UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

GUILHERME LUIS NASCIMENTO

MAX VINICIUS DANGUI ARAUJO

IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE ATRAVÉS DO SINAL DA VOZ

CURITIBA

2022

GUILHERME LUIS NASCIMENTO
MAX VINICIUS DANGUI ARAUJO

IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE ATRAVÉS DO SINAL DA VOZ

Laryngeal Pathologies Identification Through Voice Signal

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do Curso de Bacharelado em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Marcelo de Oliveira Rosa

CURITIBA

2022

GUILHERME LUIS NASCIMENTO
MAX VINICIUS DANGUI ARAUJO

IDENTIFICAÇÃO DE PATOLOGIAS DA LARINGE ATRAVÉS DO SINAL DA VOZ

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do Curso de Bacharelado em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná.

Data de aprovação: 23/junho/2022

Marcelo de Oliveira Rosa
Doutor
Universidade Tecnológica Federal do Paraná

Glauber Gomes de Oliveira Brante
Doutor
Universidade Tecnológica Federal do Paraná

Ana Paula Dassie-Leite
Doutora
Universidade Estadual do Centro-Oeste

CURITIBA
2022

Dedicamos este trabalho às nossas famílias,
por todo o apoio ao longo dessa caminhada.

AGRADECIMENTOS

Acima de todos, agradeço a meu pai, Emerson, e minha mãe, Erika, por dedicarem suas vidas para me proporcionar este momento. Estendo também meus agradecimentos a todos os familiares e amigos, ao coautor Max, pela parceria de pesquisa e durante o curso, ao orientador Marcelo, por todo o conhecimento compartilhado com extrema paciência e dedicação, e aos professores e colegas que fizeram parte desta trajetória na UTFPR.

Guilherme Luis Nascimento

Agradeço, primeiramente, à minha mãe, Maria Cândida, que infelizmente não pôde presenciar fisicamente a conclusão desta etapa em minha vida, por todo o esforço, dedicação e instrução até onde foi possível. À minha irmã, Ronize, pelo suporte, incentivo e por sempre acreditar em mim. Também deixo meus agradecimentos ao Guilherme, minha dupla durante a elaboração deste trabalho e durante alguns anos de curso, ao Professor Marcelo Rosa pela orientação, bem como meu pai Max, Tia Malu, aos primos Carlos e Flávio e aos demais colegas que acompanharam minha jornada até este momento.

Max Vinicius Dangui Araujo

RESUMO

Visando estudar uma solução para as adversidades dos métodos de diagnóstico atuais, esta monografia buscou identificar e diagnosticar patologias da laringe de maneira não invasiva. A pesquisa se desenvolveu através da extração de medidas acústicas – como *pitch*, *jitter*, *shimmer* e outras - e posterior aplicação em métodos de aprendizado de máquina, no caso os classificadores Máquinas de Vetor de Suporte (MVS) e Redes Neurais Artificiais (RNA). Baseados nos dados estatísticos dos modelos testados, foram feitas análises da efetividade e da possibilidade de aplicação em diagnósticos médicos.

Palavras-chave: identificação de patologias da laringe; classificação através do aprendizado de máquina; máquinas de vetores de suporte; redes neurais artificiais; saúde vocal.

ABSTRACT

Aiming to study a solution to the adversities of current diagnostic methods, this monography sought to identify and diagnose laryngeal pathologies in a non-invasive way. The research was developed through the extraction of acoustic measures - such as pitch, jitter, shimmer and others - and subsequent application in machine learning methods, in this case the Support Vector Machines (SVM) and Artificial Neural Networks (ANN) classifiers. Based on the statistical data of the models tested, analyses of the effectiveness and the possibility of application in medical diagnoses were carried out.

Keywords: identification of laryngeal pathologies; classification through machine learning; support vector machines; artificial neural networks; vocal health.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

Siglas

A/D	Analógico-digital
A_i	Amplitude pico-a-pico
ANN	*Artificial Neural Networks*
APQ	Quociente de Perturbação da Amplitude
dB	Decibel
DP	Desvio padrão
F_0	Frequência fundamental ou pitch
F_{HI}	Frequência Fundamental Máxima
F_{LO}	Frequência Fundamental Mínima
FN	Falso Negativo
FP	Falso Positivo
Hz	Hertz
JITA	*Jitter* absoluto
JITT	Porcentagem de *Jitter*
MVS	Máquina de Vetores de Suporte
NHR	Relação Ruido-Harmônico
PPC	Proeminência do Pico Cepstral
RAP	Perturbação Média Relativa
ReLu	*Rectified Linear unit*
RNA	Redes Neurais Artificiais
SHDB	*Shimmer*
SHIM	Porcentagem de *Shimmer*
SVM	*Support Vector Machines*

T_0	Período Fundamental
TanH	Tangente Hiperbólica
T_i	Tempo em determinado instante
UTFPR	Universidade Tecnológica Federal do Paraná
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo

LISTA DE SÍMBOLOS

Letras Gregas

φ	Transformação que leva os dados de um espaço de entrada para um espaço de saída
Θ	Bias para Redes Neurais Artificiais
η	Taxa de aprendizagem

Notações

$/a/$	Fonema vocálico a

SUMÁRIO

1 INTRODUÇÃO

1.1 Tema

A história da comunicação humana é um ramo dos estudos evolutivos que ainda divide opiniões. Qual foi a primeira linguagem utilizada pelos homens primitivos é uma incógnita para os pesquisadores, que argumentam com hipóteses de que começamos a nos comunicar através de grunhidos, gestos ou uma combinação de ambos (BORDENAVE, 1997). Contudo, há um ponto em comum entre tantas teorias: a voz teve um papel essencial, tanto no surgimento quanto na evolução da comunicação.

Milhares de anos depois, a voz segue como um dos pilares mais importantes para o mundo que vivemos. Ela é o instrumento para cerca de um terço das atividades profissionais (VILKMAN, 2004 apud MINISTÉRIO DA SAÚDE, 2018), um dos meios mais precisos de transmissão de emoções (KRAUS, 2017) e ainda ligada ao desenvolvimento humano, como no amadurecimento dos bebês, impulsionando o desenvolvimento cerebral (WEBBA *et al.*, 2015) e redução da bradicardia e apneia em recém-nascidos prematuros (DOHENY *et al.*, 2012).

Como todo som, a voz nada mais é que uma variação da pressão e densidade do ar, que se propaga somente por um meio material (SERWAY; JEWETT, 2016). Nos seres humanos, para Kent e Read (2002), esse processo de produção da fala envolve três subsistemas: respiratório, onde o pulmão fornece a energia para a produção do som; laringeal, no qual estão localizadas as pregas vocais, que regulam a passagem do ar e modulam o sinal na frequência; e, por último, o articulador, formado pelas passagens acima da laringe (LADEFOGED; JOHNSON, 2011) e os órgãos articuladores (língua, lábios, mandíbula e velum), que são responsáveis pela ressonância e modelagem do ar, respectivamente.

Para que o processo ocorra de forma eficiente e o resultado seja sempre satisfatório, todos os subsistemas precisam estar livres de quaisquer patologias. Rouquidão, cansaço e baixa intensidade são alguns dos vários exemplos de sintomas de uma voz anômala (PRZYSIEZNY; PRZYSIEZNY, 2015), que podem se originar do mau uso da voz, deficiências musculares ou em nervos e até transtornos psicológicos (SODRÉ, 2016). Além das dores e desconforto, o prejuízo ao enfermo estende-se à sua comunicação, afetando suas relações pessoais, a vida profissional - como é caso para radialistas, locutores, cantores e profissões de atendimento ao público (ROSA, 1998) – e até a saúde mental, já que pessoas com problemas no trato vocal podem sofrer preconceito e exclusão social, o que posteriormente pode ocasionar transtornos psicológicos graves. Assim, é de extrema importância identificar essas patologias com rapidez e precisão, sobretudo pelas chances de recuperação sem a necessidade de tratamentos mais agressivos serem muito maiores quando a identificação é feita logo no início (ANDRADE, 2003).

Em contrapartida, o diagnóstico de uma disfonia é um processo trabalhoso, exigindo muito conhecimento prévio, observação e intuição do profissional. Atualmente, os métodos mais comuns são dependentes de uma análise perceptiva do médico, que tenta identificar patologias

ao escutar a voz do paciente - exame feito por um fonoaudiólogo - ou observar diretamente as cordas vocais, no caso de um diagnóstico laringológico; ambos os exames são subjetivos e, no caso do segundo, muito incômodo ao enfermo (ROSA, 1998).

Incentivados pelas dificuldades citadas e com o auxílio de ferramentas computacionais, foi estudada uma forma alternativa de diagnóstico nesse trabalho. A voz é uma onda longitudinal, com componentes de amplitude e tempo, da qual se pode obter medidas como a intensidade e frequência, por exemplo. Isso é possível através da análise gráfica do sinal, obtido no processo de digitalização da mesma, que envolve as etapas de conversão A/D, filtragem, amostragem e quantização do sinal (KENT; READ, 2002). Esses indicadores são geralmente retirados de um fonema sustentado, como /a/, que permitem uma análise focada na laringe e apresentam menor complexidade de sinal em relação a palavras inteiras ou frases, ainda que também não sejam senoidais (SPEAKS, 2018).

Quando quantificados, os parâmetros avaliativos podem finalmente definir de forma padronizada o que é uma voz saudável ou não. Comparando os dados considerados normais aos de um paciente - levando sempre em consideração características como a idade, sexo e hábitos de cada pessoa (ANDRADE, 2003) -, é possível encontrar variações entre eles, indicando a existência de uma patologia. No caso positivo, também é possível tentar determinar qual é esta doença, ao aferir as mesmas informações com um banco de dados de vozes patológicas pré-estabelecido, além de acompanhar a evolução do tratamento indicado pelo médico (ROSA, 1998). Este diagnóstico acontece através da aplicação do aprendizado de máquina, onde algoritmos comparadores são treinados para identificarem essas alterações, construindo assim uma avaliação rápida, com maior grau de exatidão e sem causar desconforto ao paciente, qual foi explorada detalhadamente nesta pesquisa.

1.1.1 Delimitação do Tema

O estudo realizado tem como objetivo a detecção de patologias da laringe através do sinal da voz, solucionando os problemas recorrentes dos métodos de avaliação tradicionais.

Isso foi feito a partir do estudo de um banco de dados de vozes pré-definido, que conta com vozes saudáveis e de pessoas com disfonias já identificadas. Dele, utilizando o *software* Praat, retirou-se dos sinais parâmetros através das análises no tempo e frequência, resultando em dados como *pitch*, *jitter*, *shimmer* e variações, e espectral, que produz medidas como a proeminência do pico cepstral (CPP, do inglês *Cepstral Peak Prominence*).

Após a obtenção desses dados, a comparação para o diagnóstico das vozes patológicas foi feita via *software* MATLAB, utilizando os modelos de Máquina de Vetores de Suporte e Redes Neurais Artificiais. No final, discutiu-se os resultados da classificação por ambos os métodos.

1.2 Problemas e Premissas

A identificação das patologias da voz inicia-se através de uma entrevista com o paciente, levantando informações como o histórico médico e sintomas apresentados, seguida então de uma análise sonora, na qual o fonoaudiólogo escuta a voz do paciente e estuda fatores como o tom, sonoridade (ou intensidade) e qualidade, em busca de desvios do que é considerado padrão nessas qualificações (PINDZOLA; PLEXICO; HAYNES, 2015). Por fim, faz-se necessária a observação direta da laringe e trato vocal através de um diagnóstico laringológico, mediante o uso de equipamentos introduzidos na boca e garganta ou nariz do paciente, como um espelho na laringoscopia indireta e um endoscópio (rígido ou flexível) ou estroboscópio, na videolaringoscopia, onde o especialista pode observar o movimento das pregas vocais e a posição, aparência e configuração da estrutura em geral (ROSA, 1998; PINDZOLA; PLEXICO; HAYNES, 2015).

Ainda que tradicionais, esses métodos trazem consigo vários problemas, podendo levar a dificuldades no diagnóstico e desconforto do paciente. Sodré (2016) diz que os exames feitos pela boca ou nariz causam um grande desconforto e podem ser mais difíceis de serem realizados em crianças, além de ser corriqueiro não poder observar as pregas em funcionamento no caso nasal, já que o paciente não consegue articular os fonemas ou frases necessárias para a avaliação. Em adição, os testes são dependentes dos sentidos do médico, ou seja, subjetivos. Identificar um distúrbio na voz através da audição ou visão é um processo impreciso e que exige não só conhecimento, mas também experiência do analista, dificultando a sua realização por profissionais menos experientes. A dificuldade de definir um padrão nas classificações da voz, ainda que existam metodologias como a GRBAS e PERC, completa a lista de empecilhos que comprometem o diagnóstico (ROSA, 1998).

Embora solucione os problemas de desconforto, subjetividade e falta de padronização no diagnóstico, o método de identificação de patologias pela voz também possui obstáculos. Em diversos estudos apontados por Awan, Roy e Dromey (2009), é possível concluir que a análise a partir do tempo e frequência possui sua acurácia vulnerável à presença de ruído do sinal. Além disso, ela só pode ser aplicada a fonemas sustentados, método que, em casos, pode apresentar uma maior dificuldade na identificação de patologias (YIU et al, 2000 apud AWAN et al., 2009), além de inflar os valores dos parâmetros, devido a segmentos não sonoros e padrões de entonação.

Como o foco da pesquisa é identificar problemas apenas na laringe, a utilização de um fonema vocálico faz-se necessária, pois evita uma maior interferência do trato vocal e seus articuladores (língua e lábios, por exemplo) no sinal. Assim, para tornar a identificação das disfonias mais precisa, também decidiu-se por utilizar a análise cepstral, que possui a “capacidade de produzir estimativas de aperiodicidade e/ou ruído aditivo sem a identificação de limites de ciclo individuais” (AWAN; ROY; DROMEY, 2009).

1.3 Objetivos

1.3.1 Objetivo Geral

Através da análise matemática do sinal da voz, método não invasivo e que permite a padronização das avaliações, determinar a existência ou não de patologias na laringe utilizando algoritmos classificadores. Em adição, nos casos positivos, diagnosticar pelo mesmo método qual é a patologia em específico.

1.3.2 Objetivos Específicos

- Estudar a viabilidade do método não invasivo para a classificação da condição da voz e identificação de doenças;
- Estudar a aplicação do aprendizado de máquina no reconhecimento de doenças, em específico os métodos de Máquina de Vetores de Suporte e Redes Neurais Artificiais;
- Analisar a eficiência do método e de cada um dos classificadores.

1.4 Justificativa

Para Bordenave (1997), a comunicação está tão presente e enraizada em no nosso dia a dia que “somente percebemos a sua essencial importância quando, por um acidente ou uma doença, perdemos a capacidade de nos comunicar”. Considerando que a voz é um dos principais pilares da comunicação humana, podemos compreender a relevância dos cuidados com a mesma, afinal, qualquer disfunção pode afetar não só a saúde, mas também a vida social e profissional das pessoas.

Os estudos realizados nesse trabalho permitem que os principais problemas dos métodos tradicionais aplicados na identificação de patologias na laringe sejam evitados, possibilitando um diagnóstico não invasivo, rápido, preciso e padronizado. Além disso, a quantificação dos parâmetros avaliativos facilita com que os médicos determinem o melhor curso de tratamento para cada condição, bem como acompanhem a eficácia e evolução do mesmo.

Junto dos avanços técnicos, a aplicação dessa metodologia vai de encontro aos problemas encontrados em países com estrutura pública de saúde saturada e atendimento lento. A popularização da avaliação laringológica, redução de custos operacionais e tempo de espera e a importância da descoberta precoce de doenças pode não só ser um ponto positivo para o paciente e profissionais da saúde, mas também para o sistema de saúde público e privado, elevando consideravelmente a qualidade e acessibilidade do serviço.

