UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Caio Sobrinho da Silva

# Análise Computacional de Disfonias: caracterizando patologias na estrutura vocal humana

**São José do Rio Preto**
**2025**

Caio Sobrinho da Silva

# Análise Computacional de Disfonias: caracterizando patologias na estrutura vocal humana

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. Eng. Rodrigo Capobianco Guido

**São José do Rio Preto**
**2025**

D229a

da Silva, Caio Sobrinho

Análise Computacional de Disfonias : caracterizando patologias na estrutura vocal humana / Caio Sobrinho da Silva. -- São José do Rio Preto, 2025

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. Disfonia. 2. Processamento Digital de Sinais. 3. Aprendizado de Máquina. 4. SVM. 5. Análise Acústica.. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

Caio Sobrinho da Silva

# Análise Computacional de Disfonias: caracterizando patologias na estrutura vocal humana

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

Profa. Dra. Adriana Barbosa Santos,
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Diego Renan Bruno
UNESP – Câmpus de São José do Rio Preto

**São José do Rio Preto**
**2025**

# Agradecimentos

Agradeço, primeiramente, a Deus e a Cristo Jesus por me tirarem da minha zona de conforto e por me conduzirem para o outro lado do estado, pois até aqui o Senhor tem me sustentado.

Sou grato aos meus pais, que pagaram um alto preço para que eu pudesse estar aqui hoje. Foram eles que me apoiaram no período do meu acidente e que continuam me apoiando até hoje. Cada lágrima derramada durante esse processo não se compara ao suor e dedicação deles.

Agradeço também ao meu orientador, Rodrigo, que me encorajou na minha primeira iniciação científica e que agora me acompanha como orientador. Obrigado pelo carinho, pela paciência e por sempre me ouvir.

"Porque o Senhor é quem dá sabedoria;

da sua boca procedem o conhecimento e o discernimento."

- Provérbios 2:6

## *Resumo*

O processamento digital de sinais de voz tem se tornado uma ferramenta indispensável para o diagnóstico não invasivo de patologias laríngeas. Este trabalho propõe o desenvolvimento de um sistema computacional para a detecção automática de disfonias, utilizando a base de dados Saarbruecken Voice Database (SVD). A metodologia compreendeu o pré-processamento dos sinais acústicos e a extração de características nos domínios do tempo e da frequência, incluindo Energia, Entropia, Taxa de Cruzamento por Zero (ZCR), Frequência Fundamental ($F_0$), Formantes, Jitter e Shimmer. Para avaliar a relevância e a capacidade discriminativa desses atributos, foi realizada uma análise estatística de dispersão por meio de diagramas de caixa (boxplots), permitindo a visualização do comportamento das características entre grupos saudáveis e patológicos. A classificação dos padrões vocais foi realizada utilizando o algoritmo Support Vector Machine (SVM). Os resultados obtidos através das matrizes de confusão e métricas de acurácia demonstraram que a combinação de descritores físicos e estatísticos, aliada a classificadores inteligentes, oferece uma abordagem robusta e eficaz para auxiliar na triagem e diagnóstico clínico de distúrbios vocais.



**Palavras-chave: Disfonia. Processamento Digital de Sinais. Aprendizado de Máquina. SVM. Análise Acústica.**

*Abstract*

Digital voice signal processing has become an indispensable tool for the non-invasive diagnosis of laryngeal pathologies. This study proposes the development of a computational system for automatic dysphonia detection utilizing the Saarbruecken Voice Database (SVD). The methodology encompassed acoustic signal pre-processing and feature extraction in both time and frequency domains, including Signal Energy, Entropy, Zero Crossing Rate (ZCR), Fundamental Frequency ($F_0$), Formants, Jitter, and Shimmer. To evaluate the relevance and discriminative capacity of these attributes, a statistical dispersion analysis was performed using boxplots, allowing for the visualization of feature behavior between healthy and pathological groups. Vocal pattern classification was carried out using the Support Vector Machine (SVM) algorithm. The results obtained through confusion matrices and accuracy metrics demonstrated that the combination of physical and statistical descriptors, coupled with intelligent classifiers, offers a robust and effective approach to aid in the screening and clinical diagnosis of vocal disorders.

**Dysphonia. Digital Signal Processing. Machine Learning. SVM. Acoustic Analysis.**

# Lista de Figuras

# Lista de Tabelas

# Lista de Abreviações

| | |
|---|---|
| $F_0$ | Frequência Fundamental |
| A3 | Método de Constância de Energia |
| B3 | Método de Constância de ZCR |
| CV | Validação Cruzada |
| DSP | Processamento Digital de Sinais |
| EDA | Análise Exploratória de Dados |
| FFT | Transformada Rápida de Fourier |
| IA | Inteligência Artificial |
| LP | Lógica Paraconsistente |
| LPA2v | Lógica Paraconsistente Anotada de 2 valores |
| LPC | Predição Linear |
| ML | Aprendizado de Máquina |
| PSD | Espectro de Densidade de Potência |
| RBF | Função de Base Radial |
| SVD | Saarbruecken Voice Database |

SVM Máquina de Vetores de Suporte

ZCR Taxa de Cruzamento por Zero

# Sumário

# Capítulo 1

# Introdução ao trabalho

## 1.1 Introdução

O destacado crescimento tecnológico observado nas últimas décadas tem possibilitado o aprimoramento substancial de diversas áreas do conhecimento, viabilizando a solução de problemas anteriormente considerados intratáveis. No contexto médico, a evolução da capacidade computacional tem fomentado o desenvolvimento de técnicas diagnósticas menos invasivas e mais eficientes. Particularmente, o diagnóstico de patologias associadas ao aparelho vocal tem sido amplamente beneficiado pelo surgimento de novas técnicas em Processamento Digital de Sinais (DSP), conforme fundamentado nos princípios clássicos de análise espectral e temporal descritos por (LYONS, 2004) e nas normas de codificação de áudio digital apresentadas por (BOSSI; GOLDBERG, 2003)

A produção da voz humana é um processo complexo que depende da interação coordenada de diversos subsistemas e órgãos (BEHRMAN, 2021). No entanto, a dinâmica natural desse processo pode ser severamente perturbada por enfermidades ou condições anômalas. Estudos indicam que, embora imprevisíveis, as alterações causadas por tais doenças tendem a não ser aleatórias, seguindo padrões matemáticos de "doenças dinâmicas"(B´ELAIR, 1995). Essa característica permite a modelagem

e a classificação dessas anomalias por intermédio de sistemas de reconhecimento de padrões (DUDA, 2000), utilizando medidas acústicas e aerodinâmicas para avaliar a eficácia de terapias e intervenções cirúrgicas (ULOZA; SAFERIS; ULOZIENE, 2005; HOLMBERG, 2003).

Nesse cenário, abordagens que exploram características nos domínios do tempo e da frequência tornaram-se essenciais. A literatura recente destaca a eficácia de descritores específicos para a caracterização de sinais, tais como a energia do sinal (GUIDO, 2016a), a taxa de cruzamento por zero (ZCR) (GUIDO, 2016b) e a entropia baseada na extração de características (GUIDO, 2018). A combinação desses atributos alimenta algoritmos de aprendizado de máquina voltados ao reconhecimento de locutores e patologias (MAK; CHIEN, 2000).

## 1.2 Objetivo Geral

O objetivo principal deste trabalho é investigar e desenvolver uma abordagem computacional híbrida para o auxílio ao diagnóstico de disfonias, fundamentada na análise acústica de sinais de voz. A proposta consiste em validar a aplicação da Lógica Paraconsistente na seleção de características relevantes e avaliar o desempenho de classificadores inteligentes na distinção entre padrões vocais saudáveis e patológicos, utilizando bases de dados clínicas padronizadas (SVD. . . , ; GUIDO, 2019).

## 1.3 Objetivos Específicos

Para concretizar o objetivo geral, foram estabelecidas as seguintes metas específicas:

- **Caracterização de Sinais Vocais:** Investigar os padrões acústicos presentes na base de dados *Saarbruecken Voice Database* (SVD), identificando as variações

espectrais e temporais que distinguem locutores normais de disfônicos (SVD. . . , ).

- **Análise de Descritores Acústicos:** Avaliar a capacidade discriminativa de um conjunto robusto de características, compreendendo medidas físicas (Energia, ZCR), medidas de perturbação (*Jitter*, *Shimmer*) e medidas de complexidade do sinal (Entropia e Formantes) na representação de patologias laríngeas (LYONS, 2004; GUIDO, 2018; BEHRMAN, 2021).

- **Tratamento da Incerteza:** Validar um método de seleção de atributos baseado em Lógica Paraconsistente, visando reduzir a dimensionalidade dos dados e mitigar as incertezas e contradições inerentes às medições biomédicas (GUIDO, 2019; AVRON A.; ARIELI, 2018).

- **Modelagem Preditiva:** Desenvolver e otimizar modelos de classificação baseados em Máquinas de Vetores de Suporte (SVM), analisando o impacto de diferentes funções *kernel* (polinomial e base radial) na separabilidade das classes vocais (DUDA, 2000).

- **Validação Clínica Computacional:** Aferir a eficácia do sistema proposto por meio de métricas estatísticas rigorosas (sensibilidade, especificidade, acurácia), buscando correlacionar os resultados computacionais com a interpretabilidade clínica das disfonias detectadas (MONTE-SERRAT, 2021).

## 1.4 Justificativa

O diagnóstico tradicional de patologias vocais frequentemente depende de avaliações perceptivas subjetivas ou de exames invasivos, como a laringoscopia. No entanto, a voz humana, resultante da interação complexa de subsistemas fisiológicos, carrega padrões acústicos sutis que podem ser imperceptíveis à audição humana, mas detec-

táveis computacionalmente. A análise por meio de Processamento Digital de Sinais (DSP) justifica-se, portanto, pela capacidade de fornecer marcadores objetivos e não invasivos, permitindo identificar alterações na dinâmica vocal que precedem sintomas clínicos evidentes (BEHRMAN, 2021; ULOZA; SAFERIS; ULOZIENE, 2005).

Diferentes patologias laríngeas alteram as propriedades físico-químicas e mecânicas das pregas vocais, impactando diretamente a estabilidade do sinal acústico. Dessa forma, o sinal de voz pode ser decomposto em descritores matemáticos específicos. Regiões de instabilidade na vibração das cordas vocais manifestam-se como perturbações de curto termo (*Jitter* e *Shimmer*) ou aumento na desordem do sinal (Entropia) (GUIDO, 2018; HOLMBERG, 2003). Essas características atuam como "assinaturas digitais"da patologia, constituindo a base estrutural necessária para a diferenciação entre vozes saudáveis e disfônicas.

