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

Edson Haruyuki Satake Junior

Processamento E Análise De Sinais Digitais Vozeados Para O Pré-Diagnóstico de Patologias Laríngeas

São José do Rio Preto - SP
2022

Edson Haruyuki Satake Junior

Processamento E Análise De Sinais Digitais Vozeados Para O Pré-Diagnóstico de Patologias Laríngeas

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

Orientador: Prof. Dr. Rodrigo Capobianco Guido

São José do Rio Preto - SP

2022

S253p

Satake Junior, Edson Haruyuki

Processamento e análise de sinais digitais vozeados para o pré-diagnóstico de patologias laríngeas / Edson Haruyuki Satake Junior. -- São José do Rio Preto, 2022

59 p. : il., tabs.

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. Ciência da computação. 2. Inteligência artificial. 3. Processamento de sinais Técnicas digitais. 4. Distúrbios da voz. 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.

Edson Haruyuki Satake Junior

Processamento E Análise De Sinais Digitais Vozeados Para O Pré-Diagnóstico de Patologias Laríngeas

Trabalho de Conclusão de Curso (TCC) apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, junto ao Conselho de Curso de Bacharelado em Ciência da Computação, do Instituto de Biociências, 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. Dr. Aleardo Manacero Junior
UNESP – Câmpus de São José do Rio Preto

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

São José do Rio Preto - SP
2022

Dedico aos meus pais, familiares e amigos.

Agradecimentos

Agradeço aos meus pais por tornarem possível esta etapa da minha vida e sempre me encorajarem a perseguir meus sonhos.

A minha família, que sempre confiou em minhas capacidades.

Aos docentes, que dedicaram momentos de suas vidas para repassar seus conhecimentos.

Aos meus amigos, por me alegrar, ajudar e escutar nos momentos difíceis.

Ao professor Guido, por toda a ajuda, paciência e apoio fornecido ao longo de todo este trabalho e curso.

E por fim, a todos aqueles, que direta ou indiretamente, me apoiaram de alguma forma neste momento importante de minha vida.

“I have no special talents, I am only passionately curious.”

Albert Einstein

Resumo

SATAKE JUNIOR, E. H. *Processamento E Análise De Sinais Digitais Vozeados Para O Pré-Diagnóstico de Patologias Laríngeas*. 2022. 59p. TCC UNESP 2022.

Deficiências vocais continuam a afetar parcelas significativas da população mundial, no entanto os processos clínicos tradicionais são comumente invasivos e submetem pacientes a possíveis traumas. Este trabalho, desenvolve um método computacional, utilizando análises acústicas e técnicas de processamento de sinais, que permite detectar a presença de patologias laríngeas por meio de sinais digitais de voz, e os classificar em saudáveis ou patológicos. São utilizadas para a discriminação dos sinais as características de fator de perturbação direcional (DPF), perturbação média relativa (RAP) e fator de *jitter* (JF), enquanto uma máquina de vetores de suporte (SVM) e um algoritmo K-vizinhos mais próximos (KNN) são utilizados como classificadores. Foram utilizados 136 sinais de voz, cuja quantidade de sinais saudáveis e patológicos são iguais, e estes correspondem a casos de Edema de Reinke. Por fim, os testes mostraram uma acurácia global média de até 67% e máxima de 85%, para a SVM, e média de 74% e máxima de 88%, para a KNN. Enquanto a acurácia média de detecção de patologias alcançou 70% e máxima de 82%, para a SVM, e média de 65% e máxima de 88%, para a KNN.

Palavras-chave: Processamento de sinais. Detecção de patologias. Patologias laríngeas. Deficiência vocal. Aprendizado de máquina. Máquina de Vetores de Suporte. K-vizinhos mais próximos. Cepstro.

Abstract

SATAKE JUNIOR, E. H. *Processing And Analysis Of Voiced Digital Signals For Pre-Diagnostic Of Laryngeal Pathologies*. 2022. 59p. TCC UNESP 2022.

Voice disorder continue to affect significant portions of global population, however the traditional clinical processes are commonly invasive and submit patients to potential trauma. This work, develop a computational method, using acoustic analysis and signal processing techniques, which allow to detect the presence of laryngeal pathologies through digital voice signals, and classify them into healthy or pathological. The characteristics of directional perturbation factor (DPF), relative average perturbation (RAP) and jitter factor (JF) are used for signal discrimination, while a support vector machine (SVM) and a K-Nearest Neighbors algorithm (KNN) are used as classifiers. 136 voice signals were used, whose quantity of healthy and pathological signals are the same, and these correspond to cases of Reinke's Edema. Lastly, the tests showed an average global accuracy of 67% and maximum of 85%, for SVM, and average of 74% and maximum of 88%, for KNN. While the average accuracy of pathologies detection reached 70% and a maximum of 82% for SVM, and average of 65% and maximum of 88% for KNN.

Keywords: Signal processing. Pathology Detection. Laryngeal Pathologies. Voice Disorder. Machine Learning. Support Vector Machine. K-Nearest Neighbors. Cepstrum.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

ACC *Accuracy*

DFT *Discrete Fourier Transform*

DPF *Directional Perturbation Factor*

FFT *Fast Fourier Transform*

IDFT *Inverse Discrete Fourier Transform*

IFFT *Inverse Fast Fourier Transform*

JF *Jitter Factor*

KNN *K-Nearest Neighbors*

PCM *Pulse Code Modulation*

RAP *Relative Average Perturbation*

RIFF *Resource Interchange File Format*

SEN *Sensitivity*

SPC *Specificity*

SVD *Saarbrucken Voice Database*

SVM *Support Vector Machine*

WAVE *Waveform Audio File Format*

Sumário

Capítulo 1

Introdução

A voz é uma das mais importantes ferramentas naturais de comunicação da humanidade a qual é utilizada a todo momento nos mais diversos locais e situações do cotidiano. Tal ferramenta é movida por um complexo sistema de geração de voz, constituído principalmente da cooperação sistemática entre o pulmão, laringe e trato vocal, que exige contínuos cuidados a fim de manter sua integridade e desfrutar plenamente do seu uso. Neste âmbito, deficiências vocais se mostram como uma grande barreira para a efetiva comunicação interpessoal e, consequentemente, representa um significativo empecilho no convívio em sociedade.

Em 2020, foi identificado que aproximadamente 17,9 milhões de adultos são afetados por patologias laríngeas nos Estados Unidos [10]. Existem diversas causas possíveis de deficiências vocais sendo as patologias laríngeas e presenças de anomalias no trato vocal as mais comuns [10]. Também se destaca que os métodos clássicos de detecção de anomalias laríngeas são desconfortáveis e traumáticos para os pacientes. Dessa forma, técnicas de detecção por meio do processamento digital dos sinais de voz se mostram uma boa alternativa aos métodos tradicionais.

1.1 Motivação e Justificativas

Devido a naturalidade do uso da voz e de sua presença constante no cotidiano, é comum que ocorram negligências no cuidado e manutenção do sistema de geração de voz, levando ao surgimento de diversas desordens vocais. Além disso, estes problemas apenas se agravam quando as pessoas tendem a postergar a realização de exames, muitas vezes devido a natureza invasiva e, comumente traumatizante, dos procedimentos clínicos para detecção e diagnose de patologias da voz como, por exemplo, a laringoscopia direta e a nasofibrolaringoscopia.

No entanto, assim como diversos outros tipos de patologias, várias desordens vocais podem ser curadas ou estabilizadas mais facilmente quando detectadas e tratadas o mais rápido possível. Desta forma, a motivação deste trabalho se baseia na possibilidade em desenvolver uma técnica não-invasiva de detecção de patologias laríngeas, que consequentemente reduza a sujeição do paciente a traumas, por meio de técnicas computacionais.

1.2 Objetivos

Este trabalho teve por objetivo desenvolver e implementar um método computacional capaz de detectar de forma não-invasiva a existência de patologias laríngeas por meio de amostras de voz. Para isso, se extraiu características destas utilizando técnicas de processamento digital de sinais e após aplicação dos pré-processamentos adequados, foram utilizadas como parâmetros de entrada em classificadores baseados em aprendizado de máquina, utilizando técnicas de validação cruzada.

1.3 Metodologia

A fim de desenvolver este trabalho foram pesquisadas, na literatura, características que pudessem ser extraídas de sinais de voz e utilizadas a fim de detectar patologias laríngeas, assim como as técnicas necessárias para os processos de extração. Em seguida, se obteve 136 sinais digitais de voz, a partir da base de dados livre *Saarbrüecken Voice Database* (SVD) [18], nas quais a quantidade de sinais saudáveis e patológicos foram iguais e cada um deles continham apenas a vogal /a/ sustentada, por alguns segundos, em tonalidade neutra. As vozes patológicas em questão foram afetadas pelo Edema de Reinke. A partir destes sinais se extraiu as seguintes características: fator de perturbação direcional (DPF), perturbação média relativa (RAP) e fator de *jitter* (JF), que foram utilizadas para gerar vetores de características. Estes foram inseridos em um classificador de máquina de vetores de suporte (SVM) e de K-vizinhos mais próximos (KNN), e os resultados obtidos validados através do uso da técnica de validação cruzada, de matrizes de confusão e das métricas de acurácia (ACC), sensibilidade (SEN) e especificidade (SPC).

1.4 Organização do trabalho

O texto deste trabalho está organizado da seguinte forma:

- No capítulo 2, são apresentados os principais conceitos e teorias necessários para a compreensão do trabalho desenvolvido. Também são apresentados alguns trabalhados publicados que envolvem a detecção de patologias laríngeas por meio do uso de técnicas de processamento de sinais, mostrando como são inúmeras as possibilidades de se realizar essa tarefa.

- No capítulo 3 é apresentado, com detalhes, todo o desenvolvimento do trabalho proposto

e de que forma os conceitos discutidos no capítulo anterior foram utilizados.

- No capítulo 4 são relatados todos os resultados obtidos no trabalho, a partir dos testes de validação e classificação que foram realizados.

- No capítulo 5 são apresentadas as conclusões sobre o trabalho.

Capítulo 2

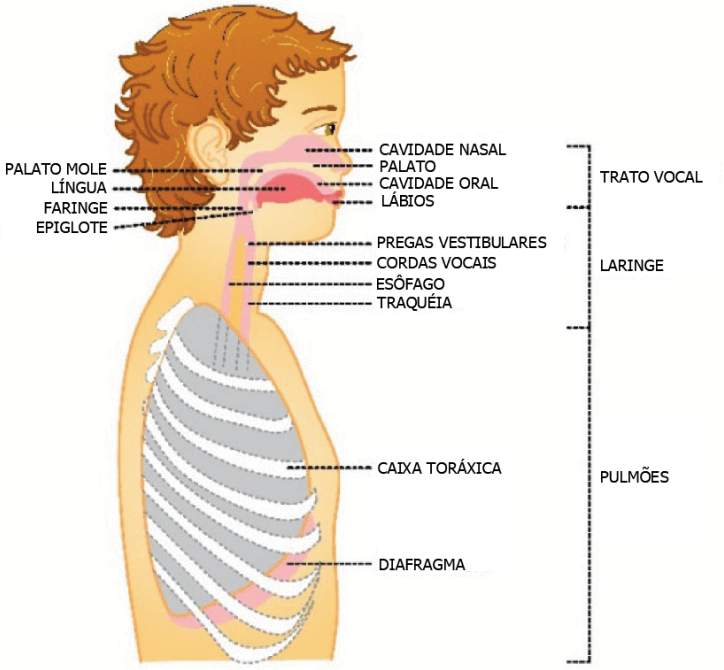
Revisão Bibliográfica

2.1 Sistema de Geração de Voz Humana e Patologias Laríngeas

O sistema de geração de voz humana é composto por três partes principais: o trato vocal, a laringe e o pulmão, como ilustrado na figura 2.1, sendo o último, a fonte de energia do sistema. A geração de voz ocorre quando o ar inspirado pelo pulmão é expelido pela compressão do diafragma, gerando uma corrente de ar estável, controlada pelos músculos da caixa torácica, que percorre da traqueia até a epiglote fazendo com que as cordas vocais vibrem gerando os impulsos sonoros da voz.

