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

VITOR OLIVEIRA PIERITZ

Identificação acústica não invasiva de disfonia vocal utilizando inteligência artificial

São José do Rio Preto
2023

VITOR OLIVEIRA PIERITZ

Identificação acústica não invasiva de disfonia vocal utilizando inteligência artificial

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Rodrigo Capobianco Guido

São José do Rio Preto

2023

P618i

Pieritz, Vitor Oliveira

Identificação acústica não invasiva de disfonia vocal utilizando inteligência artificial / Vitor Oliveira Pieritz. -- São José do Rio Preto, 2023

32 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Rodrigo Capobianco Guido

1. Ciência da computação. 2. Aprendizado do computador. 3. Inteligência artificial. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

VITOR OLIVEIRA PIERITZ

Identificação acústica não invasiva de disfonia vocal utilizando inteligência artificial

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Comissão Examinadora

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

Prof. Dr. Leandro Neves
UNESP – Câmpus de São José do Rio Preto

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

São José do Rio Preto
10 de janeiro de 2023

RESUMO

A voz é um instrumento fundamental para a comunicação dos seres humanos, sendo muitas vezes primordial para exercer algumas profissões, como por exemplo no caso de professores, atores, locutores de rádio e dubladores. Porém, nem sempre o cuidado do aparelho vocal recebe toda a atenção necessária, uma vez que para se obter um diagnóstico e tratamento adequados é necessário passar por processos invasivos e muitas vezes traumáticos, o que muitas vezes pode desestimular as pessoas a procurar tratamentos adequados. A computação e a inteligência artificial possibilitam a detecção de problemas vocais de maneira não invasiva, através de áudios e inteligência artificial, o que poderia estimular as pessoas a procurarem tratamento adequado uma vez que foi detectado uma possível patologia vocal. Dentre uma das várias patologias existentes, uma chama a atenção devido a não clareza de sintomas mas a facilidade de se obter caso não se tenha um domínio vocal adequado: A disfonia funcional vocal. Este trabalho desenvolveu e implementou um método computacional capaz de detectar a disfonia funcional vocal nos pacientes de forma não-invasiva, utilizando áudios da voz de pacientes sadios e não sadios, computação e inteligência artificial. As técnicas utilizadas para tal ato foram: Máquina de Vetores de Suporte (SVM)(do tipo kernel linear e RBF kernel) e K-Vizinhos Mais Próximos (KNN) para aprendizado de máquina, Coeficientes Cepstrais na Frequência de Mel (MFCC) e Operador de Energia de Teager (TEO) para pré-processamento de dados e extração de recursos úteis dos áudios dos pacientes e Matrizes de Confusão e Validação Cruzada para validação dos resultados obtidos. Como melhor resultado, foi encontrado a combinação de MFCC com SVM Linear resultando em uma acurácia média de 95%.

Palavras–chave: Aprendizado de Máquina. Disfonia funcional vocal. Máquina de Vetores de Suporte. K-Vizinhos Mais Próximos. Coeficientes Cepstrais na Frequência de Mel. Operador de Energia de Teager. Matrizes de Confusão . Validação Cruzada.

ABSTRACT

The voice is a fundamental instrument for the communication of human beings, often primordial to exercise some professions, for example, in the case of teachers, actors, radio announcers and voice actors. However, the care of the vocal apparatus does not always receive all the necessary attention, since to obtain an adequate diagnosis and treatment it is necessary to undergo invasive and often traumatic processes, which can often discourage people from seeking appropriate treatments. Computing and artificial intelligence make it possible to detect vocal problems in a non-invasive way through audio and artificial intelligence, which could encourage people to seek adequate treatment once a possible vocal pathology has been detected. Among one of the several existing pathologies, one draws attention due to the lack of clarity of symptoms but the ease of attaining it if one does not have an adequate vocal domain: Vocal functional dysphonia. This work developed and implemented a computational method capable of detecting functional vocal dysphonia in patients in a non-invasive way, using voice audio from healthy and unhealthy patients, computing and artificial intelligence. The techniques used for this purpose were: Support Vector Machine (SVM) (linear kernel type and RBF kernel) and K-Nearest Neighbors (KNN) for machine learning, Cepstral Coefficients in the Honey Frequency (MFCC) and Operator Teager's Energy (TEO) for data pre-processing and extraction of valuable resources from the patients' audio and Confusion Matrices and Cross-Validation for validating the results obtained. As the best result, the combination of MFCC with Linear SVM was found, resulting in an average accuracy of 95%.

Keywords: Machine Learning. Vocal functional dysphonia. Support Vector Machine. K-Nearest Neighbors. Cepstral Coefficients in the Frequency of Honey. Teager Energy Operator. Confusion Matrices. Cross Validation.

LISTA DE ILUSTRAÇÕES

Figura 1- O hiperplano de separação 12
Figura 2- 2 vetores mais próximos do hiperplano possível 12
Figura 3- Erros de classificação 13
Figura 4- A margem suave 13
Figura 5- Função Kernel com poucas dimensões 14
Figura 6- Função Kernel com muitas dimensões 14
Figura 7- Variação no Kernel RBF Gaussiano com variação em σ 15
Figura 8- Acurácia do valor de K em MFCC com KNN 25
Figura 9- Acurácia de K em MFCC com KNN 25
Figura 10- Taxa de erro no valor de K em MFCC com KNN 26
Figura 11- Acurácia do valor de K em TEO com KNN 26
Figura 12- Acurácia de K em TEO com KNN 27
Figura 13- Taxa de erro no valor de K em TEO com KNN 27

LISTA DE TABELAS

Tabela 1 – Acurácia geral de cada modelo 28

LISTA DE ABREVIATURAS E SIGLAS

FN	Falso Negativo
FP	Falso Positivo
KNN	K-Vizinhos Mais Próximos
MFCC	Coeficientes Cepstrais na Frequência de Mel
RBF	Função de Base Radial
RF	Florestas Randômicas
RFC	Classificador de Florestas Aleatórias
SVM	Máquina de Vetores de Suporte
TEO	Operador de Energia de Teager
TDF	Transformada de Fourier
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1 INTRODUÇÃO 7
1.1 Considerações Iniciais 7
1.2 Objetivos 8
1.3 Justificativa 8
1.4 Motivação 8
1.5 Metodologia 10
2. REVISÃO BIBLIOGRÁFICA 11
2.1 Técnicas de aprendizado de máquina 11
2.1.1 Máquinas de Vetores de Suporte (SVM) 11
2.1.2 K-Vizinhos Mais Próximos (KNN) 16
2.2 Pré-processamento e extração de caracerísticas 16
2.2.1 Coeficientes Cepstrais na Frequência de Mel (MFCC) 16
2.2.2 Operador de Energia de Teager(TEO) 17
2.3 Validação de resultados 18
2.3.1 Matrizes de Confusão 18
2.3.2 Validação Cruzada 19
2.4 Trabalhos correlatos e estado da arte 19
3 DESENVOLVIMENTO DO TRABALHO 21
3.1 Coleta de dados 21
3.2 Métodos de extração de características com pré processamento 22
3.3 Métodos de aprendizado de máquina com comprovação de resultados 24
3.3.1 MFCC com KNN 24
3.3.2 TEO com KNN 26
3.3.3 SVM 27
4 RESULTADOS 28
4.1 Resultados Gerais 28
5 CONCLUSÃO 30
REFERÊNCIAS 31

1 INTRODUÇÃO

1.1 Considerações iniciais

É inquestionável que a voz é um instrumento importante para a comunicação dos seres humanos, sendo também muitas vezes fundamental para exercer algumas profissões, como professores, atores, locutores de rádio e dubladores. Porém, apesar da importância da voz, é extremamente comum o descuido das pessoas com seus aparelhos vocais, uma vez que para se obter um diagnóstico e tratamento adequados é necessário passar por processos invasivos e muitas vezes traumáticos, como por exemplo a laringoscopia e a nasofibroscopia.

