

RAFAEL MOTTA SCARPA

Detecção Acústica de Acometimento por Covid-19

São José do Rio Preto
2023

RAFAEL MOTTA SCARPA

Detecção Acústica de Acometimento por Covid-19

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

São José do Rio Preto

2023

S286d Scarpa, Rafael
Detecção Acústica de Acometimento por Covid-19 / Rafael
Scarpa. -- São José do Rio Preto, 2023
46 p.

Dissertação (mestrado) - Universidade Estadual Paulista
(Unesp), Instituto de Biociências Letras e Ciências Exatas, São
José do Rio Preto
Orientador: Rodrigo Capobianco Guido

1. Processamento de sinais digitais. 2. Técnicas digitais de
processamento de sinais. 3. Inteligência artificial. 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.

RAFAEL MOTTA SCARPA

DETECÇÃO ACÚSTICA DE ACOMETIMENTO POR COVID-19

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. Rodrigo Capobianco Guido.
UNESP - Campus de São José do Rio Preto.

Prof^a. Dr^a. Luciene Cavalcanti Rodrigues.
IFSP - Campus de São José do Rio Preto.

Prof. Dr. Henrique Dezani.
FATEC - Campus de São José do Rio Preto

São José do Rio Preto

01 de setembro de 2023

AGRADECIMENTOS

Agradeço primeiramente a Deus pela saúde, sabedoria e entendimento.

Ao meu professor e orientador Dr. Rodrigo Capobianco Guido pelos ensinamentos, disponibilidade, pela compreensão, aconselhamentos assertivos e ajuda no desenvolvimento desta pesquisa realizando sugestões enriquecedoras.

Agradeço também aos meus professores da FATEC Rio Preto (Faculdade de Tecnologia de São José de Rio Preto). Em especial a professora Valéria Maria Volpe e o professor José Alexandre Ducatti por terem me incentivado a cursar o mestrado.

Aos meus amigos que apoiaram e ajudaram, especialmente ao Diego de Souza Oliveira e Vitor de Azevedo Carareto.

E por último e não menos importante, aos meus pais José Carlos Scarpa e Iranice Motta Cordovil pelo apoio e incentivo.

Muito obrigado a todos!

RESUMO

Inegavelmente, a recente pandemia do COVID-19 alterou o cotidiano de todo o mundo. Nesse sentido, inúmeras estratégias foram e ainda têm sido pesquisadas para aprimorar a detecção da enfermidade, englobando exames hematológicos, diagnóstico por imagens, entre outros. Uma possibilidade, também explorada, consiste no pré-diagnóstico por meio da ausculta respiratória. Assim, visando automatizar esse processo e oferecer um auxílio ao profissional de saúde, com base em um processo apenas modestamente investigado, caracteriza-se o objetivo deste trabalho: projetar e implementar uma técnica de processamento inteligente de sinais para detectar, de modo não-invasivo e com base na análise acústica dos ruídos respiratórios, o acometimento por COVID-19, principalmente em estágios iniciais. Valendo-se de novos conceitos disponíveis na literatura, tais como aqueles pertinentes à Engenharia Paraconsistente de Características, o método desenvolvido, de modesto custo computacional, foi testado valendo-se de uma parcela da base de sinais pública COVID-19 Sounds App. Métricas promissoras e compatíveis com as expectativas foram obtidas: classificadores baseados em distâncias Euclidianas e Support Vector Machines (SVM) alcançaram plena acurácia e significantes valores de especificidade, sensibilidade e F1 score.

Palavras-chave: Processamento de sinais. COVID-19. Sistemas inteligentes.

ABSTRACT

Undoubtedly, the recent COVID-19 pandemic has altered the daily life of the entire world. In this sense, numerous strategies have been and are still being discussed to enhance the detection of the disease, including hematological tests, imaging diagnosis, among others. One possibility, also explored, consists of pre-diagnosis through respiratory auscultation. Thus, aiming to automate this process and provide assistance to the healthcare professional, based on a moderately investigated process, the objective of this work is characterized: to design and implement an intelligent signal processing technique for a non-invasive detection of COVID-19 infection, mainly in early stages, based on the acoustic analysis of respiratory sounds. Leveraging new concepts available in the literature, such as those relevant to Paraconsistent Feature Engineering, the developed method, with modest computational costs, was tested using a portion of the COVID-19 Sounds App public signal database. Promising metrics consistent with expectations were obtained: classifiers based on Euclidean distances and Support Vector Machines (SVM) achieved full accuracy and significant values of precision, recall, and F1 score.

Keywords: Signal processing. COVID-19. Intelligent systems.

LISTA DE ILUSTRAÇÕES

Figura 1 - Trato vocal humano com suas determinadas partes nomeadas.....	11
Figura 2 - Ilustração de um período fundamental com a identificação do jitter e do shimmer. A frequência fundamental é o inverso do período fundamental.	12
Figura 3 - Espectro do sinal da fala da vogal /a/ de dois indivíduos: um contaminado pela patologia COVID e outro saudável.	12
Figura 4 - Espectro do sinal da tosse de dois indivíduos um contaminado pela patologia COVID e outro saudável.	13
Figura 5 - Espectrogramas dos sinais de voz de um adulto saudável: (a) sinal de voz de uma vogal, (b) tosse forte, e (c) respiração, isto é, inalação seguida de exalação.	14
Figura 6 - Espectrogramas dos sinais das respirações de dois indivíduos: um contaminado pela patologia COVID e outro saudável.....	14
Figura 7 - O Plano Paraconsistente onde as variáveis G1 e G2 encontram-se marcadas.....	17
Figura 8 - Estrutura da estratégia proposta.....	22
Figura 9 - Procedimento adotado na etapa de pré-processamento dos sinais de voz.	23
Figura 10 - código em linguagem C/C++ utilizado para implementar o pré-processamento dos dados, registrado na Figura 9.	23
Figura 11 - SVM utilizada como estratégia knowledge-based de classificação. sendo $(0 \leq i, j \leq 39)$	25
Figura 12 - Trecho de código em linguagem C/C++ utilizado para implementar a validação cruzada knowledge-based baseada na SVM.	28
Figura 13 - Trecho de código em linguagem C/C++ utilizado para implementar a validação cruzada pattern-matching baseada em distâncias Euclidianas.....	29
Figura 14 - Fluxograma para a estratégia de validação cruzada pattern-matching...	29
Figura 15 - Fluxograma para a estratégia de validação cruzada knowledge-based..	30
Figura 16 - P, indicado como um ponto vermelho no plano paraconsistente, para os 40 VCs de cada um dos nove grupos de sinais.	31

Figura 17 - Box-plots resultantes da etapa de classificação baseada no classificador pattern-matching por distâncias Euclidianas considerando a métrica acurácia para X = 2, 4, 6, ..., 38.	33
Figura 18 - Box-plots resultantes da etapa de classificação baseada no classificador pattern-matching por distâncias Euclidianas considerando a métrica especificidade para X = 2, 4, 6, ..., 38.	34
Figura 19 - Box-plots resultantes da etapa de classificação baseada no classificador pattern-matching por distâncias Euclidianas considerando a métrica sensibilidade para X = 2, 4, 6, ..., 38.	34
Figura 20 - Box-plots resultantes da etapa de classificação baseada no classificador pattern-matching por distâncias Euclidianas considerando a métrica F1 score para X = 2, 4, 6, ..., 38.	35
Figura 21 - Box-plots resultantes da etapa de classificação baseada no classificador knowledge-based por SVM considerando a métrica acurácia para X = 2, 4, 6, ..., 38	35
Figura 22 - Box-plots resultantes da etapa de classificação baseada no classificador knowledge-based por SVM considerando a métrica especificidade para X = 2, 4, 6, ..., 38.	36
Figura 23 - Box-plots resultantes da etapa de classificação baseada no classificador knowledge-based por SVM considerando a métrica sensibilidade para X = 2, 4, 6, ..., 38.	36
Figura 24 - Box-plots resultantes da etapa de classificação baseada no classificador knowledge-based por SVM considerando a métrica F1 score para X = 2, 4, 6, ..., 38.	37

SUMÁRIO

1	Introdução.....	9
2	Fundamentação Teórica.....	10
2.1	Sinais Digitais.....	10
2.2	Vozes Afetadas Patologicamente.....	10
2.3	Tosse Patológica.....	12
2.4	Respiração Patológica.....	13
2.5	Análise e Classificação de Sinais Patológicos.....	14
2.6	Energia de um Sinal e Operador de Teager.....	15
2.7	Zero crossing rate.....	16
2.8	Engenharia Paraconsistente de Características.....	16
2.9	Base de dados.....	17
2.10	Estado-da-arte: revisão de artigos científicos relacionados ao projeto.....	18
3	A Abordagem Proposta.....	22
3.1	A Estrutura da Abordagem Proposta.....	22
4	Testes e Resultados.....	31
4.1	Experimentos com os VCs.....	31
4.2	Experimentos com os classificadores.....	32
4.2.1	Classificador <i>pattern-matching</i> com distâncias Euclidianas.....	32
4.2.2	Classificador <i>knowledge-based</i> baseado na SVM.....	32
5	Conclusões.....	39
	REFERÊNCIAS.....	40

1 INTRODUÇÃO

A partir de 2019, o mundo mudou por completo em função do início de uma epidemia que matou milhões de seres humanos. A doença proveniente dessa tragédia foi a Síndrome Respiratória Aguda Grave (SARS - Severe Acute Respiratory Syndrome), mais conhecida por COVID-19, a qual tornou-se uma pandemia de proporções assustadoras. A SARS tem como sintomas principais a febre, a dor de cabeça, o cansaço, a tosse seca, a dor de garganta, a dificuldade para respirar, a perda da fala e a dor no peito, entre outros.

Dentre os sintomas, dois oferecem possibilidade de análise acústica: tosse seca e dor de garganta, pois elas acabam gerando rouquidão ou alterações similares na fala. Por meio dessas análises, é possível utilizar as tecnologias recentes de processamento de sinais e sistemas inteligentes para identificar padrões no sinal de voz dos indivíduos, o que serve como um teste para detecção da referida condição orgânica.

Desse modo, em função do destacado interesse atual pelo tema, assim como da área de pesquisa que é alvo de interesse deste autor e do seu orientador, esta monografia foi dedicada ao tema em questão. Os conceitos aqui apresentados servirão de base para a continuidade do trabalho, o qual resultará na dissertação de mestrado do autor.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordados conceitos teóricos fundamentais para a detecção e classificação de enfermidades por meio de sinais acústicos. Em seguida, serão abordadas as principais técnicas utilizadas nos respectivos sistemas e a base de dados que será utilizada para as pesquisas. Adicionalmente, constam também discussões envolvendo trabalhos similares documentados na literatura.

2.1 Sinais Digitais

Para viabilizar a análise computacional, os sinais de voz, que são originalmente analógicos, devem estar disponíveis em formato digital. Assim, destacam-se, nesse âmbito:

- o Teorema de Nyquist, segundo o qual a taxa de amostragem necessária para reconstruir um sinal analógico, sem perdas no meio digital, é de duas vezes a maior frequência nele presente [1].
- a quantização, ou seja, precisão de cada amostra, a qual será de 16 bits, que para o caso de sinais de voz tem se mostrado suficiente [1].

Particularmente, o formato de arquivo computacional que melhor permite manipular os sinais digitais de voz é o WAV, o qual será usado no estudo. De modo geral, trata-se de um cabeçalho com informações essenciais sobre o sinal seguido das amostras digitalizadas, isto é, raw data, conforme detalhado em [1].