Durante a respiração, as cordas vocais se encontram em um estado relaxado e a glote fechada, o que permite que o fluxo de ar, provindo do pulmão, passe sem muita obstrução não gerando vibração significativa das cordas vocais. No entanto, durante a geração de voz, essas podem se encontrar em dois estados denominados: vozeado e não-vozeado. No estado não-vozeado, as cordas vocais se aproximam gerando turbulência ao fluxo de ar. Enquanto no estado vozeado, ou seja na geração de vogais, as cordas vocais se aproxima, ficam tensas e a glote se fecha parcialmente fazendo com que o fluxo de ar seja interrompido pelas cordas, gerando uma onda de pressão quasi-periódica. Os impulsos provocados por esta pressão e a sua frequência são, então, denominados *pitch* e frequência de *pitch*, ou frequência fundamental, respectivamente. Por fim, o trato vocal molda e filtra o som gerado pelo pulmão e laringe

Figura 2.1 – Componentes do sistema humano de geração de voz.



Fonte: Extraído e adaptado de: [10]

produzindo a voz final do sistema [20].

Matematicamente, e com base nestes fatos, um sinal de voz variante no tempo $s(t)$ pode ser representado de forma simplista pela equação de convolução 2.1, na qual $e(t)$ representa os impulsos das ondas de pressão, denominado como a fonte de excitação, e $h(t)$ os efeitos do trato vocal.

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

Patologias laríngeas comumente estão relacionadas com: anomalias nas cordas vocais como nódulos, edemas e atrofia; inflamações como as provocadas pela laringite; e traumas, por exemplo devido a exposição química prolongada. A presença dessas irregularidades acabam por, comumente, causar vozes roucas, *pitchs* anormais, soprosidade, amplitudes instáveis, e diversas outras possibilidades de sintomas [2].

Estas condições se refletem nos efeitos do trato vocal e na fonte de excitação, devido as anomalias na laringe. Desta forma, é possível que análises acústicas, espectrais, entre outras,

do sinal de voz patológico possam fornecer informações suficientes para detecção destas patologias, assim como pode ser visto nos trabalhos relacionados melhor descritos na seção 2.7.

2.2 Análise de Curto-Tempo

As características dos sinais de fala variam ao longo do tempo, inviabilizando o processamento da voz como um sinal digital monolítico, devido a possibilidade de gerar resultados não condizentes com a realidade. No entanto, a forma do trato vocal se modifica relativamente devagar e, portanto, é razoável assumir que para intervalos de tempo muito pequenos, as características da voz não se alteram [16]. Logo, é necessário analisar o sinal em pequenos intervalos, ou seja uma análise de curto-tempo, e, para isso, pode ser utilizada a técnica de janelamento.

O janelamento é um processo de fragmentação de um sinal $s[\cdot]$ em vários blocos, denominados janelas, tradicionalmente de mesmos tamanhos, os quais depende das necessidades da aplicação em que a técnica é utilizada. Por exemplo, no caso da DFT, um tamanho maior irá prover uma melhor resolução da frequência e pior do tempo, enquanto o oposto ocorre para janelas menores.

Neste trabalho, são utilizadas técnicas de transformações de sinais de voz no domínio temporal para o da frequência. No entanto, estas técnicas comumente produzem vazamentos espectrais (do inglês, *spectral leakage*) quando o sinal de entrada não é perfeitamente periódico. Esses vazamentos fazem com que a magnitude das frequências, do sinal transformado para o domínio espectral, se propague para as frequências vizinhas o que gera incertezas no sinal transformado.

Uma forma de minimizar esse problema, é a aplicação de uma função janela para cada um dos fragmentos gerados no processo de janelamento. Essas funções permitem reduzir as amplitudes dos termos mais aos extremos da janela de forma gradual, a fim de reduzir o efeito das descontinuidades do sinal e, consequentemente, a intensidade dos vazamentos. No entanto, a aplicação destas funções pode levar a perda de informações presentes nos extremos de cada

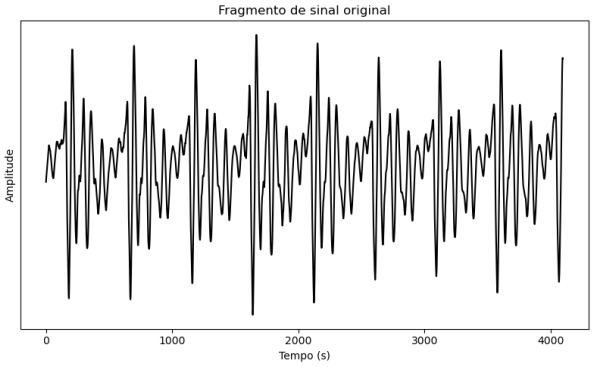
janela. Desta forma, a sobreposição de janelas pode ser adotada para tentar reduzir esta perda de informações. As figuras 2.2 e 2.3 ilustram o processo de janelamento e a aplicação da função janela.

Após estes pré-processamentos é possível realizar, então, a análise de curto-tempo através da equação de convolução 2.2:

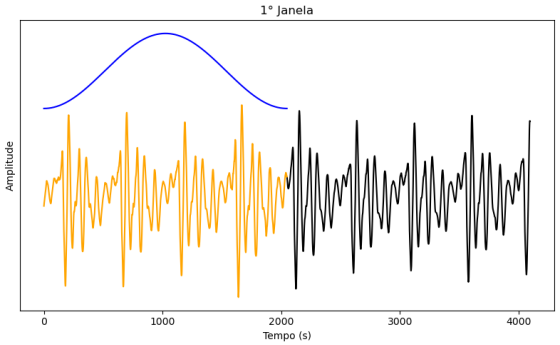
$$X_n = \sum_{m=-\infty}^{\infty} T\{s[n]w[n-m]\}, \quad \text{tal que } \exists s[n], \exists w[n-m], \tag{2.2}$$

Onde n representa um índice de tempo do sinal completo, X_n um parâmetro analisado neste instante, m o índice do somatório da convolução, $T\{\ \}$ um operador que define a natureza da análise e w uma função de janelamento.

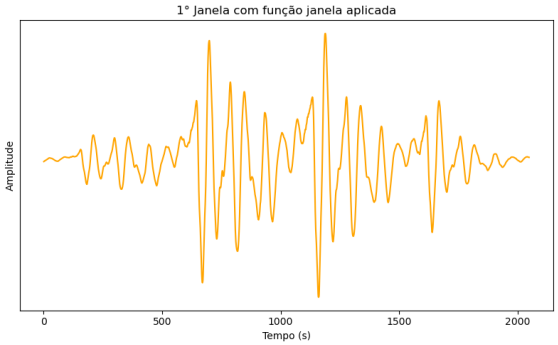
Figura 2.2 – Processo de janelamento e aplicação de função janela.



(a) Fragmento original de um sinal.



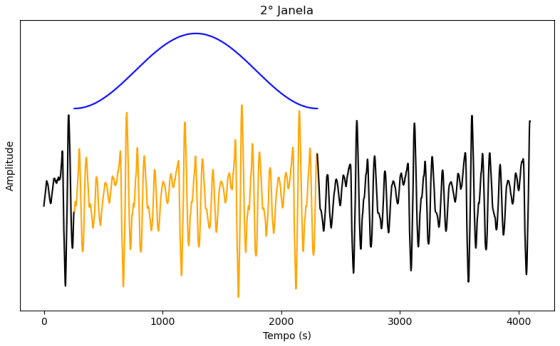
(b) 1^a Janela do fragmento (a) em laranja e função janela hamming em azul.



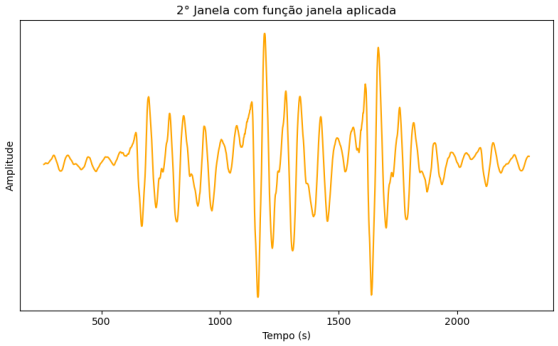
(c) Função janela w aplicada a janela (b).

Fonte: Confeccionado pelo autor.

Figura 2.3 – Segunda janela do processo de janelamento e aplicação de função janela da figura 2.2.



(a) 2^a Janela do fragmento (a) da figura 2.2 em laranja e função janela hamming em azul.



(b) Função janela w aplicada a janela (a).

Fonte: Confeccionado pelo autor.

2.3 Análise Espectral

Matematicamente, um sinal de voz é representado como uma função variante no tempo. No entanto, algumas características úteis podem não ser facilmente observáveis no domínio do tempo. Nestes casos, pode ser interessante transformá-lo para o domínio da frequência.

Uma forma de realizar a transformação de um sinal discreto no domínio do tempo para o da frequência é através da Transformada Discreta de Fourier (DFT) [16], que pode ser calculada pela equação 2.3, e sua inversa pela 2.4:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}} \tag{2.3}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{\frac{j2\pi kn}{N}} \tag{2.4}$$

Nas equações 2.3 e 2.4, é possível notar que a DFT, embora útil, possui uma complexidade de ordem $O(N^2)$ e, desta forma, o seu custo computacional aumenta significativamente para

sinais muito grandes. Para contornar este problema Cooley e Tukey propuseram e desenvolveram o algoritmo *radix-2* FFT, de complexidade $O(N \log N)$, para calcular a DFT para sinais de tamanho N iguais a potências de 2 [6, 7].

O método proposto pelos autores, em sua forma mais básica, se baseia em separar a DFT do sinal $x[n]$, vista na equação 2.3, em duas partes: na soma dos índices n pares e na dos ímpares [7]. Como pode ser visto na equação 2.5:

$$X[k] = \sum_{m=0}^{N/2-1} x[2m] e^{-\frac{2j\pi mk}{N/2}} + e^{-\frac{2j\pi k}{N}} \sum_{m=0}^{N/2-1} x[2m+1] e^{-\frac{2j\pi mk}{N/2}} \tag{2.5}$$

Devido a periodicidade dos exponenciais complexos definida pela fórmula de Euler, os elementos restantes $X[k + \frac{N}{2}]$ da DFT também podem ser obtidos pela equação 2.6:

$$X[k + \frac{N}{2}] = \sum_{m=0}^{N/2-1} x[2m] e^{-\frac{2j\pi mk}{N/2}} - e^{-\frac{2j\pi k}{N}} \sum_{m=0}^{N/2-1} x[2m+1] e^{-\frac{2j\pi mk}{N/2}} \tag{2.6}$$

Desta forma, neste trabalho, será utilizado o algoritmo mencionado para realizar o cálculo da DFT e da IDFT. O funcionamento do algoritmo *radix-2* FFT e IFFT pode ser visto nos pseudocódigos 1 e 2 respectivamente:

Algoritmo 1: Radix-2 FFT

Entrada: x, N // x: Amplitudes do sinal no domínio do tempo, N: Tamanho do sinal
Saída: X // X: Magnitudes das frequências do sinal

se *N não é potência de dois* **então**
    retorna x
$x_{\text{par}}[0], x_{\text{par}}[1], ..., x_{\text{par}}[N/2] = x[0], x[2], x[4], ..., x[2m]$ // Amplitudes dos índices pares;
$x_{\text{impar}}[0], x_{\text{impar}}[1], ..., x_{\text{impar}}[N/2] = x[1], x[3], x[5], ..., x[2m+1]$ // Amplitudes dos índices ímpares;
X_{par} = Radix-2 FFT(x_{par}, $N/2$) // Chamada recursiva;
X_{impar} = Radix-2 FFT(x_{impar}, $N/2$);
para $k = 0$ *até* $N/2$ **faça**
    $X[k] = X_{par}[k] + (X_{impar}[k]\ e^{-j2\pi k/N})$ // $j = \sqrt{-1}$;
    $X[k + N/2] = X_{par}[k] - (X_{impar}[k]\ e^{-j2\pi k/N})$;
retorna X

Algoritmo 2: Radix-2 IFFT

Entrada: X, N // X: Magnitudes das frequências do sinal (complexo), N: Tamanho do sinal

