

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Jefferson Silva Almeida

**Deteccção da Doença de Parkinson a partir de Exames de Fonação
Sustentada e Sinais de Voz utilizando técnicas de Aprendizagem de
Máquina**

Fortaleza – CE

13 de setembro de 2021

Jefferson Silva Almeida

Detecção da Doença de Parkinson a partir de Exames de Fonação Sustentada e Sinais de Voz utilizando técnicas de Aprendizagem de Máquina

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Processamento de Imagens e Sinais

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

13 de setembro de 2021

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

- A447d Almeida, Jefferson Silva.
Detecção da Doença de Parkinson a partir de Exames de Fonação Sustentada e Sinais de Voz utilizando técnicas de Aprendizagem de Máquina / Jefferson Silva Almeida. - 2021.
76 f. : il. color.
- Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2021.
Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.
1. Doença de Parkinson. 2. Processamento de voz. 3. Extração de características. 4. Aprendizado de Máquina. 5. Auxílio ao diagnóstico médico. I. Título.

CDD 004



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Jefferson Silva Almeida

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

**Orientador: Prof. Dr. Pedro Pedrosa Rebouças
Filho**

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Prof. Dr. Antonio Carlos da Silva Barros
Universidade da Integração Internacional da Lusofonia
Afro-Brasileira (UNILAB)

**Prof. Dr. Carlos Mauricio Jaborandy de Mattos
Dourado Junior**
Instituto Federal do Ceará - Campus Fortaleza (IFCE)

Fortaleza – CE
13 de setembro de 2021

*Este trabalho é dedicado a todas as pessoas que,
de alguma forma, tem contribuído para a existência de um mundo melhor,
baseado no ensino e na pesquisa científica.*

Agradecimentos

Agradeço em primeiro lugar a Deus, que iluminou meu caminho durante estes anos de estudo.

Ao meu orientador e amigo Prof. Dr. Pedro Pedrosa, pelo conhecimento compartilhado, apoio, e por acreditar no meu potencial como aluno e pesquisador.

Aos meus familiares pelo apoio e compreensão nas horas em que estive ausente, em especial a meu avô José Almeida que percebeu potencial em mim desde o nascimento, a minha avó Vera Lúcia pelo cuidar nas horas que precisei, a minha mãe Verônica pelo carinho, ao meu irmão Wesley pela companhia e apoio e a minha amada Juliana.

Aos professores que me ajudaram fornecendo a base de dados utilizada nesta pesquisa.

Ao IFCE pelo acolhimento e oportunidades, aos professores e amigos que fiz durante a vida acadêmica.

A FUNCAP pelo apoio financeiro aos meus estudos e ao LAPISCO a qual sou integrante a vários anos.

*“Dê instrução ao sábio, e ele se tornará mais sábio ainda;
ensine o justo, e ele crescerá na prudência.”
(Provérbios 9:9)*

Resumo

A doença de Parkinson é um dos distúrbios neurológicos que mais afetam as pessoas no mundo. Este estudo investiga o processamento de sinais de voz para detectar a doença de Parkinson. A metodologia aplicada avalia dezoito técnicas de extração de características e quatro métodos de aprendizado de máquina para classificar dados brutos obtidos a partir de exames de fonação sustentada e fala. A fonação sustentada relaciona-se com a tarefa de vogal /a/ vocalizada e fala com a pronúncia de uma sentença curta na língua lituana. Durante a gravação dos exames de fala utilizou-se áudio estéreo com microfones do tipo profissional acústico cardióide (AC) e um smartphone (SP). Desta forma, pode-se avaliar o desempenho utilizando diferentes qualidades de microfone e se existe a possibilidade de embarcar os algoritmos em um aplicativo para realização dos exames em smartphone. Cinco medidas foram empregadas para analisar o desempenho do classificador: Medidas de Taxa de Erro (EER), Área Sobre Curva (AUC) das Curvas de Detecção de Erro (DET) e Característica de Operação do Receptor (ROC), Acurácia, Especificidade e Sensibilidade. O melhor desempenho para o canal AC obteve acurácia de 94,55%, AUC 87,84% e EER 19,01%. Ao usar o canal SP, obtivemos acurácia de 92,94%, AUC 92,40% e EER 14,15%. Os resultados mostraram que não há diferença significativa entre os microfones de áudio cardióide (AC) e do smartphone (SP), apoiando assim a ideia de que a doença de Parkinson pode ser detectada processando no smartphone. O áudio de Fonação (P) forneceu resultados significativamente melhores do que os de Fala (S), Dublado (V) e Mudo (U). Com base nos resultados apresentados, pode-se afirmar que a tarefa de fonação é a mais adequada para uso em sistemas de auxílio ao diagnóstico da doença de Parkinson. Além de ser aquela que permitiu obter o melhor desempenho na etapa de classificação nas modalidades de microfones AC e SP.

Palavras-chave: Doença de Parkinson; Processamento de voz; Características fonológicas; Extração de características; Aprendizagem de Máquina; Auxílio ao diagnóstico médico.

Abstract

Parkinson's disease is one of the neurological disorders that most affect the people in the world. This study investigates the processing of voice signals for detecting the disease. The methodology applied uses eighteen feature extraction techniques and four machine learning methods to classify raw data obtained from sustained phonation and speech tasks. Phonation relates to the vowel /a/ voicing task and speech to the pronunciation of a short sentence in the Lithuanian language. The audio tasks recording using two microphone channels from acoustic cardioid (AC) and a smartphone (SP). In this way we can evaluate the performance and if there is the possibility to use the application on the smartphone. Five measures were employed to analyze the classifier performance: Equal Error Rate (EER) and Area Under Curve (AUC) measures from Detection Error Tradeoff (DET) and Receiver Operating Characteristic (ROC) curves, Accuracy, Specificity, and Sensitivity. The best performance for AC channel achieved an accuracy of 94.55%, AUC 87.84%, and EER 19.01%. When using the SP channel, we have achieved an accuracy of 92.94%, AUC 92.40%, and EER 14.15%. The results showed that there is no significant difference between acoustic cardioid (AC) and smartphone (SP) microphones, thus supporting the idea that Parkinson's disease can be detected by processing on the smartphone. Phonation (P) audio provided significantly better results than Speech (S), Voiced (V) and Unvoiced (U). Based on the results presented, we can say that the task of phonation is the most suitable for use. In addition to being the one that allowed to obtain the best performance in the modalities of AC and SP microphones.

Keywords: Parkinson's disease. Speech processing. Phonological features. Feature extraction. Machine Learning. Diagnosis.

Lista de ilustrações

Figura 1 – Diagrama ilustrativo da área afetada pela doença de Parkinson; denominada Mesencéfalo, região onde é produzida a Dopamina; região chamada de substância negra.	23
Figura 2 – Diagrama ilustrativo dos estágios para o diagnóstico da doença de Parkinson.	24
Figura 3 – Breve histórico de estudos relacionados a doença de Parkinson e a análise de sinais de voz para detecção da doença.	26
Figura 4 – Órgãos de produção da fala.	31
Figura 5 – Arquitetura gráfica de um perceptron multicamadas com uma camada escondida.	41
Figura 6 – Funcionamento do classificador OPF: (a) Conjunto de treinamento modelado com grafo completo (b) Computação de uma árvore de expansão mínima sobre o conjunto de treinamento (c) Floresta de caminho ótimo encontrada sobre o conjunto de treinamento, os protótipos são destacados (d) Processo de classificação de uma amostra (e) A amostra de teste é finalmente classificada.	41
Figura 7 – Arquitetura de uma Support Vector Machine usando kernel Radial Basis Function, RBF. Adaptado de (VAPNIK; CHERVONENKIS, 2015). . .	42
Figura 8 – Descrição das etapas seguidas na abordagem proposta nesta dissertação.	45
Figura 9 – Visualização de dados pelo t-SNE.: (a) AC - acústico cardíóide, fonação e extrator YA; e (b) SP-smartphone, fonação e extrator KT. Gravação de: PD - Pacientes com doença de Parkinson e HC - indivíduo saudável.	51