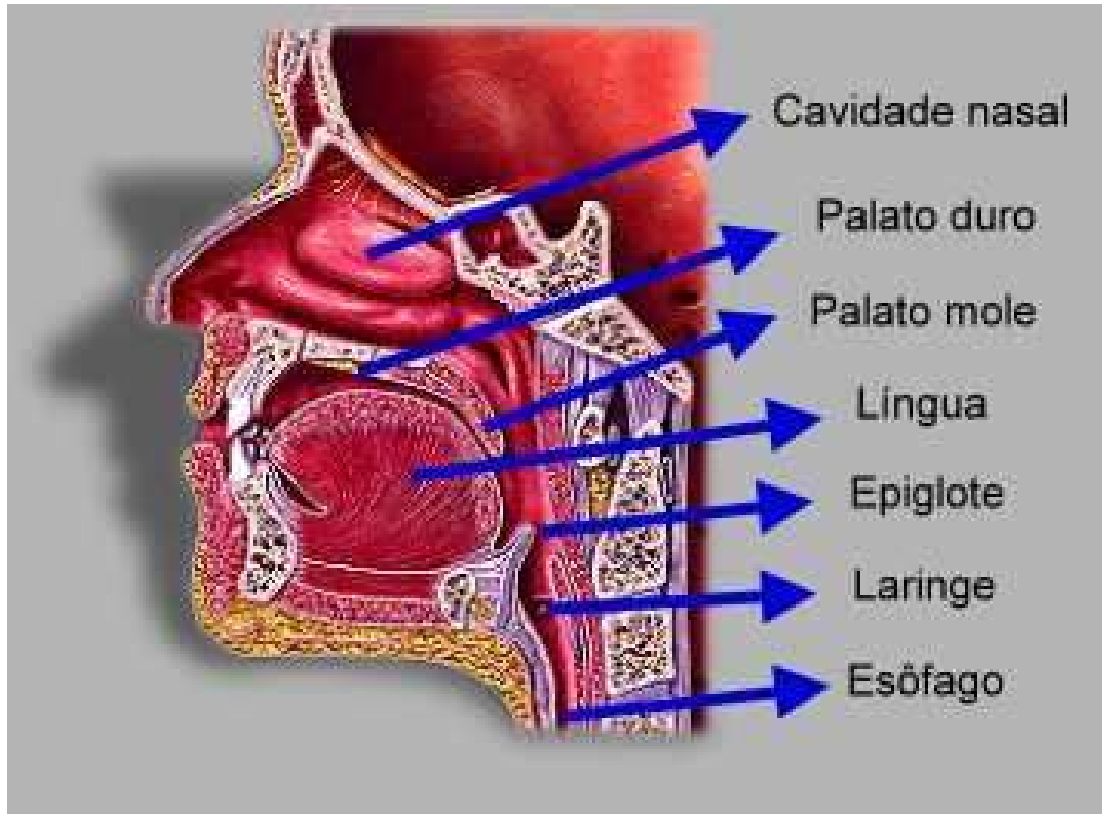
2.2 Vozes Afetadas Patologicamente

As patologias mais comumente encontradas na voz estão relacionadas diretamente com as várias partes do trato vocal dos indivíduos, isto é, cavidade nasal, palato duro, palato mole, língua, epiglote, laringe e esôfago, conforme a Figura 1. Essas patologias podem ser diferenciadas por meio da relação de padrões perceptíveis como rouquidão, aspereza e soproidade [2]. A rouquidão consiste na combinação da aspereza com a soproidade.

A aspereza tem como característica principal um aumento elevado do estiramento das pregas vocais, o que é ocasionado pelos músculos que as circundam. Diferentemente, a soproidade ocorre quando há um escape de ar na região da glotal, gerado por alguma alteração orgânica nos músculos ou pelo mau funcionamento deles. A produção do sinal de voz está diretamente relacionada ao impacto das

funções respiratórias, o que significa que a análise de voz pode ser utilizada como um indicador de doenças respiratórias, tais como a COVID-19.

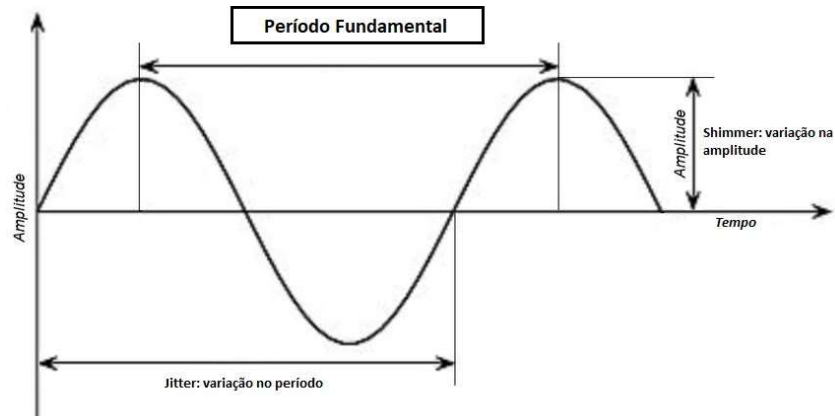
Figura 1 - Trato vocal humano com suas determinadas partes nomeadas.



Fonte: [3]

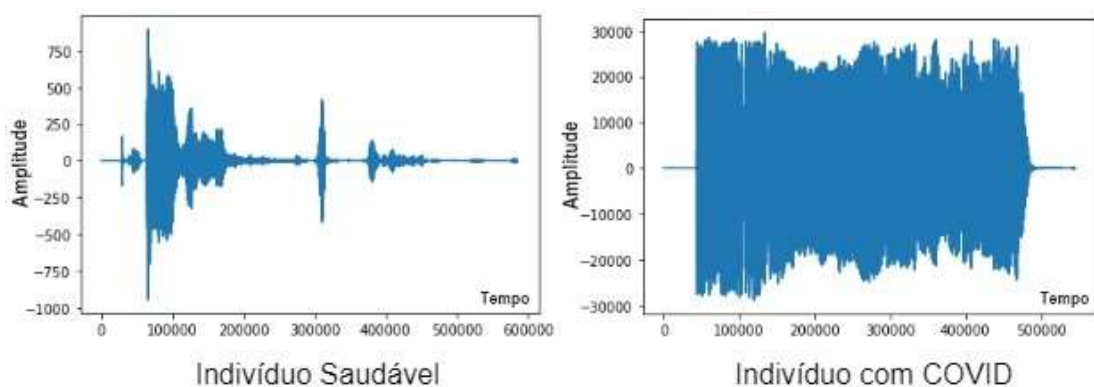
As Figuras 2 e 3, medidas de perturbação de ciclo-a-ciclo, que são também conhecidas como medidas de perturbação da voz, são particularmente úteis para o pré-diagnóstico de patologias nos órgãos fonadores [4]. Para identificar essas alterações, são usados parâmetros tais como a frequência fundamental, ou jitter, que constitui um fenômeno de perturbação no período de vibração das pregas vocais. Alterações nas pregas vocais provenientes de patologias produzem vibrações errôneas e são detectadas na alteração do ciclo da amplitude, ou frequência fundamental. Outro marcador interessante é a variação da amplitude, ou shimmer. Alterações desse parâmetro são causadas por lesões de massa nas pregas vocais, tais como nódulos, pólipos, edemas ou carcinomas. No caso de alterações no jitter, as enfermidades podem ter origem neurológica ou até mesmo psicológica.

Figura 2 - Ilustração de um período fundamental com a identificação do jitter e do shimmer. A frequência fundamental é o inverso do período fundamental.



Fonte: De autoria própria.

Figura 3 - Espectro do sinal da fala da vogal /a/ de dois indivíduos: um contaminado pela patologia COVID e outro saudável.

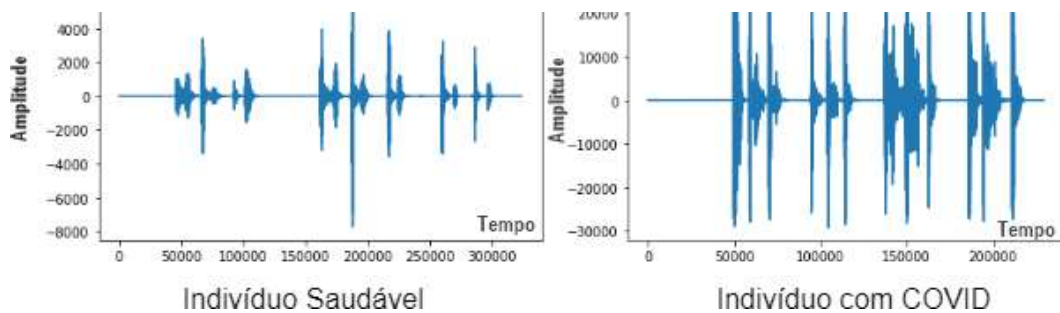


Fonte: De autoria própria.

2.3 Tosse Patológica

A tosse é uma reação do corpo humano para o livramento de substâncias danosas, sejam elas bactérias, poeiras, vírus, fungos ou outras substâncias. Essa reação pode gerar, aos poucos, irritabilidade da laringe, faringe, traqueia e pulmões, sendo uma contração espasmódica e repentina da cavidade torácica que, ao expulsar o ar dos pulmões, gera um som característico.

Figura 4 - Espectro do sinal da tosse de dois indivíduos um contaminado pela patologia COVID e outro saudável.



Fonte: De autoria própria.

Conforme a Figura 4, características acústicas da tosse dependem da velocidade do fluxo de ar, bem como das dimensões do trato vocal e das vias aéreas. Isso torna possível detectar tosses em sinais acústicos [5]. Como a estrutura física do sistema respiratório fica alterada com infecções respiratórias, é possível até classificar o estado patológico com base na tosse.

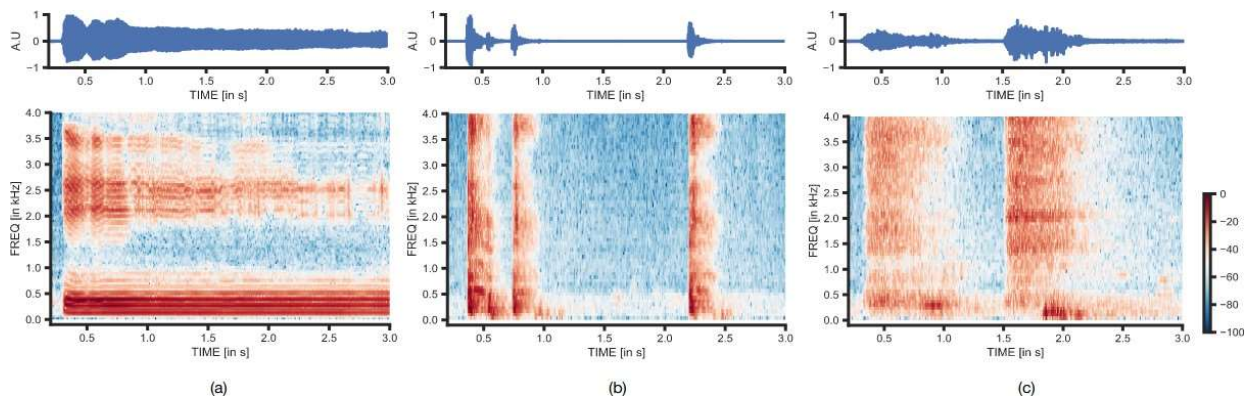
2.4 Respiração Patológica

A respiração é um processo pelo qual o corpo humano faz a inspiração e expiração dos gases, ou seja, a troca gasosa dos elementos presentes nos pulmões. Essa ação produz um ruído específico, sendo ele o som da ventilação pulmonar. A ventilação pulmonar envolve os órgãos do sistema respiratório, os quais são normalmente prejudicados pela patologia que se pretende estudar. De acordo com tal ruído apresenta padrões distintos para a comparação e análise da COVID-19 em contrates com os indivíduos não contaminados [5]. Na Figura 5 constam três espectrogramas, onde se pode observar o espectro do sinal de voz de um adulto não contaminado pela SARS, a sua tosse e o ruído da sua respiração. Complementarmente, na Figura 2.6 constam os espectrogramas dos sinais das respirações de dois indivíduos: um contaminado pela patologia COVID e outro saudável.

2.5 Análise e Classificação de Sinais Patológicos

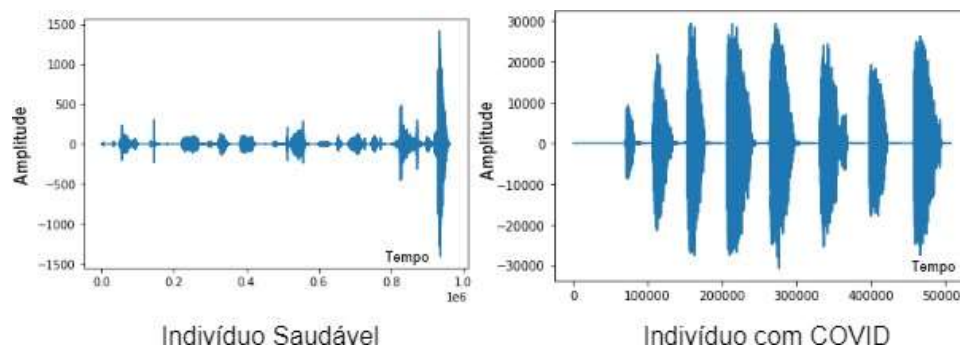
Atualmente, inúmeras técnicas vêm sendo usadas para a investigação e a classificação de sinais de voz. Dentre elas, destacam-se as wavelets e as redes neurais artificiais [6] [7]. Particularmente, de acordo com as redes neurais têm maior abrangência devido ao fato de que muitos sistemas classificadores alcançam resultados muito próximos a 100% de acurácia quando parâmetros diversos são fornecidos como entrada para tais algoritmos: frequências formantes, períodos de vibração das pregas vocais, isto é, pitch e suas variações, estabilidade do período de pitch durante a fonação, grau de similaridade do formato do pitch, espectro médio de longo tempo (LTAS), razão ruído-harmônicas (NHR) e razão harmônicas-ruído (HNR) [4]. No trabalho de mestrado a ser desenvolvido, pretende-se usar tais parâmetros e ferramentas para identificação da COVID-19.