Saída: x // x: Amplitudes do sinal no domínio do tempo

se *N não é potência de dois* **então**
  retorna X
$X_{\text{par}}[0], X_{\text{par}}[1], ..., X_{\text{par}}[N/2] = X[0], X[2], X[4], ..., X[2m]$ // Magnitudes das frequências dos índices pares;
$X_{\text{impar}}[0], X_{\text{impar}}[1], ..., X_{\text{impar}}[N/2] = X[1], X[3], X[5], ..., X[2m+1]$ // Magnitudes das frequências dos índices ímpares;
x_{par} = Radix-2 IFFT(X_{par}, $N/2$) // Chamada recursiva;
x_{impar} = Radix-2 IFFT(X_{impar}, $N/2$);
para $k = 0$ *até* $N/2$ **faça**
  $x[k] = (x_{par}[k] + (x_{impar}[k]\ e^{j2\pi k/N}))/2$ // $j = \sqrt{-1}$;
  $x[k + N/2] = (x_{par}[k] - (x_{impar}[k]\ e^{j2\pi k/N}))/2$;
retorna X

2.4 Análise Cepstral

Em vários tipos de problemas é necessário, ou interessante, se obter a frequência fundamental f_0 de um sinal de voz, uma vez que pode ser utilizada no cálculo de características do sinal. Essa determinação pode ser feita através do cepstro de potência $C_p(\tau)$ [14], definido como

$$C_p(\tau) = |\mathcal{F}\{\log|S(\omega)|^2\}|^2 \tag{2.7}$$
$$S(\omega) = \mathcal{F}\{s(t)\}$$

por Bogert et al. [4], onde $\mathcal{F}$ representa a transformada de Fourier e $s(t)$ o sinal de voz, ou através do cepstro real $C_r(\tau)$ [20], derivado do cepstro complexo $C_c(\tau)$ proposto por Oppenheim, descartando as informações das fases $j\theta(\omega)$ [15, 17]. Neste trabalho, será utilizado o cepstro

real.

$$C_c(\tau) = \mathcal{F}^{-1}\{\log \mathcal{F}\{s(t)\}\} \tag{2.8}$$

$$= \mathcal{F}^{-1}\{\log |S(\omega)| + j\theta(\omega)\}$$

$$C_r(\tau) = \mathcal{F}^{-1}\{\log |S(\omega)|\} \tag{2.9}$$

A aplicação da DFT transforma, inicialmente, o sinal no domínio temporal para o domínio espectral. Os efeitos do trato vocal, se manifestam como picos em baixas frequências representando as ressonâncias, enquanto os da fonte de excitação se manifestam como picos em frequências maiores representando os harmônicos [14].

Em seguida, pela transformada inversa, o sinal é passado para o domínio da quefrequência, que representa uma medida de tempo, mas não no mesmo sentido do domínio temporal, relacionada a frequência [4], tal que, sua relação para um sinal discreto de áudio pode ser dada por

$$\frac{f_{max}}{\tau} = f_\tau \tag{2.10}$$

onde f_{max} representa a taxa de amostragem do sinal e f_τ a frequência correspondente a τ-ésima quefrequência τ. Neste domínio, a periodicidade dos harmônicos se manifestam como um curto pico localizado próximo à quefrequência correspondente a f_0, enquanto as ressonâncias se manifestam como picos mais largos em baixas quefrequências [14]. A f_0 pode, então, ser obtida pela equação 2.10, encontrando a quefrequência do pico gerado pelos harmônicos o qual para sinais vozeados possui uma magnitude nitidamente maior que os demais.

No entanto, assim como sinais de voz são variantes no tempo, a f_0 também varia. Logo, é necessário que o cepstro seja aplicado em pequenos intervalos a fim de obter f_0 válidas. Para isso o cepstro deve ser aplicado através da análise de curto-tempo [13, 14].

2.5 Medidas de Perturbação

Neste trabalho, foram utilizadas três medidas de perturbação para a análise de sinais de voz saudáveis e patológicos. Sendo elas: o Fator de Perturbação Direcional (DPF), a Perturbação Média Relativa (RAP) e o Fator de *Jitter* (JF).

O DPF é uma medida proposta por Hecker [9], que mede a perturbação dos períodos de *pitch* considerando a direção das suas mudanças. Esta medida é definida como a porcentagem da quantidade total de diferenças de períodos em que ocorre uma mudança do sinal algébrico. A contagem dessas mudanças é feita da seguinte forma: o primeiro período é considerado como um referencial; se o segundo for menor, então é atribuído um sinal negativo a diferença, caso contrário, um positivo é atribuído. Em seguida, o segundo período passa a ser considerado o referencial e a sua diferença com o período seguinte é verificado. O processo se repete até a última diferença existente.

Dessa forma, é obtida a quantidade de mudanças algébricas do sinal, que será representado por $QMAS$, e o parâmetro pode ser computado segundo a equação 2.11, na qual N representa a quantidade de períodos de *pitch*.

$$DFP = \frac{QMAS}{N-1}100 \tag{2.11}$$

A RAP é uma medida proposta por Koike [11], que mede a flutuação dos período de *pitch*. É a razão da diferença absoluta média, entre um período e a média deste período com os seus dois vizinhos mais próximos, com o período médio. Esta medida é dada pela equação 2.12, onde T_i representa o período do sinal no i-ésimo intervalo de tempo e N a quantidade de períodos medidos.

$$RAP = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{T_{i-1}+T_i+T_{i+1}}{3} - T_i\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \tag{2.12}$$

O JF é uma medida que fornece uma relação entre a média das perturbações da frequência fundamental a partir da média dessas frequências. Esta medida é definida pela equação 2.13, onde F_i representa a frequência fundamental no i-ésimo intervalo de tempo e N a quantidade de

frequências fundamentais medidas.

$$JF = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|F_i - F_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N}F_i}10^2 \tag{2.13}$$

2.6 Classificação

No âmbito do aprendizado de máquina, a classificação é definida como o problema em se identificar a qual das diversas classes, ou categorias, definidas um certo dado, ou no caso deste trabalho: uma amostra de voz, pertence. Existem três etapas principais no processo de classificação: a etapa de treinamento, teste e validação [22].

Na primeira, um modelo de classificação é treinado inserindo dados de treinamento em um algoritmo de aprendizado, o qual pode ou não ser supervisionado. Na segunda, o modelo treinado na etapa anterior é então utilizado para tentar classificar dados de testes. E por fim, na etapa de validação o modelo treinado é analisado através das medidas estatísticas dos resultados dos testes, e os parâmetros do modelo são ajustados, retornando para a primeira etapa se for necessário melhorar o modelo.

2.6.1 Máquina de Vetores de Suporte (SVM)

No problema de classificação, as Máquinas de Vetores de Suporte (SVM) são modelos de aprendizado supervisionado que, por meio da análise de dados por algoritmos de aprendizado, são capazes de detectar padrões em um conjuntos de dados [12]. Ela se baseia na estratégia de definir o melhor hiperplano que seja capaz de classificar novos dados em uma dentre duas classes, sendo categorizada como um classificador binário.

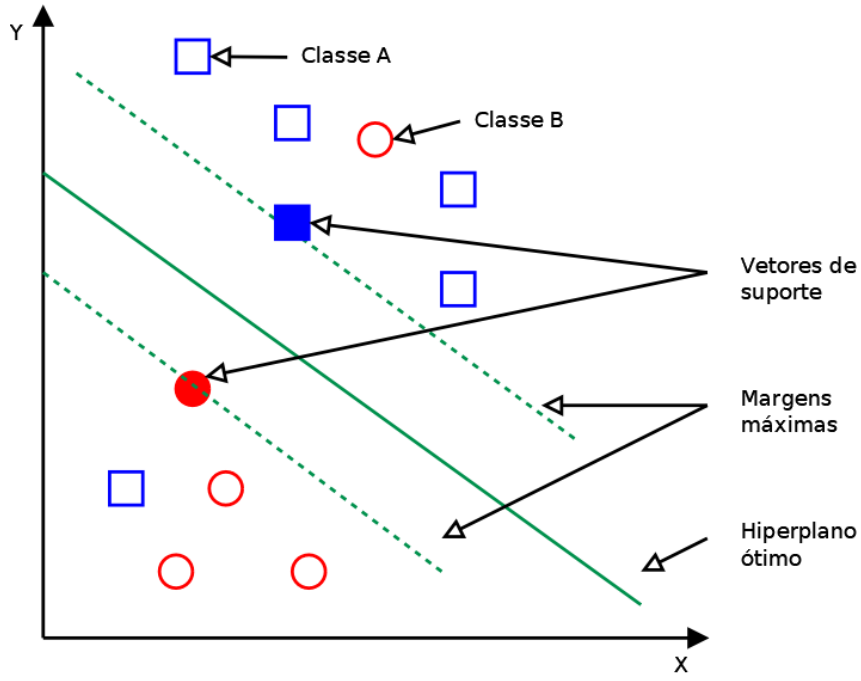
A obtenção deste hiperplano é realizada com a utilização de um conjunto de dados de treino

que são analisados pelo algoritmo de aprendizado, por meio de uma função de decisão, a fim de detectar um conjunto de dados de treinamento que melhor definem um possível hiperplano na dimensão estabelecida pelas características utilizadas na classificação. Estes dados escolhidos se encontram mais próximo da superfície separadora do que os outros e são denominados vetores de suporte. Eles delimitam uma margem em volta do hiperplano separador que visa melhor afastar os elementos de classes diferentes.

Desta forma, a SVM tenta determinar o hiperplano de forma a colocar a maior quantidade possível de dados de uma mesma classe no mesmo lado, enquanto maximiza a margem definida pelos vetores de suporte, como pode ser visto na figura 2.4.

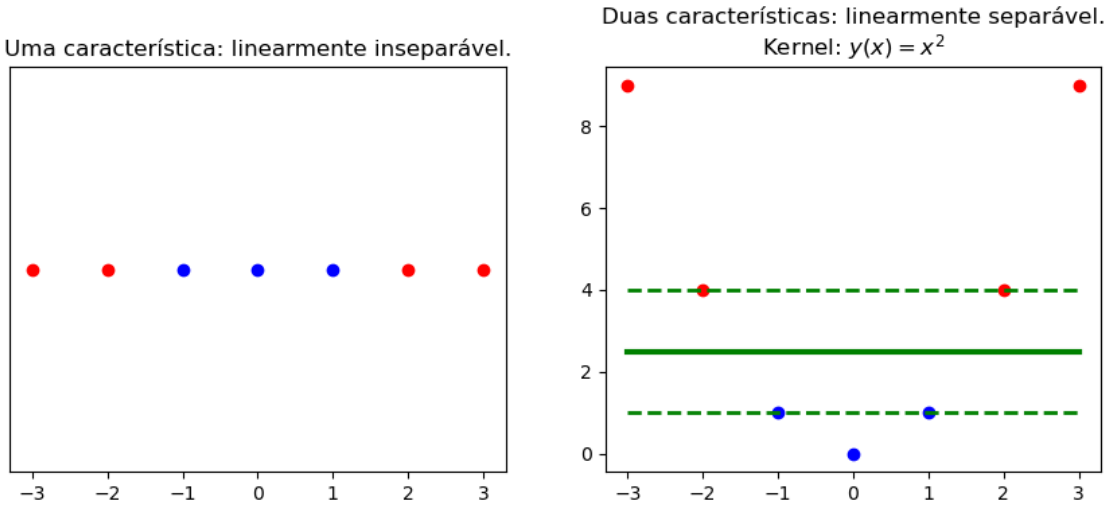
No entanto, as vezes um conjunto de dados não pode ser linearmente separáveis por um hiperplano simples. A fim de contornar o problema é possível, então, criar uma nova dimensão através da aplicação de uma função de transformação, denominada função *kernel*, nos pontos dos dados de treino como ilustrado na figura 2.5. Neste trabalho, a SVM é utilizada para tentar classificar as vozes em saudáveis ou patológicas.

Figura 2.4 – Hiperplano de uma SVM que separa duas classes definidas, uma em azul e outra em vermelho, e as margens e vetores que o definem.