No que diz respeito ao uso de Inteligência Artificial, a aplicação de técnicas de aprendizado de máquina é fundamental para lidar com a não-linearidade inerente à produção vocal. O aprendizado supervisionado, em especial mediante a Máquinas de Vetores de Suporte (SVM), destaca-se por sua robustez em problemas de classificação binária e pela capacidade de generalização em espaços de alta dimensão (DUDA, 2000). O uso deste paradigma é reforçado pela disponibilidade de bases de dados consolidadas, como a *Saarbruecken Voice Database* (SVD), que permitem o treinamento de modelos preditivos com alta acurácia e reprodutibilidade (SVD. . . , ; MAK; CHIEN, 2000).

## 1.5 Motivação

A maioria dos processos de fonação humana é controlada por interações complexas e não-lineares entre o fluxo aerodinâmico pulmonar e as estruturas mioelásticas da laringe. Tal como em sistemas biológicos moleculares, a distribuição da informação patológica no sinal de voz não é homogênea; certos segmentos temporais e caracte-

rísticas espectrais específicas contribuem mais significativamente para a distinção de uma anomalia do que outros (BEHRMAN, 2021). Estes atributos, como a entropia do sinal e as perturbações de curto termo (*jitter* e *shimmer*), funcionam como marcadores críticos da saúde vocal (GUIDO, 2018; HOLMBERG, 2003).

Assim, aplicar métodos de aprendizado de máquina para identificar e ponderar esses descritores acústicos como moduladores da classificação entre vozes saudáveis e patológicas apresenta-se como uma abordagem extremamente favorável. Essa estratégia permite não apenas a automação do diagnóstico, mas a descoberta de padrões sutis que métodos tradicionais ou subjetivos podem falhar em detectar, viabilizando o desenvolvimento de ferramentas clínicas de alta precisão e baixo custo (DUDA, 2000; ULOZA; SAFERIS; ULOZIENE, 2005).

# Capítulo 2

# Revisão Bibliográfica

## 2.1 Fundamentação Teórica

A voz humana é um fenômeno acústico contínuo. Para que possa ser analisada por um computador, ela deve ser convertida em um sinal digital discreto por meio de um processo de amostragem e quantização (LYONS, 2004). O Processamento Digital de Sinais (DSP) compreende o conjunto de técnicas matemáticas utilizadas para analisar e extrair informação desses sinais digitais (LYONS, 2004; BOSSI; GOLDBERG, 2003).

Um sinal de voz é considerado quase-estacionário: em curtos intervalos de tempo (janelas), suas propriedades estatísticas permanecem relativamente constantes, mas elas mudam ao longo do tempo. Por essa razão, a análise de voz é tipicamente realizada em segmentos de curto termo (LYONS, 2004). A análise pode ser dividida em dois domínios principais:

- **Domínio do Tempo:** A análise é feita diretamente na amplitude do sinal (a forma de onda). Características como Energia e Taxa de Cruzamento por Zero (ZCR) são extraídas aqui.

- **Domínio da Frequência (Espectral):** A análise foca em quais frequências compõem o sinal. Isso é feito aplicando transformações matemáticas (como a Trans-

formada de Fourier), permitindo a identificação de Frequência Fundamental ($F_0$) e Formantes.

Para preparar o sinal para a extração de características, técnicas de pré-processamento são cruciais. A **normalização** ajusta a amplitude máxima do sinal, removendo variações de volume de gravação. A **pré-ênfase** é um filtro que acentua as altas frequências, compensando a tendência natural da voz de ter mais energia em baixas frequências e facilitando a análise de formantes e ruídos de alta frequência (MAK; CHIEN, 2000).

### 2.1.1 Fisiologia Vocal e Biomarcadores Acústicos

A produção da voz humana é descrita pelo modelo Fonte-Filtro. A **Fonte** é o fluxo de ar vindo dos pulmões que faz as pregas vocais vibrarem, gerando a Frequência Fundamental ($F_0$). O **Filtro** é o trato vocal (faringe, boca, nariz), que molda o som da fonte, criando as ressonâncias conhecidas como Formantes ($F_1, F_2, F_3, F_4$).

[Image of source-filter model of speech production]

Patologias laríngeas (como nódulos, pólipos ou edemas) alteram fisicamente a estrutura ou a capacidade de vibração das pregas vocais (HOLMBERG, 2003; ULOZA; SAFERIS; ULOZIENE, 2005). Essa alteração física se traduz em distorções acústicas mensuráveis, que servem como **biomarcadores** para a detecção de disfonias. As características extraídas neste trabalho são projetadas para capturar essas distorções.

#### Características de Perturbação (Fonte)

Medem a instabilidade da vibração das pregas vocais (HOLMBERG, 2003):

- ***Jitter*:** Refere-se à variabilidade (perturbação) da frequência fundamental ($F_0$) ao longo do tempo. Um *Jitter* elevado indica instabilidade na vibração das pregas vocais.

- ***Shimmer*:** Refere-se à variabilidade (perturbação) da amplitude do sinal em cada ciclo. Um *Shimmer* elevado sugere flutuação na intensidade da vibração ou fechamento incompleto das pregas vocais.

**Características Espectrais (Fonte e Filtro)**

- **Frequência Fundamental ($F_0$):** É a taxa de vibração das pregas vocais (Fonte). Embora não seja um indicador direto de patologia, sua estabilidade (medida pelo *Jitter*) é crucial.

- **Formantes ($F_1$ a $F_4$):** São os picos de energia no espectro da voz, definidos pela configuração do trato vocal (Filtro). A análise dos formantes é essencial para caracterizar o som da vogal sustentada (BEHRMAN, 2021).

**Características Não-Lineares e Estatísticas**

Estas características, propostas na literatura de referência, capturam a complexidade e a desordem do sinal, que frequentemente aumentam em vozes patológicas:

- **Energia do Sinal:** Mede a intensidade (amplitude) do sinal em uma janela. Vozes disfônicas podem apresentar dificuldade em manter uma energia estável (GUIDO, 2016a).

- **Taxa de Cruzamento por Zero (ZCR):** Indica quantas vezes o sinal cruza o eixo zero. Uma ZCR alta está frequentemente associada a ruído (sopro) na voz, comum em patologias que impedem o fechamento completo das pregas vocais (GUIDO, 2016b).

- **Entropia:** No contexto de sinais, a entropia mede o grau de desordem, complexidade ou imprevisibilidade do sinal. Sinais de voz saudáveis e estáveis tendem a ser mais previsíveis (menor entropia) do que sinais ruidosos e caóticos de vozes patológicas (maior entropia) (GUIDO, 2018).

- **Métodos A3 e B3:** São descritores estatísticos que avaliam a *constância* da energia (A3) e da frequência/ZCR (B3) ao longo do sinal, sendo métodos robustos para capturar a estabilidade da fonação (GUIDO, 2018). A aplicação metodológica detalhada será explorada no Capítulo 3.

### 2.1.2 Aprendizado de Máquina

Aprendizado de máquina ou *machine learning* é o estudo de algoritmos e modelos estatísticos aplicados a sistemas computacionais para a realização de uma tarefa específica sem que sejam explicitamente programados para isso (MAK; CHIEN, 2000).

Algoritmos de aprendizado de máquina atuam das mais diversas formas e podem ser divididos de acordo com o método de aprendizagem utilizado. O método central deste trabalho é o **Aprendizado Supervisionado**.

No aprendizado supervisionado, toda entrada (vetor de características da voz) é acompanhada de uma saída desejada ou rótulo (ex: "Saudável"ou "Disfônico"). Dado um conjunto de treinamento rotulado, o algoritmo busca encontrar uma função matemática (modelo) com o menor erro na predição, visando otimizar a capacidade de generalização do modelo para dados nunca vistos (DUDA, 2000).

Para esse projeto, os dados de entrada são os descritores acústicos (*Jitter*, *Shimmer*, Entropia, etc.) e a saída será a classificação binária da voz.

#### Máquinas de Vetores de Suporte (SVM)

As Máquinas de Vetores de Suporte (SVM – *Support Vector Machine*) consistem em um algoritmo de aprendizado supervisionado fundamentado na Teoria do Aprendizado Estatístico (DUDA, 2000). Embora originalmente projetado para classificação binária, sua robustez matemática o tornou uma das ferramentas mais eficazes para reconhecimento de padrões em dados biomédicos.

O conceito central do SVM é a determinação de um **hiperplano de separação**

**ótimo**. Diferente de outros classificadores lineares que buscam apenas uma linha que separe as classes, o SVM busca o hiperplano que maximiza a **margem** — definida como a distância entre a fronteira de decisão e os pontos de dados mais próximos de cada classe, denominados **vetores de suporte**.

### Justificativa da Escolha do SVM

A escolha do SVM como classificador central deste trabalho, em detrimento de arquiteturas como Redes Neurais Artificiais (ANN) ou Árvores de Decisão, fundamenta-se em quatro pilares técnicos que endereçam as características específicas do problema de detecção de disfonias:

1. **Eficiência em Alta Dimensionalidade:** O problema abordado envolve um vetor de características complexo (*Jitter*, *Shimmer*, Entropia, Formantes). O SVM é reconhecidamente robusto à "maldição da dimensionalidade", mantendo alto desempenho mesmo quando o número de atributos é elevado (DUDA, 2000).

2. **Generalização em Bases de Dados Limitadas:** Enquanto técnicas de *Deep Learning* exigem volumes massivos de dados para evitar o superajuste (*overfitting*), o SVM baseia-se no Princípio da Minimização do Risco Estrutural (SRM). Isso garante uma capacidade de generalização superior em bases de dados de tamanho médio, como a *Saarbruecken Voice Database* (SVD) utilizada neste estudo (GUIDO, 2016a).

3. **Convergência Global:** O treinamento de um SVM é formulado como um problema de otimização convexa (programação quadrática). Diferentemente das redes neurais, que utilizam retropropagação e podem estagnar em mínimos locais (soluções subótimas), o SVM garante matematicamente a convergência para o mínimo global, resultando em um treinamento mais estável e computacionalmente eficiente para dados tabulares (DUDA, 2000).