Figura 5 - Espectrogramas dos sinais de voz de um adulto saudável: (a) sinal de voz de uma vogal, (b) tosse forte, e (c) respiração, isto é, inalação seguida de exalação.



Fonte: [5]

Figura 6 - Espectrogramas dos sinais das respirações de dois indivíduos: um contaminado pela patologia COVID e outro saudável.



Fonte: De autoria própria.

2.6 Energia de um Sinal e Operador de Teager

O conceito de energia de um sinal digital $x[n]$ de tamanho M , isto é, $E(x[n]) = \sum_{i=0}^{M-1} (x_i)^2$ pode ser aplicado em vários ramos da engenharia e da computação, inclusive no processamento digital de sinais de voz [8][9]. De modo mais aprimorado, os componentes de amplitude (AM) e frequência (FM) podem ser utilizados para a estimativa da energia. Dessa forma, a análise de sinais de vozes acometidas por patologias pode ser baseada em uma decomposição do tipo AM-FM. No artigo, demonstrou-se que o Teager Energy Operator (T) é uma ferramenta que permite a análise conjunta dos operadores AM e FM [9]. Dessa forma, T modela aproximadamente a energia de um sinal sob análise, $x[n]$, como um produto quadrado de sua amplitude (A) por sua frequência instantânea (Ω) com base na relação $T(x[n]) = x^2 - x_{i-1} \cdot x_{i+1}, \forall i \in \mathbb{Z}$.

Conforme demonstrado no artigo, T foi modificado com base em um novo conceito: signal mass (η), de tal forma que $A^2 \cdot \Omega^2$ pode ser calculado exatamente, e não somente aproximadamente, como uma função de $x[n]$, fornecendo informações precisas por meio do recém-definido Enhanced Teager Energy Operator ($\mathcal{T}$), o que é de interesse para a área deste estudo [9]. Especificamente, observou-se que

$$\mathcal{T} = \frac{T}{\frac{1}{2}\eta} = \frac{x_n^2 - x_{n-1} \cdot x_{n+1}}{\text{sinc}^2(\Omega)} = A^2 \cdot \Omega^2, \text{ onde}$$

$$\text{sinc}^2(\Omega) = \begin{cases} 1 & \text{para } x_n = 0 \\ \text{sinc}^2(\arccos(z)) & \text{para } x_n \neq 0 \text{ e } (-1 \leq z \leq 1) \\ \left(\frac{z^2 + z\sqrt{z^2-1} - 1}{|z + \sqrt{z^2-1}| \cdot \ln(|z + \sqrt{z^2-1}|)} \right)^2 & \text{caso contrário, isto é, para } |z| > 1 \end{cases}.$$

$$\text{sendo } z = \frac{x_{n-1} + x_{n+1}}{2x_n}.$$

2.7 Zero crossing rate

Zero crossing rate (ZCR) consiste, essencialmente, o número de vezes que o sinal cruza o eixo horizontal em uma dada janela de tempo, isto é, quando amostras sucessivas possuem sinais algébricos distintos [10]. Apesar de ser obtida por meio da análise temporal do sinal, o ZCR pode ser interpretado como uma medida simples da frequência fundamental do sinal. O ZCR de um sinal discreto $s[\cdot]$ com M amostras é definido como:

$$ZCR = \frac{1}{2M} \sum_{i=0}^{M-2} |sgn[s(i+1)] - sgn[s(i)]| \quad , \quad (2.1)$$

sendo sgn a seguinte função:

$$sgn[s(i)] = \begin{cases} 1, & s(i) \geq 0 \\ -1, & s(i) < 0 \end{cases} \quad . \quad (2.2)$$

2.8 Engenharia Paraconsistente de Características

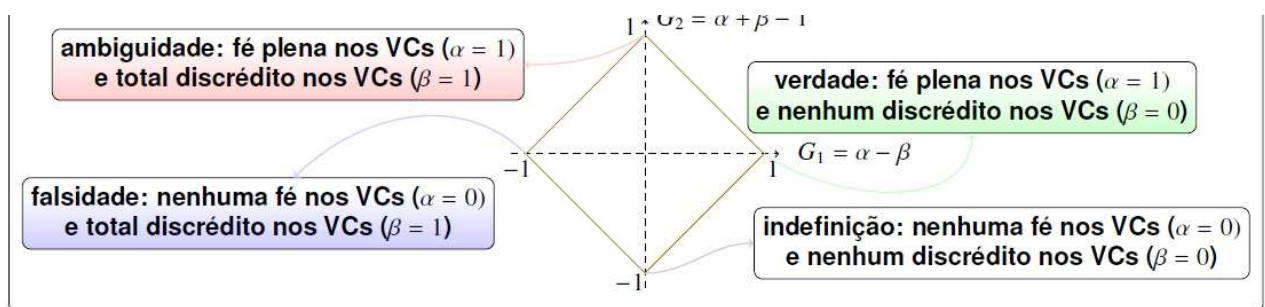
No trabalho de mestrado a ser desenvolvido, pretende-se utilizar características extraídas de forma artesanal, ao invés de aprendidas automaticamente, para facilitar as respectivas interpretações físicas. Dessa forma, é interessante utilizar uma estratégia para seleção das melhores características a serem entregues para o algoritmo classificador. Uma possibilidade para tal, que tem sido utilizada com sucesso em diversas subáreas, é a Engenharia Paraconsistente de Características (EPC) [11]. Particularmente, essa técnica faz uso dos conceitos de lógica paraconsistente para considerar dois critérios independentes para mensurar a adequação das características:

- α , que representa o valor de similaridade intraclasses;
- β , que representa a dissimilaridade interclasses.

Conforme as variáveis α e β são consideradas para se calcular o grau de certeza, $G1 = \alpha \beta$, e o grau de contradição, $G2 = \alpha + \beta - 1$, de um conjunto de vetores de características (VCs) representantes de determinadas classes, onde ($1 \leq G1$, $G2 \leq 1$) [11]. Particularmente, $G1 = 1$ e $G2 = 1$ são dois extremos que representam falsidade absoluta e verdade plena, respectivamente [12]. Do mesmo modo, $G2 = 1$

e $G_2 = 1$ são dois extremos que representam indefinição total e ambiguidade completa, respectivamente, como pode se observar no Plano Paraconsistente apresentado na Figura 7.

Figura 7 - O Plano Paraconsistente onde as variáveis G_1 e G_2 encontram-se marcadas.



Fonte: De autoria própria.

O resultado da análise é um ponto $P = (G_1, G_2)$ que é tanto melhor quanto mais próximo estiver do ponto ideal, isto é, $(G_1, G_2) = (1, 0)$, conforme detalhado em [11]. Assim, procurar-se-ão características que atraiam $(P = G_1, G_2)$ para o vértice (1, 0). No trabalho de mestrado, somente duas classes são consideradas: vozes de indivíduos saudáveis e vozes de indivíduos acometidos pela COVID-19.

2.9 Base de dados

Dentre os bancos de dados disponíveis na Internet para permitir a exequibilidade do presente estudo, foi selecionado o do aplicativo COVID-19 Sounds App (<https://www.covid-19-sounds.org/>). Este aplicativo tem como objetivo mapear informações de pacientes saudáveis e com a COVID-19, a fim de reunir o maior número de sinais patológicos da doença. O banco de dados tem como base alguns critérios que são: idade, estado de saúde, gênero, localidade, existência de doença de asma, tosse, hábitos tabagistas, teste de COVID positivo ou negativo, diabetes, doença isquêmica, dificuldade respiratória, existência e dor de garganta, febre, fadiga, dores musculares, perda de olfato, doenças pulmonares crônicas e pneumonia.

O banco de dados possui 23.8 GB de dados, sendo dividido em pastas com as vozes coletadas, incluindo sons respiratórios rápidos e lentos, sons de tosse profundos e rasos, fonação de vogais sustentadas /a/, /i/ e /o/ e contagem de números em ritmos lento e rápido, totalizando nove sinais de cada um dos 189 indivíduos matriculados. Desse modo, estão disponíveis $9 \cdot 189 = 1700$ sinais acústicos amostrados a uma taxa de 48 kHz, possibilitando, portanto, a retenção de frequência até 24KHz. Dentre os 189 indivíduos, 68 foram laudados por profissionais da saúde como sendo saudáveis e 121 como estando afetados patologicamente. Originalmente, todos os sinais acústicos da base foram gravados no formato mp3.

2.10 Estado-da-arte: revisão de artigos científicos relacionados ao projeto

O artigo contém a descrição de uma abordagem para detectar e classificar diferentes tipos de patologias de voz por meio da investigação de diferentes bandas de frequências, usando funções de correlação. Adicionalmente, a escala de Mel foi utilizada em conjunto com um banco de filtros [13]. O experimento desenvolveu-se com base em três bancos de dados de sinais gravados em diferentes idiomas, sendo eles Inglês, Alemão e Árabe e consistem na vogal /a/ sustentada de indivíduos com vozes normais e patológicas. A classificação foi realizada com base em uma máquina de vetores de suporte (SVM). Examinando-se frequências no intervalo de 1000 a 8000 Hz, as maiores acurácias encontradas foram de 99,255%, 98,941% e 95,188%, para cada idioma mencionado, respectivamente. As ideias são usadas no artigo, em que há uma proposta de detecção da COVID-19 por meio de dispositivos móveis com base em técnicas de inteligência artificial e de audição computacional [14]. Biomarcadores implementado com base na ferramenta COMPARE de inteligência artificial para diagnosticar pessoas assintomáticas e sintomáticas de forma rápida e móvel.

Na publicação o principal objetivo é a investigação de parâmetros para viabilizar o aplicativo Programa Multidimensional de Voz (MDVP), onde a detecção e classificação automática das patologias da voz é realizada com base em múltiplos bancos de dados, utilizando amostras de vogais de pessoas contaminadas e não contaminadas. Vogais sustentadas / a / de locutores saudáveis e com patologias na voz foram utilizadas para extração de 33 características quantitativas das vozes,

incluindo frequência fundamental, amplitude, energia espectral, shimmer, jitter, a presença de qualquer lacuna de sonoridade e outros recursos. O discriminante de Fisher (FDR) foi usado para selecionar as características e separar as classes, com acurácia de até 99,68% [15].

A proposta contida no artigo foi utilizar a técnica padrão derivado entrelaçado (IDP), que envolve uma derivada direcional de ordem n em uma descrição espectro-temporal dos sinais de excitação da fonte glótica [16]. Foi utilizada uma SVM com função de base radial (RBF) como kernel na etapa classificadora. O método proposto foi comparado com 12 características Mel Frequency Cepstral Coefficients (MFCC) mais seus coeficientes de velocidade e aceleração. Foram utilizados três bancos de dados e os resultados foram acurácias de 98,2%, 95,3% e 94,2%.

No trabalho, foi realizada a análise da caracterização quantitativa das alterações do COVID-19 na voz. Para isso, foi usado um sistema dinâmico para modelar a oscilação das pregas vocais. Somente pacientes sintomáticos susceptíveis foram detectados e somente anomalias na respiração foram observadas. Mapeamentos no espaço de fases foram realizados, observando-se que, para indivíduos COVID-19 negativos eles são ciclos limites ou toroides finos, indicando um maior grau de sincronização nos modos próprios de vibração e maior simetria de movimento [17]. Para pacientes COVID-19 positivos, as trajetórias são mais complexas, indicando um maior grau de assincronia e, além disso, a assimetria e amplitude de movimento são reduzidas. Uma análise do desvio-padrão foi utilizada como ferramenta na etapa final de classificação, sendo os resultados registrados na forma de curvas Receiver Operating Characteristic (ROC). Concluiu-se que a doença interrompe o arrastamento das pregas vocais, causando assimetrias nos sinais gerados. O estudo atingiu 91% no valor da área sobre a curva ROC.