Lista de tabelas

Tabela 1 – Extratores de características utilizados, em que o número de atributos é especificado, além da referência na literatura e sigla adotada neste trabalho. Itens: (1-12) biblioteca OpenSmile; (13) Descritor Essentia; (14) Biblioteca JAudio; (15) Biblioteca KTU; (16) Descritor MPEG7; (17) Biblioteca Tsanas; e (18) Biblioteca YAAFE.	39
Tabela 2 – Classificadores utilizados, em que a configuração é especificada, além da referência na literatura e sigla adotada neste trabalho. Itens: (1-3) k Nearest Neighbors; (4) Multilayer Perceptron; (5-14) Optimum Path Forest; e (15-17) Support Vector Machine.	43
Tabela 3 – Número de sujeitos no banco de dados utilizado neste trabalho.	47
Tabela 4 – Passo 1a - Escolha do extrator e classificador com menor EER (%), . . .	52
Tabela 5 – Passo 1b - Escolha do extrator e classificador com menor EER (%), . . .	53
Tabela 6 – Passo 1c - Escolha do extrator e classificador com menor EER (%), . . .	54
Tabela 7 – Passo 2a - Escolha do extrator e classificador com maior acurácia (%), . . .	55
Tabela 8 – Passo 2b - Escolha do extrator e classificador com maior acurácia (%), . . .	56
Tabela 9 – Passo 2c - Escolha do extrator e classificador com maior acurácia (%), . . .	57
Tabela 10 – Passo 2d - Escolha do extrator e classificador com maior acurácia (%), . . .	58
Tabela 11 – Passo 2e - Escolha do extrator e classificador com maior acurácia (%), . . .	59
Tabela 12 – Passo 2f - Escolha do extrator e classificador com maior acurácia (%), . . .	60
Tabela 13 – Passo 2g - Escolha do extrator e classificador com maior acurácia (%), . . .	61
Tabela 14 – Passo 2h - Escolha do extrator e classificador com maior acurácia (%), . . .	62
Tabela 15 – Passo 3 - Resultado final e comparação com outra pesquisa,	63
Tabela 16 – Matriz de confusão de resultados de classificação usando métodos de melhor desempenho: (a) AC obtido com P/YA e classificador K1; (b) SP obtido com P/KT classificador K1.	63

Lista de abreviaturas e siglas

PD	Paciente com doença de Parkinson
DP	Doença de Parkinson
HC	Pessoa saudável
AC	Microfone acústico cardíóide
SP	Microfone do <i>smartphone</i>
EER	Taxa igual do erro
AUC	Área sobre a curva
DET	Curvas de erro <i>trade-off</i>
ROC	Curva característica de operação do receptor
P	Modalidade de áudio na tarefa vocal de Fonação
S	Modalidade de áudio de Fala
V	Modalidade de áudio dublado
U	Modalidade de áudio mudo
OMS	Organização Mundial de Saúde

Sumário

1	INTRODUÇÃO	23
1.1	Motivação	25
1.2	Estado da arte	25
1.3	Objetivos	27
1.3.1	Objetivos específicos	27
1.4	Contribuições	27
1.5	Produção científica	27
1.6	Estrutura da Dissertação	29
2	FUNDAMENTAÇÃO TEÓRICA	31
2.1	O mecanismo da fala	31
2.2	Características relevantes da fala	32
2.2.1	Força	33
2.2.2	Velocidade	33
2.2.3	Amplitude	34
2.2.4	Estabilidade	34
2.2.5	Tom	35
2.2.6	Precisão	35
2.3	Extração de Características	35
2.3.1	OpenSMILE	36
2.3.2	Essentia	36
2.3.3	jAudio	37
2.3.4	KTU	37
2.3.5	MPEG7	37
2.3.6	Tsanas	38
2.3.7	YAAFE	38
2.4	Classificadores	39
2.4.1	<i>kNearest Neighbors</i>	40
2.4.2	Perceptron de Múltiplas Camadas	40
2.4.3	<i>Optimum Path Forest</i>	41
2.4.4	<i>Support Vector Machines</i>	42
2.5	Conclusão do Capítulo	42
3	METODOLOGIA	45
3.1	Ideia geral	45
3.2	Base de dados <i>crossourced</i>	46

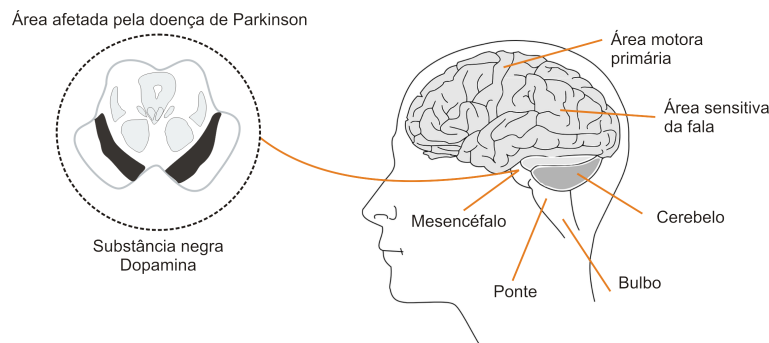
3.3	Métricas de Avaliação	47
4	RESULTADOS E DISCUSSÃO	49
5	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS . .	65
6	DECLARAÇÃO ÉTICA	67
	REFERÊNCIAS	69

1 Introdução

A Doença de Parkinson (DP), foi inicialmente chamada Paralisia do Tremor e descrita pela primeira vez em 1817 por James Parkinson (PARKINSON, 1817). Esta é uma doença neurológica e degenerativa caracterizada pela deficiência na produção de um neurotransmissor essencial ao funcionamento normal do cérebro chamado: Dopamina.

A Dopamina é produzida em uma região específica do cérebro conhecida por Mesencéfalo e fica localizada próxima ao cerebelo e o tronco encefálico. Este neurotransmissor é produzido na região chamada de substância negra, justamente por conter esta coloração escura. Mostramos na Figura 1 a região do cérebro afetada pela doença. Por meio de uma visualização em corte transversal, pode-se ver em destaque as duas regiões negras, onde é produzida o neurotransmissor.

Figura 1 – Diagrama ilustrativo da área afetada pela doença de Parkinson; denominada Mesencéfalo, região onde é produzida a Dopamina; região chamada de substância negra.



Fonte: Próprio autor.

Os neurotransmissores são substâncias químicas produzidas pelos neurônios, especificamente por células do sistema nervoso. Por meio delas, a comunicação entre as células podem ser estabelecidas através de impulsos elétricos.

A deficiência de dopamina é causada pela degeneração dos neurônios na substância negra e pode estar relacionada aos fatores idade, histórico familiar, traumatismo craniano e até mesmo devido ao contato com alguns pesticidas.

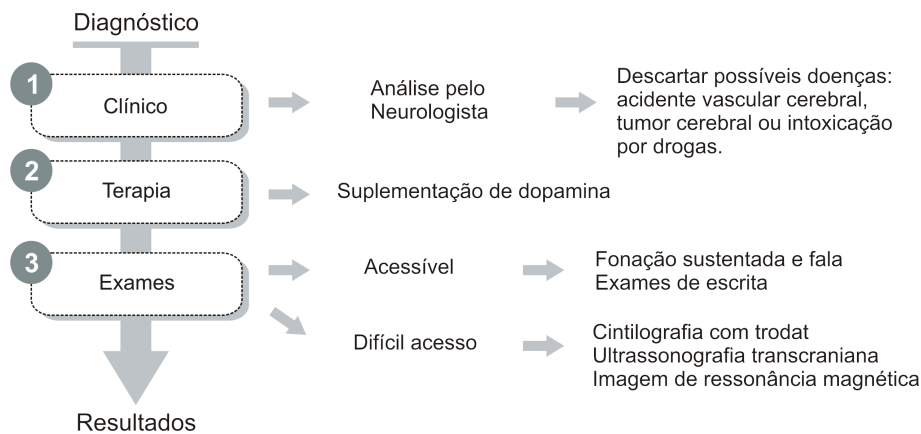
A dopamina é um importante neurotransmissor, porque auxilia os movimentos voluntários. Sua redução abaixo dos níveis considerados normais causa diversos sintomas que são classificados em dois grupos: motores e não-motores.

Os sintomas de DP relacionados ao primeiro grupo incluem movimentos de tremor involuntários, fraqueza muscular que dificulta a realização de atividades simples como por

exemplo, abotoar uma camisa, calçar os sapatos (BARBAGALLO et al., 2018). Além da rigidez muscular que devido à falta de dopamina no corpo, os músculos não recebem os sinais para relaxar. Outros sintomas manifestados são alterações na fala, na caligrafia e perda da expressão facial. No segundo grupo os sintomas podem manifestar-se com demência, depressão, ansiedade, distúrbios do sono e raciocínio lento (SHULMAN et al., 2002).

A Figura 2 mostra os estágios necessários para realizar o diagnóstico da DP. Ao todo são três etapas, iniciando (1) com a análise clínica realizada pelo neurologista. Desta forma é possível descartar outras doenças que apresentam sintomas semelhantes.

Figura 2 – Diagrama ilustrativo dos estágios para o diagnóstico da doença de Parkinson.



Fonte: Próprio autor.

Após a confirmação dos sintomas, o segundo passo (2) é a realização de terapia com uso de medicamentos (GARCIA-MONTES; BORONAT-GARCIA; DRUCKER-COLÍN, 2012), na qual o paciente com suspeita de DP é submetido à suplementação de dopamina. Se houver melhorias, há uma alta probabilidade do indivíduo ter a doença. No entanto em caso de dúvidas um terceiro passo é necessário para confirmar a doença. Nesta etapa (3) são realizados testes laboratoriais, que podem não ser acessíveis a todas as pessoas, especialmente nos países em desenvolvimento.