4. **Modelagem Não-Linear (*Kernel Trick*):** Dados biológicos, como sinais de voz patológicos, raramente são linearmente separáveis. O SVM soluciona isso por meio de funções de *kernel*, que projetam os dados em um espaço de dimensão superior onde a separação se torna linear, sem o custo computacional explícito dessa transformação.

Conforme a metodologia deste trabalho, foram explorados dois *kernels* principais para capturar a complexidade dos sinais vocais:

- **Kernel Polinomial:** Cria fronteiras de decisão curvas baseadas em polinômios de grau $d$.
- **Kernel de Base Radial (RBF):** Um dos *kernels* mais versáteis, capaz de mapear espaços de características de dimensão infinita, sendo particularmente eficaz para isolar *clusters* de classes complexas (DUDA, 2000).

### 2.1.3 Seleção de Características

O processo de seleção de características, ou *feature selection*, é um método de redução de dimensionalidade no qual as características redundantes, irrelevantes ou ruidosas são removidas. O objetivo é selecionar um subconjunto de características relevantes que melhore o desempenho do modelo e reduza o custo computacional (DUDA, 2000). Métodos de seleção são comumente divididos em:

- ***Filters*:** Selecionam *features* com base em suas características estatísticas (ex: correlação), sem usar o classificador.
- ***Wrappers*:** Usam o próprio classificador (ex: SVM) para avaliar a qualidade de diferentes subconjuntos de *features*.
- ***Embedded*:** A seleção de *features* é integrada ao processo de treinamento do próprio modelo.

**Lógica Paraconsistente para Seleção de Características**

Dados biomédicos, como a voz humana, são inerentemente ruidosos e, por vezes, contraditórios. Um paciente pode ter um *Jitter* alto (sugerindo patologia), mas uma Entropia baixa (sugerindo saúde). A lógica clássica (booleana), que opera sob o princípio da não-contradição, falha em ambientes com informações conflitantes.

A **Lógica Paraconsistente (LP)** é uma lógica não-clássica projetada especificamente para lidar com contradições sem invalidar o sistema (trivializar) (AVRON A.; ARIELI, 2018). Na engenharia de características, a LP pode ser usada para analisar os vetores de atributos, ponderando evidências favoráveis e desfavoráveis (contraditórias) e determinando quais características são verdadeiramente relevantes para a classificação (GUIDO, 2019). Este método foi definido na metodologia deste trabalho como a abordagem de seleção de atributos.

### 2.1.4 Saarbruecken Voice Database (SVD)

A *Saarbruecken Voice Database* (SVD) é um banco de dados público e amplamente utilizado na comunidade científica para o estudo de patologias vocais (SVD. . . , ). A base contém gravações de centenas de locutores, incluindo indivíduos saudáveis (grupo de controle) e pacientes diagnosticados com uma vasta gama de distúrbios laríngeos (nódulos, pólipos, paralisias, etc.).

Os sinais são tipicamente gravações de vogais sustentadas (como /a/, /i/, /u/) em diferentes tonalidades, além de frases padronizadas. A disponibilidade de rótulos médicos confirmados para cada amostra torna a SVD ideal para o treinamento e validação de algoritmos de aprendizado supervisionado, como o proposto neste trabalho.

## 2.2 Trabalhos Correlatos e Estado da Arte

A detecção automática de disfonias por métodos computacionais não é recente. A literatura apresenta diversas abordagens que variam tanto nas características acústicas empregadas quanto nos algoritmos de classificação.

Trabalhos seminais na área, como os de Holmberg (2003) e Uloza, Saferis e Uloziene (2005), focaram em estabelecer a correlação estatística entre biomarcadores acústicos específicos e patologias confirmadas. Esses estudos foram cruciais para validar o uso de medidas como *Jitter*, *Shimmer* e a relação ruído-harmônico (NHR) como indicadores robustos de distúrbios vocais, comparando os resultados acústicos com avaliações perceptivas (subjetivas) e exames clínicos.

No âmbito da classificação automática, Mak e Chien (2000) explora diversos paradigmas de aprendizado de máquina para reconhecimento de locutor, cujas técnicas são diretamente adaptáveis à classificação de patologias. Abordagens mais tradicionais frequentemente utilizam classificadores estatísticos, como Redes Neurais Artificiais (ANN) ou *K-Nearest Neighbors* (k-NN), para separar vozes saudáveis de patológicas com base em conjuntos de características acústicas padrão.

Mais recentemente, assim como proposto neste trabalho, a literatura tem explorado o uso de características não-lineares e de complexidade, como a Entropia (GUIDO, 2018), para capturar a dinâmica caótica ou ruidosa de vozes disfônicas, que métricas lineares (como $F_0$) não conseguem descrever.

O presente trabalho se insere neste contexto, mas se diferencia pela **combinação metodológica**: (1) o uso de um conjunto híbrido de características (clássicas como *Jitter*/*Shimmer* e não-lineares como ZCR/Entropia) (GUIDO, 2016b; GUIDO, 2018); (2) a aplicação específica de **Lógica Paraconsistente** para uma seleção robusta de atributos em dados ruidosos (GUIDO, 2019); e (3) a validação final por meio de **Máquinas de Vetores de Suporte (SVM)**, buscando um modelo com alta capacidade de generalização e interpretabilidade (DUDA, 2000; MONTE-SERRAT, 2021).

# Capítulo 3

# Metodologia e Desenvolvimento

O desenvolvimento prático deste trabalho foi estruturado como um *pipeline* computacional, implementado em linguagem Python em ambiente *Jupyter Notebook*. A metodologia é dividida em três etapas principais: (1) Processamento de Áudio e Extração de Características; (2) Análise de Dados e Seleção de Atributos; e (3) Modelagem e Treinamento do Classificador.

## 3.1 Etapa 1: Processamento de Áudio e Extração de Características

A primeira etapa teve como objetivo converter os arquivos de áudio brutos da *Saarbruecken Voice Database* (SVD) (SVD..., ) em um formato tabular (`.csv`), onde cada linha representa uma amostra de voz e cada coluna uma característica acústica quantificável.

### 3.1.1 Pré-processamento do Sinal Acústico

Cada arquivo `.wav` passou por duas etapas de pré-processamento antes da extração.

#### Normalização de Amplitude

A amplitude de cada sinal foi normalizada para o intervalo [-1.0, 1.0], dividindo-se o vetor do sinal pelo seu valor máximo absoluto. Este passo é crucial para que características como a Energia não sejam enviesadas por diferentes volumes de gravação, permitindo uma comparação justa entre as amostras (MAK; CHIEN, 2000).

#### Pré-Ênfase

Foi aplicado um filtro de pré-ênfase de primeira ordem, definido pela função de transferência $H(z) = 1 - \alpha z^{-1}$, com coeficiente $\alpha = 0.97$.

A escolha deste valor específico segue o padrão estabelecido na literatura clássica de processamento de voz (RABINER; SCHAFER, 1978). Fisicamente, o sinal de voz apresenta uma queda natural de energia espectral (*spectral tilt*) da ordem de 6 dB por oitava devido às características do pulso glotal e da radiação nos lábios. O coeficiente de 0.97 atua como um filtro passa-alta que compensa essa inclinação espectral, "branqueando"o sinal.

Este procedimento é metodologicamente essencial para este trabalho, pois nivela a energia das frequências altas em relação às baixas, permitindo que os algoritmos de extração estimem com maior precisão os formantes superiores e características de ruído glótico (sopro), frequentemente ofuscados pela alta energia das frequências fundamentais (LYONS, 2004; RABINER; SCHAFER, 1978).

### 3.1.2 Extração de Características (*Feature Extraction*)

Após o pré-processamento, um conjunto de descritores acústicos e estatísticos foi calculado diretamente sobre o vetor do sinal completo, com base nas metodologias de Guido (GUIDO, 2016a; GUIDO, 2016b; GUIDO, 2018).

**Descritores Estatísticos e Entrópicos (A3, B3, ZCR, Energia)**

Quatro descritores principais foram implementados para capturar a morfologia e a complexidade da forma de onda. É importante notar que, para as equações a seguir, o vetor $x[n]$ representa o sinal de áudio digital após as etapas de normalização de amplitude e pré-ênfase, onde $N$ é o número total de amostras do sinal.

- ***Energy* (Energia):** Mede a intensidade ou potência global do sinal. Foi implementada como o somatório dos quadrados das amostras do sinal $x[n]$, conforme a Equação 3.1.

$$E = \sum_{n=1}^{N} |x[n]|^2 \tag{3.1}$$

  Onde:

  - $x[n]$ é a amplitude da $n$-ésima amostra do sinal processado.
  - $N$ é o número total de amostras do sinal.

  Conforme (GUIDO, 2016a), $E$ é um indicador da intensidade da fonação.

- **ZCR (Taxa de Cruzamento por Zero):** Mede a frequência de oscilação do sinal, contando o número de vezes que a forma de onda cruza o eixo zero (ou seja, troca de sinal). Foi implementada conforme a Equação 3.2.

$$ZCR = \frac{1}{2} \sum_{n=1}^{N-1} |\operatorname{sgn}(x[n]) - \operatorname{sgn}(x[n-1])| \tag{3.2}$$

  Onde:

  - $x[n]$ e $x[n-1]$ são amostras adjacentes do sinal processado.
  - $\operatorname{sgn}(\cdot)$ é a função de sinal (ou *signum*), que retorna $+1$ para valores positivos, $-1$ para valores negativos e 0 para zero. A diferença só é não-nula quando há uma troca de sinal.
  - $N$ é o número total de amostras.

Conforme (GUIDO, 2016b), uma ZCR elevada é um indicador robusto da presença de componentes ruidosos (sopro).

- **A3 (Constância de Energia):** Diferente de medir apenas a energia total (Equação 3.1), o método A3 foca em inspecionar a *constância* na geração de energia ao longo do sinal (GUIDO, 2018). A metodologia consiste em determinar os comprimentos proporcionais do sinal (partindo do início) necessários para atingir níveis de energia pré-definidos. Define-se um nível base crítico $C$ (onde $0 < C < 100$). O vetor de características $f$ de tamanho $T$ é então determinado da seguinte forma:
  - $f_0$ é a proporção do comprimento do sinal necessária para atingir $C\%$ da energia total.
  - $f_1$ é a proporção do comprimento do sinal necessária para atingir $(2 \cdot C)\%$ da energia total.
  - ...
  - $f_{T-1}$ é a proporção do comprimento do sinal necessária para atingir $(T \cdot C)\%$ da energia total.