O artigo contém, como ideia básica, uma análise dinâmica da glote durante a produção de voz, visando a identificação de recursos vocais significativos para a detecção do COVID-19 [18]. Para isso, foram analisadas diversas vozes, de pessoas sintomáticas e assintomáticas, como também pessoas saudáveis. Um método baseado em um modelo de localização e discriminação de anomalias no sinal de voz foi proposto, com base em um filtro glottal flow waveform (GFW) em conjunto com um

algoritmo para detectar deslocamento e velocidade de cada prega vocal. Dessa forma, mapeando esses deslocamentos de modo assertivo usando como um classificador do tipo convolutional neural network (CNN), três experimentos de validação cruzada foram realizados para detecção da COVID-19. Uma acurácia global de 90% foi obtida.

No artigo, os autores buscaram identificar a COVID-19 por meio da detecção de padrões na voz de pessoas infectadas. Para isso, usou-se uma base de dados comunitária, a qual é incrementada por meio de um aplicativo. Tosse, respiração e fala são os sinais utilizados para a detecção. O locutor voluntário, além de gravar a sua voz, responde a perguntas chaves para a pesquisa: gênero, idade, doenças respiratórias prévias, tabagismo, entre outras. Notou-se que as amostras de tosse são menos assertivas em comparação com as amostras de respiração. Uma rede perceptron multicamadas (MLP) alimentada com características oriundas da escala Mel foi utilizada, obtendo-se uma acurácia global de 66,74%. Os autores creem que, com o passar do tempo e treinamento, os resultados possam ser aprimorados [5].

No artigo, procurou-se implementar um algoritmo de aprendizagem profunda para detecção da COVID-19, baseado em técnicas de reconhecimento de fala. Realizou-se uma análise de tosses, respirações e fala com base em uma rede neural recorrente (RNN) como classificador, implementada em linguagem de programação Python sob ambiente TensorFlow. A maior acurácia, de 97% para verdadeiros-positivos e 88,2% para verdadeiros-negativos, foi obtida com base na análise das tosses e características obtidas a partir da escala Mel [19].

No trabalho, utilizou-se uma estratégia para reconhecimento de fala para identificação do COVID-19 [20]. A pesquisa foi realizada com base em aplicativos conhecidos: o OpenSmile para tratar a Paralinguística Computacional, o Praat para tratamento dos sinais de extração de características e uma rede neural convolucional (CNN) pré-implementada, a qual foi treinada por transferência de aprendizagem. Uma acurácia global média de 83% foi obtida.

No artigo, os autores analisaram a contaminação por COVID-19 com base no ciclo respiratório e na fala, além de informações sobre o estado físico, fisiológico, psicológico e emocional do locutor [21]. Aplicando redes neurais perceptron e uma

SVM com núcleos RBF, comparam-se padrões de entrada com outros previamente armazenados. Escala Mel, jitter e shimmer foram utilizados, refletindo uma acurácia global de 50% na detecção da doença.

Finalmente, no artigo realizou-se um estudo para detecção de pessoas infectadas pelo COVID-19 com base em tosses forçadas gravadas e enviadas via telefone celular [22]. Uma CNN foi utilizada para classificação e levou em conta não somente as vozes, mas também informações tais como a degradação muscular, mudanças no comportamento, humor e mudanças no trato respiratório. O modelo proporcionou uma acurácia de 97% com sensibilidade de 100%.

Diante dos trabalhos revisados, nota-se que já existe, embora modestamente, pesquisas se consolidando para a detecção de patologias diversas e da COVID-19 por meio da análise de sinais de voz. Considerável parte das técnicas utilizam parâmetros comuns na área de processamento de voz, tais como aqueles oriundos da escala Mel (MFCCs), embora outras se utilizem de características aprendidas automaticamente. Em nenhum dos trabalhos revisados foi identificada maior preocupação na seleção das características, motivando, assim, o emprego da EPC. Complementarmente, estratégias mais recentes, tais como o Operador de Energia de Teager Aprimorado, também estão sendo considerados neste trabalho, conforme descreve-se doravante [9].

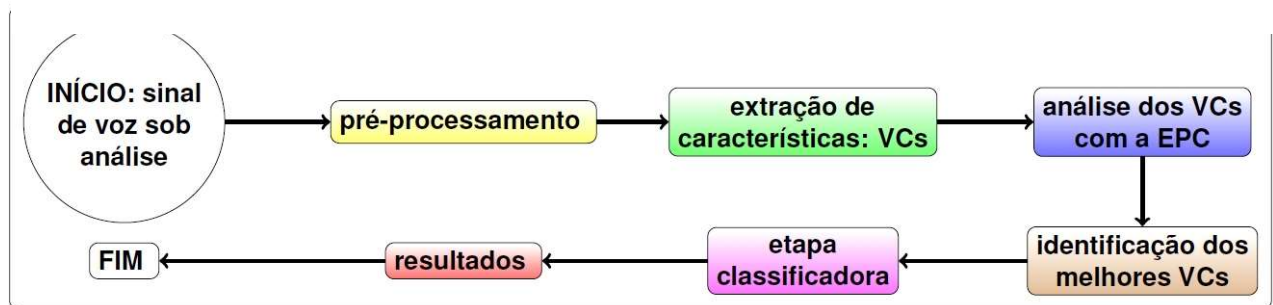
3 A ABORDAGEM PROPOSTA

Apresenta-se, neste Capítulo, a estratégia geral proposta neste trabalho, com os devidos detalhes que permitem reproduzi-la.

3.1 A Estrutura da Abordagem Proposta

A abordagem proposta para o desenvolvimento deste trabalho, ilustrada na Figura 8, deu-se por meio de um procedimento sequencial, dividido em etapas, especificamente da forma como segue.

Figura 8 - Estrutura da estratégia proposta



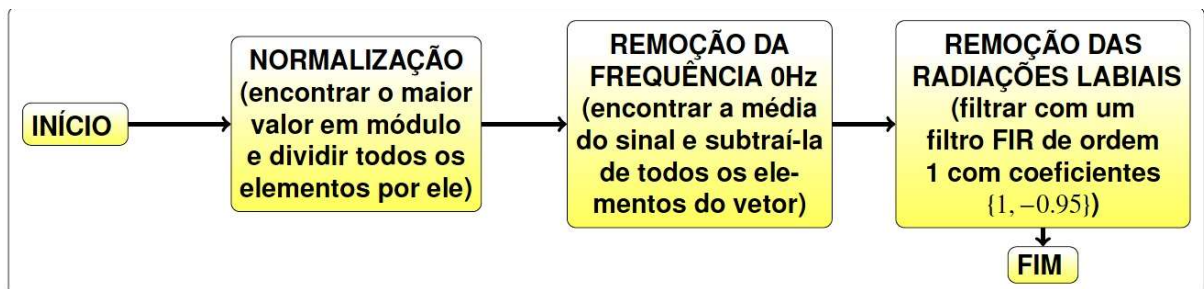
Fonte: De autoria própria.

• **ETAPA 1 |pré-processamento|**: a partir da base de dados, contendo nove sinais de voz para cada um dos 189 indivíduos matriculados, sendo 68 saudáveis e 121 afetados patologicamente, totalizando 1700 sinais, foram selecionados sinais de 40 indivíduos, sendo 20 de cada uma das duas classes. A escolha dos 40 indivíduos dentre os 189 deu-se com base na qualidade dos sinais acústicos disponíveis: inúmeras gravações continham ruídos e sinais quase inaudíveis. Assim, acusticamente e com o auxílio de editores de áudio que possibilitaram visualizar as formas de onda dos sinais em questão, foram retidos somente aqueles que se mostravam “limpos” e bem definidos. Após a escolha, os $9 \cdot 40 = 360$ sinais foram convertidos para o formato wave, com quantização de 16 bits, para posterior processamento.

Os dados brutos dos 360 sinais selecionados foram extraídos, valendo-se de um trecho de código escrito em linguagem C/C++ disponibilizado pelo orientador deste

trabalho. Particularmente, os sinais foram normalizados em amplitude, forçando-os a repousarem sobre o intervalo de 1 a 1, e as respectivas médias foram removidas objetivando excluir a frequência 0 Hz [4]. Posteriormente, as altas frequências são destacadas por meio do procedimento de pré-ênfase baseado na convolução discreta com o filtro digital passa-altas, de resposta ao impulso finita com ordem 1 e com coeficientes 1, 0.95 [23][1]. O respectivo fluxograma para este procedimento encontra-se na Figura 9 e trecho de código-fonte em linguagem C/C++ contendo a implementação correspondente encontra-se na Figura 10.

Figura 9 - Procedimento adotado na etapa de pré-processamento dos sinais de voz.



Fonte: De autoria própria.

Figura 10 - código em linguagem C/C++ utilizado para implementar o pré-processamento dos dados, registrado na Figura 9.

```

double maior_em_modulo=st[0];
for(int i=1;i<M;i++)
    if(sqrt(st[i]*st[i])>maior_em_modulo)
        maior_em_modulo=sqrt(st[i]*st[i]);
for(int i=0;i<M;i++)
    st[i]/=maior_em_modulo;
//calculando a média e removendo-a para excluir a frequência 0Hz
double media=0;
for(int i=0;i<M;i++)
    media+=st[i];
media/=((double)(M));
for(int i=0;i<M;i++)
    st[i]-=media;
//removendo as radiações labiais
double* s=new double[M];
for(int k=1;k<M;k++)
    s[k-1]=st[k]-0.95*st[k-1];
M--;
  
```

Fonte: De autoria própria.

• **ETAPA 2 |obtenção dos VCs|**: cada um dos $9 \cdot 40 = 360$ sinais pré-enfatizados na

Etapa 1 foi convertido em um VC de tamanho quatro, como segue:

- **Característica #0**: energia total do sinal sob análise, normalizada em função do tamanho do sinal;
- **Característica #1**: desvio-padrão dos respectivos quadrados das amplitudes do sinal sob análise;
- **Característica #2**: ZCR do sinal sob análise;
- **Característica #3**: raiz quadrada da energia do Operador de Teager Aprimorado (T) calculado a partir do sinal sob análise.

As quatro características mencionadas foram escolhidas, como fruto de resultados anteriores na área de processamento de sinais de voz, por possibilitarem informações nos domínios temporal e espectral dos sinais sob análise [4][8][9][10].

• **ETAPA 3 |pré-análise dos VCs|**: com base na EPC, os 40 VCs de cada um dos nove subgrupos foram submetidos, separadamente, ao exame paraconsistente visando verificar a posição do ponto P em relação ao vértice $(1,0)$ do plano paraconsistente, conforme explicado na Seção 2.8 do Capítulo anterior. Diante do resultado desta análise, conforme detalhado no próximo Capítulo, optou-se por testar dois classificadores modestos, que certamente atendem às necessidades do experimento, conforme segue.