Os testes laboratoriais de voz e caligrafia estão entre os mais acessíveis. Justamente porque dispensam o uso de equipamentos relativamente caros, como por exemplo scanners de ressonância magnética. Testes de voz e caligrafia são relevantes porque são aplicados com base nos primeiros sintomas que surgem no indivíduo com suspeita de DP.

Os testes de caligrafia podem ser conduzidos em papel com o auxílio por sistemas de visão computacional (SOUZA et al., 2018; PEREIRA et al., 2016). Testes com toque de dedo podem ser empregados para avaliar sintomas de tremor, como para pacientes com doença de Huntington, que tem sintomas semelhantes aos da DP (LAURAITIS; MASKELIŪNAS;

DAMAŠEVIČIUS, 2018). Os testes vocais contemplam os sinais sustentados de fonação e fala, e tem-se mostrado uma área promissora para realização de pesquisa científica.

1.1 Motivação

Recentemente, a Organização Mundial de Saúde (OMS) descreveu as desordens neurológicas como uma das mais significantes ameaças à saúde pública. Entre as desordens neurológicas mais comuns está a Doença de Parkinson (DP), acidente vascular cerebral, esclerose múltipla, cefaléia, demência, epilepsia, entre várias outras doenças. Estima-se que neste momento a cada 60 pessoas, 16 estão sofrendo devido a alguma doença neurológica (WHO, 2006).

Outro fato a considerar é que a perda de neurônios dopaminérgicos evolui rapidamente, atingindo cerca de 50% no momento do diagnóstico clínico, e pode progredir rapidamente ao longo de 5 anos (FEARNLEY; LEES, 1991). Por esse motivo, é imprescindível o diagnóstico em tempo ágil para que seja iniciado o tratamento da doença. É preciso oferecer alternativas a realização de exames mais complexos, porque muitas vezes o paciente não possui recursos financeiros para custear estes exames e enfrenta longas filas à espera de uma vaga no sistema único de saúde pública.

Portanto, ter mais testes disponíveis é de grande importância, pois permite agilizar o processo de diagnóstico e tratamento (YADAV; LI, 2015).

No campo da Ciência da Computação almejamos o estudo e compreensão de novos extratores de características e classificadores, avaliando o desempenho dos mesmos nos exames de voz. Além de contribuir com a análise de diferentes qualidades de microfone, bem como o processamento realizado com base em diferentes tipos de áudio.

1.2 Estado da arte

Os estudos iniciais envolvendo sinais de voz para detecção de DP ocorreram em 1978, quando dois ouvintes experientes usaram gravações em fita de alta fidelidade de amostras de fala de 200 pacientes, ambos saudáveis e com DP (LOGEMANN et al., 1978).

Em 1998, descobriu-se que a voz é o déficit primário, mais frequentemente afetado nos estágios iniciais da DP, quando comparado à articulação e fluência (HO et al., 1999). E o mais importante é que segundo Harel, Cannizzaro e Snyder (2004), a doença de Parkinson pode ser detectada em sua fase incipiente.

Técnicas estatísticas foram utilizadas para avaliar a autocorrelação do sinal residual em pacientes com doença de Parkinson e foram identificadas diferenças significativas na amplitude dos passos entre os pacientes com DP e o grupo saudável (PEREIRA et al.,

2006).

Uma medida de disfonia⁰ foi proposta para discriminar pessoas com DP e pessoas saudáveis a partir de variações na frequência da voz (LITTLE et al., 2009).

Em 2013, descobriu-se que as vogais sustentadas, palavras e sentenças de um conjunto de tarefas de fala carregam informações discriminativas de DP que podem ser utilizadas usando ferramentas de aprendizado de máquina (SAKAR et al., 2013).

Os trabalhos utilizando sinal de voz continuaram nos anos seguintes com a contribuição de vários autores (TSANAS et al., 2012; ZHAO et al., 2014; VERIKAS et al., 2015; UPADHYA; CHEERAN, 2018; CERNAK et al., 2017).

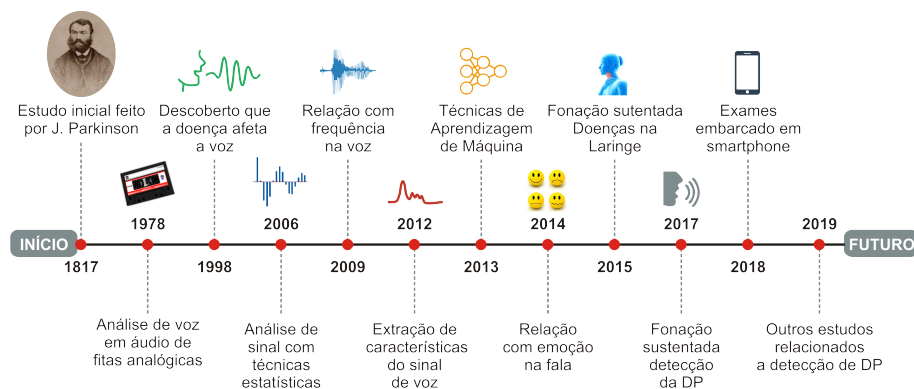
As tendências mais notáveis são o uso de tecnologia de smartphone para gravação de amostras de fala no cotidiano para capturar os sinais de comprometimento da fala em pessoas (RUSZ et al., 2018), e o desenvolvimento de software para análise de sinais de fala patológicos considerando fonação, articulação, prosódia e inteligibilidade (OROZCO-ARROYAVE et al., 2018).

O desenvolvimento de métodos eficazes de extração de características abre caminho para a análise de *big data* usando dados *crowdsourced*¹ de usuários de smartphones em todo o mundo (KE et al., 2018). Porém será preciso antes desenvolver métodos capaz de detectar a doença, independente da linguagem falada pelo paciente com DP.

Por outro lado, métodos de inteligência computacional baseados em heurística combinados com novas arquiteturas de redes neurais prometem aumentar a precisão da classificação em sistemas de suporte à decisão para aplicações biomédicas (WOŹNIAK; POŁAP, 2017; WOŹNIAK; POŁAP, 2018).

A Figura 3 apresenta um resumo de trabalhos relacionados a este tema.

Figura 3 – Breve histórico de estudos relacionados a doença de Parkinson e a análise de sinais de voz para detecção da doença.



Fonte: Próprio autor.

⁰ Disfonia: alteração ou enfraquecimento da voz.

¹ *Crowdsourced*: dados compartilhados por cooperação internacional.

1.3 Objetivos

O principal objetivo desta dissertação é avaliar métodos de extração e classificação para detectar a Doença de Parkinson em indivíduos por meio de exames de voz.

1.3.1 Objetivos específicos

Os objetivos específicos que devem ser alcançados neste trabalho, são:

- Propor uma metodologia viável para detecção da doença;
- Utilizar um dataset novo recentemente utilizado na literatura;
- Analisar a performance de diferentes tarefas vocais;
- Aplicar diferentes extratores de características;
- Avaliar o desempenho de diferentes classificadores;
- Investigar o desempenho com diferentes microfones;
- Apresentar uma solução eficiente para solucionar o problema;
- Em Ciência da Computação, avaliar novos extratores e classificadores;

1.4 Contribuições

Neste trabalho estamos utilizando a base de dados proposta por Vaiciukynas *et al.* (VAICIUKYNAS *et al.*, 2017) com os mesmos conjuntos de recursos de áudio. A principal contribuição de nossa abordagem avalia o desempenho com o uso de outros classificadores e o processamento de áudio sem a fusão de conjuntos de recursos de várias modalidades. Seguindo essa ideia, buscamos uma metodologia simplificada e objetiva para detectar a doença.

As principais contribuições desta dissertação são:

- A avaliação de diferentes classificadores em relação ao encontrado na literatura utilizando esta base de dados;
- Variedade de opções para embarcar o software em smartphone e/ou computadores portáteis considerando as combinações extrator e classificador;
- Apresentação de uma abordagem eficaz para detecção da doença;

1.5 Produção científica

Durante esta pesquisa de mestrado, os seguintes trabalhos científicos foram aceitos e publicados, a saber:

Artigos publicados em revistas:

- **ALMEIDA, J. S.**; REBOUÇAS FILHO, P. P.; CARNEIRO, TIAGO; WEI, WEI; DAMAŠEVIČIUS; MASKELIŪNAS, RYTIS; DE ALBUQUERQUE, VICTOR HUGO C.. **Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques.** Pattern Recognition Letters, v. 125, p. 55-62, 2019.
- **ALMEIDA, J. S.**; MARINHO, LEANDRO B.; SOUZA, JOÃO WELLINGTON M.; ASSIS, E. A.; REBOUÇAS FILHO, P. P.. **Localization System for Autonomous Mobile Robots Using Machine Learning Methods and Omnidirectional Sonar.** IEEE Latin America Transactions, v. 16, p. 368-374, 2018.

