

Universidade Estadual Paulista
Instituto de Biociências, Letras e Ciências Exatas
Departamento de Ciência da Computação e Estatística

Thales Cândido Rosa

Detecção e caracterização computacional de vozes
afetadas patologicamente

São José do Rio Preto - SP

2023

Thales Cândido Rosa

Detecção e caracterização computacional de vozes
afetadas patologicamente

Monografia apresentada ao Programa de
graduação em Ciência da Computação da
UNESP para obtenção do título de Bacharel.

Orientador: Prof. Dr. Rodrigo Capobianco Guido

Prof^a. Dr^a. Renata Spolon Lobato

Prof. Dr. Wallace Correa de Oliveira Casaca

São José do Rio Preto - SP

2023

R788d Rosa, Thales Cândido
Detecção e caracterização computacional de vozes afetadas
patologicamente / Thales Cândido Rosa. -- São José do Rio Preto,
2023
66 p.

Trabalho de conclusão de curso (Bacharelado - Ciência da
Computação) - Universidade Estadual Paulista (Unesp), Instituto de
Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Rodrigo Capobianco Guido

1. Processamento de sinais. 2. Vozes afetadas patologicamente. 3.
Acústica. 4. Escala Mel. 5. Escala Bark. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de
Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Thales Cândido Rosa

Detecção e caracterização computacional de
vozes afetadas patologicamente

Trabalho de Conclusão de Curso (TCC)
apresentado como parte dos requisitos
para obtenção do título de Bacharel em
Ciência da Computação, junto ao Conse-
lho de Curso de Bacharelado em Ciência
da Computação, do Instituto de Bi-
ociências, Letras e Ciências Exatas da
Universidade Estadual Paulista "Júlio de
Mesquita Filho", Câmpus de São José do
Rio Preto.

Comissão examinadora

Prof. Dr. Rodrigo Capobianco Guido
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof^a. Dr^a. Renata Spolon Lobato
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Wallace Correa de Oliveira Casaca
UNESP – Câmpus de São José do Rio Preto

São José do Rio Preto - SP

10 de janeiro de 2023

Este trabalho de pesquisa só foi possível
através do apoio e suporte da minha família.
Dedico este trabalho a ela.

Agradecimentos

Agradeço primeiramente a Deus por ter me guardado, ajudado e sustentado até aqui.

Agradeço aos meus pais, meu irmão e minha namorada por sempre terem me apoiado.

Agradeço ao meu orientador Prof. Dr. Rodrigo Capobianco Guido pelos ensinamentos fornecidos, pelo apoio e pela contribuição para minha formação acadêmica e pessoal.

Agradeço, também, à Universidade Estadual Paulista "Júlio de Mesquita Filho", pelo curso de graduação ministrado por professores de excelência.

“Tudo é aliado do homem que sabe querer.”

Machado de Assis

Resumo

ROSA, T. C. *Detecção e caracterização computacional de vozes afetadas patologicamente*. 2023. 66p. TCC UNESP 2023.

Atualmente, a detecção precoce de patologias na voz é importante e necessária, pois pode ser decisiva no sucesso do tratamento e recuperação de pacientes com tais patologias. Dessa forma, este trabalho busca construir um modelo computacional de detecção não-invasiva de vozes afetadas pelas patologias laringite, paralisia das cordas vocais e disfonia orgânica utilizando algoritmos de aprendizado de máquina a fim de detectar e caracterizar tais patologias com o auxílio de uma base de dados contendo gravações de vozes patológicas e normais. Foram realizados testes com 1308 gravações provenientes da *Saarbrücken Voice Database* que implicaram em uma acurácia de até 86,2%.

Palavras-chave: Processamento de sinais. Vozes afetadas patologicamente. Acústica. Escala *Mel*. Escala *Bark*.

Abstract

ROSA, T. C. *Computational portrayal and detection of pathologically affected voices*. 2023. 66p. TCC UNESP 2023.

Nowadays, the early detection of voice pathologies is crucial and decisive on the successful patient treatment and recovery from such diseases. Nevertheless, this work aims to build a computational model for a non-invasive detection of voices affected by laryngitis, vocal cord paralysis and organic dysphonia with use of machine learning algorithms as well as voice databases containing healthy and pathological speech recordings. Tests were carried out using 1308 recordings from Saarbrücken Voice Database which resulted in an accuracy of up to 86,2%.

Keywords: Signal processing. Pathologically affected voices. Acoustics. Mel scale. Bark scale.

Lista de Figuras

Figura 2.1 - Ilustração do processo de quantização de um sinal.	20
Figura 2.2 - Representação gráfica do algoritmo KNN	28
Figura 2.3 - Representação gráfica de uma árvore de decisão.	29
Figura 2.4 - Representação gráfica de uma rede neural artificial	31
Figura 2.5 - Ilustração representando curvas ROC e suas respectivas ASC.	36
Figura 3.1 - Sinal de áudio de uma amostra contendo a amplitude e o tempo em segundos.	42
Figura 3.2 - Extração dos MFCCs e BFCCs para compor o vetor de características X	43
Figura 3.3 - Aplicação do algoritmo de aprendizado de máquina automatizado para o vetor de características X e o vetor y que contém as classificações de cada amostra.	44
Figura 3.4 - Validação cruzada k -fold dos classificadores obtidos pelo algoritmo de aprendizado de máquina automatizado.	45
Figura 4.1 - Curvas ROC obtidas no conjunto de teste I.	48
Figura 4.2 - Curvas ROC obtidas no conjunto de teste II.	50
Figura 4.3 - Curvas ROC obtidas no conjunto de teste III.	52
Figura 4.4 - Curvas ROC obtidas no conjunto de teste IV.	54
Figura 4.5 - Curvas ROC obtidas no conjunto de teste V com laringite.	56
Figura 4.6 - Curvas ROC obtidas no conjunto de teste V com disfonia orgânica.	57
Figura 4.7 - Curvas ROC obtidas no conjunto de teste V com paralisia das cordas vocais.	58

Lista de Tabelas

Tabela 2.1 - Bandas críticas da escala <i>Bark</i>	24
Tabela 2.2 - Representação de uma matriz de confusão	34
Tabela 4.1 - Matriz de confusão do melhor caso do teste do conjunto I.	47
Tabela 4.2 - Matriz de confusão do pior caso do teste do conjunto I.	47
Tabela 4.3 - Matriz de confusão do melhor caso do teste do conjunto II.	49
Tabela 4.4 - Matriz de confusão do pior caso do teste do conjunto II.	49
Tabela 4.5 - Matriz de confusão do melhor caso do teste do conjunto III.	51
Tabela 4.6 - Matriz de confusão do pior caso do teste do conjunto III.	51
Tabela 4.7 - Matriz de confusão do melhor caso do teste do conjunto IV.	53
Tabela 4.8 - Matriz de confusão do pior caso do teste do conjunto IV.	53
Tabela 4.9 - Matriz de confusão do melhor caso do teste do conjunto V com laringite.	55
Tabela 4.10 -Matriz de confusão do pior caso do teste do conjunto V com laringite.	55
Tabela 4.11 -Matriz de confusão do melhor caso do teste do conjunto V com disfonia orgânica.	56
Tabela 4.12 -Matriz de confusão do pior caso do teste do conjunto V com disfonia orgânica.	56
Tabela 4.13 -Matriz de confusão do melhor caso do teste do conjunto V com paralisia das cordas vocais.	58
Tabela 4.14 -Matriz de confusão do pior caso do teste do conjunto V com paralisia das cordas vocais.	58

Lista de Abreviaturas

APQ	<i>Amplitude Perturbation Quotient</i>
ASC	Área Sob a Curva
AVPD	<i>Arabic Voice Pathology Database</i>
BFCC	<i>Bark Frequency Cepstral Coefficients</i>
HUPA	Hospital Universitário Príncipe de Astúrias
KNN	<i>K-Nearest Neighbors</i>
MEEI	<i>Massachusetts Eye and Ear Infirmary</i>
MDVP	<i>Multidimensional Voice Program</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
OSELM	<i>Online Sequential Extreme Learning Machine</i>
PCA	<i>Principal Component Analysis</i>
PPQ	<i>Pitch Perturbation Quotient</i>
RAP	<i>Relative Average Perturbation</i>
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
ROC	<i>Receiver Operating Characteristic</i>
SVD	<i>Saarbruecken Voice Database</i>
SVM	<i>Support Vector Machine</i>
WAVE	<i>Waveform Audio File Format</i>

Sumário

1	Introdução	16
1.1	Considerações iniciais	16
1.2	Objetivos	17
1.3	Organização do trabalho	17
2	Fundamentação teórica	18
2.1	Anomalias na fala	18
2.1.1	Laringite	19
2.1.2	Disfonia orgânica e funcional	19
2.1.3	Paralisia das cordas vocais	20
2.2	Filtros digitais e convolução	20
2.3	Transformada de <i>Fourier</i>	21
2.4	Transformada discreta de cosseno	22
2.5	Escala <i>Mel</i>	22
2.5.1	Coeficientes cepstrais de <i>Mel</i>	23
2.6	Escala <i>Bark</i>	23
2.6.1	Coeficientes cepstrais de <i>Bark</i>	23
2.7	Coeficientes de perturbação	25
2.8	Classificação	26
2.8.1	Análise de componentes principais (PCA)	27
2.8.2	K vizinhos mais próximos (KNN)	27
2.8.3	Árvore de decisão	28
2.8.4	<i>Random Forest</i>	29

2.8.5	<i>Ada Boost</i>	30
2.8.6	<i>Extra Trees</i>	30
2.8.7	Máquina de Vetores de Suporte (SVM)	30
2.8.8	Redes neurais artificiais	31
2.8.9	Redes neurais convolucionais	32
2.8.10	Aprendizado de máquina automatizado	33
2.8.11	Validação cruzada	33
2.8.12	Matriz de confusão e métricas	34
2.8.13	Curva ROC e área sob a curva	35
2.9	Trabalhos relacionados	36
3	Metodologia	41
3.1	Obtenção dos dados	41
3.2	Extração das características	42
3.3	Construção do classificador	44
3.4	Teste do classificador e análise dos resultados	45
4	Testes e Resultados	46
4.1	Testes do conjunto I - Coeficientes de perturbação	46
4.2	Testes do conjunto II - Coeficientes de perturbação e MFCCs	48
4.3	Testes do conjunto III - Coeficientes de perturbação e BFCCs	50
4.4	Testes do conjunto IV - Coeficientes de perturbação, MFCCs e BFCCs	52
4.5	Testes do conjunto V - Coeficientes de perturbação, MFCCs e BFCCs com clas- sificadores individuais	54
4.5.1	Laringite	54
4.5.2	Disfonia orgânica	55
4.5.3	Paralisia das cordas vocais	57
4.6	Conclusão	59
5	Conclusões e Trabalhos Futuros	60

Capítulo 1

Introdução

1.1 Considerações iniciais

A fala é de caráter primário na comunicação entre os indivíduos por ser uma habilidade básica imprescindível ao comportamento humano. Anomalias na fala são doenças que afetam a capacidade de comunicação efetiva de uma pessoa e podem causar fortes consequências psicológicas e emocionais [1].

A detecção precoce de doenças desse gênero é de grande importância para permitir o sucesso no tratamento e recuperação desses pacientes [1]. Ademais, processos computadorizados de detecção de patologias na voz podem permitir uma alternativa simples e acessível aos métodos tradicionais de diagnóstico [19].

Além disso, técnicas de inteligência artificial e aprendizado de máquina podem facilitar a identificação de vozes afetadas patologicamente por serem facilmente implementáveis e adequadas para o pré-diagnóstico de tais enfermidades [1].