• **ETAPA 4 |classificação|**: considerando-se separadamente cada um dos nove grupos de sinais e utilizando X VCs para servir de templates, ou casos de treinamento, e o restante, isto é, $(40 - X)$, para avaliar a técnica de classificação utilizada, foram testados os seguintes classificadores, considerando que o valor de $X = \{2, 4, 6, \dots, 38\}$ será discutido no Capítulo seguinte:

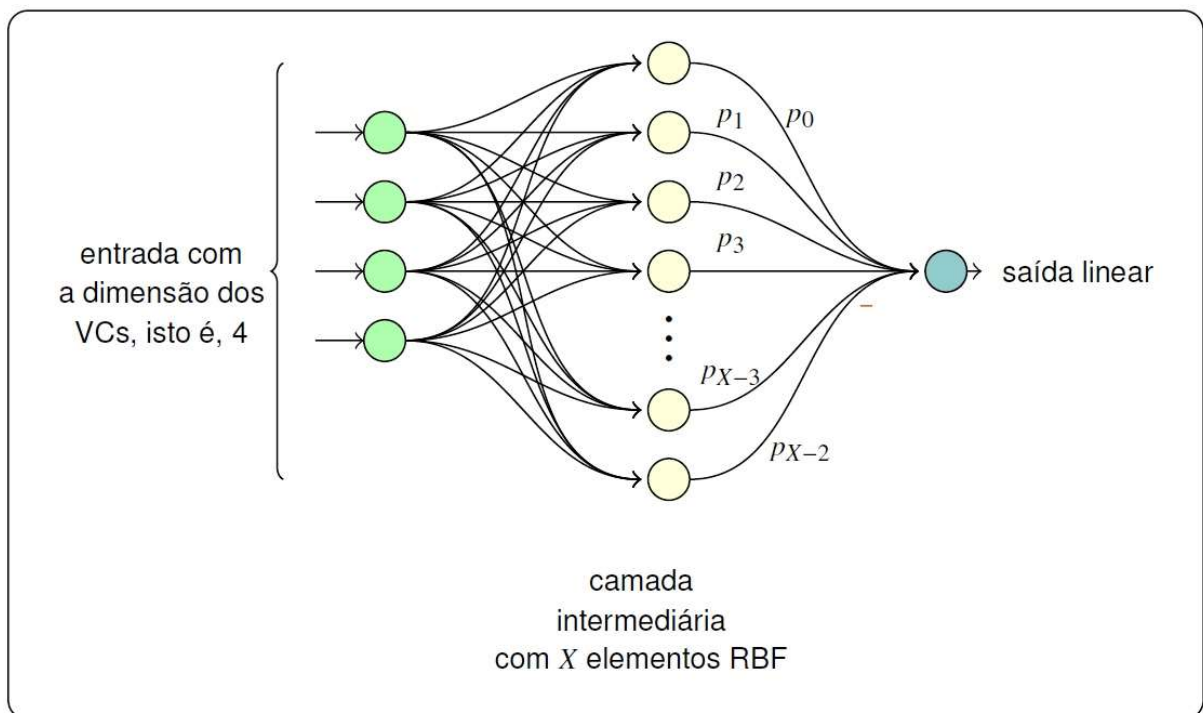
- **Classificador C_A** : critério *pattern-matching* de distâncias Euclidianas, o qual implica que a classe atribuída ao VC em teste é a do VC-modelo para o qual ele dista minimamente, considerando-se todos os VCs-modelos;
- **Classificador C_B** : critério *knowledge-based* do tipo *Support Vector Machine* (SVM). Neste caso, optou-se pela estrutura apresentada na Figura 11, de forma que

o treinamento da SVM foi realizado com base em uma estratégia semi-supervisionada: inicialmente, de modo plenamente não-supervisionado, os núcleos de função de base radial (RBF) dos elementos intermediários foram ajustados de forma que o i -ésimo elemento apresente saída máxima, isto é, 1 para o i -ésimo caso de treinamento.

Especificamente, o i -ésimo elemento intermediário da SVM possui a seguinte função de transferência:

$1 + e^{-\text{distância euclidiana } \{i\text{-ésimo elemento de treinamento; } j\text{-ésimo elemento de treinamento}\}}$

Figura 11 - SVM utilizada como estratégia *knowledge-based* de classificação. sendo $(0 \leq i, j \leq 39)$.



Fonte: De autoria própria.

Posteriormente, de modo totalmente supervisionado, os valores 0 e 1 foram utilizados para rotular vozes de indivíduos saudáveis e daqueles afetados patologicamente, respectivamente, de modo a estabelecer um sistema linear de dimensão $(X) \times (X)$ para o qual a solução implica os pesos p_i .

Notavelmente, cada um dos casos de treinamento refere-se a uma equação do sistema. Assim, a SVM encontra-se pronta para o uso: cada caso de teste é

apresentado na camada de entrada e atravessa a camada intermediária produzindo X 1 valores que são combinados linearmente com os pesos p_i e produzem um valor de saída que implica o resultado da classificação. Se esse valor estiver mais próximo de 0 ou 1, rotula-se o elemento de teste como sendo a voz de um indivíduo saudável ou afetado patologicamente, respectivamente.

• **ETAPA 5 [cômputo dos resultados]:** os resultados são computados considerando-se um processo de validação cruzada no qual todas as possibilidades foram testadas, ou seja:

– $X = 2$ VCs selecionados aleatoriamente para servir de modelos ou elementos de treinamento, sendo um de cada classe, e $40 - X = 40 - 2 = 38$ VCs restantes, sendo 19 de cada classe, para servirem como elementos de teste;

– $X = 4$ VCs selecionados aleatoriamente para servir de modelos ou elementos de treinamento, sendo dois de cada classe, e $40 - X = 40 - 4 = 36$ VCs restantes, sendo 18 de cada classe, para servirem como elementos de teste;

– $X = 6$ VCs selecionados aleatoriamente para servir de modelos ou elementos de treinamento, sendo três de cada classe, e $40 - X = 40 - 6 = 34$ VCs restantes, sendo 17 de cada classe, para servirem como elementos de teste;

– $X = 38$ VCs selecionados aleatoriamente para servir de modelos ou elementos de treinamento, sendo 19 de cada classe, e $40 - X = 40 - 38 = 2$ VCs restantes, sendo um de cada classe, para servirem como elementos de teste.

Finalmente, vencida a Etapa 5, foram elaborados box-plots em função de X , no qual destacam-se [7]:

• pior acurácia (Ac), melhor acurácia, acurácia média e os respectivos quartis, considerando que a acurácia implica o número total de classificações corretas, isto é, verdadeiros positivos (VP) somados aos verdadeiros negativos (VN), dentre todas as classificações, ou seja, a soma dos verdadeiros positivos com os falsos positivos (FP) com os verdadeiros negativos e com os falsos negativos (FN):

$$Ac = \frac{VP + VN}{VP + FP + VN + FN}$$

- pior especificidade (Es), melhor especificidade, especificidade média e os respectivos quartis, considerando que a especificidade implica a capacidade de classificar corretamente VCs de indivíduos saudáveis, ou seja, patológicos:

$$Es = \frac{VN}{VN + FP}$$

- pior sensibilidade (Se), melhor sensibilidade, sensibilidade média e os respectivos quartis, considerando que a sensibilidade implica a capacidade de classificar corretamente VCs de indivíduos enfermos, ou seja,

$$Se = \frac{VP}{VP + FN}$$

- pior F1 score, melhor F1 score, F1 score médio e os respectivos quartis, lembrando que o F1 score define a qualidade geral do classificador, isto é,

$$F1 \text{ score} = \frac{2 \cdot P \cdot Se}{P + Se} \quad ,$$

sendo P a precisão, isto é, $P = \frac{VP}{VP + FP}$.

Notavelmente, todas as rotinas que implementam os algoritmos e procedimentos descritos neste Capítulo foram implementadas em linguagem C/C++. Para fins complementares, os trechos dos códigos-fonte que foram construídos para implementar os procedimentos de validação cruzada para as abordagens *pattern-matching* baseada em distâncias Euclidianas e *knowledge-based* baseada na SVM, respectivamente, constam nas Figuras 13 e 12, além dos respectivos fluxogramas constantes nas Figuras 14 e 15.

Figura 12 - Trecho de código em linguagem C/C++ utilizado para implementar a validação cruzada *knowledge-based* baseada na SVM.

```

accuracy=0;
for(long i=0;i<2;i++)
    for(long j=0;j<2;j++)
        if(i==j)
            accuracy+=confusion_matrix[i][j];
accuracy/=(number_of_examples_per_class_for_testing*2);
if(accuracy>highest_accuracy)
{
    printf("\n\naccuracy: %.3f percent",(accuracy*100));
    highest_accuracy=accuracy;
    printf("\nBEST CONFUSION MATRIX TILL THE MOMENT:\n");
    for(long i=0;i<2;i++)
    {
        for(long j=0;j<2;j++)
            printf("\t%d",confusion_matrix[i][j]);
        printf("\n\n");
    }
    printf("CORRESPONDING TRAINING SET:\n");
    for(long i=0;i<number_of_examples_per_class_for_training*2;i++)
    {
        for(long j=0;j<dimension_of_each_vector;j++)
            printf("\t%.5f",train[i][j]);
        printf("\n");
    }
}
if(accuracy<lowest_accuracy)
{
    printf("\n\naccuracy: %.3f percent",(accuracy*100));
    lowest_accuracy=accuracy;
    printf("\nWORST CONFUSION MATRIX TILL THE MOMENT:\n");
    for(long i=0;i<2;i++)
    {
        for(long j=0;j<2;j++)
            printf("\t%d",confusion_matrix[i][j]);
        printf("\n\n");
    }
}

```

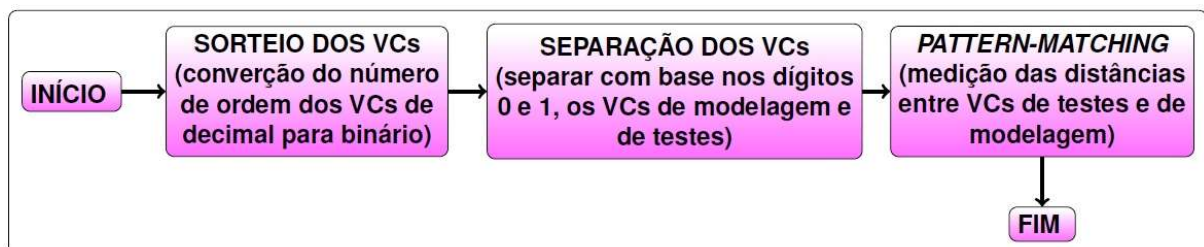
Fonte: De autoria própria.

Figura 13 - Trecho de código em linguagem C/C++ utilizado para implementar a validação cruzada *pattern-matching* baseada em distâncias Euclidianas.

```
static double** train=new
double*[number_of_examples_per_class_for_training*number_of_classes];
for(long i=0;i<number_of_examples_per_class_for_training*number_of_classes;i++)
    train[i]=new double[dimension_of_each_vector];
static double** test=new
double*[number_of_examples_per_class_for_testing*number_of_classes];
for(long i=0;i<number_of_examples_per_class_for_testing*number_of_classes;i++)
    test[i]=new double[dimension_of_each_vector];
double menor;
long training_vector;
long testing_vector;
long class_of_training_vector;
long class_of_testing_vector;
int** confusion_matrix=new int*[number_of_classes];
for(long i=0;i<number_of_classes;i++)
    confusion_matrix[i]=new int[number_of_classes];
for(long i=0;i<number_of_classes;i++)
    for(long j=0;j<number_of_classes;j++)
        confusion_matrix[i][j]=0;
long ppl_train=0;
long ppl_test=0;
double accuracy=0;
double highest_accuracy=0;
long iterations=0;
int* m=new
int[number_of_examples_per_class_for_training+number_of_examples_per_class_for_testing];
for(int i=0;i<number_of_examples_per_class_for_training;i++)
    m[i]=1;
for(int
i=number_of_examples_per_class_for_training;i<number_of_examples_per_class_for_training+
number_of_examples_per_class_for_testing;i++)
    m[i]=0;
printf("\n\n\n");
```

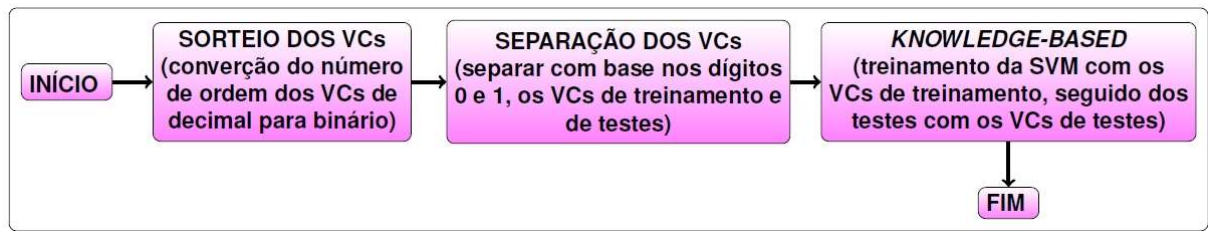
Fonte: De autoria própria.

Figura 14 - Fluxograma para a estratégia de validação cruzada *pattern-matching*



Fonte: De autoria própria.

Figura 15 - Fluxograma para a estratégia de validação cruzada *knowledge-based*



Fonte: De autoria própria.