Vale também ressaltar que, apesar de outros pesquisadores já terem apresentado trabalhos neste campo - que serviram de referência teórica e foram citados ao longo desta monografia - , ainda se faz necessário a continuação desses estudos, pois há espaço para a melhora dos resultados obtidos, como o aumento da sensibilidade e especificidade dos exames. Isso ocorre através da aplicação de diferentes métodos de cálculo para a obtenção dos parâmetros avaliativos e o uso de novos algoritmos comparadores, devido a acelerada evolução das técnicas de aprendizado de máquina. A adição da análise cepstral, por exemplo, escolhida para complementar a lista de medidas acústicas a serem obtidas dos sinais, apresenta a possibilidade de melhora dos resultados na identificação de qual é a patologia existente, considerando o estudo apresentado por Awan, Roy e Dromey (2009).

1.5 Procedimentos Metodológicos

Inicialmente, faz-se necessário a compreensão da estrutura e funcionamento do aparelho da fala, patologias que afligem o subsistema laringeal, conceitos matemáticos para a extração de medidas acústicas e métodos de classificação através do aprendizado de máquina, fundamentos buscados na literatura de cada campo.

Com o auxílio do *software* Praat, foram obtidos os parâmetros dos sinais a serem analisados, disponíveis no banco de vozes Modelo 4337, da KayPENTAX Corp (Kay Elemetrics Corp, Lincoln Park, NJ). Apenas as amostras com o fonema /a/ foram utilizadas.

Em seguida, agora somente pelo MATLAB - mais especificamente através do ajuste de *scripts* fornecidos pelo aplicativo *Classification Learner*, disponível na biblioteca do *software* -, foram desenvolvidos os algoritmos classificadores. Os mesmos passaram primeiramente por um treinamento e então realizaram o diagnóstico de vozes anômalas e quais são suas patologias. Por fim, obteve-se medidas de desempenho para a compreensão da eficiência dos métodos.

1.6 Estrutura do Trabalho

Três etapas formam esta monografia. Inicialmente, foi introduzido o tema principal de estudo, contextualizando a importância da voz e traçando os problemas nos métodos atuais de diagnósticos de patologias na laringe, bem como a solução desenvolvida no trabalho.

Em sequência, a base teórica para a compreensão da pesquisa foi apresentada. Neste segundo capítulo, quatro tópicos foram abordados: a anatomia e fisiologia do aparelho da fala; as patologias que afetam a laringe, especificamente as que estão presentes no banco de dados de vozes utilizado; as medidas acústicas obtidas do sinal da voz; e os classificadores, no caso os métodos de Máquina de Vetores de Suporte e Redes Neurais Artificiais.

Por último, discutiu-se o processo de implementação dos parâmetros avaliativos nos algoritmos classificadores, a relevância dos mesmos para o processo e os resultados alcançados, quantificando o desempenho do método na identificação de vozes anômalas e suas patologias.

2 REFERENCIAL TEÓRICO

Nesta etapa, abordaremos com mais profundidade os conceitos teóricos, que servem de base para esta pesquisa. Em específico, são apresentados: a anatomia e fisiologia do aparelho da fala, detalhando a composição dos subsistemas e o papel de cada um na formação da voz; as patologias que foram analisadas pelos métodos de classificação; as medidas acústicas a serem usadas na identificação das anomalias vocais; e, por último, os modelos de Máquinas de Vetores de Suporte e Redes Neurais Artificiais, responsáveis pela comparação dos dados e a determinação do diagnóstico final.

2.1 Anatomia e fisiologia do aparelho da fala

Por não existir um único órgão com a função da produção vocal, faz-se necessário a operação integrada de vários sistemas para que o processo de vocalização seja possibilitado (ANDRADE, 2003). A união dessas estruturas é denominada aparato vocal e é composto pelos pulmões, traqueia, laringe e cavidade supraglotal (cavidades nasal e oral, língua, lábios e dentes) (RAZERA, 2004).

Os órgãos desses sistemas têm a comunicação como função secundária. Suas atividades primárias são permitir a respiração e participar do sistema digestivo. Dessa forma, a produção dos fonemas é regulada pelas condições fisiológicas desses órgãos (ROSA, 1998).

O aparato vocal é uma estrutura complexa com diversos componentes. Nesta seção, são apresentados os subsistemas de forma simplificada e as etapas do processo de formação da voz.

2.1.1 Subsistemas

Podemos subdividir o aparato vocal em três grupos com a finalidade de facilitar sua compreensão fisiológica.

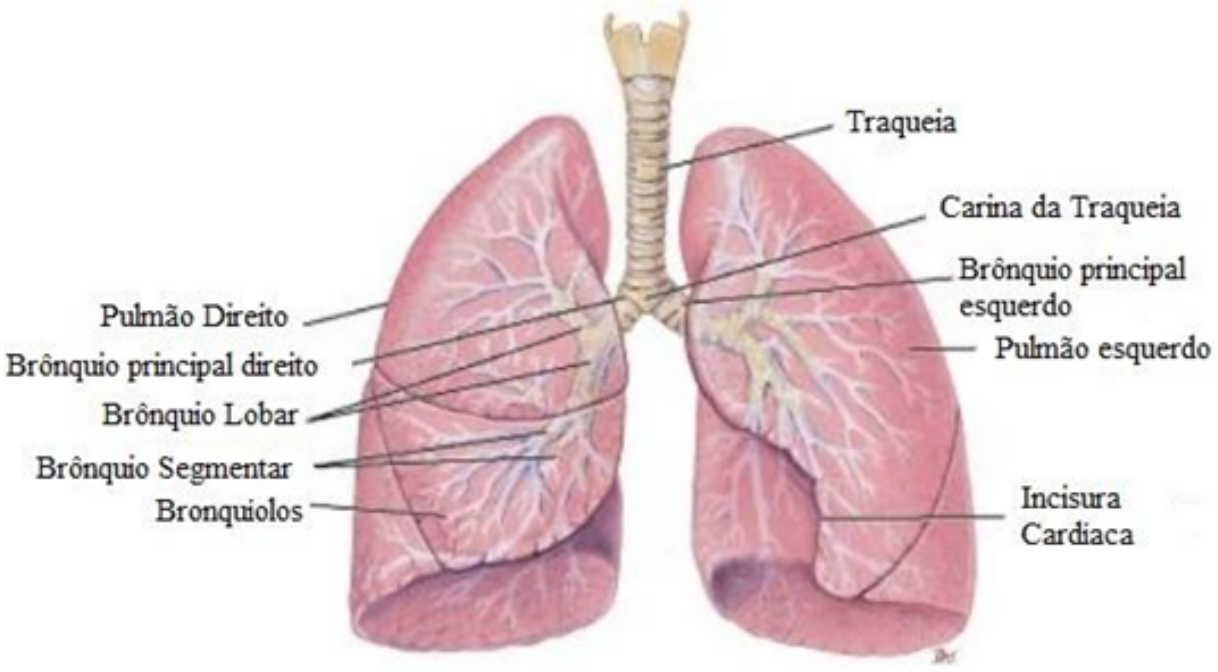
O primeiro subsistema é composto pelos pulmões e pela traqueia, ilustrado pela Figura 1. Os pulmões são estruturas vascularizadas essenciais para a produção da voz. Estes têm como função principal a hematose, trocas gasosas que permitem o fluxo de oxigênio presente no ar para o sangue e de dióxido de carbono presente no sangue para o ambiente externo (RAZERA, 2004).

A traqueia é um conduto cilíndrico formado por anéis de cartilagem (RAZERA, 2004). É a estrutura que permite o escoamento dos gases advindos do processo da respiração para dentro e fora dos pulmões.

Além destas estruturas há também o diafragma (composto por musculatura, tendão e membrana). Este, localizado abaixo dos pulmões, separa a cavidade torácica da cavidade abdo-

minal e está ligado diretamente ao sistema respiratório e, dessa forma, com o sistema fonatório (ROSA, 1998).

Figura 1 – Anatomia do sistema respiratório



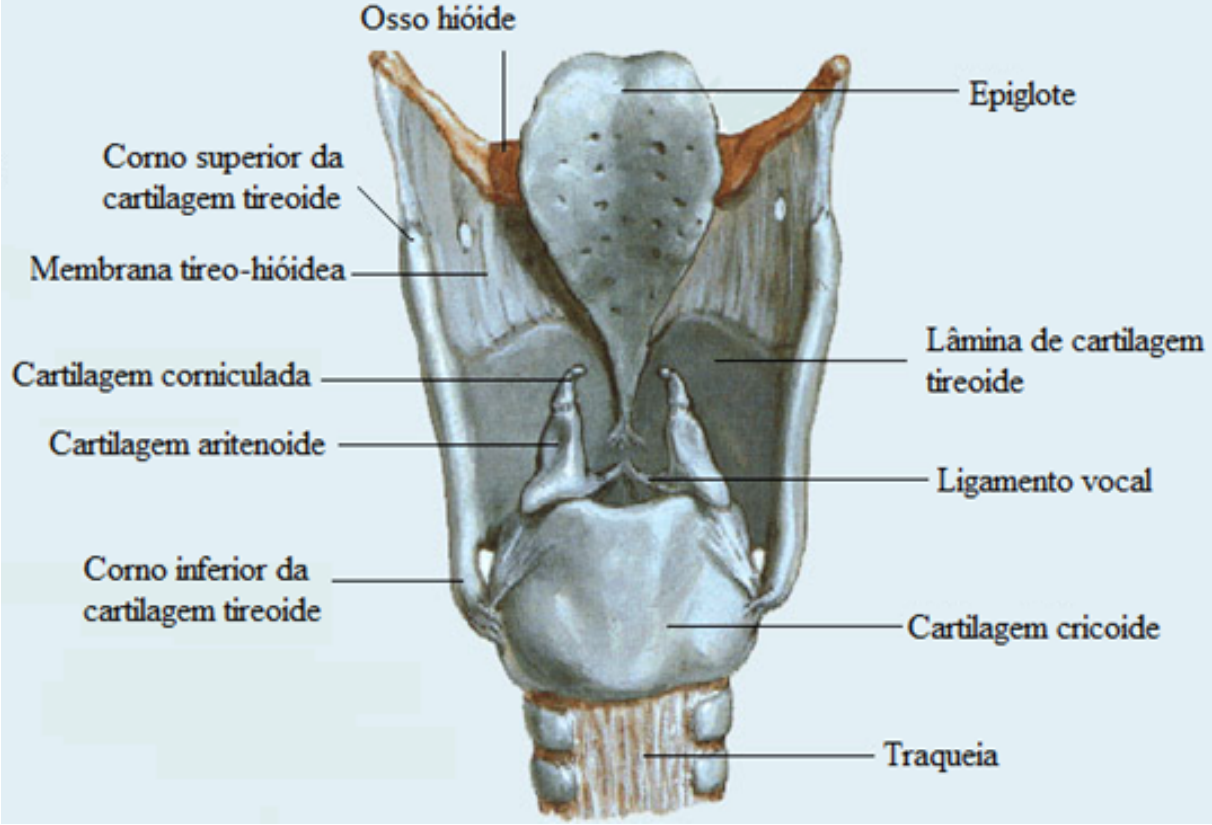
Fonte: Anatomia em foco[1]**.**

O segundo subsistema é formado pela laringe (Figura 2. É uma estrutura localizada acima da traqueia e é constituída por cartilagens, músculos, membranas (COSTA, 2012) e o ossos. Possui diversas funções, dentre as mais importantes podemos citar as atividades respiratória, deglutitória e fonatória (ANDRADE, 2003).

Na atividade respiratória e deglutitória, a laringe atua como um esfíncter, estrutura chamada de epiglote, evitando que qualquer corpo, além do ar, se dirija aos pulmões e encaminha o bolo alimentar para o esôfago. Além disso, o movimento das pregas vocais auxilia o movimento respiratório. Na fonação, a elasticidade, tensão, ampliação e diminuição da abertura glótica somado ao esforço respiratório, promove as variações nos tons da voz (COSTA, 2012).

[1] Disponível em: <https://www.anatomiaemfoco.com.br/sistema-respiratorio/pulmao-humano/> Acesso em: 11 de agosto de 2021.

Figura 2 – Vista posterior da laringe



Fonte: COSTA, 2012.

A laringe pode ser dividida em três regiões: supraglote, glote e infraglote. A primeira é composta por todo o espaço acima da glote até o orifício superior da laringe. A segunda é constituída pelas pregas vocais. E a terceira é a região posicionada abaixo das pregas vocais até a base laríngea (JUNQUEIRA, 1999).

Na glote estão localizadas um dos principais componentes para a produção da voz, as pregas vocais (Figura 3). Assim, as estruturas que formam esse membro, bem como a manutenção da saúde laríngea, têm um papel de destaque nas análises patológicas em geral.

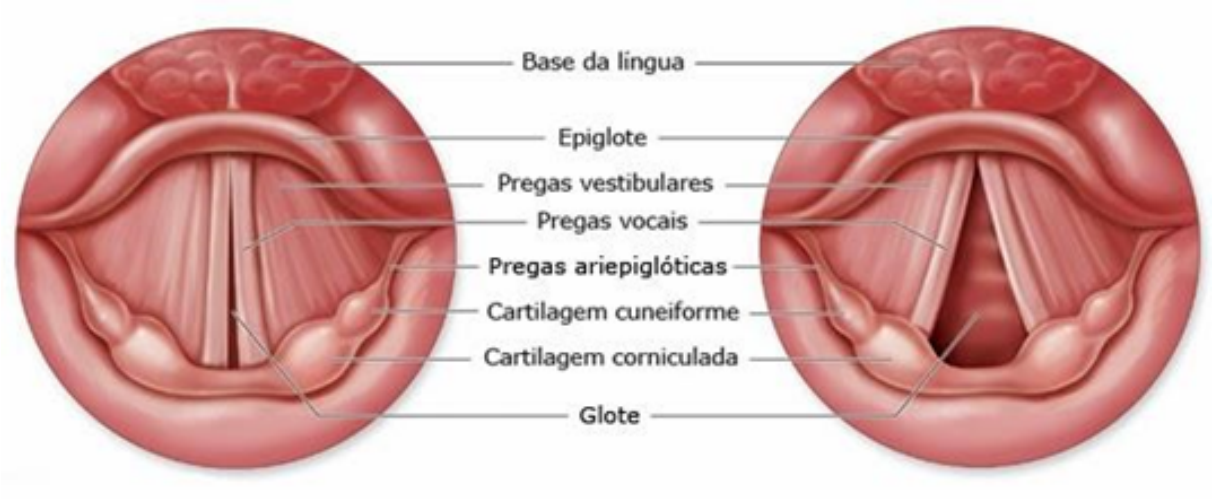
Segundo Costa (2012), o componente mais importante dessa região é a cartilagem tireóidea. O formato de duas lâminas laterais tem em sua união um ângulo denominado proeminência laríngea (Figura 2); esse ângulo apresenta variações de acordo com o sexo, observando-se valores próximos a 90ara o sexo masculino e 120ara o feminino. A variação angular é responsável pela definição do tamanho das pregas vocais e contribui para a demarcação das frequências por elas emitidas.

Por ser um componente de extrema importância, as pregas vocais devem ser analisadas com um maior nível de detalhes. Sua composição pode ser subdividida em três camadas: superficial, intermediária e profunda (HIRANO, 1977 apud ANDRADE, 2003).

De acordo com Andrade (2003), a camada superficial, denominada espaço de Reinke, é flexível e por isso é a que mais vibra durante a fonação. As camadas intermediária e profunda

são constituídas principalmente por fibras, elásticas e de colágeno respectivamente. A junção destas camadas é nomeada de ligamento vocal.

