).

Segundo Souza (2009), as regiões da matriz são ilustradas na Tabela 2 São caracterizadas como: VP, valores positivos que o sistema julgou positivos, sendo verdadeiros positivos; FN, valores positivos que o sistema julgou negativos, sendo falsos negativos; VN, valores negativos que o sistema julgou como negativos, sendo verdadeiros negativos; FP, valores negativos que o sistema julgou positivos, sendo falsos positivos.

Assim, num caso de classificação binária (patológico = 1 e saudável = 0), VP e VN denotam o número de sujeitos saudáveis e patológicos que são classificados corretamente. Enquanto isso, FP e FN denotam o número de sujeitos saudáveis e patológicos classificados incorretamente (Hossin & Sulaiman, 2015). Através da utilização da matriz de confusão é que as medidas de avaliação serão calculadas, por exemplo: acurácia (exatidão), precisão, sensibilidade (*recall*), especificidade e medida F (Souza, 2009).

Tabela 2 – Diagrama de uma matriz confusão (Adaptado de Teixeira, Fernandes, & Alves, 2017).

		Valor Previsto (predito pelo teste)	
		Patológico (1)	Saudável (0)
Valor Verdadeiro (confirmado por análise)	Patológico (1)	Verdadeiro Positivo (**VP**)	Falso Negativo (**FN**)
	Saudável (0)	Falso Positivo (**FP**)	Verdadeiro Negativo (**VN**)

Acurácia/Exatidão (Acc) – a Equação (3.9) representa a razão de predições corretas pela soma de todas as predições. Esta medida apresenta grande sensibilidade aos desbalanceamentos do conjunto de dados (Hossin & Sulaiman, 2015).

$$Acc = \frac{VP + VN}{(\mathrm{VP} + \mathrm{VN} + \mathrm{FP} + \mathrm{FN})} \tag{3.9}$$

Precisão (P) – a Equação (3.10) simboliza a razão dos verdadeiros positivos pela soma de todos os positivos. Representa a porcentagem de sujeitos patológicos classificados

corretamente como patológicos (sem considerar os casos negativos) (Hossin & Sulaiman, 2015).

$$P = \frac{VP}{(\mathrm{VP} + FP)} \tag{3.10}$$

Sensibilidade (Sn) – a Equação (3.11) estabelece a razão de verdadeiros positivos. Isto é, denota a capacidade do modelo em predizer corretamente a condição que têm (patológicos diagnosticados corretamente) (Hossin & Sulaiman, 2015)(Souza, 2009).

$$Sn = \frac{VP}{(\mathrm{VP} + \mathrm{FN})} \tag{3.11}$$

Especificidade (S) – a Equação (3.12) representa a razão de verdadeiros negativos. Isto é, denota a capacidade do modelo em predizer corretamente a condição de saudáveis (saudáveis diagnosticados corretamente) (Hossin & Sulaiman, 2015).

$$Sp = \frac{VN}{(\mathrm{VN} + \mathrm{FP})} \tag{3.12}$$

Medida F (MF) – a Equação (3.13) corresponde à média aritmética da sensibilidade e especificidade. Esta medida, em conjuntos desbalanceados, possibilita uma avaliação mais confiável dos modelos (Hossin & Sulaiman, 2015).

$$MF = 2 * \frac{(\mathrm{P} * \mathrm{S})}{P + S} \tag{3.13}$$

Capítulo 4: Identificação e Tratamento de *Outliers*

A mineração de dados realiza a extração de informações em conjuntos de dados por meio de técnicas estatísticas e computacionais (Campos, 2015). É uma aplicação que surgiu para resolver algumas das dificuldades enfrentadas por técnicas de análise tradicionais. As fontes das dificuldades nas análises são: a escalabilidade do conjunto de dados, com até *petabytes*; a alta dimensionalidade dos dados, com centenas ou milhares de atributos; a complexidade e a distribuição (Tan, Steinbach, & Kumar, 2006).

Dentro da mineração de dados, a detecção de anomalias é a busca por dados cujas características se diferem das demais observações (Berton, 2011). O termo *outlier* pode ser definido como uma observação com comportamento inconsistente em relação ao restante das observações a qual é derivada (Hawkins, 1980). Uma segunda definição diz que um *outlier* é uma observação que parece desviar-se acentuadamente dos outros membros da amostra em que ela ocorre (Grubbs, 1969).

As presenças de observações atípicas influenciam nos cálculos da média, desvio-padrão, histograma. Como resultado, provocam a distorção de conclusões e generalizações sobre o conjunto de dados analisado (Lima et al, 2017). Portanto, a identificação da qualidade das observações é aferida ao estudar o modo como impactam os resultados (Muñoz-Garcia, Moreno-Rebollo, & Pascual-Acosta, 1990).

O aparecimento de anomalias em base de dados é causado maioritariamente por erros humanos, de instrumentos, desvios em populações, comportamento fraudulento, mudanças ou falhas no comportamento de sistemas (Barnett & Lewis, 1994). A identificação de anomalias é implementada em sistemas com diferentes finalidades, como a detecção de invasão em redes de computadores, verificação de fraude bancária, detecção de doenças, em estatísticas desportivas e detectando erros de medição (Campos, 2015).

A Figura 9, apresenta uma série temporal, os pontos distantes do limite são *outliers*. Estes estão indicados pelos pontos destacados com círculo (Mata, 2017). Uma forma simplista de solucionar o problema seria eliminar as observações atípicas. No entanto, deve-se considerar que eles contêm informações relevantes, que caracterizam os dados e permitem conhecer a população como um todo. Logo, a técnica recomendada é o tratamento ou correção dos mesmos (Figueira, 1998).

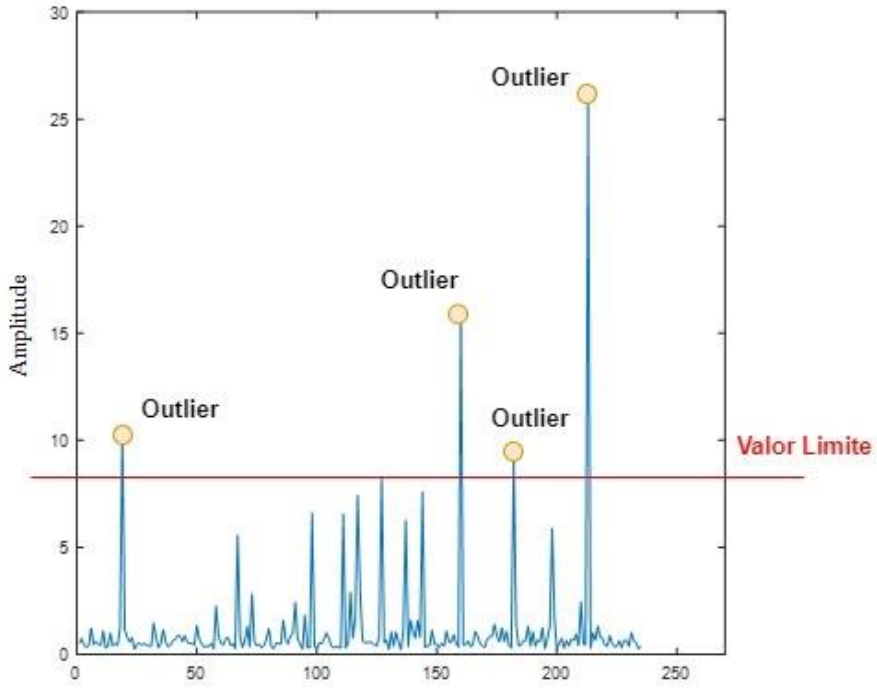

Figura 9 – Representação de dados anômalos.

4.1. Técnicas de Identificação de Anomalias

Os testes utilizados neste trabalho são o do Desvio-Padrão e do Diagrama da Caixa. Nestes métodos, a estimação da probabilidade de ter *outilers* é facilitada se os dados tiverem uma distribuição normal. A função de densidade de probabilidade de uma distribuição normal é dada pela Equação (4.1). Considere, a variável contínua (x), o desvio-padrão (σ) e a média (μ) (Triola, 2017).

$$f(\mathrm{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{4.1}$$

A Figura 10 ilustra a função densidade de probabilidade de uma distribuição normal. Portanto, cerca de 68%, 95% e 99,7% dos dados que estão localizados dentro de $\pm 1\sigma$, $\pm 2\sigma$ e $\pm 3\sigma$ de distância da média, respectivamente (Seo, 2006).

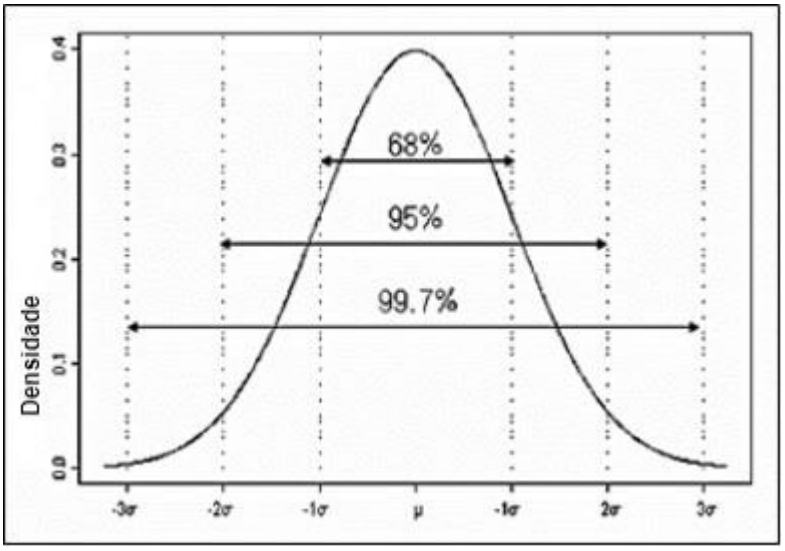

Figura 10 – Ilustração de uma função de densidade de probabilidade (Adaptado de Seo, 2006).

Uma regra típica em uma distribuição de dados normal que considera apenas uma dimensão é sempre identificar como anomalia as observações que desviam mais de três desvios-padrão de distância da média, uma vez que uma instância localizada nessa região tem alta probabilidade de ser um *outlier* (Howell, 2010).

As operações de normalização de dados são úteis na identificação de *outliers* quando se utiliza um método que é bastante eficaz em uma distribuição normal. É importante notar que os *outliers* esperados antes e depois da transformação são diferentes (Seo, 2006).

4.2. Testes de Identificação

4.2.1. Desvio Padrão

O desvio-padrão de um conjunto de valores amostrais é uma medida da variação dos valores em relação à média. O desvio-padrão amostral é calculado de acordo com a Equação (4.2) onde x é o valor da amostra, $\bar{x}$ é a média das amostras e n é o número total de amostras e sempre expresso com a mesma unidade dos valores originais (Triola, 2017).

$$s = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{n-1}} \tag{4.2}$$

O método do Desvio- Padrão (DP) é uma técnica largamente empregada no tratamento de anomalias devido à sua simplicidade. Sua utilidade é limitada a dados com distribuição normal e razoavelmente simétricos. Podem ser consideradas como *outliers* nos dados as observações que estão fora do intervalo de dois ou três desvios-padrão, acima e abaixo da média das observações (Seo, 2006). Em geral, é definido como na Equação (4.3).

$$\bar{x} \pm 2s \text{ ou } \bar{x} \pm 3s \quad (4.3)$$

4.2.2. Diagrama de Caixa

Um *Box Plot* (BP) também conhecido como Diagrama de Caixa é um método muito comum e de fácil utilização. É uma ferramenta que possibilita uma visualização simplificada da semelhança entre os parâmetros e grupos de vozes (Gonçalves, 2015).

O Diagrama de Caixa foi desenvolvido por Tukey e é aplicável a dados simétricos ou com leve assimetria, como aqueles que têm uma distribuição normal (Lima, Maroldi, Silva, Hayashi, & Hayashi, 2017). Este método é menos sensível a valores *outliers*, pois não depende da média das amostras ou do desvio-padrão e não faz suposições distributivas (Tukey, 1970).

A partir da Figura 11 é ilustrado o diagrama de caixa. O primeiro quartil (Q1) corresponde a 25% das menores medidas e o terceiro quartil (Q3) corresponde a 75% das menores medidas. A altura da caixa corresponde à amplitude do intervalo interquartil (IQR = Q3 - Q1) e o segundo quartil (Q2) é equivalente à mediana. Portanto, a caixa representa 50% de todos os valores observados, concentrados na tendência central dos valores (McGill, Tukey, & Larsen, 1978).

Valores que excedem os limites do *Box Plot*, calculados a partir dos quartis, são considerados *outliers*. Os limites inferiores e inferiores são localizados a uma distância de 1.5 IQR abaixo de Q1 e acima de Q3 [Q1-1.5IQR, Q3 + 1.5IQR] (Tukey, 1970; Gonçalves, 2015).

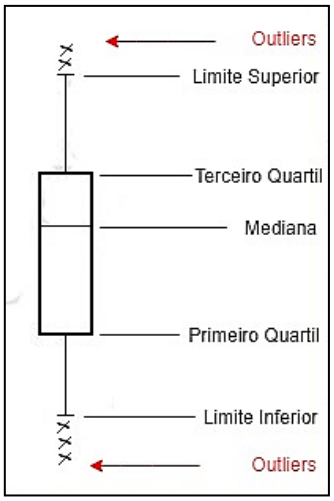

Figura 11- Diagrama de Caixa.

4.3. Métodos de Normalização

Algumas ferramentas de modelação têm seu desempenho beneficiado pela normalização, por exemplo, Rede Neural, KNN, *clustering*, porque tais operações de normalização são destinadas a minimizar problemas como redundância de dados e resultados distorcidos na presença de anomalias (Campos, 2015).

Algumas transformações foram feitas no conjunto de dados. Assim, o que se busca é fazer uma escala para estabilizar a variância, diminuir a assimetria e aproximar a variável da distribuição normal (Pino, 2014). Alguns dos métodos mais utilizados de normalização de dados são a transformação Z-score e a potência.