Produções com co-autorias:

- MARINHO, LEANDRO B.; REBOUÇAS FILHO, P. P.; **ALMEIDA, J. S.**; SOUZA, JOÃO WELLINGTON M.; SOUZA JUNIOR, A. H.; ALBUQUERQUE, VICTOR HUGO C. . **A novel mobile robot localization approach based on classification with rejection option using computer vision.** COMPUTERS ELECTRICAL ENGINEERING, v. 68, p. 26-43, 2018.
- de Souza, João W. M.; Alves, Shara S. A.; REBOUCAS, E. de S.; **ALMEIDA, J. S.**; REBOUÇAS FILHO, P. P.. **A New Approach to Diagnose Parkinson's Disease Using a Structural Cooccurrence Matrix for a Similarity Analysis.** doi.org/10.1155/2018/7613282, v. 1, p. 1-8, 2018.
- Rebouças Filho, Pedro Pedrosa; LIMA MOREIRA, FRANCISCO DIEGO; MARINHO, LEANDRO BEZERRA; **ALMEIDA, JEFFERSON SILVA.** **Novel and powerful mosaic constructor for territorial analysis using mobile robots via Binary Robust Invariant Scalable Keypoints.** MEASUREMENT, v. 114, p. 448-470, 2018.

Capítulos de livros publicados:

- de Souza, João W. M.; **ALMEIDA, JEFFERSON S.**; Rebouças Filho, Pedro Pedrosa. **A New Approach for the Diagnosis of Parkinson s Disease Using a Similarity Feature Extractor.** Advances in Intelligent Systems and Computing. 736ed.: Springer International Publishing, 2018, v. 736, p. 21-31.
- da Silva, Suane Pires P.; MARINHO, LEANDRO B.; **ALMEIDA, JEFFERSON S.**; Rebouças Filho, Pedro Pedrosa. **A Novel Approach for Mobile Robot Localization in Topological Maps Using Classification with Reject Option from Structural Co-occurrence Matrix.** Lecture Notes in Computer Science. 1ed.: Springer International Publishing, 2017, v. 10424, p. 3-15.

Artigos publicados em congressos:

- da Silva, Suane Pires P.; NOBREGA, Raul Victor M.; MEDEIROS, A. G.; MARINHO, LEANDRO B.; **ALMEIDA, J. S.**. An Innovative Method for Location of Mobile Robot developed with Topological Maps and Classification with Reject Option using CNN in Omnidirectional Images. In: International Joint Conference on Neural Networks - IJCNN, 2018, Rio de Janeiro. An Innovative Method for Location of Mobile Robot developed with Topological Maps and Classification with Reject Option using CNN in Omnidirectional Images, 2018.
- SILVA, S. P. P.; MARINHO, LEANDRO B.; **ALMEIDA, J. S.**; REBOUÇAS FILHO, P. P.. A Novel Approach for Mobile Robot Localization in Topological Maps using Classification with Reject Option from Structural Co-occurrence Matrix. In: 17th International Conference on Computer Analysis of Images and Patterns - CAIP 2017, 2017, Ystad - Sweden. 17th International Conference on Computer Analysis of Images and Patterns - CAIP 2017, 2017.
- SILVA, S. P. P.; MARINHO, LEANDRO B.; **ALMEIDA, J. S.**; REBOUÇAS FILHO, P. P.. A Novel Approach for Robot Localization in Topological Maps using Classification with Reject Option from Structural Co-occurrence Matrix in Omnidirectional Images. In: 16th IEEE International Conference On Machine Learning And Applications - ICMLA 2017, 2017, Cancun - México. 16th IEEE International Conference On Machine Learning And Applications, 2017, 2017.
- SOUZA, JOÃO WELLINGTON M.; **ALMEIDA, JEFFERSON S.**; REBOUÇAS FILHO, P. P. . A new approach for the diagnosis of Parkinson's disease using a similarity feature extractor. In: ISDA 2017 - International Conference on Intelligent Systems Design and Applications, 2017, Delhi - Índia. Intelligent Systems Design and Applications, 2017.

1.6 Estrutura da Dissertação

Esta dissertação está organizada da seguinte forma. Uma revisão bibliográfica sobre técnicas de Pré-processamento de Sinal de Voz, Extração de Características em Sinal de Voz, e Rede Neural Artificial é apresentada no Capítulo 2. No Capítulo 3 por sua vez, é abordado a metodologia avaliada neste trabalho. No Capítulo 4 são mostrados os resultados obtidos e realizadas as suas discussões. No Capítulo 5, por fim, são apresentadas as conclusões e contribuições desta dissertação.

2 Fundamentação Teórica

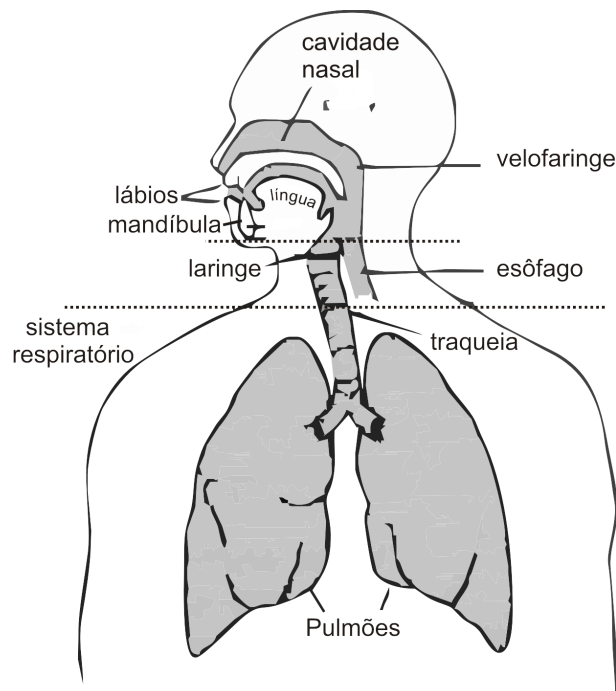
Neste Capítulo apresentamos uma breve revisão abordando a teoria sobre métodos e técnicas utilizados. Na seção 2.1 apresentamos os mecanismos de produção da voz e na seção 2.2 as características da fala. Na seção 2.3 comentamos sobre os métodos de extração de características e na seção 2.4 apresentamos brevemente a teoria dos métodos de classificação.

2.1 O mecanismo da fala

Os aspectos principais de produção da fala estão relacionados com seis órgãos, denominados subsistemas principais e ilustrados na Figura 4.

O sistema respiratório, formado pelos pulmões, via aérea, caixa torácica, diafragma e estruturas associadas, fornece o suprimento básico de ar para gerar sons.

Figura 4 – Órgãos de produção da fala.



Fonte: Adaptado de (BERNTHAL; BANKSON; FLIPSEN, 2009).

A laringe, composta de várias cartilagens e músculos, gera os sons da fala pela vibração das pregas vocais, e permite que o ar passe dos pulmões para o trato vocal (as cavidades oral e nasal) para sons sem voz.

O velofaringe e as estruturas associadas do orifício velofaríngeo - une ou separa as cavidades oral e nasal de modo que o ar passe pela cavidade oral, pela cavidade nasal ou por ambos.

A língua, principalmente um complexo de músculos, é o principal articulador da cavidade oral; é capaz de assumir uma variedade de formas e posições na pronúncia de vogal e consoante. Os lábios, junto com a mandíbula, são os mais visíveis dos articuladores; eles estão envolvidos na produção de vogais e consoantes.

A mandíbula, a estrutura óssea maciça e seus músculos associados, sustentam os tecidos moles da língua e do lábio inferior. Participa da produção da fala, auxiliando na movimentação da língua e lábios e fornecendo suporte esquelético para esses órgãos.

A função básica do sistema respiratório na fala é empurrar o ar para dentro da via aérea composta pela laringe e pelas cavidades oral e nasal. A função básica da laringe é regular o fluxo de ar dos pulmões para criar tanto segmentos sonoros quanto surdos. A via aérea superior, muitas vezes chamada de trato vocal, vai da laringe à boca ou ao nariz e é o local do que é comumente chamado de articulação da fala. Na maior parte, esse processo é realizado por movimentos dos articuladores: língua, lábios, mandíbula e velofaringe.

A articulação da fala é, portanto, uma questão de alongamento, encurtamento e constrição² do tubo conhecido como trato vocal.

Todo este processo é controlado pelo sistema nervoso, que deve traduzir a mensagem a ser comunicada em um padrão de sinais que correm para os vários músculos do mecanismo da fala. Quando esses músculos se contraem, uma variedade de coisas pode acontecer: o ar pode ser expelido dos pulmões, as pregas vocais podem começar a vibrar, a velofaringe pode se fechar, a mandíbula pode baixar ou os lábios podem se projetar (BERNTHAL; BANKSON; FLIPSEN, 2009).