Devido a este cenário, o uso da computação e da inteligência artificial possibilitaria a detecção de problemas vocais de maneira não invasiva através de áudios, o que diminuiria os possíveis problemas causados pela negligência em buscar o tratamento adequado, uma vez que ao se detectar o problema, estimula os pacientes a buscarem o tratamento correto sem prolongação.

Dentre as várias patologias vocais que existem, uma chama atenção devido a sua natureza: A disfonia funcional vocal. A disfonia funcional vocal é uma má qualidade de voz, sem dificuldades anatômicas, neurológicas ou outras dificuldades orgânicas óbvias que afetam a laringe, caixa de voz ou qualquer órgão vocal(ROY, 2003).

BEHLAU et al(2015) explica que existem dois tipos de disfonias vocais, a disfonia funcional e a orgânica (sendo esta sendo originada por consequências não diretas na voz). Como é uma patologia em que não se é muito óbvio o motivo de seu surgimento, além do mal uso do aparelho vocal, a disfonia funcional vocal está mais propensa a ser negligenciada do que em comparação com patologias que apresentam sintomas mais agudos.

No campo científico, ao se tratar de usar inteligência artificial para detectar doenças vocais, a patologia mais pesquisada é a de doença de Parkinson. O trabalho de HEDGE et al (2018), que consiste em ser uma pesquisa sobre várias técnicas de detecção de patologias vocais em processamentos de voz, mostrou que o mal de Parkinson foi abordado 19 vezes das 91 pesquisas.

Apesar de não ser citado no trabalho de HEDGE, o trabalho de DANKOVICOVÁ et al(2018) apresenta uma maneira de detectar a disfonia vocal usando os classificadores Máquina de Vetores de Suporte (SVM), Classificador de Florestas Aleatórias (RFC) e K-Vizinhos Mais Próximos (KNN).

1.2 Objetivos

O objetivo deste trabalho é desenvolver e implementar um método computacional capaz de detectar a disfonia funcional vocal nos pacientes de forma não-invasiva, utilizando áudios da voz de pacientes sadios e não sadios, computação e inteligência artificial. Além disso, visa esboçar um pequeno estado da arte do uso atual de inteligência artificial para classificar áudios de pacientes com patologias vocais.

1.3 Justificativa

Apesar da importância da voz, não é da natureza das pessoas dar a devida atenção a problemas e sintomas apresentados pelos órgãos do aparelho vocal, evitando ou adiando muitas vezes a busca de diagnósticos e tratamentos adequados quando os sintomas começam a aparecer. Mesmo não sendo claro quais são todos os motivos de tal comportamento, pode-se afirmar que as técnicas de diagnóstico da fonoaudiologia são invasivas e até traumáticas para alguns pacientes, o que os desencorajam a procurar ajuda profissional adequada para cuidarem de sua voz. Desta forma, uma maneira não invasiva de detectar patologias vocais pode ser utilizada a fim de amenizar este problema, encorajando os pacientes procurarem ajuda profissional adequada, algo que pode ser feito graças à inteligência artificial.

1.4 Motivação

Apesar das várias patologias vocais existentes, a disfonia vocal é uma que chama bastante a atenção, devido a sua origem nem sempre ser por doenças. Sama et al (2001) descreve que a disfonia funcional refere-se ao comprometimento da produção de voz na ausência de alterações mucosas ou neurogênicas devido a doenças da laringe, com mais de 40% dos pacientes disfônicos não apresentarem doença orgânica ou mucosa identificável. Roy(2003) complementa estas informações, descrevendo que a disfonia vocal é um distúrbio da voz na ausência de patologia laríngea estrutural ou neurológica, de natureza multidisciplinar da voz, sendo

a atividade mal regulada dos músculos intrínsecos e extrínsecos da laringe como principal causa da disfonia vocal, bem como traços de personalidade específicos.

BEHLAU et al(2015) acrescenta que existem dois tipos de disfonias: A orgânica e a funcional. A orgânica caracteriza-se por ser causada devido a complicações não relacionadas diretamente com a voz,como refluxo gastroesofágico, paralisia das pregas vocais e doenças sistêmicas, por exemplo, Parkinson e esclerose lateral amiotrófica. Já as disfonias funcionais são decorrentes de eventos fonotraumáticos, comportamentos abusivos ou mau uso da voz, podendo ser citados alguns eventos como a má técnica vocal e/ou desequilíbrio muscular, com ou sem envolvimento psicoemocional.

Este tipo de situação, onde a pessoa não tem conhecimentos adequados de como utilizar a própria voz, é extremamente comum no nosso país e no mundo; uma vez que muitas vezes os profissionais da voz não recebem instruções adequadas de como utilizar apropriadamente a sua voz ou nem ao menos percebem que estão se prejudicando. Usando um dos profissionais da voz como exemplo, no caso os professores, PENTEADO (2007) afirma a necessidade da sociedade influenciar os professores a buscarem tratamento adequado:

> "[...] é preciso melhor compreender as maneiras pelas quais a sociedade, por meio das relações de trabalho, da cultura e da qualidade de vida, influencia o processo saúde-doença-cuidado de professores e, para tanto, faz-se necessário buscar conhecer as percepções que os trabalhadores professores têm acerca da sua voz, dos cuidados com a saúde vocal e as maneiras de interpretarem e de lidarem com os seus problemas vocais."

Assim, a disfonia vocal seria uma patologia facilmente negligenciada pelas pessoas, uma vez que seus sintomas nem sempre são claros e facilmente detectáveis e não existe uma cultura de as pessoas cuidarem de sua voz por precaução. Por isto é de suma importância a detecção não invasiva desta patologia, uma vez que as pessoas recorreriam a esta ferramenta rápida capaz de detectar possíveis problemas de disfonia vocal e evitar problemas de saúde maiores causados pela negligência ou por não quererem se sujeitar a medidas invasivas e traumatizantes para o paciente.

Esta ferramenta também auxiliaria nos cenários em que os pacientes nem ao menos fazem ideia que estão com algum problema vocal, como no caso dos professores como apresentado por PENTEADO(2007). O uso de inteligência artificial

em áudios dos pacientes para poder classificar a disfonia vocal seria de extrema utilidade para a sociedade de uma maneira geral, uma vez que agregaria maior consciência de saúde-doença de maneira rápida e não invasiva.

1.5 Metodologia

Para realizar o objetivo proposto por este trabalho, pesquisou-se na literatura trabalhos que continham alguma com a aplicação de aprendizado de máquina na detecção de patologias vocais. Em seguida, coletou-se áudios da base de dados Saarbrüecken Voice Database[3] de pessoas sadias e pessoas que apresentam disfonia funcional vocal . Por fim, para desenvolver este método computacional de previsão, é necessário dividir em 3 etapas: Pré-processamento, métodos de aprendizado de máquina e métodos de comprovação de resultados.

Para pré-processamento, este trabalho utilizou Coeficientes Cepstrais na Frequência de Mel (MFCC) e Operador de Energia de Teager(TEO). Nesta etapa este trabalho já se destaca em relação ao trabalho de DANKOVICOVÁ et al(2018), uma vez que não usaram estas tratativas para a classificação do seu trabalho. Os métodos de aprendizado de máquina utilizados são o SVM(tanto modo de Kernel linear quanto Função de Base Radial, RBF) e o KNN, o que é bem similar ao de DANKOVICOVÁ et al(2018).

Por fim, os resultados obtidos são validados utilizando matrizes de confusão e Validação Cruzada, o que é necessário para avaliar a efetividade dos algoritmos de aprendizado de máquina e algoritmos de classificação. Este ponto aqui também é um diferencial em relação ao trabalho de DANKOVICOVÁ et al(2018), que utiliza Validação Cruzada mas não utiliza de matrizes de confusão para validação.