1.2 Objetivos

Motivado pelas considerações anteriores, este trabalho tem a finalidade de construir um modelo computacional de detecção não-invasiva de vozes afetadas patologicamente utilizando ferramentas de aprendizado de máquina e uma base de dados de sinais pública contendo sinais de vozes previamente laudados, além de avaliar o desempenho do modelo criado usando técnicas e métricas para avaliação de classificadores.

1.3 Organização do trabalho

O texto vindouro do presente trabalho está organizado da seguinte forma:

- No Capítulo 2 apresentam-se alguns conceitos e teorias relacionados ao trabalho desenvolvido, bem como trabalhos publicados que envolvem a área de detecção de vozes afetadas patologicamente, evidenciando como são diversas as maneiras de se realizar essa tarefa.
- No Capítulo 3 apresenta-se, com detalhes, todo o desenvolvimento do trabalho proposto e a maneira que os conceitos discutidos no capítulo anterior foram empregados.
- No Capítulo 4 relatam-se todos os resultados obtidos no trabalho, a partir dos testes de reconhecimento que foram realizados.
- No Capítulo 5 apresentam-se as conclusões sobre o trabalho, bem como propostas para pesquisas futuras.

Capítulo 2

Fundamentação teórica

2.1 Anomalias na fala

Anomalias na fala podem ser causadas pela presença de infecções epiteliais, mudanças sistêmicas, estresse mecânico, irritações superficiais, além de alterações epiteliais, neurológicas ou musculares [43]. Essas patologias podem causar deformidade e prejudicar a mobilidade das pregas vocais, causando vibrações irregulares que modificam a voz e fazendo com que ela soe tensa, áspera, fraca ou ofegante [43].

Até os dias de hoje, existem duas abordagens para a detecção de patologias na voz: subjetiva e objetiva. Aquela consiste na análise perceptual auditiva e exame visual das pregas vocais, tipicamente com base na laringo-estroboscopia [19]. Embora ambas abordagens possuam diferentes métodos de diagnóstico, estão sujeitas ao viés subjetivo de quem realiza as análises e exigem a presença física do paciente, o que pode ser um problema dependendo da patologia e o seu grau de infectividade [19].

Por outro lado, as abordagens objetivas são baseadas em análises computadorizadas dos sinais acústicos emitidos pela voz do paciente, permitindo identificar perturbações imperceptíveis ao ouvido humano [27]. Além disso, tais abordagens possuem um menor custo e podem ser usadas a partir de dispositivos como *smartphones* que são de fácil acesso para a população geral [27].

Este trabalho buscou identificar três diferentes anomalias na fala causadas por suas respectivas patologias, sendo elas laringite, disfonia orgânica e paralisia das cordas vocais.

2.1.1 Laringite

Trata-se da inflamação da laringe que tem como sintoma mais comum a rouquidão, além de febre, dificuldade de engolir, dor no pescoço e tosse [45]. O tipo mais comum é a aguda, a qual os sintomas têm duração de até três semanas, sendo a crônica com sintomas se estendendo por mais de três semanas [45].

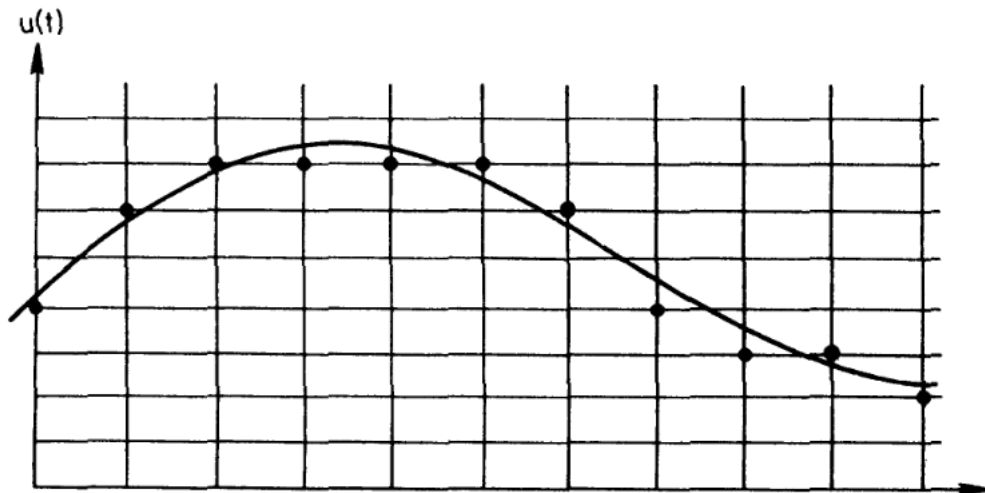
2.1.2 Disfonia orgânica e funcional

Disfonia, de forma geral, é a dificuldade de produzir sons vocais e pode ser subdividida em duas categorias: caso haja uma causa anatômica identificável, a disfonia é classificada como orgânica, caso contrário é classificada como funcional [24].

A disfonia orgânica pode ser causada por uma série de doenças que afetam o sistema neurológico ou os músculos laríngeos e as principais alterações na voz são rouquidão, aspereza, tensão vocal, soprosidade e alterações da frequência e intensidade vocal [24].

A disfonia funcional, por sua vez, pode ser causada por disfunção das pregas vocais ou por motivos psicológicos e as principais alterações na voz são rouquidão, soprosidade, fadiga vocal, redução do tempo de fonação e alterações de timbre [24].

Figura 2.1 – Ilustração do processo de quantização de um sinal.



Fonte: Extraído de: [18]

2.1.3 Paralisia das cordas vocais

Também conhecida como paralisia do nervo laríngeo recorrente, é o ferimento de um ou ambos nervos laríngeos recorrentes que podem causar rouquidão, problemas na respiração e dificuldade de ingerir líquidos devido à mobilidade debilitada dos músculos laríngeos [44].

2.2 Filtros digitais e convolução

Filtros digitais são ferramentas amplamente usadas em diferentes tipos aplicações por possibilitarem a quantização de sinais contínuos tipicamente analógicos, isto é, transformar sinais contínuos em uma sequência de valores igualmente espaçados para permitir o processamento digital desses sinais [18], confirme a Figura 2.1.

Hamming define convolução de dados com coeficientes como sendo o processo de utilizar um filtro não-recursivo y_n em uma sequência u_n de n valores igualmente espaçados, onde c_k um

conjunto de coeficientes que são constantes do filtro representado pela equação 2.1.

$$y_n = \sum_{k=-\infty}^{\infty} c_k u_{n-k} \quad (2.1)$$

É importante observar que na prática os valores do filtro deverão ser finitos para que seja possível filtrar um sinal e que usualmente o número valores não nulos de c_k é razoavelmente menor do que os número de valores em u_n [18].

2.3 Transformada de *Fourier*

A transformada de *Fourier* é uma transformada integral que decompõe uma função temporal em componentes de frequência, que correspondem a um valor complexo da frequência original [26].

O seu uso em sinais que variam ao longo do tempo, como sinais de voz, pode ser otimizado ao dividir o sinal em pequenas janelas de tempo que, em geral, se alteram pouco. A análise do sinal contido nessas janelas de tempo é conhecida como análise de curto-tempo [26].

A transformada discreta de Fourier (DFT) estima a transformada de Fourier para uma amostra finita de valores, como um sinal digital de áudio [17]. A transformada discreta de Fourier é definida pela fórmula 2.2 onde N é o tamanho da amostra e K é o número de coeficientes.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}kn}, \quad k = 1, 2, \dots, K \quad (2.2)$$

Para obter o periodograma, basta remover a parte complexa do resultado obtido em 2.2 como descrito em 2.3. O periodograma representa uma estimativa da densidade espectral do sinal e é útil para obter características de frequência e amplitude de um sinal em uma janela de tempo [39].

$$P_k = \frac{1}{N} |X_k|^2 \quad (2.3)$$

2.4 Transformada discreta de cosseno

Ahmed *et al* [2] define a transformada discreta de cosseno como uma transformada matemática que expressa uma sequência finita de pontos em termos da soma de uma função cosseno que oscila em diferentes frequências.

A transformada é definida formalmente como uma função reversível $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$, é representada pela equação 2.4, onde N é um número natural, x_n representa uma sequência de pontos onde x_n é par quando $n = -\frac{1}{2}$ ou $n = N - \frac{1}{2}$ e X_n representa a transformada discreta de cosseno resultante e X_k é par quando $k = 0$ e ímpar quando $k = N$.

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right] \quad \text{para } k = 0, 1, \dots, N-1 \quad (2.4)$$

2.5 Escala *Mel*

A escala Mel é uma escala perceptual de tons definida experimentalmente por Stevens *et al* a fim de medir o tom de um sinal sonoro de maneira subjetiva [38]. Essa escala matemática foi dividida em 1000 unidades subjetivas denominadas *Mels* e a conversão da frequência para a escala *Mel* é dada pela equação 2.5 onde f é a frequência em *Hertz* [38].

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.5)$$

2.5.1 Coeficientes cepstrais de *Mel*

Os coeficientes cepstrais de *Mel* permitem obter características cepstrais locais do sinal analisado [20]. Para isso, é usada uma técnica de janelamento com a transformada de *Fourier*, descrita em 2.3. Depois, obtém-se o periodograma descrito em 2.3, transforma-se as frequências em *Hertz* para a escala *Mel* e por fim, aplica-se a transformada discreta de cosseno, definida pela equação 2.4, a fim de converter o periodograma para o domínio do tempo, obtendo os coeficientes cepstrais de *Mel* [20].

2.6 Escala *Bark*

Trata-se de uma escala psicoacústica proposta por Zwicker em 1961 buscando medir a intensidade sonora perceptível ao ouvido humano dividindo o espectro sonoro audível em 24 faixas de frequência determinadas experimentalmente cuja distância entre as bandas são perceptualmente iguais [47].

Assim como a escala *Mel*, a escala *Bark* pode ser usada para extrair características foneticamente importantes por captar as faixas de frequência audíveis ao ouvido humano [20].

As bandas críticas de *Bark* e suas respectivas frequências centrais, de recorte e suas larguras estão definidas na Tabela 2.1.

2.6.1 Coeficientes cepstrais de *Bark*

Os coeficientes cepstrais de *Bark* são obtidos de maneira semelhante aos de *Mel*, descrito em 2.5.1 e também permitem obter características cepstrais locais do sinal analisado, porém

Tabela 2.1 – Bandas críticas da escala *Bark*

Número	Frequência central (Hz)	Frequência de recorte (Hz)	Largura da banda (Hz)
		20	
1	50	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150
9	1000	1080	160
10	1170	1270	190
11	1370	1480	210
12	1600	1720	240
13	1850	2000	280
14	2150	2320	320
15	2500	2700	380
16	2900	3150	450
17	3400	3700	550
18	4000	4400	700
19	4800	5300	900
20	5800	6400	1100
21	7000	7700	1300
22	8500	9500	1800
23	10500	12000	2500
24	13500	15500	3500

Fonte: Extraído e adaptado de: [47]

transforma-se as frequências do periodograma em *Hertz* para a escala *Bark*.

2.7 Coeficientes de perturbação

Neste trabalho, foram utilizados coeficientes de perturbação apresentados, definidos por Teixeira e Gonçalves [41], onde N é número de períodos, T_i é o comprimento do período glotal extraído em i e A_i é a amplitude extraída em i , tal que $i \in \mathbb{N}$ e $0 \leq i \leq N - 1$.