2.2 Características relevantes da fala

No livro intitulado *Motor Speech Disorders*, Duffy (2013) explica que existem características salientes que permitem examinar desordens relacionadas a voz. Características salientes são aquelas que contribuem de forma mais direta e influente para o diagnóstico. Entre estas características, podemos destacar:

- Força;
- Velocidade de movimento;
- Amplitude de movimento;
- Estabilidade;
- Tom;

² Constrição: diminuição do diâmetro ou o volume de (músculo, órgão, etc).

- Precisão.

A seguir são detalhados a forma com que cada um afeta os mecanismos relacionados a voz.

2.2.1 Força

Os músculos tem força suficiente para cumprir sua função, mais uma reserva de força. Esta reserva permite que o músculo execute forças de contração por repetidas vezes sem causar fadiga. Porém, quando o músculo está fraco não consegue contrair ao nível desejado.

A fraqueza muscular pode alterar três das principais válvulas relacionados aos mecanismos da fala: laringe, velofaríngeo e articulatório. Pode interferir em todos os componentes da produção da fala (respiração, fonação, ressonância, articulação e prosódia) (DUFFY, 2013).

2.2.2 Velocidade

Os movimentos realizados durante a fala são relativamente rápidos, especialmente os movimentos laríngeos, velofaríngeos e articulatórios em que a válvula expira o ar para produzir aproximadamente 14 fonemas por segundo. Por exemplo, esta velocidade de pronúncia dos fonemas é típica de uma fala conversacional. Estes movimentos podem ocorrer como contrações únicas ou repetitivas. Esta dinâmica da fala é coordenada pela ativação dos neurônios motores alfa.

A velocidade excessiva é incomum em pessoas que possuem lesões por esforço repetitivos, DORTs (Distúrbios Osteomusculares Relacionados ao Trabalho), porém pode ocorrer na disartria hipocinética. A disartria pode ser tratada como um problema de linguagem, em que o indivíduo tem dificuldade de coordenar os movimentos dos músculos que estão envolvidos no processo da fala (DIAS; LIMONGI, 2003).

Dias e Limongi (2003) explica que durante as fases iniciais da doença de Parkinson (DP) o indivíduo pode apresentar alterações discretas na qualidade da voz e da articulação. As alterações da voz e da fala nos portadores de DP constituem, em conjunto, o que se denomina de disartria hipocinética ou disartrofonía. A disartrofonía pode causar monotonia e redução da intensidade da voz, articulação imprecisa e distúrbios do ritmo. Existem tratamentos que favorecem a recuperação da intensidade vocal em pacientes com DP, como por exemplo, o método de Lee Silverman (DIAS; LIMONGI, 2003).

Os efeitos observados na alteração de velocidade da fala são percebidos após análise de exames acústicos e fisiológicos.

2.2.3 Amplitude

Os mecanismos da fala possuem estruturas que realizam movimentos únicos e repetitivos. Estas estruturas percorrem distâncias de forma bastante precisa, porém existem variações consideradas pequenas.

Em patologias relacionadas a voz, por exemplo, a disartria hipocinética é observado a diminuição da amplitude dos movimentos realizados pelos mecanismos da voz e o aumento da velocidade.

As anomalias na amplitude de movimento geralmente têm uma grande influência nas características prosódicas³ da fala, resultando, algumas vezes, em variações prosódicas restritas ou excessivas (DUFFY, 2013).

2.2.4 Estabilidade

O corpo humano em repouso produz naturalmente na musculatura uma oscilação mensurável na faixa de frequência de 8 a 12 Hz. Durante o movimento normal é pouco comum a ocorrência de oscilações em partes do corpo nas pessoas saudáveis. Porém em situações em que o indivíduo está submetido a fadiga extrema, sob estresse emocional este tremor fisiológico visível pode ser percebido.

Em pessoas com DP, a estabilidade motora é comprometida e surgem movimentos involutários ou hipercinesias. O tremor é o movimento involutário mais comum que se manifesta na doença neurológica. Consiste em oscilações repetitivas e rítmicas de uma parte do corpo, geralmente com frequência de 3 a 12 Hz. Os tremores podem ocorrer em quatro situações: (1) em repouso, (tremor de repouso); (2) quando uma parte do corpo é mantida contra a gravidade (tremor postural); (3) durante o movimento (tremor de ação); ou (4) próximo ao final de um movimento (tremor terminal) (DUFFY, 2013).

Os tremores leves podem não afetar significativamente a fala, dependendo da respiração, ressonância ou articulação do indivíduo. Porém comumente afeta a fonação, e quando grave, pode afetar a prosódia; durante a produção de vogais sustentadas os efeitos causados pelo tremor são mais perceptíveis. Desta forma, os efeitos do tremor na fala podem ser ouvidos durante a fala, registrados em exames do mecanismo oral e mensurados fisiologicamente ou acusticamente (DUFFY, 2013).

A doença pode manifestar movimentos de tremor aleatórios, imprevisíveis, que podem variar em velocidade, duração e amplitude. Podem ser graves ao ponto de interromper ou alterar a direção do movimento pretendido. Eles são a principal fonte de fala anormal nas disartrias⁴ hipercinéticas. Além do mais, pode afetar todos os componentes

³ Prosódia: parte da gramática tradicional que se dedica às características da emissão dos sons da fala.

⁴ Disartria: distúrbio da articulação da fala (dificuldade na produção de fonemas).

de produção da fala, comprometendo a precisão e frequentemente a prosódia (emissão de sons da fala).

2.2.5 Tom

Em pacientes com DP, o tônus muscular pode ser reduzido ou excessivo. Tônus muscular é o estado involuntário de contração natural dos músculos corporais. Este estado permite o músculo entrar em ação rapidamente sempre que necessário. Sem o tônus muscular os músculos estariam em estado de total repouso, e isso dificultaria a ação imediata. Alterações no tônus podem afetar diretamente as válvulas e os mecanismos de produção da fala.

Quando o tônus (tom) é reduzido constantemente, associa-se esta anormalidade a disartrias flácidas. O tom variável associa-se com disartria espástica ou hipocinética. Os efeitos de tom anormal podem ser percebidos durante a fala (DUFFY, 2013).

2.2.6 Precisão

Em um indivíduo sadio a fala é suficientemente executada com precisão, de forma a assegurar o significado linguístico e emocional.

Movimentos imprecisos dos mecanismos podem resultar em vários erros na fala. A causa pode estar no sistema motor e na execução dos movimentos ou no neurotransmissor, que programa os movimentos.

Imprecisões na fala podem surgir por alteração no tônus, força, estabilidade e velocidade. Por exemplo, se a força e amplitude de movimento forem excessivas ou diminuirão.

É pouco comum que apenas uma única característica neuromuscular se manifeste em pessoas com disartria (DUFFY, 2013).

Levando em consideração as características da fala mencionadas anteriormente, Duffy (2013) afirma que estas observações podem ser úteis para estabelecer um diagnóstico de patologias relacionadas a fala.

Uma maneira comumente utilizada para obter informações a partir de sinal de voz é utilizando os extratores de características, que são explicados na seção a seguir.

2.3 Extração de Características

Nesta pesquisa avaliamos diversas bibliotecas de extração de características específicas para sinal de áudio, fala e música. Nas subseções um breve comentário foi realizado, a fim de mostrar o que possui os extratores com base na explicação da documentação.

2.3.1 OpenSMILE

OpenSMILE (*Speech and Multimedia Interpretation by Large-space Extraction*) é uma biblioteca de código aberto para extração de características de sinais de áudio, fala e música (EYBEN et al., 2013). É aplicada principalmente na área de reconhecimento de emoções. O projeto OpenSMILE existe desde 2008 e é mantido pela empresa alemã audEERING GmbH desde 2013. Exemplos de tais características codificadas na fala humana são a emoção de um falante (SCHULLER et al., 2010), idade, gênero e personalidade, bem como estados de fala como depressão, intoxicação ou distúrbios patológicos vocais. A biblioteca apresenta os seguintes conjuntos de características:

- MFCC (*Mel-frequency Cepstral Coefficients*) para reconhecimento de fala;
- PLP (*Cepstral Coefficients*) para reconhecimento de fala;
- Prosódia (altura e intensidade);
- *INTERSPEECH 2009 Emotion Challenge*;
- *INTERSPEECH 2010 Paralinguistic*;
- *INTERSPEECH 2011 Speaker State*;
- *INTERSPEECH 2012 Speaker Trait*;
- *INTERSPEECH 2013 ComParE*.