A transformação Z-score para uma variável aleatória x, com média μ e desvio padrão σ, é dada pela Equação (4.4).

$$z = \frac{(\mathrm{x} - \mu)}{\sigma} \tag{4.4}$$

O índice z pondera a distância entre um ponto e a média, em termos do desvio padrão. O conjunto normalizado possui média 0 e desvio padrão 1. E as propriedades (assimetria e curtose) do conjunto original são conservadas (Triola, 2017).

A transformação potência é uma transformação muito utilizada e transformações, como a logarítmica e a raiz quadrada, são casos particulares desta. Pressupõe-se que após a transformação as observações resultantes sejam independentes e normalmente distribuídas

(Pino, 2014). Considere que y simboliza os dados originais, a transformação de Box-Cox consiste em encontrar um λ tal que os dados transformados $y^{(\lambda)}$ se aproximem de uma distribuição normal (Pino, 2014). A transformação de Box-Cox para obter a normalidade é dada pela Equação (4.5).

$$y^{(\lambda)} = \frac{(y^{\lambda} - 1)}{\lambda}, \lambda \neq 0$$
$$y^{(\lambda)} = \log(y), \lambda = 0 \quad (4.5)$$

Os dados analisados não apresentam zeros, por isso a equação (4.5) pode ser simplifica. No caso em que $\lambda = 0$, a transformação é chamada de transformação logarítmica e é dada pela Equação (4.6), onde y simboliza os dados originais e y_i o resultado da normalização.

$$y_i = \log(y) \quad (4.6)$$

No caso em que $\lambda = \frac{1}{2}$ a transformação é chamada de transformação raiz quadrada é obtida conforme a Equação (4.7).

$$y_i = 2\sqrt{(y)} \quad (4.7)$$

Capítulo 5: Desenvolvimento Prático no Tratamento de *Outliers*

Neste capítulo é descrita a metodologia empregada no tratamento de *outliers*. E são apresentados os resultados obtidos e conclusões.

5.1. Grupos de Testes

Foram selecionados parâmetros de interesse extraídos conforme o trabalho desenvolvido por (Fernandes et al., 2019), como descrito na seção 2.4. Nomeadamente, 194 sujeitos saudáveis, 69 sujeitos com Disfonia, 41 sujeitos com Laringite Crónica e 169 sujeitos com Paralisia das Cordas Vocais.

Considerou-se a independência do género feminino e masculino na identificação de patologias para os parâmetros estudados. Isto é, há a concatenação entre dados de indivíduos do sexo feminino e masculino. Conforme observado em Teixeira et al., (2018) a separação dos género, antes do processo de classificação não traz melhora significativa na identificação de patologias, exceto para o parâmetro *jitter* absoluto. A Tabela 3 indica as patologias, a abreviatura, a quantidade de pacientes, as médias e o desvio padrão das idades, em cada grupo patológico de teste (sem distinção de género nas amostras).

Tabela 3 - Descrição dos dados utilizados.

Grupo de Teste	Abreviação	Nº de sujeitos	Média de Idade	Desvio Padrão de Idade
Controlo	C	194	38,06	14,36
Disfonia	D	69	47,38	16,27
Laringite Crónica	LC	41	49,69	13,47
Paralisia da Prega Vocal	PCV	169	57,75	13,77

A identificação das patologias foi realizada individualmente, isto é, cada patologia foi comparada com o grupo de controlo (Controlo x Disphonia, Controlo x Laringite Crónica e Controlo x Paralisia das Cordas Vocais).

5.2. Arquitetura dos Classificadores

Aos grupos teste foram aplicadas duas redes neurais, com arquiteturas distintas, para classificar cada sujeito em controle (saudável) ou patológico (com a patologia testada).

A primeira RNA é uma arquitetura MLP ou feed-forward com 8 nós na camada oculta, e 1 nó na camada de saída que contém a classificação como patológica (saída = 1) ou não patológica (saída = 0). A função de ativação para a camada oculta é a função Logístico-Sigmoidal e na camada de saída é a função de transferência Linear. A função de treinamento utilizada atualiza o peso e estados de bias de acordo com a otimização de Levenberg-Marquardt. Os parâmetros de entrada são *jitter* relativo (jitter), *shimmer* relativo (Shim) e HNR, para 9 arquivos de fala. A matriz de entrada consiste em 27 colunas x N linhas, onde 27 correspondem a 9 observações para os 3 parâmetros e N é o número de sujeitos.

Os sujeitos foram divididos para cada grupo do MLP, em conjuntos de treinamento, validação e teste. Essas relações estão descritas na Tabela 4, o conjunto de dados foi dividido em 70% para treinamento, 15% de validação e 15% para teste. O conjunto de validação é usado para parar o treinamento antecipadamente, com a finalidade de evitar o *overfitting* e o conjunto de testes para avaliar o desempenho por meio da taxa de acerto. Como os valores iniciais da RNA dependem da semente, a acurácia final será diferente a cada iteração. Portanto, 20 sessões de treinamento da RNA foram realizadas e o melhor resultado foi mantido.

A segunda RNA é uma arquitetura LSTM com 3 neurónios na camada de entrada e 2 na camada de saída (Guedes et al, 2018). A função de Transferência Máxima Suave (*Softmax*) é empregada e realiza a classificação binária (0 ou 1). Os parâmetros de entrada para o LSMT são *jitter* relativo (*jitter*), *shimmer* relativo (Shim) e autocorrelação. A matriz de entrada está organizada em 3 colunas x N linhas, onde 3 correspondem aos parâmetros de entrada e N é o número de sujeitos multiplicado por 9 arquivos de fala (cada arquivo é considerado uma instância).

Os sujeitos foram divididos para cada grupo do LSTM em conjuntos de treinamento, validação e teste. Essas relações estão descritas na Tabela 4, onde o conjunto de dados foi dividido em 70% para treinamento, 15% de validação e 15% para teste. O conjunto de validação é usado para parar o treinamento antecipadamente, com a finalidade de evitar o *overfitting*, e o conjunto de testes para avaliar o desempenho por meio da taxa de acerto. Novamente, a RNA apresentou 20 sessões de treinamento e apenas o melhor resultado foi retido.

Tabela 4 – Base de dados para cada modelo de reconhecimento usando MLP e LSTM.

Classificador	Modelo de Reconhecimento	Total sujeitos (100%)	Treinamento	Validação	Teste
MLP	Controlo x Disfonia	263	185	39	39
	Controlo x Laringite	235	165	35	35
	Controlo x Paralisia	363	255	54	54
LSTM	Controlo x Disfonia	1143	802	156	185
	Controlo x Laringite	727	524	93	110
	Controlo x Paralisia	3267	2360	416	491

5.3.Pré-Processamento do Conjunto de Dados

5.3.1. Normalização *a priori*

O pré-processamento consiste em normalizar o conjunto de dados, identificar os *outliers* e alterar seu valor por um valor-limite determinado de acordo com o método utilizado para a identificação de *outliers*. Os métodos do Diagrama da Caixa e do Desvio Padrão foram usados para o conjunto de dados de cada modelo de reconhecimento. Para cada método foram experimentadas nenhuma normalização, normalização com Z-score, normalização Logarítmica e normalização da Raiz Quadrada.

O método aplicado para o tratamento de *outliers* é o de preenchimento, depois que um *outlier* é identificado. Seu valor é substituído pelo valor limite, estabelecido de acordo com o método escolhido. O limite é calculado pelos métodos, Diagrama de Caixa, onde o limite é calculado a partir do intervalo interquartílico ou, Desvio Padrão, onde o limite é calculado pela distância em desvios padrão da média. A substituição do *outlier* pelo Valor Limite Inferior (VLI) ocorre para valores menores que o VLI. Além disso, a substituição pelo Valor Limite Superior (VLS) ocorre para valores maiores que o VLS.

Para o caso de novos sujeitos (amostras) serem adicionados ao conjunto de dados, a verificação do processo de reconhecimento deve ser feita pelo valor limite anteriormente determinado, com o conjunto de dados original. Isto é, não é necessário recalcular os limites.

5.3.2. Normalização *a posteriori*

O pré-processamento consiste em identificar os *outliers*, alterar seu valor por um valor-limite determinado de acordo com o método utilizado para a identificação de *outliers e*

posteriormente normalizar o conjunto de dados. Os métodos do Diagrama da Caixa e do Desvio Padrão foram usados para o conjunto de dados de cada modelo de reconhecimento. Para cada método foram experimentadas nenhuma normalização, normalização com Z-score, normalização Logarítmica e normalização da Raiz Quadrada. O método de Normalização *a posteriori* descrito é semelhante aquele com Normalização *a priori*, diferenciando apenas quanto a ordem das etapas de aplicação.

5.4. Resultados e Discussões

Inicialmente, será discutido o procedimento de normalização e, em seguida, será analisado o método de tratamento de *outliers*. Para o procedimento de normalização, independentemente do tipo de rede utilizada, seja rede MLP ou rede LSTM, o melhor resultado obtido foi com a aplicação do método Logarítmico que possui o conjunto de dados com uma distribuição mais próxima da gaussiana, considerando a maioria dos conjuntos de dados.

A partir do conjunto de dados referente à classificação com a MLP e com objetivo de identificar o método de normalização mais apropriado foi apresentada e Figura 12. Nesta, os grupos Controlo x Paralisia são comparados, com base em uma análise crítica do parâmetro de *jitter* relativo.

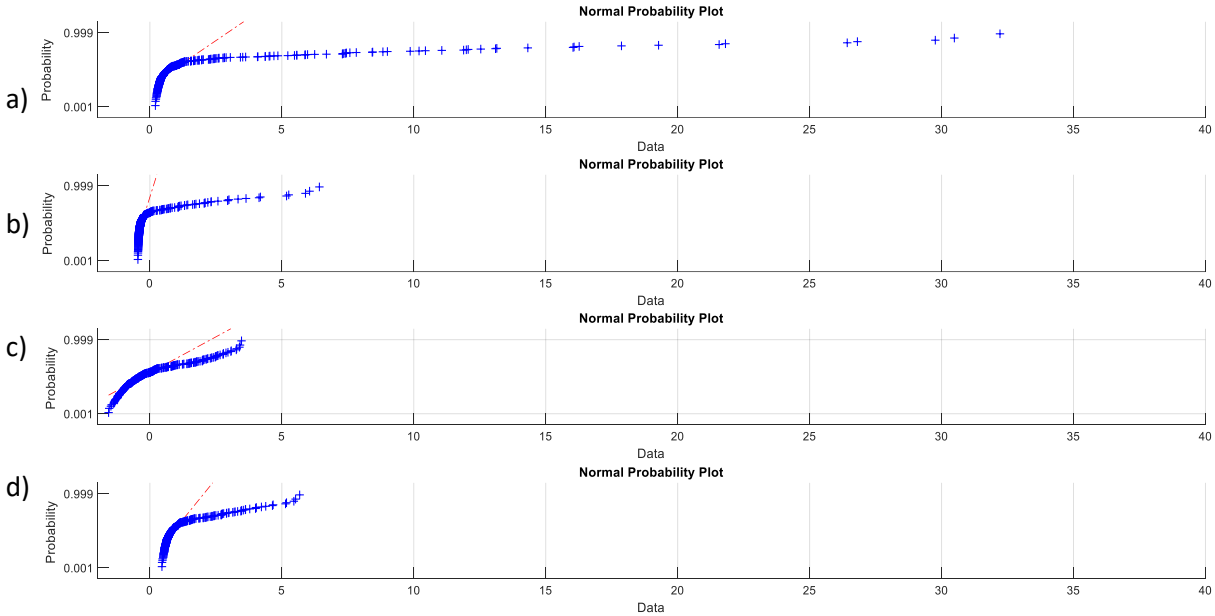

Figura 12 – Normalização dos dados de entrada (jitter): (a) original; (b) normalização do Z-score; (c) normalização Logarítmica; (d) normalização da Raiz Quadrada.

A Figura 12 contém um gráfico de probabilidade normal que compara a distribuição dos dados com a distribuição normal. Na qual, a linha vermelha indica a distribuição normal teórica e cada ponto azul representa uma instância. Se o conjunto de dados tiver uma distribuição normal, então os pontos irão sobrepor a linha de referência. A normalização Logarítmica tem um comportamento quase linear que se assemelha à linha vermelha. Assim, enfatiza-se que a normalização obtida a partir do Logaritmo é a que mais aproxima os dados de uma distribuição normal. Mesmo que após a normalização com a técnica Logarítmica, os dados apresentem características mais aproximadas à de uma distribuição normal, esse fato pode não levar a uma melhoria efetiva na capacidade de classificação das Redes Neuronais. Assim, para fins de comparação, ambas as redes foram testadas para todos os métodos de normalização e também na ausência de normalização.

Para demonstrar os resultados das técnicas de processamento de *outliers* e seu desempenho, a Figura 13 mostra um exemplo com o parâmetro *jitter* para o conjunto de dados Controlo x Laringite com MLP, com os valores originais (ausência de normalização) e após a aplicação dos métodos BP e DP.

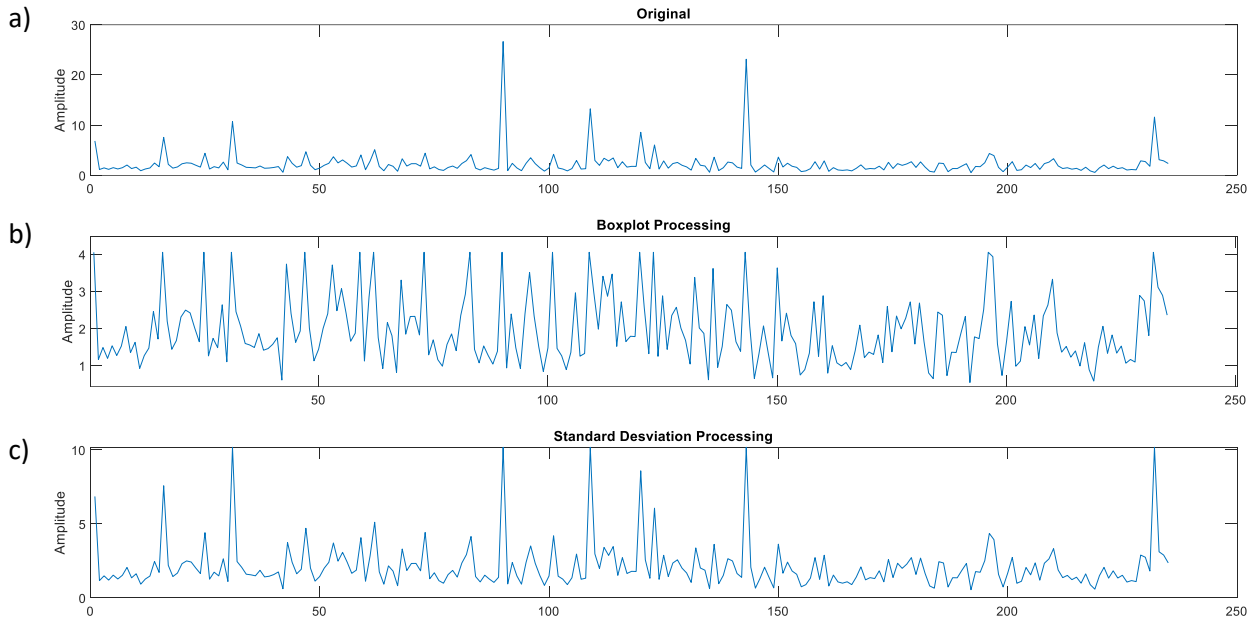

Figura 13 – Parâmetros jitter do conjunto de dados MLP: (a) original; (b) processamento de Box Plot; c) processamento de Desvio Padrão.

