IICONAIS
Congresso Nacional de Inovações em Saúde
doity.com.br/conais2021

SOCEPIS
Sociedade Cearense de Pesquisa e Inovações em Saúde

DETECÇÃO E CLASSIFICAÇÃO DE VOZES SAUDÁVEIS E PATOLÓGICAS UTILIZANDO *RANDOM FOREST*

Danilo Rangel Arruda Leite[1], Ronei Marcos de Moraes[2], Leonardo Wanderley Lopes[1,3]

[1] Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba (danilorangel@buscapb.com.br)

[2] Departamento de Estatística. Programa de Pós-graduação em Modelos de Decisão em Saúde/Universidade Federal da Paraíba (ronei@de.ufpb.br)

[3] Departamento de Fonoaudiologia/Universidade Federal da Paraíba (lwlopes@hotmail.com)

Resumo

O sistema automático de classificação de vozes saudáveis e patológicas tem recebido atenção significativa nas pesquisas de detecção e diagnóstico precoce de distúrbios vocais. Neste trabalho, propomos um método para classificar as vozes saudáveis e patológicas. Para implementar este sistema, utilizamos gravações de áudio de vozes normais e patológicas. Extraímos os Coeficientes Cepstrais de Frequência-Mel (MFCC – Mel Frequency Cepstral Coefficients) dos sinais de voz e utilizamos uma técnica de visualização para explorar a capacidade desses recursos em discriminar vozes saudáveis e patológicas. Neste estudo, utilizamos o Random Forest (RF) para classificar os recursos extraídos. Os resultados desse experimento evidenciam que o RF é um modelo que pode ser utilizado para classificação do desvio de voz, obtendo uma precisão na classificação de 0,80 com kappa de 0,46, sensibilidade e especificidade de 0,50 e 0,86, respectivamente.

Palavras-chave: Aprendizagem de máquina. Distúrbios da voz. Classificação de patologias. *Random forest.*

Área Temática: Temas livres.

Modalidade: Trabalho completo

1 INTRODUÇÃO

A voz é um dos principais meios de comunicação do ser humano, é uma ferramenta de grande complexidade que envolve aspectos fisiológicos, perceptuais, aerodinâmicos, acústicos e emocionais e a sua emissão deve ser agradável, sem esforços e conforme aos interesses profissionais, sociais e pessoais do interlocutor. Qualquer alteração na sua emissão pode ser classificada como distúrbio de voz ou desvio vocal (LOPES, 2020).

A presença de desvio na voz pode causar mudanças significativas nos padrões vibratórios, afetando assim, a qualidade da produção normal da voz, podendo influenciar

negativamente na qualidade de vida de um indivíduo, prejudicando a comunicação no seu trabalho, entre outros (LOPES, 2020; GUAN, 2019). Nesse sentido, alterações vocais devem ser diagnosticadas e tratadas o mais precocemente possível. A análise da voz possibilita alcançar resultados que mostram a condição de um desvio vocal com mais eficiência (GUAN, 2019).

Nesse cenário, a análise acústica é um procedimento não invasivo que utiliza técnicas de processamento digital de sinal de voz, podendo contribuir com medidas acústicas para a construção de ferramentas de classificação do desvio vocal. Os métodos tradicionais de diagnóstico de patologias, baseiam-se na experiência do profissional e em equipamentos caros. Sendo assim, deve-se ressaltar a importância de utilizar meios complementares para a avaliação fonoaudiológica, para possibilitar uma avaliação mais precisa e de qualidade (GUAN, 2019; LEITE, 2020).

Nessa perspectiva, os sistemas médicos assistidos por computador estão sendo cada vez mais utilizados para auxiliar os profissionais a diagnosticar e classificar os distúrbios de voz com métodos não invasivos e com menor custo (PHAM, 2018). Sendo assim, este trabalho tem como objetivo utilizar os recursos discriminativos dos sinais de voz para classificar vozes saudáveis ou patológicas, a partir das medidas cepstrais do sinal e analisar a eficiência do modelo de Aprendizado de Máquina RF (CHEN, 2020; XING, 2016; BREIMAN, 2001).