Jitter absoluto (*jitta*): é a diferença média absoluta entre períodos consecutivos representado na equação 2.6.

$$jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (2.6)$$

Jitter local (*jitt*): é a diferença média absoluta entre períodos consecutivos dividido pela média dos períodos representado na equação 2.7.

$$jitt = \frac{\frac{N}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (2.7)$$

Perturbação relativa média (*rap*): é a diferença média absoluta entre um período e a média de seus dois vizinhos dividido pela média dos períodos, representado na equação 2.8.

$$rap = \frac{\frac{N}{N-1} \sum_{i=1}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (2.8)$$

Quociente de perturbação de período de cinco pontos (*ppq5*): é a diferença média absoluta entre um período e os seus quatro vizinhos mais próximos dividido pela média dos períodos representado na equação 2.9.

$$rap = \frac{\frac{N}{N-1} \sum_{i=2}^{N-2} \left| T_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (2.9)$$

Shimmer médio local em decibéis (*shimdb*): é a média absoluta do logaritmo base 10 da

razão entre as amplitudes consecutivas multiplicado por 20, conforme representado na equação 2.10.

$$shimdb = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \times \log_{10} \frac{A_{i+1}}{A_i} \right| \quad (2.10)$$

Shimmer local (*shim*): é a diferença média absoluta entre as amplitudes dos períodos consecutivos dividido pela amplitude média do sinal representado na equação 2.11.

$$shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (2.11)$$

Quociente de perturbação de amplitude de três pontos (*apq3*): é a diferença média absoluta entre a amplitude de um período e a média das amplitudes de seus vizinhos dividido pela amplitude média do sinal, representado na equação 2.12.

$$apq3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| A_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (2.12)$$

Quociente de perturbação de amplitude de cinco pontos (*apq5*): é a diferença média absoluta entre a amplitude de um período e a média das amplitudes de seus dois vizinhos mais próximos dividido pela amplitude média do sinal, representado na equação 2.13.

$$apq5 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| A_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (2.13)$$

2.8 Classificação

Classificação é o trabalho de se identificar a qual grupo previamente definido um determinado dado pertence [42]. A classificação pode ser dividida nas etapas de treinamento, teste e validação. A etapa de treinamento consiste em treinar um modelo de classificação usando um conjunto de dados. A etapa de teste utiliza o modelo já treinado para classificar os mesmos

dados usados na etapa anterior. Em seguida, a etapa de validação utiliza os resultados do teste com medidas estatísticas para analisar o modelo e assim, ajustar os seus parâmetros para buscar uma melhor acurácia [42].

Nas seções de 2.8.2 a 2.8.12, alguns conceitos relacionados à classificação são explorados.

2.8.1 Análise de componentes principais (PCA)

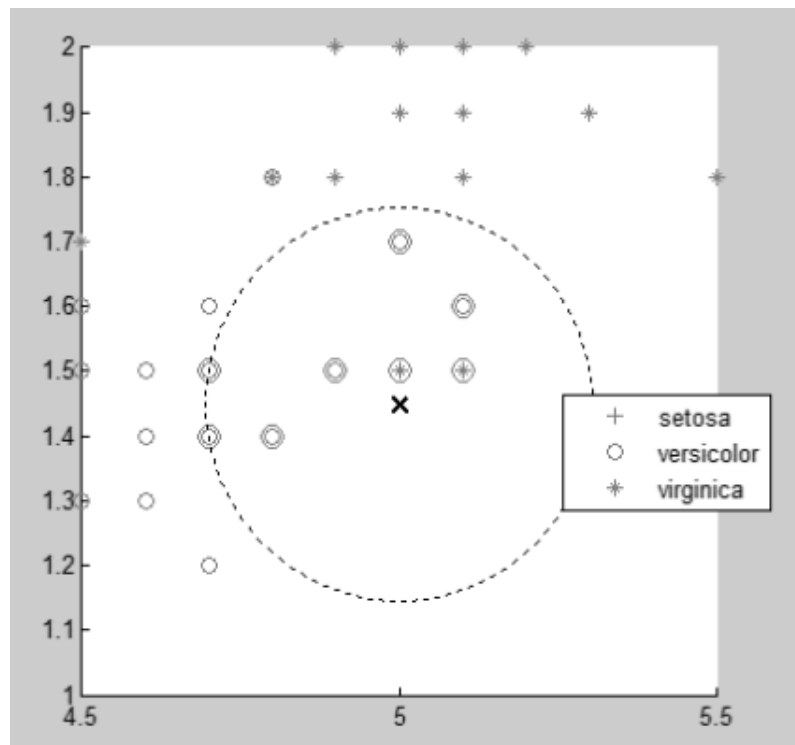
Trata-se de uma técnica estatística comumente utilizada para analisar amplas bases de dados contendo um grande número de dimensões a fim de facilitar a interpretação dos dados obtidos preservando as informações contidas e permitindo a visualização de dados multidimensionais [25]. Isso é feito criando um novo sistema de coordenadas onde a variação dos dados iniciais é descrita com o uso de menos dimensões do que havia anteriormente.

2.8.2 K vizinhos mais próximos (KNN)

K vizinhos mais próximos é um algoritmo de aprendizagem de máquina supervisionado, ilustrado na Figura 2.2, que consiste em encontrar os k vizinhos mais próximos para um dado k em uma base de dados baseando-se principalmente na distância euclidiana entre o novo dado e seus vizinhos mais próximos a fim de classificá-lo [10].

Como o algoritmo KNN depende fortemente da distância entre os elementos para classificá-los, dados contendo características representadas em diferentes grandezas ou em escalas muito distantes necessitam ser normalizados para permitir uma maior acurácia na classificação [31].

Figura 2.2 – Representação gráfica do algoritmo KNN

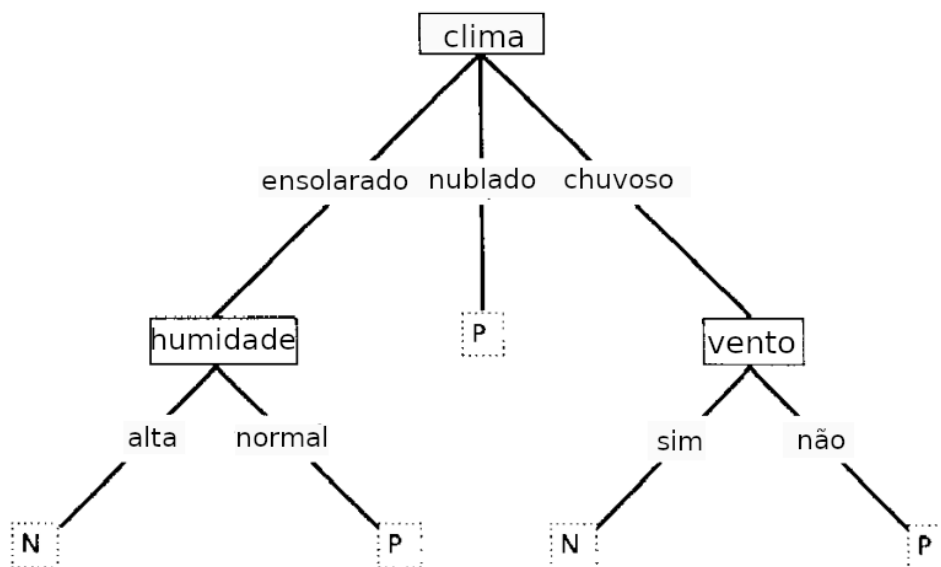


Fonte: Extraído de: [10]

2.8.3 Árvore de decisão

É uma abordagem de aprendizado de máquina supervisionado onde uma árvore de decisão é usada para prever valores com base em um conjunto de dados usados para o seu treinamento [32]. O método cria uma árvore onde cada nó não folha é nomeado com uma característica e cada um de seus nós filhos é nomeado com um valor possível do vetor de características fornecido ao classificador [32]. Cada nó folha da árvore corresponde a uma classificação e possui a probabilidade distribuída obtida pelo número de elementos que receberam essa classificação durante a etapa de treinamento [32]. A imagem 2.3 mostra uma representação de uma árvore de decisão para um exemplo classificando um clima como agradável ou não.

Figura 2.3 – Representação gráfica de uma árvore de decisão.



Fonte: Adaptado de: [32]

2.8.4 *Random Forest*

Random Forest é um algoritmo de aprendizado de máquina desenvolvido por Ho [21] em 1995 que opera construindo múltiplas árvores de decisão durante o seu treinamento, permitindo o aumento arbitrário de complexidade e diminuindo a perda de generalização que comumente ocorre em árvores de decisão tradicionais.

A essência do algoritmo é a de construir múltiplas árvores de decisão em subespaços aleatoriamente selecionados no espaço de características, o que permite generalizar a classificação de maneira complementar e o resultado pode ser monotonicamente melhorado [21].

2.8.5 *Ada Boost*

Ada Boost é um método de classificação criado por Freund e Schapire [14] em 1995 que pode ser usado em conjunto com diversos algoritmos de aprendizado a fim de produzir um modelo com melhor desempenho. O método recebe o valor de saída de classificadores que se deseja melhorar, denominados classificadores fracos, que são combinados à uma soma de pesos, onde os pesos são determinados durante a etapa de treinamento de acordo com a performance de cada classificador e o valor final obtido é um resultado com melhor acurácia [14].

2.8.6 *Extra Trees*

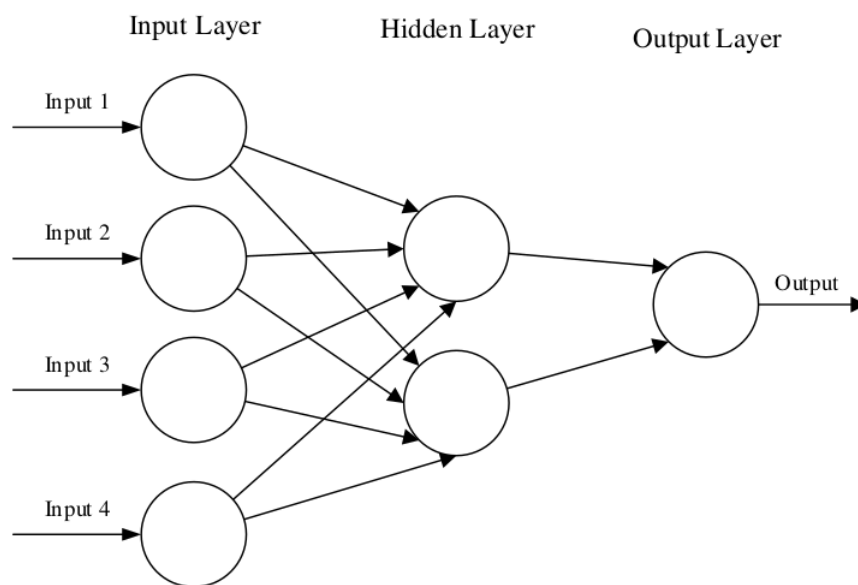
Extra Trees ou *Extremelly Randomized Trees* é um algoritmo de aprendizado de máquina proposto por Geurts, Ernst e Wehenkel [16] em 2006, que consiste em criar árvores de decisão da mesma maneira que o algoritmo *Random Forest*, porém de maneira fortemente aleatória.

A aleatoriedade do método é garantida pelo fato de também escolher o ponto de separação de uma característica de forma aleatória ao construir as árvores de decisão, diferente de um classificador *Random Forest* [16]. Essa diferença permite que as árvores de decisão sejam construídas de maneira independente dos valores do vetor de características, permitindo uma maior eficiência e consequentemente, uma melhor performance [16].