A RNA deve enfrentar maiores dificuldades na classificação dos sujeitos patológicos na presença de *outliers*, porque estes podem distorcer os resultados obtidos. Na Figura 13, na

representação referente aos dados originais, alguns pontos que têm picos muito grandes, sendo este um comportamento diferente dos demais pontos do conjunto. Logo, correspondem a valores *outliers*. Na representação do tratamento de *outliers* com o método Box Plot, o limite superior parece ser muito baixo, deixando pequena a diferença de posição entre os pontos localizados nos limites máximo e mínimo. Na representação do tratamento de *outliers* com o método DP, os picos dos *outliers* são reforçados. Portanto, o método mais conveniente parece ser o DP, pois torna a identificação dos *outliers* visualmente mais fácil. No entanto, não se pode concluir sobre a acurácia do método ainda, é necessário levar em consideração os resultados da classificação.

A Tabela 5 apresenta a melhoria na acurácia no conjunto de testes após a aplicação *a priori* dos métodos de normalização, combinados com os métodos de detecção e tratamento de *outliers*, para MLP e LSTM. A primeira coluna mostra a melhor acurácia sem processamento de *outliers* (SP). Nas demais colunas é apresentada a acurácia máxima, para cada método e modelo de reconhecimento. No qual, os métodos são identificados como *Box Plot* (BP), *Box Plot* com normalização Z-score (BP1), *Box Plot* com normalização Logarítmica (BP2), *Box Plot* com normalização Raiz Quadrada (BP3), Desvio Padrão (DP), DP com normalização Z-score (DP1), DP com normalização Logarítmica (DP2), DP com normalização Raiz Quadrada (DP3).

Tabela 5 – Comparação da acurácia da MLP e LSTM na identificação de patologias.

Classificador	Modelo de Reconhecimento	SP (%)	BP (%)	BP1 (%)	BP2 (%)	BP3 (%)	DP (%)	DP1 (%)	DP2 (%)	DP3 (%)
MLP	Controlo Disfonia	69	69	77	64	62	74	80	77	72
	Controlo Laringite	80	74	83	80	74	71	89	86	83
	Controlo Paralisia	72	72	78	74	70	76	78	70	74
LSTM	Controlo Disfonia	59	68	67	68	64	66	67	66	61
	Controlo Laringite	63	68	76	62	67	65	68	70	65
	Controlo Paralisia	68	71	72	69	63	70	71	71	69

SP – Sem processamento de *outlier*, BP – Método Box Plot, BP1 – Método Box Plot com normalização Z-score, BP2 – Método Box Plot com normalização Logarítmica, BP3 – Método Box Plot com normalização Raiz Quadrada, DP – Método Desvio Padrão, DP1 – DP com normalização Z-score, DP2 – DP com normalização Logarítmica, DP3 – DP com normalização Raiz Quadrada.

No caso de Controlo x Disfonia usando o MLP, a acurácia aumentou de 69% para 80% com o DP1. No Controlo x Laringite, a acurácia aumentou de 80% para 89% com o método DP2. No reconhecimento entre Controle x Paralisia, usando MLP, a acurácia aumentou de 72% para 78%, novamente com o método DP1, também com o método BP1 foi alcançada a mesma acurácia.

Para a LSTM, no reconhecimento entre Controlo x Disfonia, a acurácia aumentou de 59% para 68% com os métodos BP e BP2. Os métodos BP1 e DP1 alcançaram uma acurácia de 67%, sendo esta, muito próxima do máximo valor de acurácia. No caso de Controlo x Laringite, a acurácia aumentou de 63% para 76%, com o método BP1. No caso de Controlo x Paralisia, a melhoria foi de 68% para 72% com o método BP1. Nesse caso, o método DP1 atingiu 71%, obtendo uma acurácia muito próxima da máxima.

A MLP parece funcionar bem com o método DP, em particular o DP1. Por outro lado, a LSTM parece funcionar bem com o método BP, em particular o BP1. Utilizando a MLP, o desvio padrão e o método de normalização do Z-score (DP1) alcançaram a melhor acurácia na identificação da patologia vocal. O BP1 alcançou uma acurácia semelhante. Quando a classificação é feita com o LSTM, o *Box Plot* com normalização Z-score (BP1) obteve o melhor desempenho. O DP1 apresentou um desempenho similar.

Em relação à normalização, considerando a maioria dos casos em estudo, o Z-score teve melhor desempenho que os métodos Logarítmicos e da Raiz Quadrada. Parece que, embora o método de normalização Logarítmica seja melhor para normalizar o conjunto de dados, o Z-score permite uma melhor acurácia na classificação. Com relação à detecção e tratamento dos *outliers*, os métodos DP e BP executam com acurácia semelhante quando combinados com a normalização Z-score.

Na Tabela 6, diferente do realizado e descrito anteriormente, referente à metodologia na identificação de *outliers*, foi experimentada uma variante na qual o processo de normalização dos dados é implementado após a identificação das observações anômalas. Na qual, a primeira coluna mostra a melhor acurácia sem processamento de *outliers* (SP). Nas demais colunas é apresentada a acurácia máxima, para cada método e modelo de reconhecimento. No qual, os métodos são identificados como Box Plot (BP), Box Plot com normalização Z-score (BP1), Box Plot com normalização Logarítmica (BP2), Box Plot com normalização Raiz Quadrada (BP3), Desvio Padrão (DP), DP com normalização Z-score (DP1), DP com normalização Logarítmica (DP2), DP com normalização Raiz Quadrada (DP3).

Tabela 6 – Comparação da acurácia da MLP com normalização posterior a identificação de outliers.

Classificador	Modelo de Reconhecimento	SP (%)	BP (%)	BP1 (%)	BP2 (%)	BP3 (%)	DP (%)	DP1 (%)	DP2 (%)	DP3 (%)
MLP	Controlo Disfonia	69	69	82	77	72	74	77	82	72
	Controlo Laringite	80	74	80	78	83	71	75	80	80
	Controlo Paralisia	72	72	80	72	70	76	76	74	76

SP – Sem processamento de *outlier*, BP – Método Box Plot, BP1 – Método Box Plot com normalização Z-score, BP2 – Método Box Plot com normalização Logarítmica, BP3 – Método Box Plot com normalização Raiz Quadrada, DP – Método Desvio Padrão, DP1 – DP com normalização Z-score, DP2 – DP com normalização Logarítmica, DP3 – DP com normalização Raiz Quadrada.

Deste modo, são apresentados os resultados com normalização *a posteriori* na Tabela 6, para o conjunto de dados referente a classificação com MLP e para cada um dos modelos de reconhecimento. No caso de Controlo x Disfonia usando o MLP, a acurácia aumentou de 69% para 82% com o BP1 e também DP2. No Controlo x Laringite, a acurácia aumentou de 80% para 83% com o método BP3. No reconhecimento entre Controle x Paralisia, usando MLP, a acurácia aumentou de 72% para 80%, com o método BP1.

A MLP parece funcionar bem com o método BP, particularmente o BP1 e BP3. Utilizando a MLP, o *Box Plot* e o método de normalização da raiz quadrada (BP3) alcançou a melhor acurácia na identificação da patologia vocal. Em relação à normalização, considerando a maioria dos casos em estudo, o Z-score teve melhor desempenho que os métodos Logarítmicos e da Raiz Quadrada, pois este permite uma melhor acurácia na classificação. Com relação à detecção e tratamento dos *outliers*, os métodos BP1 e DP2 executam com acurácia semelhante.

De modo geral, na Tabela 6 foram obtidos valores com menor variação entre a acurácia obtida e os métodos aplicados. No entanto, ao comparar caso a caso a Tabela 5 e Tabela 6 não se percebe uma diferença muito significativa. Entretanto, em alguns casos houve alguma melhoria (Em Controlo x Disfonia, o método BP2 obteve 64% com a normalização na etapa inicial e 77% quando a normalização foi feita como etapa final, sendo um aumento percentual de 13 pontos).

Portanto, o melhor método é aquele que permite obter maior acurácia, isto é, o método a normalização Z-score aplicado *a priori* e combinado com o método de identificação de *outliers Box Plot*, no caso Controlo x Laringite, o método DP1 obteve 89% de acurácia na classificação.

5.5.Conclusão

O pré-processamento do conjunto de dados para o procedimento de reconhecimento de patologias vocais foi analisado e comparado experimentalmente a melhoria na acurácia, usando três modelos de reconhecimento para Disfonia, Laringite Crónica e Paralisia de Cordas Vocais, com duas arquiteturas de redes neuronais diferentes, as arquiteturas MLP e LSTM.

O pré-processamento consiste na identificação e tratamento dos *outliers*. Os métodos *Box Plot* (BP) e Desvio Padrão (DP) foram comparados. Como esses métodos podem lidar melhor com a distribuição gaussiana, foram realizados procedimentos de normalização usando o Z-score, os métodos Logarítmicos e de Raiz Quadrada. Considerou-se a aplicação da normalização anterior ou posterior ao método de identificação de anomalias.

Em relação à normalização *a priori*, embora o método Logarítmico tenha apresentado melhor ajuste com uma curva normal, o Z-score obteve uma melhoria maior na acurácia. Em relação à identificação e correção *outliers*, ambos os métodos, BP e DP mostraram acurácia semelhante com a normalização do Z-score. O método DP1 melhora na acurácia, entre 3 e 11 pontos em porcentagem, enquanto o método BP1 melhora na acurácia, entre 3 e 13 pontos em percentagem.

Em relação à normalização *a posteriori*, embora o método o Z-score obteve uma melhoria maior na acurácia. Em relação à identificação e correção *outliers*, ambos os métodos, BP1 e DP2 mostraram acurácia semelhante com a normalização do Z-score. O método BP1 melhora na acurácia, entre 8 e 13 pontos em porcentagem, enquanto o método DP2 melhora na acurácia, entre 2 e 13 pontos em porcentagem.

Como conclusão final, ambos métodos BP1 ou SD1 com normalização *a priori*, são recomendáveis para pré-processamento dos conjuntos de dados para reconhecimento de patologias vocais.

Capítulo 6: Seleção de Parâmetros

Neste capítulo, serão abordados os problemas enfrentados ao analisar bases de dados com grandes dimensões e também os principias métodos de seleção de parâmetros.

6.1.Definição de Seleção de Parâmetros

Desde o início dos anos 70, a seleção de parâmetros tem sido objeto de investigação e desenvolvimento constante nas áreas de mineração e redução de dados, aprendizado de máquina e reconhecimento de padrões (H. D. Lee, 2005). A seleção de parâmetros é empregada para solucionar problemas como a alta dimensionalidade, o risco de *overfitting* e os resultados tendenciosos, encontrados em conjuntos de buscas com excesso de informações devido a presença de muitos parâmetros (Guyon & Elisseeff, 2003; Ventura, 2013).

Os algoritmos de aprendizagem são afetados por informação de treino redundante e pouco relevante (Barbosa, 2013; Fernandes, 2017). Portanto, a seleção e a remoção de características pouco relevantes do conjunto inicial de dados reduzem a dimensão dos dados, o gasto computacional e resulta no aperfeiçoamento da previsão dos preditores (Guyon & Elisseeff, 2003).

O problema de Seleção de Parâmetros (SP) parte do problema geral de redução da dimensionalidade, cujo objetivo é selecionar o melhor subconjunto (S) de *features* ou parâmetros dentro do conjunto disponível (X), de modo que (6.1) (Fernandes, 2017):

$$S \subseteq X \tag{6.1}$$

Todos os métodos de seleção de parâmetros são constituídos por dois componentes principais. O primeiro componente é o critério de avaliação da qualidade de um subconjunto de parâmetros, também chamado de função de custo que define se um subconjunto é melhor que outro e qual parâmetro adicionar ou eliminar. O segundo componente é a estratégia de busca dentro de um conjunto de parâmetros candidatos (Bishop, 1995).

Deste modo, as dimensões da seleção de parâmetros podem ser subdivididas em: os pontos de partida ou a direção em que a busca será realizada, a estratégia da busca, o critério para avaliação dos subconjuntos gerados e o critério de paragem (Apolónia, 2018).

6.2. Direção de Busca

A seleção progressiva (*forward selection*) é uma estratégia de busca incremental linear que seleciona variáveis disponíveis individuais, começando com o subconjunto vazio e adicionando um parâmetro de cada vez (May, Dandy, & Maier, 2011; Doak, 1992). Os parâmetros são progressivamente adicionados (existe uma ordenação parcial dos estados, pois cada estado possui um parâmetro a mais em relação ao estado anterior) ao grupo de treinamento e se houver melhora na performance do algoritmo, o parâmetro específico será incluído ao subgrupo final (Guyon & Elisseeff, 2003). A seleção não considera todas as combinações possíveis e busca apenas em um subconjunto pequeno, sendo possível que o algoritmo encontre um conjunto ótimo local de variáveis de entrada e interrompa antecipadamente (May et al., 2011)

A eliminação retrógrada (*backward elimination*) inicia com o conjunto total de parâmetros e remove um de cada vez (Doak, 1992), sendo o oposto da abordagem de seleção progressiva. Isto é, inicialmente todos os parâmetros de entrada disponíveis são selecionados e testados. E então são gradualmente eliminados aqueles que não são importantes para otimização da performance do modelo (May et al., 2011)

A seleção bidirecional efetua duas pesquisas paralelas, por iteração, uma para adicionar um parâmetro e outra para excluir (Doak, 1992). Esta abordagem, pode ser vantajosa sempre que não haja informação sobre o número de parâmetros do subconjunto ótimo, o qual pode estar localizado na região central do espaço de busca (H. D. Lee, 2005).