Os resultados deste estudo poderão auxiliar o profissional da voz nos procedimentos de avaliação e monitoramento dos distúrbios de voz, assim como contribuir na produção do conhecimento e treinamento de novos profissionais, sejam eles acadêmicos ou profissionais, além de auxiliar na construção de ferramentas para classificar distúrbios de voz (HOSSAIN, 2016; LOPES, 2018; MIRAMONT, 2020). Ressalta-se que as medidas cepstrais são consideradas robustas para a avaliação de sinais de indivíduos com distúrbio de voz e como técnica comumente escolhida enquanto recurso para treinar diferentes tipos de classificadores (PISHGAR, 2018; MIRAMONT, 2020).

2 MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

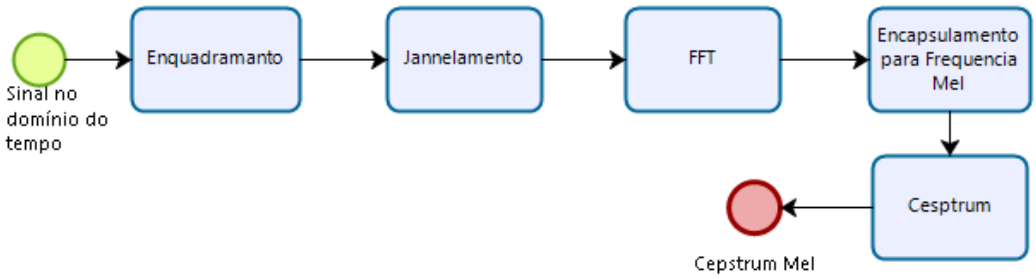
O MFCC é uma técnica aplicada em short-term spectral features, baseada no comportamento auditivo humano para extrair características acústicas do sinal de voz com base no domínio do tempo. Baseiam-se na diferença de frequências que o ouvido humano consegue distinguir e essa representação da frequência humana é feita utilizando a uma escala chamada Mel (TIWARI, 2010)

MFCC são representações da parte real do cepstro de um janelamento em curto período

de tempo de sinais acústicos, originário de uma Fast Fourier Transform (FFT) de um sinal (TIWARI, 2010). Esta técnica além de extrair características para criação de vetores acústicos do sinal de voz, também reduz a quantidade de informações sem utilidade, tais como: efeitos de reverbação e partes sem registros (TIWARI, 2010).

Para calcular os coeficientes, alguns passos devem ser seguidos, são eles: o enquadramento e janelamento do sinal, utilização dos dados no domínio da frequência usando-se de uma FFT, aplicação da escala logarítmica (Mel) para construção de um conjunto de coeficientes cepstrais e, finalmente, utilização de uma Transformada Discreta do Cosseno (DCT) (TIWARI, 2010). O fluxo do MFCC pode ser visualizado pelo diagrama da Figura 1.

Figura 1. Diagrama de funcionamento do MFCC



Fonte**:** Autores, 2021.

O processo de extração de recursos tem como entrada um sinal de áudio no domínio do tempo e tem como saída os coeficientes no domínio cepstral Mel. Para melhor compreensão da técnica MFCC é necessário entender o que é frequência em escala Mel. A audição humana não possui a mesma resposta para todas as frequências, tendo uma percepção linear até 1000Hz e uma percepção logarítmica a partir de 1000Hz (TIWARI, 2010; PISHGAR, 2018). Algumas equações podem ser utilizadas para conversão da frequência em Hz para frequência na escala mel que tentam se aproximar da percepção humana, uma delas é a equação 1:

$$mel(f) = 2595 * log_{10}\left(1 + \frac{f}{700}\right) \qquad (1)$$

em que $mel(f)$ é o resultado da frequência em $mels$ e f é a frequência nominal em Hz.

3 RANDOM FOREST (RF)

O RF é um método de conjunto utilizado para resolver problemas de classificação e regressão. Algoritmo do tipo ensemble learning, que combina várias árvores de decisão (MORAES E MARTÍNEZ 2015; BREIMAN, 2001), treinadas individualmente, na tentativa de

produzir um melhor modelo preditivo para resolver o mesmo problema, diminuindo a variância e o viés. Tem a vantagem de ser eficiente para grandes conjuntos de dados. É um algoritmo típico de Bootstrap Aggregation, também conhecido como Bagging (CHEN et al. 2020; LEITE et al. 2020).