2.8.7 Máquina de Vetores de Suporte (SVM)

Máquina de Vetores de Suporte é um método de aprendizado de máquina supervisionado voltado para problemas de classificação binários ou multi-classes que busca implementar a ideia

Figura 2.4 – Representação gráfica de uma rede neural artificial



Fonte: Extraído de: [30]

de que um vetor de dados entrada é mapeado não-linearmente a um espaço de características de alta dimensão, para então construir uma superfície de decisão linear nesse espaço de características [11].

Dentre as principais vantagens, o algoritmo SVM é um método não paramétrico que retém as informações dos dados de treinamento e é capaz de representar funções complexas que geralmente necessitam de métodos paramétricos para serem representadas [34].

2.8.8 Redes neurais artificiais

Redes Neurais Artificiais são sistemas computacionais inspirados no funcionamento de um sistema nervoso biológico. Trata-se de uma rede de nós computacionais interconectados, denominados neurônios, que trabalham entrelaçados e de maneira distribuída aprendendo coletivamente com a entrada de informações a fim de otimizar a saída dessas informações [30].

A Figura 2.4 contém um exemplo da estrutura básica de uma rede neural artificial com uma

camada de entrada, uma camada oculta e uma camada de saída [30]. As camadas de entrada são alimentadas usualmente por vetores multidimensionais e os dados são repassados para as camadas ocultas que então tomarão decisões em conjunto de como uma mudança estocástica melhoraria ou pioraria o resultado final [30].

Por fim, a camada de saída é responsável por receber os dados transformados pelas camadas ocultas e é o último conjunto de neurônios a modificar de alguma forma os dados inseridos na rede neural [30]. O empilhamento de múltiplas camadas ocultas em uma rede neural é conhecido como *deep learning* e é usado para extrair progressivamente o máximo de características possível dos dados de entrada [30].

2.8.9 Redes neurais convolucionais

As redes neurais convolucionais são análogas às redes neurais artificiais, definidas na seção 2.8.8, no que diz respeito ao uso de camadas de neurônios para progressivamente aprimorar o próprio aprendizado. No entanto, as redes neurais convolucionais são compostas de *perceptrons* multicamada, ou seja, é uma rede de nós totalmente conectados, a qual cada neurônio de uma determinada camada é conectado a todos os neurônios da camada seguinte, tornando-as propensas ao sobre-ajuste de dados [30].

Redes neurais artificiais são amplamente usadas no reconhecimento de padrões e principalmente usando imagens, visto que o seu aspecto foi inspirado no padrão de conectividade dos neurônios do córtex visual de animais [30].

2.8.10 Aprendizado de máquina automatizado

Aprendizado de máquina automatizado ou *AutoML* é definido como o processo de automatizar a implementação de aprendizado de máquina para a solução de problemas reais usando o próprio aprendizado de máquina para selecionar o melhor modelo computacional e os melhores parâmetros, medindo a acurácia e o tempo de computação para um determinado conjunto de dados [23].

Esse tipo de automação permite a criação de soluções simples e com bom desempenho que implementam aprendizado de máquina de maneira ágil e otimizada por pessoas que não são especialistas em aprendizado de máquina [23].

2.8.11 Validação cruzada

Trata-se de uma técnica de validação de um modelo que busca analisar como ele generaliza os resultados considerando-se uma determinada base de dados. A técnica de validação cruzada *holdout* divide aleatoriamente uma base de dados em dois conjuntos, um que será usado para treinar o modelo e outro será usado para testá-lo. A acurácia do modelo é representada por uma pontuação dada pela porcentagem de acertos do modelo na fase de teste [34].

Outra técnica de validação cruzada é chamada *k-fold*, que consiste em dividir a base de dados em k conjuntos, onde um deles será usado para teste e o restante para treinamento. Isso se repetirá k vezes, sendo que um conjunto diferente será usado para teste em cada uma das iterações e a média das pontuações de todas as iterações corresponde a uma estimativa melhor do que uma única pontuação [34].

Tabela 2.2 – Representação de uma matriz de confusão

		Classificação testada	
		P	N
		VP	FP
Classificação prevista	PP	VP	FP
	PN	FN	VN

Fonte: do próprio autor

2.8.12 Matriz de confusão e métricas

Uma matriz de confusão é uma tabela que permite observar a acurácia de um algoritmo de classificação, no qual são apresentados o número de ocorrências de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos obtidos [37].

Essa representação permite observar a precisão de um modelo ao analisar a porcentagem ou os números absolutos de ocorrência de cada possibilidade de classificação de um dado [37]. A tabela 2.2 contém um exemplo de uma matriz de confusão, onde P e N são os números de ocorrências positivas e negativas, respectivamente. PP é o número de ocorrências positivos detectadas pelo classificador e PN o número de ocorrências negativas detectadas. VP e FP são os números de verdadeiros positivos e falsos positivos e FN e VN são os números de falsos negativos e verdadeiros negativos.

Algumas métricas são necessárias para melhor analisar o desempenho de um classificador. Dentre elas, as mais importantes são a acurácia, sensibilidade e especificidade.

Acurácia é a métrica que define a proporção de elementos classificados corretamente pelo classificador em relação ao total de elementos, definida pela equação 2.14.

$$Acuracia = \frac{VP + VN}{P + N} \quad (2.14)$$

Sensibilidade é definida como a proporção de verdadeiros positivos que foram corretamente

identificados pelo classificador [6]. A formula que define a sensibilidade está representada na equação 2.15

$$Sensibilidade = \frac{VP}{P} \quad (2.15)$$

Especificidade é definida como a razão entre o número de verdadeiros negativos identificados pelo classificador e o total de negativos da base de dados [6]. A formula que define a especificidade está representada na equação 2.16.

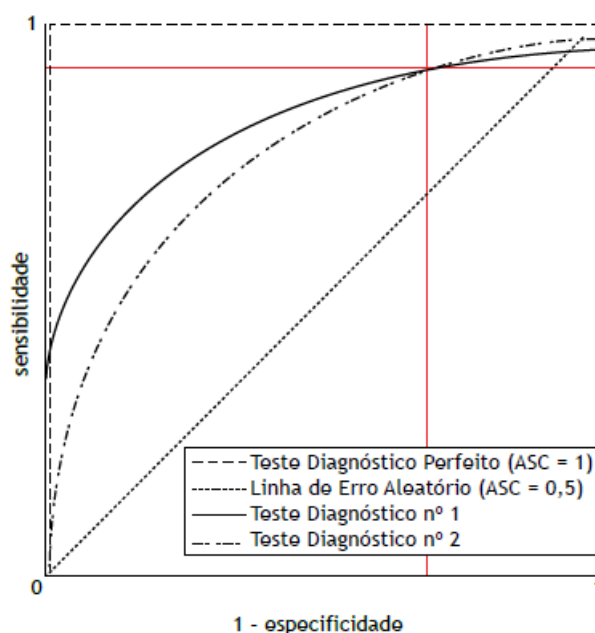
$$Especificidade = \frac{VN}{N} \quad (2.16)$$

2.8.13 Curva ROC e área sob a curva

Curva ROC, ou curva da característica de operação do receptor, é uma representação gráfica do desempenho de um classificador mostrando a relação entre a sensibilidade, ou taxa de verdadeiros positivos, e a especificidade, taxa de especificidade, ou taxa de verdadeiros-negativos, nos eixos y e x, respectivamente [15].

A área sob a curva (ASC) mede a habilidade de um classificador de diferenciar as amostras e representa a probabilidade de uma amostra positiva apresentar uma pontuação mais alta do que uma amostra negativa [15]. A figura 2.5 representa duas curvas ROC e o valor de suas respectivas ASC.

Figura 2.5 – Ilustração representando curvas ROC e suas respectivas ASC.



Fonte: Extraído de: [15]

2.9 Trabalhos relacionados

Nesta seção, alguns trabalhos relacionados ao tema proposto são abordados a fim de expor as referências utilizadas para o trabalho, bem como o estado da arte.

O trabalho realizado por Hossain *et al* [22] propõe um *framework* que interage com dispositivos conectados à Internet e são transmitidos via computação em nuvem para detecção de patologias na voz de maneira eficiente, permitindo a sua implementação em *smart cities*. O *software* utiliza sinais de voz e sinais eletroglotográficos (EGG) para classificar o locutor como normal ou patológico. O algoritmo de classificação utiliza o modelo de mistura gaussiana e permitiu uma acurácia de 93% utilizando a base de dados SVD.

Satake Junior [35], por sua vez, desenvolveu um método computacional para detectar patologias laríngeas a partir de sinais de voz e classificá-las como saudáveis ou patológicas. Os parâmetros utilizados para a classificação foram o fator de perturbação direcional, perturbação média relativa e fator de *jitter*. Foram utilizados os modelos de classificação SVM, que trouxe

uma acurácia média de 70%, e modelo KNN, que permitiu uma acurácia média de 65%.

Reid *et al* [33] desenvolveram e validaram um algoritmo de aprendizado de máquina que converte os sinais de voz em espectrogramas e os insere em uma rede neural convolucional e classificadora de imagens nomeada *VGG19*, que retorna um mapa de características. Esse dado é então classificado como patológico ou normal com o uso de um classificador SVM. O algoritmo foi treinado com gravações de locutores saudáveis e com disfonia da base de dados SVD e o seu teste revelou uma acurácia de 96,5%.

Fonseca *et al* [13] utilizaram informações como energia do sinal, taxa de *zero-crossing* e entropia do sinal para construir dois classificadores cujo objetivo é identificar vozes de pacientes com edema de Reinke e laringite. A abordagem utilizada foi a de criar dois classificadores separados e combinar os seus valores de saída, o que resultou em uma acurácia de até 95%.

Amami e Smiti [7] propuseram um método para classificar vozes com diferentes patologias que utiliza um algoritmo de agrupamento baseado em densidade chamado *DBSCAN* em conjunto com um algoritmo de aprendizado de máquina incremental que implementa um classificador SVM. O classificador construído usou gravações da base de dados MEEI e obteve uma acurácia de 98%.

Ali *et al* [4] desenvolveram um método baseado na intensidade da voz, capaz de detectar vozes patológicas automaticamente durante uma fala contínua. O algoritmo define os picos de intensidade do sinal formando um contorno da voz e discrimina entre patológico e saudável com base na área sob o contorno. A acurácia obtida pelo método foi de 100%. O motivo pelo qual o método se mostra eficaz é que as patologias detectadas prejudicam a mobilidade das cordas vocais do locutor, resultando em uma voz mais fraca que o normal.

Syed *et al* [40] realizaram um experimento comparando a utilização de classificadores baseados em redes neurais convolucionais (CNN) e baseados em redes neurais recorrentes (RNN) com 27 camadas cada. Foram extraídas 13 características do coeficiente cepstral da frequência de Mel (MFCC) de gravações da base de dados SVD e essas características foram utilizadas para testar e validar o método. O resultado foi uma acurácia de 87,11% usando as redes CNN e de 86,52% com as redes RNN.

Teixeira e Gonçalves [41] construíram um algoritmo que mede fatores de perturbação de

jitter, *shimmer* e a frequência fundamental do sinal para classificar vozes patológicas através da detecção de rouquidão na fala do paciente. Neste trabalho foram usadas gravações da base de dados SVD e a acurácia obtida foi alta, permitindo concluir que os parâmetros de *jitter* e *shimmer* são relevantes para o diagnóstico de vozes afetadas patologicamente.