A seleção randômica/aleatória não possui uma direção específica na qual a busca deva ocorrer. O objetivo dessa abordagem é evitar que a busca fique em um mínimo local, por meio da não fixação de como os subconjuntos serão gerados (H. D. Lee, 2005).

6.3. Estratégia de Busca

As estratégias de busca podem ser locais que começam em um local de início e se movem de forma incremental ou serem globais que consideram muitas combinações (May et al., 2011).

A busca exaustiva ou completa analisa todas as combinações possíveis de variáveis de entrada e seleciona o conjunto ótimo. Uma avaliação exaustiva de todas essas combinações

possíveis de parâmetros pode ser viável quando a dimensionalidade do conjunto candidato é baixa, mas se torna inviável à medida que a dimensionalidade aumenta (May et al., 2011).

A busca aleatória ou não-determinística, tem em seu processo de busca a eleição do subconjunto ótimo de parâmetros de modo aleatório. Ao longo das etapas de seleção, a seleção pode ser completamente aleatória ou sequencial (porém, sempre há alguma aleatoriedade na escolha) (Kumar & Minz, 2014). O propósito desta abordagem é conseguir escapar dos mínimos locais do espaço de busca (Doak, 1992).

A busca heurística é utilizada em problemas de otimização, onde o espaço de busca é grande. Os algoritmos com estas características têm facilidade em encontrar com eficiência soluções globais. Este tipo de algoritmos implementa uma pesquisa que combina avaliação aleatória de soluções em todo o espaço de busca, com um mecanismo para aumentar o foco da pesquisa em regiões que levam a boas soluções (May et al., 2011).

A busca sequencial seleciona um ou mais parâmetros, num processo progressivo e iterativo. Apresenta uma abordagem completa, fácil de implementar, mas pode não obter o subconjunto ótimo (Kumar & Minz, 2014).

6.4. Parada de Busca

O critério de parada da busca é uma decisão importante a ser tomada. Alguns possíveis critérios são: a) parar de remover ou adicionar parâmetros quando nenhuma das alternativas melhora a acurácia da estimativa para a classificação, b) continuar revisando o subconjunto de parâmetros enquanto a acurácia não se degrada, c) continuar gerando subconjuntos candidatos até que o outro extremo do espaço de busca seja alcançado e escolher o melhor desses subconjuntos, d) parar quando o subconjunto de parâmetros selecionado separar perfeitamente todas as classes (assumindo que não há ruídos nos dados) e e) ordenar os parâmetros segundo alguma pontuação de importância e utilizar um parâmetro de sistema para determinar o ponto de parada, por exemplo, o número de parâmetros desejado para o subconjunto. Essa alternativa é mais robusta que a anterior (Apolónia, 2018).

6.5. Critério de Seleção

Os critério de seleção apresentam três diferentes abordagens: métodos do tipo filtro, *wrapper e embedded* (Guyon & Elisseeff, 2003). Deste modo, a relação entre a acurácia do classificador e o subconjunto de parâmetros selecionado pode ser avaliado independentemente

ou em função do classificador (Kohavi & John, 1997). Os métodos do tipo híbrido realizam a seleção de parâmetros combinando os diferentes métodos anteriormente citados (Fernandes, 2017).

Os filtros selecionam um subgrupo de parâmetros numa etapa do pré-processamento dos dados. Como não dependem do algoritmo de aprendizagem, são em geral mais rápidos. Estimam um índice de relevância para cada parâmetro individualmente, que reflete a importância do parâmetro de forma isolada para a solução do problema e ignoram a possibilidade de interdependência de parâmetros, podendo não identificar o melhor subgrupo de parâmetros para a solução. Entretanto, são teoricamente menos propensos a *overfitting* (Guyon & Elisseeff, 2003)

Os *wrappers* selecionam parâmetros através da execução de um algoritmo específico de aprendizagem de máquina para avaliar as combinações de subconjunto de parâmetros a serem considerados (Kohavi & John, 1997). Se o número de parâmetros for insuficiente, o risco de *overfitting* aumenta (Kumar & Minz, 2014).

Os algoritmos de seleção do tipo *embedded* combinam as vantagens dos dois métodos anteriores. Formam uma categoria de seletores de parâmetros onde a seleção está "embutida" no processo de forma que depende do modelo de aprendizagem (May et al., 2011). Isto é, seleção e classificação de parâmetros seguindo a critérios gerados durante o processo de treinamento do algoritmo de aprendizado (Guyon & Elisseeff, 2003; Ventura, 2013). Tem baixo custo computacional, consegue detectar dependências entre parâmetros (a influência individual de cada parâmetro no desempenho do modelo) e conseguem obter a solução ótima mais rapidamente (Bolón-Canedo et al., 2012; Guyon & Elisseeff, 2003).

Os critérios de determinação dos parâmetros relevantes empregados nesta dissertação fundamentam-se na aplicação de técnicas como a Correlação, o ReliefF, Test t de Welch e a análise de Regressão Multilinear.

6.5.1. Coeficiente de Correlação Linear de Pearson

Um dos esquemas mais simples de filtragem é a avaliação de cada parâmetro individualmente, baseada na sua correlação (Lee, 2005). Um bom subconjunto de parâmetros deve apresentar pouca correlação entre os parâmetros e forte correlação com a saída. Em outras

palavras, um parâmetro é importante se estiver correlacionado ou preditivo da classe; caso contrário, é irrelevante (Hall, 1999; Kohavi & John, 1997)

Deste modo, considerou-se o Coeficiente Linear de Pearson na determinação da relevância dos atributos. O qual, é definido na Equação (6.2) como $rho(\mathrm{a},\mathrm{b})$ (Gibbons & Chakraborti, 2003).

$$rho(\mathrm{a},\mathrm{b}) = \frac{\sum_{i=1}^{n}(\mathrm{X}_{a,i} - \overline{X_a})(\mathrm{Y}_{b,j} - \overline{Y_b})}{\left\{\sum_{i=1}^{n}(\mathrm{X}_{a,i} - \overline{X_a})^2 \sum_{j=1}^{n}(\mathrm{Y}_{b,j} - \overline{Y_b})2\right\}^{\frac{1}{2}}} \tag{6.2}$$

Para a coluna X_a na matriz X e a coluna X_b na matriz Y são calculadas as médias nas Equações (6.3) e (6.4), onde n é o comprimento de cada colunas.

$$\overline{Xa} = \frac{\sum_{i=1}^{n}(\mathrm{X}_{a,i})}{n} \tag{6.3}$$

$$\overline{Xb} = \frac{\sum_{j=1}^{n}(\mathrm{X}_{\mathrm{b,j}})}{n} \tag{6.4}$$

6.5.2. ReliefF

O algoritmo ReliefF desenvolvido por Kononenko (1994) é um dos únicos algoritmos de abordagem filtro capaz de avaliar dependências de atributos (Cavique, Mendes, Funk, & Santos, 2013). Utiliza o conceito de vizinhos mais próximos para obter estatísticas de atributos que indiretamente representam interações (Urbanowicz, Meeker, La Cava, Olson, & Moore, 2018).

O algoritmo penaliza os preditores que atribuem valores diferentes aos vizinhos da mesma classe e recompensa os preditores que atribuem valores diferentes aos vizinhos de classes diferentes (Kononenko, Šimec, & Robnik-Šikonja, 1997). O peso de cada recurso reflete sua capacidade de distinguir entre os valores de classe que podem variar no intervalo de [-1, +1]. Um atributo relevante apresenta valores positivos de peso (Kohavi & John, 1997).

O ReliefF identifica duas observações vizinhas mais próximas do alvo; um com a mesma classe, chamada de ocorrência de acerto mais próxima H e a outra com a classe oposta, chamada de ocorrência de erro mais próxima M. A última etapa do ciclo atualiza o peso de um atributo A em W, se este tiver um valor diferente entre a observação de destino R_i e as observações H ou de M. A função $diff$ calcula a diferença no valor do elemento A entre duas instâncias I_1 e I_2, onde $I_1 = Ri$ e I_2 é H ou M, ao executar atualizações de peso (Urbanowicz et al., 2018).

Para atributos discretos $diff$ é definido como na Equação (6.5)

$$\begin{aligned} &diff\,(\mathrm{A}, \mathrm{I}_1, \mathrm{I}_2) = 0, \mathrm{valor}(\mathrm{A}, \mathrm{I}_1) = \mathrm{valor}(\mathrm{A}, \mathrm{I}_2) \\ &diff\,(\mathrm{A}, \mathrm{I}_1, \mathrm{I}_2) = 1, \mathrm{outros} \end{aligned} \tag{6.5}$$

No caso de atributos contínuos é definido como na Equação (6.6)

$$diff\,(\mathrm{A}, \mathrm{I}_1, \mathrm{I}_2) = \frac{valor(\mathrm{A}, \mathrm{I}_1) - valor(\mathrm{A}, \mathrm{I}_2)}{\max(\mathrm{A}) - \min(\mathrm{A})} \tag{6.6}$$

6.5.3. Teste t de Welch

O Teste t de Welch é uma medida que avalia os subconjuntos de parâmetros de acordo com a abordagem filtro. O Teste t de Welch é uma alternativa viável para o Teste t clássico (Welch, 1947). Utilizado no caso de não se presumir que as duas amostras de dados são de populações com variações iguais, a estatística de teste sob a hipótese nula tem uma distribuição aproximada de t Student com um número de graus de liberdade dado pela aproximação de Satterthwaite (Welch, 1947). O Teste t Welch define a estatística t como na Equação (6.7) (Williams, 1984):

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \tag{6.7}$$

Onde $\bar{x}$ e $\bar{y}$ são as médias da amostra, s_x e s_y são os desvios padrão da amostra e n são os tamanhos das amostras. No caso em que se assume que as duas amostras de dados são de populações com variações iguais, a estatística de teste sob a hipótese nula tem distribuição t de Student com v graus de liberdade. Os graus de liberdade são determinados pela Equação (6.8) , onde $v_1 = N_1 - 1$ e $V_2 = N_2 - 1$ são os graus de liberdade relacionados à primeira e à segunda variação estimada (Gliozzo, 2016).

$$v = \frac{\left(\frac{s^2_1}{N_1} + \frac{s^2_2}{N_2}\right)}{\frac{s^4_1}{N^2_1 v_1} + \frac{s^4_2}{N^2_2 v_2}} \tag{6.8}$$

Após o cálculo de v e t , esses valores podem ser usados com a distribuição t para testar a hipótese nula de que as duas médias são iguais (Gliozzo, 2016).

6.5.4. Análise de Regressão Multilinear

A Regressão por etapas é um método sistemático para adicionar e remover os parâmetros de um modelo multilinear com base em sua significância estatística em uma regressão. Enquanto uma regressão simples de duas variáveis resulta na equação de uma reta, um problema de três variáveis implica num plano, e um problema de k variáveis implica um hiperplano. Um modelo multilinear pode ser representado pela Equação (6.9) (Song & Lu, 2017):

$$Y_c = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{6.9}$$

- Onde, a corresponde ao intercepção do eixo y;
- b_i, corresponde ao coeficiente angular da i-ésima variável;
- k, corresponde ao número de variáveis independentes.
- y , é a variável dependente.

O método de Regressão Multilinear emprega uma abordagem *embedded* na avaliação dos parâmetros. Este, inicia com um conjunto inicial e, em seguida, compara o poder explicativo

de modelos cada vez maiores e menores. Em cada etapa, o p-valor de uma estatística-F é calculado para testar os modelos com e sem um determinado parâmetro candidato. Se o parâmetro não estiver atualmente no modelo, a hipótese nula é de que o parâmetro teria um coeficiente zero se adicionado. Se houver evidência suficiente para rejeitar a hipótese nula, o parâmetro será adicionado. Por outro lado, se um parâmetro está atualmente no modelo, a hipótese nula é que o termo tenha um coeficiente zero. Se não houver evidências suficientes para rejeitar a hipótese nula, o parâmetro será removido (Draper & Smith, 1998).

O modelo inicial foi ajustado considerando que o critério de entrada típico para que um parâmetro entre no modelo é um p-valor menor de 0,05 e para que um parâmetro seja retirado do modelo é um p-valor ser maior que 0,10 (Draper & Smith, 1998).

Capítulo 7: Seleção de Parâmetros

Esta seção destina-se a descrever a metodologia e os resultados obtidos na seleção de parâmetros.

7.1. Método de Seleção de Parâmetros

Os parâmetros acústicos utilizados nesta etapa foram extraídos das gravações de fala da base SVDT (Barry & Pützer, 2018). Contém 20 parâmetros, sendo estes o *jitter* absoluto, *jitter* relativo, *shimmer* absoluto, *shimmer* relativo, autocorrelação, HNR, NHR, 13 coeficientes de MFCC. Cada parâmetro é obtido 3 vogais (/a/ /i/ e /u/) e pronunciadas 3 tons (alto, normal e baixo), totalizando 20x3x3 =180 parâmetros analisados.

O pré-processamento dos dados emprega a técnica BP1, método de identificação de *outliers Box Plot* com tratamento do tipo preenchimento e normalização Z-score (conforme o Capítulo 4 foi a técnica com melhor acurácia na classificação).

A classificação é realizada por uma RNA, com arquitetura do tipo MLP. Esta apresenta 180 elementos na camada de entrada e um na camada de saída. A saída da rede realiza a classificação binária em 0 (saudável) e 1 (patológico).