O algoritmo usa as folhas, ou decisões finais, de cada nó para chegar a uma conclusão própria. Isso aumenta a precisão do modelo, uma vez que está observando os resultados de muitas árvores de decisão diferentes e encontrando uma média (LEITE et al. 2020).

Para utilizar o RF a fim de resolver problemas de classificação, o RF utiliza o coeficiente de Gini para determinar qual das ramificações é mais provável de ocorrer, medindo o grau de heterogeneidade dos dados. Logo, pode ser utilizado para medir a impureza de um nó. Este índice num determinado nó é dado por:

$$Gini(S) = 1 - \sum_{i=1}^{c} (p_i)^2 \qquad (2)$$

onde o p_i representa a frequência relativa da classe que está sendo observada no conjunto de dados e c representa o número de classe.

O coeficiente de entropia nos diz quão (im)puro é o conjunto de dados. Por exemplo, a entropia é zero quando o conjunto de dados for completamente homogêneo, ou um quando for igualmente dividida (BREIMAN, 2001; MITCHELL, 1997). Dado um conjunto de dados (S) que pode ter c classes distintas, a entropia de S será dada por:

$$Entropia(S) = \sum_{i=1}^{c} - p_i \, log_2(p_i) \qquad (3)$$

Outra característica relacionada à entropia é o ganho de informação. O ganho de informação é baseado na redução da entropia depois que um conjunto de dados é dividido em um atributo. Construir uma árvore de decisão envolve encontrar o atributo que retorna o maior ganho de informação, ou seja, os ramos mais homogêneos (LEITE et al., 2020).

4 TRABALHOS CORRELATOS

5 MÉTODO

Toda parte computacional, foi desenvolvida utilizando a linguagem de programação Python versão 3.6 (PYTHON 2019), assim como a biblioteca librosa versão 0.8.0 (MCFEE, 2015) para extração de recursos do sinal de voz. Librosa é uma biblioteca python utilizado para

análise de música e áudio.

5.1. DESIGN DO ESTUDO

Este trabalho trata-se de um estudo descritivo, observacional, transversal e retrospectivo. Foram avaliadas 305 amostras de indivíduos (240 mulheres e 65 homens) com idade média de 36,00 ± 12,13 anos. Todos foram atendidos no laboratório de voz de uma instituição de ensino superior entre abril de 2012 e dezembro de 2017, e todos os participantes assinaram o termo de consentimento livre e esclarecido. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da instituição de origem (1141943/14).

5.2. CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

Utilizou-se um conjunto de amostra de sujeitos que atenderam aos seguintes critérios: adultos entre 18 e 65 anos; o conjunto recebeu avaliação laringológica, incluindo laudo otorrinolaringológico escrito nas duas semanas anteriores à sessão de coleta de dados, para confirmação do diagnóstico de distúrbio de voz; não se realizou tratamento vocal (terapia ou cirurgia) antes da coleta de dados.

Foram aplicados os critérios de exclusão relacionados a: pacientes com distúrbios cognitivos ou neurológicos que impossibilitassem a utilização de procedimentos de gravação; usuários profissionais da voz e indivíduos que haviam recebido terapia vocal formal anteriormente ou que haviam se submetido a cirurgia na região da cabeça ou pescoço nos últimos 18 meses; participantes cujos sinais de voz apresentaram duração inferior a 5 segundos, presença de corte de pico no sinal acústico e relação sinal-ruído (SNR) abaixo de 30dB SPL (DELIYSKI, 2005).

Inicialmente, tínhamos uma população de 507 indivíduos (403 mulheres e 104 homens). Desses 507 participantes, 202 foram excluídos pelos seguintes critérios: 89 indivíduos não possuíam laudo otorrinolaringológico registrado no banco de dados; 39 sujeitos com recorte de pico de sinal na tarefa de voz alta; 39 indivíduos eram usuários profissionais da voz; 23 sujeitos produziram vogal sustentada com duração inferior a 5 segundos; e 12 sujeitos apresentaram sinais com SNR abaixo de 30dB SPL. Assim, utilizamos amostras de 305 sujeitos nesta pesquisa.