Para o trabalho ser realizado, foi utilizado o Google Colaboratory, uma vez que é uma ferramenta online e gratuita que facilita a implementação de códigos por não exigir nenhuma configuração e possuir acesso gratuito a GPUs. Devido o uso de Google Colaboratory, também será utilizado a linguagem Python, uma vez que apresenta algumas bibliotecas que auxiliam na implementação de classificadores, como por exemplo o Pandas e o Librosa.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo serão apresentados e definidos alguns dos conceitos para a compreensão e entendimento do trabalho. Os tópicos que serão descritos são: Técnicas de aprendizado de máquina (Máquinas de Vetores de Suporte e K-Vizinhos Mais Próximos), pré-processamento e extração de características(Coeficientes Cepstrais na Frequência de Mel e Operador de Energia de Teager), validação de resultados (Matrizes de Confusão e Validação Cruzada), trabalhos correlatos e estado da arte.

2.1 Técnicas de aprendizado de máquina

O aprendizado de máquina é um campo dentro da inteligência artificial que visa construir automaticamente modelos analíticos a partir de dados de entrada. É importante notar que GOYAL et al(2020) em seu trabalho concluiu que caso deseja-se alcançar uma boa precisão com baixo tempo de computação, então classificadores tradicionais de aprendizado de máquina como SVM (kernel da função de base radial ,RBF), KNN e Florestas Randômicas(RF) são as melhores escolhas. Graças a esta informação, foi selecionado o uso de SVM (do tipo linear e do tipo RBF) e KNN para este trabalho.

2.1.1 Máquinas de Vetores de Suporte (SVM)

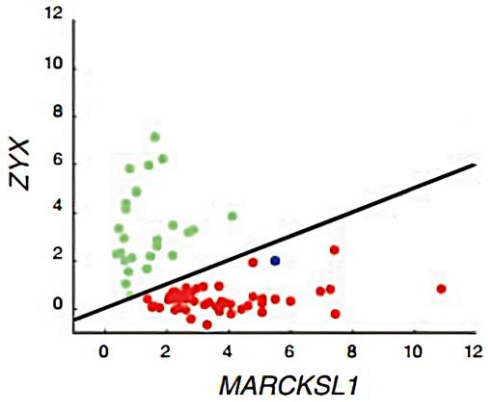
Segundo Noble (2006), Máquinas de Vetores de Suporte (SVM) é um algoritmo computacional capaz de aprender por exemplos ao atribuir rótulos para objetos, sendo geralmente aplicado com sucesso em aplicações biológicas, mais comumente usadas na classificação automática de perfis de expressão de genes de microarray.

As Figuras 1- 6 representam o estudo de Noble (2006) sobre o uso de SVM para conseguir prever se o paciente possui leucemia linfoblástica aguda (pontos vermelhos), leucemia mielóide (pontos verdes) ou algo desconhecido (pontos azuis), a partir dos genes ZYX e MARCKSL1. Um valor grande em algum dos eixos indica que o gene é altamente expresso e vice-versa.

As ideias básicas por trás do algoritmo SVM podem ser compreendidas ao entender quatro conceitos: O hiperplano de separação, o hiperplano de margem máxima, a margem suave e a função kernel.

O hiperplano de separação tem como objetivo separar 2 classes que serão classificadas por meio de um hiperplano. Geometricamente, esta regra corresponde a traçar uma linha entre os dois aglomerados de vetores das duas classes. Posteriormente, basta os novos objetos serem posicionados seguindo uma das duas classes. A Figura 1 ilustra bem como é feita esta separação.

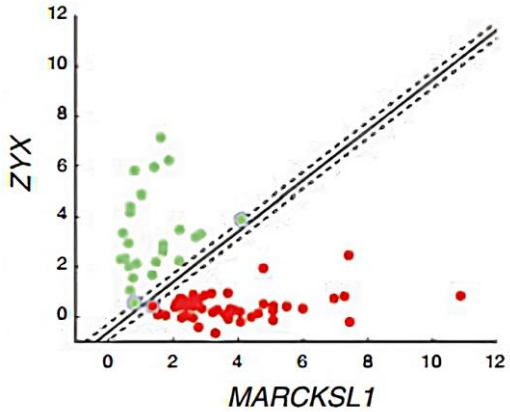
Figura 1 – O hiperplano de separação



Fonte: Adaptada de (NOBLE, 2006, p.1566)

O hiperplano de margem máxima tem como objetivo separar estas duas classes sem erros, mantendo a distância entre os vetores o mais próximos do hiperplano possível, conforme ilustrado na Figura 2.

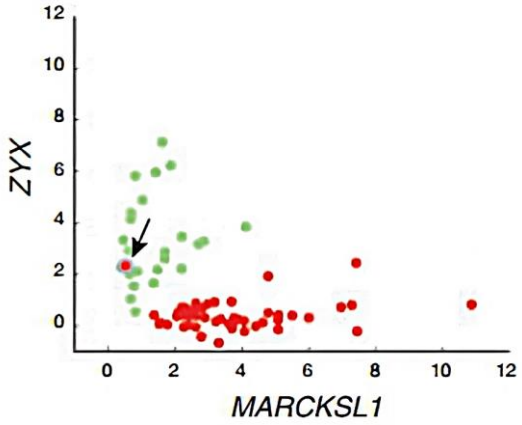
Figura 2 – 2 vetores mais próximos possível do hiperplano



Fonte: Adaptada de (NOBLE, 2006, p.1566)

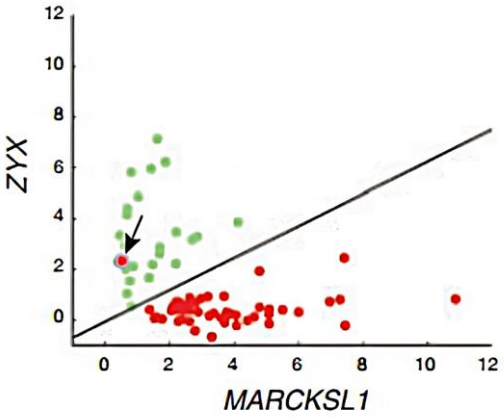
A margem suave serve para que pequenos erros de classificação, como mostrado na Figura 3, não afetem o resultado final, permitindo que alguns pontos dos dois vetores não sejam afetados pelo hiperplano. A Figura 4 mostra como seria a solução apresentada pela Figura 3.

Figura 3 - Erros de classificação



Fonte: Adaptada de (NOBLE, 2006, p.1566)

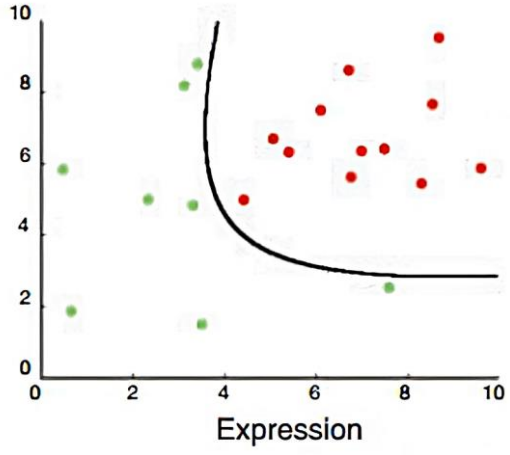
Figura 4 – A margem suave



Fonte: Adaptada de (NOBLE, 2006, p.1566)

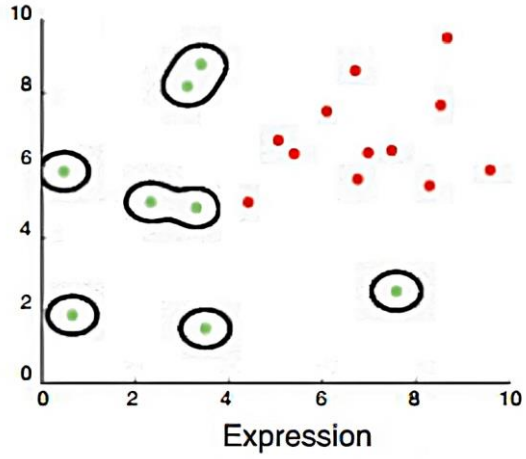
Em alguns casos, a margem suave pode não ser útil por não conseguir fazer nenhum ponto separando as duas classes. A função kernel fornece uma solução para este problema ao adicionar dimensões aos dados, sendo elas adquiridas a partir do uso de valores da expressão original ao serem elevados ao quadrado. A quantidade de dimensões usadas pela função kernel afeta diretamente o quão restrita é a margem entre as classes, como pode ser vista a diferença na Figura 5 e Figura 6.