Fonte: Confeccionado pelo autor.

Figura 2.5 – Separação de duas classes de dados unidimensionais, uma em azul e outra em vermelho. [A esquerda]: dados unidimensionais linearmente inseparáveis. [A direita]: dados linearmente separáveis devido a nova dimensão gerada pela função *kernel* $y(x) = x^2$.



Fonte: Confeccionado pelo autor.

2.6.2 K-Vizinhos Mais Próximos

No problema de classificação, o algoritmo dos K vizinhos mais próximos é um método de classificação que utiliza informações acerca da vizinhança geográfica de uma nova amostra sendo classificada para decidir a qual classe ela pertence, ao invés de procurar por limites lineares ou não-lineares capazes de separar as classes presentes [12].

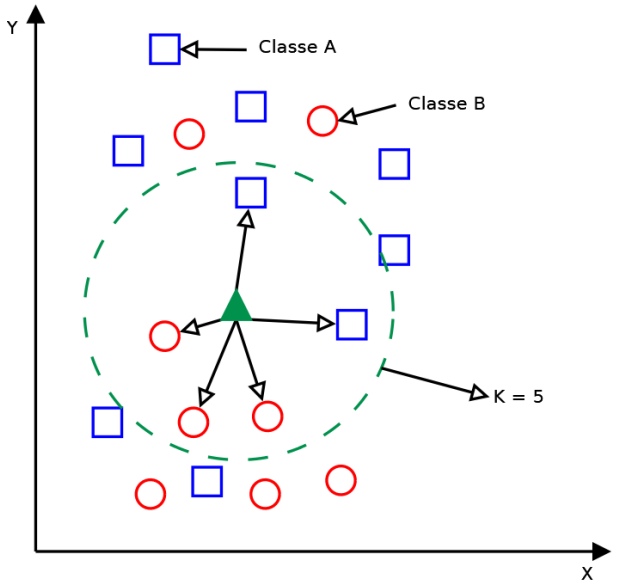
Isso é realizado por meio do cálculo de um valor de distância, como a euclidiana, manhattan e de minkowski, entre a amostra a ser classificada e as pertencentes no conjunto de treinamento, juntamente com a utilização de um parâmetro de afinação, representado por K, que define a quantidade de distâncias a serem utilizadas no processo de decisão da KNN.

O processo é realizado mediante a seleção das K amostras de treino que possuam a menor distância em relação a qual está sendo classificada. As K amostras, então, realizam um processo de votação para decidir a qual classe a nova amostra pertencerá tal que, cada uma irá votar em

sua própria classe. Desta forma, a nova amostra irá pertencer a classe mais votada pelas K amostras de treinamento mais próximas, conforme ilustrado na figura 2.6.

No entanto, devido a K ser um parâmetro de afinação, não há uma fórmula analítica para se obter o seu valor apropriado. Assim, é necessário que se verifique experimentalmente os possíveis valores a serem adotados. Porém, é um fato conhecido que valores muito grandes costumam gerar excessos de generalização, e muito pequenos casos de sobreajuste [12]. Além disso, devido ao fato da KNN utilizar valores de distância é importante que se realize a normalização das características utilizadas no cálculo da distância, a fim de evitar a presença de viés a favor das características com escalas maiores.

Figura 2.6 – Dados de duas classes diferentes, uma em azul e outra em vermelho, com a distância máxima considerada para K = 5, e o novo ponto a ser classificado em verde. As setas saindo do novo ponto representam os votos.



Fonte: Confeccionado pelo autor.

2.6.3 Validação Cruzada

A validação cruzada, no âmbito dos problema de classificação, é uma técnica de validação de modelos que permite testar o quão bem o modelo de classificação consegue prever as classes

corretas de novos dados de entrada [12]. A fim de realizar esta análise, o conjunto de dados é divido em dois subconjuntos: um de treinamento e outro de teste.

A técnica é composta de duas etapas: na primeira, uma técnica de aprendizado de máquina é utilizada no conjunto de treinamento para gerar um modelo de classificação treinado, enquanto, na segunda, o modelo gerado é aplicado no conjunto de treinamento e a quantidade de acertos e erros é obtida. Desta forma, é possível calcular a acurácia e outras medidas estatísticas que servirão de base para decidir se o modelo classifica bem, ou não, dados que não foram utilizados no treinamento.

Para aumentar a confiabilidade do modelo, e consequentemente dos resultados, os subconjuntos de treino e teste podem ser modificados e utilizados em outras iterações de treinamento e teste, a fim de considerar combinações diferentes de dados e detectar se o modelo contém viés ou sobre-ajuste.

2.6.4 Matriz de Confusão e Métricas

A matriz de confusão, também conhecida como matriz de contingência, é basicamente uma tabela que permite a visualização da performance de um algoritmo de aprendizado de máquina. Esta matriz é sempre quadrada de ordem N, na qual N é a quantidade de classes definidas para o problema específico. Suas colunas representam as classes esperadas, ou seja, a classificação conhecida, e suas linhas as classes preditas pelo classificador. No entanto a configuração oposta também pode ser utilizada.

A matriz se inicia preenchida por zeros e para cada dado de teste, inserido no classificador, a entrada correspondente ao resultado da classificação obtida é atualizada conforme a figura 2.7. Desta forma, é possível verificar pelas diagonais quantos dados foram correta e incorretamente classificados para o modelo treinado. Assim, é evidente que os valores da diagonal principal da figura 2.7 devem ser maximizados para um melhor resultado.

Figura 2.7 – Matriz de confusão.

		Verdadeira Classe	
	População Total (P + N)	Classe Positiva Conhecida (P)	Classe Negativa Conhecida (N)
Classe Predita	Predito Positivo (PP)	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Predito Negativo (PN)	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Confeccionado pelo autor.

Foram utilizadas três métricas para a análise dos resultados obtidos: a acurácia, a sensibilidade e a especificidade.

A acurácia, ou acurácia global, é uma medida de quantas amostras estão sendo corretamente classificadas em relação a população total e é calculada pela equação 2.14:

$$\text{ACC} = \frac{VP + VN}{P + N} \tag{2.14}$$

Esta medida indica quantos acertos o classificador obteve, mas não permite detectar se uma classe está sendo melhor classificada do que a outra.

A sensibilidade é uma razão da quantidade de amostras de teste pertencentes a classe positiva, ou patológica, que foram corretamente classificadas e é calculada pela equação 2.15:

$$\text{SEN} = \frac{VP}{P} \tag{2.15}$$

Enquanto a especificidade é idêntica à sensibilidade, porém, para amostras de teste pertencentes a classe negativa, ou saudável, e é calculada pela equação 2.16:

$$\text{SPC} = \frac{VN}{N} \tag{2.16}$$

Estas medidas permitem verificar a performance do classificador para cada uma das classes

separadamente, desta forma, complementando a informação não fornecida pela acurácia.

2.7 Trabalhos Relacionados

Nesta subseção, estão dispostos alguns trabalhos relacionados com o problema abordado e técnicas utilizadas para fim de referência e exposição do estado da arte. Dentre estes, serão apresentados: um trabalho de conclusão de curso, uma teses e três artigos.

No trabalho realizado por Sato [19], o objetivo foi propor, elaborar e desenvolver um algoritmo para a detecção não-invasiva de patologias laríngeas. Foram utilizadas a técnica de autocorrelação, para a obtenção da média e variância das distâncias entre picos do sinal de voz, e a variância de entropia do sinal. Após extração das características, foi utilizado um classificador SVM, tal que 15 vozes saudáveis e 15 patológicas foram utilizadas para o treinos e testes, utilizando a técnica de validação cruzada do tipo *hold-up*. Os resultados indicaram uma acurácia global de 73.33% para o método proposto.

Fonseca *et al* [8] propõem um algoritmo para discriminação de vozes saudáveis e patológicas que utiliza a transformada *wavelet* discreta de Daubechies (DWT-db) e os coeficientes de predição lineares (LPC) para obtenção de características de sinais de voz, e a máquina de vetores de suporte por mínimos quadrados (LS-SVM) como opção de classificador. No trabalho foram utilizadas 60 amostras das quais 48 foram usadas no treinamento e 12 na validação, tal que metade de cada conjunto era de vozes patológicas. Os resultados, então, apontaram uma acurácia por volta de 91% e uma baixa complexidade computacional relacionada ao comprimento do sinal de voz.

No artigo de Chen *et al* [5], é proposto um novo método de classificação de vozes patológicas baseado na transformada de Hilbert-Huang (HHT) e nos coeficientes cepstrais de predição linear (LPCC), juntamente com um classificador k-vizinhos mais próximos (KNN). O método se baseia em suavizar os sinais e decompor as mudanças de tendências de diferentes escalas nas, então, denominadas *Intrinsis Modal Functions* (IMFs), por meio da *Empirical Mode*

Decomposition (EMD). São, então, obtidas 12 características das IMFs e nove dos LPCCs. Por fim, os resultados demonstram uma acurácia de 93.3% e boa confiabilidade para o método proposto.

Teixeira *et al* [21] utilizaram dos conceitos de *jitter* e *shimmer* relativos, relação harmônico-ruído (HNR) e dos coeficientes cepstrais na frequência de Mel (MFCC) na detecção e classificação de vozes patológicas por meio do uso de uma SVM. O estudo utilizou 473 amostras de voz, sendo 279 delas compreendidas entre três tipos de patologias: disfonia, laringite crônica e paralisia das cordas vocais. Foram adotados três grupos: (I) grupo consistindo nos parâmetros de *jitter*, *shimmer* e HRN para vogais sustentadas, (II) consistindo nos MFCCs de vogais sustentadas e (III) consistindo nos MFCCs de uma sentença em alemão. Os resultados demonstraram que o melhor resultado foi obtido para o grupo (I) com uma acurácia de 71%.

No trabalho de Alves *et al* [1], foi estudada a possibilidade de detecção de patologias relacionadas com as pregas vocais por meio do uso de características cepstrais multibanda, da vogal sustentada /a/, em dois tipos de classificadores: SVM e KNN. Foram utilizadas as características: MFCCs, distâncias cepstrais, diferenças de amplitude (DAP) e quefrequência (DQP) entre os dois primeiros picos cepstrais, a energia desses picos (EP1 e EP2) e a enegia cepstral entre esses picos (EEP). A obtenção do MFCCs se baseou na decomposição dos sinais de voz em sub-bandas, por meio da aplicação de transformadas *wavelet*, e realização de análises cepstrais. Como entrada dos classificadores foram utilizadas 21 características: 13 MFCCs, DAP, DQP, EP1, EP2, EEP e três distâncias cepstrais. O conjunto de dados foi organizado em seis pares de subconjuntos de voz sendo eles: patológicas/controle, nódulo/controle, edema/controle, neurológicas/controle, nódulo/neurológicas, edema/neurológicas e edema/nódulo. Os resultados foram ,então, obtidos por meio da utilização de uma validação cruzada do tipo *leave-one-out* na qual os quatro primeiros pares obteram uma acurácia de 100% e o restante de 99.08%, 98.86%, e 88.72%, respectivamente.

Este trabalho, então, segue o mesmo objetivo dos trabalhos citados anteriormente, buscando um método não-intrusivo de detecção de patologias laríngeas por meio de análises acústicas de sinais de voz, visando identificar e obter características que suficientemente discriminem vozes saudáveis de patológicas. Tal que, para a classificação, será adotada como base um classificador

SVM e um KNN, cuja confiabilidade será verificada por meio das medidas de sensibilidade, especificidade, e acurácia.

Ainda neste trabalho é utilizada a técnica da análise cepstral para a detecção de frequências fundamentais de forma mais precisa que a técnica de autocorrelação utilizada por Sato [19] e menos custosa do que os demais trabalhos enunciados. Também são utilizadas características mais simples como a DPF, RAP e JF, com o intuito de reduzir o custo computacional da extração de características presente nos trabalhos [1, 5, 8, 21].

Capítulo 3

Metodologia

Neste capítulo serão detalhadas as etapas do método proposto para a realização deste trabalho, se utilizando dos conceitos apresentados no capítulo 2. Todas as etapas do desenvolvimento serão descritas, sendo elas: a obtenção das amostras de voz utilizadas, a extração dos seus dados brutos, a verificação da consistência dos dados extraídos, a extração de características, preparação e utilização dos dados para a classificação e descrição do processo de análise dos resultados.