5.3. CARACTERÍSTICAS DOS PARTICIPANTES

Setenta e oito sujeitos não apresentavam alteração vocal, sem queixa vocal e confirmada por avaliação laringológica. Duzentos e vinte e sete pacientes foram diagnosticados com distúrbios vocais, incluindo queixa vocal e distúrbio laríngeo, a saber: nódulos nas pregas

vocais, 82 (36%); cisto de prega vocal, 38 (17%); distúrbios da voz secundários a distúrbio do refluxo gastroesofágico, 23 (10%); fechamento glótico incompleto sem causa orgânica ou neurológica, 31 (14%); pólipos de pregas vocais, 19 (8%); paralisia unilateral de prega vocal, 18 (8%); sulco vocal, 9 (4%); e edema de Reinke, 7 (3%).

5.4. PROCEDIMENTOS

Este estudo foi realizado em um desenho retrospectivo e transversal, utilizando dados de um Laboratório de voz que recebe pacientes para diagnóstico e terapia vocal. O sexo, a idade, as queixas vocais e o diagnóstico laríngeo dos sujeitos foram inseridos em um banco de dados. Os procedimentos de coleta de dados para as amostras de voz foram conduzidos de acordo com o protocolo de avaliação da voz de rotina para as avaliações iniciais dos indivíduos no laboratório. Todos os dados foram registrados digitalmente (sinais de voz) ou por escrito em prontuário (anamnese e laudo do exame laríngeo) no banco de dados do laboratório.

A vogal /Ɛ/ foi selecionada para este estudo por dois motivos. Em primeiro lugar, é uma vogal oral, média verdadeira, aberta e não arredondada e é considerada a vogal mais média do português brasileiro (GONÇALVES, 2009), semelhante ao /aɛ/ inglês. Em segundo lugar, permite uma posição mais neutra e intermediária do trato vocal e, portanto, é comumente utilizada em testes de exame visual da laringe no Brasil.

Durante a realização da extração das medidas acústicas analisadas neste estudo, as amostras foram editadas pela biblioteca librosa selecionando 3 segundos centrais de cada amostra de vogal sustentada. O objetivo desse procedimento foi excluir seções com maior variabilidade relacionadas à fase de início e deslocamento da voz. Assim, realizamos a análise acústica apenas dos 3 segundos centrais de todas as amostras.

Foram extraídas medidas cepstrais, a partir de amostras de 3 segundos da vogal sustentada /Ɛ/ em arquivos com, no mínimo, 6 segundos de duração, com taxa de amostragem de 44.100 Hz. Utilizamos a técnica MFCC por ser eficiente na extração das características do sinal de voz.

O modelo criado inclui quatro diferentes etapas, como coleta de dados, extração de recursos do sinal de voz, construção do modelo e avaliação de desempenho. As etapas que descrevem a construção do modelo proposto, são listadas a seguir: **i.** No pré-processamento, foram utilizados arquivos de voz com no mínimo 6 segundos, foram extraídos 3 segundos do meio do arquivo de áudio da vogal sustentada /Ɛ/, utilizando uma taxa de amostragem de 44.100 Hz. **ii**. Foi utilizada a biblioteca librosa para extração de recursos do MFCC e armazenar os dados em um vetor e retornar uma matriz de áudio para ser utilizada na classificação; **iii.** O

modelo de aprendizado de máquina RF (BREIMAN, 2001) foi investigado e analisados seus resultados, conforme apresentados na Tabela 1. Os parâmetros utilizados no RF, foram: Estimadores = 10, critério = Gini; para treinar o modelo, dividimos o dataset utilizando validação cruzada de 4 vezes em três parte: 65% para conjunto de treinamento, 15% para validação e 20% para teste; **iv.** Foi analisado o desempenho do modelo para verificar sua eficiência na classificação dos dados, conforme descrito em análise dos dados.