Foram testadas e comparadas diferentes combinações de função de treino, de ativação e número de neurónios na camada escondida, foi selecionada aquela que permitiu melhoria na acurácia da classificação. A função de treino escolhida foi a Propagação reversa (*Back propagation*) do gradiente conjugado em escala que atualiza o peso e valores do *bias* de acordo com o método do gradiente conjugado em escala. A função de ativação utilizada na camada escondida é a função de transferência Sigmoide Simétrica e da camada de saída é a função de transferência Linear.

Esta dissertação se insere num conjunto de trabalhos relacionados que utilizam diferentes processos e patologias, especificamente, Paralisia das Cordas Vocais, Laringite Crónica e Disfonia (Guedes et al., 2018; Fernandes et al., 2018; Fernandes et al., 2019). Neste trabalho foi utilizada a Paralisia das Cordas Vocais, por ser a patologia com maior número de indivíduos. Os pacientes foram subdivididos em duas classes, nomeadamente, saudáveis (194 sujeitos) e patológicos com Paralisia das Cordas Vocais (169 sujeitos), totalizando 363 sujeitos analisados. Considera-se que as classes estão aproximadamente balanceadas.

A saída dada pela RNA nem sempre é exatamente zeros ou uns e, portanto, teve que ser pós-processada (arredondamento) para necessariamente estar entre zero e um. O conjunto de dados foi dividido em três subconjuntos: treino, validação e teste. As quantidades de cada conjunto de dados foram de 70%, 15%, 15%, respectivamente, conforme a Tabela 7.

Tabela 7 – Conjunto de dados utilizados na classificação com MLP.

Classificador	Modelo de Reconhecimento	Total sujeitos (100%)	Treinamento	Validação	Teste
MLP	Controlo	194	136	29	29
	Paralisia	169	119	25	25
	Controlo x Paralisia	363	255	54	54

Foram implementados algoritmos que aplicam técnicas seletoras de parâmetros descritas no capítulo 6. Estas técnicas serão caracterizadas quanto a estratégia, a direção, critério de parada de busca e na avaliação do desemprenho.

7.1.1. Seleção Baseada na Correlação

Foram desenvolvidos 3 algoritmos com diferentes abordagens (método de busca) para a correlação (C1, C2, C3). O algoritmo C1 é um método de seleção sequencial progressiva e utiliza a correlação linear para determinar a relevância dos parâmetros. Para um parâmetro ser considerado relevante, é necessário que este seja fortemente correlacionado com a saída.

No final é devolvido um conjunto de parâmetros que serão usadas na entrada da rede neuronal. A Figura 14 e o pseudocódigo descrevem o funcionamento de C1. O espaço de busca inicia vazio (i=0) e a cada iteração é incrementado um parâmetro na entrada da RNA. O critério de parada é atingido quando todos os parâmetros forem incrementados (i=180). O desempenho do modelo é obtido por meio do cálculo da acurácia. A cada iteração é calculado e armazenado um novo valor de acurácia (para o conjunto de teste). Após a finalização do algoritmo o subconjunto com a máxima acurácia no teste é selecionado.

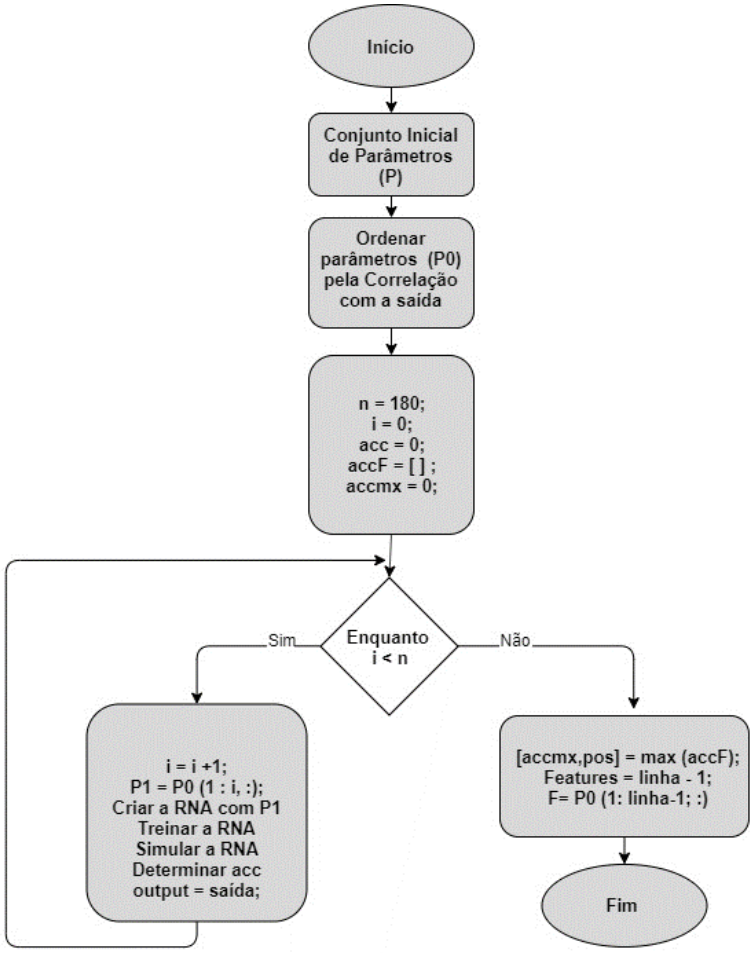

Figura 14 – Diagrama do algoritmo C1.

Algoritmo C1: Correlação 1

1- Entrada: conjunto inicial de parâmetros P = (180 parâmetros x 363 sujeitos);
2- Determinação da importância dos parâmetros: correlação linear do conjunto inicial P de parâmetros;
3- Ordenação: os parâmetros de P são ordenados de acordo com a correlação entre o parâmetro e a saída. Os parâmetros são ordenados de modo decrescente, ou seja, dos parâmetros com maior correlação com a saída para os com menor correlação com saída
4- Saída: conjunto de parâmetros ordenados P0;
5- Inicialização: n=180; i=0; acc=0; accF=[ ]; accmx = [ ];
6- Seleção Sequencial: os elementos de P1 são incrementados

```
Enquanto i < n
    Incrementa i
    P1 = P0 (1: i);
    Cria a RNA com P1 de entrada
    Determina a acurácia
    accF(i) = acc;
    i = i + 1;
Fim Enquanto
    [accmx, pos] = max (accF);
    Features = 1: pos;
    F = P0 (1: pos, :);
    acc = accmx;
```

7- Finaliza.

O algoritmo C2 é um método de seleção sequencial, onde a correlação linear determina a relevância dos parâmetros. Com relação à direção de busca, o algoritmo apresenta duas variantes: *forward selection* (seleção progressiva) e *backward elimination* (eliminação retrógrada).

A principal diferença em relação ao algoritmo anterior é que C2 possui um critério de paragem diferente, pretende-se assim selecionar menores subconjuntos representativos de parâmetros. Relativo ao critério de parada, foram implementadas duas variantes o (i=3) e o (i=20). O primeiro busca a máxima acurácia e sua posição, após a identificação destes continua procurando até mais 3 parâmetros, mesmo sem melhoria na acurácia. O segundo busca a máxima acurácia e sua posição, após a identificação destes continua procurando até mais 20 parâmetros, mesmo sem melhoria na acurácia. O processo termina quando Fim =1, pois o valor da acurácia máxima (accmx) é menor que o da acurácia anterior (accold).

O algoritmo C2 considera atributos relevantes quando tem grande correlação com a classe. O desempenho do modelo é obtido por meio do cálculo da acurácia. A cada iteração é calculado e armazenado um novo valor de acurácia (para o conjunto de teste). Após a finalização do algoritmo o subconjunto com a máxima acurácia no teste é selecionado. A Figura 15 corresponde a versão do algoritmo C2 que utilizam a seleção progressiva com critério de parada de busca, i=3. E a Figura 16 corresponde a variante de C2 com eliminação retrógrada dos parâmetros e critério de parada da busca, i=3, onde o espaço de busca inicia cheio (linha=180) e é decrementado um parâmetro a cada iteração.

O algoritmo C3 é o método de seleção sequencial que utiliza a correlação linear para determinar a relevância dos parâmetros. Diferentemente de C2, o algoritmo C3 considera um atributo relevante se este possuir forte correlação com a saída e pequena correlação entre atributos (atributo-atributo).

Com relação à direção de busca, apresenta duas variantes: *feed-forward* (seleção progressiva) e *backward elimitation* (eliminação retrógrada). A principal diferença com relação ao algoritmo C1 é que C3 possui um critério de paragem diferente, pretende-se assim selecionar menores subconjuntos representativos de parâmetros. Relativo ao critério de parada, foram implementadas duas variantes o (i=3) e o (i=20). O primeiro busca a máxima acurácia e sua posição, após a identificação destes continua procurando até mais 3 parâmetros, mesmo sem melhoria na acurácia. O segundo busca a máxima acurácia e sua posição, após a identificação destes continua procurando até mais 20 parâmetros, mesmo sem melhoria na acurácia. O

processo termina quando Fim =1, pois o valor da acurácia máxima (accmx) é menor que o da acurácia anterior (accold).

A Figura 17 descreve a versão de C3 com seleção progressiva e critério de parada, i=3. E a Figura 18 descreve a versão de C3 com eliminação retrógrada e critério de parada, i=3, onde o espaço de busca inicia cheio (linha=180) e é decrementado um parâmetro a cada iteração.

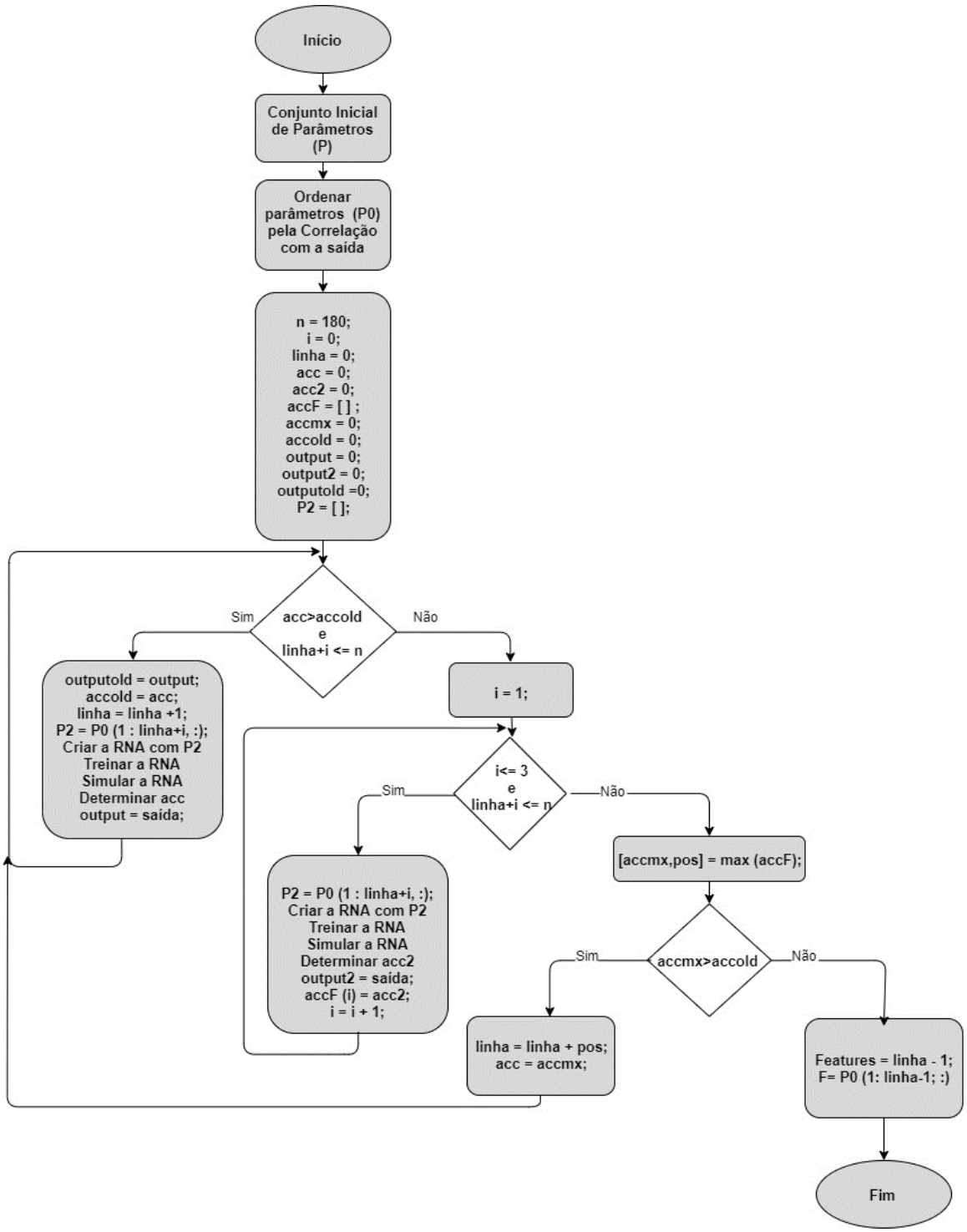

Figura 15 – Diagrama do algoritmo C2 com seleção sequencial progressiva e i=3.