2.3.2 Essentia

Descritor Essentia é uma biblioteca de ferramentas com código aberto para análise, descrição e síntese de áudio e música (BOGDANOV et al., 2013). A biblioteca apresenta os seguintes conjuntos de características:

- Blocos padrão de processamento de sinal : FFT, DCT, cortador de estrutura, janelamento, envelope, alisamento;
- Filtros (FIR IIR) : baixo / alto / passagem de banda, rejeição de banda, remoção de DC, volume igual;
- Descritores estatísticos : mediana, média, variância, potência, momentos crus e centrais, disseminação, curtose, assimetria, achatamento;
- Descritores no domínio do tempo: duração, loudness, LARM, Leq, loudness de Vickers, taxa de cruzamento zero, tempo de ataque de log e outros descritores de envelope de sinal;
- Descritores espectrais : Bark / Mel / ERB bandas, MFCC, GFCC, LPC, picos espectrais, complexidade, rolloff, contraste, HFC, inharmonicity e dissonância;
- Descritores tonais : função de saliência de pitch, melodia e pitch predominantes, características relacionadas a HPCP (chroma), acordes, tecla e escala, frequência de sintonização;
- Descritores de ritmo : detecção de batidas, BPM, detecção de início, transforma-

ção de ritmo, batimento de volume.

2.3.3 jAudio

Os recursos do jAudio foram introduzidos em 2005. Esta biblioteca foi projetada para extrair recursos de sinais unidimensionais ou multidimensionais para classificação automática de música. (MCENNIS; MCKAY; FUJINAGA, 2006). Internamente, esta biblioteca possui um compilado de vários extratores e atributos, no qual podemos mencionar alguns deles:

- Magnitude Spectrum, Power Spectrum, Spectral Centroid, Spectral Roll off Point;
- FFT Bin Frequencies, Zero Crossings;
- Zernike Moments, Zernike Moments Beat Histogram;
- Spectral Flux, Compactness, Spectral Variability;
- RMS, Mean, standard Deviation;
- Fraction of Low Energy Windows;
- Beat Histogram, Beat Histogram Labels;
- MFCC, LPC;
- Harmonic Spectral Centroid;
- Area Moments, Area Moments MFCC, Area Moments Constant QMFCC, Area Polynomial Approximation Constant QMFCC.

2.3.4 KTU

Os recursos do KTU foram introduzidos para o exame de patologia da voz por Gelzinis, Verikas e Baciauskiene (2008) na Universidade de Tecnologia de Kaunas e, posteriormente, expandidos para incluir recursos adicionais. A variante duradoura desse conjunto de recursos foi criada combinando subconjuntos de recursos, que contemplam:

- MFCC (Mel-frequency Cepstral Coefficients);
- PLPCC (Perceptual Linear Predictive Cepstral Coefficients);
- além de outros atributos.

2.3.5 MPEG7

O descritor MPEG-7 contém ferramentas de descrição para áudio. A extração de descritores de baixo nível é normativa e baseia-se no próprio sinal de áudio. Com a ajuda de descritores de baixo nível, é possível pesquisar e filtrar conteúdo de áudio em relação a, por exemplo, espectro, harmonia, timbre e melodia (CRYSANDT; TUMMARELLO; PIAZZA, 2004).

Nos descritores de áudio de baixo nível (LLD - *Low Level Descriptors*) existem dezessete parâmetros temporais e espectrais que podem ser divididos nos seguintes grupos:

- Básico: forma de onda instantânea e valores de potência;
- Espectro básico: espectro de energia *log-frequency* e características espectrais (por exemplo, centróide espectral, espalhamento espectral);
- Parâmetros do sinal: frequência fundamental e harmonicidade dos sinal
- Timbral temporal: Tempo de ataque do registro e centroide temporal
- Timbral espectral: características espectrais especializadas em um espaço de frequência linear
- Representações de base espectral: um número de características usadas em conjunto para o reconhecimento de som para projeções em um espaço de baixa dimensão.

2.3.6 Tsanas

Esta *toolbox* apresenta um número de algoritmos de processamento de sinal de voz, visando a caracterização objetiva da voz e, em particular, a avaliação de distúrbios da voz (TSANAS et al., 2010). Esses algoritmos são direcionados principalmente para:

- quantificar a amplitude (variantes shimmer);
- a frequência (variantes de jitter);
- aumento do ruído (medidas sinal-ruído).

A caixa de ferramentas foi desenvolvida e só foi validada em configurações com a vogal /a/ sustentada. Está disponível na versão Matlab. Os algoritmos podem ser generalizáveis para outras vogais sustentadas, mas definitivamente não são apropriadas para fala conversacional.

2.3.7 YAAFE

YAAFE (*Yet Another Audio Features*) é uma caixa de ferramentas de extração de recursos de áudio (MATHIEU et al., 2010). Está disponível em Python e Matlab. Entre os atributos disponível, podemos mencionar:

- Propagação Perceptual
- Nitidez Perceptual
- Variação espectral
- Diminuição Espectral
- Estatísticas de forma espectral
- MFCC, LPC, FFT
- OBSI, OBSIR

- Modulação de amplitude

Na Tabela 1 apresentamos o extrator, o tamanho expresso de acordo com o número de atributos a qual dispõe, e a sigla utilizada.

Tabela 1 – Extratores de características utilizados, em que o número de atributos é especificado, além da referência na literatura e sigla adotada neste trabalho. Itens: (1-12) biblioteca OpenSmile; (13) Descritor Essentia; (14) Biblioteca JAudio; (15) Biblioteca KTU; (16) Descritor MPEG7; (17) Biblioteca Tsanas; e (18) Biblioteca YAAFE.

Item	Extrator de características	N. de atributos	Referência	Sigla
1	avec2011	1941	(EYBEN et al., 2013)	AV1
2	avec2013	2268	(EYBEN et al., 2013)	AV2
3	emo_large	6552	(EYBEN et al., 2013)	EL
4	emobase	988	(EYBEN et al., 2013)	EM1
5	emobase2010	1582	(EYBEN et al., 2013)	EM2
6	IS13_ComParE	6373	(EYBEN et al., 2013)	ED
7	IS09_emotion	384	(EYBEN et al., 2013)	IS1
8	IS10_paraling	1582	(EYBEN et al., 2013)	IS2
9	IS10_paraling_compat	1582	(EYBEN et al., 2013)	IS3
10	IS11_speaker_state	4368	(EYBEN et al., 2013)	IS4
11	IS12_speaker_trait	5757	(EYBEN et al., 2013)	IS5
12	IS12_speaker_trait_compat	6125	(EYBEN et al., 2013)	IS6
13	Essentia_descriptors	1915	(BOGDANOV et al., 2013)	IS7
14	jAudio_features	1794	(MCENNIS; MCKAY; FUJINAGA, 2006)	JA
15	KTU_features	1267	(GELZINIS; VERIKAS; BACAUSKIENE, 2008)	KT
16	MPEG7_descriptors	527	(CRYSANDT; TUMMARELLO; PIAZZA, 2004)	MP
17	Tsanas	339	(TSANAS et al., 2010)	TS
18	YAAFE_features	1885	(MATHIEU et al., 2010)	YA

2.4 Classificadores

A inspiração para o funcionamento das Redes Neurais Artificiais (RNAs) é baseada no funcionamento do cérebro humano, área aprofundada pelo estudo pioneiro de Haykin (2002), que introduziu a idéia dos neurônios tal como são conhecidos hoje.

McCulloch e Pitts (1988) foram os primeiros a formular um modelo matemático simples para um neurônio. McCulloch um psiquiatra, Pitts, matemático, se associaram um ano antes da publicação do trabalho conjunto. O trabalho é a unificação de estudos da neurofisiologia com a lógica matemática, que propõe conexões entre neurônios artificiais ajustadas corretamente e operando de forma síncrona para a computação de funções.

RNAs são sistemas paralelos distribuídos, compostos por unidade de processamento, denominadas neurônios, dispostas em camadas e cujo funcionamento baseia-se no cálculo de funções matemáticas. As unidades, entre as camadas, são interconectadas e, na maioria dos modelos, estão associadas a pesos, os quais armazenam o conhecimento adquirido na forma de valores numéricos (LUDERMIR, 2007).

Nas subseções a seguir explicaremos em uma breve revisão os tipos de classificador utilizados.

2.4.1 *k*Nearest Neighbors

O *k*Vizinhos mais próximo (*k*Nearest Neighbors, *k*NN) é um algoritmo de aprendizado supervisionado cuja ideia básica consiste em encontrar as *k* amostras mais próximas das amostras não classificadas e usar esses rótulos para classificar uma amostra desconhecida.

Formalmente, dado um conjunto de *N* pontos em um espaço de característica *n*-dimensional

$$\{(x_1, x_2, \dots, x_n) | x_i \text{ é real para todo } 1 \leq i \leq n\}, \quad (2.1)$$

considerando os subespaços *m*-dimensionais

$$\{(x_1, x_2, \dots, x_n) | x_i = 1 \text{ para } i \in I, x_i = 0 \text{ para } i \notin I\} \quad (2.2)$$

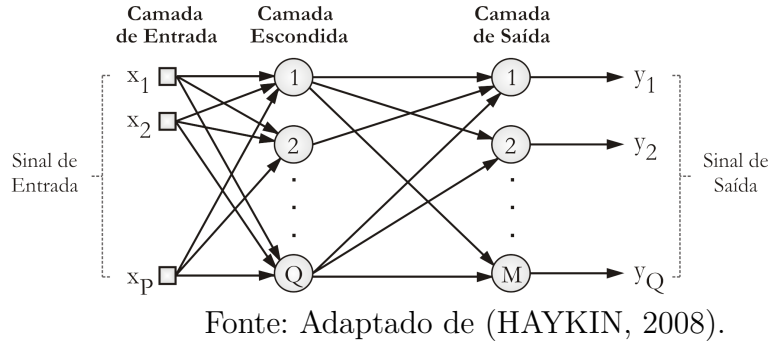
onde *I* é um subconjunto de *m*-elementos de $\{1, 2, \dots, n\}$ e $m < n$.