Fang *et al* [12] desenvolveram um classificador para detectar vozes patológicas dentre 8 patologias comuns usando somente os coeficientes cepstrais de *Mel* de gravações de locutores saudáveis e patológicos sustentando um som vocálico. Ao testar o modelo construído com as gravações da base de dados MEEI foram obtidas acurácias de até 99,32%, demonstrando que os MFCCs são suficientes para detectar vozes afetadas patologicamente.

Cesari *et al* [8] construíram um algoritmo capaz de estimar a probabilidade da presença de uma patologia na voz calculando coeficientes de perturbação de *jitter*, *shimmer* e a proporção harmônico-ruído para avaliar quantitativamente a qualidade da voz do locutor. Foram usadas gravações das vozes de 208 voluntários, sendo 48 saudáveis e 150 patológicos e a melhor acurácia obtida usando um modelo de aprendizado de máquina foi de 77,4%.

Mohammed *et al* [28] propuseram uma ferramenta de extração de características para detecção de vozes patológicas baseada em *deep learning* em conjunto com um classificador de redes neurais convolucionais. Os experimentos foram realizados com gravações da base de dados SVD e apresentaram uma acurácia de até 95,41% mostrando que há possibilidade de aplicação do método em ambientes clínicos.

Al-Dhief *et al* [3] construíram um modelo de classificação de vozes patológicas e saudáveis usando uma tecnologia de aprendizado extremo sequencial online denominada *OSELM*. As características extraídas foram os coeficientes cepstrais de *Mel* a partir de gravações da base de dados SVD e a acurácia máxima obtida foi de 85%, demonstrando que o algoritmo é capaz de classificar vozes patológicas efetivamente.

Alves *et al* propuseram um algoritmo para detectar patologias de origem neurológica que afetam as cordas vocais usando características cepstrais em conjunto com uma transformada *wavelet*. Diferentes tipos de classificadores de aprendizado de máquina foram empregados e a maior acurácia obtida utilizando validação cruzada foi de 100%.

Zhang *et al* [46] desenvolveram um método de extração de características de vozes pa-

tológicas usando um filtro de sub-bandas com uma transformada *wavelet* na escala *Bark*. Em seguida, foram testados 14 classificadores de aprendizado de máquina que obtiveram uma acurácia máxima de 99,56% com o uso das gravações da base de dados MEEI.

No trabalho de Al-Nasheri *et al* [5] o objetivo foi encontrar parâmetros do *Multidimensional Voice Program (MDVP)* para detectar e classificar automaticamente patologias na voz em diferentes bases de dados para então classificar quais parâmetros configuram uma melhor detecção e classificação das doenças.

Os parâmetros MDVP foram definidos pela ferramenta *Multidimensional Voice Program*, um software comercial desenvolvido pela *Key Elemetrics* em 1993 que é capaz de analisar objetivamente a qualidade da voz utilizando parâmetros quantitativos que medem a frequência fundamental, amplitude e balanço de energia espectral além da presença de lacunas de sonoridade e diplofonia [29].

Ainda nesse artigo, o método proposto utilizou gravações de pessoas saudáveis e pacientes disfônicos sustentando a vogal "a" das bases de dados SVD, MEEI e AVPD. No experimento, os parâmetros MDVP coletados pelo software *Key Pentax CSL Model 3000*, um software comumente usado para a extração dessas características, foram inseridos em um classificador *SVM* para discriminar as gravações entre patológicas e normais. Em seguida, foi utilizada a razão de discriminação de Fisher (*FDR*) para ordenar os parâmetros que melhor detectaram as gravações patológicas.

Os três parâmetros que trouxeram melhores resultados foram a variação da amplitude de pico em cada período (*vAm*), variação da frequência fundamental fonatória (*PFR*) e o quociente de perturbação de amplitude (*APQ*), sendo esses os responsáveis por classificar as gravações patológicas com uma acurácia de 99.68%.

Em comparação com os trabalhos listados nesta seção, o projeto descrito nesta monografia é similar por valer-se da construção de classificadores de vozes patológicas fazendo o uso de coeficientes cepstrais na extração de características. Os trabalhos de Syed *et al*, Fang *et al*, Al-Dhief *et al* e Alves *et al* fazem o uso dos coeficientes de *Mel*, assim como foi feito neste trabalho. Por outro lado, o trabalho de Zhang *et al* fez uso da transformada *wavelet* e da filtragem de sub-bandas na escala *Bark* para classificar vozes patológicas, abordagens que não foram exploradas

neste trabalho.

Em complemento, este trabalho diferencia-se dos demais por fazer uso dos coeficientes cepstrais de *Bark* na extração de características, alternativo ao trabalho de Zhang *et al* que usou a escala *Bark* para aplicar a transformada *wavelet*.

Capítulo 3

Metodologia

Neste capítulo serão descritas as etapas realizadas neste trabalho fazendo uso dos conceitos apresentados no Capítulo 2. Serão descritas as seguintes etapas: obtenção dos dados a serem usados, extração das características, construção do classificador e análise dos resultados.

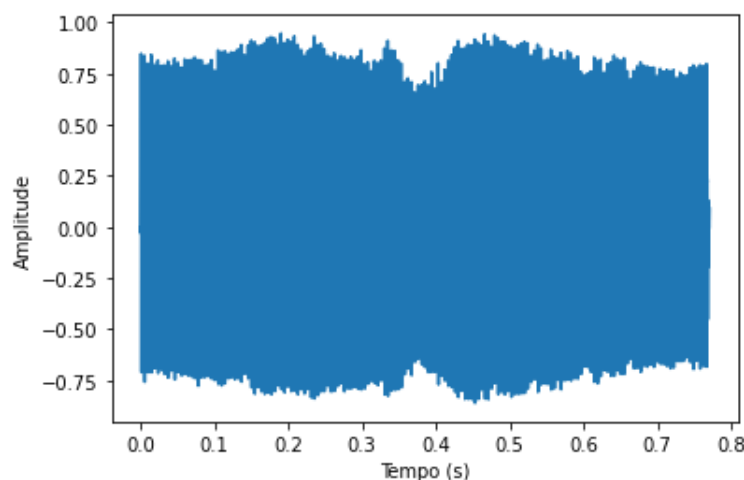
O objetivo deste trabalho é a criação de um método computacional capaz de classificar sinais de voz com três diferentes patologias.

3.1 Obtenção dos dados

Os dados utilizados para construção do método computacional são gravações de voz da vogal /a/ em tom normal sustentada por aproximadamente um segundo. As gravações estão no formato WAVE, amostradas em 50kHz e quantizadas em 16 bits. Esses dados foram obtidos a partir do *Saarbrücken Voice Database* (SVD), mantido pela *Universität des Saarlandes*, na cidade de *Saarbrücken*, Alemanha.

Foram usadas 1308 gravações, sendo 869 (66,4%) de pacientes saudáveis, 210 (16,1%) de pacientes com paralisia das cordas vocais, 134 (10,2%) de pacientes com laringite e 95 (7,3%) de pacientes com disfonia orgânica.

Figura 3.1 – Sinal de áudio de uma amostra contendo a amplitude e o tempo em segundos.



Fonte: do próprio autor.

Durante a obtenção das gravações, foram selecionadas somente as gravações cujo paciente foi diagnosticado apenas com uma das patologias acima, excluindo os casos em que o diagnóstico confirmava duas ou mais doenças das que serão identificadas, pois o classificador trabalha apenas com uma única classificação para cada amostra.

3.2 Extração das características

Na extração das características, as informações extraídas foram os coeficientes cepstrais de Mel (MFCC), definidos em 2.5.1, os coeficientes cepstrais de Bark (BFCC) definidos em 2.6.1, os coeficientes de perturbação definidos em 2.7 além da duração, a frequência média em Hertz e o desvio padrão da frequência.

A obtenção dos coeficientes de perturbação foi feita com o uso da biblioteca *Parselmouth* que abstrai o software *Praat* em *Python*. *Praat* é um software de análise e síntese de fala, desenvolvido por Boersma e Weenink no Instituto de Ciências Fonéticas da Universidade de Amsterdã [9].

Uma ferramenta para normalizar os coeficientes de *Mel* e *Bark* obtidos é a análise de compo-

Figura 3.2 – Extração dos MFCCs e BFCCs para compor o vetor de características X.

```
X = []

for nome_arquivo in os.listdir('/dataset/'):
    audio = parselmouth.Audio(nome_arquivo)
    sinal = audio.as_array()
    amostragem = audio.sampling_frequency
    mfcc = spafe.features.mfcc.mfcc(sinal, sr=amostragem)
    bfcc = spafe.features.bfcc.bfcc(sinal, sr=amostragem)

    # Normalizacao usando PCA
    pca = sklearn.decomposition.PCA(n_components=2)
    mfcc_norm = pca.fit_transform(mfcc_f)
    bfcc_norm = pca.fit_transform(bfcc_f)

    # Adiciona as caracteristicas desejadas a X
    X.append(
        np.concatenate([
            coefs_perturbacao,
            mfcc_norm,
            bfcc_norm
        ])
    )
```

Fonte: do próprio autor.

nentes principais (PCA), definida na seção 2.8.1. Neste caso, foi usada a solução da biblioteca *Scikit Learn* para reduzir a dimensionalidade do vetor de características e otimizar o processo de aprendizado do classificador.

O primeiro passo para extrair os MFCCs e BFCCs foi recortar todas as gravações para que todas tenham a mesma duração, a fim de manter os vetores de característica da mesma proporção para todo o conjunto de dados. Em seguida, os MFCCs e BFCCs foram efetivamente extraídos com o uso da biblioteca *Spafe* e submetidos a um PCA, conforme está representado no algoritmo em Python da figura 3.2, onde as gravações estão dispostas no diretório */dataset/* e os coeficientes de perturbação estão na variável *coefs_perturbacao*.

Figura 3.3 – Aplicação do algoritmo de aprendizado de máquina automatizado para o vetor de características X e o vetor y que contém as classificações de cada amostra.

```
# Separacao dos dados de treino e teste
X_treino, X_teste, y_treino, y_teste = train_test_split(
    X,
    y,
    test_size=0.4
)

# Execucao do algoritmo de aprendizado de maquina automatizado
automl = autosklearn.classification.AutoSklearnClassifier(
    time_left_for_this_task=240,
    per_run_time_limit=60,
    n_jobs=8
)
automl.fit(X_train, y_train)
```

Fonte: do próprio autor.

3.3 Construção do classificador

A construção do classificador foi feita usando aprendizado de máquina automatizado para selecionar os melhores classificadores e parâmetros para um determinado conjunto de dados e tempo limite de computação. Esse método possibilita verificar quais grupos de características são relevantes para classificar vozes patológicas pois permite testá-los individualmente obtendo a melhor acurácia em diferentes modelos de aprendizado de máquina.

A biblioteca de aprendizado de máquina automatizado usada neste trabalho foi a *auto-sklearn* que utiliza os classificadores da biblioteca *sklearn*. Essa biblioteca é capaz de classificar os modelos testados e seus parâmetros de acordo com as melhores acurácias obtidas e os menores custos computacionais.

A figura 3.3 mostra o código em *Python* da separação dos dados de treinamento e teste e a aplicação do algoritmo de aprendizado de máquina automatizado para para o conjunto de dados construídos na etapa anterior, o tempo limite total definido em 240 segundos, o tempo limite por tarefa definido em 60 segundos e o número máximo de trabalhos concorrentes definido em 8.