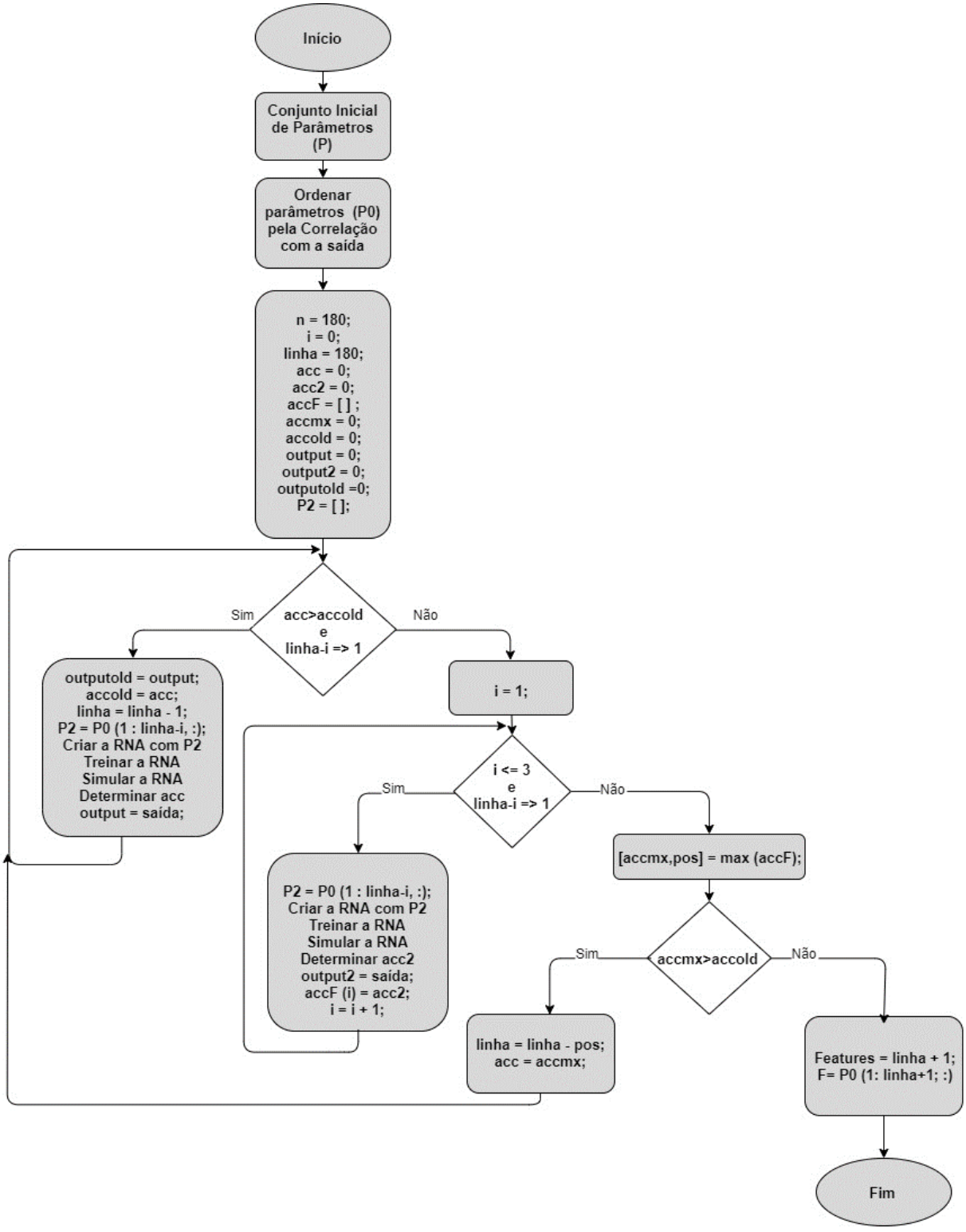

Figura 16 – Diagrama do algoritmo C2 com seleção sequencial regressiva e i=3.

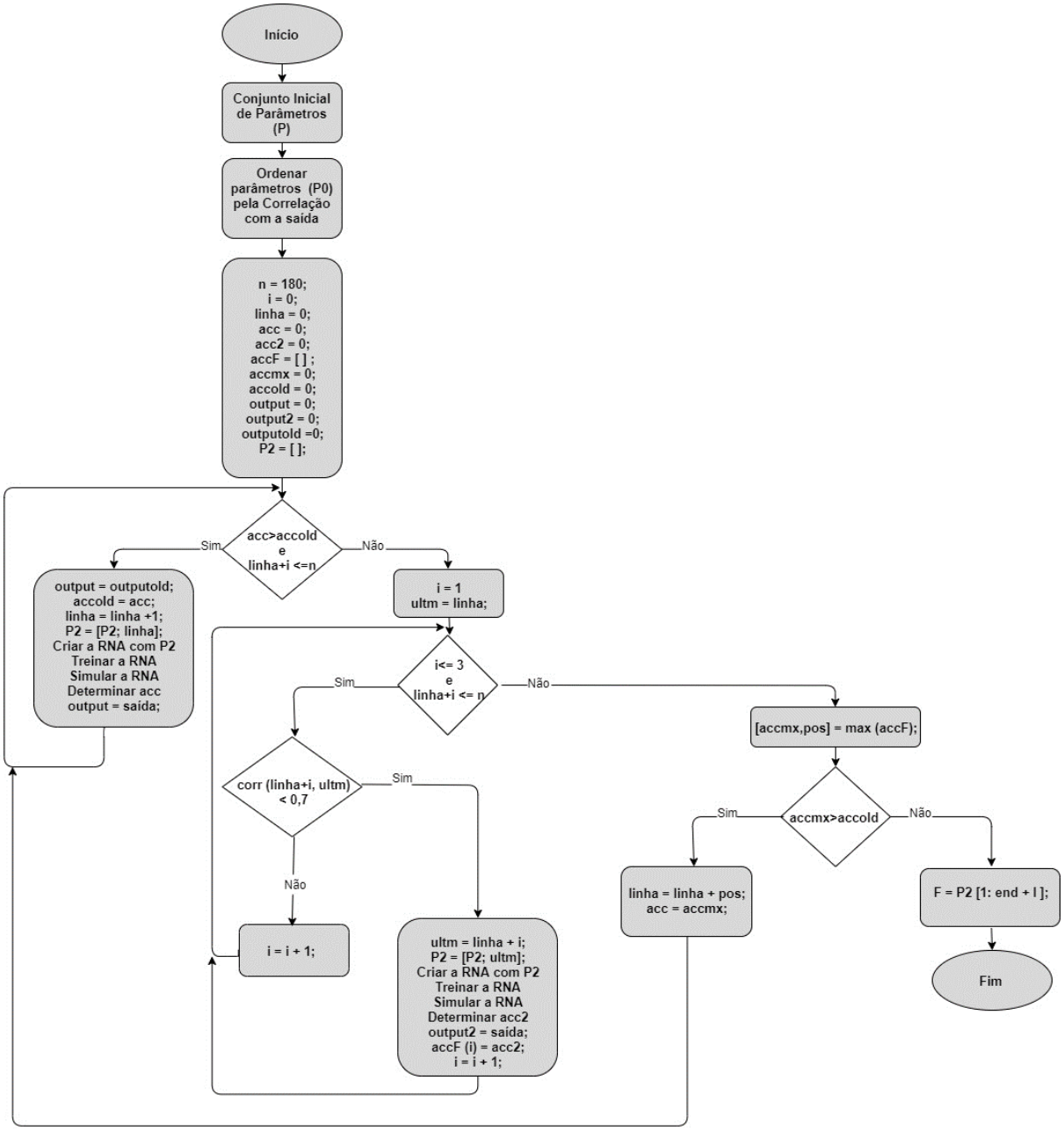

Figura 17 – Diagrama do algoritmo C3 com seleção sequencial progressiva e i=3.

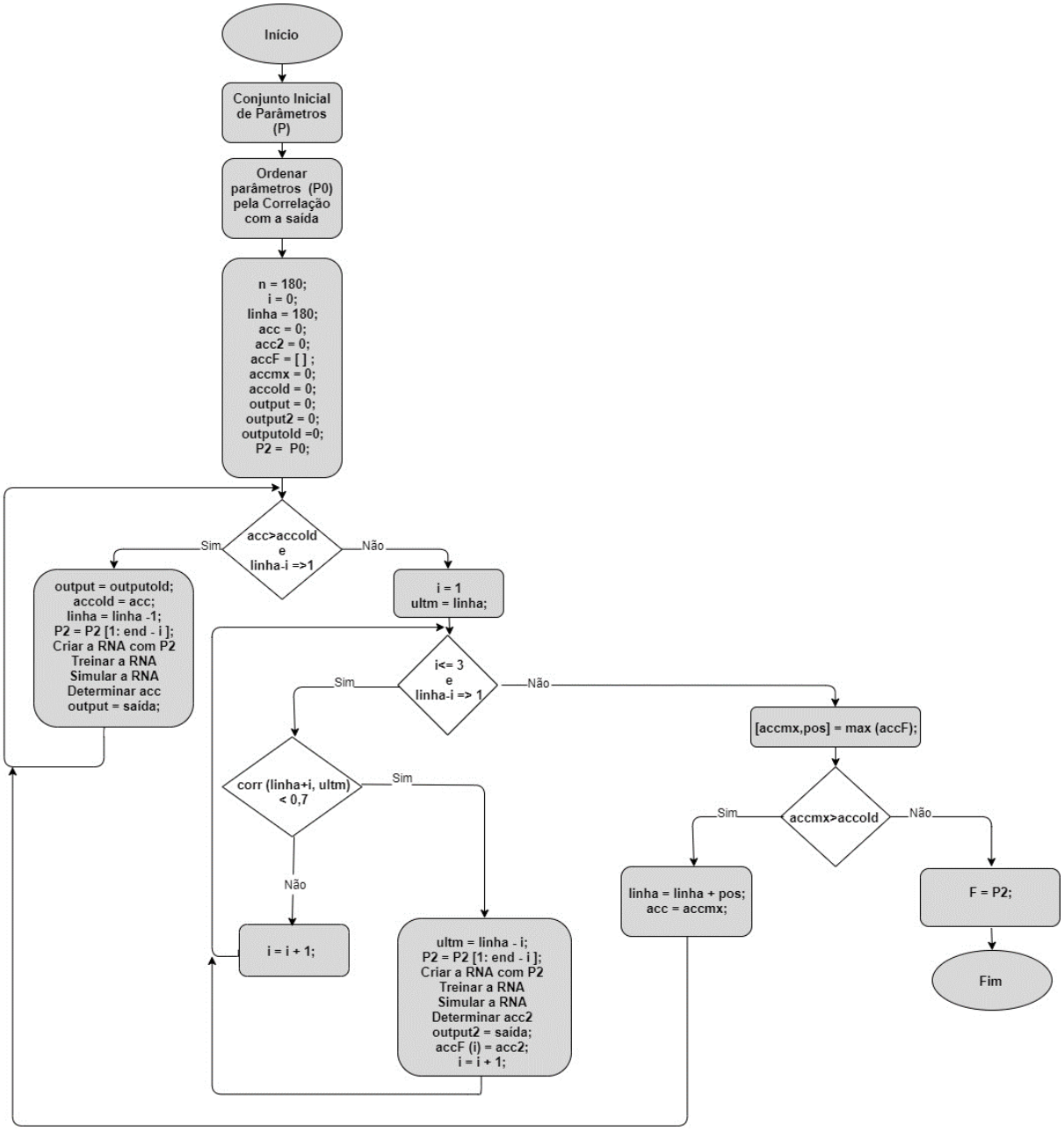

Figura 18 – Diagrama do algoritmo C3 com eliminação retrógrada e espaço de busca i=3.

7.1.2. Seleção Baseada no ReliefF

O algoritmo R1 é um método de seleção sequencial progressiva que utiliza o ReliefF para determinar a relevância dos parâmetros. Para um parâmetro ser considerado relevante é necessário que ele possua grandes valores positivos de peso. Após a aplicação do algoritmo ReliefF é devolvido um conjunto de parâmetros que será usado na entrada da rede neuronal.

O funcionamento do algoritmo R1 é descrito pela Figura 14. Este ocorre de modo semelhante ao algoritmo C1, diferenciando-se apenas no modo como ordena os parâmetros de interesse.

7.1.3. Seleção Baseada no Test t de Welch

O algoritmo TW1 é um método de seleção sequencial progressiva, onde a determinação da relevância dos parâmetros fundamenta-se no Test t de Welch, classificando os melhores atributos baseados na capacidade destes em separar as classes. Um atributo relevante é aquele que apresenta elevada significância estatística segundo o Test t de Welch (um grande valor positivo para a variável p-valor). No final é devolvido um conjunto de parâmetros que será usado na entrada da rede neuronal. O funcionamento do algoritmo TW1 é descrito pela Figura 14.

7.1.4. Seleção Baseada em Regressão Multlinear

O algoritmo RM1 é um método de seleção sequencial progressiva. A determinação da relevância dos parâmetros fundamenta-se numa análise de Regressão Multilinear. Em cada etapa, o p-valor de uma estatística do teste F é calculado para testar o desempenho do modelo. Para um atributo ser relevante é necessário que ele tenha elevada significância estatística. O algoritmo considera que o critério típico para um parâmetro entrar no modelo é um p-valor menor de 0,05 e para que um parâmetro sair é um p-valor ser maior que 0,10. Posteriormente, é devolvido um conjunto de parâmetros que é empregado na entrada da rede neuronal. A Figura 14 descreve o funcionamento de TW1.

7.2.Resultados e Discussão

O grupo (sujeitos com Paralisia das Cordas Vocais e sujeitos saudáveis) foi analisado pelos métodos de seleção de atributos C1 e C2 para selecionar subconjuntos representativos, ao longo do conjunto total de 180 parâmetros.

Foram testadas diferentes configurações de MLP e a melhor arquitetura foi escolhida usando a acurácia no treino e conjunto de validação. Quanto à direção de busca, foram aplicados a seleção progressiva e a eliminação retrógrada de atributos selecionados por C1 e C2. Quanto aos critérios de parada da busca, foram experimentadas diferentes situações: C1 busca por todo o conjunto atributos, C2 busca a máxima acurácia e posteriormente procurar até mais 3

parâmetros, além da posição desta (i=3) e C2 busca a máxima acurácia e posteriormente procurar até mais 20 parâmetros, além da posição desta (i=20).

Na Tabela 8 são apresentados os resultados são referentes à da MLP com a melhor arquitetura. Esta tabela mostra na linha superior os métodos de seleção utilizados e na primeira coluna as características correspondes a cada método experimentado. Tem-se o tipo de seleção, o número de nós nas camadas de entrada, oculta e saída [E, O, S], a função de ativação nas camadas ocultas (FAT-O) e de saída (FAT-S), a função de treino (FT), o número de parâmetros selecionados, a acurácia e a medida F determinados para o conjunto de teste resultantes dos Métodos C1 e C2. A direção de busca pode ser Seleção Progressiva (SP) ou Eliminação Retrógrada (ER).

Tabela 8 – Comparação entre métodos C1 e C2.