O objetivo da execução destas etapas é desenvolver um método computacional capaz de detectar vozes patológicas e, desta forma, distinguir, e consequentemente classificar, sinais de voz patológicos de saudáveis, por meio de técnicas computacionais. Para este fim, são adotados arquivos de áudio no formato WAVE.

3.1 Coleta de Dados

Devido ao autor não possuir acesso à pessoas com condições patológicas de voz, os sinais de áudio foram obtidos por meio do *Saarbrüecken Voice Database* (SVD), uma base de dados livre que contêm diversas amostras de vozes saudáveis e patológicas, no idioma alemão, mantida pela Universidade do Sarre, em Sarbruque na Alemanha. Todas as amostras da base possuem laudo

médico especializado, comprovando a veracidade dos dados.

Neste trabalho, primeiramente, foram coletados 136 sinais vozeados de diferentes pessoas, idades e sexos, tal que, 50% correspondem a vozes saudáveis, enquanto o restante a patológicas. Cada sinal é composto pela vogal /a/, no idioma alemão, sustentada por aproximadamente um a dois segundos, amostrado em 50kHz, quantizado em 16 *bits*, e armazenado no formato WAVE sem compressão.

As vozes patológicas em questão, correspondem a casos clinicamente comprovados de Edema de Reinke. Este edema é caracterizado por uma lesão difusa que surge na camada superficial da prega vocal, na qual é comum que apresente acúmulo de fluidos. Esta patologia apresenta grande correlação com o uso intensivo da voz, abusos vocais e tabagismo [3].

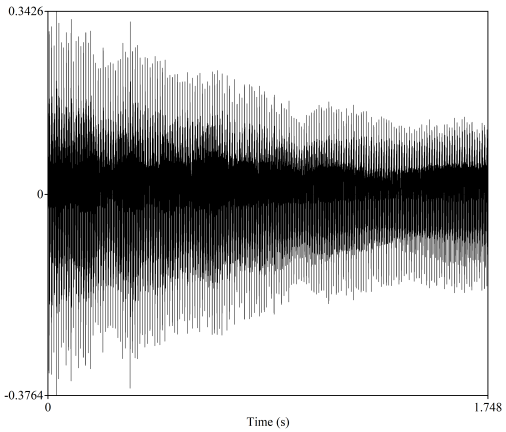
3.2 Extração dos Dados Brutos e Verificação de Consistência

Arquivos WAVE possuem cabeçalhos, no entanto, eles não são necessário para a análise dos sinais de voz. Desta forma, após a coleta dos sinais, foi utilizada uma rotina para extrair os dados brutos, ou amplitudes, dos seus respectivos arquivos. Esta rotina gera como saída um arquivo de texto puro contendo as amplitudes do sinal em ordem temporal ascendente. Desta forma, possibilitando visualizar a forma de onda de cada sinal, quando representado visualmente em um gráfico de amplitude por tempo.

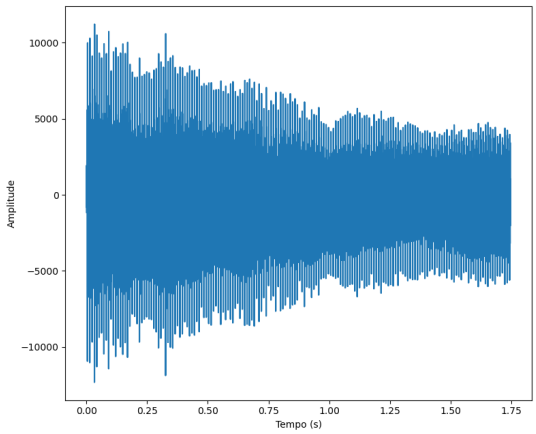
Em seguida, a fim de averiguar se os dados foram corretamente extraídos e, desta forma, comprovar que o sinal obtido corresponde ao original, as formas de ondas obtidas foram comparadas com as visualizações do programa de edição e análise de áudio *Praat*, o qual é bem aceito pela comunidade científica. A comparação, então evidenciou que os sinais extraídos são equivalente aos originais e, portanto, não houve erros na extração.

A figura 3.1 ilustra a comparação para um sinal de voz específico, dentre os obtidos na base de dados SVD.

Figura 3.1 – Comparação entre as formas de onda de um arquivo de áudio original e a representação gráfica das suas amplitudes, extraídas pela rotina C/C++.



(a) Um dos sinais de voz obtidos representado graficamente pelo prorama *Praat*.



(b) Sinal de voz da figura (a), extraído pela rotina C/C++ e representado graficamente pela biblioteca Python *matplotlib*.

Fonte: Confeccionado pelo autor.

3.3 Extração de Características

Após verificar que os sinais extraídos são consistentes com os arquivos de áudio originais, foi possível iniciar a extração das características escolhidas para os sinais selecionados. Como pode ser visto na seção 2.5, as características escolhidas medem as perturbações do período de *pitch* e da frequência fundamental dos sinais de voz ao longo do tempo. Desta forma, é implementada a análise cepstral de curto-tempo para cada um dos sinais, tal que é definida pela

equação 3.1, originada da união das equações 2.2 e 2.9, na qual n representa a n-ésima janela.

$$C_r(n,\tau) = \mathcal{F}^{-1}\{\log|\mathcal{F}\{s(t)w(t-m)\}|\} \tag{3.1}$$

$$C_r(n,\tau) = X_n$$

$$\mathcal{F}^{-1}\{\log|\mathcal{F}\{\ \}|\} = \sum_{m=-\infty}^{\infty} T\{\ \}$$

Para essa implementação, foram utilizados os seguintes parâmetros na análise de curto-tempo:

- Tamanho da janela: 2048 pontos, equivalente a $\frac{2048}{50kHz} = 40,96ms$ do sinal. Esse tamanho foi escolhido por abordar um intervalo razoável para a análise e permitir a aplicação da FFT por ser uma potência de 2;
- Tamanho do passo: 256 pontos, equivalente a $\frac{500}{50kHz} = 5,12ms$ do sinal. Desta forma, o sinal em questão terá sua frequência fundamental verificada a cada 5,12ms do sinal;
- Tamanho da sobreposição: 1792 pontos, equivalendo a uma sobreposição de 87,5% da janela. Apenas os 1792 pontos mais a esquerda de uma janela se sobrepõem a antecessora.
- Função janela w utilizada: Hamming, definida como

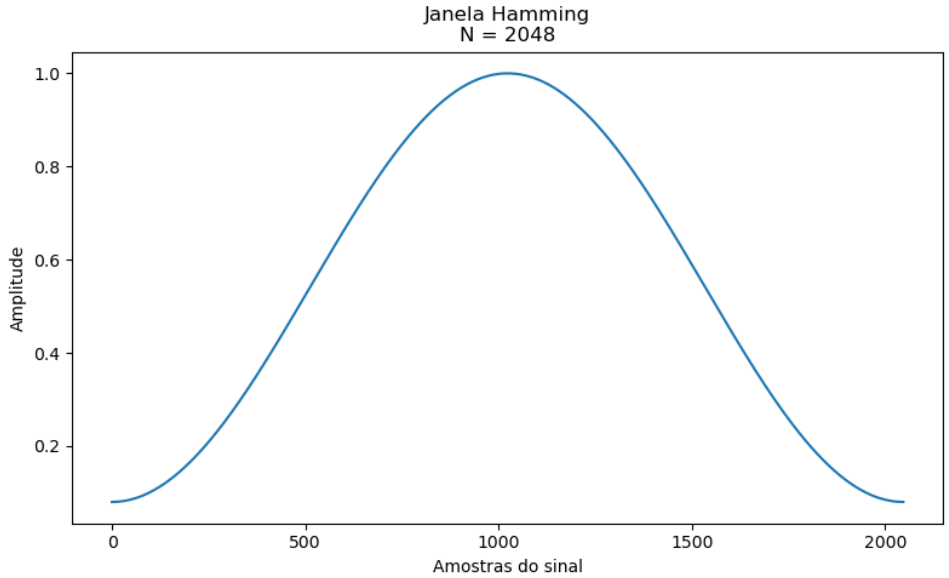
$$w(n) = 0,54 - 0,46\cos\left(\frac{2\pi n}{N-1}\right); \quad 0 \leq n < N;\ N = 2048.$$

Assim, foram calculados o espectro e cepstro real para cada uma das janelas, de todos os sinais selecionados. O maior pico de amplitude no cepstro foi, então, procurado no intervalo de quefrequência [225, 700], buscando, assim, a frequência fundamental dentro do intervalo [71Hz, 222Hz], por meio da equação 2.10. Desta forma, foi obtido um valor de frequência fundamental (f_0) a cada 5,12ms de duração do sinal, enquanto os períodos de *pitch* (T_0) correspondentes foram obtidos por meio da relação:

$$T_0 = \frac{1}{f_0}$$

Nas figuras 3.2, 3.3 e 3.4, é possível observar, respectivamente, a função janela de Hamming, uma das janelas de espectro do sinal da figura 3.1(b) e o seu cepstro normalizado de forma que só tenha amplitudes nulas ou positivas, sendo que neste último o pico da maior amplitude está bem visível.

Figura 3.2 – Função janela de Hamming, para um sinal de 2048 amostras de amplitudes.



Fonte: Confeccionado pelo autor.

Figura 3.3 – Uma das janelas do espectro de um sinal de voz.

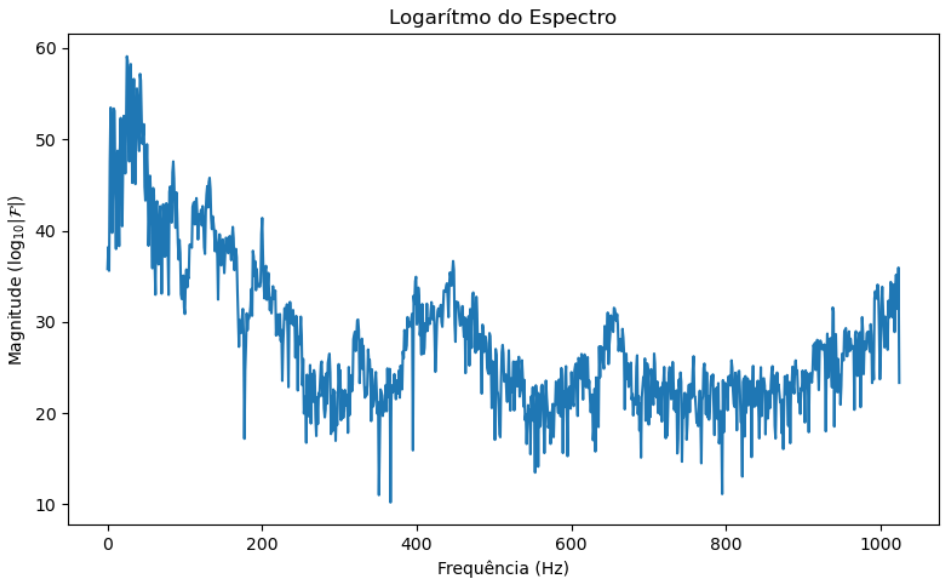
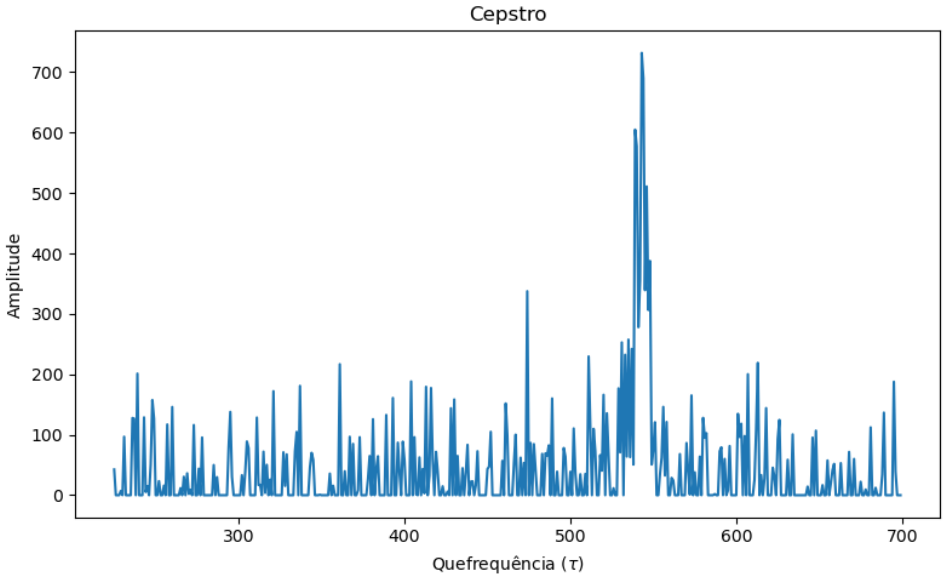


Figura 3.4 – Normalização de uma das janelas do cepstro de um sinal de voz, com um pico no índice 543 da quefrequência, indicando uma frequência fundamental de 92,081Hz e período de *pitch* de 0,1086ms aproximadamente.