Figura 3 – Anatomia das pregas vocais

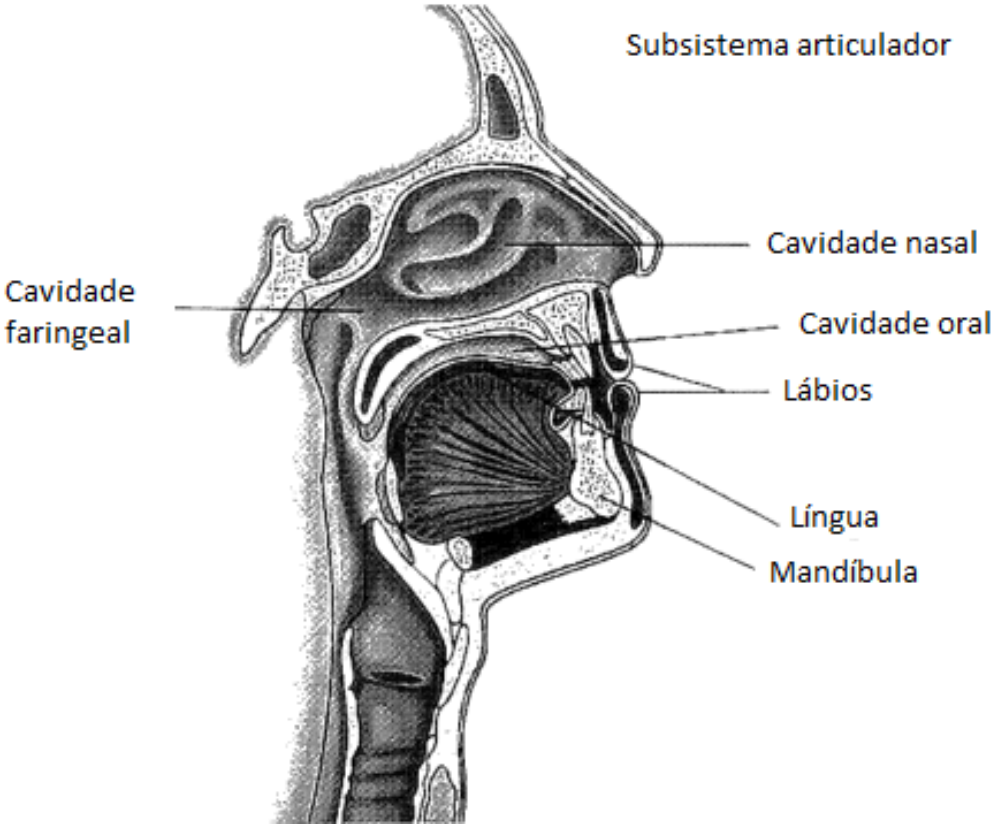


Fonte: Voz instrumento[2].

O terceiro subsistema é o trato vocal, demonstrado pela Figura 4. Segundo Razera (2004) é representada pela porção superior do aparato vocal e acima das pregas vocais. O sinal originário da glote corresponde a um tom de baixa intensidade, o qual precisa de amplificação e de modelagem para que os fonemas sejam caracterizados (ROSA, 1998).

[2] Disponível em: <http://vozinstrumento.blogspot.com/p/conhecendo-voz.html> Acesso em: 12 de agosto de 2021.

Figura 4 – Trato vocal



Fonte: Adaptado de Kent e Read (2002).

O trato vocal é composto pelas cavidades nasal e oral, língua, dentes e lábios. Estas cavidades atuam como cadeia de ressonadores para os sons originários da laringe (RAZERA, 2004) e os demais componentes como modeladores do sinal.

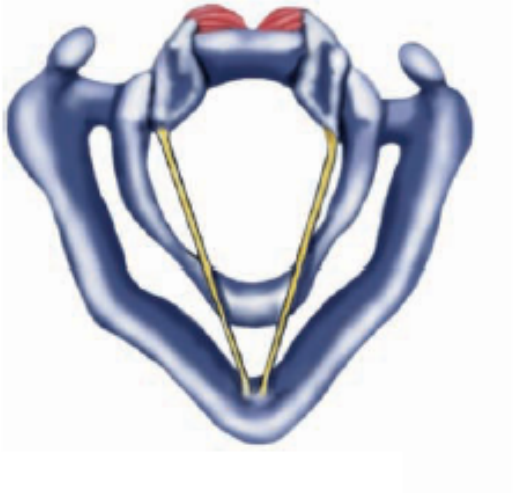
2.1.2 Formação da voz

Nós, seres humanos, respiramos para fornecer oxigênio aos sistemas do corpo (SEIKEL; KING; DRUMRIGHT, 2009), entretanto, também utilizamos o sistema respiratório para outra função: a produção da fala. Na primeira atribuição, os músculos inspiratórios trabalham em sincronia com o aumento do volume da cavidade torácica durante a entrada do ar, mas deixam de atuar durante a expiração; já na segunda, estes músculos continuam a trabalhar durante a expiração (CALLOU; LEITE, 2009), permitindo com que seja produzido um fluxo de ar pelos pulmões, que é expelido do órgão e viaja através da traqueia até chegar na laringe (LADEFOGED; JOHNSON, 2011).

Agora no segundo subsistema, o ar encontra a glote, onde estão localizadas as pregas vocais. Nesse momento, para compreendermos a modulação da voz, devemos observar o estado das pregas. Se relaxada, a mesmas estarão em formato triangular (SODRÉ, 2016), como mostra a Figura 5, permitindo com que o ar passe com facilidade e sem sofrer grandes altera-

ções em direção ao subsistema fonador. Isso ocorre durante a respiração, por exemplo, e os sons resultantes são chamados de surdos ou desvozeados (CALLOU; LEITE, 2009).

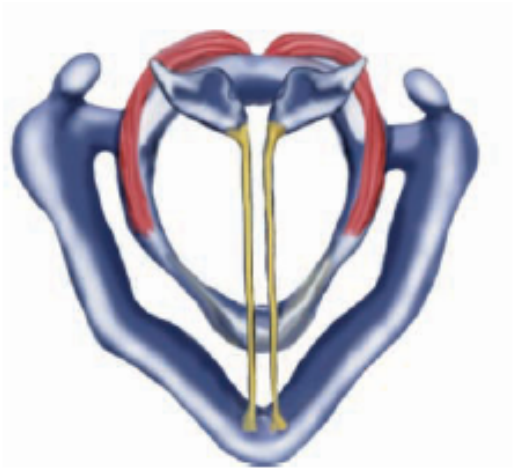
Figura 5 – Pregas vocais relaxadas



Fonte: SEIKEL, 2009.

Por outro lado, se elas estiverem contraídas, o fluxo de ar é barrado, aumentando a pressão local até o ponto em que a as pregas são forçadas a se abrirem, como mostra a Figura 6. Devido a sua construção elástica e as foças aerodinâmicas, as pregas acabam vibrando, modulando periodicamente o fluxo de ar (SODRÉ, 2016). Com a passagem do ar e a redução da pressão, as pregas voltam a se fechar, reiniciando o processo, que pode ser completado em cerca 9 milissegundos, de acordo com Seikel, King e Drumright (2009). Os sons produzidos dessa forma são chamados de sonoros ou vozeados (CALLOU; LEITE, 2009).

Figura 6 – Pregas vocais contraídas



Fonte: SEIKEL, 2009.

Essa vibração ocorre em uma frequência primária – que em média é de 212 Hz para mulheres adultas, 132 Hz para homens adultos e em torno de 300 Hz para crianças - ou em harmônicas da mesma, quais são importantes para a diferenciação de fonemas. A diferença

entre os valores fundamentais é causada por fatores como a massa e comprimento das pregas, variando entre sexos e idades (SEIKEL; KING; DRUMRIGHT, 2009).

Também é importante destacar que as pregas apresentam outras configurações além da abdução e adução completa. Isso permite a produção de sons mais específicos, como uma voz sussurrada ou estridente (DAVENPORT; HANNAHS, 2010). Essas variações são chamadas de registros vocais, sendo que a mais comum - utilizada diariamente por nós - é conhecida como fonação modal ou voz modal (SEIKEL; KING; DRUMRIGHT, 2009).

Apesar de agora modulado na frequência, o fluxo de ar que passa pela glote ainda não forma a voz como conhecemos (SILVA, 2007). Para que isso aconteça, o som deve passar pelo terceiro e último subsistema, o articulador, responsável pela ressonância e modelagem do ar (SODRÉ, 2016).

Após deixar a glote, o fluxo de ar pode ser direcionado para o trato nasal ou oral pelo véu palatino (LADEFOGED; JOHNSON, 2011). Se direcionado para o nariz e boca, há a produção de sons nasais; já se seguir apenas pela boca, forma-se os sons orais (DAVENPORT; HANNAHS, 2010). Em ambos os casos, o ar passa por uma ressonância (CALLOU; LEITE, 2009) e é moldado pelos articuladores, como o movimento da língua e lábios, que criam obstáculos a esse fluxo. Ao contrário das consoantes, as vogais envolvem pouca ou nenhuma constrição em sua produção (SILVA, 2007), ou seja, o som produzido na laringe sofre pouca alteração, permitindo uma análise mais aprofundada da atuação do órgão.

2.2 Patologias

Tom, sonoridade e qualidade (este último analisado através de fatores como aspereza, rouquidão e soprosidade) são atributos que formam nossa percepção da voz. Assim, qualquer alteração na frequência fundamental, amplitude e complexidade do sinal - medidas que descrevem matematicamente as características citadas - pode ser considerado um distúrbio da voz (PINDZOLA; PLEXICO; HAYNES, 2015).

Esses distúrbios afetam os três subsistemas e sua origem varia amplamente, podendo ser o resultado do mau uso da voz, alguma deformidade ou doenças, partindo de casos simples, como resfriado, alergias e laringite, a situações mais graves, como refluxo gástrico, câncer, doença de Parkinson e esclerose lateral amiotrófica (LANIER, 2010). Seja qual for a causa, dificuldade e cansaço ao falar, rouquidão, projeção vocal limitada, dor e até a quase ou completa ausência da voz (PRZYSIEZNY; PRZYSIEZNY, 2015) são sintomas causados por estas enfermidades.

No caso do subsistema laringeal, foco desta pesquisa, Pindzola, Plexico e Haynes (2015) afirmam que a laringe deve ser capaz de variar a pressão do fluxo de ar de forma alternada e regular, realizando aberturas e fechamentos de diferentes graus e ajustes durante o processo da fala, utilizando durante esse processo a quantidade correta de energia, para evitar tensões indevidas. As pregas vocais também devem possuir um tamanho e formato aproximadamente

igual (seguindo o padrão para a idade e sexo da pessoa), além de mover-se de forma síncrona. Qualquer variação nessas atividades e propriedades faz com que o a produção da fala não seja realizada de forma eficiente.

Essas alterações na laringe são classificadas de acordo com sua origem: se causadas pelo mau uso da voz ou transtornos psicológicos, elas são chamadas de comportamentais ou funcionais; já se forem resultado de irregularidades físicas na anatomia do corpo, elas ficam conhecidas como orgânicas (ROSA, 1998), possuindo as subdivisões estrutural, para as causadas por alguma lesão ou anormalidade física, e neurogênica, resultantes de problemas no sistema nervoso (SODRÉ, 2016).

As patologias a serem analisadas pelo algoritmo classificador são restritas à disponibilidade dos bancos de dados utilizados. Além disso, foi determinado uma taxa de amostragem mínima de 20 casos por doença, uma estimativa para assegurar o funcionamento ideal da técnica aplicada. Assim, considerando estas imposições, foram selecionadas um total de oito patologias para este estudo, quais são descritas a seguir.

2.2.1 Ceratose - Leucoplasia

De acordo com Rosa (1998), ceratose caracteriza-se pelo surgimento de placas rosadas nas pregas vocais, enquanto leucoplasia são lesões de coloração esbranquiçada que ocorrem na membrana superficial da mucosa, causando rigidez na região. A união das duas não é por acaso, já que essas condições possuem similaridades: ocorrem com frequência nas cordas vocais, apresentam displasia epitelial e, em alguns casos, podem ser cancerígenas (GILLIS *et al.*, 1983).

O consumo abusivo de tabaco e álcool é relacionado à patologia, mas no geral as causas são diversas, apresentando casos que aparentam estar relacionados com a anemia por deficiência de ferro ou mesmo sem nenhum fator que indique origem das doenças (GILLIS *et al.*, 1983). Já segundo Gillis *et al.* (1983), a leucoplasia também pode ser consequência de uma infecção viral, refluxo laringofágico e traumas nas pregas vocais.

2.2.2 Compressão ventricular

Também conhecida como disfonia ventricular, esta patologia caracteriza-se pela vibração das pregas vestibulares, seja ao mesmo tempo ou no lugar das pregas vocais (SODRÉ, 2016).

Normalmente, as pregas vestibulares não realizam o trabalho de fonação, portanto são chamadas de pregas vocais falsas (SEIKEL; KING; DRUMRIGHT, 2009). Entretanto, elas passam a realizar essa função para compensar algum tipo de lesão nas pregas vocais (SODRÉ,

2016), resultando em uma voz muito rugosa, com pouca variação de amplitude e baixa frequência (ROSA, 1998).

2.2.3 Edema de Reinke

O Espaço de Reinke é a região entre a camada superficial da lâmina própria e o ligamento vocal (HIRANO, 1976 apud TAVALUC; TAN-GELLER, 2019). Quando a área apresenta um acúmulo de fluído gelatinoso, ocorre um aumento considerável de tamanho (de forma bilateral) nas pregas vocais (ROSA, 1998), impedindo que as mesmas vibrem de forma adequada.

A patologia, também conhecida como degeneração polipoidal, faz com que a frequência fundamental seja diminuída, resultando em uma voz mais grave, além de causar soprosidade (ROSA, 1998). Segundo Gainor, Chowdhury e Sataloff (2011), as causas mais comuns da doença são o fumo, abuso da voz, refluxo laringofaríngeo e hipotireoidismo.

2.2.4 Hiperfunção

Estabelece-se como hiperfunção o mau uso das pregas vocais devido a forças musculares excessivas ou desequilibradas (HILLMAN et al, 1989 apud STEPP, 2011). Segundo Sodré (2016), esta condição ainda pode ser separada em dois casos:

- Hiperadução: quando as pregas se unem com muita força, causando uma voz de má qualidade e forçada, resultando em fadiga, dor e desconforto aos enfermos;
- Hiperabdução: quando as pregas não conseguem contrair-se completamente para a fonação, levando a uma voz fraca e soprosa. Os pacientes que sofrem da condição enfrentam fadiga ao falar.

As causas podem ser o uso excessivo para o primeiro caso e condições emocionais, como o estresse, para o segundo.

2.2.5 Paralisia

Doença que afeta a capacidade de adução e abdução das pregas vocais (SEIKEL; KING; DRUMRIGHT, 2009), devido a alterações nos núcleos do tronco cerebral, nervo vago ou do nervo laríngeo recorrente (SODRÉ, 2016). Ela pode ocorrer de forma unilateral, quando um dos lados das pregas vocais não consegue realizar o movimento de contração (ROSA, 1998), e bilateral, onde ambas as pregas não conseguem fechar, caso mais grave e mais provável a ser irreversível (PLOTT, 1964 apud MICHAELS, 2012).

Dependendo da posição da paralisia, a patologia causa soprosidade, devido a incapacidade de fechamento das pregas para realizar a fonação (ROSA, 1998), ocorrendo fonação tanto na inspiração quanto na expiração, segundo Seikel, King e Drumright (2009).

2.2.6 Refluxo laringofaríngeo

Refluxo gastroesofágico é fluxo contrário de alimentos ou líquidos do estômago para o esôfago e as passagens respiratórias (SEIKEL; KING; DRUMRIGHT, 2009). Apesar de não ser uma doença da laringe, ela causa tosse, pigarro e irritação (MICHAELS, 2012), levando a lesões no órgão e nas cordas vocais, sendo também associada ao desenvolvimento de úlceras nas pregas vocais (SEIKEL; KING; DRUMRIGHT, 2009).

2.2.7 Edema das pregas vocais

Em termos gerais, edema é o acúmulo excessivo de fluídos nos tecidos do corpo (TAMPARO; LEWIS, 2011). No caso desta patologia, isso ocorre nas pregas vocais, gerando um inchaço na região.

Suas consequências são muito similares às do Edema de Reinke, entretanto, o banco de vozes Modelo 4337, da KayPENTAX Corp (Kay Elemetrics Corp, Lincoln Park, NJ), que foi utilizado no desenvolvimento desse trabalho, separa as duas doenças. Supõe-se que as razões para isso estejam nas possíveis diferenças nas causas das doenças.