Em ambas as Figuras 14 e 15, nota-se que o sorteio dos VCs, para a escolha daqueles que são utilizados para modelagem ou treinamento daqueles que são usados nos testes, foi realizada por um mecanismo que converte o número de ordem de cada vetor, ou seja, $\{1, 2, 3, \dots\}$, para o valor correspondente na base 2, ou seja, $\{0000, 0001, 0010, \dots\}$, visando facilitar a separação e o controle para que um mesmo VC nunca seja usado no treinamento ou modelagem e nos testes, numa mesma instância do experimento.

Diante de todo o exposto, os testes foram realizados e os resultados encontram-se descritos no próximo Capítulo.

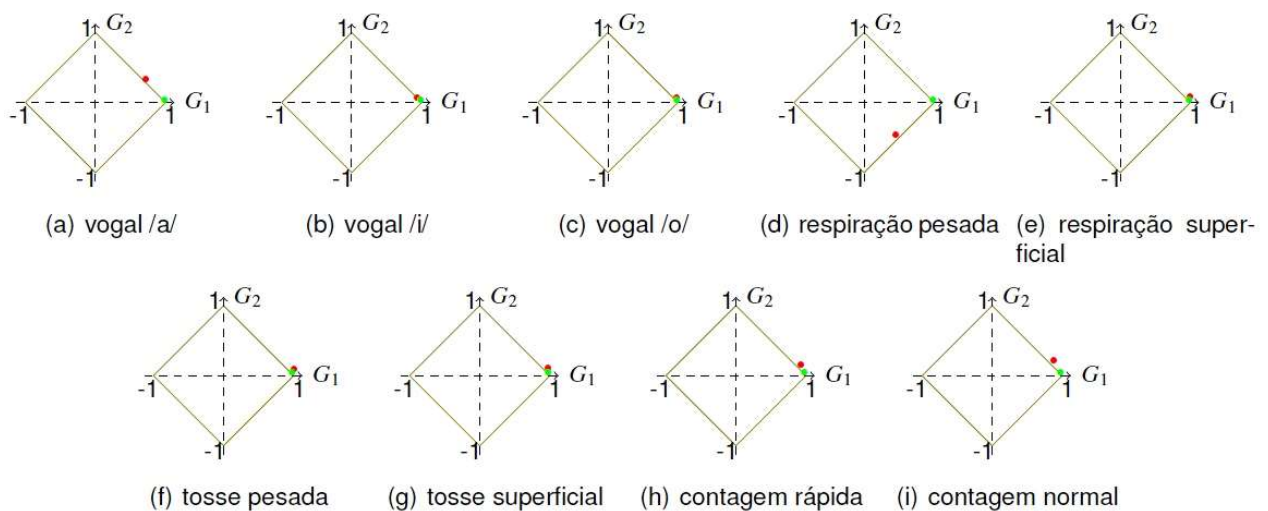
4 TESTES E RESULTADOS

Neste Capítulo, são apresentados os testes realizados e os respectivos resultados obtidos, confirmando a aplicabilidade da estratégia proposta.

4.1 Experimentos com os VCs

Os experimentos referentes à ETAPA 3 do Capítulo anterior revelaram localizações favoráveis do ponto P e uma proximidade considerável dele com o vértice ótimo, exceto para a vogal /a/ e para a respiração pesada, conforme a Figura 16.

Figura 16 - P , indicado como um ponto vermelho no plano paraconsistente, para os 40 VCs de cada um dos nove grupos de sinais.



Fonte: De autoria própria.

O vértice ótimo $(1, 0)$ está indicado como um ponto verde. Em função do quadrante no qual P se localiza dentro do plano paraconsistente, para todos os nove grupos de sinais, é coerente estimar que um classificador modesto, tal qual os baseados em critérios de distâncias ou SVMs, por exemplo, possibilitará a classificação [11].

Particularmente e de acordo com os resultados presentes na Figura 16, as distâncias dos pontos P , em vermelho naquela Figura, até o vértice ótimo (1,0) são as seguintes:

- **vogal /a/:** $\sqrt{(1 - 0.723)^2 + (0 - 0.236)^2} = \sqrt{(0.277)^2 + (0.236)^2} = 0.364$
- **vogal /i/:** $\sqrt{(1 - 0.939)^2 + (0 - (-0.023))^2} = \sqrt{(0.061)^2 + (0.023)^2} = 0.065$
- **vogal /o/:** $\sqrt{(1 - 0.978)^2 + (0 - (-0.022))^2} = \sqrt{(0.022)^2 + (0.022)^2} = 0.031$
- **respiração pesada:** $\sqrt{(1 - 0.447)^2 + (0 - (-0.553))^2} = \sqrt{(0.553)^2 + (0.553)^2} = 0.782$
- **respiração superficial:** $\sqrt{(1 - 0.993)^2 + (0 - (-0.007))^2} = \sqrt{(0.007)^2 + (0.007)^2} = 0.009$
- **tosse pesada:** $\sqrt{(1 - 0.997)^2 + (0 - (-0.003))^2} = \sqrt{(0.003)^2 + (0.003)^2} = 0.004$
- **tosse superficial:** $\sqrt{(1 - 0.979)^2 + (0 - 0.016)^2} = \sqrt{(0.021)^2 + (0.016)^2} = 0.026$
- **contagem rápida:** $\sqrt{(1 - 0.931)^2 + (0 - 0.068)^2} = \sqrt{(0.069)^2 + (0.068)^2} = 0.098$
- **contagem normal:** $\sqrt{(1 - 0.873)^2 + (0 - 0.123)^2} = \sqrt{(0.127)^2 + (0.123)^2} = 0.176$

Assim sendo, a menor distância, que é a mais favorável para a melhor separação entre as classes, é 0.004, correspondente aos sinais de tosse pesada. Portanto, somente tais VCs é que foram considerados doravante.

4.2 Experimentos com os classificadores

4.2.1 Classificador *pattern-matching* com distâncias Euclidianas

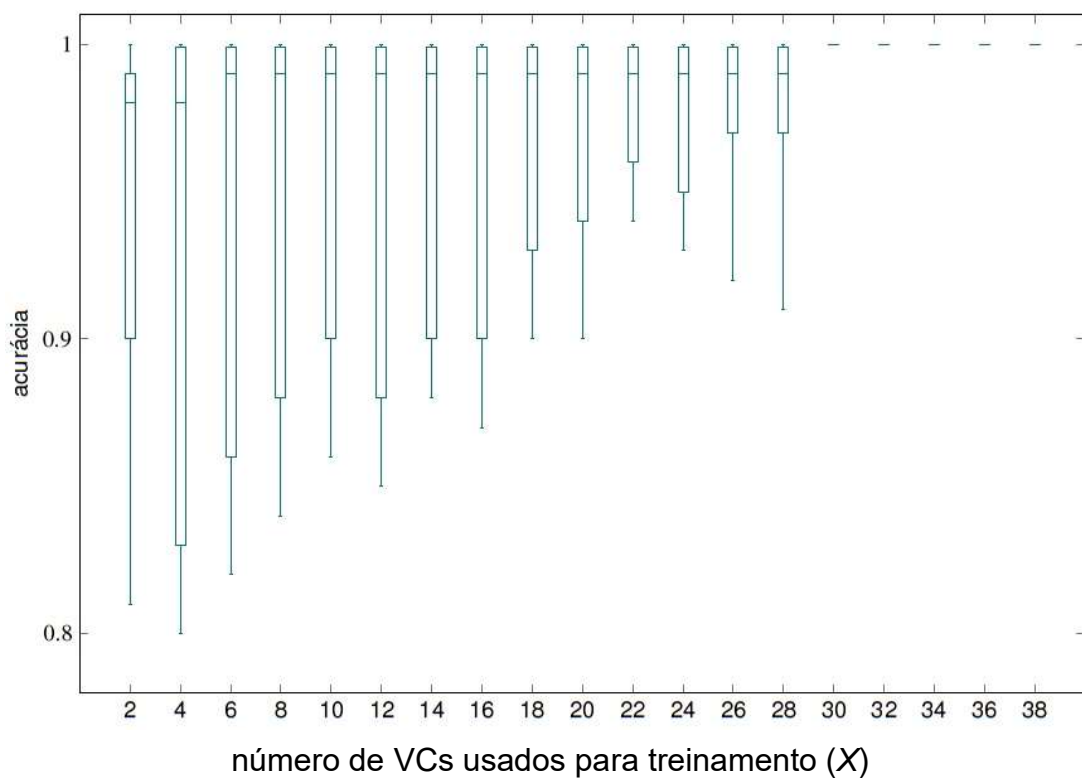
No caso da abordagem *pattern-matching*, X VCs oriundos das tosses pesadas, sendo metade de cada classe, foram utilizados para servir de modelos. Os demais VCs, ou seja, $40 - X$, foram utilizados para testes, adotando-se o procedimento de validação cruzada repetido 1000 vezes, como mencionado no Capítulo anterior.

Os resultados, constantes nos *box-plots* das Figuras de 17 até 20, corresponderam às expectativas em função da localização do ponto P na Figura 16 (f): foi possível obter significativa separação entre as classes, até mesmo com acurácia plena, baseado em um classificador modesto.

4.2.2 Classificador *knowledge-based* baseado na SVM

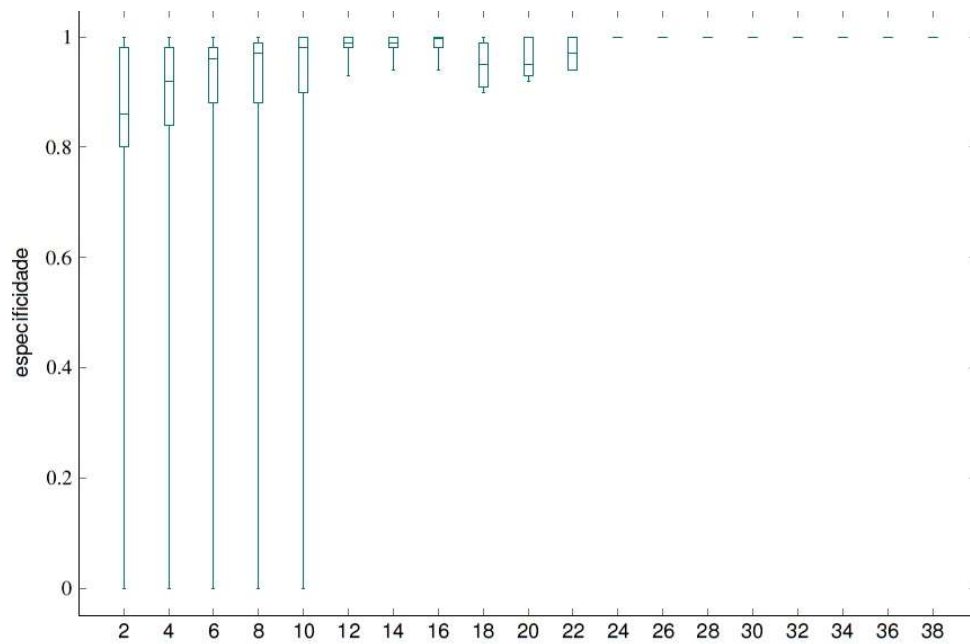
Nesta abordagem, a estrutura da SVM da Figura 11 foi utilizada variando-se o valor de X , isto é o número de elementos na camada intermediária, para que ele corresponda ao número de casos de treinamento. Assim como no caso anterior, um procedimento de validação cruzada repetido 1000 vezes foi executado, conduzindo aos resultados constantes nos *box-plots* das Figuras de 21 até 24.

Figura 17 - Box-plots resultantes da etapa de classificação baseada no classificador *pattern-matching* por distâncias Euclidianas considerando a métrica **acurácia** para $X = 2, 4, 6, \dots, 38$.