Entrada	**C1**	**C2 (i=3)**		**C2 (i=20)**	
Tipo de Seleção	SP	SP	ER	SP	ER
Arquitetura [E, O, S]	[170,20,1]	[3,25,1]	[174,25,1]	[3,25,1]	[159,25,1]
FAT – O	Tansig	Tansig	Tansig	Tansig	Tansig
FAT – S	Purelin	Purelin	Purelin	Purelin	Purelin
FT	Trainscg	Trainscg	Trainscg	Trainscg	Trainscg
Parâmetros	170	3	174	3	159
Acurácia	90,59	87,08	85,37	85,57	85,57
Medida F	90,64	87,02	85,75	86,95	86,95

Tabela 8 – Comparação entre métodos C1 e C2. SP – Seleção Progressiva, ER – Eliminação Retrógrada.

Como pode ser observado, os melhores resultados foram obtidos através da seleção SP com o algoritmo C1, o qual selecionou 170 parâmetros e obteve 90,59% de acurácia e 90,64% de medida F.

Na análise a partir do algoritmo C3 com i=3, são testados diferentes coeficientes de correlação entre atributos (atributo-atributo). Os resultados são apresentados na Tabela 9.

Tabela 9 – Comparação entre os métodos C3 com i=3.

Entrada	**C3 (i=3)**		**C3 (i=3)**		**C3 (i=3)**	
Correlação atb-atb <	0,7		0,8		0,9	
Tipo de Seleção	SP	ER	SP	ER	SP	ER
Arquitetura [E, O, S]	[7,25,1]	[177,25,1]	[8,25,1]	[179,25,1]	[6,25,1]	[174,25,1]
FAT – O FAT – S FT	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg
Parâmetros	7	177	8	179	6	174
Acurácia	89,55	87,36	85,57	85,29	85,18	86,46
Medida F	90,01	88,38	86,95	87,17	86,66	86,96

SP – Seleção Progressiva, ER – Eliminação Retrógrada.

Em relação ao algoritmo C3 com i=3, como pode ser observado, os melhores resultados foram obtidos com SP e correlação atributo-atributo menor que 0,7, que resultou na seleção de 7 parâmetros e obteve 89,55% de acurácia e 90,01% de medida F.

Uma outra análise é feita ao considerar o algoritmo de seleção de atributos C3 com i=20. Foram testados diferentes coeficientes de correlação entre atributos (atributo-atributo), os resultados são apresentados por meio da Tabela 10.

Em relação ao algoritmo C3 com i=20, como pode ser observado, os melhores resultados foram obtidos com SP e com correlação atributo-atributo menor que 0,7, resultando na seleção de 20 parâmetros e obteve 88,73 % de acurácia e 89,18% de medida F.

Tabela 10 – Comparação do método C3 com i=20.

Entrada	**C3 (i=20)**		**C3 (i=20)**		**C3 (i=20)**	
Correlação atb-atb <	0,7		0,8		0,9	
Tipo de Seleção	SP	ER	SP	ER	SP	ER
Arquitetura [E, O, S]	[19,25,1]	[165,25,1]	[20,25,1]	[179,25,1]	[6,25,1]	[171,25,1]
FAT – O **FAT – S** **FT**	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg	Tansig Purelin Trainscg
Parâmetros	19	165	20	179	6	171
Acurácia	88,49	87,04	88,73	85,51	88,27	87,03
Medida F	88,73	87,71	89,18	86,61	89,17	88,52

SP – Seleção Progressiva, ER – Eliminação Retrógrada.

Ao analisar todos algoritmos de seleção fundamentados em correlação, conclui-se que que C1 apresentou maior acurácia na classificação. Portanto, o método de busca que considera todo o conjunto de atributos demonstrou ser mais eficiente.

Com a intenção de comparar os algoritmos, a partir do método de busca mais eficiente, foram testadas as seguintes situações. Inicialmente, todos os 180 parâmetros foram classificados pela RNA sem que fosse aplicado método de seleção (SS) e depois foram testados os algoritmos desenvolvidos (C1, R1, TW1, RM1). Os resultados desta comparação estão apresentados na Tabela 11.

Tabela 11 – Comparação entre os algoritmos C1, R1, TW1 e RM1.

Entrada	SS	C1	R1	TW1	RM1
Tipo de Seleção	SP	SP	SP	SP	SP
Arquitetura [E, O, S]	[180,20,1]	[170,20,1]	[30,20,1]	[92,20,1]	[80,20,1]
FAT – O **FAT – S**	Tansig Purelin	Tansig Purelin	Tansig Purelin	Tansig Purelin	Tansig Purelin
FT	Trainscg	Trainscg	Trainscg	Trainscg	Trainscg
Parâmetros	180	170	30	92	80
Acurácia	83,10	90,59	92,21	90,75	90,79
Medida F	83,61	90,64	92,34	90,83	90,65

SP – Seleção Progressiva.

Depois de comparar todos os métodos de seleção desenvolvidos, os melhores resultados foram obtidos pelo algoritmo R1, o qual selecionou através da SP 30 parâmetros e obteve 92,21% de acurácia e 92,34% de medida F.

A Tabela 12 apresenta a caracterização do melhor conjunto selecionado pelo método R1. Na tabela a linha superior mostra a identificação dos tons (alto, baixo e normal) no qual as vogais (/a/, /i/ ou /u/) foram produzidas, na primeira coluna estão identificados os parâmetros em estudo.

Segundo a aplicação do método R1, em geral os coeficientes mel cepstrais foram os mais selecionados. Portanto, os parâmetros selecionados com maior frequência são o MFCC9, MFCC10, MFCC11. Enquanto que entre as vogais, a vogal selecionada com maior frequência foi a vogal /a/. E também existiram parâmetros que não foram selecionados nem mesmo uma vez como Jitt (*jitter* relativo), HNR, NHR, MFFC2.

Tabela 12- Caracterização dos parâmetros selecionados por R1.

Vogais	**/a/**	**/i/**	**/u/**
Jitt			
Jitta		A	
Shim			A
ShidB	B		
Autocor.	A		A
HNR			
NHR			
MFCC1	N		N
MFCC2			
MFCC3		N	N
MFCC4		B	
MFCC5	B		
MFCC6		B	N
MFCC7		A	A
MFCC8			N
MFCC9	B, N	N	
MFCC10	B, N	A, N	B
MFCC11	N	B	
MFCC12	A, B	B	
MFCC13			N

Vogal: /a/, /i/ e /u/. Tons: B – Baixo, N – Normal, A – Alto.

7.3.Conclusão

O método de busca mais eficiente foi o que considera como espaço de busca todos 180 parâmetros analisados, contra os demais métodos de busca com critérios de parada diferentes. Ao comparar todos os algoritmos desenvolvidos com relação a acurácia e medida F do conjunto de teste, pode-se destacar que o algoritmo R1 teve melhor desempenho. Na acurácia teve um aumento de 9 pontos percentuais e na medida F de 8 pontos percentuais, contra o método sem seleção de parâmetros. Foram selecionados 30 parâmetros por meio da seleção progressiva.

Porém, deve ser considerado que espaço de busca é relativamente pequeno, isto é, 180 parâmetros de entrada. Caso a quantidade de parâmetros de entrada fosse muito maior, o tempo de processamento necessário para selecionar os parâmetros seria demasiado longo. Portanto, neste caso é indicado o algoritmo C3 (i=3) e correlação atributo-atributo menor que 0,7. Na acurácia obteve um aumento de 7 pontos percentuais e na medida F de 6 pontos percentuais, contra o método sem seleção de parâmetros. Foram selecionados 7 parâmetros por meio da seleção progressiva.

Como conclusão final, o método R1 e C3 (i=3) com Seleção Progressiva e correlação atributo-atributo menor que 0,7 são recomendáveis para selecionar os parâmetros dos conjuntos de dados para reconhecimento de patologia vocal.

Capítulo 8: Conclusão e Trabalhos Futuros

Neste capítulo por meio de uma breve discussão serão apresentadas as considerações finais e a conclusão e os trabalhos futuros.

8.1.Conclusão Geral

O foco desta dissertação foi o estudo das novas abordagens por meio da utilização dos parâmetros *jitter*, *shimmer*, autocorreção, HNR, NHR e MFCCs aplicadas a vogais sustentadas para a classificação de patologias da voz.

As contribuições produzidas nesta dissertação podem ser divididas em duas partes: o pré-processamento do conjunto de dados que engloba a normalização, identificação e tratamento de *outliers*, descrita no Capítulo 5 e a seleção de subconjuntos representativos, descrita no Capítulo 7.

Com relação ao pré-processamento, *os* métodos de identificação de *outliers* empregados foram o Diagrama da Caixa e o Desvio Padrão. Para cada método foram experimentadas nenhuma normalização, normalização com Z-score, normalização Logarítmica e normalização da Raiz Quadrada. Para a correção dos *outliers* é o método do preenchimento que consiste em alterar o valor do ponto anômalo por um valor-limite determinado de acordo com o método BP e DP. Ao considerar a ordem das etapas de aplicação o pré-processamento pode ser diferenciado em Normalização *a priori* (ocorre como a primeira etapa a normalização, seguida da identificação e posterior tratamento das anomalias) e Normalização *a posteriori* (a normalização é a etapa final do processamento).

Em relação à normalização *a priori*, embora o método Logarítmico tenha apresentado melhor ajuste com uma curva normal, o Z-score obteve uma melhoria maior na acurácia. Em relação à identificação e correção *outliers*, ambos os métodos, BP e DP mostraram acurácia semelhante com a normalização do Z-score. O método DP1 melhora na acurácia, entre 3 e 11 pontos em porcentagem, enquanto o método BP1 melhora na acurácia, entre 3 e 13 pontos em percentagem.

Em relação à normalização *a posteriori*, o método Z-score obteve uma maior na acurácia. Em relação à identificação e correção *outliers*, ambos os métodos, BP1 e DP2

mostraram acurácia semelhante com a normalização do Z-score. O método BP1 melhora na acurácia, entre 8 e 13 pontos em porcentagem, enquanto o método DP2 melhora na acurácia, entre 2 e 13 pontos em porcentagem.

Como conclusão, o modelo que apresentou os melhores resultados de acurácia nos experimentos em relação à normalização foi o método Z-score. Em relação ao método de identificação e tratamento de *outliers*, foram os métodos, BP1 e DP1 que obtiveram acurácia semelhante com a normalização do Z-score.

Posteriormente, foram utilizadas as técnicas de Seleção de Parâmetros que ordenam os atributos segundo uma métrica de importância. Deste modo, os parâmetros relevantes são selecionados de acordo com o critério estabelecido pelos testes: Correlação, ReliefF, Test t de Welch e Regressão Multilinear. Os algoritmos desenvolvidos que utilizam a correlação são: o C1 que busca por todo o conjunto de atributos, C2 (i=3), busca, além da máxima acurácia, até 3 parâmetros a mais da posição desta e C2 (i=20), busca, além da máxima acurácia, até 20 parâmetros a mais da posição desta. Os demais algoritmos são o R1, TW1 e RM1 que procuram por todo o conjunto de atributos e baseiam-se no ReliefF, no Test t de Welch e na análise de Regressão Multilinear, respectivamente.

O método de busca mais eficiente foi o que considera como espaço de busca todos 180 parâmetros analisados, contra os demais métodos de busca com critérios de parada diferentes. Ao comparar todos os algoritmos desenvolvidos com relação a acurácia e medida F do conjunto de teste, pode-se destacar que o algoritmo R1 teve melhor desempenho. Na acurácia teve um aumento de 9 pontos percentuais e na medida F de 8 pontos percentuais, contra o método sem seleção de parâmetros. Foram selecionados 30 parâmetros por meio da seleção progressiva.

Porém, deve ser considerado que o espaço de busca é relativamente pequeno, isto é, 180 parâmetros de entrada. Caso a quantidade de parâmetros de entrada fosse muito maior, o tempo de processamento necessário para selecionar os mesmos seria demasiado longo. Portanto, neste caso é indicado o algoritmo C3 (i=3) com correlação atributo-atributo menor que 0,7. Na acurácia obteve um aumento de 7 pontos percentuais e na medida F de 6 pontos percentuais, contra o método sem seleção de parâmetros. Foram selecionados 7 parâmetros por meio da seleção progressiva.

Como conclusão, o método R1 e C3 (i=3) com Seleção Progressiva e correlação atributo-atributo menor que 0,7 são recomendáveis para selecionar os parâmetros dos conjuntos de dados para reconhecimento de patologia vocal.

Os métodos e algoritmos desenvolvidos nesta dissertação mostraram-se promissores para implementações futuras.

8.2.Trabalhos Futuros

Aos trabalhos futuros sugere-se que sejam experimentados sinais de fala contínua na classificação de vozes patológicas. Realizar estudos acerca da utilização de outros atributos, como os deltas MFCC's, a idade e gênero dos sujeitos.

Além disto, para melhorar o desempenho do modelo aplicar técnicas de redução da dimensionalidade dos dados, como o PCA (Análise de Componentes Principais).

Sugere-se ainda, com a finalidade de aumentar variabilidade nos dados, adicionar mais doenças relacionadas a Laringite Crónica, Disfonia e Paralisia das Cordas Vocais ou sujeitos com mais de uma doença.

Referências

Alcaim, A. & Oliveira, C. A. S. (2011). Fundamentos de sinais de voz e imagem. In *Interciência: PUC-Rio* (1 ed.). Rio de Janeiro.

Almeida, N. C. de. (2010). *Sistema inteligente para diagnóstico de patologias na laringe utilizando máquinas de vetor de suporte*. Universidade Federal do Rio Grande do Norte.

Alves, N. F. R. (2016). *Diagnóstico inteligente de patologias da laringe*. Instituto Politécnico de Bragança.

Apolónia, J. D. B. de B. (2018). *Seleção de atributos de dados inconsistentes*. Universidade Aberta - UAB, Lisboa.

Arjmandi, M. K. (2011). Identification of Voice Disorders Using Long-Time Features and Support Vector Machine With Different Feature Reduction Methods. *Journal of Voice*, *25*(6), 275–289.

Barbosa, S. R. T. G. da S. (2013). *Caracterização de patologias da pele por ultrassons*. Universidade de Coimbra.

Barry, W. J., & Pützer, M. (2018). *Saarbrücken Voice Database*. Retrieved from http://www.stimmdatenbank.coli.uni-saarland.de

Behlau, M., Azevedo, R., & Madazio, G. (2010). *Voz: o livro do especialista.* (2nd ed.; Revinter, Ed.). Rio de Janeiro.