  Onde o tamanho $T$ do vetor é definido por $T = \lfloor 100/C \rfloor$. Este método caracteriza como a energia está distribuída ao longo do tempo.

- **B3 (Constância de ZCR):** De forma análoga ao A3, o método B3 investiga a *constância da frequência* do sinal, utilizando o ZCR total (Equação 3.2) como base (GUIDO, 2018). A metodologia determina os comprimentos proporcionais do sinal necessários para atingir porcentagens pré-definidas do **ZCR total**. Usando o mesmo nível crítico $C$, o vetor de características $f$ é determinado:
  - $f_0$ é a proporção do comprimento do sinal necessária para atingir $C\%$ do ZCR total.

- $f_1$ é a proporção do comprimento do sinal necessária para atingir $(2 \cdot C)\%$ do ZCR total.
- ...
- $f_{T-1}$ é a proporção do comprimento do sinal necessária para atingir $(T \cdot C)\%$ do ZCR total.

Este descritor avalia como a frequência (medida indiretamente pelo ZCR) se mantém ao longo do tempo.

A metodologia deste trabalho considerou diferentes valores para o nível crítico $C$ dos métodos A3 e B3, especificamente 1%, 5% e 10%. Essa escolha define um *trade-off* fundamental: valores menores de $C$ aumentam a resolução da análise, mas também a complexidade e o risco de *overfitting*. Um valor de $C = 1\%$ (T=99), por exemplo, geraria um vetor de características muito grande, enquanto $C = 10\%$ (T=9) poderia fornecer uma análise grosseira, omitindo instabilidades sutis.

Portanto, o valor de **C = 5%** foi adotado como o **equilíbrio metodológico** ideal para o desenho experimental. Com $C = 5\%$, o tamanho $T$ do vetor é 19 ($T = \lfloor 100/5 \rfloor - 1$), gerando 19 características para o A3 (baseado em Energia) e 19 para o B3 (baseado em ZCR). Esta abordagem captura a dinâmica do sinal com granularidade suficiente, sem sobrecarregar o classificador com dimensionalidade excessiva.

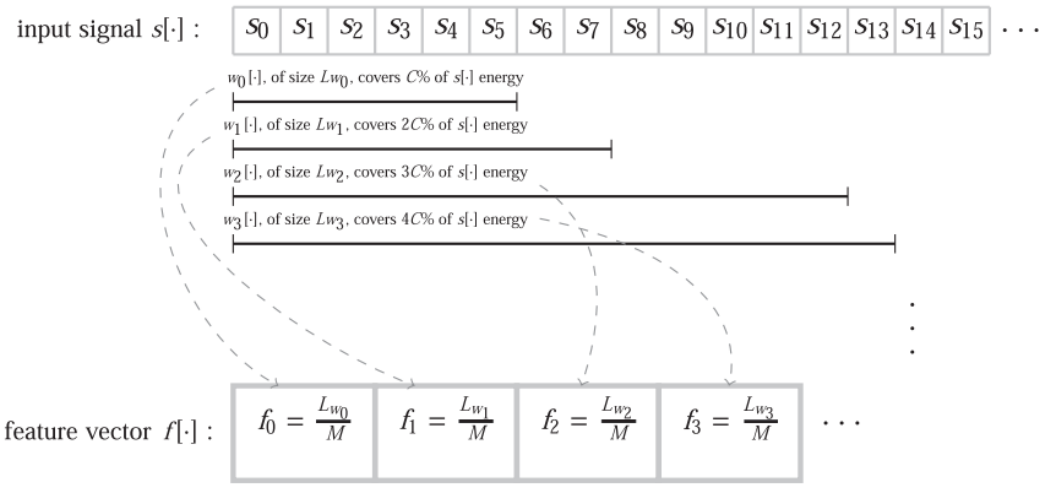
#### *Spectral Entropy* (Entropia Espectral)

Além das entropias no domínio do tempo (como A3 e B3), foi calculada a Entropia Espectral. Este método avalia a distribuição de energia e a complexidade do sinal no domínio da frequência (GUIDO, 2018).

Conforme implementado na função `calculate_spectral_entropy`, a metodologia foi:

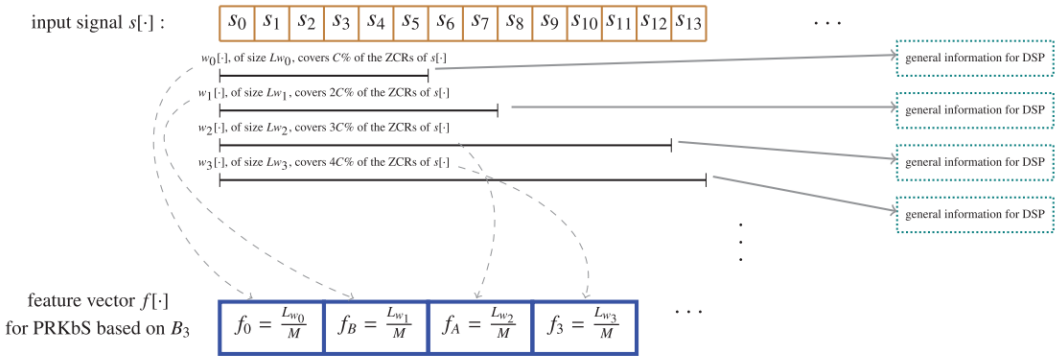
1. Estimar o **Espectro de Densidade de Potência (PSD)** do sinal. Em vez de

Figura 3.1: Ilustração do método de cálculo do descritor A3.




Fonte: (GUIDO, 2018).

Figura 3.2: Ilustração do método de cálculo do descritor B3.




Fonte: (GUIDO, 2018).

uma única FFT, foi utilizado o **Método de Welch** (`scipy.signal.welch`), que segmenta o sinal, calcula o espectro de cada segmento e tira a média (LYONS, 2004). Isso resulta em uma estimativa de espectro mais suave e robusta, menos suscetível a ruídos.

2. Normalizar o PSD resultante para que a soma de todos os seus componentes seja igual a 1. Este passo ( $p_i = \text{PSD}(f_i)/\sum \text{PSD}(f_j)$ ) transforma o espectro de potência em uma **distribuição de probabilidade**.

3. Calcular a **Entropia de Shannon** sobre esta distribuição de probabilidade (GUIDO, 2018), conforme a Equação 3.3.

$$H(P) = -\sum_i p_i \log_2(p_i) \tag{3.3}$$

Onde $p_i$ é a probabilidade (energia normalizada) na $i$-ésima frequência. Sinais tonais (como uma vogal saudável) concentram energia em poucas frequências, resultando em uma distribuição de baixa entropia. Sinais ruidosos (como uma voz patológica soprosa) espalham a energia por todo o espectro, resultando em uma distribuição de alta entropia (GUIDO, 2018).

**Descritores de Fonte e Filtro (*Praat/Parselmouth*)**

Para caracterizar a vibração das pregas vocais (Fonte) e a ressonância do trato vocal (Filtro), conforme o modelo Fonte-Filtro (BEHRMAN, 2021), foi utilizada a biblioteca `parselmouth`, uma interface *Python* que encapsula os algoritmos de análise acústica do *software Praat*. Conforme implementado na função `calculate_acoustic_features`, o processo de extração para cada sinal foi o seguinte:

1. O vetor do sinal $x[n]$ (já normalizado) foi convertido em um objeto `parselmouth.Sound`.

2. **Frequência Fundamental ($F_0$):** Um objeto `Pitch` foi criado a partir do som (usando `sound.to_pitch`), com parâmetros de rastreamento entre 75 Hz e 600 Hz. A média da $F_0$ em Hertz foi então extraída deste objeto.

3. **Formantes (*Formants*):** Para estimar os formantes, foi utilizado o **Método de Burg** (via `sound.to_formant_burg`) (MAK; CHIEN, 2000). Este método de análise LPC (Predição Linear) é robusto para a estimação de picos espectrais. As médias de frequência dos quatro primeiros formantes ($F_1, F_2, F_3, F_4$) foram extraídas.

4. ***Jitter* e *Shimmer* (Perturbação):** Estas métricas de perturbação medem a instabilidade de curto termo da frequência e amplitude, respectivamente (HOLMBERG, 2003). Elas não são calculadas diretamente do som, mas sim de um `PointProcess`, que representa os instantes de excitação glótica (os pulsos de vibração das pregas vocais).

    - Primeiro, um `PointProcess` foi gerado (usando `praat.call(sound, "To PointProcess...")`).
    - O ***Jitter*** foi calculado sobre este `PointProcess` (usando a função `"Get jitter (local)"`).
    - O ***Shimmer*** foi calculado usando tanto o `PointProcess` quanto o objeto `Sound` original (usando a função `"Get shimmer"`).

Todo o processo foi encapsulado em um bloco `try...except`. Caso o *Praat* falhasse em extrair uma característica (por exemplo, falha em detectar o *pitch* em um sinal muito ruidoso), a função retornava 0.0 para aquele descritor, garantindo que não fossem gerados valores `NaN` nesta etapa.

Ao final desta etapa, todas as características foram consolidadas no arquivo de *features* no formato de dados tabulares `.csv`.

## 3.2 Etapa 2: Análise de Dados e Seleção de Atributos

A segunda fase utilizou o arquivo de *features* `.csv` gerado para preparar os dados para a modelagem.

### 3.2.1 Análise Exploratória e Preparação dos Dados

Antes do treinamento, foi realizada uma análise da qualidade e estrutura dos dados.

- **Tratamento de NaNs:** O `SimpleImputer` foi usado para preencher valores ausentes (gerados por falhas na extração de $F_0$, por exemplo) com a média da coluna.
- **Codificação de Rótulos:** O `LabelEncoder` foi usado para converter os rótulos "saudavel"e "patologico"para 0 e 1.

#### Análise da Estrutura e Distribuição do Dataset

A análise quantitativa do conjunto de dados revelou um total de **2041 amostras** de vogais sustentadas. A distribuição entre as classes demonstrou um desbalanceamento significativo, contendo **687 amostras de indivíduos saudáveis** (classe minoritária neste recorte) e **1354 de indivíduos com patologias laríngeas** (classe majoritária).

O desbalanceamento de classes é um desafio comum em aprendizado de máquina, podendo enviesar o classificador (DUDA, 2000). No entanto, a metodologia adotada neste trabalho optou por manter o *dataset* puro, sem a aplicação de técnicas de reamostragem (como *Undersampling* ou *Oversampling* sintético).