O *k*NN requer pouco esforço na fase de treinamento. Em contrapartida, o custo de cálculo para a classificação de uma amostra desconhecida é relativamente elevado. O número de vizinhos é o parâmetro no *k*NN (HO, 1998).

2.4.2 Perceptron de Múltiplas Camadas

O Perceptron de Múltiplas Camadas (*Multilayer Perceptron*, MLP) é um tipo de rede neural artificial comumente usada na solução de problemas não linearmente separáveis (HAYKIN, 2008). Sua arquitetura é composta de camadas de neurônios, de tal modo que cada camada está totalmente ligada à seguinte, Figura 5. O MLP tem uma ou mais camadas escondidas, que são formadas por neurônios ocultos. De acordo com Haykin (2008), os neurônios ocultos interferem entre as camadas de entrada e saída, permitindo que a rede extraia estatísticas de alta ordem. Os pesos de conexão na rede medem o grau de correlação entre níveis de atividade dos neurônios que eles conectam. Além disso, um algoritmo de treinamento deve ser aplicado para ajustar os pesos. O algoritmo de treinamento mais popular é o *Backpropagation*.

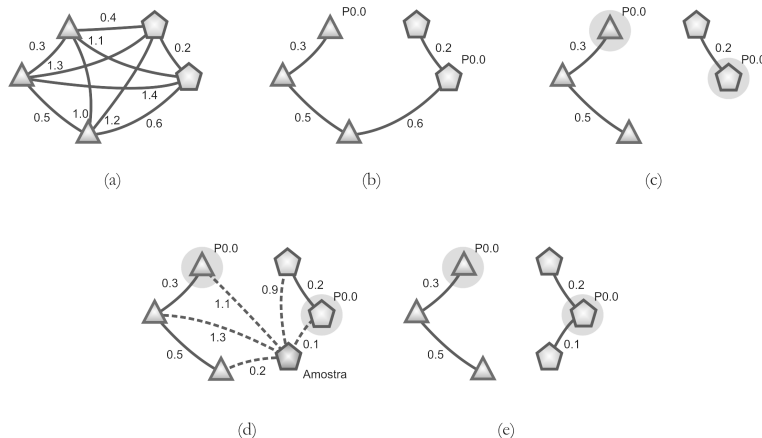
Figura 5 – Arquitetura gráfica de um perceptron multicamadas com uma camada escondida.



2.4.3 Optimum Path Forest

O OPF (*Optimum Path Forest*, conhecido por Floresta de Caminhos Ótimos, é baseado na teoria de grafos. Basicamente, o OPF estabelece que durante a etapa de treinamento, acontece um processo de competição entre algumas amostras de referência, chamadas de protótipos, na busca de particionar o grafo em uma floresta de caminhos ótimos, e posteriormente, realizar a classificação (PAPA; aO; SUZUKI, 2009). A Figura 6 ilustra de forma interativa este processo, onde temos dois grupos de objetos que deverão ser classificados quanto a sua geometria. As linhas representam a distância entre os objetos próximos. Ao inserir uma amostra desconhecida, calcula-se a distância entre todos os objetos do grupo, de forma a agregar a amostra ao objeto mais próximo, encontrando assim a qual grupo pertence.

Figura 6 – Funcionamento do classificador OPF: (a) Conjunto de treinamento modelado com grafo completo (b) Computação de uma árvore de expansão mínima sobre o conjunto de treinamento (c) Floresta de caminho ótimo encontrada sobre o conjunto de treinamento, os protótipos são destacados (d) Processo de classificação de uma amostra (e) A amostra de teste é finalmente classificada.



Fonte: Próprio autor.

2.4.4 Support Vector Machines

As SVMs (*Support Vector Machines*) são embasadas pela teoria de aprendizado estatístico, desenvolvida por Vapnik (1995) a partir de estudos iniciais em (VAPNIK; CHERVONENKIS, 2015). As SVMs lineares são eficazes na classificação de conjuntos de dados linearmente separáveis, porém, em situações de não linearidade, é comum a utilização de outros *kernels*, como por exemplo: polinomial, RBF, gaussiano ou sigmoidal.

Independentemente da forma como uma máquina de vetor de suporte é implementada, ela difere da abordagem convencional para o desenho de um perceptron multicamada de uma maneira fundamental. Na abordagem convencional, a complexidade do modelo é controlada mantendo o número de características, isto é, neurônios ocultos. Por outro lado, a máquina de vetor de suporte oferece uma solução para o projeto de uma máquina de aprendizado controlando a complexidade do modelo independentemente da dimensionalidade, como mostrado na Figura 7.

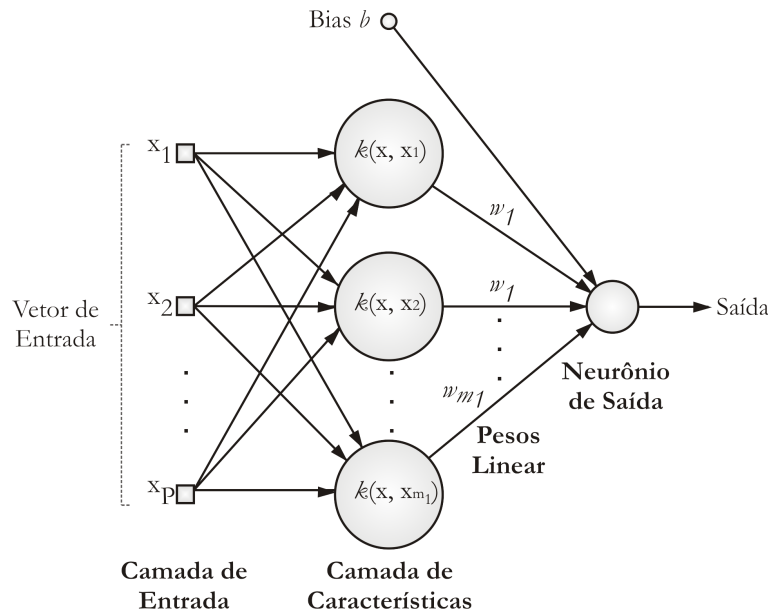


Figura 7 – Arquitetura de uma Support Vector Machine usando kernel Radial Basis Function, RBF. Adaptado de (VAPNIK; CHERVONENKIS, 2015).

A Tabela 2 apresenta os classificadores avaliados nesta pesquisa. Foram especificados as configurações, referência e sigla adotadas.

2.5 Conclusão do Capítulo

A fundamentação teórica foi importante para compreensão dos aspectos relacionados a fala. Ficou claro que a doença altera o funcionamento normal dos mecanismos de produção da fala. Em virtude disto, escolhemos avaliar os extratores classificadores específicos para

Tabela 2 – Classificadores utilizados, em que a configuração é especificada, além da referência na literatura e sigla adotada neste trabalho. Itens: (1-3) k Nearest Neighbors; (4) Multilayer Perceptron; (5-14) Optimum Path Forest; e (15-17) Support Vector Machine.

Item	Classificador	Configuração	Referência	Sigla
1	KNN	$k = 1$	(FUKUNAGA; NARENDRA, 1975)	K1
2	KNN	$k = 3$	(FUKUNAGA; NARENDRA, 1975)	K3
3	KNN	$k = 5$	(FUKUNAGA; NARENDRA, 1975)	K5
4	MLP	IL/HL/OL	(HAYKIN, 2009)	M
5	OPF	Bray-Curtis	(PAPA et al., 2007)	OB
6	OPF	Canberra	(PAPA et al., 2007)	OC
7	OPF	Bray-Curtis	(PAPA et al., 2007)	OB
8	OPF	Canberra	(PAPA et al., 2007)	OC
9	OPF	Chi-Square	(PAPA et al., 2007)	OCS
10	OPF	Euclidean	(PAPA et al., 2007)	OE
11	OPF	Mahalanobis	(PAPA et al., 2007)	OMS
12	OPF	Manhattan	(PAPA et al., 2007)	OMN
13	OPF	Squared Chi-Squared	(PAPA et al., 2007)	OSCS
14	OPF	Squared Chord	(PAPA et al., 2007)	OSC
15	SVM	Linear	(VAPNIK, 1998)	SL
16	SVM	Poly	(VAPNIK, 1998)	SP
17	SVM	RBF	(VAPNIK, 1998)	SR
Configuração: Parâmetro; Topologia: IL-entrada, HL-camada escondida, OL-saída; Distância ou Kernel.				

tarefa vocal. Neste sentido, avaliaremos as tarefas de fonação sustentada e de fala. O desempenho da classificação irá depender do extrator e da relevância dos atributos na detecção da doença.