Figura 5- Função Kernel com poucas dimensões



Fonte: Adaptada de (NOBLE, 2006, p.1566)

Figura 6- Função Kernel com muitas dimensões



Fonte: Adaptada de (NOBLE, 2006, p.1566)

Apesar de aumentar a quantidade de dimensões nas funções de kernel para cuidar da necessidade de margens suave poder resolver o problema de separar duas classes, no entanto, pode ser problemático. Isto acontece devido à chamada maldição de dimensionalidade: como o número de variáveis em consideração aumenta, o número de soluções possíveis também aumenta, mas exponencialmente. Consequentemente, torna-se mais difícil para qualquer algoritmo selecionar a solução correta.

HEDGE et al (2018) complementa com a informação de que o SVM tem suas raízes na teoria da aprendizagem estatística e mostra resultados promissores em

muitas aplicações práticas, que funcionam muito bem para dados de alta dimensão. A ideia básica é construir um hiperplano com um máximo de margem que separa as características de diferentes classes ao representar o limite de decisão como um conjunto de vetores de suporte.

GOSH et al (2019) explica que existem dois tipos de classificadores usados para os classificadores SVM: os do tipo lineares e os do tipo não lineares. A SVM do tipo linear é extremamente similar a citada acima, onde usa-se uma margem de classificação entre as duas classes bem grande, sendo ela o hiperplano. Também pode haver 2 tipos de margem, margem dura e margem suave. Já nos classificadores SVM não lineares, os tipos mais comuns de kernel são: Função de kernel polinomial, função de kernel sigmoid e o kernel RBF. Como este trabalho se limita em usar apenas o SVM do tipo linear e o SVM do tipo kernel da função de base radial (RBF), a explicação será limitada a apenas estes casos.

GOSH et al (2019) explica que o kernel da função de base radial (RBF) é o kernel mais popular para classificação de dados de treinamento e SVM. A sua função é similar à função de calcular distância euclidiana, usando dois pontos de referência com um parâmetro livre σ, cuja a fórmula é dada por:

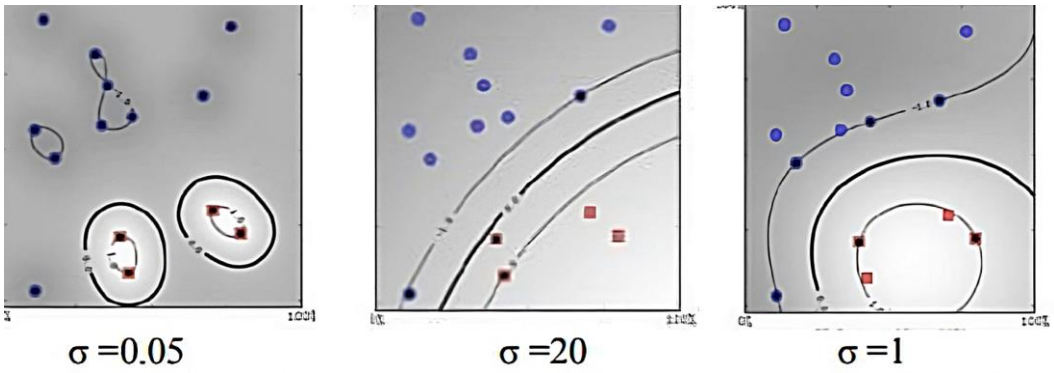
$$K(x, x') = \exp(-\gamma||x - x'||^2) \tag{2.1}$$

O parametro γ pode ser escrito como :

$$\gamma = \frac{1}{2\sigma^2} \tag{2.2}$$

A Figura 7 mostra um gráfico comparativo dependendo do valor de σ.

Figura 7- Variação no Kernel RBF Gaussiano com variação em σ



Fonte: (GOSH et al, 2019)

2.1.2 K-Vizinhos Mais Próximos (KNN).

Segundo GOYAL et al(2020), os K-Vizinhos Mais Próximos é um classificador de aprendizado de máquina supervisionado no qual, para cada ponto de dados de teste, K pontos de dados de treinamento mais próximos são encontrados e verificados quais são as classes que ocorrem com mais frequência nesses pontos, atribuindo a elas como pontos de dados de teste. Estes pontos K mais próximo dos pontos de dados são descobertos com a ajuda da distância euclidiana, porém o valor de K influencia a precisão do desempenho do modelo.

Se o valor de k for pequeno, então os resultados dos testes serão mais sensíveis a informações desnecessárias. Se k for muito alto, o custo computacional será aumentado significativamente. Portanto, cabe ao usuário ter parcimônia e decidir adequadamente quais são os valores ideais, mas geralmente variando entre 5 e 10.

2.2 Pré-processamento e extração de caracerísticas

Para poder processar qualquer informação no reconhecimento automático de voz, a extração de recursos é necessária. Isso é importante porque os sons geralmente são compostos com informações desnecessárias para análise, como por exemplo, ruídos de fundo, emoções ou os sons gerados por um ser humano devido ao trato vocal (incluindo dentes, língua, etc). Estas informações poluem as informações importantes que irão ser analisadas e por isso, precisam ser filtradas. Para tal feito, foi selecionado dois tipos de técnica de extração de características: Coeficientes Cepstrais na Frequência de Mel (MFCC) e Operador de Energia de Teager(TEO)

2.2.1 Coeficientes Cepstrais na Frequência de Mel (MFCC)

A análise de sons utilizando a frequência Mel é baseada experimentos de percepção do ouvido humano, isto é importante pois é provado que os ouvidos humanos são mais sensíveis e têm resolução mais alta aos sons de baixa frequência em comparação com sons de frequência alta. Assim, o banco de filtros é projetado para enfatizar a baixa frequência sobre a alta frequência, a fim de emular este comportamento dos ouvidos humanos. Esta escala de percepção de frequências com

distâncias iguais possibilita a extração de recursos nos áudios. Em outras palavras, a frequência da escala Mel é proporcional à logaritmos de frequência lineares, refletindo a percepção humana sobre os sons. A equação abaixo converte uma escala de frequência linear em uma escala de frequência Mel, onde f é a frequência em Hertz e log está na base 10 (JOSHI, CHEERAN, 2014).

$$\mathrm{Mel(f)} = 2595\log(1 + \frac{\mathrm{f}}{700}) \tag{2.3}$$

HEDGE(2018) complementa estas informações ao escrever que o MFCC é muito bom para conseguir identificar a existência de patologia mas não no caso de identificá-la. Além disso, ele resume os passos gerais para extrair os recursos do MFCC sendo elas:

i. Cálculo dos coeficientes discretos da transformada de Fourier(TDF).
ii. Filtragem com filtro triangular espaçado de Mel.
iii. Cálculo de energias de sub-bandas.
iv. Cálculo dos coeficientes de transformada discreta de cosseno.