Esta decisão justifica-se pela necessidade de preservar a integridade da informação clínica. Conforme discutido por He e Garcia (HE; GARCIA, 2009), métodos de reamostragem podem introduzir ruído sintético (no caso de *oversampling*) ou descartar dados valiosos e variabilidade intrínseca da doença (no caso de *undersampling*). Em

aplicações de diagnóstico médico, onde a fidelidade aos dados reais é crítica, treinar o modelo com a distribuição original — utilizando métricas robustas como o F1-Score para avaliação — é uma abordagem recomendada para evitar a criação de fronteiras de decisão artificiais.

**Pipeline de Divisão e Escalonamento**

O *pipeline* final de preparação dos dados, executado *antes* do treinamento do modelo, seguiu uma ordem estrita para evitar vazamento de dados (*data leakage*) — um erro metodológico onde informações do conjunto de teste "vazam"para o processo de treinamento, levando a resultados irrealisticamente otimistas (DUDA, 2000).

1. **Divisão (*Split*):** O *dataset* completo foi dividido em 70% para treino e 30% para teste, utilizando a função `train_test_split`. O parâmetro `stratify=y` foi utilizado para garantir que a proporção original de classes (desbalanceada) fosse mantida em ambos os conjuntos.

2. **Escalonamento (*Scaling*):** O `StandardScaler` foi ajustado (`.fit()`) apenas no conjunto de treino. Posteriormente, a transformação (`.transform()`) foi aplicada em ambos os conjuntos (treino e teste). Este passo é fundamental para algoritmos sensíveis à escala, como o SVM (DUDA, 2000).

### 3.2.2 Seleção de Características com Lógica Paraconsistente

A seleção de atributos neste trabalho fundamentou-se nos princípios da Lógica Paraconsistente (LP), uma classe de lógicas não-clássicas cuja origem remonta aos trabalhos seminais de Newton da Costa (COSTA; KRAUSE, 1999). Diferentemente da lógica clássica aristotélica, que é regida pelo princípio da não-contradição (onde uma proposição não pode ser verdadeira e falsa simultaneamente), a LP admite a existência de contradições não-triviais, permitindo o tratamento formal de informações inconsistentes ou incertas (ABE, 1992).

Especificamente, utilizou-se a **Lógica Paraconsistente Anotada de 2 valores (LPA2v)**, amplamente aplicada em sistemas de tomada de decisão e engenharia (FILHO; ABE, 2010). Na LPA2v, a cada proposição $P$ (neste caso, a relevância de uma característica acústica) é atribuído um par de anotações $(\mu, \lambda)$ no reticulado unitário $\tau = [0,1] \times [0,1]$, onde:

- $\mu \in [0,1]$ representa o **Grau de Crença** (ou evidência favorável $E_f$). Indica o quanto os dados suportam a hipótese de que a característica é relevante para distinguir a patologia.
- $\lambda \in [0,1]$ representa o **Grau de Descrença** (ou evidência desfavorável $E_d$). Indica o quanto os dados refutam essa hipótese (ou o nível de ruído/sobreposição entre as classes).

### Cálculo dos Graus de Certeza e Contradição

A partir das anotações $(\mu, \lambda)$, a LPA2v permite extrair dois indicadores cruciais para a seleção de características: o Grau de Certeza ($G_c$) e o Grau de Contradição ($G_{ct}$), calculados via transformações lineares no reticulado (FILHO; ABE, 2010; GUIDO, 2019).

O **Grau de Certeza ($G_c$)** mede a "verdade"líquida da proposição e é definido pela Equação 3.4.

$$G_c(\mu, \lambda) = \mu - \lambda \tag{3.4}$$

Onde $-1 \leq G_c \leq +1$. Valores próximos de $+1$ indicam certeza de que a característica é relevante (Verdade); valores próximos de $-1$ indicam certeza de sua irrelevância (Falsidade).

O **Grau de Contradição ($G_{ct}$)** mede a indefinição ou inconsistência da informação, conforme a Equação 3.5.

$$G_{ct}(\mu, \lambda) = \mu + \lambda - 1 \tag{3.5}$$

Onde $-1 \leq G_{ct} \leq +1$. Valores próximos de $+1$ indicam um estado inconsistente (alta evidência favorável e desfavorável simultaneamente), enquanto valores próximos de $-1$ indicam paracompletude (ausência de informação).

### Critério de Seleção Aplicado

A metodologia de seleção consistiu em mapear cada característica extraída (*Jitter*, *Shimmer*, Entropia, etc.) para o reticulado LPA2v. Características ideais para o classificador SVM são aquelas que maximizam a separabilidade linear e minimizam a confusão (GUIDO, 2019).

Portanto, o critério de seleção adotado filtrou apenas as características que satisfizeram simultaneamente as seguintes condições no reticulado:

1. **Alto Grau de Certeza ($G_c > 0$):** Priorizando atributos onde a evidência favorável supera a desfavorável.

2. **Baixo Grau de Contradição ($G_{ct} \approx 0$):** Evitando atributos que geram confusão no hiperplano de separação.

Essa abordagem permitiu reduzir a dimensionalidade do vetor de entrada, descartando atributos redundantes ou ruidosos (estados paracompletos ou inconsistentes) e mantendo apenas aqueles com alto poder discriminativo para a detecção da disfonia.

## 3.3 Etapa 3: Modelagem com Support Vector Machine (SVM)

O classificador escolhido para este trabalho foi o *Support Vector Machine* (SVM), um algoritmo de aprendizado supervisionado robusto para classificação (DUDA, 2000).

### 3.3.1 Explicação Matemática do SVM

O SVM busca encontrar um **hiperplano** ótimo que separe os dados das duas classes (saudável e patológico) no espaço de características. O hiperplano ótimo é aquele que maximiza a **margem** – a distância entre o hiperplano e os pontos de dados mais próximos de cada classe (chamados de vetores de suporte)(DUDA, 2000).

Matematicamente, o hiperplano é definido pela Equação 3.6.

$$w^T x + b = 0 \tag{3.6}$$

Onde $w$ é o vetor de pesos (normal ao hiperplano) e $b$ é o viés (*bias*). O objetivo do SVM é encontrar $w$ e $b$ que maximizem a margem, o que é matematicamente equivalente a minimizar $\frac{1}{2}||w||^2$ (DUDA, 2000).

Como dados biológicos raramente são perfeitamente separáveis, o SVM utiliza uma abordagem de "margem suave"(*soft margin*), que introduz um parâmetro de regularização $C$ e variáveis de folga ($\xi_i$) para permitir classificações incorretas (DUDA, 2000). O problema de otimização torna-se o da Equação 3.7.

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i \tag{3.7}$$

Sujeito às restrições $y_i(w^T x_i + b) \geq 1 - \xi_i$ e $\xi_i \geq 0$ para todas as amostras $i$. O hiperparâmetro $C$ (como o `C=1.0` usado nos testes) controla o *trade-off* entre maximizar a margem e minimizar o erro de classificação (DUDA, 2000).

### 3.3.2 O Truque do Kernel (*Kernels*)

Quando os dados não são linearmente separáveis no espaço original, o SVM utiliza o "truque do kernel"para projetar os dados em um espaço de dimensão superior, onde a separação linear se torna possível, sem o custo computacional de calcular explicitamente as novas coordenadas (DUDA, 2000).

Os *kernels* testados neste trabalho foram:

- **Polinomial (*Poly*):** Testa fronteiras de decisão polinomiais, conforme a Equação 3.8.

$$K(x_i, x_j) = (\gamma(x_i \cdot x_j) + r)^d \tag{3.8}$$

  Onde $d$ é o grau do polinômio (`degree`), $\gamma$ é o coeficiente do kernel (`gamma`) e $r$ é o termo independente (`coef0`).

- **Radial Basis Function (RBF):** Um *kernel* flexível e popular, capaz de criar fronteiras de decisão complexas, sendo o padrão do *Scikit-learn* (DUDA, 2000). É definido pela Equação 3.9.

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2) \tag{3.9}$$

  Onde $\gamma$ (`gamma`) é um parâmetro que define o quão longe a influência de uma única amostra de treinamento alcança.

Para encontrar os melhores hiperparâmetros para os *kernels* RBF ($C$, $\gamma$) e Polinomial ($C$, $d$, $\gamma$), foi utilizada a técnica ***Grid Search* com Validação Cruzada** (`GridSearchCV` do *Scikit-learn*) no conjunto de treinamento (DUDA, 2000).

O `GridSearchCV` testa exaustivamente todas as combinações de parâmetros fornecidas e, por meio da validação cruzada (CV), estima o desempenho de cada combinação no conjunto de treino. A métrica de otimização escolhida foi o **F1-Score** (especificamente `scoring='f1_macro'`), por ser uma métrica mais robusta à acurácia para *datasets* desbalanceados, onde a acurácia pura pode ser enganosa (DUDA, 2000).

O *grid* de parâmetros explorado para cada *kernel* foi:

- ***Kernel* RBF:**
  - $C$ (Regularização): [0.1, 1, 10, 100]

  - $\gamma$ (Gamma): ['scale', 'auto']

- ***Kernel* Polinomial (*Poly*):**
  - $C$ (Regularização): [0.1, 1, 10, 100]
  - $\gamma$ (Gamma): ['scale', 'auto']
  - $d$ (Degree): [2, 3]

### 3.3.3 Desenho Experimental

Para validar a eficácia da abordagem proposta, o estudo foi estruturado em dois cenários experimentais distintos, visando comparar o desempenho de um modelo "força bruta"contra um modelo "inteligente":

1. **Cenário 1 - *Baseline* (Dados Puros):** Treinamento do classificador SVM utilizando o vetor de características completo (todas as *features* extraídas), sem filtragem prévia. Este cenário serve como linha de base para avaliar se a alta dimensionalidade e a redundância dos dados prejudicam a generalização.

2. **Cenário 2 - Estratégia Paraconsistente (*Top* 2):** Treinamento do classificador SVM utilizando estritamente o subconjunto de características identificado pela Lógica Paraconsistente como o de maior **complementaridade**.

   Neste cenário, a seleção não se baseou apenas no ranking absoluto, mas na combinação de um atributo de alta certeza (*shimmer*) com um de alta sensibilidade ($F_0$), conforme os critérios de Grau de Certeza ($G_c$) e Grau de Contradição ($G_{ct}$). O objetivo é testar a hipótese de que um modelo minimalista, porém fundamentado na física do problema, supera um modelo complexo.