Em seguida, foram calculados os valores de DPF, RAP e JF, utilizando as equações 2.11, 2.12 e 2.13, a partir dos valores de frequência fundamental e período de *pitch* calculados na análise cepstral. Estes valores foram combinados para construir quatro conjuntos de vetores de características diferentes, descritos na tabela 3.1, a fim de verificar se existia alguma separação linear possível entre os sinais saudáveis e patológicos utilizando algum dos conjuntos sem, ainda, utilizar um classificador não linear.

Tabela 3.1 – Conjuntos de vetores de características.

Conjunto de vetores de características	Características contidas
I	DPF e RAP
II	DPF e JF
III	RAP e JF
IV	DPF, RAP e JF

Observando-se os gráficos das figuras 3.5, 3.6 e 3.7 foi possível afirmar que para os conjuntos I, II e III, não era possível separar linearmente, por completo, os sinais patológicos dos

saudáveis. No entanto, os dois primeiros conjuntos demonstraram uma separação e agrupamento promissores para a aplicação de um classificador SVM. Enquanto o terceiro demostrou um agrupamento preocupante em torno dos valores nulos de RAP e JF, indicando que outro algoritmo de aprendizado talvez seja mais indicado, como o classificador KNN.

Figura 3.5 – Representação gráfica do plano DPF-RAP, do conjunto de vetores de características I.

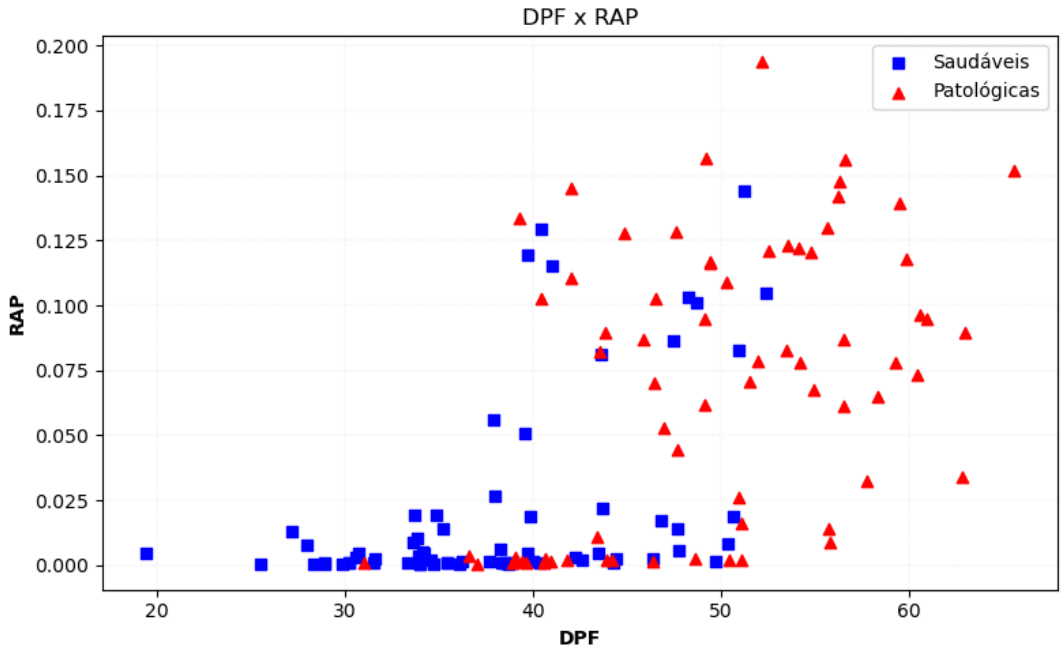


Figura 3.6 – Representação gráfica do plano DPF-JF, do conjunto de vetores de características II.

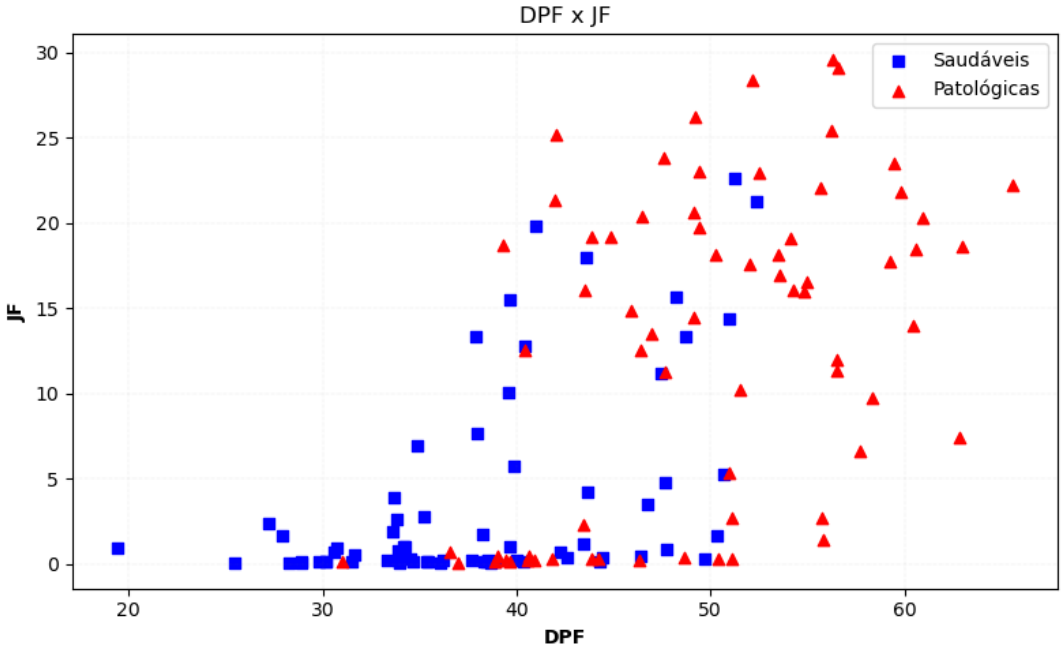
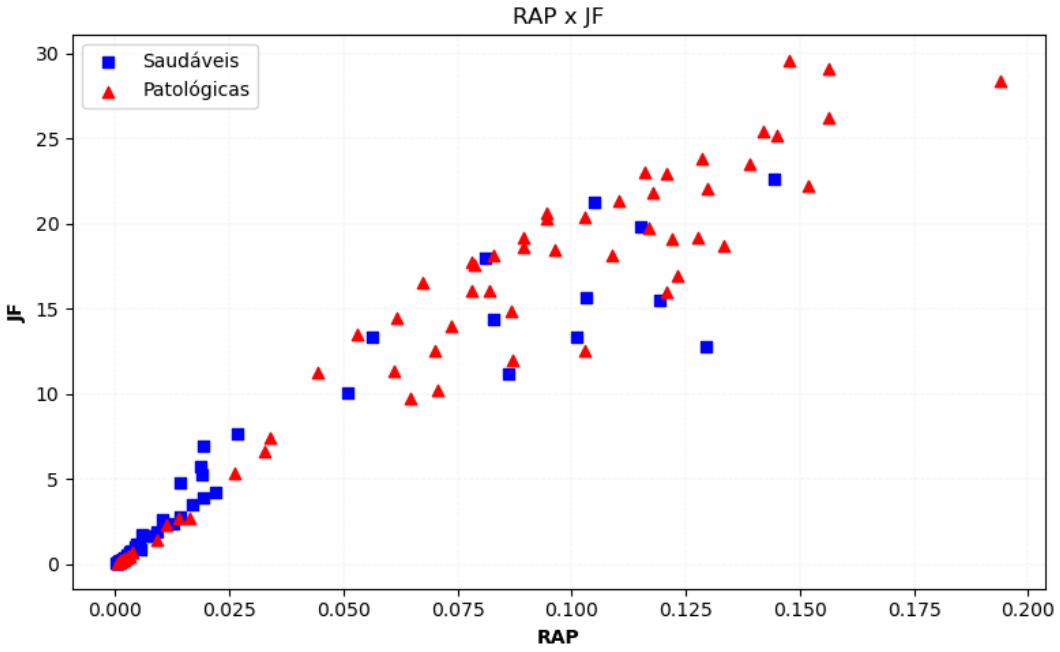
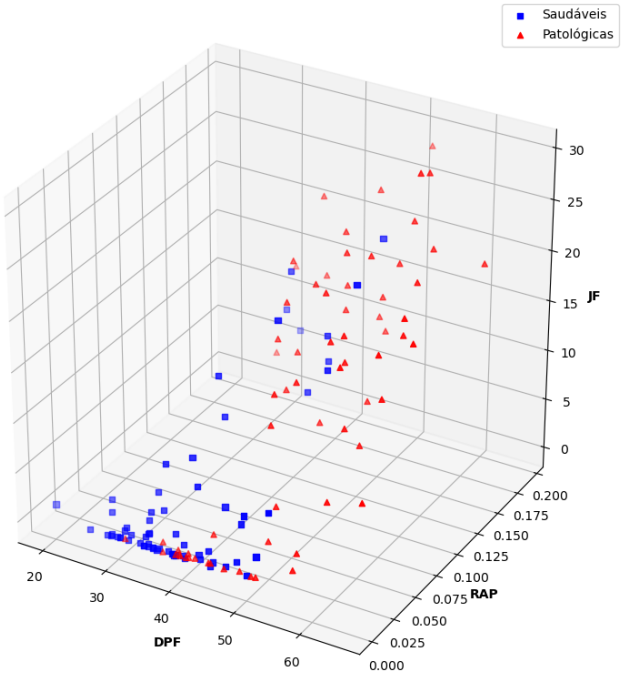


Figura 3.7 – Representação gráfica do plano RAP-JF, do conjunto de vetores de características III.



Por outro lado, o conjunto IV ilustrado pelo gráfico da figura 3.8 também demonstrou a mesma condição vista nos conjuntos I e II. Desta forma, o uso de um classificador SVM para este conjunto também se mostrou promissor.

Figura 3.8 – Representação gráfica do plano DPF-RAP-JF, do conjunto de vetores de características IV.



3.4 Classificação

Após a extração das características, foram obtidos 136 vetores de características para cada um dos conjuntos, sendo que os I, II e III, possuem duas características cada, enquanto o IV possui três. Nesta etapa, se buscou identificar quais seleções de vetores, presentes nos quatro conjuntos descritos, melhor separa as duas classes de sinais.

Para isso, cada conjunto foi dividido em duas partes: um para o treinamento do modelo de aprendizado e outro para a realização dos testes do modelo treinado, tal que, cada parte possuía metade do conjunto de vetores de características e era composta por quantidades iguais de cada uma das classes. Em outras palavras, cada conjunto foi dividido em dois subconjuntos com 68 sinais sendo que destes 34 eram saudáveis e os outros 34 patológicos.

No entanto, verificar todas as possibilidades de combinações se mostrou inviável, já que seria necessário analisar $\binom{136}{68} = \frac{136!}{(136-68)68!}$ combinações. Desta forma, foi utilizada a técnica de validação cruzada de Monte Carlo, na qual, os sinais de cada conjunto de treino e teste são escolhidos aleatoriamente sem repetição em cada uma das execuções do classificador, porém, mantendo a proporção estabelecida.