2.2.2 Operador de Energia de Teager(TEO)

O Operador de Energia de Teager (TEO) é um filtro passa-alta não linear que suprime o sinal de fundo de baixa frequência enquanto melhora o sinal de alta frequência transitório, tendo sido usado com sucesso em várias aplicações para processamento de fala e processamento de imagens. (SUBASI; YILMAZ; TUFAN, 2011).

Segundo GUIDO (2019), TEO pode ser calculado a partir de uma analogia do sistema massa-mola, onde o TEO (T) aproximadamente modela a energia de um sinal discreto (x[.]) como o produto ao quadrado de sua amplitude (A) por sua frequência instantânea (Ω), como descrita na Equação a seguir.

$$x_n^2 - x_{n-1} \cdot x_{n+1} = A^2 \cdot sin^2(\Omega) = T \tag{2.4}$$

2.3 Validação de resultados

Para garantir a estabilidade dos modelos de aprendizado de máquina, é necessário que elas passem por algum tipo de validação dos resultados. No caso deste trabalho, serão usados dois tipos de validação: As Matrizes de Confusão e Validação Cruzada. Ambas são técnicas muito úteis quando é necessário avaliar a generalização de um modelo, ou seja, avaliando um novo conjunto de dados formados pelo aprendizado de máquina.

2.3.1 Matrizes de Confusão

Segundo DÜNTSCH et al(2019), a soma dos elementos diagonais de uma matriz de confusão é amplamente utilizada para medir o sucesso de uma classificação baseada em um algoritmo ou observação humana; em contraste com uma medição "verdadeira", ou seja, uma classificação definida por um especialista da área. A ideia principal é que um algoritmo (ou um observador) consiga formar suas próprias classes de equivalência dos dados e é forçado a atribuir a estas classes as categorias dadas pela medição verdadeira

No ramo computacional, a matriz de confusão é usada para avaliar a eficácia de um aprendizado de máquina de um algoritmo de classificação. A avaliação funciona com uma tabela de frequência de classificação com quatro combinações diferentes de valores previstos e reais. Esses quatro valores podem ser: Verdadeiro Positivo (VP), o que significa que o algoritmo previu positivo e é verdadeiro; Falso Positivo (FP), o que significa que o algoritmo previu positivo e é falso, Falso Negativo (FN), o que significa que o algoritmo previu negativo e é falso e Verdadeiro Negativo (VN), o que significa que o algoritmo previu negativo e é verdade.

A partir destes valores, é possível extrair informações para melhorar calcular a eficácia do aprendizado de máquina. Geralmente as medidas mais utilizadas são especificidade, precisão, f1 score, sensibilidade e acurácia. No caso deste trabalho, a medida a ser adotada será a acurácia, uma vez que é a melhor medida para medir percentualmente a relevância de uma determinada classe. Segundo, BERRE et al(2019), a fórmula a ser usada para esta medição será

$$Acurácia(\%) = \frac{VP+VN}{VP+VN+FP+FN} \tag{2.5}$$

Os parâmetros da equação são: Verdadeiro Positivo(VP), Falso Positivo(FP), Verdadeiro Negativo(VN) e Falso positivo(FN). Ou seja, a acurácia é calculada a partir da soma da quantidade de Verdadeiros Positivos com a quantidade de Falsos Positivos dividido pela soma de Verdadeiros Positivos, Falsos Positivos, Verdadeiros Negativos e Falsos positivos.

2.3.2 Validação Cruzada

A Validação Cruzada divide aleatoriamente os exemplos de entrada de dados em um número de tamanhos iguais, criando subconjuntos e treinando o classificador várias vezes(k vezes), exceto um dos subconjuntos, que será usado para validação após cada procedimento de treinamento. Desta forma um procedimento de Validação Cruzada k vezes terão k-1 subconjuntos são usados para treinamento e 1 subconjunto para teste. Isso é repetido k vezes para usar todas as combinações de conjuntos de treinamento e validação. Com esta abordagem é possível que todo o dado seja classificado e validado, enquanto mantem um conjunto de treinamento separado (GRANHOLM et al,2012).

2.4 Trabalhos correlatos e estado da arte

HEDGE et al(2018) realizou uma pesquisa sobre abordagens de aprendizado de máquina para detecção automática de distúrbios de voz onde explica sucintamente as definições do que seria algum tipo de distúrbio vocal, os tipos de patologias, principais bases de dados utilizadas, principais técnicas de extração de características e principais técnicas de aprendizado de máquina utilizadas. Esta pesquisa também mostrou e comparou a acurácia de 98 trabalhos, bem como levanta discussões, como a importância da pesquisa para Alzheimer e doença de Parkinson.

ISLAM et al(2020) complementa o trabalho anterior, reforçando a necessidade a necessidade de técnicas não invasivas de processamento de sinal para detectar deficiência de voz poder ser de uso da população em geral, além de ser mais detalhista em como a voz é formada, condições médicas para a incapacidade vocal, métodos medicinais atuais e outras técnicas de aprendizado de máquina que não foram contempladas no trabalho de HEDGE et al(2018).

A respeito de trabalhos atuais envolvendo a detecção automática de distúrbios de voz, os principais trabalhos estão relacionados à doença de Parkinson, como pode ser citado o trabalho de GOYAL et al (2020). Neste trabalho, faz-se uma análise comparativa de compara vários classificadores de aprendizados de máquina, aprendizado profundo e métodos ensemble baseados em medições como precisão, sensibilidade, especificidade, precisão e pontuação F1. Além disso, concluiu-se que caso deseja-se alcançar uma boa precisão com baixo tempo de computação, então classificadores tradicionais de aprendizado de máquina como SVM (RBF), KNN e RF são as melhores escolhas.

A respeito do uso de inteligência artificial para a detecção acústica não invasiva de disfonia funcional vocal exclusivamente, o trabalho de DANKOVICOVA et al(2018). utiliza SVM, RFC e KNN para detectar disfonia fucional vocal, bem como a base de dados *Saarbruecken* Voice Database. Os resultados mostrados chegaram a até 91,3%.

3 DESENVOLVIMENTO DO TRABALHO

Utilizando dos conceitos apresentados no capítulo 2, este capítulo detalhará as etapas do método proposto para este trabalho. Estas etapas podem ser divididas em: Coleta de dados, métodos de extração de características e pré processamento e Métodos de aprendizado de máquina e comprovação de resultados.

Vale ressaltar que a escolha das técnicas de aprendizado de máquina, pré-processamento e validação de resultados foram escolhidos com o objetivo de complementar o trabalho de DANKOVICOVÁ et al(2018), uma vez que este estudou como detectar a disfonia vocal usando os classificadores Máquina de Vetores de Suporte (SVM), Classificador de Florestas Aleatórias (RFC) e K-Vizinhos Mais Próximos (KNN). O uso de um diferente conjunto de pré-processamento, aprendizado de máquina e validação de resultados visa procurar algum conjunto melhor para a detecção de disfonia vocal do que os apresentados por DANKOVICOVA et al(2018).

3.1 Coleta de dados

Os áudios foram extraídos do conjunto de dados de *Saarbruecken Voice Database*, mantida pela Universidade do Sarre, em Sarbruque na Alemanha. Estes áudios correspondem a arquivos WAV com sinais de fala provenientes de diversas pessoas, saudáveis e doentes, em diferentes períodos de tempo e diferentes tons e frequências. A quantidade total de áudios da base de dados é de 24.681, provenientes de mais de 2000 pessoas com diferentes patologias vocais, pois cada paciente possui vários áudios . Isto foi feito pois a base de dados possui áudios de uma mesma pessoa de diferentes formas, sendo normalmente feitas: Gravação de vogais (i, a, u) produzidas em diferentes tons(normal, agudo e grave), gravações das vogais (i, a, u) com tom ascendente-decrescente e gravações de frases.