Berton, L. (2011). *Caracterização de classes e detecção de outliers em redes complexas*. Universidade de São Paulo.

Bishop, C. M. (1995). *Neural Network for Pattern Recognition*. Clarendon Press, Oxford.

Bodén, M. (2001). A guide to recurrent neural networks and backpropagation. *Dallas Project*, *2*(2), 1–10.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonic-to-noide ratio of a sample sound. In P. of the institute of phonetic Sciences (Ed.), *Proceedings of the Institute of Phonetic Sciences* (Vol. 17, pp. 97–110). Retrieved from http://isip.lzu.edu.cn/Members/sunny/speech-processing/papers-on-audio-speech-language-processing/formant-extraction/Proceedings_1993.pdf

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2012). A review of feature

selection methods on synthetic data (Springer-Verlag London Limited 2012; Vol. 34). https://doi.org/10.1007/s10115-012-0487-8

Brockmann-Bauser, M. (2011). *Improving jitter and shimmer measurements in normal voices*. Medical School.

Campos, G. O. (2015). *Estudo, avaliação e comparação de técnicas de detecção não supervisionada de outliers*. Universidade de São Paulo.

Carvalho, R. T. S., Cavalcante, C. C., & Cortez, P. C. (2011). Wavelet Transform and Artificial Neural Networks Applied to Voice Disorders Identification. *IEEE*.

Cavique, L., Mendes, A. B., Funk, M., & Santos, J. M. A. (2013). A feature selection approach in the study of azorean proverbs. In *Exploring Innovative and Successful Applications of Soft Computing* (pp. 38–58). https://doi.org/10.4018/978-1-4666-4785-5.ch003

Cordeiro, H. T. (2016). *Reconhecimento de patologias da voz usando técnicas de processamento da fala* (Universidade Nova de Lisboa). Retrieved from https://run.unl.pt/bitstream/10362/19915/1/Cordeiro_2016.pdf

Cortez, P., & Neves, J. (2000). *Redes Neuronais Artificiais*. Universidade do Minho.

Dajer, M. E. (2010). *Análise de sinais de voz por padrões visuais de dinâmica vocal*. Universidade de São Paulo.

Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28*(4), 357–366.

Dibazar, A. A., Berger, T. W., & Narayanan, S. S. (2006). Pathological Voice Assessment. *International Conference of the IEEE Engineering in Medicine and Biology Society*, (0), 1–2. https://doi.org/10.1109/IEMBS.2006.259835

Dibazar, A. A., Narayanan, S., & Berger, T. W. (2002). Feature analysis for automatic detection of pathological speech. *IEEE Conference Publication*, (0), 1–2. https://doi.org/10.1109/IEMBS.2002.1134447

Doak, J. (1992). An evaluation of feature selection methodsand their application to computer security permalink. *Computer Science*, 1–90.

Fernandes, C. A. F. S. (2017). *Algoritmo do Tipo Filter-Wrapper de Seleção de Features para*

Utilização na Seleção de Genes Dissertação. Universidade de Coimbra.

Fernandes, J. F. (2018). *Determinação da autocorrelação, hnr e nhr para análise acústica vocal*. Instituto Ppolitécnico de Bragança.

Fernandes, J., Silva, L., Teixeira, F., Guedes, V., & Santos, J. (2019). Parameters for vocal acoustic analysis - Cured database. *Procedia Computer Science*, 1–8.

Fernandes, J., Teixeira, F., Guedes, V., Junior, A., & Teixeira, J. P. (2018). Harmonic to noise ratio measurement - Selection of window and length. *Procedia Computer Science*, *138*, 280–285. https://doi.org/10.1016/j.procs.2018.10.040

Ferreira, J. F. T. de S. (2012). *Tecnologia de apoio em tempo-real ao canto- relação entre parâmetros perceptivos da voz cantada com fenómenos acústicos objectivos.* Faculdade de Engenharia da Universidade do Porto.

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. *Comput Biol Med.*, *37*(4), 571–578. https://doi.org/10.1016/j.compbiomed.2006.08.008

Gibbons, J. D., & Chakraborti, S. (2003). Nonparametric Statistical Inference. In *Fundamentals of Biostatistics* (4 ed.). https://doi.org/10.5005/jp/books/10313_14

Gliozzo, J. (2016). *Network-based methods for outcome prediction in the "sample space."* Retrieved from http://arxiv.org/abs/1702.01268

Godino-Llorente, J. I., Gómez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, *53*(10), 1943–1953. https://doi.org/10.1109/TBME.2006.871883

Gonçalves, A. A. (2015). *Patologias da laringe com análise acústica vocal*. Instituto Politécnico de Bragança.

Graves, A. (2002). Supervised Sequence Labelling with Recurrent Neural Networks. *Transplantation*, *73*(5), 751–755. https://doi.org/10.1097/00007890-200203150-00016

Guedes, V. D. O. (2019). *Deep Learning aplicado a classificação de patologias da voz*. Instituto Politécnico de Bragança.

Guedes, V., Junior, A., Fernandes, J., Teixeira, F., & Teixeira, J. P. (2018). Long short term

memory on chronic aryngitis Classification. *Procedia Computer Science - Elsevier*, *128*, 250–257.

Guimarães, I. (2007). *A ciência e a arte da voz humana*. Alcoitão: Escola Superior de Saúde do Alcoitão.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. In L. P. Kaelbling (Ed.), *Journal of Machine Learning Research - JMLR* (Vol. 3). https://doi.org/10.1162/153244303322753616

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. University of Waikato.

Haykin, S. (2001). *Redes Neurais. Principios e prática.* (2 edição; L. H.- M.q.o.f, Ed.).

Haykin, S. C. (1999). Neural networks: a comprehensive foundation. In *The Knowledge Engineering Review* (1 ed., Vol. 13). https://doi.org/10.1017/S0269888998214044

Hochreiter, S., & Schmidhuber, J. (1997). *LONG SHORT-TERM MEMORY*. *9*(8), 1735–1780.

Hosseini, P. T., & Almasganj, F. (2008). Local Discriminant Wavelet Packet Basis for Voice Pathology Classif i cation. *IEEE*.

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Howell, D. C. (2010). *Statistical methods for psychology* (7 edition). Warsworth Cengage Learning.

Huche, F., & Allali, A. (2005). *A Voz - Patologia vocal de origem orgânica.* (5th ed., Vol. 3; A. Editora, Ed.).

Ittichaichareon, C., Suksri, S., & Yingthawornsuk, T. (2012). Speech recognition using MFCC. *… Conference on Computer …*, 135–138. https://doi.org/10.13140/RG.2.1.2598.3208

Iwata, S. (1972). Periodicities of pitch perturbations in normal and pathologic larynges. *The Laryngoscope*, *82*(1). https://doi.org/https://doi.org/10.1002/lary.5540820112

Kohavi, R., & John, H. G. (1997). Artificial intelligence wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-

X

Kononenko, I., Šimec, E., & Robnik-Šikonja, M. (1997). Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Applied Intelligence*, *7*(1), 39–55. https://doi.org/10.1023/A:1008280620621

Kumar, V., Abbas, A. K., Fausto, N., & Aster, J. C. (2010). *Robbins and Cotran Patologia – bases patológicas das doenças.* (8 ed). Rio de Janeiro: Brasil: Elsevier Editora Ltda.

Kumar, V., & Minz, S. (2014). Feature selection: a literature review. *International Journal of Industrial and Systems Engineering*, *4*(3), 211–229. https://doi.org/10.1504/IJISE.2013.052279

Lee, H. D. (2005). *Seleção de atributos importantes para a extração de conhecimento de bases de dados*. Universidade de São Paulo.

Lee, J. W., Kang, H. G., Choi, J. Y., & Y. I. Son. (2013). An Investigation of Vocal Tract Characteristics for Acoustic Discrimination of Pathological Voices. *BioMed Research International*, 11.

Lieberman, P., & Affiliations, V. (1963). Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges. *The Journal of the Acoustical Society of America*, *35*, 344. https://doi.org/https://doi.org/10.1121/1.1918465

Lima, L. F. M., Maroldi, A. M., Silva, D. V. O. da, Hayashi, C. R. M., & Hayashi, M. C. P. I. (2017). Métricas científicas em estudos bibliométricos: detecção de outliers para dados univariados. In *Em Questão* (Vol. 23). https://doi.org/10.19132/1808-5245230.254-273

Logan, B. (2000). *Mel frequency cepstral coefficients for music modeling*. 13.

Lopes, J. M. D. S. (2008). *Ambiente de análise robusta dos principais parâmetros qualitativos da voz* (Faculdade de Engenharia da Universidade do Porto). Retrieved from http://repositorio-aberto.up.pt/bitstream/10216/58997/1/000136648.pdf

Machado, A. F. (2013). *Conversão de voz inter-linguística*. Universidade de São Paulo.

May, R., Dandy, G., & Maier, H. (2011). Review of input variable selection methods for artificial neural networks (University ofAdelaide). https://doi.org/10.5772/16004

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, *32*(1), 12–16. https://doi.org/10.1080/00031305.1978.10479236

MEEI. (1994). *Voice disorders database, (Version 1.03 cd-rom)*. Retrieved from https://www.masseyeandear.org/about

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: an empirical comparison between svm and ann. *Expert Systems with Applications*, *40*(2), 621–633. https://doi.org/10.1016/j.eswa.2012.07.059

Muda, L., Begam, M., & Elamvazuthi, I. (2010). *Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques*. *2*(3), 138–143. Retrieved from http://arxiv.org/abs/1003.4083

Oppenheim, A. V, Schafer, R. W., & Buck, J. R. (1999). *Discrete-Time Signal Processing - Second Edition.pdf* (2nd ed.). New Jersey.

Pino, F. A. (2014). A questão da não normalidade: uma revisão. *Rev. de Economia Agrícola*, *61*(2), 17–33.

Rabiner, L. R. & Schafer, R. W. (2009). Theory and application of digital speech processing. In *Bulletin of the Seismological Society of America* (Preliminar). https://doi.org/10.1785/0120190038

Rosa, M. D. O., Pereira, J. C., & Grellet, M. (2000). Adaptive Estimation of Residue Signal for Voice Pathology Pathology Diagnosis. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *47*(1), 96–104.

Scalassara, P. R., Dajer, E., Maciel, C. D., Pereira, C., & Guido, R. C. (2009). Relative entropy measures applied to healthy and pathological voice characterization. *Science Direct*, *207*, 95–108. https://doi.org/10.1016/j.amc.2007.10.068

Seo, S. M. S. (2006). A review and comparison of methods for detecting outliers in univariate data sets (University of Pittsburgh). Retrieved from http://d-scholarship.pitt.edu/7948/

Silva, A. C. P. da, Esteves, S. S., Telma Feliciano, Centro, Freitas, S. V., & Sousa, C. A. (2017). Paralisia das cordas vocais - 8 anos de experiência no Centro Hospitalar do Porto. *Revista Portuguesa de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço*, *54*, 169–173. Retrieved from url: https://www.journalsporl.com/ index.php/sporl/article/view/363.

Song, X., & Lu, H. (2017). Multilinear regression for embedded feature selection with application to fMRI analysis. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*,

(2), 2562–2568.

Souza, C. (2009). Análise de Poder Discriminativo Através de Curvas ROC. Retrieved June 13, 2019, from Análise de Poder Discriminativo Através de Curvas ROC

Teixeira, J. P., Fernandes, J., Teixeira, F., & Fernandes, P. O. (2018). Acoustic analysis of chronic laryngitis - Statistical analysis of sustained speech parameters. *BIOSIGNALS 2018 - 11th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 11th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2018*, *4*(Biostec), 168–175. https://doi.org/10.5220/0006586301680175

Teixeira, J. P., & Fernandes, P. O. (2015). Acoustic analysis of vocal dysphonia. *Procedia Computer Science*, *64*, 466–473. https://doi.org/10.1016/j.procs.2015.08.544

Teixeira, J. P., Fernandes, P. O., & Alves, N. (2017). Vocal Acoustic Analysis - Classification of Dysphonic Voices with Artificial Neural Networks. *Procedia Computer Science*, *121*, 19–26. https://doi.org/10.1016/j.procs.2017.11.004

Teixeira, J. P., Ferreira, D. B., & Carneiro, S. M. (2011). *Análise acústica vocal - Determinação do jitter e shimmer para diagnóstico de patologias da fala* (Escola Superior de Tecnologia e Gestão). Retrieved from https://bibliotecadigital.ipb.pt/bitstream/10198/7282/1/artigo_publicado.pdf

Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of jitter and shimmer measurements. *Procedia Technology*, *16*, 1190–1199. https://doi.org/10.1016/j.protcy.2014.10.134

Teixeira, J. P., & Gonçalves, A. (2016). Algorithm for Jitter and Shimmer Measurement in Pathologic Voices. *Procedia Computer Science*, Vol. 100, pp. 271–279. https://doi.org/10.1016/j.procs.2016.09.155

Triola, M. F. (2017). Introdução à estatística. In *Elementary Statistics* (12 ed). Rio de Janeiro: Pearson Education INC.

Tukey, J. W. (1970). Exploratory data analysis. In *SAGE Publications* (Vol. 1). Addison-Wesley Publishing Company.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, *85*, 189–203. https://doi.org/10.1016/j.jbi.2018.07.014

Ventura, B. V. O. C. (2013). *Uso de algoritmos de aprendizagem de máquina e estratégias de seleção de atributos para otimizar a identificação de ceratocone a partir de propriedade biomecânicas da córnea* (Vol. 1). Universidade Federal de Alagoas (UFAL).

Welch, A. B. L. (1947). The Generalization of Student ' s ' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*(12), 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Williams, R. B. (1984). *Introduction to Statistics*. Macmillan International Higher Education.

Xu, M., Duan, L. Y., Cai, J., Chia, L. T., Xu, C., & Tian, Q. (2004). HMM-based audio keyword generation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3333*, 566–574. https://doi.org/10.1007/978-3-540-30543-9_71

Yu, D., & Deng, L. (2015). Automatic speech recognition. A deep learning approach. In Springer (Ed.), *Signals and Communication Technology*. https://doi.org/10.1109/ICETA.2016.7802064