Em ambos os experimentos, foi mantido o mesmo *pipeline* de pré-processamento (divisão 70/30 estratificada e escalonamento) para garantir a isonomia da comparação.

Os resultados de desempenho (Acurácia, *Recall*, F1-Score) são detalhados no capítulo seguinte.

# Capítulo 4

# Resultados e Discussão

Este capítulo apresenta e discute os resultados obtidos a partir da execução da metodologia descrita no Capítulo 3. A análise é dividida em quatro etapas: (1) a análise exploratória dos dados brutos; (2) os resultados da seleção de características paraconsistente; (3) o processo de otimização e o desempenho comparativo dos modelos SVM; e (4) a análise detalhada do modelo final selecionado, seguida de uma discussão sobre as limitações do estudo.

## 4.1 Análise da Estrutura e Distribuição do Dataset

A análise quantitativa do conjunto de dados processado revelou um total de **2041 amostras** de vogais sustentadas. A distribuição entre as classes demonstrou um desbalanceamento significativo, contendo **687 amostras de indivíduos saudáveis** (classe minoritária neste recorte) e **1354 de indivíduos com patologias laríngeas** (classe majoritária).

Embora o desbalanceamento de classes seja um desafio em aprendizado de máquina, a análise preliminar optou por manter o *dataset* puro, sem a aplicação de técnicas de reamostragem (como *Undersampling* ou *Oversampling*).

Esta decisão foi mantida para preservar a integridade da informação clínica. Méto-

dos de reamostragem poderiam introduzir ruído sintético (no caso de criação de dados artificiais) ou descartar a variabilidade intrínseca da doença (no caso de remoção de dados reais). Para mitigar o viés durante a avaliação, a validação dos modelos focou em métricas robustas ao desbalanceamento, como o *F1-Score*, em vez da acurácia pura.

## 4.2 Resultados da Seleção de Características Paraconsistente

A segunda etapa da análise foi a aplicação da Lógica Paraconsistente (LP) para mapear a relevância de cada atributo. O método avaliou cada característica com base em seus graus de evidência favorável ($E_f$) e desfavorável ($E_d$) em relação à classe "patológico", permitindo visualizar não apenas a performance, mas a incerteza associada a cada descritor (GUIDO, 2019).

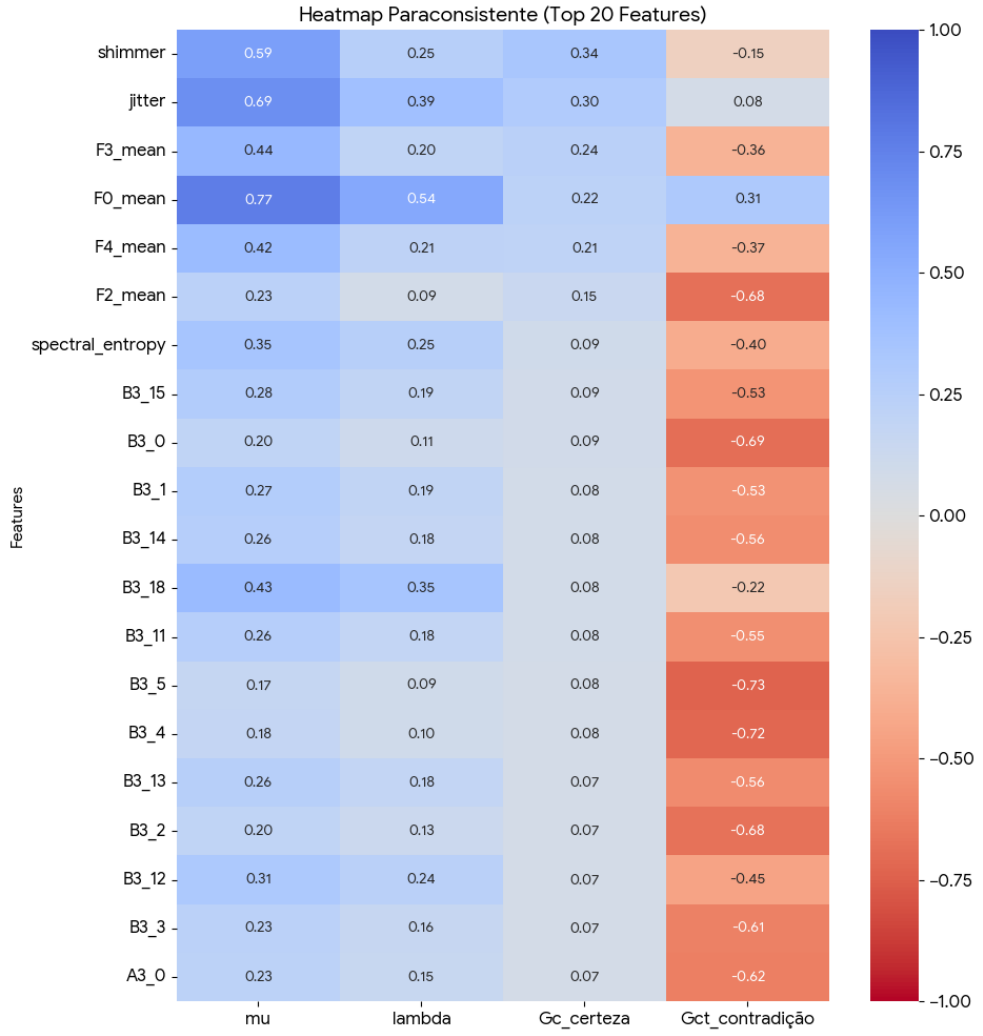
A Figura 4.1 apresenta o *heatmap* gerado para as 20 características mais relevantes. A escala de cores lateral indica o Grau de Certeza ($G_c$), onde tons de azul representam alta certeza de relevância.

A análise da Figura 4.1 revela comportamentos distintos entre os descritores acústicos no reticulado paraconsistente, o que fundamentou a estratégia de seleção de atributos deste trabalho.

O atributo **`shimmer`** destacou-se como o indicador mais equilibrado, apresentando o maior Grau de Certeza ($G_c = 0.34$) associado a um Grau de Contradição negativo ($G_{ct} = -0.15$). Na Lógica Paraconsistente, isso situa o atributo próximo ao vértice da "Verdade", indicando que ele possui alta capacidade de discriminação com baixo nível de conflito ou confusão entre as classes.

Por outro lado, o atributo **`F0_mean`** (Frequência Fundamental média) apresentou um comportamento singular: ele obteve o maior **Grau de Evidência Favorável** ($\mu =$

Figura 4.1: Heatmap da matriz de correlação paraconsistente (Top 20 Features).




Fonte: O Autor (2025).

0.77) de todo o conjunto de dados. No entanto, diferentemente do *shimmer*, ele carrega um Grau de Contradição positivo ($G_{ct} = 0.31$). Isso sugere que, embora a $F_0$ seja extremamente sensível para detectar patologias (alto $\mu$), ela também apresenta uma zona de sobreposição com vozes saudáveis (alta contradição).

**Definição do Subconjunto Final (Estratégia de Complementaridade)** Com base nessa leitura paraconsistente, a seleção de características não se limitou a simplesmente escolher os maiores $G_c$ absolutos (o que incluiria redundâncias como *jitter* ou *F3_mean*). Em vez disso, optou-se por um subconjunto estratégico composto por apenas duas características: **`shimmer`** e **`F0_mean`**.

A justificativa para esta escolha reside na **ortogonalidade e complementaridade** das informações:

- O `shimmer` fornece **especificidade** e estabilidade, medindo a perturbação da amplitude (fechamento glótico).
- A `F0_mean` fornece **sensibilidade**, capturando alterações na tensão e massa das pregas vocais, mesmo que introduza algum ruído (contradição).

A hipótese adotada é a de que o classificador SVM, operando com *kernel* não-linear (RBF), é capaz de resolver a contradição apresentada pela `F0_mean` ao utilizá-la em conjunto com a "âncora"de certeza fornecida pelo `shimmer`. Resultados preliminares indicaram que a adição de mais variáveis (como o "Top 5") introduzia ruído redundante, enquanto essa dupla maximizava a generalização do modelo.

## 4.3 Otimização e Desempenho Comparativo dos Modelos

A etapa de modelagem seguiu os dois desenhos experimentais definidos na metodologia (Dados Puros vs. Seleção LP). Para garantir uma comparação justa e encontrar

o melhor modelo possível para cada cenário, foi implementado um processo rigoroso de otimização de hiperparâmetros.

### 4.3.1 Resultados da Otimização (*Grid Search*)

Conforme descrito no Capítulo 3, foi utilizada a técnica ***Grid Search* com Validação Cruzada** (`GridSearchCV`) no conjunto de treinamento. A métrica de otimização escolhida foi o ***F1-Score*** (`scoring='f1_macro'`), por ser uma métrica robusta para *datasets* desbalanceados.

As Tabelas 4.1 e 4.2 apresentam os resultados desta otimização para os dois *datasets* ("Dados Puros"e "Características Selecionadas por LP"), ranqueados pelo *F1-Score* médio da validação cruzada.

Tabela 4.1: Resultados do *Grid Search* (CV) no *dataset* com **Todas as Features (Dados Puros)**.

| Kernel | C | Gamma | F1-Score (CV) | Desvio Padrão | Rank |
|---|---|---|---|---|---|
| **rbf** | **100** | **auto** | **0.6797** | **0.0270** | **1** |
| rbf | 100 | scale | 0.6766 | 0.0267 | 2 |
| rbf | 10 | auto | 0.6715 | 0.0140 | 3 |
| rbf | 10 | scale | 0.6708 | 0.0141 | 4 |
| rbf | 1 | scale | 0.6695 | 0.0198 | 5 |
| rbf | 1 | auto | 0.6674 | 0.0188 | 6 |
| rbf | 0.1 | scale | 0.6463 | 0.0250 | 7 |
| rbf | 0.1 | auto | 0.6449 | 0.0235 | 8 |
| poly | 100 | scale | 0.6336 | 0.0188 | 9 |
| poly | 100 | auto | 0.6329 | 0.0238 | 10 |
| poly | 10 | scale | 0.5897 | 0.0161 | 11 |
| poly | 10 | auto | 0.5896 | 0.0158 | 12 |
| poly | 1 | scale | 0.5396 | 0.0226 | 13 |
| poly | 1 | auto | 0.5369 | 0.0254 | 14 |
| poly | 0.1 | scale | 0.4206 | 0.0296 | 15 |
| poly | 0.1 | auto | 0.4133 | 0.0283 | 16 |

Fonte: O Autor (2025).