Para cada paciente, há informações correspondentes a idade, sexo, data da gravação, etc... e, claro naturalmente, o seu estado de saúde. Assim, os áudios escolhidos são esses relacionados a um paciente que aparece no conjunto de dados.

Apesar da base de dados contar com 359 pacientes de disfonia vocal, não foi possível utilizar todos. Isto acontece pois alguns pacientes não possuem áudio ou existe uma variação na quantidade de áudios por pacientes, podendo ser entre 26 e 28, fazendo com que aqueles pacientes que tenham apenas 26 áudios tivessem

colunas preenchidas por Nan(Not A Number). Isto afeta diretamente os métodos de classificação, então optou-se por ignorar pacientes que tivessem apenas 26 áudios. O resultado final conta um quadro novo de 271 pacientes.

A etapa de coleta de dados iniciou-se carregando o arquivo csv da base de dados em forma de planilha, utilizando um quadro de dados do pandas, uma das bibliotecas utilizadas em Python e limpar todas as informações que não seria útil para o estudo, como por exemplo, a data, ID do paciente antigo, diagnósticos e doenças. O quadro de dados resultante foi composto por 4 colunas: ID, condição de saúde, gênero e idade. Após isso, iniciou-se a etapa de carregar os áudios dos pacientes para só assim iniciar a etapa de extração de características, descritas a seguir.

3.2 Métodos de extração de características com pré processamento.

Para poder processar qualquer informação no reconhecimento automático de fala, é necessário a extração de características. Isto é importante porque os sons geralmente são compostos com informações desnecessárias para análise, poluindo as informações importantes. Entre os principais tipos de poluição, podem ser descritos: Ruídos de fundo, emoções e os sons gerados por um ser humano devido ao trato vocal (incluindo dentes, língua, etc). Para tratar disso, foram aplicados os dois métodos de pré-processamento e de extração de características já descritas neste documento: O MFCC e o TEO.

Para ler os áudios e realizar dois tipos de extração de recursos de forma independente, foi criada a função em Python chamada “GettingSignalInformation (data, fextraction)”. O primeiro parâmetro é o quadro de dados de todos os pacientes e o segundo um número inteiro indicando se será usado o MFCC ou TEO. Esta função pega a lista de áudios de um mesmo paciente de um diretório correto (isto é possível pois cada arquivo começa com o ID do paciente). Então, para cada elemento da lista, o áudio é lido usando Librosa e chamado para a extração de recursos correspondente. Depois de trabalhar com todos os áudios de um paciente, um dict final é anexado a uma lista e no final todas essas informações são transformadas em dados de pandas frame.

A extração de características é dividida em mais duas funções, uma para MFCC e outra para TEO. Ambas as funções esperam como parâmetros o nome do sinal e das características, porém o MFCC também recebe o src entregue pela Librosa.

No caso do MFCC, o sinal é transformado em 13 features (as dimensões inferiores), pois representam o envelope dos espectros dos áudios, com o nome do parâmetro acrescentado um índice ao final. Tudo é salvo em um dict e devolvido. No caso da TEO, os passos são semelhantes, porém devido ao fato de trabalharmos com KNN e SVM, dimensões muito grandes podem ser um problema. Então reduziu-se para um total de 6 features, obtendo algumas medidas estatísticas, sendo elas: Média, variação padrão, maior valor, menor valor, maior e diferença menor entre recursos sequenciais.

Depois deste processo, foi necessário tratar os problemas de pré-processamento. Iniciando com o problema de colunas serem preenchidas por Nan(Not A Number), que já foi descrita na parte 3.1 deste trabalho. Além disso, foi necessário também codificar variáveis categóricas, uma vez que muitos algoritmos de aprendizado de máquina não podem trabalhar diretamente com dados categóricos, sendo necessário converter para números. Assim, no processo de tratamento de colunas Nan os dados foram divididos em duas etapas, uma para variáveis dependentes e outra para variáveis independentes.

É importante notar que a quantidade de pessoas saudáveis e não saudáveis não era proporcional quando retiradas no banco de dados. Após tratar o problema de colunas serem preenchidas por Nan, o número de pessoas saudáveis era de 211 enquanto de pessoas não saudáveis era de 60.

Este desequilíbrio na quantidade de pessoas saudáveis e não saudáveis causa o problema de muitos algoritmos de aprendizado de máquina ignorarem a amostragem menor. Uma forma de contornar esse problema é por meio da reamostragem, que consiste em adicionar cópias de instâncias da classe sub-representada, chamada over-sampling, ou excluir instâncias da classe super-representada, chamada de subamostragem. Nesse caso, como não há muitos dados, a melhor opção foi o oversampling. A fraqueza de fazer isso aleatoriamente é que ele duplica registros aleatórios da classe minoritária, o que pode causar excesso de

arquivos duplicados em uma mesma amostragem. Para tratar tal problema foi utilizado a classe SMOTE(Synthetic Minority Over-sampling Technique), uma vez que ela funciona selecionando exemplos que estão próximos no espaço de recursos, desenhando uma linha entre os exemplos no espaço de recursos e desenhando uma nova amostra em um ponto ao longo dessa linha. Após a aplicação da sobreamostragem, o número de pacientes não saudáveis aumentou para 211, a mesma quantidade de pacientes saudáveis.

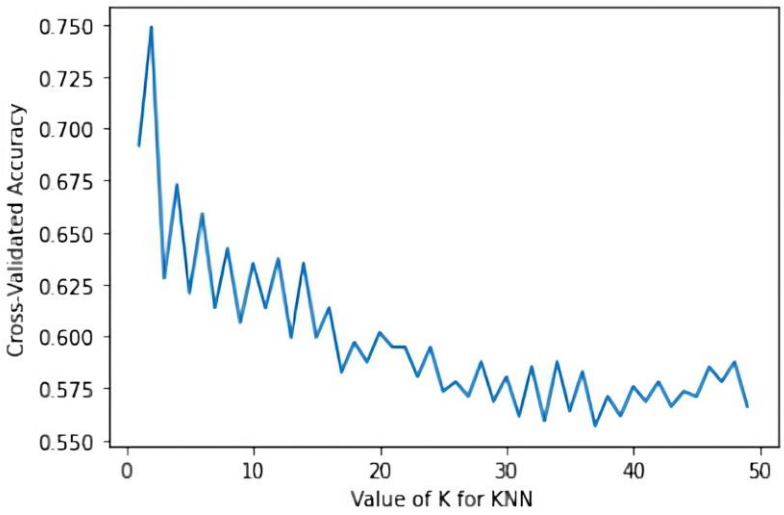
3.3 Métodos de aprendizado de máquina com comprovação de resultados

Após a etapa de coleta de dados e pré processamento, inicia-se a etapa de aprendizado de máquina junto com a comprovação de resultados. Aqui existe uma combinação entre utilizar algum método de extração de característica(MFFC ou TEO) junto com um modelo de aprendizado de máquina(SVM ou KNN) e comprovação de resultado(Validação Cruzada e/ou matriz de confusão). Entre as combinações utilizadas, estão elas: MFCC com KNN, MFCC com SVM kernel linear, MFCC com RBF Kernel, TEO com KNN, TEO com SVM kernel linear e TEO com RBF Kernel. É importante notar que as funções utilizadas pertencem a bibliotecas em Python já existentes, como Pandas e Librosa.

3.3.1 MFCC com KNN

Na combinação MFCC com KNN, é necessário encontrar qual seria o melhor K para poder utilizar o algoritmo de KNN, desta forma, foi testada a diferentes valores de K entre 1 e 50 utilizando Validação Cruzada e foi notável que os melhores valores estão entre 1 e 5, sendo o melhor de todos para K o número 2. A Figura a seguir mostra a acurácia da Validação Cruzada em relação ao valor de K.