Figura 3.4 – Validação cruzada *k-fold* dos classificadores obtidos pelo algoritmo de aprendizado de máquina automatizado.

```
# Obtencao das acuracias para cada iteracao da validacao
acuracias = sklearn.model_selection.cross_val_score(
    classificador, X, y, scoring='accuracy', cv=10
)

# Calculo da media e desvio padrao
media = np.mean(acuracias)
desvio = np.std(acuracias)
print(media, desvio)

# Predicao usando validacao cruzada
y_pred = sklearn.model_selection.cross_val_predict(
    classificador, X, y, cv=10
)

# Construcao da matriz de confusao
matriz_conf = sklearn.metrics.confusion_matrix(y, y_pred)
print(matriz_conf)
```

Fonte: do próprio autor.

3.4 Teste do classificador e análise dos resultados

A etapa de teste do classificador e análise dos resultados consiste em submeter o classificador de melhor desempenho obtido pelo algoritmo de aprendizado de máquina automatizado à uma validação cruzada *k-fold* usando $k = 10$. As métricas obtidas foram a média e o desvio padrão da acurácia calculada durante a validação cruzada, além das matrizes de confusão do melhor e pior caso, a curva ROC e sua respectiva área abaixo da curva.

O algoritmo em *Python* representado na figura 3.4 mostra a execução dos testes de validação cruzada *k-fold* e construção da matriz de confusão no melhor classificador obtidos pelo algoritmo de aprendizado de máquina automatizado, representado pela variável *classificador*.

Capítulo 4

Testes e Resultados

A realização dos testes foi feita usando 4 conjuntos de teste: um usando apenas os coeficientes de perturbação, outro usando os coeficientes de perturbação somados aos MFCCs, outro usando os coeficientes de perturbação somados aos BFCCs e por fim um usando todos os três conjuntos de coeficientes.

Os resultados de cada conjunto de teste estão apresentados da seção 4.1 à seção 4.4 contendo os seguintes dados: acurácia média e desvio padrão, matrizes de confusão do melhor e pior caso, curva ROC e sua respectiva área abaixo da curva.

4.1 Testes do conjunto I - Coeficientes de perturbação

O classificador selecionado pelo algoritmo de aprendizado de máquina automatizado foi o *Extra Trees* combinado a um algoritmo de seleção de características no pré-processamento. A acurácia média obtida na validação cruzada *k-fold* foi 0,631 com desvio padrão de 0,019 e a matriz de confusão do melhor caso está representada na tabela 4.1, a matriz de confusão do pior caso na tabela 4.2, a curva ROC e a área abaixo da curva na figura 4.1.

Ao analisar a matriz de confusão para este conjunto, é possível notar uma concentração

Tabela 4.1 – Matriz de confusão do melhor caso do teste do conjunto I.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	90	6	4	2
	Paralisia	1	1	0	0
	Laringite	0	0	1	0
	Disfonia	0	0	0	0

Fonte: do próprio autor

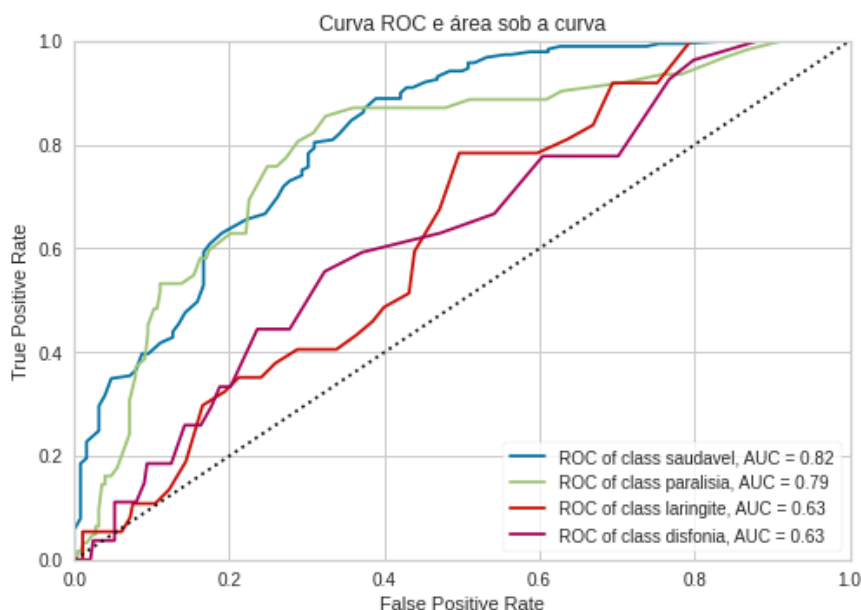
Tabela 4.2 – Matriz de confusão do pior caso do teste do conjunto I.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	0	0	0	0
	Paralisia	24	9	3	0
	Laringite	22	12	0	0
	Disfonia	24	9	1	1

Fonte: do próprio autor

das classificações como saudável, grande parte delas correta, porém com um grande número de falsas classificações, que revela um classificador com desempenho mediano.

Figura 4.1 – Curvas ROC obtidas no conjunto de teste I.



Fonte: do próprio autor.

4.2 Testes do conjunto II - Coeficientes de perturbação e MFCCs

O classificador selecionado também foi o *Extra Trees*, porém combinado a um pré-processador de aprendizado de máquina que também implementa o algoritmo *Extra Trees* para otimizar o vetor de características. A média da acurácia obtida na validação cruzada foi de 0,652 com desvio padrão de 0,027. As matrizes de confusão do melhor e pior caso estão representadas nas tabelas 4.3 e 4.4, respectivamente.

A matriz de confusão do conjunto II permite observar uma concentração de classificações saudáveis e a existência de um número alto de gravações falsamente classificadas como saudáveis. Além disso, houve uma pequena melhora na acurácia obtida. A curva ROC e a respectiva área abaixo da curva está representada na figura 4.2.

Tabela 4.3 – Matriz de confusão do melhor caso do teste do conjunto II.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	98	4	0	0
	Paralisia	2	0	0	0
	Laringite	1	0	0	0
	Disfonia	0	0	0	0

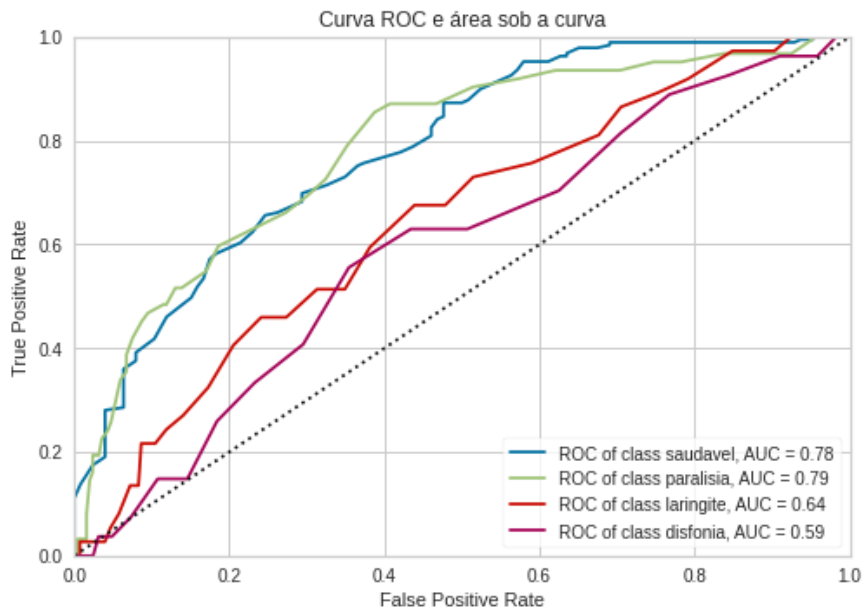
Fonte: do próprio autor

Tabela 4.4 – Matriz de confusão do pior caso do teste do conjunto II.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	0	0	0	0
	Paralisia	27	9	0	0
	Laringite	25	9	0	0
	Disfonia	27	8	0	0

Fonte: do próprio autor

Figura 4.2 – Curvas ROC obtidas no conjunto de teste II.



Fonte: do próprio autor.

4.3 Testes do conjunto III - Coeficientes de perturbação e BFCCs

O classificador com o melhor desempenho foi o *Random Forest* combinado a um pré-processador de aprendizado de máquina implementando o algoritmo *Extra Trees* para otimizar o vetor de características. A acurácia na validação cruzada *k-fold* foi de 0,651 com desvio padrão de 0,03 e a matriz de confusão do melhor caso está representada na tabela 4.5 e a do pior caso na tabela 4.6. A curva ROC e a respectiva área abaixo da curva está representada na figura 4.3.

Assim como os testes anteriores, o resultado apresentou um número alto de amostras falsamente classificadas como saudáveis e a acurácia se manteve muito próxima dos outros conjuntos.

Tabela 4.5 – Matriz de confusão do melhor caso do teste do conjunto III.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	95	6	1	0
	Paralisia	2	0	0	0
	Laringite	1	0	0	0
	Disfonia	0	0	0	0

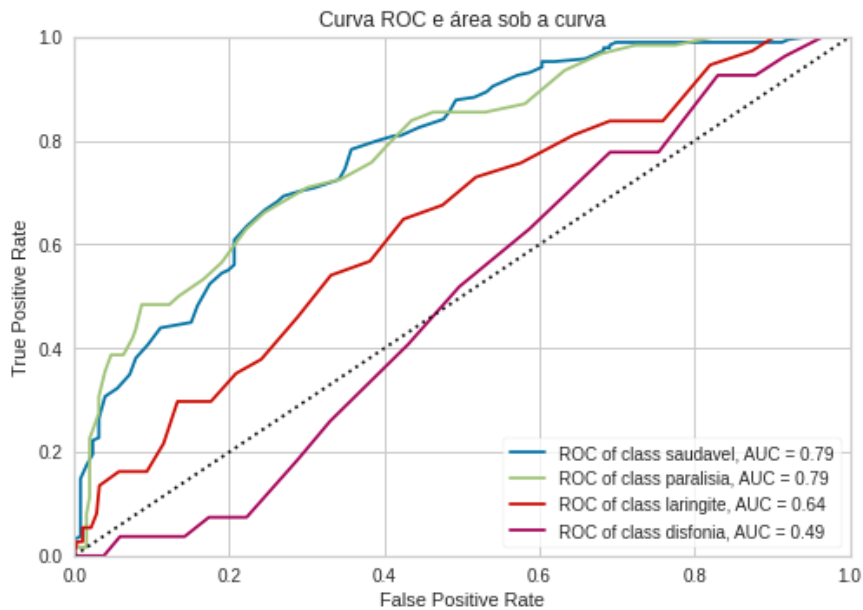
Fonte: do próprio autor

Tabela 4.6 – Matriz de confusão do pior caso do teste do conjunto III.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	0	0	0	0
	Paralisia	28	8	0	0
	Laringite	25	9	0	0
	Disfonia	27	7	1	0

Fonte: do próprio autor

Figura 4.3 – Curvas ROC obtidas no conjunto de teste III.



Fonte: do próprio autor.

4.4 Testes do conjunto IV - Coeficientes de perturbação, MFCCs e BFCCs

Ao adicionar todos os conjuntos de coeficientes, o classificador com o melhor desempenho foi o *Random Forest* e sem nenhum pré-processamento do vetor de características. A acurácia foi de 0,652 e desvio padrão de 0,023 na validação cruzada *k-fold*. A matriz de confusão do melhor caso do conjunto IV está representada na tabela 4.7 e a do pior caso está representada na tabela 4.8. A curva ROC e a respectiva área abaixo da curva está representada na figura 4.4.

Novamente o teste não demonstrou grandes alterações mesmo combinando todas as características, o que permite concluir que as características utilizadas podem ser capazes de identificar vozes patológicas porém não fornecem características suficientes para diferenciar as patologias.