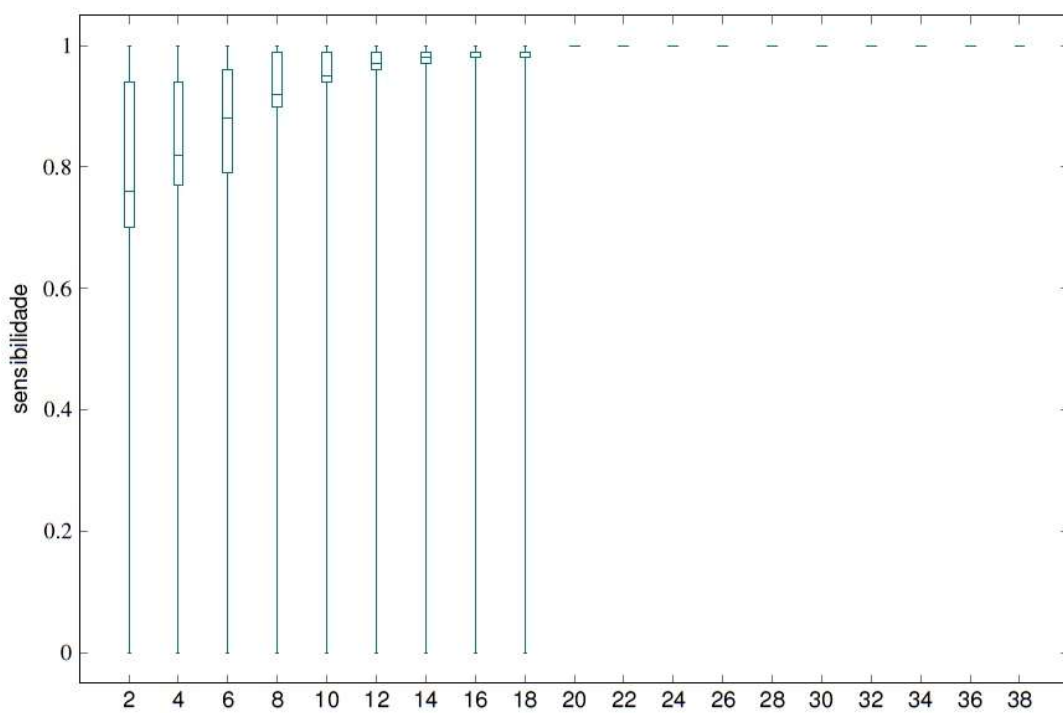
Fonte: De autoria própria.

Figura 18 - Box-plots resultantes da etapa de classificação baseada no classificador *pattern-matching* por distâncias Euclidianas considerando a métrica **especificidade** para $X = 2, 4, 6, \dots, 38$.



número de VCs usados para treinamento (X)

Figura 19 - Box-plots resultantes da etapa de classificação baseada no classificador *pattern-matching* por distâncias Euclidianas considerando a métrica **sensibilidade** para $X = 2, 4, 6, \dots, 38$.



número de VCs usados para treinamento (X)

Figura 20 - Box-plots resultantes da etapa de classificação baseada no classificador *pattern-matching* por distâncias Euclidianas considerando a métrica **F1 score** para $X = 2, 4, 6, \dots, 38$.

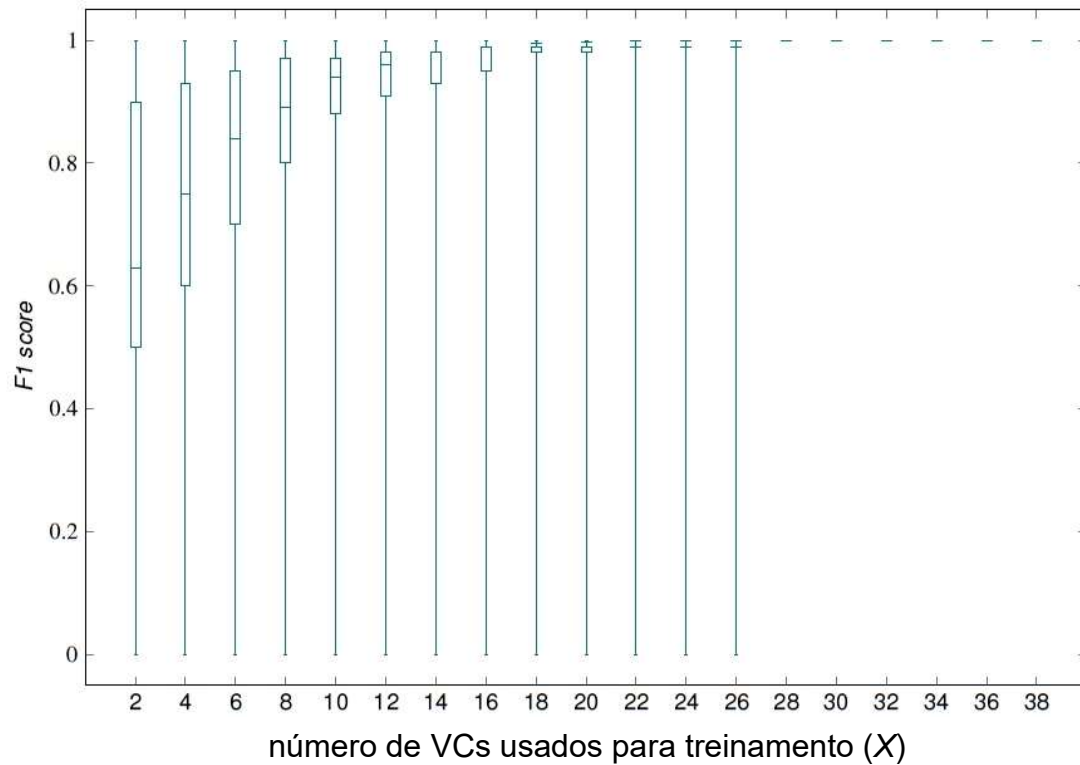


Figura 21 - Box-plots resultantes da etapa de classificação baseada no classificador *knowledge-based* por SVM considerando a métrica **acurácia** para $X = 2, 4, 6, \dots, 38$

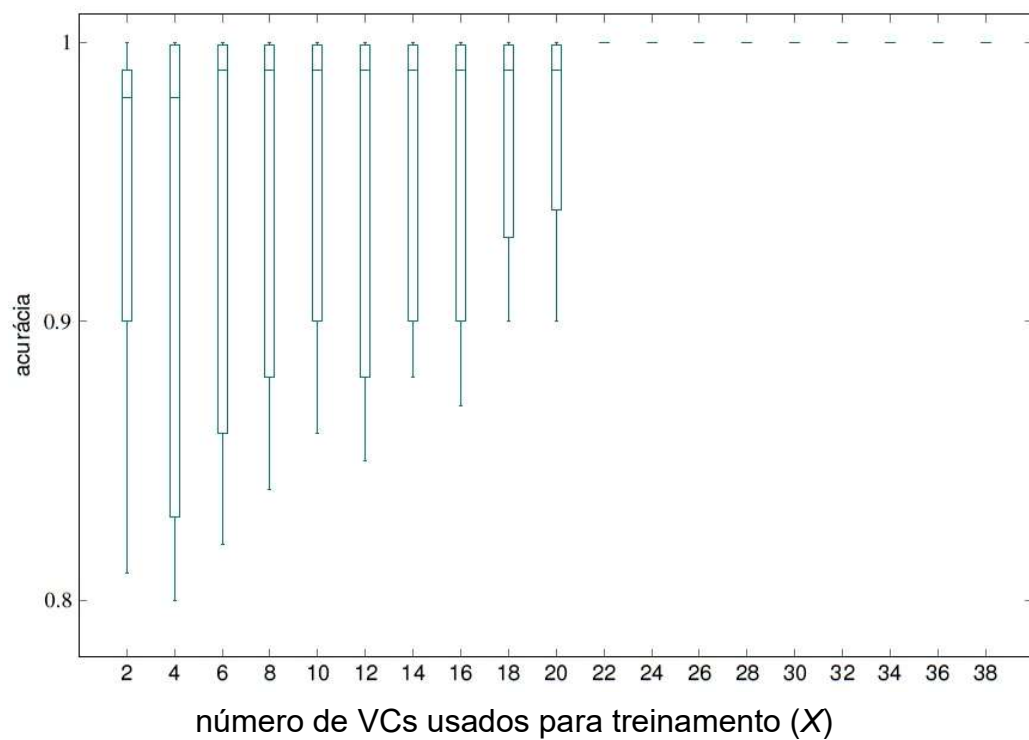


Figura 22 - Box-plots resultantes da etapa de classificação baseada no classificador *knowledge-based* por SVM considerando a métrica **especificidade** para $X = 2, 4, 6, \dots, 38$.

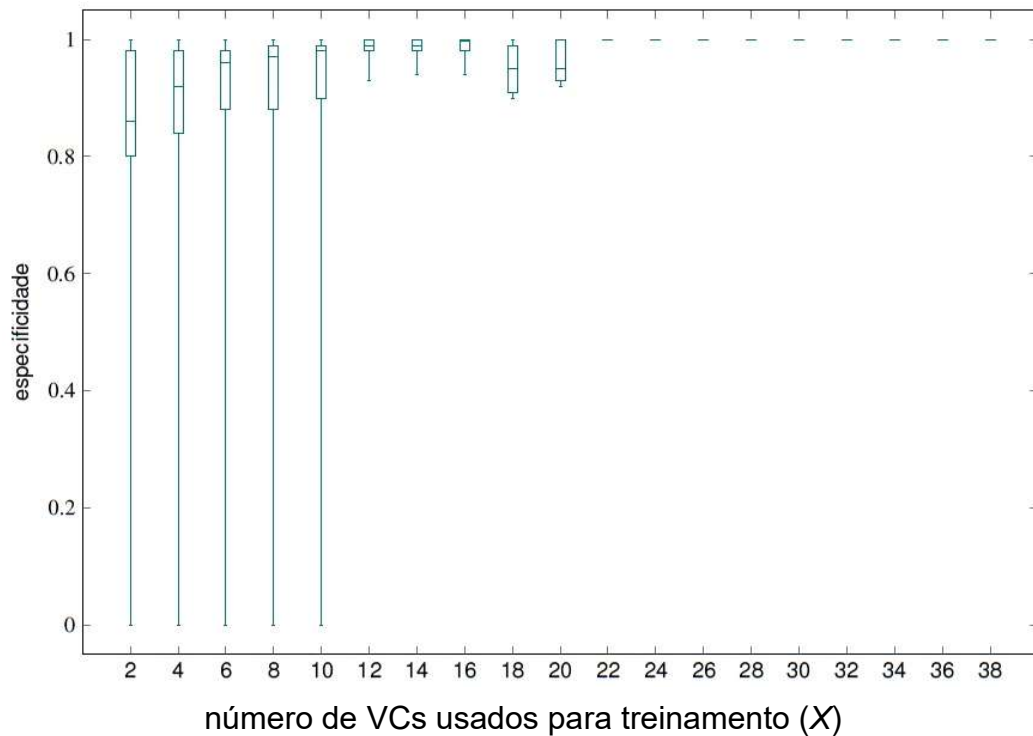


Figura 23 - Box-plots resultantes da etapa de classificação baseada no classificador *knowledge-based* por SVM considerando a métrica **sensibilidade** para $X = 2, 4, 6, \dots, 38$.