5.5. ANÁLISE DOS DADOS

Esta pesquisa investigou o algoritmo RF para classificar desvio vocal utilizando recursos extraídos do MFCC e analisar seus resultados. Para avaliar a precisão dos resultados obtidos através do classificador, foram utilizadas 4 medidas consagradas no meio científico (HOSSAIN, 2016; MORAES, 2009; LOPES, 2017), são elas: acurácia; coeficiente de Kappa (COHEN, 1960); sensibilidade; especificidade. Essas medidas estão relacionadas com os resultados de classificação e diagnóstico verdadeiro.

O teste é considerado positivo (desvio) ou negativo (saudável), e o desvio presente ou ausente. O teste está correto quando ele é positivo na presença do desvio (Verdadeiro Positivo-VP) ou negativo na ausência do desvio (Verdadeiros Negativo-VN). Além disso, o teste está errado quando ele é positivo na ausência do desvio (Falso Positivo-FP), ou negativo quando o desvio está presente (Falso Negativo-FN).

O Kappa é um método estatístico para avaliar o nível de concordância ou reprodutibilidade entre dois conjuntos de dados. O coeficiente Kappa é calculado por:

$$kappa = \frac{P(O) - P(E)}{1 - P(E)} \qquad (4)$$

em que:

P(O): proporção observada de concordâncias (soma das respostas concordantes dividida pelo total);

P(E): proporção esperada de concordâncias (soma dos valores esperados das respostas concordantes dividida pelo total).

Quanto maior o valor de Kappa, mais forte a concordância, Landis e Koch (1977) sugerem a seguinte interpretação para o Kappa (LANDIS, 1977):

Tabela 1. Interpretação do Kappa

Valores de Kappa	Interpretação
<0	Ausência de concordância
0-0,19	Concordância pobre
0,20-0,39	Concordância leve
0,40-0,59	Concordância moderada
0,60-0,79	Concordância substantiva
0,80-1,00	Concordância quase perfeita

Fonte: Landis e Koch (1977), com adaptações

Acurácia (ACC) mede a capacidade do teste de identificar corretamente quando há e quando não há presença do desvio. É definida como a relação entre o número de casos corretamente classificados e todos os casos expostos ao classificador:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \qquad (5)$$

A medida de sensibilidade (SENS) é a capacidade do teste em identificar corretamente o desvio entre aqueles que o possuem. É definida pela relação entre o número de casos corretamente classificados com a presença do distúrbio e a quantidade total de casos com o distúrbio (ANISHA, 2020).

$$SENS = \frac{VP}{VP + FN} \qquad (6)$$

A medida de especificidade (ESP) é a capacidade do teste em excluir corretamente aqueles que não possuem o desvio. É definida pela relação entre o número de casos corretamente classificados como saudável e a quantidade total de casos saudáveis (ANISHA, 2020):

$$ESP = \frac{VN}{VN + FP} \qquad (7)$$

A representação das medidas de sensibilidade e especificidade é mais clara quando se trata da discriminação entre um sinal de voz com normal e um sinal de voz com desvio. Quando há discriminação entre classes com desvios, é necessário que seja definido, no classificador, qual grupo de sinais terá sua correta classificação medida pela sensibilidade e especificidade.

6 RESULTADOS E DISCUSSÃO

A necessidade de estudar patologias da laringe de forma objetiva, já era considerada desde a década de 1970. Em 1979, Davis (1979) chegou à conclusão de que os métodos envolvendo a análise acústica da voz são mais objetivos que os métodos auditivos, na avaliação da voz. Assim, propôs que fosse utilizado meios computacionais para calcular a filtragem inversa do sinal de voz, de forma a avaliar precocemente casos de patologia e monitorizar o progresso durante a terapia da voz.

Nesse cenário, alguns estudos descrevem o desenvolvimento de classificadores de aprendizado de máquina para detecção e classificação de patologias de voz.