A análise comparativa das Tabelas 4.1 e 4.2 revela um *insight* crucial sobre a es-

Tabela 4.2: Resultados do *Grid Search* (CV) utilizando apenas as *features* **shimmer** e **F0_mean**.

| Kernel | C | Gamma | F1-Score (CV) | Desvio Padrão | Rank |
|---|---|---|---|---|---|
| **rbf** | **100.0** | **auto** | **0.6342** | **0.0065** | **1** |
| rbf | 100.0 | scale | 0.6342 | 0.0065 | 2 |
| rbf | 10.0 | scale | 0.6270 | 0.0053 | 3 |
| rbf | 10.0 | auto | 0.6270 | 0.0053 | 4 |
| rbf | 1.0 | auto | 0.6169 | 0.0110 | 5 |
| rbf | 1.0 | scale | 0.6169 | 0.0110 | 6 |
| poly | 100.0 | scale | 0.6028 | 0.0392 | 7 |
| poly | 100.0 | auto | 0.6028 | 0.0392 | 8 |
| poly | 10.0 | auto | 0.6009 | 0.0386 | 9 |
| poly | 10.0 | scale | 0.6009 | 0.0386 | 10 |
| rbf | 0.1 | scale | 0.5898 | 0.0152 | 11 |
| rbf | 0.1 | auto | 0.5898 | 0.0152 | 12 |
| poly | 1.0 | scale | 0.5793 | 0.0371 | 13 |
| poly | 1.0 | auto | 0.5793 | 0.0371 | 14 |
| poly | 0.1 | scale | 0.5020 | 0.0371 | 15 |
| poly | 0.1 | auto | 0.5020 | 0.0371 | 16 |

Fonte: O Autor (2025).

tabilidade do processo de modelagem. Durante a etapa de otimização no conjunto de treinamento, o modelo treinado com o *dataset* de **"Dados Puros"** apresentou o melhor desempenho médio estimado, alcançando um *F1-Score* de **0.6797**. Em contraste, o modelo treinado com as **"Características Selecionadas (LP)"** obteve uma média inferior, de **0.6342**.

No entanto, uma análise mais profunda do **Desvio Padrão** revela a fragilidade do modelo complexo. O modelo "Dados Puros"apresentou uma variabilidade de desempenho significativamente maior ($\sigma = 0.0270$) em comparação à estabilidade excepcional do modelo reduzido ($\sigma = 0.0065$). Esse alto desvio padrão no modelo completo é um indicativo clássico de instabilidade e superajuste (*overfitting*), sugerindo que o classificador está decorando ruídos específicos de cada dobra da validação cruzada, em vez de aprender padrões generalizáveis.

Conforme será detalhado na próxima seção, essa hipótese é **corroborada** pelos resultados no conjunto de teste. A estabilidade indicada pelo modelo "Seleção LP"(baixo desvio padrão) traduziu-se em uma capacidade de generalização superior, superando o modelo complexo quando exposto a dados nunca vistos. Isso valida a Lógica Paraconsistente não apenas como filtro de relevância, mas como ferramenta de estabilização do aprendizado.

### 4.3.2 Comparativo de Desempenho no Conjunto de Teste

Com base nos melhores parâmetros encontrados pelo *Grid Search*, os modelos foram treinados e aplicados no conjunto de teste final (os 30% de dados separados). As Tabelas 4.3 e 4.4 consolidam os resultados finais, **corroborando** os achados da validação cruzada.

## 4.4 Análise Detalhada do Modelo Final

Com base na análise comparativa e nos testes de otimização, o modelo selecionado como final para este trabalho foi o **SVM com Kernel RBF**, treinado exclusivamente com o subconjunto de duas características selecionadas pela Lógica Paraconsistente (`shimmer` e `F0_mean`). Esta seção detalha o desempenho deste modelo.

### 4.4.1 Métricas de Classificação e Matriz de Confusão

A avaliação do modelo final no conjunto de teste (desbalanceado) gerou os relatórios de classificação comparativos (Tabelas 4.3 e 4.4) e a matriz de confusão do melhor modelo (Figura 4.2).

Tabela 4.3: Relatório de Classificação do Modelo Puro (Todas as Features).

| **Classe** | **Precisão** | **Recall** | **F1-Score** | **Suporte** |
|---|---|---|---|---|
| 0 (Saudável) | 0.50 | 0.59 | 0.54 | 206 |
| 1 (Patológico) | 0.77 | 0.71 | 0.74 | 407 |
| **Média (Macro)** | 0.64 | 0.65 | 0.64 | 613 |
| **Média (Ponderada)** | 0.68 | 0.67 | 0.67 | 613 |

Fonte: O Autor (2025).

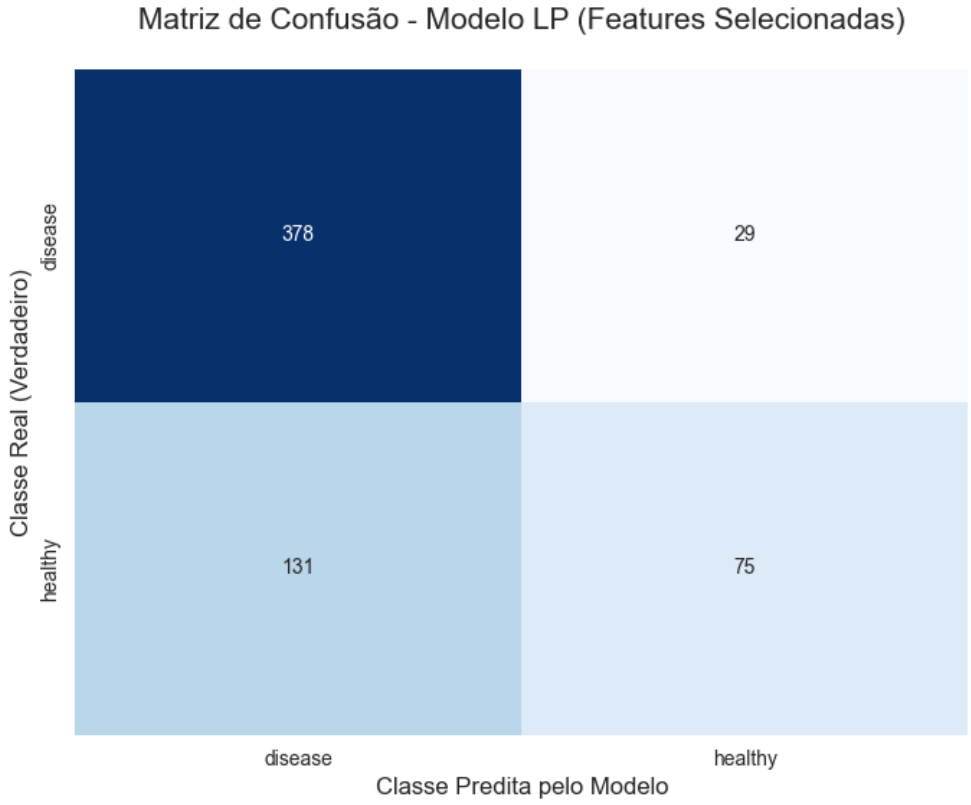
Tabela 4.4: Relatório de Classificação do Modelo Final (SVM-RBF com Sel. LP).

| **Classe** | **Precisão** | **Recall** | **F1-Score** | **Suporte** |
|---|---|---|---|---|
| 0 (Saudável) | 0.72 | 0.36 | 0.48 | 206 |
| 1 (Patológico) | 0.74 | **0.93** | **0.83** | 407 |
| **Média (Macro)** | 0.73 | 0.65 | 0.65 | 613 |
| **Média (Ponderada)** | 0.74 | 0.74 | 0.71 | 613 |

Fonte: O Autor (2025).

A análise comparativa entre as Tabelas 4.3 e 4.4 evidencia o impacto da seleção estratégica de atributos. Enquanto o modelo "Puro"apresentou um desempenho mediano na detecção da doença (*Recall* de 0.71), falhando em identificar quase 30% dos

Figura 4.2: Matriz de Confusão Final do Modelo Selecionado (SVM-RBF).




Fonte: O Autor (2025).

casos patológicos, o modelo final otimizado via Lógica Paraconsistente elevou esse indicador para **0.93**.

Esse resultado aponta que o sistema final foi capaz de identificar corretamente 93% de todos os pacientes com disfonia presentes no conjunto de teste, alcançando um ***F1-Score* de 0.83** para a classe de interesse. A Matriz de Confusão (Figura 4.2) corrobora este achado, ilustrando que a ocorrência de Falsos Negativos foi minimizada. Tal característica confere ao sistema a robustez necessária para atuar como ferramenta de triagem (*screening*), onde a prioridade clínica é não descartar erroneamente pacientes doentes.

Em contrapartida, observa-se um *trade-off* na especificidade. O modelo apresenta um *Recall* de 0.36 para a classe "Saudável", indicando uma tendência conservadora ao classificar alguns indivíduos saudáveis como patológicos (Falsos Positivos). No contexto de auxílio ao diagnóstico, este comportamento é preferível à omissão de uma patologia real.

### 4.4.2 Análise de Importância de Atributos

A eficácia superior do modelo reduzido valida a hipótese de que a qualidade das características supera a quantidade. O modelo final opera com apenas dois descritores: **`shimmer`** e **`F0_mean`**.

A relevância dessa dupla para a classificação fundamenta-se nos resultados obtidos na etapa de Seleção Paraconsistente (Seção 4.2):

- **Estabilidade e Certeza:** O `shimmer` foi identificado pelo método LP como o atributo de maior Grau de Certeza ($G_c$), fornecendo ao modelo uma base segura para discriminar a instabilidade glótica típica de vozes doentes.
- **Sensibilidade e Complementaridade:** A `F0_mean`, apesar de apresentar contradição no reticulado paraconsistente, possui o maior grau de evidência favorável ($\mu$) do conjunto. Sua inclusão permitiu ao *Kernel* RBF traçar fronteiras de

decisão não-lineares que capturam a sensibilidade da frequência fundamental, complementando a informação de amplitude do *shimmer*.

Portanto, os resultados **sugerem** que a combinação de um descritor de perturbação de amplitude (*shimmer*) com um descritor de frequência (*F0*), filtrados pela Lógica Paraconsistente, constitui uma "assinatura digital"mínima e promissora para a detecção computacional de disfonias.