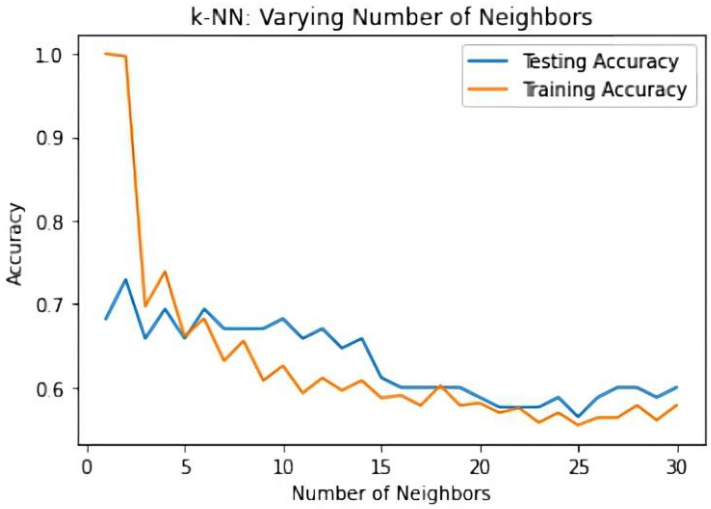
Figura 8- Acurácia do valor de K em MFCC com KNN



Fonte: Confeccionado pelo autor.

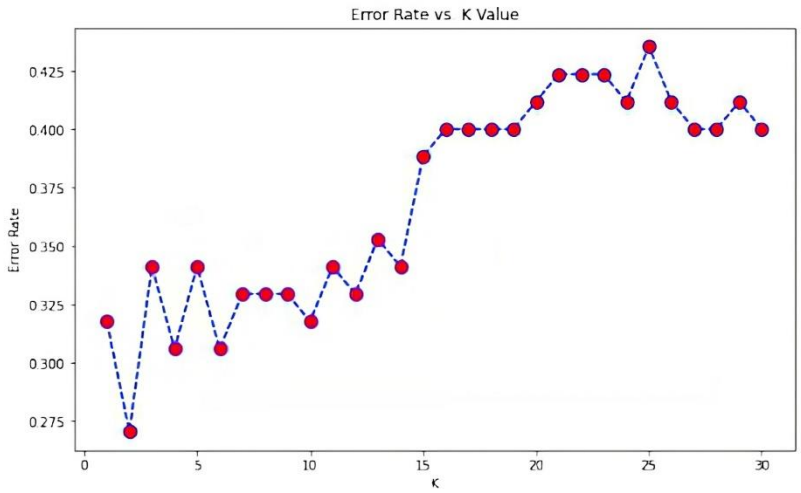
Uma vez que percebeu-se que K com valor 2 era o melhor valor, usou-se a Matriz de Confusão com uma divisão aleatória. Para avaliar melhor este modelo, realizou-se uma Validação Cruzada, dividindo os dados entre dados de teste e dados de treinamento, a fim de procurar o melhor número de vizinhos em relação a acurácia. As Figuras a seguir mostram os gráficos gerados em relação a acurácia com o melhor número de vizinhos no modelo KNN e a taxa de erro em relação ao valor de K.

Figura 9- Acurácia de K em MFCC com KNN



Fonte: Confeccionado pelo autor.

Figura 10- Taxa de erro no valor de K em MFCC com KNN

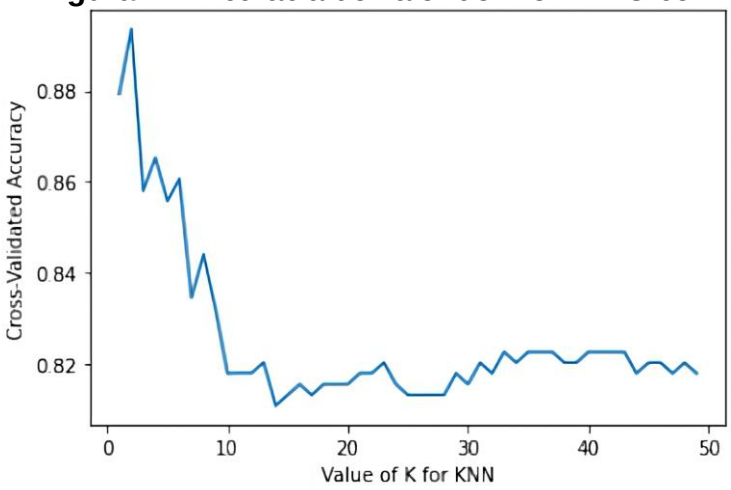


Fonte: Confeccionado pelo autor.

3.3.2 TEO com KNN

Na combinação TEO com KNN também foi testada a diferentes valores de K entre 1 e 50 e foi que o melhor valor para K também é o número 2. A Figura a seguir mostra a acurácia da Validação Cruzada em relação ao valor de K..

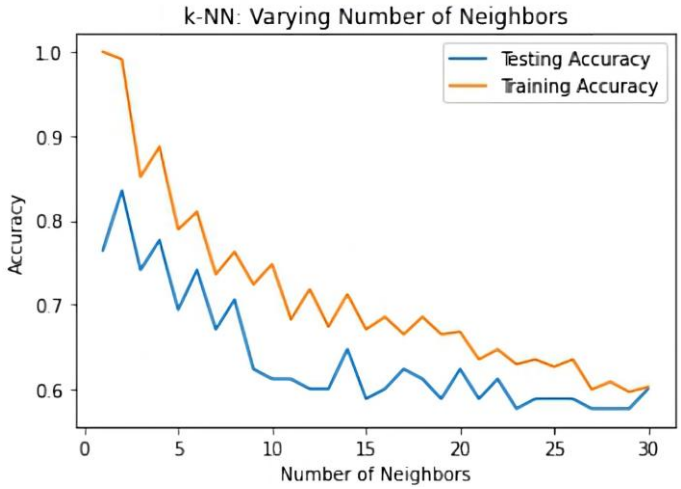
Figura 11- Acurácia do valor de K em TEO com KNN



Fonte: Confeccionado pelo autor.

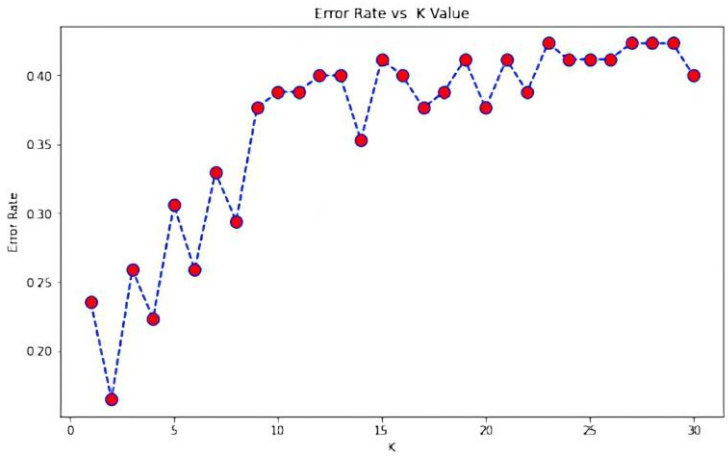
Uma vez que percebeu-se que K com valor 2 também era o melhor valor a ser usado, usou-se a Matriz de Confusão com uma divisão aleatória. Assim como no método MFCC com KNN, realizou-se Validação Cruzada holdout, dividindo os dados entre dados de teste e dados de treinamento, a fim de procurar o melhor número de vizinhos em relação a acurácia. As Figuras a seguir mostram os gráficos gerados em relação a acurácia com o melhor número de vizinhos no modelo KNN e a taxa de erro em relação ao valor de K.

Figura 12- Acurácia de K em TEO com KNN



Fonte: Confeccionado pelo autor.

Figura 13- Taxa de erro no valor de K em TEO com KNN



Fonte: Confeccionado pelo autor.