Assim, cada uma das combinações foram processadas por um classificador SVM com *kernel* Ke de base radial do tipo Gaussiano, definido pela equação 3.2,

$$Ke(\boldsymbol{x}, \boldsymbol{x'}) = \exp\left(- \frac{\|\boldsymbol{x} - \boldsymbol{x'}\|^2}{2\sigma^2} \right), \quad \sigma = 1 \tag{3.2}$$

tal que $\boldsymbol{x}$, $\boldsymbol{x'}$ representam vetores de características e σ um parâmetro livre. E por um classificador KNN com normalização máximo-mínimo, definido pela equação 3.3,

$$V' = \frac{V - min(F)}{max(F) - min(F)}(V_{max} - V_{min}) + V_{min}, \quad V_{max} = 1, \quad V_{min} = -1 \tag{3.3}$$

tal que V' é o novo valor normalizado, V o valor atual, $max(F)$ e $min(F)$ são, respectivamente, o maior e menor valores calculados da característica F, e V_{max} e V_{min} definem o intervalo em

que V' deve estar contido.

Enquanto, paro o cálculo das distâncias utilizadas pela KNN, foi utilizada a distância euclidiana, definida pela equação 3.4,

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2} \tag{3.4}$$

tal que, p e q são dois vetores de características distintos, n o tamanho destes vetores e p_i e q_i são as i-ésimas características de cada vetor.

Por fim, para o classificador KNN, o valor do parâmetro de afinação K foi escolhido a partir dos resultados de várias execuções do algoritmo, os quais demostraram que o melhor valor é tal que $K = 7$, uma vez que valores maiores até 23 não haviam mudanças significativas, menores de sete demonstravam piores resultados e acima de 23 começavam a aparecer sinais de excessiva generalização.

Por fim, os classificadores retornaram os valores de acurácia (ACC), sensibilidade (SEN) e especificidade (SPC). Além disso, as matrizes de confusão com a maior ACC, as médias e desvios padrões das ACC, SEN e SPC, de cada conjunto de iterações também foram armazenadas ao fim das execuções.

Capítulo 4

Resultados

Definidos os conjuntos de testes foi possível prosseguir para as etapas de treinamento, teste e validação do modelo de classificação. O processo foi realizado utilizando os classificadores SVM e KNN, de forma que, a classe positiva representasse um sinal patológico e a negativa um saudável.

Assim padronizado, foram realizadas cinco rotinas de testes, em cada classificador, nas quais foram executadas 100, 200, 500, 1000 e 5000 repetições, respectivamente, para cada um dos conjuntos de vetores de características descritos na tabela 3.1. Em cada repetição, os conjuntos de treino e teste foram submetidos a validação cruzada de Monte Carlos, na qual as quantidades de sinais de treino e teste nos conjuntos não sofreram modificações, assim como a razão entre sinais saudáveis e patológicos. Para execução do procedimento de classificação, foi, então, calculada a média e desvio padrão dos valores de ACC, SEN e SPC como métricas de validação dos resultados.

4.1 Testes do conjunto I – DPF e RAP

No primeiro teste, foi utilizado o conjunto I de vetores de características, contendo os valores de DPF e RAP dos sinais. Os resultados das rotinas de testes para este conjunto podem ser

vistos na tabela 4.1. Através destes, é possível notar que apesar da representação gráfica dos vetores, vista na figura 3.5, demonstrar uma possibilidade promissora de discriminação para o classificador SVM, os resultados não foram tão bons quanto o esperado. Por outro lado, para o classificador KNN, foram obtidas melhores porcentagens de ACC e ótimas medidas de SPC em relação a SVM.

Tabela 4.1 – Resultados dos testes de validação dos classificadores SVM e KNN para o conjunto I (DPF e RAP) de vetores de características.

Resultados dos testes de validação para o conjunto I (DPF e RAP) de vetores de características.					
SVM					
Nº de Iterações	100	200	500	1000	5000
ACC média (%)	61,250002	59,882650	60,400000	60,191176	60,168824
SEN média (%)	61,676471	60,750000	61,664706	61,017647	61,222353
SPC média (%)	60,823529	59,014706	59,135294	59,364706	59,115294
Desvio padrão da ACC	6,282358	7,249734	7,088368	7,190710	7,043936
Desvio padrão da SEN	11,071804	11,337950	10,500592	10,669072	10,400167
Desvio padrão da SPC	9,935657	11,412392	11,137487	11,895050	11,399616
KNN, K = 7					
Nº de Iterações	100	200	500	1000	5000
ACC média (%)	72,573518	72,779409	72,385288	72,619111	72,583817
SEN média (%)	63,411765	64,426471	63,994118	64,397059	64,334118
SPC média (%)	81,735294	81,132353	80,776471	80,841176	80,833529
Desvio padrão da ACC	3,543348	3,801784	3,986198	3,827668	3,818356
Desvio padrão da SEN	8,661818	9,593453	9,648206	9,219451	9,227044
Desvio padrão da SPC	7,387560	8,215198	8,243912	8,309428	8,128111

A matriz de confusão de maior ACC obtida neste conjunto, para a SVM foi:

$$\begin{pmatrix} 33 & 9 \\ 1 & 25 \end{pmatrix} \Rightarrow \text{ ACC} = 85,294100\%;\ \text{SEN} = 97,058824\%;\ \text{SPC} = 73,529412\%$$

Enquanto para a KNN foi:

$$\begin{pmatrix} 29 & 5 \\ 5 & 29 \end{pmatrix} \Rightarrow \text{ACC} = 85,294100\%; \text{SEN} = 85,294118\%; \text{SPC} = 85,294118\%$$

Devido ao fato dos valores de ACC, SEN e SPC médios da SVM se manterem em torno de 59%, enquanto os desvios padrões da sensibilidade e especificidade se mantiveram acima

de 10, indica uma instabilidade indesejada na efetividade da detecção de patologias. Desta forma, mesmo obtendo uma ACC máxima alta de 85%, não se pode concluir que este conjunto é suficientemente adequado quando classificado por uma SVM.

No entanto, ao utilizar um classificador KNN os desvios padrões diminuíram perceptivelmente, assim como houve um aumento significativo nos valores de ACC e SPC, enquanto a SEN sofreu um pequeno aumento. Desta forma, mesmo obtendo uma SEN máxima menor, de 85%, este conjunto se mostra mais aceitável quando classificado por uma KNN.

4.2 Testes do conjunto II – DPF e JF

No segundo teste, foi utilizado o conjunto II de vetores de características, contendo os valores de DPF e JF dos sinais. Os resultados das rotinas de testes para este conjunto podem ser vistos na tabela 4.2. Por meio destes resultados, é possível verificar que este conjunto correspondeu as expectativas esperadas, quando classificado por uma SVM, ao contrário do conjunto anterior, e manteve bons resultados com a KNN.

Para a SVM os valores médios de ACC, SEN e SPC tiveram um aumento aproximado em torno de 7%, 9% e 5%, respectivamente. Enquanto os desvios padrões diminuíram em até 3% para as três métricas, provendo uma melhor estabilidade do que o conjunto anterior. Enquanto para a KNN houveram pequenas melhorias percentuais em todas as métricas utilizadas.

A matriz de confusão de maior ACC obtida neste conjunto, para a SVM foi:

$$\begin{pmatrix} 28 & 4 \\ 6 & 30 \end{pmatrix} \Rightarrow \text{ ACC} = 85,294100\%;\ \text{SEN} = 82,352941\%;\ \text{SPC} = 88,235294\%$$

Enquanto para a KNN foi:

$$\begin{pmatrix} 28 & 3 \\ 6 & 31 \end{pmatrix} \Rightarrow \text{ACC} = 86,764700\%; \text{SEN} = 82,352941\%; \text{SPC} = 91,176471\%$$

Tabela 4.2 – Resultados dos testes de validação dos classificadores SVM e KNN para o conjunto II (DPF e JF) de vetores de características.

Resultados dos testes de validação para o conjunto II (DPF e JF) de vetores de características.					
SVM					
N^o de Iterações	100	200	500	1000	5000
ACC média (%)	67,558819	67,632349	67,449997	67,502938	67,264116
SEN média (%)	70,794118	70,691176	70,547059	70.432353	70,108824
SPC média (%)	64,323529	64,573529	64,352941	64,573529	64,419412
Desvio padrão da ACC	4,815724	4,589217	4,683844	4,825040	4,865660
Desvio padrão da SEN	7,857107	8,194536	8,121618	8,319003	8,147625
Desvio padrão da SPC	9,259906	9,269438	9,156930	9,210524	9,427546
KNN, K = 7					
N^o de Iterações	100	200	500	1000	5000
ACC média (%)	74,676471	74,102937	74,379411	74,185586	74,291468
SEN média (%)	66,117647	64,647059	65,188235	65,250000	65,266471
SPC média (%)	83,235294	83,558824	83,570588	83,141176	83,316471
Desvio padrão da ACC	3,665805	3,729064	3,757640	3,883247	3,740330
Desvio padrão da SEN	8,428263	8,355193	8,813260	8,928780	8,619599
Desvio padrão da SPC	8,175874	7,667326	7,868358	7,849856	7,924891

Ao contrario do conjunto anterior, os resultados médios se mostraram mais próximos do máximo obtido e com desvios menores, abaixo de 10, para ambos os classificadores. Na SVM, houve um aumento significativo da SEN, indicando uma capacidade de detecção de patologias razoavelmente satisfatória, enquanto, por outro lado, na KNN se manteve estável. Desta forma, é possível afirmar que este conjunto é suficientemente adequado quando utilizado qualquer um dos classificadores testados, além de ter apresentado os melhores resultados dentre os quatro testados.

4.3 Testes do conjunto III – RAP e JF

No terceiro teste, foi utilizado o conjunto III de vetores de características, contendo os valores de RAP e JF dos sinais. Os resultados das rotinas de testes para este conjunto podem ser vistos na tabela 4.3. Através destes resultados, foi possível verificar que, assim como esperado, este conjunto não foi capaz de discriminar sinais saudáveis de patológicos através da SVM. No

entanto, a KNN foi capaz de separar as classes de forma razoável.

Todas as métricas de validação adotadas apresentaram uma piora significativa em comparação com os testes anteriores para a SVM, enquanto para a KNN houve apenas piora na ACC e SPC. As médias em torno de 50%, evidenciaram a incapacidade de discriminar os sinais com o conjunto em questão, por meio de um classificador SVM de base radial do tipo Gaussiano, chegando a ser equiparável com um modelo aleatório. Enquanto a KNN demonstrou ser capaz de obter resultados razoáveis com este conjunto.

Tabela 4.3 – Resultados dos testes de validação dos classificadores SVM e KNN para o conjunto III (RAP e JF) de vetores de características.

Resultados dos testes de validação para o conjunto III (RAP e JF) de vetores de características.					
SVM					
Nº de Iterações	100	200	500	1000	5000
ACC média (%)	50,294118	50,742648	50,964707	50,713235	50,837059
SEN média (%)	48,794118	50,882353	51,623529	51,532353	51,587647
SPC média (%)	51,794118	50,602941	50,305882	49,894118	50,086471
Desvio padrão da ACC	6,221654	7,393472	7,280630	7,172427	7,049967
Desvio padrão da SEN	12,095000	11,627189	12,079979	11,858382	12,132993
Desvio padrão da SPC	15,438637	15,617050	16,213851	16,411617	16,616507
KNN, K = 7					
Nº de Iterações	100	200	500	1000	5000
ACC média (%)	71,544107	72,316166	71,929403	72,074994	71,867638
SEN média (%)	64,911765	65,338235	64,470588	64,823529	64,369412
SPC média (%)	78,176471	79,294118	79,388235	79,326471	79,365882
Desvio padrão da ACC	3,506156	3,669263	3,587804	3,825129	3,699919
Desvio padrão da SEN	8,529002	8,014857	8,373737	8,840883	8,649981
Desvio padrão da SPC	7,242848	6,863491	7,375561	7,163495	7,183549

A matriz de confusão de maior ACC obtida neste conjunto, para a SVM foi:

$$A = \begin{pmatrix} 26 & 8 \\ 8 & 26 \end{pmatrix} \Rightarrow \text{ ACC} = 76,470600\%;\ \text{SEN} = 76,470588\%;\ \text{SPC} = 76,470588\%$$

Enquanto para a KNN foi:

$$\begin{pmatrix} 27 & 3 \\ 7 & 31 \end{pmatrix} \Rightarrow \text{ACC} = 85,294100\%; \text{SEN} = 79,411765\%; \text{SPC} = 91,176471\%$$

Devido as baixíssimas médias e aos desvios padrões da SEN e SPC entre 10 e 16, se torna óbvio que a SVM não é adequada para a discriminação das classes nesse conjunto. A KNN mesmo obtendo resultados similares aos conjuntos anteriores também apresentou uma piora de até 3%. Desta forma, este conjunto apresentou os piores resultado dentre os quatro testados e, consequentemente, se mostrou ineficaz em discriminar vozes patológicas de saudáveis, no caso da SVM, e no da KNN quando comparado com os outros conjuntos.