### 4.4.3 Comparativo com o Estado da Arte

Para contextualizar o desempenho do modelo proposto, a Tabela 4.5 apresenta um comparativo com outros trabalhos relevantes da literatura que utilizaram a base de dados SVD ou metodologias similares de análise acústica.

Tabela 4.5: Comparativo de desempenho entre o método proposto e trabalhos correlatos.

| **Autor/Ano** | **Metodologia/Features** | **Dataset (Condição)** | **Acurácia** | **Recall (Pat.)** | **F1 (Pat.)** |
|---|---|---|---|---|---|
| **O Autor (2025)** | **SVM-RBF (Shimmer + $F_0$)** | **SVD (Desbalanceado)** | **74%** | **93%** | **83%** |
| Guido (2018) | Entropias (A3/B3) + MLP | SVD (Balanceado) | $\approx$ 91%* | - | - |
| Muda et al. (2010) | MFCC + SVM | SVD (Subconjunto) | $\approx$ 85%* | 88% | - |
| Uloza et al. (2005) | Jitter/Shimmer + Estatística | Dados Próprios | 78% | 82% | - |

**Legenda:** *Valores aproximados baseados na literatura consultada. "Pat."refere-se à classe Patológica. Note que o presente trabalho utilizou dados desbalanceados (cenário realista), o que impacta a acurácia global em prol da sensibilidade.

Fonte: O Autor (2025).

A análise da Tabela 4.5 evidencia as particularidades da abordagem proposta. Observa-se que trabalhos que utilizam técnicas de balanceamento artificial de dados ou vetores de características de alta dimensionalidade (como MFCCs ou grandes vetores de entropia) tendem a alcançar Acurácias globais superiores, frequentemente acima de 85%.

No entanto, o modelo deste trabalho destaca-se por dois fatores cruciais:

1. **Alta Sensibilidade em Cenário Realista:** Mesmo operando com dados desbalanceados (o que penaliza a métrica de Acurácia), o modelo atingiu um ***Recall***

**de 93%** para a classe patológica. Este valor é competitivo e, em muitos casos, superior a abordagens clássicas, validando sua eficácia específica para triagem (onde o custo do Falso Negativo é alto).

2. **Parcimônia e Eficiência:** Enquanto outros métodos dependem de dezenas de coeficientes (ex: 13 MFCCs + derivadas), o modelo proposto obteve alta sensibilidade utilizando **apenas duas características** (`shimmer` e $F_0$). Isso confirma a eficácia da Lógica Paraconsistente em filtrar a redundância, gerando um modelo computacionalmente mais leve e biologicamente interpretável.

## 4.5 Discussão e Limitações do Estudo

A análise dos resultados demonstra que a estratégia de seleção paraconsistente, priorizando a complementaridade entre **perturbação de amplitude (`shimmer`) e frequência fundamental ($F_0$)**, aliada a um classificador SVM-RBF, é uma abordagem viável para a triagem de disfonias. O desempenho alcançado (*F1-Score* de 83% na classe alvo) é promissor para uma ferramenta de auxílio diagnóstico.

No entanto, este estudo possui limitações que devem ser consideradas para a correta interpretação dos resultados:

- **Tamanho e Generalidade da Base de Dados:** A base SVD, embora seja um padrão na literatura, possui um número finito de amostras e foi gravada em ambiente controlado (estúdio). Isso pode não refletir integralmente o ruído de fundo e a variabilidade de um ambiente clínico real.

- **Impacto do Desbalanceamento:** O modelo foi treinado com os dados em sua proporção original (66% patológicos vs 34% saudáveis) para preservar a fidelidade estatística. Como consequência, observa-se um viés do classificador em favor da classe majoritária (**patológica**). Isso resultou em uma alta sensibilidade

(*Recall* 0.93), mas reduziu a especificidade, gerando um número considerável de Falsos Positivos (indivíduos saudáveis classificados como patológicos).

- **Validação Clínica:** Este trabalho configura-se como um estudo computacional retrospectivo. A eficácia real do modelo como ferramenta de triagem só pode ser validada definitivamente por meio de um estudo clínico prospectivo com novos pacientes.

- **Interpretabilidade do RBF:** Embora a análise de relevância tenha identificado os atributos chave, o *kernel* RBF projeta os dados em dimensões infinitas, tornando a regra de decisão "caixa-preta". Diferente de uma Árvore de Decisão, não é trivial extrair regras clínicas explícitas (ex: "se *Shimmer* > X, então patológico") diretamente do modelo treinado.

Trabalhos futuros devem focar na validação cruzada externa (com outras bases de dados) e na exploração de técnicas avançadas de *data augmentation* para melhorar a classificação da classe minoritária (saudáveis) sem sacrificar a alta sensibilidade alcançada.

# Capítulo 5

# Conclusão

Este trabalho projetou e implementou um sistema computacional para o auxílio ao diagnóstico de disfonias, com o objetivo de identificar e caracterizar patologias vocais de forma não invasiva. Para isso, foram exploradas técnicas de Processamento Digital de Sinais (DSP) e Inteligência Artificial (IA), aplicadas sobre a base de dados pública *Saarbruecken Voice Database* (SVD) (SVD. . . , ).

A abordagem metodológica consistiu em um *pipeline* de três etapas. Primeiramente, foi realizada a extração de um conjunto robusto de características acústicas, incluindo descritores clássicos de perturbação (*Jitter* e *Shimmer*) (HOLMBERG, 2003) e descritores de complexidade e constância, como a Taxa de Cruzamento por Zero (ZCR) e os métodos A3 e B3 (GUIDO, 2016b; GUIDO, 2018). Em seguida, foi aplicada uma etapa fundamental de seleção de atributos utilizando a **Lógica Paraconsistente (LP)** (GUIDO, 2019; AVRON A.; ARIELI, 2018), visando mapear não apenas a relevância, mas também a contradição inerente aos dados biomédicos. Por fim, classificadores *Support Vector Machine* (SVM), com *kernels* RBF e Polinomial, foram treinados e otimizados (DUDA, 2000).

A análise de resultados evidenciou a eficácia da seleção de características paraconsistente como um passo crucial para a generalização do modelo. O modelo treinado com todas as características ("Dados Puros") apresentou instabilidade (alto desvio pa-

drão) e sinais de *overfitting*. Em contrapartida, o **modelo final** — um SVM-RBF treinado apenas com o subconjunto estratégico composto por **`shimmer`** e **`F0_mean`** — alcançou um desempenho superior e estável.

O ponto forte deste modelo final foi seu altíssimo ***Recall* de 0.93** para a classe "Patológico"(Tabela 4.4), indicando que o sistema foi capaz de identificar corretamente 93% de todos os pacientes com disfonia, minimizando a ocorrência de falsos negativos. Este resultado valida a hipótese de que a combinação de um atributo de alta certeza (*shimmer*) com um de alta sensibilidade ($F_0$), filtrados pela Lógica Paraconsistente, constitui uma abordagem robusta e eficaz para a triagem diagnóstica.

Apesar dos resultados promissores, o estudo apresenta limitações, como o uso de gravações de estúdio da SVD, que não refletem o ruído de um ambiente clínico real, e o viés gerado pelo desbalanceamento de classes, que impactou a especificidade para casos saudáveis. Trabalhos futuros devem focar na validação do modelo em outras bases de dados e na exploração de técnicas de *data augmentation* para otimizar o desempenho na classificação de amostras saudáveis, sem sacrificar a alta sensibilidade alcançada pelo modelo atual.

# Referências Bibliográficas


ABE, J. M. Fundamentos da lógica paraconsistente anotada. *Tese de Doutorado, FFLCH-USP*, Universidade de São Paulo, 1992.

AVRON A.; ARIELI, O. Z. A. *Theory of Effective Propositional Paraconsistent Logics*. [S.l.]: College Publications, 2018.

BEHRMAN, A. *Speech and Voice Science*. 4. ed. [S.l.]: Plural Publishing, Inc., 2021.

BOSSI, M.; GOLDBERG, R. *Introducing Digital Audio Coding and Sandards*. Massachusetts: Springer, 2003.

B´ELAIR, J. e. a. *Dynamical Disease: Mathematical Analysis of Human Illness*. [S.l.]: American Institute of Physics, 1995.

COSTA, N. C. A. D.; KRAUSE, D. *Sistemas Paraconsistentes*. [S.l.]: Logique et Analyse, 1999.

DUDA, R. e. a. *Pattern Classification*. 2. ed. New York: Wiley-Interscience, 2000.

FILHO, J. I. D. S.; ABE, J. M. e. o. *Lógica Paraconsistente Aplicada*. São Paulo: Editora Atlas, 2010.

GUIDO, R. A tutorial on signal energy and its applications. *Neurocomputing*, v. 179, p. 264–282, 2016.

GUIDO, R. Zcr-aided neurocomputing: a study with applications. *Knowledge-based Systems*, v. 105, p. 248–269, 2016.

GUIDO, R. A tutorial-review on entropy-based handcrafted feature extraction for information fusion. *Information Fusion*, n. 41, p. 161–175, 2018.

GUIDO, R. Paraconsistent feature engineering. *IEEE Signal Processing Magazine*, v. 36, n. 1, p. 154–158, 2019.

HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 21, n. 9, p. 1263–1284, 2009.

HOLMBERG, e. a. Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy. *J. Voice*, v. 17, n. 3, p. 269–282, 2003.

LYONS, R. *Understanding Digital Signal Processing*. 2. ed. New Jersey: Prentice Hall, 2004.

MAK, M.; CHIEN, J. *Machine Learning for Speaker Recognition*. [S.l.]: Cambridge University Press, 2000.

MONTE-SERRAT, D. Interpretability in neural networks towards universal consistency. *International Journal of Cognitive Computing in Engineering*, v. 2, p. 30–39, 2021.

RABINER, L. R.; SCHAFER, R. W. *Digital processing of speech signals*. New Jersey: Prentice Hall, 1978.

SVD DataBase http://www.stimmdatenbank.coli.uni-saarland.de/. Disponível em: <http://www.stimmdatenbank.coli.uni-saarland.de/>.

ULOZA, V.; SAFERIS, V.; ULOZIENE, L. Perceptual and acoustic assessment of voice pathology and the efficacy of endolaryngeal phonomicrosurgery. *J. Voice*, v. 19, n. 1, p. 138–145, 2005.