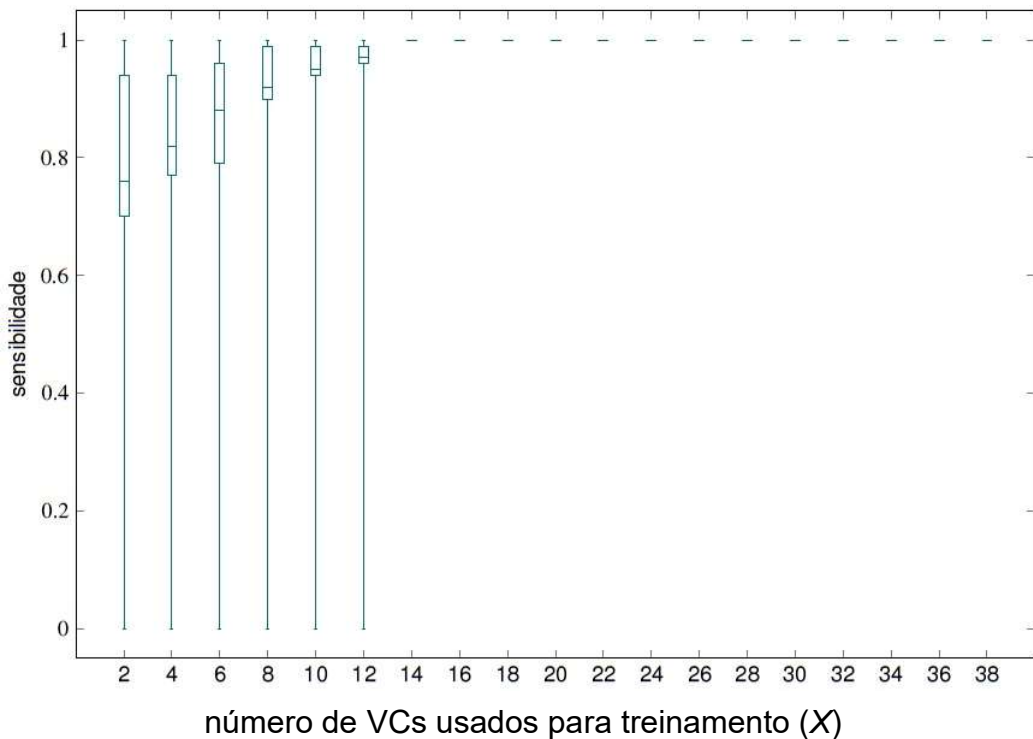
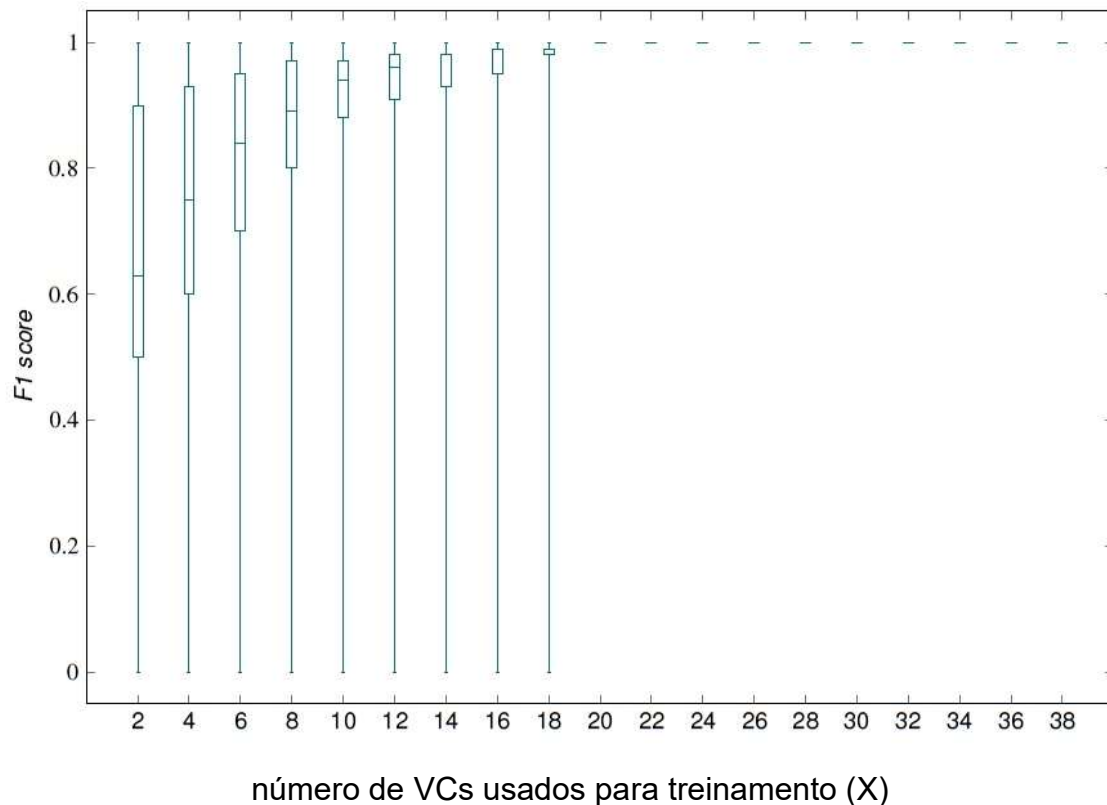


Figura 24 - Box-plots resultantes da etapa de classificação baseada no classificador *knowledge-based* por SVM considerando a métrica **F1 score** para $X = 2, 4, 6, \dots, 38$.



Fonte: De autoria própria.

Independentemente do classificador escolhido, ou seja, distâncias Euclidianas ou SVM, nota-se nas Figuras de 17 até 20 que a variabilidade nas quatro métricas estudadas são menores à medida que o número de VCs usados no treinamento cresce. Em termos de acurácia, essa tendência mostra-se clara a partir de 30 casos de treinamento, sendo 15 de cada uma das duas classes, no caso do classificador por distâncias. De modo similar, o mesmo ocorre a partir de 22 casos de treinamento, isto é, 11 para cada uma das duas classes, no caso da SVM.

Semelhantemente ao que ocorre com a acurácia, a especificidade e a sensibilidade também mostram-se transitando bruscamente de uma variação ampla para outra que é mínima, para ambos os classificadores testados. Caracterizando-se por uma avaliação mais global, o mesmo padrão é percebido para a métrica *F1 score*. Assim, de modo geral, nota-se maior segurança ao usar o modelo baseado na SVM pois ele mostrou-se mais capaz de generalizar os resultados com base em um menor

número de casos de treinamento, embora o modelo de distâncias Euclidianas também tenha se mostrado plenamente aplicável, embora com maior necessidade de casos de treinamento.

Finalmente, é importante lembrar que, em casos envolvendo avaliações de saúde, como ocorre na presente dissertação, não somente a acurácia deve ser levada em conta, mas também e principalmente, o número de falsos negativos que implicam em rotular indivíduos enfermos como estando saudáveis, trazendo a eles uma falsa e perigosa tranquilidade. Assim, é fundamental que a especificidade do sistema seja expressiva, conforme ocorre no presente caso.

No próximo Capítulo, por meio do qual conclui-se esta dissertação, os resultados e as conclusões são destacados, comentando-se também sobre os planos para experimentos futuros.

5 CONCLUSÕES

De modo geral, este estudo representa um significativo avanço na pesquisa envolvendo a detecção de COVID-19 por meio da análise de características sonoras. A abordagem proposta, que envolveu a análise, identificação, classificação e padronização das características sonoras associadas aos indivíduos contaminados e não contaminados, revelou um potencial considerável na aplicação da metodologia proposta.

A análise minuciosa das características vocais, avaliadas em nove modalidades de locuções, permitiu a criação de critérios de classificação distintos - o *pattern-matching* baseado em distâncias Euclidianas e o *knowledge-based* utilizando SVMs. Os testes de validação permitiram observar que o classificador SVM superou o método de distâncias Euclidianas, com uma ênfase particular nos resultados altamente sensíveis e específicos para os sinais de tosse pesada.

Ao conduzir o procedimento de validação cruzada repetidamente ao longo de 1000 iterações, este estudo não apenas reforçou a confiabilidade dos resultados obtidos, mas também estabeleceu uma base sólida para futuras investigações. As expectativas inicialmente fixadas foram plenamente cumpridas, indicando que a EPC apresenta um potencial promissor como método de pré-seleção de VCs, especialmente no contexto da COVID-19.

À medida que avançamos, é crucial reconhecer que este estudo representa um passo inicial em uma jornada de descoberta mais ampla. Trabalhos subsequentes, portanto, poderão aprofundar a eficácia da abordagem proposta, refinando ainda mais as classificações. Uma possibilidade registrada aqui é o uso de estratégias de *deep learning* em bases de dados relativamente maiores. Tal possibilidade, embora interessante, não foi utilizada neste trabalho pelo fato de haver baixo número de casos efetivamente utilizáveis na base de dados. Embora a possibilidade do uso de estratégias de *augmentation* existisse, preferiu-se não as adotar, visando assegurar que o trabalho não seria realizado com base em sinais sintetizados, mas sim, somente sinais de origem efetivamente humana.

REFERÊNCIAS

- 1 OPPENHEIM, Alan. **Discrete-time Signal Processing**. Pearson, United States, 1998.
- 2 BLESS, Diane. **Measurement of vocal function**. Otolaryngologic Clinics of North America, 24(5):1023–1033, 1991.
- 3 VIANA, Lisley. **Curso de fisiologia da voz – parte 02**. page 1, 2022.
- 4 FONSECA, Everthon Silva. GUIDO, Rodrigo Capobianco. SCALASSARA, Paulo Rogério. MACIEL, Carlos Dias. PEREIRA, José Carlos. **Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders**. Computers in Biology and Medicine, 37(4):571–578, 2007. Wavelet-based Algorithms for Medical Problems.
- 5 SHARMA, Neeraj. KRISHNAN, Prashant. KUMAR, Rohit. RAMOJI, Shreyas. CHETUPALLI, Srikanth Raj. R. Nirmala. GHOSH, Prasanta Kumar. GANAPATHY, Sriram. **Coswara – A Data-base of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis**. arXiv e-prints, page arXiv:2005.10548, May 2020.
- 6 E.S. Fonseca, R.C. Guido, A.C. Silvestre, and J.C. Pereira. **Discrete wavelet transform and support vector machine applied to pathological voice signals identification**. pages 5 pp.–, 2005.
- 7 HAYKIN, Simon. **Neural Networks: A Comprehensive Foundation**. Pearson, United States, 1998.
- 8 GUIDO, Rodrigo Capobianco. **A tutorial on signal energy and its applications**. Neuro-computing, 179:264–282, 2016.
- 9 GUIDO, Rodrigo Capobianco. **Enhancing teager energy operator based on a novel and appealing concept: Signal mass**. Journal of the Franklin Institute, 356(4):2346–2352, 2019.
- 10 GUIDO, Rodrigo Capobianco. **Zcr-aided neurocomputing: A study with applications**. Knowledge-Based Systems, 105:248–269, 2016.
- 11 GUIDO, Rodrigo Capobianco. **Paraconsistent feature engineering [lecture notes]**. IEEE Signal Processing Magazine, 36(1):154–158, 2019.
- 12 CARNIELLI, Marcelo Esteban Coniglio Walter. **Paraconsistent Logic: Consistency, Contradiction and Negation**. Springer, New York, 2016.
- 13 AHMED, Al-nasheri. MUHAMMAD, Ghulam. ALSULAIMAN, Mansour. ZULFIQAR, Ali. **Investigation of voice pathology detection and classification on different frequency regions using correlation functions**. Journal of Voice, 31(1):3–15, 2017.
- 14 UDHAYA, S.M. GANESAN, R., KATIRAVAN, Jeevaa. RAMAKRISHNAN, M., RUHIN, R. **Mobile application based speech and voice analysis for covid-19**

detection using computational audit techniques. International Journal of Pervasive Computing and Communications, ahead-of-print, 10, 2020.

15 AHMED, Al-nasheri. GHULAM, Muhammad. MANSOUR, Alsulaiman. ZULFIQAR Ali. TAMER, A. Mesallam. MOHAMED, Farahat. KHALID, H. Malki. MOHAMED, A. Bencherif. **An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification.** Journal of Voice, 31(1):113.e9–113.e18, 2017.

16 GHULAM, Muhammad. MANSOUR, Alsulaiman. ZULFIQAR, Ali. TAMER, A. Mesallam. MOHAMED, Farahat, KHALID, H. Malki. AHMED, Al-nasheri. MOHAMED A. Bencherif. **Voice pathology detection using interlaced derivative pattern on glottal source excitation.** Biomedical Signal Processing and Control, 31:156–164, 2017.

17 MAHMOUD, Ismail. SOHAM, Deshmukh. SINGH, Rita. **Detection of covid-19 through the analysis of vocal fold oscillations,** 10, 2020.

18 SOHAM, Deshmukh. MAHMOUD, Ismail. SINGH, Rita. **Interpreting glottal flow dynamics for detecting covid-19 from voice,** 10, 2020.

19 HASSAN, Abdelfatah. SHAHIN, Ismail, ALSABEK, Mohamed Bader. **Covid-19 detection system using recurrent neural networks.** pages 1–5, 2020.

20 SHIMON, Carmi. SHAFAT, Gabi. DANGOOR, Inbal. BEN-SHITRIT, Asher. **Artificial intelligence enabled preliminary diagnosis for covid-19 from voice cues and questionnaires.** The Journal of the Acoustical Society of America, 149:1120–1124, 02, 2021.

21 POPADINA, Alina O., SALAH, Al-Majeed. JALAL, Karam. **Voice analysis framework for asthma-covid-19 early diagnosis and prediction: Ai-based mobile cloud computing application.** pages 1803–1807, 2021.

22 LAGUARTA, Jordi. HUETO, Ferran. SUBIRANA, Brian. **Covid-19 artificial intelligence diagnosis using only cough recordings.** IEEE Open Journal of Engineering in Medicine and Biology, 1:275–281, 2020.

23 QUATIERI, Thomas. **Discrete-Time Speech Signal Processing: Principles and Practice.** Pearson, United States, 2008.