Tabela 4.7 – Matriz de confusão do melhor caso do teste do conjunto IV.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	98	4	0	0
	Paralisia	2	0	0	0
	Laringite	1	0	0	0
	Disfonia	0	0	0	0

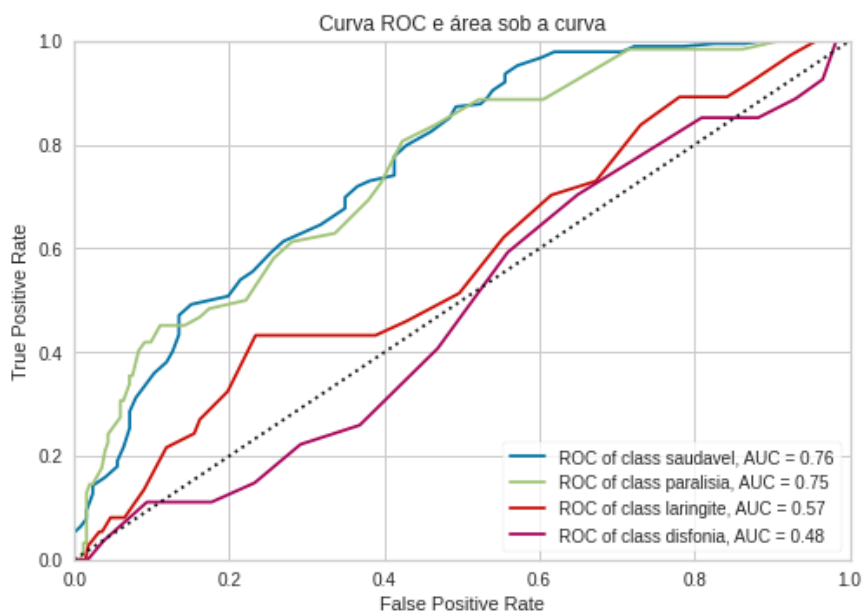
Fonte: do próprio autor

Tabela 4.8 – Matriz de confusão do pior caso do teste do conjunto IV.

		Classificação testada			
		Saudável	Paralisia	Laringite	Disfonia
Classificação prevista	Saudável	0	0	0	0
	Paralisia	28	7	1	0
	Laringite	24	10	0	0
	Disfonia	29	5	1	0

Fonte: do próprio autor

Figura 4.4 – Curvas ROC obtidas no conjunto de teste IV.



Fonte: do próprio autor.

4.5 Testes do conjunto V - Coeficientes de perturbação, MFCCs e BFCCs com classificadores individuais

Por fim, foram testadas as mesmas características testadas na seção 4.4, porém treinando e testando as gravações de cada patologia com as gravações de indivíduos saudáveis separadamente e apresentando os seus resultados.

4.5.1 Laringite

Ao executar o algoritmo de aprendizado de máquina automatizado, o modelo de classificador selecionado com as gravações de pacientes com laringite foi o *Random Forest* combinado a um algoritmo de seleção de características para o pré-processamento.

Tabela 4.9 – Matriz de confusão do melhor caso do teste do conjunto V com laringite.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	73	2
	Laringite	0	0

Fonte: do próprio autor

Tabela 4.10 – Matriz de confusão do pior caso do teste do conjunto V com laringite.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	32	0
	Laringite	40	2

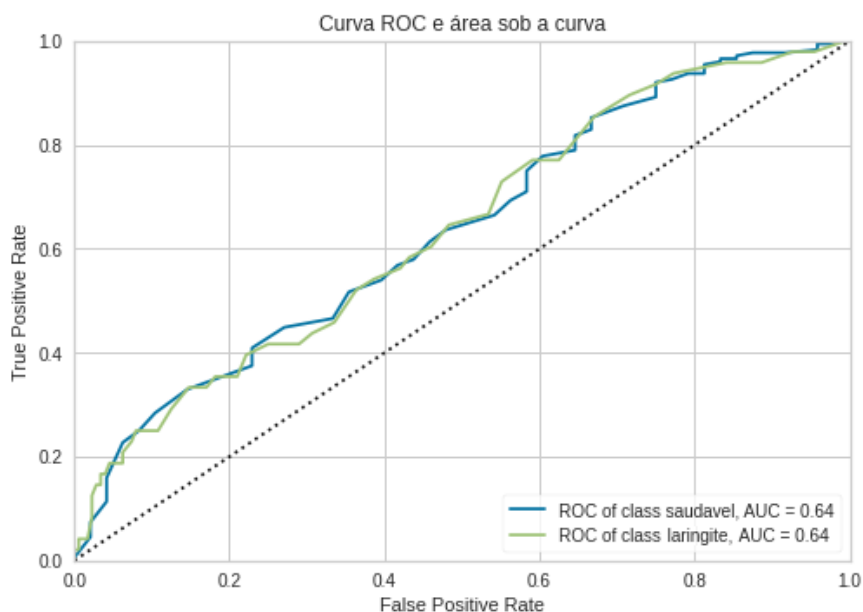
Fonte: do próprio autor

A acurácia média obtida foi de 0,829 com desvio padrão de 0,143 e as matrizes de confusão do melhor caso e pior caso estão representadas nas tabelas 4.9 e 4.10, respectivamente. A curva ROC e a área abaixo da curva está representada na figura 4.5.

4.5.2 Disfonia orgânica

O modelo selecionado para o grupo das gravações de pacientes com disfonia orgânica foi um classificador de árvore de decisão combinado ao algoritmo *Ada Boost* para melhorar seu desempenho e usando um algoritmo de seleção de características no pré-processamento. A acurácia obtida foi de 0,862 com desvio padrão de 0,203 e as matrizes de confusão do melhor caso e pior caso estão representadas nas tabelas 4.11 e 4.12, respectivamente. A curva ROC e a área abaixo da curva estão representadas na figura 4.6.

Figura 4.5 – Curvas ROC obtidas no conjunto de teste V com laringite.



Fonte: do próprio autor.

Tabela 4.11 – Matriz de confusão do melhor caso do teste do conjunto V com disfonia orgânica.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	80	3
	Laringite	0	0

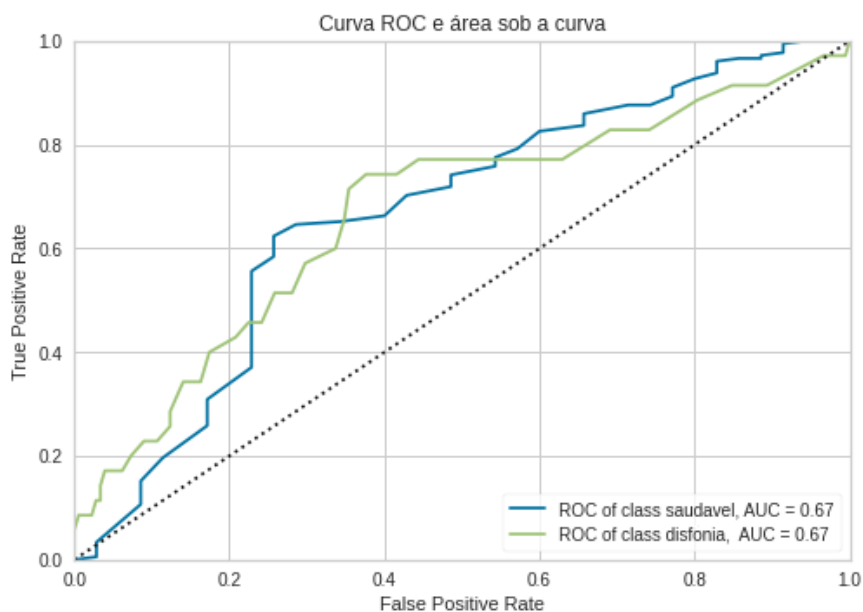
Fonte: do próprio autor

Tabela 4.12 – Matriz de confusão do pior caso do teste do conjunto V com disfonia orgânica.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	22	0
	Laringite	46	14

Fonte: do próprio autor

Figura 4.6 – Curvas ROC obtidas no conjunto de teste V com disfonia orgânica.



Fonte: do próprio autor.

4.5.3 Paralisia das cordas vocais

Para as gravações de pacientes com paralisia das cordas vocais, o modelo selecionado foi o classificador *Random Forest* com o algoritmo de seleção de características como pré-processador do modelo.

A acurácia obtida foi de 0,804 com desvio padrão de 0,141 e as matrizes de confusão do melhor caso e pior caso estão representadas nas tabelas 4.13 e 4.14, respectivamente. A curva ROC e a área abaixo da curva estão representadas na figura 4.7.

Tabela 4.13 – Matriz de confusão do melhor caso do teste do conjunto V com paralisia das cordas vocais.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	70	1
	Laringite	0	0

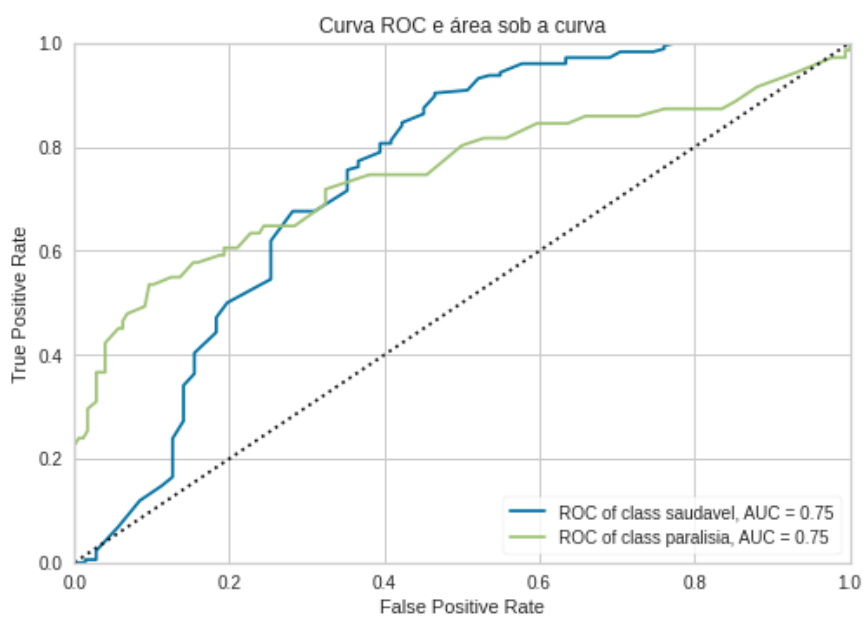
Fonte: do próprio autor

Tabela 4.14 – Matriz de confusão do pior caso do teste do conjunto V com paralisia das cordas vocais.

		Classificação testada	
		Saudável	Laringite
Classificação prevista	Saudável	19	0
	Laringite	51	0

Fonte: do próprio autor

Figura 4.7 – Curvas ROC obtidas no conjunto de teste V com paralisia das cordas vocais.



Fonte: do próprio autor.

4.6 Conclusão

Os resultados obtidos no grupo de teste V mostram uma melhora significativa em todas as métricas analisadas, indicando que as características extraídas são relevantes para classificar um sinal de voz como afetado patologicamente com as doenças estudadas neste trabalho.

No entanto, os resultados obtidos nos grupos I, II, III e IV revelam um desempenho mediano e insuficiente para construir um classificador confiável, o que indica que os coeficientes cepstrais estudados neste trabalho não carregam informações suficientes para diferenciar gravações de pacientes com cada uma das patologias abordadas em um único classificador.

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho foi proposto um método computacional para identificar e caracterizar vozes afetadas patologicamente com laringite, disfonia e paralisia das cordas vocais, fazendo uso de algoritmos de aprendizado de máquina automatizado e características cepstrais do sinal da voz.