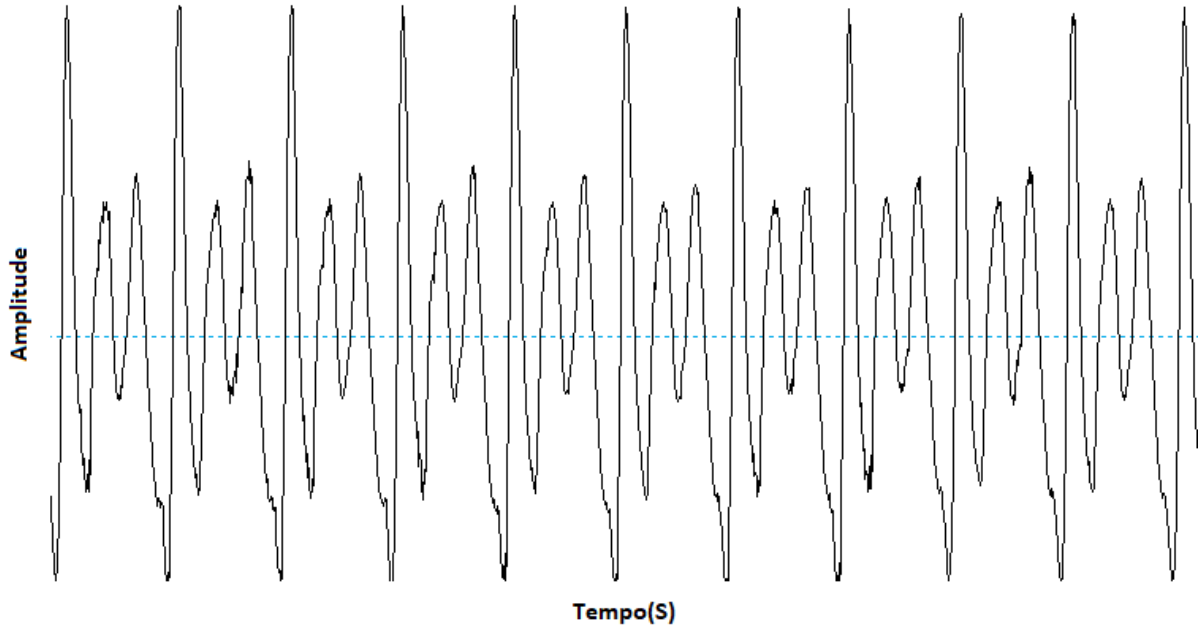
2.2.8 AP squeezing

A-P *squeezing* é uma patologia relacionada a problemas de hiperfuncionalidade das pregas vocais. Suas características são semelhantes a da hiperfunção, já descrita no texto, mas a separação das duas condições parte do banco de vozes da KayPENTAX Corp (Kay Elemetrics Corp, Lincoln Park, NJ). Supõe-se também que as razões para isso estejam nas possíveis diferenças nas causas das doenças.

2.3 Análise acústica

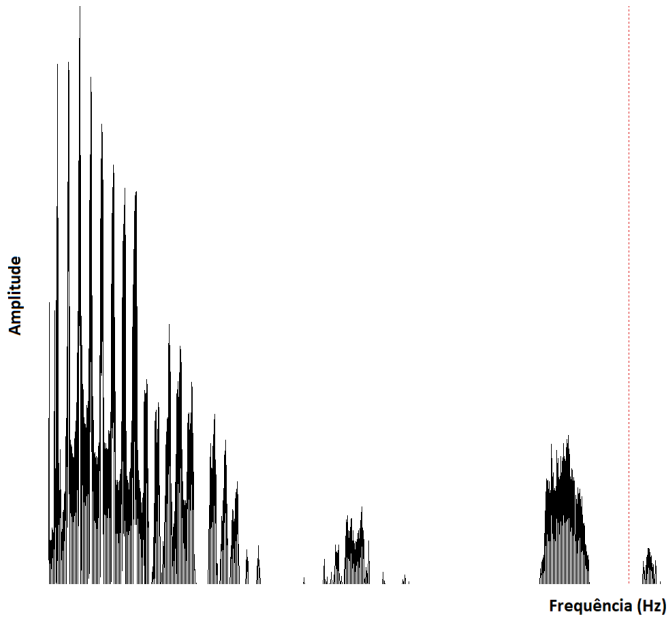
Um sinal da voz pode ser digitalizado e então analisado graficamente, onde, através de métodos matemáticos, podemos obter parâmetros que o caracterizam, chamados de medidas acústicas. Elas podem ser alcançadas através de uma análise do sinal no tempo ou na frequência, como nas Figuras 7 e 8, respectivamente.

Figura 7 – Exemplo de um sinal de voz analisado graficamente no tempo



Fonte: Autoria Própria.

Figura 8 – Exemplo de um sinal de voz analisado graficamente na frequência



Fonte: Autoria Própria.

O método de identificação de doenças da laringe através da fala é dependente dessas medidas, pois elas permitem compararmos parâmetros de um sinal diagnosticado previamente - seja normal ou patológico - com os de uma nova amostra a ser avaliada. Assim, nesta seção foram descritas as medidas importantes para esse processo e que foram utilizadas nesta pesquisa. A escolha das mesmas foi influenciada pelas teses dos autores Rosa (1998) e Sodré (2016).

2.3.1 Frequência Fundamental (Pitch)

Frequência fundamental, também chamada de *pitch* e representada pela variável f_0, é a medida que retrata a quantidade de vibrações – ou ciclos de adução e abdução - que as pregas vocais realizam por segundo (SODRÉ, 2016).

Medida em Hertz (Hz), ela pode variar de acordo com os diferentes sons produzidos, portanto trata-se da média entre os máximos e mínimos observados. Esses valores são definidos pelas características físicas das pregas, como comprimento, massa e tensão (ALMEIDA, 2010), que mudam de acordo com a idade e sexo, levando às diferentes faixas de frequência fundamental que são consideradas padrões para esses grupos: 212 Hz para mulheres adultas, 132 Hz para homens adultos e em torno de 300 Hz para crianças (SEIKEL; KING; DRUMRIGHT, 2009).

Alterações na laringe causam variações no valor de *pitch*, podendo indicar - ao compará-lo com amostras já diagnosticadas - a existência ou não de uma patologia na laringe (SODRÉ, 2016).

2.3.2 Frequência Fundamental Máxima (Fhi) e Mínima (Flo)

A frequência fundamental máxima (F_{HI}) representa o maior valor de frequência fundamental medida em um sinal de voz; analogamente, a frequência fundamental mínima (F_{LO}) é o menor valor observado (SODRÉ, 2016). Como são parâmetros derivados do *pitch*, ambos também são medidos em Hertz.

Juntas, elas formam o que pode ser chamado de alcance ou gama de *pitch*. Reduções nessa banda podem indicar a existência de uma patologia na laringe, de acordo com Seikel, King e Drumright (2009).

2.3.3 Período Fundamental (T0)

Representado por T_0, o período fundamental é determinado pelo inverso da frequência fundamental, dado pela Equação 1:

$$T_0 = 1/f0 \tag{1}$$

2.3.4 Jitter absoluto (JITA)

Sodré (2016) elucida que o *jitter* absoluto quantifica a variação do período fundamental. Podemos definir como a média das diferenças absolutas de dois períodos subsequentes. Sua unidade de medida é o segundo (s).

Podemos descrever através da Equação 2 (HORII, 1979 apud SODRÉ, 2016):

$$JITA = \frac{1}{N-1} \sum_{i=1}^{N-1} \mid T_1 - T_{i+1} \mid \tag{2}$$

Onde

- N: número total de amostras no período avaliado;
- T_i: tempo no instante i (período da amostra de f_0);

2.3.5 Porcentagem de jitter (JITT)

Esta medida estima a variação do período do tom de voz dentro da amostra *jitter* analisada em porcentagem. Em outras palavras, diferença média absoluta entre períodos subsequentes dividido pelo período médio (SODRÉ, 2016).

Podemos expressar matematicamente esta medida - medida em porcentagem (%) - através da Equação 3 (MURRY; DOHERTY, 1980 apud SODRÉ, 2016):

$$JITT = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \mid T_1 - T_{i+1} \mid}{\frac{1}{N} \sum_{i=1}^{N-1} T_i} \tag{3}$$

Onde:

- N: número total de amostras no período avaliado;
- T_i: tempo no instante i (período da amostra de f_0);

2.3.6 Shimmer (SHDB)

Esta medida, por sua vez, demonstra a variação da amplitude pico-a-pico do sinal analisado (SODRÉ, 2016) ou ainda definir como uma medida de estabilidade fonatória.

É expressa em decibéis (dB) e definida pela Equação 4 (SODRÉ, 2016):

$$SHDB = \frac{1}{N-1} \sum_{i=1}^{N-1} 20log \mid \frac{A_{i+1}}{A_i} \mid \tag{4}$$

Onde:

- N: número de períodos extraídos da frequência fundamental f_0;
- A_i: amplitude pico-a-pico.

2.3.7 Porcentagem de shimmer (SHIM)

Avalia a variação percentual da amplitude pico-a-pico. É descrita pela Equação 5 e é expressa em porcentagem (SODRÉ, 2016):

$$SHIM = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} \mid A_i - A_{i+1} \mid}{\sum_{i=1}^{N-1} A_i} \tag{5}$$

- N: número de períodos extraídos da frequência fundamental F_0;
- A_i: amplitude pico-a-pico.

2.3.8 Desvio Padrão (DP)

Esta medida é referida ao desvio padrão de todas as amostras e f_0 coletadas. É expressa em Hertz e definida pela Equação 6 (SODRÉ, 2016):

$$STD = \sqrt{\frac{\sum_{i=1}^{N}(f_i - \overline{f_0})}{N-1}} \tag{6}$$

Onde:

- N: número total de amostras no período avaliado;
- f_i: frequência no instante i.

2.3.9 Perturbação Média Relativa (RAP)

De acordo com Scalassara (2009), os sons vocálicos estáveis apresentam mudanças sutis e suaves no período de *pitch*, assim, uma forma de medir esses distúrbios na frequência é através da perturbação média relativa (RAP). Esta medida estima as variações de período de tom de voz da amostra. Em outras palavras, é a diferença média entre um período, sua média e a média de seus dois períodos mais próximos – o anterior e o posterior (SODRÉ, 2016).

Esta medida é expressa em Hertz e pode ser descrita pela Equação 7 (SODRÉ, 2016):

$$RAP = \frac{\frac{1}{N-2}\sum_{k=2}^{N-1} \mid \frac{T_{k-1} + T_k + T_{k+1}}{3} \mid - T_i}{\frac{1}{N}\sum_{k=1}^{N-1} T_0} \tag{7}$$

Onde:

- N: número total de amostras no período avaliado;
- T_i: tempo no instante i (período da amostra de f_0);

2.3.10 Quociente de Perturbação da Amplitude (APQ)

A forma mais comum de cálculo de perturbações na frequência (*Shimmer*) é pelo quociente de perturbação da amplitude (APQ) (SCALASSARA, 2009).

Esta medida avalia as variações de amplitude com um fator de amortecimento de 11 períodos (SODRÉ, 2016), ou seja, analisa uma janela de 5 períodos anteriores e posteriores. É expressa em porcentagem e pode ser definida através da Equação 8:

$$APQ = \frac{\frac{1}{N} - 10 \sum_{i=6}^{N-5} \mid A_i \mid -(\sum_{k=i-5}^{i+5} \frac{A_k}{11})}{\frac{1}{N} \sum_{i=1}^{N} A_i} \tag{8}$$

2.3.11 Relação Ruído-Harmônico (NHR)

Esta medida busca avaliar o grau de rouquidão nos sinais de vogais sustentadas. Isso é feito pela observação da quantidade de harmônicos que é substituída por ruído (ARAÚJO et al., 2002 apud SCALASSARA, 2009). De uma forma mais simplificada, é o quociente da medida da energia acústica das componentes harmônicas pela energia acústica das componentes de ruído (YUMOTO, 1982 apud SODRÉ, 2016).

Podemos representar NHR através da Equação 9:

$$NHR = \frac{N \int_0^T f_A^2(\tau) d\tau}{\sum_{i=1}^{N} \int_0^{T_i} [f_i(\tau) - f_A(\tau)]^2 d\tau} \tag{9}$$

2.3.12 Proeminência do Pico Cepstral (PPC)

De acordo com (LOPES *et al.*, 2019), como complemento às medidas de *jitter* e *shimmer*, a análise cepstral é uma alternativa muito boa ao se avaliar perturbação de sinais, visto que é capaz de determinar f_0 e produzir estimativas de ruído aditivo. As medidas cepstrais são mais confiáveis que as medidas tradicionais de perturbação para obtenção de medidas referentes ao desvio vocal.

A proeminência do pico cepstral (PPC), medido em decibel (dB), tem como objetivo quantificar o grau de periodicidade vocal acima dos ruídos (HILDEBRAND; CLEVELAND; ERICKSON, 1994 apud GOMES, 2019).

Como exposto por Murton, Hillman e Mehta (2020), o cepstro é definido pela transformada inversa de Fourier do espectro acústico, ou de forma simplificada, “espectro de um espectro”. A forma de onda passa por uma transformada de Fourier, passando para o domínio espectral. O resultado então é submetido a um logaritmo e, em seguida, a uma nova transformada Fourier, agora inversa, chegando então no domínio cepstral. Os picos de harmônicos periódicos são representados como um único grande pico e a amplitude - ou proeminência -

deste em relação a uma linha de regressão, traçada através do cepstro, é definida como proeminência do pico cepstral (MURTON; HILLMAN; MEHTA, 2020).

2.4 Classificadores

De acordo com Izenman (2008), o conceito de aprendizado de máquina parte do subcampo da ciência da computação chamado inteligência artificial. Para que uma máquina desenvolva a capacidade de racionalizar como os seres humanos, é imperativo que ela saiba aprender com suas experiências, criando então a capacidade de melhorar suas decisões com o passar do tempo.

Esse processo acontece através de algoritmos e sistemas e são divididos em duas categorias (IZENMAN, 2008): o aprendizado supervisionado, onde o código é alimentado com uma base de dados de entrada e com resultados corretos de saída, buscando então uma função que aproxima essas variáveis; e o aprendizado não supervisionado, onde não há uma variável de saída definida para guiar a definição do resultado.

O uso dessa metodologia é amplo, sendo comumente empregado em problemas de classificação e regressão, por exemplo. No primeiro, que é o caso aplicado nesta monografia, o algoritmo deve analisar os dados recebidos e então separá-los em suas determinadas categorias, enquanto no segundo o mesmo deve prever um resultado de acordo com os valores de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

A concepção do classificador, em termos gerais, está ligada a separação dos dados ao dispor em três grupos (IZENMAN, 2008):

- Treinamento: conjunto responsável pela formação do classificador, ou seja, faz parte da primeira etapa do processo. Com ele, o algoritmo aprende a relação entre os parâmetros que estão em sendo avaliados e as classes determinadas (LAROSE; LAROSE, 2014);

- Validação: deve-se a esse grupo a avaliação dos resultados iniciais e participação na definição do modelo ideal para a análise das informações;

- Teste: possui a função de fazer a avaliação final da eficiência das previsões do algoritmo, finalizando a construção do classificador.

Segundo Goodfellow, Bengio e Courville (2016), é suposto que os dados destes grupos são independentes entre si e os grupos de treinamento e teste são separados identicamente, ou seja, possuem a mesma distribuição probabilística. Em média, o primeiro conjunto sempre reune a maior quantidade de dados, enquanto os outros dois dividem igualmente o restante. Para garantir uma maior diversidade dos dados presentes no grupo de treinamento, também pode ser aplicado a técnica de validação cruzada ou *cross-validation* (STONE, 1974 e EFRON,

1979 apud IZENMAN, 2008), que examina de forma iterativa todas as divisões do banco de dados como grupo de teste.

A precisão desse processo é um dos principais fatores na avaliação de um desses algoritmos. No classificador, o erro de previsão é dado pela probabilidade de classificar erroneamente a classe de um dado (IZENMAN, 2008). Se este for maior que a classe de treinamento, temos um erro de treinamento ou aprendizado; se for menor, um erro de teste. Isso leva ao conceito de generalização, que é a capacidade de prever com precisão novos dados (BISHOP, 1995).