4.4 Testes do conjunto IV – DPF, RAP e JF

No último teste, foi utilizado o conjunto IV de vetores de características, contendo os valores de DPF, RAP e JF dos sinais. Os resultados das rotinas de testes para este conjunto podem ser vistos na tabela 4.4. Assim como no conjunto II, através dos resultados, é possível verificar que este conjunto também correspondeu as expectativas esperadas.

As métricas desta rotina de testes alcançaram valores quase idênticos ao do conjunto II, o que era esperado devido ao fato do conjunto III ter se mostrado ineficaz na separação das classes. Ao mesmo tempo, estes resultados demonstraram que a utilização de uma nova característica, fora das três utilizadas e que seja capaz de discriminar melhor as classes, pode ser capaz de permitir uma melhora significativa da classificação.

A matriz de confusão de maior ACC obtida neste conjunto, para a SVM foi:

$$\begin{pmatrix} 28 & 6 \\ 6 & 28 \end{pmatrix} \Rightarrow \text{ACC} = 82,352900\%; \text{SEN} = 82,352941\%; \text{SPC} = 82,352941\%$$

Enquanto para a KNN foi:

$$\begin{pmatrix} 30 & 4 \\ 4 & 30 \end{pmatrix} \Rightarrow \text{ACC} = 88,235300\%; \text{SEN} = 88,235294\%; \text{SPC} = 88,235294\%$$

Tabela 4.4 – Resultados dos testes de validação dos classificadores SVM e KNN para o conjunto IV (DPF, RAP e JF) de vetores de características.

Resultados dos testes de validação para o conjunto IV (DPF, RAP e JF) de vetores de características.					
SVM					
N^o de Iterações	100	200	500	1000	5000
ACC média (%)	67,035291	67,397060	67,035291	67,279407	67,362351
SEN média (%)	70,017647	70,191176	70,017647	70,376471	70,257647
SPC média (%)	64,052941	64,602941	64,052941	64,182353	64,467059
Desvio padrão da ACC	4,799667	4,856415	4,799667	4,724181	4,845597
Desvio padrão da SEN	7,897914	8,824404	7,897916	7,980392	8,134518
Desvio padrão da SPC	9,423360	9,560682	9,423360	9,304687	9,469985
KNN, K = 7					
N^o de Iterações	100	200	500	1000	5000
ACC média (%)	74,102938	74,080881	74,326471	74,185292	74,323528
SEN média (%)	65,500000	66,102941	65,347059	65,155882	65,344118
SPC média (%)	82,705882	82,058824	83,305882	83,214706	83,302941
Desvio padrão da ACC	3,881803	3,759964	4,085938	3,727745	3,764751
Desvio padrão da SEN	9,166965	8,869902	8,649145	8,588093	8,763564
Desvio padrão da SPC	8,436553	8,171281	8,1542221	8,120403	7,897983

Assim como nos conjuntos I e II, as melhores matrizes de confusão obtiveram métricas acima de 80% e, como os resultados foram quase idênticos ao do conjunto II, tanto para a SVM quanto para a KNN, é possível afirmar que este conjunto também é razoavelmente adequado para a classificação de sinais saudáveis e patológicos.

Capítulo 5

Conclusões

Neste trabalho foi proposto um método computacional capaz de classificar sinais de voz, da vogal /a/ sustentada, em patológicos, afetados pelo Edema de Reinke, e saudáveis utilizando técnicas de aprendizado de máquina. Para a SVM com *kernel* de base radial do tipo Gaussiano, os conjuntos de características deixaram um pouco a desejar em termos da ACC e SPC, enquanto obteve valores de SEN satisfatórios para os conjuntos II e IV. Enquanto para o classificador KNN ocorreu o oposto, os conjuntos deixaram a desejar em relação a SEN, porém obtiveram bons resultados de ACC e SPC.

Em geral, foi possível verificar que o conjunto III, constituído pelo par das características RAP e JF, demonstrou ser o pior para a separação de vozes patológicas e saudáveis, e que o classificador KNN se mostrou melhor em detectar a ausência de patologias laríngeas e pior em detectar a sua presença, enquanto no caso da SVM foi observado a situação contrária.

Também se notou que a menor SEN da KNN e a menor SPC da SVM ocorreram devido ao fato de alguns vetores de características patológicos estarem muito próximos ao agrupamento dos saudáveis, o que impediu os classificadores de obterem melhores resultados. No entanto, foi verificado que estes vetores patológicos correspondiam a casos pós-cirúrgicos de remoção do edema, casos iniciais ou de disfonias remanescentes, segundo os laudos médicos, e, portanto, não representam um grande risco no problema de classificação. Desta forma, tanto o classificador SVM quanto o KNN se mostraram suficientemente adequados, sendo que ambos poderiam obter resultados melhores se fossem desconsiderados os casos excepcionais citados.

Portanto, foi possível concluir que o objetivo deste trabalho foi alcançado e que as características e classificadores utilizados são adequados para o problema de classificação dos sinais, mesmo não obtendo resultados tão satisfatórios quanto aos dos trabalhos que utilizaram técnicas mais sofisticadas [1, 5, 8].

Além disso, foram obtidos, através do método proposto, ACC e SEN médias de 67% e 70%, para a SVM, e de 74% e 65%, para a KNN, respectivamente, e máximas entre 85% a 88% de ACC e 82% a 88% de SEN para ambos classificadores. Resultado semelhante a de outros trabalhos nos quais também são utilizados conceitos e técnicas de menor complexidade [19, 21].

Por fim, trabalhos futuros neste âmbito podem incluir: a utilização de quantidades maiores de sinais saudáveis e patológicos através da adoção de mais de uma base de dados ou da coleta manual de novas amostras; a extração de características mais complexas como os coeficientes cepstrais de frequência de Mel (MFCC); a utilização de técnicas mais robustas como a codificação preditiva linear (LPC) ou a análise *wavelet*; e a aplicação de classificadores mais sofisticados, nos próprios resultados obtidos neste trabalho por exemplo, como as redes neurais e a clusterização *k-mean*.

Referências

[1] ALVES, M.; SILVA, G.; BISPO, B. C.; DAJER, M. E.; RODRIGUES, P. M. Voice Disorders Detection Through Multiband Cepstral Features of Sustained Vowel. **Journal of Voice**, [s. l.], v. 0, n. 0, 2021. Disponível em: <https://www.jvoice.org/article/S0892-1997(21)00042-4/fulltext>. Acesso em: 6 jul. 2021.

[2] AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. **Voice Disorders**. (Pratice Portal). [s.d.]. Disponível em: <https://www.asha.org/practice-portal/clinical-topics/voice-disorders/>. Acesso em: 8 jul. 2021.

[3] BEHLAU, M. **Voz: O livro do especialista**. Rio de Janeiro: Revinter, 2008.

[4] BOGERT, B. P.; HEALY, J. R.; TUKEY, J. W. The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking. In: PROCEEDINGS OF THE SYMPOSIUM ON TIME SERIES ANALYSIS 1963, [s. l.]. **Anais**[...]. [s.l: s.n.]

[5] CHEN, L.; WANG, C.; CHEN, J.; XIANG, Z.; HU, X. Voice Disorder Identification by using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN). **Journal of Voice**, [s. l.], v. 35, n. 6, p. 932.e1-932.e11, 2021. Disponível em: <https://www.jvoice.org/article/S0892-1997(20)30101-6/fulltext>. Acesso em: 6 jul. 2021.

[6] COOLEY, J. W. The re-discovery of the fast Fourier transform algorithm. **Mikrochimica Acta**, [s. l.], v. 93, n. 1–6, p. 33–45, 1987.

[7] COOLEY, J. W.; TUKEY, J. W. An algorithm for the machine calculation of complex Fourier series. **Mathematics of Computation**, [s. l.], v. 19, n. 90, p. 297–301, 1965.

[8] FONSECA, E. S.; GUIDO, R. C.; SCALASSARA, P. R.; MACIEL, C. D.; PEREIRA, J. C. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. **Computers in Biology and Medicine**, [s. l.], v. 37, n. 4, p. 571–578, 2007.

[9] HECKER, M. H. L.; KREUL, E. J. Descriptions of the Speech of Patients with Cancer of the Vocal Folds. Part I: Measures of Fundamental Frequency. **The Journal of the Acoustical Society of America**, [s. l.], v. 49, n. 4B, p. 1275–1282, 1971.

[10] ISLAM, R.; TARIQUE, M.; ABDEL-RAHEEM, E. A Survey on Signal Processing Based Pathological Voice Detection Techniques. **IEEE Access**, [s. l.], v. 8, p. 66749–66776, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9055386/>. Acesso em: 25 jan. 2022.

[11] KOIKE, Y. Application of Some Acoustic Measures for the Evaluation of Laryngeal Dysfunction. **The Journal of the Acoustical Society of America**, [s. l.], v. 42, n. 5, p. 1209–1209, 1967.

[12] KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York: Springer, 2013.

[13] NOLL, A. M. Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection. **The Journal of the Acoustical Society of America**, [s. l.], v. 36, n. 2, p. 296–302, 1964.

[14] NOLL, A. M. Cepstrum Pitch Determination. **The Journal of the Acoustical Society of America**, [s. l.], v. 41, n. 2, p. 293–309, 1967.

[15] OPPENHEIM, A. V.; SCHAFER, R. W.; STOCKHAM, T. G. Nonlinear filtering of multiplied and convolved signals. **Proceedings of the IEEE**, [s. l.], v. 56, n. 8, p. 1264–1291, 1968.

[16] RABINER, L. R.; SCHAFER, R. W. Introduction to Digital Speech Processing. **Foundations and Trends® in Signal Processing**, [s. l.], v. 1, n. 1–2, p. 1–194, 2007.

[17] RANDALL, R. B. A history of cepstrum analysis and its application to mechanical problems. **Mechanical Systems and Signal Processing**, [s. l.], v. 97, p. 3–19, 2017.

[18] Saarbruecken Voice Database. Instituto de Fonética, Universidade do Sarre, Alemanha. Disponível em: <http://stimmdb.coli.uni-saarland.de/index.php4>. Acesso em: 1 ago. 2021.

[19] SATO, L. A. F. **Processamento digital de sinais acústicos com aplicações biomédicas: detecção de anomalias laríngeas**. 2018. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Univerisdade Estadual Paulista, São José do Rio Preto - SP, 2018.

[20] SUKHOSTAT, L.; IMAMVERDIYEV, Y. A Comparative Analysis of Pitch Detection Methods Under the Influence of Different Noise Conditions. **Journal of Voice**, [s. l.], v. 29, n. 4, p. 410–417, 2015.

[21] TEIXEIRA, F.; FERNANDES, J.; GUEDES, V.; JUNIOR, A.; TEIXEIRA, J. P. Classification of Control/Pathologic Subjects with Support Vector Machines. **Procedia Computer Science**, [s. l.], v. 138, p. 272–279, 2018.

[22] THARWAT, A. Classification assessment methods. **Applied Computing and Informatics**, [s. l.], v. 17, n. 1, p. 168–192, 2021.