Em suma, foi possível observar que os coeficientes cepstrais utilizados são capazes de identificar vozes patológicas conforme observado nos testes realizados na seção 4.5 e pelos trabalhos relacionados trazidos na seção 2.9. No entanto, também foi possível observar que os coeficientes de perturbação trazidos na seção 2.7, além dos coeficientes cepstrais de *Mel* e de *Bark* não são capazes de diferenciar as patologias entre si com o uso de um único classificador.

Por fim, os trabalhos futuros neste âmbito podem incluir o uso de outras bases de dados, como as bases de dados disponíveis publicamente MEEI, HUPA e AVPD, ou pelo uso da coleta manual de novas amostras. Outro possível aprimoramento seria a extração de outras características como energia do sinal, taxa de *zero-crossing* e entropia do sinal, além do uso de técnicas mais robustas como a análise *wavelet* para a extração de mais características.

Referências

- [1] ABDULMAJEED, N.Q.; AL-KHATEEB, B.; MOHAMMED, M.A. *A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions* Journal of Intelligent Systems, 31(1), pp.855-875. 2022.
- [2] AHMED, N.; NATARAJAN, T.; RAO, K. R. *Discrete Cosine Transform* IEEE Transactions on Computers, vol. C-23, no. 1, pp. 90-93, 1974.
- [3] AL-DHIEF, F. T.; BAKI, M. M.; LATIFF, N. M. A. A.; MALIK, N. N. N. A.; SALIM, N. S.; ALBADER, M. A. A; MAHYUDDIN, N. M.; MOHAMMED, M. A. *Voice pathology detection and classification by adopting online sequential extreme learning machine* IEEE Access, v. 9, p. 77293-77306, 2021.
- [4] ALI, Z.; ALSULAIMAN, M.; ELAMVAZUTHI, I.; MUHAMMAD, G.; MESALLAM, T. A.; FARAHAT, M.; MALKI, K. H. *Voice pathology detection based on the modified voice contour and SVM* Biologically Inspired Cognitive Architectures, v. 15, p. 10-18. 2016.
- [5] AL-NASHERI, A.; MUHAMMAD, G.; ASULAIMAN, M.; ALI, Z.; MESALLAM, T. A.; FARAHAT, M.; MALKI, K. H.; BENCHERIF, M. A. *An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification* Journal of Voice, Elsevier, 2017.
- [6] ALTMAN, D.G.; BLAND, J.M. *Diagnostic tests. 1: Sensitivity and specificity* BMJ: British Medical Journal, 308(6943), p.1552. 1994.

- [7] AMAMI, R.; SMITI, A. *An incremental method combining density clustering and support vector machines for voice pathology detection* Computers & Electrical Engineering, v. 57, p. 257-265. 2017.
- [8] CESARI, U.; DE PIETRO, G.; MARCIANO, E.; NIRI, C.; SANNINO, G.; VERDE, L. *Voice disorder detection via an m-Health system: Design and results of a clinical study to evaluate Vox4Health* BioMed research international. 2018.
- [9] BOERSMA, Paul; WEENINK, David. *Praat: doing phonetics by computer* Versão 6.2.23. Disponível em: <http://www.praat.org/retrieved>. Acesso em 8 de outubro de 2022.
- [10] CHOMBOON, K.; CHUJAI, P.; TEERARASSAMEE, P.; KERDPRASOP, K.; KERDPRASOP, N. *An empirical study of distance metrics for k-nearest neighbor algorithm* Proceedings of the 3rd international conference on industrial application engineering (pp. 280-285). 2015.
- [11] CORTES, C.; VAPNIK, V. *Support-vector networks* Machine learning, 20(3), pp. 273-297. 1995.
- [12] FANG, S.; TSAO, Y.; HSIAO, M. J.; CHEN, J. Y.; LAI, Y. H.; LIN, F. C.; WANG, C. T. *Detection of pathological voice using cepstrum vectors: A deep learning approach* Journal of Voice, v. 33, n. 5, p. 634-641. 2019.
- [13] FONSECA, E. S.; GUIDO, R. C.; JUNIOR, S. B.; DEZANI, H.; GATI, R. R.; PEREIRA, D. C. M. *Acoustic investigation of speech pathologies based on the discriminative para-consistent machine (DPM)*. Biomedical Signal Processing and Control, v. 55, p. 101615, 2020.
- [14] FREUND, Y.; SCHAPIRE, R. E. *A desicion-theoretic generalization of on-line learning and an application to boosting* Vitányi, P. (eds) Computational Learning Theory. Euro-COLT 1995. Lecture Notes in Computer Science, vol 904. Springer, Berlin, Heidelberg. 1995.

- [15] GARCÍA, J. P.; FERREIRA, J. C.; PATINO, C. M. *Análise ROC: uma aliada na pandemia* Jornal Brasileiro de Pneumologia, v. 47, 2021.
- [16] GEURTS, P.; ERNST, D.; WEHENKEL, L. *Extremely randomized trees* Mach Learn. 63(1), 3–42. 2006.
- [17] GRAPS. A *An introduction to wavelets* IEEE computational science and engineering. 2(2):50-61. 1995.
- [18] HAMMING, R. W. *Digital filters* Courier Corporation. 1998.
- [19] HARAR, P.; GALAZ, Z.; ALONSO-HERNANDEZ, J. B.; MEKYSKA, J.; BURGET, R.; SMEKAL, Z. *Towards robust voice pathology detection* The Natural Computing Applications Forum. 2018.
- [20] HASAN, M. R.; JAMIL, M.; RAHMAN, M. G. R. M. S. *Speaker identification using mel frequency cepstral coefficients* 3rd International Conference on Electrical & Computer Engineering - ICECE 2004, 1(4), pp. 565-568. 2004.
- [21] HO, T. K. *Random decision forests* Proceedings of 3rd international conference on document analysis and recognition. IEEE, p. 278-282. 1995.
- [22] HOSSAIN, M. S.; MUHAMMAD, G.; ALAMRI, A. *Smart healthcare monitoring: a voice pathology detection paradigm for smart cities* Multimedia Systems, Springer-Verlag, 2017.
- [23] HUTTER F.; CARUANA R.; BARDENET R.; BILENKO M.; GUYON I.; KEGL B.; LAROCHELLE H. *AutoML 2014 ICML* AutoML 2014 Workshop ICML. 2014.
- [24] ISAK P.; GUPTA M.; SWIERCZYNSKA-KREPA M.; SWIERCZYNSKI Z. *Voice Disorders: Functional and Organic Dysphonia* McMaster Textbook of Internal Medicine. Kraków: Medycyna Praktyczna. Disponível em: <https://empendium.com/mcmtextbook/chapter/B31.II.3.4.2>. Acesso em 22 out 2022.

- [25] JOLLIFF, I. T.; CADIMA, J. *Principal component analysis: a review and recent developments* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 374 (2065). 2016.
- [26] KAISER, G.; HUDGINS, L.H. *A friendly guide to wavelets* Boston Birkhauser. 1994.
- [27] MEKYSKA, J.; JANOUSOVA, E.; GOMEZ-VILDA, P.; SMEKAL, Z.; REKTOROVA, I.; ELIASOVA, I.; KOSTALOVA, M.; MRACKOVA, M.; ALONSO-HERNANDEZ, J.B.; FAUNDEZ-ZANUY, M. and LOPEZ-DE-IPINA, K. *Robust and complex approach of pathological speech signal analysis* Neurocomputing, 2015.
- [28] MOHAMMED, M. A.; ABDULKAREEM, K. H.; MOSTAFA, S. A.; KHANAPI ABD GHANI, M.; MAASHI, M. S.; GARCIA-ZAPIRAIN, B.; OLEAGORDIA, I.; ALHAKAMI, H.; AL-DHIEF, F. T. *Voice pathology detection and classification using convolutional neural network model* Applied Sciences, 10(11), 3723. 2020.
- [29] NICASTRI, M.; CHIARELLA, G.; GALLO, L. V.; CATALANO, M.; CASSANDRO, E. *Multidimensional Voice Program (MDVP) and amplitude variation parameters in euphonic adult subjects. Normative study* ACTA Otorhinolaryngologica Italica, 2004.
- [30] O'SHEA, K.; NASH, R. *An Introduction to Convolutional Neural Networks* arXiv preprint arXiv:1511.08458. 2015.
- [31] PIRYONESI, S. M.; EL-DIRABY, T. E. *Role of data analytics in infrastructure asset management: Overcoming data size and quality problems* Journal of Transportation Engineering, Part B: Pavements, 146(2), p.04020022. 2020.
- [32] QUINLAN, J. R. *Induction of decision trees* Machine Learning. 1: 81–106. 1986.
- [33] REID, J.; PARMAR, P.; LUND, T.; AALTO, D. K.; JEFFERY, C. C. *Development of a machine-learning based voice disorder screening tool* American Journal of Otolaryngology, 43(2), p. 103327. 2022.
- [34] RUSSELL, S. J. *Artificial intelligence a modern approach* Pearson Education, Inc. 2010.

- [35] SATAKE JUNIOR, E. H. *Processamento E Análise De Sinais Digitais Vozeados Para O Pré-Diagnóstico de Patologias Laríngeas* 2022. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Univerisdade Estadual Paulista, São José do Rio Preto - SP, 2022.
- [36] SOUZA, L. M.; FONSECA, E.; GUIDO, R. C. *Predição Linear e Transformada Wavelet Discreta para Identificação de Patologias na Voz* INFOIP: Revista Online de Informática, Inovação e Pesquisa, v. 2, p. 3, 2010.
- [37] STEHMAN, S. V. *Selecting and interpreting measures of thematic classification accuracy* Remote sensing of Environment, 62(1), pp.77-89. 1997.
- [38] STEVENS, S. S.; VOLKMANN, J.; NEWMANN, E. B. *A scale for the measurement of the psychological magnitude pitch* Journal of the Acoustical Society of America. 8 (3): 185–190. 1937
- [39] STOICA, P.; MOSES, R. L. *Spectral analysis of signals* Vol. 452, pp. 25-26. Upper Saddle River, NJ: Pearson Prentice Hall. 2005,
- [40] SYED, S. A.; RASHID, M.; HUSSAIN, S.; ZAHID, H. *Comparative analysis of CNN and RNN for voice pathology detection* BioMed Research International, 2021.
- [41] TEIXEIRA, J. P.; GONÇALVES, A. *Algorithm for jitter and shimmer measurement in pathologic voices* Procedia Computer Science, 100, pp.271-279. 2016.
- [42] THARWAT, A. *Classification assessment methods*, Applied Computing and Informatics, Vol. 17 No. 1, pp. 168-192. 2021.
- [43] TITZE, I. R. *Principles of Voice Production* Allyn & Bacon. 1994.
- [44] TUCKER, H. M. Vocal cord paralysis—1979: etiology and management. *The Laryngoscope*, v. 90, n. 4, p. 585-590, 1980.
- [45] WOOD, J. M.; ATHANASIADIS, T.; ALLEN, J. *Laryngitis* The BMJ. BMJ. 349: g5827. 2014.

- [46] ZHANG, X. J.; ZHU, X. C.; WU, D.; XIAO, Z. Z.; TAO, Z.; ZHAO, H. M. *Nonlinear Features of Bark Wavelet Sub-band Filtering for Pathological Voice Recognition* Engineering Letters, 29(1). 2021.
- [47] ZWICKER, E. *Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)* The Journal of the Acoustical Society of America, 33(2), pp.248-248. 1961.