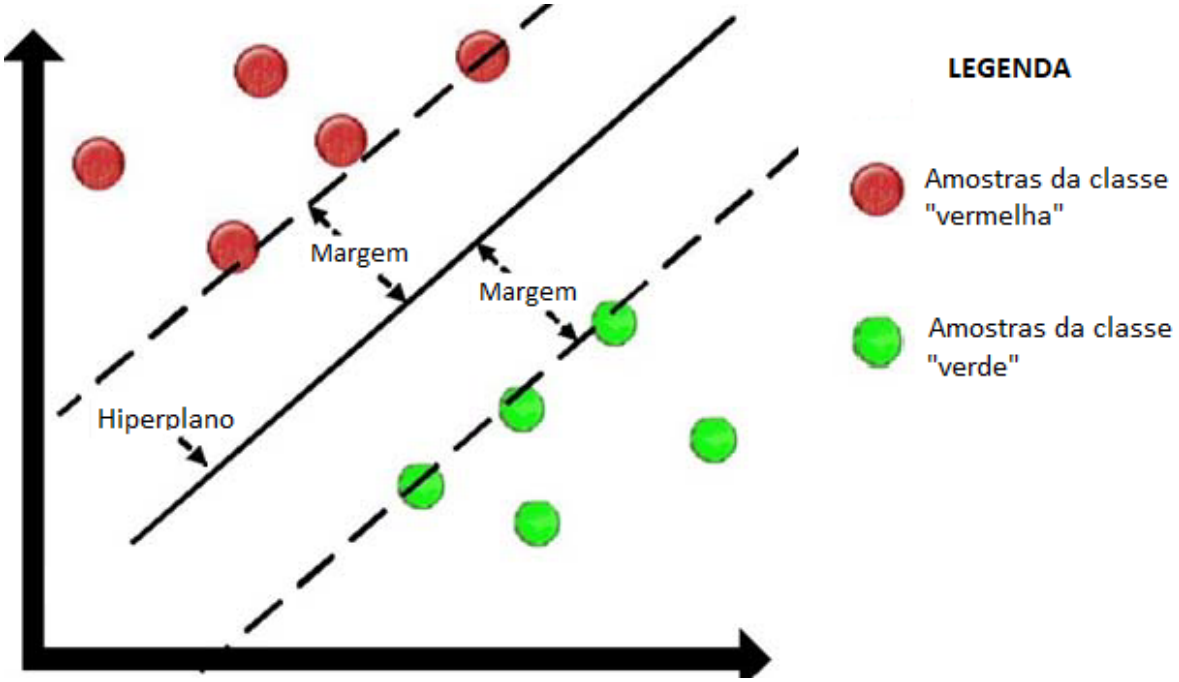
A escolha de um modelo também influencia diretamente no desenvolvimento de um método efetivo de classificação de dados. Um modelo muito generalista sofre de *underfitting*, não conseguindo produzir um valor baixo de erro no treinamento; já um critério muito específico aos dados leva ao *overfitting*, causando uma variação muito alta entre os erros de treinamento e teste (GOODFELLOW; BENGIO; COURVILLE, 2016). Para tentar evitar esses problemas, utiliza-se do princípio de Navalha de Ockham, que conduz a escolha do modelo mais simples possível (IZENMAN, 2008).

Neste trabalho, foram estudados e aplicados os modelos de Máquina de Vetores de Suporte e Redes Neurais Artificiais, quais são explicados com mais detalhes na sequência. A avaliação dos mesmos foi realizada através da implementação de uma matriz de confusão e o cálculo de medidas de desempenho, técnicas que também são detalhadas a seguir.

2.4.1 Máquina de Vetores de Suporte

Máquina de Vetores de Suporte (MVS), do inglês *Support Vector Machines* (SVM), é um método aprendizado supervisionado que separa os dados em duas classes através de uma fronteira (BOYLE, 2011), como pode ser exemplificado na Figura 9.

Figura 9 – Exemplo de separação de dados através de um vetor e suas margens máximas



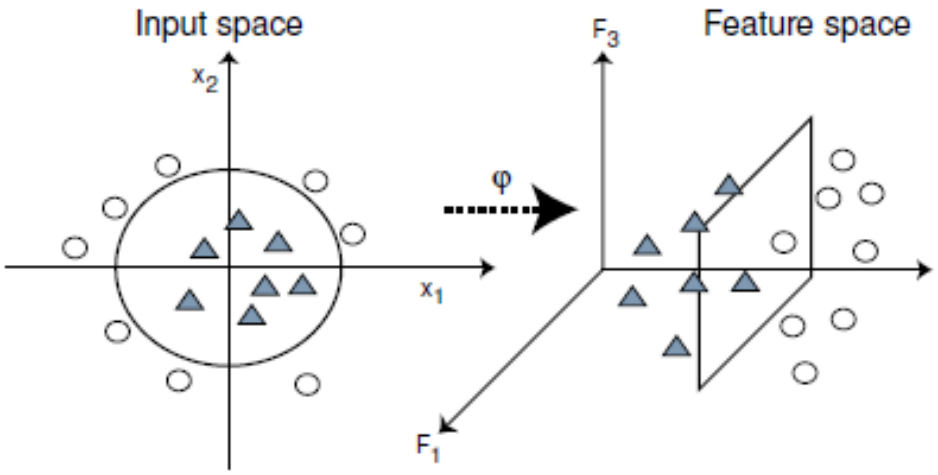
Fonte: Adaptado de BOYLE, 2011.

A metodologia foi criada especificamente para resolver problemas de classificações binárias - mas existem variações para o uso em múltiplas classes ou em situações de regressão - e é aplicada amplamente, variando de projetos em reconhecimento de caligrafia a até classificação de cânceres e previsão de estrutura secundária em proteínas, por exemplo (IZENMAN, 2008). De acordo com Boyle (2011), uma das vantagens do MVS é redução dos problemas de *overfitting*, devido a pouca flexibilidade do limite de decisão em relação aos dados.

O algoritmo tem como objetivo definir uma função que separe os dois grupos (BOYLE, 2011), entretanto, são várias as que se encaixam nessa condição (SODRÉ, 2016). Desta forma, a equação apropriada deve maximizar a distância entre o dado mais próximo em cada um dos lados da divisa (BOYLE, 2011), criando as margens que podem ser observadas também na Figura 9. Como cada elemento pode possuir várias variáveis (exemplo: *pitch*, *jitter* e *shimmer*, para um único sinal de voz) e ser apresentado vetorialmente, a fronteira passa a ser chamada de hiperplano, podendo este assumir infinitas dimensões (BOYLE, 2011).

Para casos não lineares, onde o hiperplano não consegue separar em duas classes de forma convencional, é necessário que seja feita transformações dos dados para que estes se encaixem no modelo. Isso é feito através dos métodos de *Kernel*, funções que levam os dados para um espaço multidimensional, permitindo a classificação a partir de uma equação linear (CRUZ, 2007 apud ALVES, 2016). Esse processo é representado pela Figura 10, que mostra o *input space* (espaço de entrada, em português), um espaço de duas dimensões com os dados de treinamento separados por uma função não linear, e o *feature space* (espaço de características, em português), que apresenta os dados após uma transformação φ classificados a partir de um hiperplano.

Figura 10 – Espaços de entrada e características



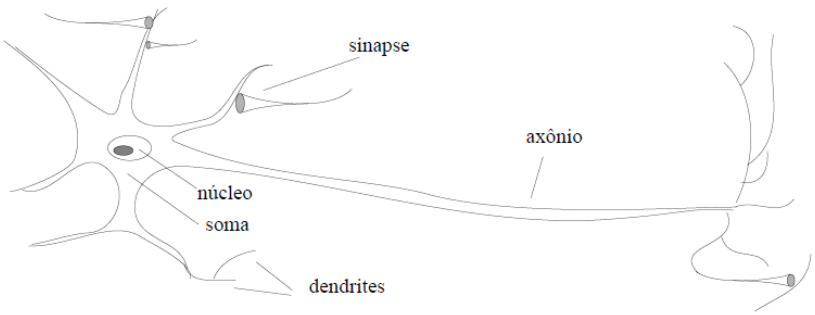
Fonte: BOYLE, 2011.

Neste trabalho foram testados diferentes *kernels*, disponibilizados pelo aplicativo *Classification Learner* do *software* MATLAB.

2.4.2 Redes Neurais Artificiais

Nós humanos aprendemos com base em eventos previamente vividos. De acordo com Haykin (2001), a habilidade de processamento, plasticidade e rápido desenvolvimento do órgão é baseada na experiência adquirida de eventos aos quais somos expostos. O uso de redes neurais artificiais (RNA) é baseado no cérebro humano e na forma em que este processa informações, diferente dos computadores tradicionais. O cérebro tem a capacidade de organizar sua estrutura - composta por células chamadas neurônios - de forma a processar mais rapidamente as informações, mais até que a maioria dos computadores. Representado pela Figura 11, um neurônio em desenvolvimento é uma estrutura muito flexível e capaz de se adaptar a diferentes situações e estímulos (HAYKIN, 2001).

Figura 11 – Estrutura de um neurônio biológico

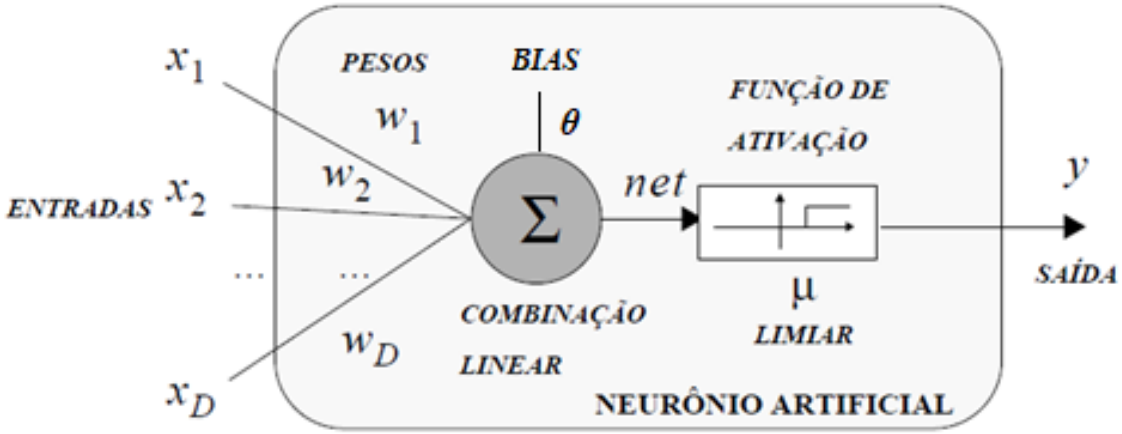


Fonte: RAUBER, 2005.

O neurônio biológico recebe sinais por conexões elétricas chamadas de sinapses, que realizam a transmissão de informação de um neurônio a outro. Há duas categorias de sinapses: a excitatória e a inibitória. A primeira favorece o disparo das informações de um neurônio a outro, enquanto que a segunda ocorre o inverso, inibe a transmissão do sinal. Essa diferenciação é feita através de pesos, sendo um valor positivo para a primeira categoria e negativo para a segunda (SODRÉ, 2016).

Um neurônio artificial possui a mesma lógica: a entrada recebe as informações, que passarão por um processamento e resultarão em uma informação de saída. A representação desse processo pode ser feita através do modelo de McCulloch e Pitts (RAUBER,2005 apud MCCULLOCH; PITTS, 1943), qual sua principal característica é a apresentação de uma lógica de aprendizagem capaz de adaptar os pesos das sinapses de modo que um problema de classificação seja resolvido, como representado na Figura 12.

Figura 12 – Modelo de neurônio artificial



Fonte: Adaptado de RAUBER, 2005.

Esse processo consiste da introdução de informações nas entradas x_1,x_2,...,x_D, que possuem pesos w_1,w_2,...,w_D e são processadas através de uma combinação linear, que pode ainda conter ou não um valor Θ chamado de *Bias*, cuja função é alterar o valor passado a diante no neurônio (SODRÉ, 2016). O resultado disso é um valor chamado *net* (Equação 10), que é então aplicado em uma função de ativação F (Equação 11).

$$net = \sum_{i=1}^{D} x_i w_i + \Theta \tag{10}$$

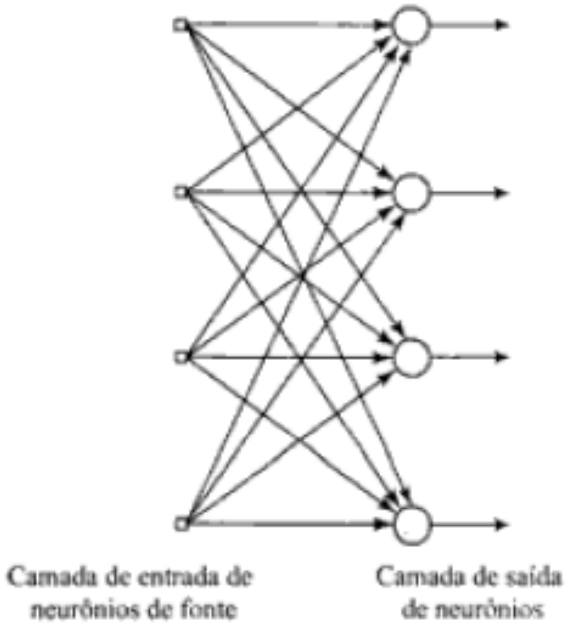
$$y = F(net) \tag{11}$$

Por fim, a classificação das informações introduzidas nas entradas do neurônio acontece através de um limiar μ (RAUBER, 2005). Se o resultado da saída ultrapassar esse valor, o neurônio dispara a sinapse, caso contrário não transmite a informação.

Haykin (2001), expõe que, para um bom desempenho, as redes neurais empregam a interligação maciça de células computacionais simples, podendo apresentar três tipos de topologias:

- Redes alimentadas diretamente ou *feedforward* com camada única: apresenta apenas uma camada composta por neurônios, que está ligada diretamente a entrada e a saída. Nessa arquitetura, representada pela Figura 13, a informação é passada sempre adiante.

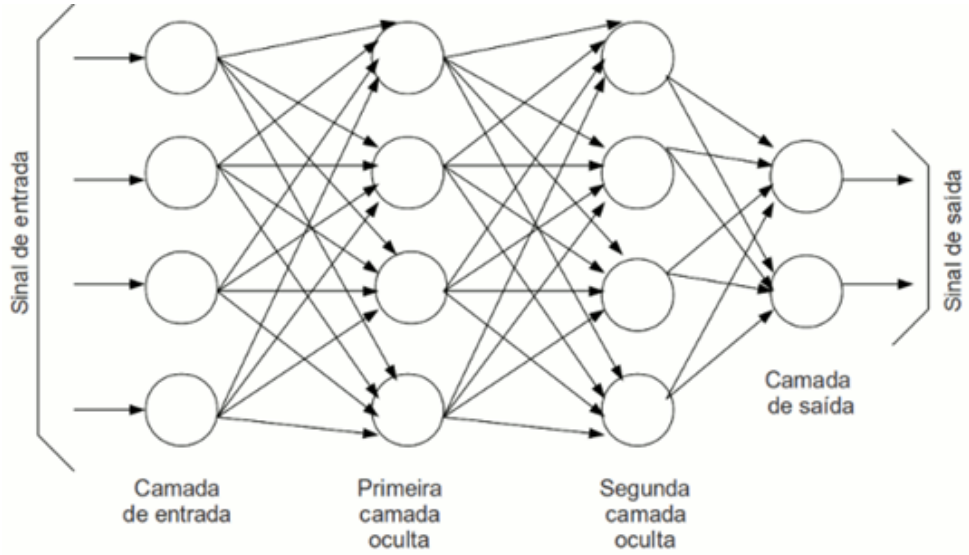
Figura 13 – Exemplo de rede *feedforward* com camada única



Fonte: HAYKIN, 2001.