Em 2016, utilizou-se o classificador SVM com a técnica de MFCC, obtendo-se um resultado de 95% de precisão (HOSSAIN, 2016). Em 2018, os métodos Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (K-NN) e Gradient Boosting foram utilizados com uma amostra de 50 vozes normais e 150 com distúrbios de voz (PHAM, 2018). Em 2019, investigaram-se os métodos Support Vector Machine (SVM), Convolutional Neural Network (CNN), CNN com SVM e Autoencoders (AE) com SVM com a técinca MFCC (GUAN, 2019).

Em 2020, Chen et al. proporam um RF difusa para reconhecimento de emoções de fala. Os resultados experimentais mostraram que as precisões de reconhecimento da proposta são 87,34% RF, maiores do que as da rede neural de retropropagação que obteve 74,50% (CHEN, 2020).

Miramont et al. (2020) desenvolveram um modelo de classificação de patologias utilizando uma Rede Neural Convolucional, obtendo uma precisão de até 95,41%. Nesse estudo foi utilizado o conjunto de dados Saarbruecken Voice Database (SVD).

O MFCC tem sido comumente utilizado na classificação automática de vozes saudáveis e patológicas (PISHGAR, 2018) e para treinar diferentes tipos de classificadores. Esta técnica pode ser considerada como uma abordagem da estrutura da percepção auditiva humana, baseada no comportamento auditivo humano para extrair características acústica do sinal de voz. A representação do resultado em frequência da audição humana é feita aplicando a frequência mel.

Alguns estudos descreveram a utilização do RF, bem como SVM, KNN e CNN, também com resultados considerados bons (HOSSAIN, 2016; Wu, 2018; PHAM, 2018), evidenciando que RF é um modelo eficiente para ser utilizado para classificação do desvio de voz (ILIOU, 2009).

O classificador de RF tende a superar a maioria dos outros métodos de classificação em

termos de precisão, evitando problemas de sobreajuste. É um algoritmo do tipo *Ensemble Learning*, que agrupa os resultados ou previsões de várias árvores de decisão (CHEN, 2020), (BREIMAN, 2001), treinadas individualmente, na tentativa de produzir um melhor modelo preditivo para resolver o mesmo problema, diminuindo a variância e o viés.

Sendo assim, nesta pesquisa foi investigado o algoritmo RF para identificar os tipos de sinais de voz e classificar os distúrbios de voz. Os resultados do experimento realizado para detecção e classificação de vozes saudáveis e patológicas, demonstraram que, em termos de acurácia, o modelo RF tem um resultado considerado bom, com acurácia de 0,80 de valores de classificação correta, conforme apresentados na Tabela 1, coeficiente de Kappa e sua respectiva interpretação (grau de concordância), por sua vez, foi classificado com um grau de concordância moderada, de 0,46 (LANDIS, 1977).

Tabela 2. Valores de classificação

Modelo	Kappa	Acurácia	Sensibilidade	Especificidade
RF	0,46	0.80	0.50	0.86

A arquitetura do modelo de classificação RF utilizou o coeficiente de Gini como medida, com número de estimadores de 10. Número de estimadores indica a quantidade de árvores construídas pelo algoritmo antes de tomar uma votação ou fazer uma média de predições (BREIMAN, 2001; LEITE, 2020).

6 CONCLUSÃO

Este estudo analisou o modelo de aprendizado de máquina RF para classificar classificação de vozes saudáveis e patológicas, utilizando o MFCC. Foi observado que o modelo de classificação RF obteve resultados satisfatórios, ficando com a acurácia de 80%, com coeficiente de Kappa com um grau de concordância moderado de 0,46, sensibilidade e especificidade de 0,50 e 0,86 respectivamente. Os resultados desse experimento evidenciam que o RF é um modelo que pode ser utilizado para classificação do desvio de voz.

REFERÊNCIAS

ANISHA, C., ARULANAND, N. **Early Prediction of Parkinson's Disease (PD) Using Ensemble Classifiers**. In Innovative Trends in Information Technology (ICITIIT)-IEEE, p. 1–6, 2020.

BREIMAN, L. **Random Forests. In Machine Learning**, p. 5–32, 2001.