3.3.3 SVM

Diferente do algoritmo de KNN, em que existe uma necessidade de encontrar o melhor K para poder utilizar o algoritmo, o algoritmo de SVM não tem esta necessidade. Desta forma, bastou importar sua função após as tratativas feitas pelos métodos de extração de características e pré processamento. Vale lembrar que no caso do SVM, o método de comprovação de resultados utilizado foi a matriz de confusão.

4 RESULTADOS

Neste capítulo, é disposto em forma de tabela a comparação de acurácia média, melhor acurácia e variância com todos com as combinações de aprendizado de máquina (SVM Linear, SVM RBF e KNN), métodos de extração de características (MFCC e TEO) e métodos de comprovação de resultado (Matrizes de Confusão e Validação Cruzada) apresentados no capítulo 3 deste trabalho. Em cada combinação, foram feitas 10000 iterações.

4.1 Resultados Gerais

Como mostrado na Tabela a seguir, ao utilizar MFCC em comparação com o KNN, o modelo SVM apresentou melhoras de cerca de 20% e 22% para o modelo Kernel Linear e RBF Kernel, respectivamente. O mesmo acontece ao utilizar TEO, apresentando melhora de cerca de 21% e 19% para o modelo Kernel Linear e RBF Kernel,respectivamente.

Tabela 1- Acurácia geral de cada modelo

	Acurácia Média	Melhor Acurácia	Variância
MFCC com KNN	71%	90%	19%
TEO com KNN	81%	96%	15%
MFCC com SVM Linear	95%	100%	5%
MFCC com SVM RBF	93%	100%	7%
TEO com SVM Linear	92%	99%	70%
TEO com SVM RBF	90%	99%	9%

Fonte: Confeccionado pelo autor.

Assim como visto no estado da arte, estes resultados já eram esperados, não fugindo muito das pesquisas relacionadas por outros autores citados neste trabalho. Porém vale notar que o MFCC com SVM RBF possui uma acurácia média (93%)

melhor do que apresentada por DANKOVICOVA et al (2018), cujos resultados mostrados chegaram a 91,3%

5 CONCLUSÃO

Este trabalho teve como principal objetivo propor um método computacional capaz de detectar a disfonia funcional vocal nos pacientes de forma não-invasiva, utilizando áudios da voz de pacientes sadios e não sadios, computação e inteligência artificial. Entre os modelos estudados, destaca-se que aqueles que usavam o aprendizado de máquina SVM, tanto Kernel Linear quanto o RBF, obtiveram uma acurácia média melhor (95% com MFCC) em relação ao uso do aprendizado de máquina KNN, que teve a pior acurácia média deste estudo (71% com MFCC). Este trabalho também complementou o estudo de outros autores, uma vez que usou uma combinação de aprendizado de máquina, extração de características e comprovação de resultados em uma patologia que nem sempre é estudada por outros autores, e quando estudado, não seguia estas combinações.

Como perspectiva de futuro, existem ainda muitos métodos de aprendizado de máquina que precisam ser estudados e testados para a detecção de patologias vocais de forma não-invasiva, uma vez que não se sabe ao certo qual é o melhor método para cada patologia vocal. É necessário testar diferentes combinações de aprendizado de máquina, extração de características e métodos de comprovação para diferentes patologias vocais, para assim agregar mais conhecimento nesta área da computação.

REFERÊNCIAS

ROY, N. Functional dysphonia, Current Opinion in Otolaryngology & Head and Neck Surgery: June 2003 - Volume 11 - Issue 3 - p 144-148.

BEHLAU M.; MADAZIO G.; OLIVEIRA G. Functional dysphonia: strategies to improve patient outcomes. Patient Relat Outcome Meas.doi: 10.2147/PROM.S68631. PMID: 26664248; PMCID: PMC4671799. 2015, Dec.

HEGDE, S.; SHETTY, S., RAI, S.; & DODDERI, T. A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders. Journal of Voice. doi:10.1016/j.jvoice.2018.07.014. 2018.

BENBA, A.; JILBAB, A.; HAMMOUCH, A. Detecting Patients with Parkinson's disease using Mel Frequency Cepstral Coefficients and Support Vector Machines. International Journal on Electrical Engineering and Informatics. Volume 7. 297-307. 10.15676/ijeei.2015.7.2.10. 2015

JEANCOLAS, L.; BENALI, H.; BENKELFAT, B.-E.; MANGONE, G.; CORVOL, J.-C.; VIDAIHET, M.; … PETROVSKA-DELACRETAZ, D. (2017). Automatic detection of early stages of Parkinson’s disease through acoustic voice analysis with mel-frequency cepstral coefficients. International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). doi:10.1109/atsip.2017.8075567. 2017

DANKOVICOVÁ, Z.; SOVÁK, D.; DROTÁR, P.; VOKOROKOS, L. Machine Learning Approach to Dysphonia Detection.Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, 040 01 Košice, Slovakia. doi:10.3390/app8101927. 2018.

SAMA, A.; CARDING, P. N.; PRICE, S.; KELLY, P.; WILSON, J. A. . The Clinical Features of Functional Dysphonia. The Laryngoscope, 111(3), 458–463. doi:10.1097/00005537-200103000-00015. 2001.

PENTEADO, R. Z. Relações entre saúde e trabalho docente: percepções de professores sobre saúde vocal.Rev Soc Bras Fonoaudiol. doi: 10.1590/S1516-80342007000100005 . 2007.

Saarbruecken Voice Database. Instituto de Fonetica, Universidade do Sarre, Alemanha. Disponível em: <http://stimmdb.coli.uni-saarland.de/index.php4>. Acesso em: 22 jul. 2022.

NOBLE, W. S. What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. doi:10.1038/nbt1206-1565. 2006

GOSH, S.; DASGUPTA, A.; SWETAPADMA, A.. A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. 2019 International Conference on Intelligent Sustainable Systems (ICISS). doi:10.1109/iss1.2019.8908018. 2019.

GOYAL, J.; KHANDNOR, P.; ASERI, T. C. A Comparative Analysis of Machine Learning classifiers for Dysphonia-based classification of Parkinson's Disease. International Journal of Data Science and Analytics. doi:10.1007/s41060-020-00234-0. 2020.

JOSHI, S.; CHEERAN, A.N. . MATLAB Based Feature Extraction Using MFCC for ASR. 2014.

SUBASI, A.; YILMAZ, A.; TUFAN, K. Detection of generated and measured transient power quality events using teager energy operator. Energy Conversion and Management, v. 52, p.1959–1967, 04. 2011.

GUIDO, R. C. Enhancing Teager Energy Operator Based on a Novel and Appealing Concept: Signal Mass. Journal of the Franklin Institute. doi:10.1016/j.jfranklin.2018.12.007. 2019.

DÜNTSCH, I.; GEDIGA, G.Confusion Matrices and Rough Set Data Analysis. J. Phys.: Conf. Ser. 1229 012055. 2019.

BERRE, C. L.; SANDBORN, W. J.; ARIDHI, S.; DEVIGNES, M.-D.; FOURNIER, L.; SMAÏL, M. T.; … PEYRIN-BIROULET, L. (2019). Application of Artificial Intelligence to Gastroenterology and Hepatology.Gastroenterology.doi:10.1053/j.gastro.2019.08.058

GRANHOLM, V.; NOBLE, W.; KÄLL, L. A cross-validation scheme for machine learning algorithms in shotgun proteomics. BMC Bioinformatics, 13(Suppl 16), S3. doi:10.1186/1471-2105-13-s16-s3.2012.

ISLAM, R.; TARIQUE, M.; ABDEL-RAHEEM, E. A Survey on Signal Processing Based Pathological Voice Detection Techniques. IEEE Access, [s. l.], v. 8, p. 66749–66776,. Disponível em: <https://ieeexplore.ieee.org/document/9055386/>. 2020Acesso em: 22 jul.2022.