- Redes alimentadas diretamente ou *feedforward* com múltiplas camadas: possui a mesma lógica de propagação adiante da topologia anterior, mas apresenta duas ou mais camadas de neurônios interligadas (Figura 14), tornando a rede capaz de obter informações importantes de forma mais global, por conta das maiores interações neurais (HAYKIN, 2001 apud CHURCHLAND; SEJNOWSKI, 1992). As camadas que não possuem ligação com as entradas ou saídas são chamadas de camadas ocultas ou *hidden layers* (RAUBER, 2005).

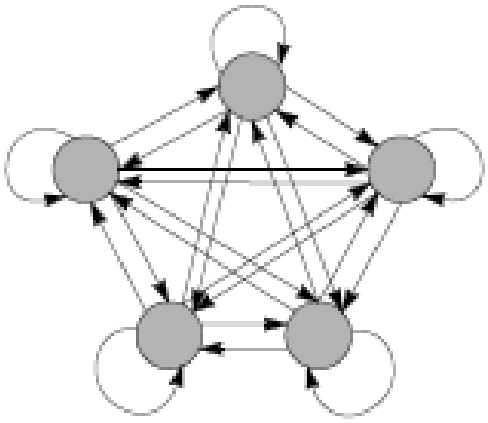
Figura 14 – Exemplo de rede *feedforward* com múltiplas camadas



Fonte: Monolito Nimbus[3].

- Redes recorrentes ou realimentadas: apresenta ao menos uma realimentação entre as camadas, repassando a saída de um neurônio para sua própria entrada, o que resulta em um impacto considerável no aprendizado da rede (HAYKIN, 2001). A Figura 15 mostra um exemplo dessa topologia.

Figura 15 – Exemplo de rede recorrente ou realimentada



Fonte: RAUBER, 2005.

A topologia, número de neurônios e camadas variaram de acordo as necessidades de cada rede, enquanto o número de entradas representam as variáveis da amostra analisada - no caso deste trabalho, são as medidas acústicas - e as saídas as classes existentes no problema aplicado. Dessa forma, a RNA tem a predisposição de armazenar o conhecimento, obtido das experiências as quais foi exposta, e de torná-lo disponível para uso. Esses fatores tornam possível a resolução de problemas de grande complexidade.

Por fim, o aprendizado da rede acontece a partir da comparação do resultado de saída com o resultado desejado, atualizando-se os pesos a partir da diferença desses dois valores (ROSA, 1998). Isso ocorre através da minimização do erro quadrático médio (Equação 12) pelo método da descida do gradiente (Equação 13) (RAUBER, 2005):

$$E = \frac{1}{n}\sum_{i}^{n}(t_i - y_i)^2 \tag{12}$$

$$w^{(l+1)} = w^{(l)} - \eta \nabla E^{(l)} \tag{13}$$

Onde:

- n: número de amostras aplicadas na rede neural;
- t_i: valor desejado de saída;
- y_i: valor de saída;
- $w^{(l+1)}$: peso corrigido na iteração $l + 1$;
- $w^{(l)}$: peso na iteração l;
- η: taxa de aprendizagem;
- $\nabla E^{(l)}$: gradiente do erro quadrático médio na iteração l.

As iterações, também chamadas de épocas, irão parar quando o erro quadrático médio for menor do que o desejado ou um número máximo iterações seja definido.

A taxa de aprendizagem serve para controlar a velocidade de alteração do valor do peso a partir do escalar η (RAUBER, 2005). Se o valor for alto demais, o algoritmo poderá ter dificuldades para atingir a limiar; se for muito baixo, o treinamento exigirá muitas épocas para ser concluído.

Neste trabalho foi utilizada a arquitetura *feedforward* com múltiplas camadas, analisando os resultados para diferentes números de camadas, neurônios e funções de ativação.

2.4.3 Matriz de confusão e medidas de desempenho

Matriz de confusão é um método avaliativo utilizado no aprendizado de máquina (seja qual for o modelo) para compreender o comportamento do algoritmo e definir sua capacidade de previsão. Ela é formada por uma tabela, que relaciona o resultado previsto pelo classificador com verdadeira condição da amostra (SODRÉ, 2016), como pode ser visto na Tabela 1.

[3] Disponível em: <https://www.monolitonimbus.com.br/redes-neurais-artificiais/> Acesso em: 18 de agosto de 2021.

Tabela 1 – Matriz de confusão

Matriz de confusão		Resultado Real	
		VERDADEIRO	**FALSO**
Resultado Previsto	**VERDADEIRO**	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	FALSO	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Autoria Própria.

Essa tabela permite rotular os resultados obtidos em quatro categorias (LAROSE; LAROSE, 2014; SODRÉ, 2016):

- Verdadeiro Positivo (VP): são as classificações previstas corretamente como verdadeiras;
- Falso Positivo (FP): são as classificações previstas erroneamente como verdadeiras, mas que na realidade são falsas;
- Falso Negativo (FN): são as classificações previstas erroneamente como falsas, mas que na realidade são verdadeiras;
- Verdadeiro Negativo (VN): são as classificações previstas corretamente como falsas.

A partir dos dados fornecidos por ela é possível calcular medidas de desempenho, que quantificam os resultados do classificador. Para este trabalho, decidiu-se por utilizar as seguintes medidas, todas elas dadas em porcentagem (SODRÉ, 2016):

- Acurácia: demonstra a capacidade do modelo em categorizar uma amostra em sua classe correta. Isso resulta na parcela de verdadeiros positivos e verdadeiros negativos em relação ao total, dada pela equação 14:

$$Acurácia = 100 * \frac{VP + VN}{VP + FP + FN + VN} \tag{14}$$

- Precisão: define a capacidade de classificar uma amostra como verdadeiro positivo entre todos os resultados positivos, representada pela equação 15:

$$Precisão = 100 * \frac{VP}{VP + FP} \tag{15}$$

- Sensibilidade: define a capacidade de classificar uma amostra como verdadeiro positivo entre os resultados verdadeiramente positivos e falsamente negativos, ou seja, mostra as chances da avaliação atestar corretamente a presença de uma patologia quando se possui uma doença. Essa medida é representada pela equação 16:

$$Sensibilidade = 100 * \frac{VP}{VP + FN} \tag{16}$$

- Especificidade: representada pela equação 17, define a capacidade de classificar uma amostra como verdadeiro negativo entre os resultados verdadeiramente negativos e falsamente positivos. No caso, o parâmetro nos mostra a probabilidade do resultado dar corretamente negativo quando não se possui uma doença.

$$Especificidade = 100 * \frac{VN}{VN + FP} \tag{17}$$

É importante destacar que uma matriz de confusão também pode assumir mais de duas classes, como acontece na tentativa de identificar qual é a doença presente nas amostras confirmadas como patológicas, por exemplo. Para esse caso, optou-se por fazer uma análise um contra todos, onde a classe a ser examinada é considerada como verdadeira e todas as outras como falso, reduzindo a dimensão da matriz para dois e tornando possível a definição dos parâmetros VP, FP, VN e FN, que agora estarão aptos a serem aplicados nas mesmas equações já descritas acima.

Por outro lado, as Equações 14, 15, 16 e 17 são aplicadas para cada classe e não representam o classificador como um todo. Dessa forma, é necessário a emprego de alguma técnica especial para obter uma medida única. Nesta pesquisa foi determinado o uso da metodologia *macro-averaging*, que calcula uma média aritmética das medidas obtidas em cada classe (SOKOLOVA; LAPALME, 2009).

3 RESULTADOS E DISCUSSÕES

Devido a existência de várias características que constituem um algoritmo classificador - todas podendo influenciar profundamente no resultado final -, a construção de um modelo ótimo é um processo que exige a comparação de diversas configurações diferentes. Dessa forma, são apresentados nesse capítulo os resultados obtidos dos classificadores Máquina de Vetores de Suporte e Redes Neurais Artificiais em diferentes cenários, tanto na discriminação de uma voz saudável ou patológica quanto na tentativa de identificar qual é a doença presente, nos casos conhecidamente positivos.

Inicialmente, o tamanho dos grupos de treinamento e teste - as funções utilizadas no *software* MATLAB reúnem o processo de validação junto do treinamento - para os dois modelos avaliados foi definido em conjunto com a técnica de *cross-validation*. Ensaiou-se divisões de quatro e cinco grupos no banco de dados, que nos levam a uma separação de 75% das amostras para treinamento e 25% para teste no primeiro caso e 80% para treinamento e 20% para teste no segundo.

Para o método de Máquina de Vetores de Suporte em específico, foi testado o impacto de seis diferentes *kernels*, presentes na biblioteca do aplicativo *Classification Learner*, do *software* MATLAB: Linear, Quadrático, Cúbico, Gaussiano Grosseiro, Gaussiano Médio e Gaussiano Fino.

Já a variação de características em Redes Neurais Artificiais ocorreu em várias frentes:

- Camadas ocultas: um número de camadas elevado pode acentuar os erros em vez de diminuí-los. Assim, considerando que a quantidade de dados a serem processados é relativamente pequena, optou-se por usar poucas camadas escondidas, testando apenas dois valores para essa característica das redes: uma ou duas camadas ocultas.

- Neurônios: a quantidade de neurônios por camada oculta foi calculada a partir da média do número de entradas e saídas, chamada de N. Com o objetivo de analisar o impacto que essa variável causaria, foram testados três cenários: $0{,}5N$, N e $2N$ neurônios.

  Sabendo da existência de 13 entradas - cada uma constituída por uma medida acústica -, duas saídas para a classificação entre voz normal e patológica e oito saídas para a identificação das patologias, obteve-se os resultados da Tabela 2. Para a definição do número de neurônios, aplicou-se um arredondamento sempre para cima.

Tabela 2 – Número de neurônios para Redes Neurais Artificiais

Número de neurônios				
Tipo da classificação	**Condição da voz**		**Identificação de patologias**	
Valores	**Real**	**Arredondado**	**Real**	**Arredondado**
0.5N	3,75	4	5,25	6
N	7,5	8	10,5	11
2N	15	15	21	21

Fonte: Autoria Própria.

- Função de ativação: com base nas funções disponíveis no MATLAB, foram selecionadas duas que se adequam melhor ao tipo de dados que foram apresentados: ReLu e TanH (Equações 18 e 19, respectivamente). Elas foram aplicadas nas células das camadas ocultas, enquanto a camada de saída utilizou-se a função Softmax (Equação 20), qual é padronizada pela função de treinamento do software.

$$ReLu = f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{18}$$

$$TanH = f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{19}$$

$$Softmax = f(x) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}, i = 1, ..., J \tag{20}$$

Todos esses dados foram reunidos em tabelas, contendo as configurações dos arranjos, os parâmetros obtidos através da matriz de confusão e as medidas de desempenho, e são apresentados na sequência.

Por fim, também é importante explanar as características banco de dados utilizado - Modelo 4337, da KayPENTAX Corp (Kay Elemetrics Corp, Lincoln Park, NJ) -. O mesmo foi filtrado para reduzir classes com poucas amostras e possui um total de 730 vozes, que podem ser separadas por sua condição (normal ou patológica) e, nos casos positivos, de acordo com a doença presente. As Tabelas 3 e 4 descrevem a frequência dessas classes.

Tabela 3 – Número de amostras - Separadas pela condição da voz

	Número de amostras
Normais	53
Patológicas	677
Total	730

Tabela 4 – Número de amostras - Separadas por patologias

	Número de amostras
AP Squeezing	153
Ceratose - Leucoplasia	24
Compressão ventricular	100
Edema das pregas vocais	39
Edema de Reinke	22
Hiperfunção	241
Paralisia	53
Refluxo Laringofaríngeo	45
Total	677

3.1 Classificação da condição da voz em amostras normais ou patológicas

Inicialmente, foi projetado o classificador para identificar a condição das amostras de voz disponíveis, ou seja, se a mesma é saudável ou indica a existência de uma patologia na laringe.

Através de um script no software MATLAB, testou-se todas as características já descritas para os dois modelos de classificação, que receberam ambos os dados da tabela já descritos pela Tabela 3. Os resultados foram apresentados nos tópicos a seguir.

3.1.1 Resultados para Máquina de Vetores de Suporte

A Tabela 5 mostra as configurações de cada teste utilizando o método de Máquinas de Vetores de Suporte. Os resultados são apresentados pela Tabela 6.

Tabela 5 – Configurações para classificação da condição da voz com MVS

Identificação do teste	Cross-Validation	Divisão dos grupos		Kernel
		Treinamento (%)	Teste (%)	
1	4	75	25	Linear
2	4	75	25	Quadrático
3	4	75	25	Cúbico
4	4	75	25	Gaussiano Grosseiro
5	4	75	25	Gaussiano Médio
6	4	75	25	Gaussiano Fino
7	5	80	20	Linear
8	5	80	20	Quadrático
9	5	80	20	Cúbico
10	5	80	20	Gaussiano Grosseiro
11	5	80	20	Gaussiano Médio
12	5	80	20	Gaussiano Fino

Tabela 6 – Resultados para classificação da condição da voz com MVS

Identificação do teste	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)
1	94,7945	95,9712	98,5229	47,1698
2	94,7945	95,7082	98,8183	43,3962
3	95,4795	97,2141	97,9321	64,1509
4	92,7397	92,7397	100	0
5	95,0685	95,5903	99,2614	41,5094
6	95,4795	96	99,2614	47,1698
7	94,9315	96,1095	98,5229	49,0566
8	95,0685	95,8512	98,966	45,283
9	96,1644	97,0972	98,8183	62,2642
10	92,7397	92,7397	100	0
11	94,7945	95,5777	98,966	41,5094
12	95,2055	95,7265	99,2614	43,3962

Como se pode observar, a acurácia ultrapassa os 90% para todos os *kernels* testados. Esses valores podem ser considerados satisfatórios, mas não necessariamente indicam um bom resultado do classificador, especialmente quando se avalia um banco de dados desba-

lanceado, ou seja, que possui uma grande diferença no número de amostras em cada classe. Assim, faz-se necessário analisar em conjunto as outras medidas de desempenho.

A precisão também segue a acurácia e apresentou valores acima de 90%, indicando uma boa taxa de acerto nos casos em que o classificador definiu como positivo, isto é, patológicos. Por outro lado, é possível que este parâmetro também tenha sido fortemente influenciado pela grande quantidade de positivos no banco de dados.

Os dados de sensibilidade apresentam valores ainda mais altos, com o menor sendo 97,93%. Isto confirma que, ao menos nas condições deste teste, o modelo MVS teve um bom desempenho na identificação da existência de uma patologia em amostras que realmente possuem uma doença. Por considerar os falsos negativos em seu cálculo, também mostra que a chance de classificar uma amostra patológica como saudável é baixa.

Por fim, a especificidade mostra a eficiência em classificar uma amostra saudável quando não se possui a doença. Analisando os resultados, é possível perceber uma maior dificuldade do modelo com essa classificação, comprometendo o desempenho do mesmo quando se trata da classe normal. Isto pode ser consequência da pequena quantidade de casos normais.

Em relação as diferentes configurações testadas, a divisão da validação cruzada em quatro ou cinco grupos fez pouca diferença nos resultados de cada medida de desempenho, não indicando um valor ótimo para a avaliação, como mostra a Tabela 7.

Tabela 7 – Média dos resultados para diferentes valores de *cross-validation* - Classificação da condição da voz com MVS

	Média dos resultados			
Cross-Validation	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
4	94,7260	95,5373	98,9660	40,5660
5	94,8174	95,5170	99,0891	40,2516

Quanto aos *kernels*, a variação nos resultados também foi muito pequena (Tabela 8), tendo o Gaussiano Grosseiro e Cúbico como as únicas excessões. O primeiro apresentou uma sensibilidade de 100% e uma especificidade de 0% para ambos os testes de *cross-validation*; isso aconteceu por não conseguir classificar uma única amostra como normal, inviabilizando seu uso para esse tipo de classificação no atual estado. Já o segundo conseguiu diferenciar-se com um resultado um pouco superior em especificidade, tornando-o o mais indicado para utilização a partir das características desse teste.