CHEN, L. et al. **Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction**. Information Sciences, 509, 150–163, 2020. https://doi.org/10.1016/j.ins.2019.09.005

COHEN, J. **A Coefficient of Agreement for Nominal Scales**. Educational and Psychological Measurement, 20(1), 37–46, 1960. https://doi.org/10.1177/001316446002000104.

DELIYSKI, D. D., SHAW, H. S., EVANS, M. K. **Adverse effects of environmental noise on acoustic voice quality measurements**. Journal of voice, 19(1), 15–28, 2005. https://doi.org/10.1016/j.jvoice.2004.07.003

GONÇALVES, M. I. R., VIEIRA, V. P., CURCIO, D. **Transfer function of Brazilian Portuguese oral vowels: a comparative acoustic analysis**. Brazilian Journal of Otorhinolaryngology, 75(5), 680–684,2009. http://dx.doi.org/10.1590/S1808-86942009000500012

GUAN, H., LERCH, A. **Learning Strategies for Voice Disorder Detection**. In IEEE 13th International Conference on Semantic Computing (ICSC), 2019.

HOSSAIN, M. S. e MOHAMMAD, G. **Healthcare Big Data Voice Pathology Assessment Framework**. In IEEE ACCESS, 2016.

ILIOU, T., ANAGNOSTOPOULOS, C. **Comparison of different classifiers for emotion recognition**. Proceedings of Panhellenic Conference on Informatics, p. 102–106, 2009.

LANDIS J. R., KOCH, G. G.. **The measurement of observer agreement for categorical data**. Biometrics, 33(1), 159–174, 1977.

LEITE, DANILO RANGEL A. et al. **Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest**. J. Health Inform, v.12, 196-201, 2020.

LOPES, L. W. et al. **Accuracy of traditional and formant acoustic measurements in the evaluation of vocal quality**. CoDAS, 30(5), e20170282, 2018. https://doi.org/10.1590/2317-1782/20182017282.

LOPES, L. W. et al. **Evidence of Internal Consistency in the Spectrographic Analysis Protocol. Journal of voice**, 2020. https://doi.org/10.1016/j.jvoice.2020.07.013

LOPES, L. W. et al. **Classificação espectrográfica do sinal vocal: relação com o diagnóstico laríngeo e a análise perceptivo-auditiva**. Audiology - Communication Research 25, e2194, 2020. https://doi.org/10.1590/2317-6431-2019-2194

LOPES, L. W. et al. **Accuracy of Acoustic Analysis Measurements in the Evaluation of Patients With Different Laryngeal Diagnoses**. Journal of Voice, 31(3), 382.e15-382.e26, 2017. https://doi.org/10.1016/j.jvoice.2016.08.015

MCFEE, B. et al. **Librosa: Audio and music signal analysis in python**. In Proceedings of the 14th python in science conference, p. 18–25, 2015.

MIRAMONT, J. et al. **Voice Signal Typing Using a Pattern Recognition Approach**. Journal of voice, 2020. https://doi.org/10.1016/j.jvoice.2020.03.006

MITCHELL, T. M. **Machine Learning**. s.l.: s.n, 1997.

MORAES, R. **Computational Intelligence Applications for Data Science**. Science Direct, Out, pp. 1-2, 2015.

MORAES, R. M., MACHADO, L. **Gaussian Naive Bayes for Online Training Assessment in Virtual Reality-Based Simulator**. Mathware & Soft Computing, 16(2), 123–132, 2009.

PHAM, M., LIN, J. E ZHANG, Y. **Diagnosing Voice Disorder with Machine Learning**. IEEE International Conference on Big Data, 2018.

PISHGAR, M., KARIM, F. E MAJUMDAR, S. **Pathological Voice Classification Using Mel-Cepstrum Vectors and Support Vector Machine**. Electrical Engineering and Systems Science, 2018.

PYTHON (2019). Python.org. https://www.python.org/

TIWARI, V. **MFCC and its applications in speaker recognition**. International Journal on Emerging Technologies 1(1), 19–22, 2010.

WU, H., SORAGHAN, J., LOWIT, A. E DI-CATERINA, G. **A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks**. Interspeech, 2018.

XING, F. E YANG, L. **Machine Learning and Medical Imaging**. New York: Elsevier, 2016.