Tabela 8 – Média dos resultados para diferentes *kernels* - Classificação da condição da voz com MVS

	Média dos resultados			
Kernel	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
Linear	94,863	96,04035	98,5229	48,1132
Quadrático	94,9315	95,7797	98,89215	44,3396
Cúbico	95,82195	97,15565	98,3752	63,20755
Gaussiano Grosseiro	92,7397	92,7397	100	0
Gaussiano Médio	94,9315	95,584	99,1137	41,5094
Gaussiano Fino	95,3425	95,86325	99,2614	45,283

3.1.2 Resultados para Redes Neurais Artificiais

A Tabela 9 mostra as configurações de cada teste utilizando o método de Redes Neurais Artificiais. Os resultados são apresentados pela Tabela 10.

Tabela 9 – Configurações para classificação da condição da voz com RNA

Identificação do teste	Cross-Validation	Divisão dos grupos		Camadas	Neurônios	Função de Ativação	
		Treinamento (%)	Teste (%)			Camadas ocultas	Camada de saída
13	4	75	25	1	4	ReLu	Softmax
14	4	75	25	1	4	TanH	Softmax
15	4	75	25	1	8	ReLu	Softmax
16	4	75	25	1	8	TanH	Softmax
17	4	75	25	1	15	ReLu	Softmax
18	4	75	25	1	15	TanH	Softmax
19	4	75	25	2	4	ReLu	Softmax
20	4	75	25	2	4	TanH	Softmax
21	4	75	25	2	8	ReLu	Softmax
22	4	75	25	2	8	TanH	Softmax
23	4	75	25	2	15	ReLu	Softmax
24	4	75	25	2	15	TanH	Softmax
25	5	80	20	1	4	ReLu	Softmax
26	5	80	20	1	4	TanH	Softmax
27	5	80	20	1	8	ReLu	Softmax
28	5	80	20	1	8	TanH	Softmax
29	5	80	20	1	15	ReLu	Softmax
30	5	80	20	1	15	TanH	Softmax
31	5	80	20	2	4	ReLu	Softmax
32	5	80	20	2	4	TanH	Softmax
33	5	80	20	2	8	ReLu	Softmax
34	5	80	20	2	8	TanH	Softmax
35	5	80	20	2	15	ReLu	Softmax
36	5	80	20	2	15	TanH	Softmax

Tabela 10 – Resultados para classificação da condição da voz com RNA

Identificação do teste	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)
13	95,4795	96,9388	98,2275	60,3774
14	96,1644	96,4235	99,5569	52,8302
15	96,4384	97,2424	98,966	64,1509
16	97,2603	97,9562	99,1137	73,5849
17	96,1644	97,3723	98,5229	66,0377
18	96,8493	97,2543	99,4092	64,1509
19	95,8904	97,3646	98,2275	66,0377
20	96,0274	96,4183	99,4092	52,8302
21	95,4795	96,6667	98,5229	56,6038
22	95,6164	96,4029	98,966	52,8302
23	96,1644	96,9609	98,966	60,3774
24	96,5753	97,3837	98,966	66,0377
25	96,1644	97,0972	98,8183	62,2642
26	95,3425	97,7712	97,1935	71,6981
27	96,8493	98,0882	98,5229	75,4717
28	96,7123	97,9442	98,5229	73,5849
29	96,7123	97,9442	98,5229	73,5849
30	95,6164	96,807	98,5229	58,4906
31	96,0274	97,6471	98,0798	69,8113
32	96,4384	97,6574	98,5229	69,8113
33	96,3014	97,654	98,3752	69,8113
34	95,7534	96,8116	98,6706	58,4906
35	97,1233	98,3776	98,5229	79,2453
36	96,0274	97,2303	98,5229	64,1509

Assim como modelo de MVS, os testes com RNA apresentaram valores altos e com pouca variação entre si para os parâmetros acurácia, precisão e sensibilidade, independentemente das configurações citadas. Dessa forma, também é possível concluir que o modelo teve um bom desempenho em classificar corretamente amostras como patológicas, ainda que o desiquilíbrio do banco de dados possa ter influenciado nos resultados.

A quarta medida acaba nos dando uma visão melhor da real capacidade do classificador, pois avalia o desempenho para os casos negativos. Comparado ao modelo anterior, as redes neurais testadas conseguiram um resultado melhor em especificidade, variando de 52,0302% (testes 14 e 22) e a 79,2453% (teste 35). Essa melhora deve-se a característica do método de classificação, que - por conta das camadas e neurônios - consegue criar um hiperplano de separação de dados mais flexível, adaptando-se com mais facilidade às características dos dados.

Quanto ao desempenho das configurações em si, notou-se apenas diferenças nos valores de especificidade entre características diferentes. Nas Tabelas 11, 12, 13 e 14, pode-se ver uma melhora nesse quesito aumentando a quantidade de divisões no banco de dados, utilizando apenas uma camada, dobrando a quantidade N de neurônios e implementando células com a função de ativação ReLu, respectivamente. Ainda assim, as variações são mínimas e nenhuma característica é consideravelmente superior que as outras.

Tabela 11 – Média dos resultados para diferentes valores de *cross-validation* - Classificação da condição da voz com RNA

	Média dos resultados			
Cross-Validation	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
4	96,1758	97,0321	98,9045	61,3208
5	96,2557	97,5858	98,3998	68,8679

Tabela 12 – Média dos resultados para diferentes quantidades de camadas - Classificação da condição da voz com RNA

	Média dos resultados			
Camadas	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
1	96,3128	97,4033	98,6583	66,3522
2	96,1187	97,2146	98,6460	63,8365

Tabela 13 – Média dos resultados para diferentes quantidades de neurônios - Classificação da condição da voz com RNA

	Média dos resultados			
Número de neurônios	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
4	95,9418	97,1648	98,5045	63,2076
8	96,3014	97,3458	98,7075	65,5660
15	96,4041	97,4163	98,7445	66,5094

Tabela 14 – Média dos resultados para diferentes funções de ativação - Classificação da condição da voz com RNA

	Média dos resultados			
Função de ativação - Camada Oculta	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
ReLu	96,2329	97,4462	98,5229	66,9811
TanH	96,1986	97,1717	98,7814	63,2075

3.2 Classificação de doenças em amostras patológicas

Na segunda etapa da pesquisa, ajustou-se os modelos de aprendizado de máquina para testar a sua capacidade de identificar qual doença está presente nas amostras previamente diagnosticadas como patológicas, alterando então a classificação de duas para múltiplas classes.

Os dados introduzidos em ambos os métodos de classificação foram descritos pela Tabela 4 e os resultados foram apresentados nos tópicos a seguir.

3.2.1 Resultados para Máquina de Vetores de Suporte

A Tabela 15 mostra as configurações de cada teste utilizando o método de Máquinas de Vetores de Suporte. Os resultados são apresentados pela Tabela 16.

Tabela 15 – Configurações para classificação de doenças com MVS

Identificação do teste	Cross-Validation	Divisão dos grupos		Kernel
		Treinamento (%)	Teste (%)	
37	4	75	25	Linear
38	4	75	25	Quadrático
39	4	75	25	Cúbico
40	4	75	25	Gaussiano Grosseiro
41	4	75	25	Gaussiano Médio
42	4	75	25	Gaussiano Fino
43	5	80	20	Linear
44	5	80	20	Quadrático
45	5	80	20	Cúbico
46	5	80	20	Gaussiano Grosseiro
47	5	80	20	Gaussiano Médio
48	5	80	20	Gaussiano Fino

Tabela 16 – Resultados para classificação de doenças com MVS

Identificação do teste	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)
37	83,8257	4,4391	12,3963	87,4619
38	82,2009	5,9804	10,3578	86,7816
39	77,548	3,9911	4,0399	84,8212
40	80,4284	3,636	7,6841	85,5723
41	83,6411	4,3985	12,1369	87,3768
42	83,8996	4,4498	12,5	87,5
43	83,8996	4,4498	12,5	87,5
44	82,3117	5,387	10,5134	86,8109
45	78,3973	4,1891	5,5441	85,249
46	80,7238	3,491	8,0394	85,6525
47	83,6411	6,172	12,1668	87,3751
48	83,8996	4,4498	12,5	87,5

De início, observa-se que a grande maioria dos testes de acurácia média apresentam valores acima de 80%, com apenas dois cenários – ambos com o *kernel* cúbico – falhando em atingir esse limiar. Por ser uma análise mais complexa, devido a maior quantidade de classes existentes, todos são valores satisfatórios, mas que também têm sua importância reduzida devido ao mesmo problema da análise com duas classes: o desbalanceamento do banco de dados. Como indica a Tabela 4, algumas doenças possuem uma frequência de amostras muito maior que outras, inviabilizando a conclusão do desempenho do classificador apenas pela acurácia.

A necessidade de analisar outras medidas de desempenho além da acurácia fica evidente quando se observa os resultados de precisão e sensibilidade média. Ambos os parâmetros apresentam resultados muito baixos, evidenciando uma grande dificuldade do modelo em identificar corretamente todas as doenças.

Já a especificidade média segue a linha da acurácia com resultados altos, mas que não representam corretamente a capacidade do modelo em diagnosticar uma patologia com eficiência. A análise multiclasse obriga o uso de uma técnica para obter resultados que representem o classificador como um todo – no caso, foi escolhida a técnica de *macro-averaging* -. Nela,

cada classe é estudada individualmente, tornando-a positiva e o restante negativo; assim, toda previsão que corretamente não pertence a esta classe positiva é considerada como verdadeiro negativo, independentemente dessa classificação em si estar certa ou não. Isso eleva os valores de verdadeiro negativo e nos leva ao caso observado: resultados altos para especificidade, mas que não indicam um bom desempenho do classificador.

Essa disparidade entre os valores altos de acurácia e especificidade e baixos de precisão e sensibilidade deixam claro o impacto negativo de um banco de dados desbalanceado. Na classificação, independentemente da configuração de validação cruzada e *kernel*, a grande quantidade de amostras de hiperfunção fez com que a maioria das previsões fossem da doença, arruinando o desempenho do método na identificação das outras patologias. Também é válido considerar o impacto das medidas acústicas, que podem não ter sido capazes de diferenciar de forma mais eficiente uma doença da outra.

Em relação aos cenários com diferentes valores de *cross-validation* e *kernels*, novamente notou-se pouca diferença entre os resultados, como atesta as Tabelas 17 e 18. É possível apenas destacar os *kernels* Cúbico e Gaussiano Grosseiro, que apresentaram valores de sensibilidade abaixo da média do restante, o que não é favorável ao desempenho do classificador.

Tabela 17 – Média dos resultados para diferentes valores de *cross-validation* - Classificação de doenças com MVS

	Média dos resultados			
Cross-Validation	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
4	81,9240	4,4825	9,8525	86,5856
5	82,1455	4,6898	10,2106	86,6813

Tabela 18 – Média dos resultados para diferentes *kernels* - Classificação de doenças com MVS

	Média dos resultados			
Kernel	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
Linear	83,86265	4,44445	12,44815	87,48095
Quadrático	82,2563	5,6837	10,4356	86,79625
Cúbico	77,97265	4,0901	4,792	85,0351
Gaussiano Grosseiro	80,5761	3,5635	7,86175	85,6124
Gaussiano Médio	83,6411	5,28525	12,15185	87,37595
Gaussiano Fino	83,8996	4,4498	12,5	87,5

3.2.2 Resultados para Redes Neurais Artificiais

A Tabela 19 mostra as configurações de cada teste utilizando o método de Redes Neurais Artificiais. Os resultados são apresentados pela Tabela 20.

Tabela 19 – Configurações para classificação de doenças com RNA

Identificação do teste	Cross-Validation	Divisão dos grupos		Camadas	Neurônios	Função de Ativação	
		Treinamento (%)	Teste (%)			Camadas ocultas	Camada de saída
49	4	75	25	1	6	ReLu	Softmax
50	4	75	25	1	6	TanH	Softmax
51	4	75	25	1	11	ReLu	Softmax
52	4	75	25	1	11	TanH	Softmax
53	4	75	25	1	21	ReLu	Softmax
54	4	75	25	1	21	TanH	Softmax
55	4	75	25	2	6	ReLu	Softmax
56	4	75	25	2	6	TanH	Softmax
57	4	75	25	2	11	ReLu	Softmax
58	4	75	25	2	11	TanH	Softmax
59	4	75	25	2	21	ReLu	Softmax
60	4	75	25	2	21	TanH	Softmax
61	5	80	20	1	6	ReLu	Softmax
62	5	80	20	1	6	TanH	Softmax
63	5	80	20	1	11	ReLu	Softmax
64	5	80	20	1	11	TanH	Softmax
65	5	80	20	1	21	ReLu	Softmax
66	5	80	20	1	21	TanH	Softmax
67	5	80	20	2	6	ReLu	Softmax
68	5	80	20	2	6	TanH	Softmax
69	5	80	20	2	11	ReLu	Softmax
70	5	80	20	2	11	TanH	Softmax
71	5	80	20	2	21	ReLu	Softmax
72	5	80	20	2	21	TanH	Softmax

Tabela 20 – Resultados para classificação de doenças com RNA

Identificação do teste	Acurácia (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)
49	82,0901	9,1027	11,005	86,9544
50	82,0162	9,1189	11,2471	87,0909
51	81,2777	11,1569	12,0483	86,9342
52	80,1329	7,8224	9,8375	86,606
53	78,3973	5,6649	6,4489	85,8176
54	77,3264	3,6045	3,7201	85,3031
55	82,9025	10,412	12,0586	87,2799
56	80,6499	7,5948	9,5132	86,5829
57	79,616	6,0269	7,4751	86,1264
58	78,1758	6,9744	7,2406	85,7483
59	76,6987	3,1298	3,0175	84,9468
60	76,5879	3,6685	3,8475	84,7823
61	82,4963	9,9989	11,5199	87,2799
62	82,0901	8,6115	11,1257	87,0528
63	80,6499	6,3946	9,3717	86,488
64	80,9084	9,1249	10,579	86,7907
65	78,065	4,9575	5,6925	85,5848
66	77,8065	5,3416	5,2946	85,5403
67	82,6071	11,434	12,9528	87,4195
68	81,3146	10,9761	11,2084	86,8884
69	79,3575	4,9284	6,7539	85,8529
70	78,9513	6,9248	7,5793	86,1025
71	76,514	4,4284	4,6217	84,7963
72	76,3663	2,5749	2,5209	84,7502

Ainda que as redes neurais apresentem uma mais flexibilidade para classificações com múltiplas classes, o desempenho das redes neurais testadas para a identificação de doenças manifestou tendências semelhantes ao do modelo MVS, exibindo valores altos de acurácia e especificidade média e o oposto para precisão e sensibilidade média.

Da mesma forma, as conclusões também são análogas: para as duas primeiras medidas citadas, os valores são inflados pelo desbalanceamento das classes - quase a totalidade do banco de dados utilizado foi classificada pela rede neural como hiperfunção -; já as duas últimas medidas mostram a realidade do modelo, que acabou sofrendo grande influência das amostras de hiperfunção e não conseguiu fazer um ajuste dos pesos para identificar corretamente as doenças de menor frequência, como ceratose-leucoplasia, edema das cordas vocais, edema de Reinke, paralisia e refluxo gástrico. Ainda, também devemos novamente considerar a possibilidade de que as medidas acústicas coletadas não tenham sido suficientemente relevantes para um ajuste mais preciso da RNA.

Focando nas configurações, constata-se resultados símeis para diferentes valores de cross-validation (Tabela 21), não indicando alguma prevalência entre quatro ou cinco divisões no banco de dados. Já para quantidade de camadas (Tabela 22), neurônios (Tabela 23) e função de ativação usada (Tabela 24) é possível notar uma pequena superioridade nas configurações com uma camada, seis neurônios e que utilizam a função ReLu.

Tabela 21 – Média dos resultados para diferentes valores de *cross-validation* - Classificação de doenças com RNA

	Média dos resultados			
Cross-Validation	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
4	79,6560	7,0231	8,1216	86,1811
5	79,7606	7,1413	8,2684	86,2122

Tabela 22 – Média dos resultados para diferentes quantidades de camadas - Classificação de doenças com RNA

	Média dos resultados			
Camadas	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
1	80,2714	7,5749	8,9909	86,4536
2	79,1451	6,5894	7,3991	85,9397

Tabela 23 – Média dos resultados para diferentes quantidades de neurônios - Classificação de doenças com RNA

	Média dos resultados			
Número de neurônios	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
6	82,0209	9,6561	11,3288	87,0686
11	79,8837	7,4192	8,8607	86,3311
21	77,2203	4,1713	4,3955	85,1902

Tabela 24 – Média dos resultados para diferentes funções de ativação - Classificação de doenças com RNA

	Média dos resultados			
Função de ativação da camada oculta	**Acurácia**	**Precisão**	**Sensiblidade**	**Especificidade**
ReLu	80,0560	7,3029	8,5805	86,2901
TanH	79,3605	6,8614	7,8095	86,1032

4 CONCLUSÃO

Motivados pelos obstáculos enfrentados pelos profissionais da saúde e desconforto para os pacientes durante o diagnóstico padrão da laringe, esta pesquisa buscou - através da análise do sinal da fala utilizando inteligência artificial - avaliar a viabilidade e confiabilidade da aplicação de um método alternativo para identificar a condição da laringe e distinguir qual é a enfermidade presente nos casos confirmados como patológicos. Para o primeiro caso, onde buscou-se avaliar uma amostra de fonema sustentado como saudável ou patológica, os resultados mostraram um bom desempenho na classificação, com eficiência nas previsões patológicas e poucos falsos negativos. Essa conclusão pode ser empregue tanto para Máquina de Vetores de Suporte quanto para Redes Neurais Artificiais, ainda que as seguintes diferenças tenham sido observadas entre dos dois métodos: o segundo apresentou uma performance média levemente melhor nas medidas acurácia e precisão e consideravelmente acima em sensibilidade, enquanto o primeiro se saiu superior por apenas 0,3754% em sensibilidade, como mostra a Tabela 25.

Tabela 25 – Média dos resultados para MVS e RNA - Classificação da condição da voz

	Média dos resultados			
Classificador	**Acurácia**	**Precisão**	**Sensibilidade**	**Especificidade**
Máquinas de Vetores de Suporte	94,7717	95,5271	99,0276	40,4088
Redes Neurais Artificiais	96,2158	97,3089	98,6521	65,0943

Já quando o foco foi identificar qual das oito doenças citadas no tópico 2.2 estava presente em cada amostra, a performance não seguiu a mesma tendência positiva do primeiro caso. Configurados para uma categorização mais complexa, ambos os classificadores apresentaram dificuldades para diferenciar as doenças, resultando em valores muito baixos de precisão e sensibilidade, como exibe a Tabela 26. Os números de acurácia e especificidade foram inflados pelo banco de dados desbalanceado e também não indicaram uma boa eficiência do modelo, seja MVS ou RNA.

Tabela 26 – Média dos resultados para MVS e RNA - Classificação de doenças

	Média dos resultados			
Classificador	**Acurácia**	**Precisão**	**Sensibilidade**	**Especificidade**
Máquinas de Vetores de Suporte	82,03473333	4,58613	10,03155833	86,63344167
Redes Neurais Artificiais	79,70826667	7,08218	8,194991667	86,19662917

É importante destacar que os bons resultados da classificação binária podem ter sido influenciados pelo desbalanceamento do banco de dados utilizado, qual também pode ter dificultado o desempenho da classificação multiclasse. Além disso, também é plausível considerar que os problemas na identificação das doenças tenham origem na inaptidão das medidas acústicas utilizadas para diferenciarem as classes, impedindo com que os classificadores tivessem um bom resultado. Por fim, ainda que largamente utilizadas em redes neurais, as funções de

ativação ReLu e TanH podem não terem sido as mais adequadas para um problema multiclasse, com um espaço amostral que demanda flexibilidade.

De qualquer forma, o método estudado indica ser viável e possui potencial para ser aplicado no dia a dia, solucionando os problemas das técnicas atuais e dando mais conforto e confiabilidade ao paciente. Entretanto, devido às adversidades citadas acima, novas pesquisas e testes devem ser feitos para aprimorar os resultados – especialmente na classificação de doenças - antes de colocá-lo em prática.

Em trabalhos futuros, a utilização de um banco de dados equilibrado, com mais amostras normais e frequências semelhantes para cada doença, contribuirá para confirmar a eficácia da classificação de duas classes e aprimorar a de múltiplas. Já um estudo focado nas medidas acústicas, analisando o impacto de cada uma na classificação e introduzindo novos parâmetros avaliativos na tentativa de aperfeiçoar a diferenciação das patologias, pode colaborar para um classificador mais leve e eficiente.

Para máquinas de vetores de suporte, a aplicação de outros *kernels* têm potencial para beneficiar o desempenho do classificador. Outras possibilidades de redes neurais também podem ser exploradas, com números diferentes de camadas e neurônios, outros tipos de arranjos e atribuição de diferentes funções de ativação e número de neurônios para cada camada.

REFERÊNCIAS

ALMEIDA, N. C. de. **Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte**. 2010. 119 p. Dissertação (Mestrado em Ciências) — Universidade Federal do Rio Grande do Norte, Natal, RN, 2010. Disponível em: https://repositorio.ufrn.br/jspui/bitstream/123456789/15149/1/NathaleeCA_DISSERT.pdf. Acesso em: 9 de junho de 2020.

ALVES, N. F. R. **Diagnóstico Inteligente de Patologias da Laringe**. 2016. 81 p. Dissertação (Mestre emTecnologia Biomédica) — Instituto Politécnico de Bragança, Bragança, Portugal, 2016. Disponível em: https://bibliotecadigital.ipb.pt/bitstream/10198/13915/1/Tese_V5.pdf. Acesso em: 31 de julho de 2021.

ANDRADE, L. M. de O. **Determinação dos limiares de normalidade dos parâmetros acústicos da voz**. 2003. Dissertação (Mestrado em Bioengenharia) — Bioengenharia, Universidade de São Paulo, São Carlos, SP, 2003. Disponível em: https://teses.usp.br/teses/disponiveis/82/82131/tde-28062005-102634/pt-br.php. Acesso em: 26 de agosto de 2020.

AWAN, S. N.; ROY, N.; DROMEY, C. Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. **Clinical linguistics & phonetics**, v. 23, n. 11, p. 825–841, 2009. Disponível em: https://doi.org/10.3109/02699200903242988. Acesso em: 21 de julho de 2020.

BISHOP, C. M. **Neural Networks for Pattern Recognition**. 1. ed. United States of America: Oxford University Press, 1995.

BORDENAVE, J. E. D. **O Que É Comunicação**. 22. ed. São Paulo, SP: Brasiliense, 1997.

BOYLE, B. H. **Support Vector Machines: Data analysis, machine learning and applications**. 1. ed. New York, NY: Nova Science Publishers, Inc, 2011.

CALLOU, D.; LEITE, Y. **Iniciação à fonética e à fonologia**. 11. ed. Rio de Janeiro, RJ: Zahar, 2009.

COSTA, W. C. de A. **Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas**. 2012. 178 p. Tese (Doutorado em Engenharia Elétrica) — Universidade Federal de Campina Grande, Campina Grande. PB, 2012. Disponível em: http://dspace.sti.ufcg.edu.br:8080/xmlui/handle/riufcg/1416?show%AFfull. Acesso em: 15 de julho de 2021.

DAVENPORT, M.; HANNAHS, S. J. **Introducing Phonetics and Phonology**. 3. ed. London, UK: Hodder Education Publishers, 2010.

DOHENY, L. *et al.* Reduced frequency of apnea and bradycardia episodes caused by exposure to biological maternal sounds. **Pediatrics International**, v. 54, n. 2, p. e1–e3, 2012. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1442-200X.2012.03575.x. Acesso em: 10 de julho de 2020.

GAINOR, D.; CHOWDHURY, F. R.; SATALOFF, R. T. Reinke edema: Signs, symptoms, and findings on strobovideolaryngoscopy. **Ear, Nose and Throat Journal**, v. 90, n. 4, p. 142+, 2011. Disponível em: https://go.gale.com/ps/i.dop=AONE&id=GALE%7CA264677635&v=2.1&it=r&userGroupName=anon%7E388ccd9a. Acesso em: 31 de julho de 2021.

GILLIS, T. M. *et al.* Natural history and management of keratosis, atypia, carcinoma in situ, and microinvasive cancer of the larynx. **The American Journal of Surgery**, v. 146, n. 4, p. 512–516, 1983. Disponível em: https://doi.org/10.1016/0002-9610(83)90243-X. Acesso em: 31 de julho de 2021.

GOMES, L. T. de A. **Pico cepstral nas disfonias comportamentais: dados preliminares**. 2019. 29 p. Monografia (Bacharel em Fonoaudiologia) — Universidade Federal do Rio Grande do Norte, Natal, RN, 2019. Disponível em: https://monografias.ufrn.br/jspui/bitstream/123456789/10188/1/PicoCepstralDisfoniasComportamentais_Gomes_2019.pdf. Acesso em: 21 de agosto de 2021.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed. Cambridge, MA: MIT Press, 2016.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2. ed. Porto Alegre, RS: Bookman, 2001.

IZENMAN, A. J. **Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning**. 1. ed. New York, NY: Springer Science+Business Media, 2008.

JUNQUEIRA, R. R. A. **Alterações mínimas da laringe: Um diagnóstico diferencial**. 1999. 178 p. Monografia (Especialização em Voz) — Centro de Especialização em Fonoaudiologia Clínica, Rio de Janeiro. RJ, 1999. Disponível em: http://sp.cefac.br/alunminus/cefac/biblioteca/publicacoes/arquivos/0000108_TA94.PDF. Acesso em: 11 de agosto de 2021.

KENT, R. D.; READ, C. **Acoustic Analysis of Speech**. 2. ed. Clifton Park, NY: Delmar Cengage Learning, 2002.

KRAUS, M. W. Voice-only communication enhances empathic accuracy. **American Psychologist**, v. 72, n. 7, p. 644–654, 2017. Disponível em: https://www.apa.org/pubs/journals/releases/amp-amp0000147.pdf. Acesso em: 9 de julho de 2020.

LADEFOGED, P.; JOHNSON, K. **A course in Phonetics**. 6. ed. [*S.l.*]: Wadsworth Cengage Learning, 2011.

LANIER, W. **Speech Disorders**. 1. ed. United States of America: Gale Cengage Learning, 2010.

LAROSE, D. T.; LAROSE, C. D. **Discovering Knowledge in Data: An Introduction to Data Mining**. 2. ed. Hoboken, NJ: John Wiley & Sons, 2014.

LOPES, L. W. *et al.* Medidas cepstrais na avaliação daintensidade do desvio vocal. **CoDAS**, v. 31, n. 4, 2019. Disponível em: https://doi.org/10.1590/2317-1782/20182018175. Acesso em: 21 de agosto de 2021.

MICHAELS, L. **Pathologyof the Larynx**. 1. ed. [*S.l.*]: Springer-Verlag Berlin Heidelberg 1984, 2012.

MURTON, O.; HILLMAN, R.; MEHTA, D. Cepstral peak prominence valuesfor clinical voice evaluation. **American Journal of Speech-Language Pathology**, v. 29, n. 4, p. 1596–1607, 2020. Disponível em: https://doi.org/10.1590/2317-1782/20182018175. Acesso em: 21 de agosto de 2021.

PINDZOLA, R. H.; PLEXICO, L. W.; HAYNES, W. O. **Diagnosis and Evaluation in Speech Pathology**. 9. ed. [*S.l.*]: Pearson Education, 2015.

PRZYSIEZNY, P. E.; PRZYSIEZNY, L. T. S. Work-related voice disorder. **Brazilian Journal of Otorhinolaryngology**, v. 81, p. 202–211, abr. 2015. Disponível em: http://dx.doi.org/10.1016/j.bjorl.2014.03.003. Acesso em: 12 de julho de 2020.

RAUBER, T. W. **Redes Neurais Artificiais**. Vitória, ES: Universidade Federal do Espírito Santo - Departamento de Informática, 2005.

RAZERA, D. E. **Determinadores de Pitch**. 2004. 97 p. Dissertação (Mestrado em Engenharia Elétrica) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2004. Disponível em: https://teses.usp.br/teses/disponiveis/18/18133/tde-02022016-164138/publico/Dissert_Razera_DanielE.pdf. Acesso em: 11 de agosto de 2021.

ROSA, M. de O. **Análise acústica da voz para Pré-diagnóstico de Patologias da Laringe**. 1998. 261 p. Dissertação (Mestrado em Engenharia Elétrica) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 1998. Disponível em: https://teses.usp.br/teses/disponiveis/18/18133/tde-11122015-144509/pt-br.php. Acesso em: 12 de julho de 2020.

SAÚDE, M. da. **Distúrbio de Voz Relacionado ao Trabalho (DVRT)**. 1. ed. Brasília, DF: Ministério da Saúde, 2018. v. 11. (Protocolos de Complexidade Diferenciada, v. 11). Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/disturbio_voz_relacionado_trabalho_dvrt.pdf. Acesso em: 10 de junho de 2020.

SCALASSARA, P. R. **Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias de Laringe**. 2009. 256 p. Tese (Doutorado em Engenharia Elétrica) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2009. Disponível em: https://teses.usp.br/teses/disponiveis/18/18152/tde-03122009-085230/publico/Scalassara.pdf. Acesso em: 10 de agosto de 2021.

SEIKEL, J. A.; KING, D. W.; DRUMRIGHT, D. G. **Anatomy & Physiology for Speech,Language, and Hearing**. 4. ed. Clifton Park, NY: Delmar Cengage Learning, 2009.

SERWAY, R. A.; JEWETT, J. W. **Princípios de Física, volume 2: Movimentos Ondulatórios e Termodinâmica**. 5. ed. São Paulo, SP: Cengage Learning, 2016. v. 2.

SILVA, A. H. P. **Língua Portuguesa I: Fonética e Fonologia**. 1. ed. Curitiba, PR: IESDE Brasil S.A, 2007.

SODRÉ, B. R. **Reconhecimento de Padrões Aplicados a Identificação de Patologias da Laringe**. 2016. 108 p. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) — Universidade Tecnológica Federal do Paraná, Curitiba, PR, 2016. Disponível em: http://repositorio.utfpr.edu.br/jspui/handle/1/2013. Acesso em: 30 de junho de 2020.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, v. 45, n. 4, p. 427–437, jul. 2009. Disponível em: https://doi.org/10.1016/j.ipm.2009.03.002. Acesso em: 12 de maio de 2022.

SPEAKS, C. E. **Introduction to Sound: Acoustics for the Hearing and Speech Sciences**. 4. ed. San Diego, CA: Plural Publishing, Inc, 2018. Disponível em: http://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=1724994&lang=pt-br&site=eds-live&scope=site. Acesso em: 25 de junho de 2020.

STEPP, C. E. *et al.* Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. **Ear, Nose and Throat Journal**, v. 54, n. 5, p. 1260+, 2011. Disponível em: https://go.gale.com/ps/i.do?p=AONE&id=GALE%

7CA269777486&v=2.1&it=r&userGroupName=anon%7E9b2cfedb. Acesso em: 31 de julho de 2021.

TAMPARO, C. D.; LEWIS, M. A. **Diseases of the Human Body**. 5. ed. Philadelphia, PA: F. A. Davis Company, 2011.

TAVALUC, R.; TAN-GELLER, M. Reinke's edema. **Otolaryngologic Clinics of North America**, v. 52, n. 4, p. 627–635, 2019. Disponível em: https://doi.org/10.1016/j.otc.2019.03.006. Acesso em: 31 de julho de 2021.

VOICE Disorders Database. Versão 1.03. Lincoln Park, NJ: Massachusetts Eye and Ear Infirmary e Kay Elemetrics Corp, 1994. 1 CD. [*S.l.*: *s.n.*].

WEBBA, A. R. *et al.* Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 112, n. 10, p. 3152–3157, mar. 2015. Disponível em: https://www.pnas.org/content/112/10/3152. Acesso em: 10 de julho de 2020.