Para que tenha-se uma ideia da relevância desta pesquisa, torna-se interessante explicar matematicamente a combinação das tarefas vocais, tipos de microfone e dos métodos e técnicas utilizados.

Ao todo, considerando as variáveis envolvidas, temos:

- 2 tipos de microfones
- 4 tipos de áudio
- 18 extratores
- 4 classificadores

Com isso obteve-se 576 combinações diferentes de mecanismos de classificação avaliadas neste trabalho.

$$2 \times 4 \times 18 \times 4 = 576$$

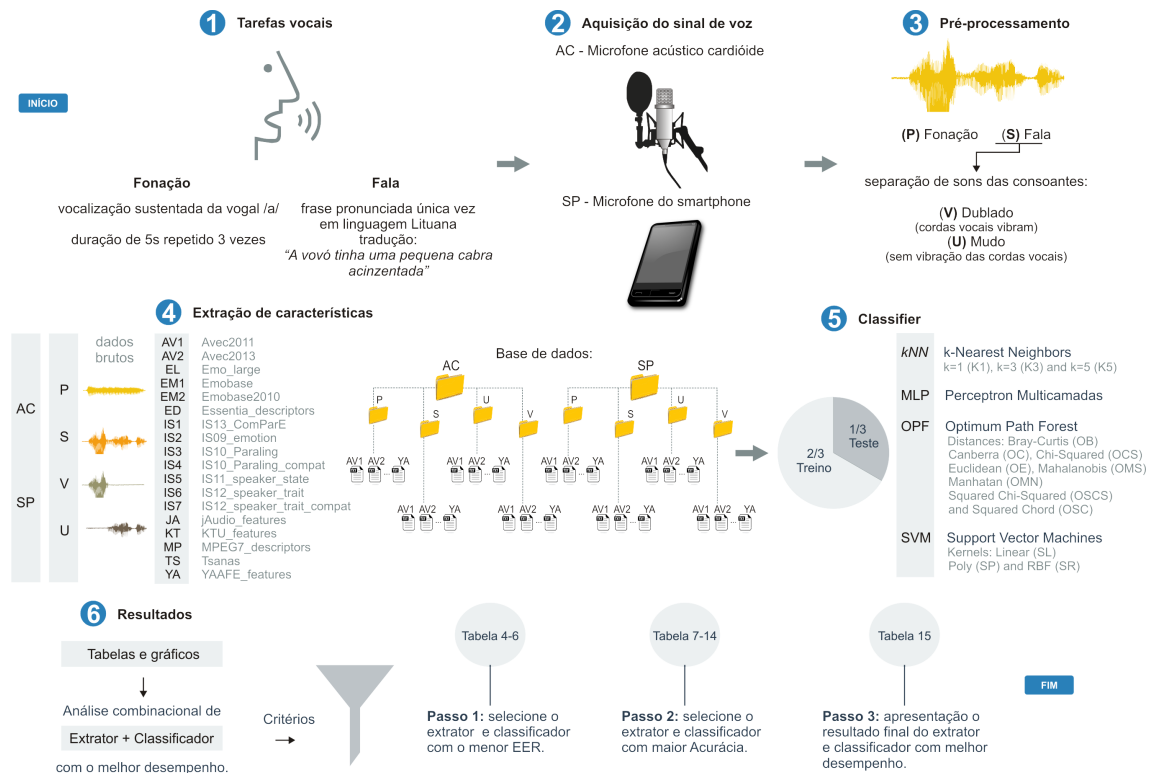
3 Metodologia

Neste capítulo explicamos a metodologia da abordagem proposta e descrevemos a ideia geral do trabalho. Na seqüência, apresentamos um resumo dos extratores de características e técnicas de aprendizado de máquina, que contempla as métricas utilizadas para avaliação.

3.1 Ideia geral

Para detectar a DP, a abordagem proposta baseia-se em sintomas motores explicitamente relacionados à voz. O primeiro passo para implementar esta metodologia foi adquirir a base de dados disponibilizada por (VAICIUKYNAS et al., 2017). Esta base de dados contém áudios de voz obtidos a partir de exames de pacientes saudáveis e pacientes com DP. A Figura 8 mostra o diagrama descrevendo o passo seguido nesta abordagem.

Figura 8 – Descrição das etapas seguidas na abordagem proposta nesta dissertação.



Fonte: Próprio autor.

O exame de voz começa no Passo 1, com duas tarefas vocais realizadas por cada paciente com orientação e acompanhamento médico. As tarefas vocais utilizadas incluem fonação e fala; respectivamente pela ação de pronunciar uma vogal e pronúncia de sentença.

Os equipamentos utilizados na aquisição de sinal de voz são um microfone profissional (AC) e o microfone do aparelho *smartphone* (SP). Assim, podemos avaliar o desempenho dos classificadores com tipos diferentes de microfones, AC e SP respectivamente.

O passo 3 compreende o pré-processamento dos sinais de áudio adquiridos no primeiro passo. A fala do paciente gravada em áudio é particionada em outros dois arquivos de áudio, chamados Dublado e Mudo. Estes dois tipos de áudio referem-se aos sons produzidos durante a fala, relacionados aos sons característicos de vogais ou consoantes.

No Passo 4, a extração de características contendo todos os extratores é aplicada, o que gera outros arquivos para cada conjunto de características e modalidades. Ao todo, 144 arquivos de extração foram gerados.

Essa abordagem usa a técnica de *hold-out* para particionar a base de dados e quatro classificadores, como mostrado na Etapa 5. Foram usadas dez iterações. As configurações para o k-Nearest Neighbors (kNN), Perceptron Multi-Layer (MLP), Optimum-Path Forest (OPF) e Support Vector Machine (SVM) foram escolhidas e treinados usando ferramentas da biblioteca OpenCV versão 2.49 com configurações padrão para os hiper-parâmetros. Três kernels de função diferentes foram usados para o classificador SVM.

No MLP o número de neurônios na camada oculta variou de 1 a 50. Por sua vez, o OPF foi treinado com várias métricas de distância. Na etapa 6 foi feita uma seleção pelas métricas EER (Equal Error Rate) e Acurácia, ambas utilizadas para avaliar os resultados e escolher a melhor combinação de extrator e classificador.

As curvas ROC (Receiver Operating Characteristic) e DET (Detection Error Tradeoff) foram aplicadas para obter AUC (Area Under Curve) e EER (Equal Error Rate), respectivamente. O EER foi utilizado por dois motivos: permite identificar a combinação de extrator e classificador que melhor apresenta um balanço entre a taxa de falso negativo (FNR) e taxa de falso positivo (FPR). Portanto, a métrica EER será a métrica padrão para comparar os resultados do presente trabalho com os resultados de (VAICIUKYNAS et al., 2017).

3.2 Base de dados *crossourced*

O conjunto de dados usado neste trabalho foi gerado por (VAICIUKYNAS et al., 2017) a partir de duas tarefas vocais e organizado em duas modalidades diferentes: fonação e fala. As tarefas de áudio foram gravadas usando áudio stereo com um microfone profissional do tipo acústico cardíóide (modelo AKG Perception 220, com faixa de frequência de 20-20000Hz) e um smartphone (microfone interno do Samsung Galaxy Note 3). O microfone foi posicionado a cerca de 10cm de distância da fonte de áudio do participante. O formato

de áudio usado foi o mono PCM wav (16 bits a 44.1 kHz taxa de amostragem). A modalidade de fonação contém uma voz sustentada de vogal / a / vocalizada em um nível de intensidade confortável. A duração da tarefa foi de cerca de 5s e foi repetida 3 vezes. Modalidade de fala contém a pronúncia de uma frase na língua lituana por um falante nativo, "turėjo senelė žilą oželį". O software Praat foi usado para dividir a fala de áudio, gerando partes sonoras e surdas. A base de dados contém dados gravados de 99 sujeitos, tanto masculinos como femininos, para gravações acústicas de cardióide (CA) e de smartphone (SP). Mais informações são mostradas na Tabela 3.

Tabela 3 – Número de sujeitos no banco de dados utilizado neste trabalho.

Exames	Fonação		Fala	
	AC	SP	AC	SP
HC masculino	11 (33)	11 (33)	11	11
HC feminino	24 (72)	24 (72)	24	24
HC total	35 (105)	35 (105)	35	35
PD masculino	30 (89)	30 (90)	29	30
PD feminino	34 (101)	34 (102)	34	34
PD total	64 (190)	64 (192)	63	64
Total	99 (295)	99 (297)	98	99

Nota: PD - Paciente com doença de Parkinson, HC - Indivíduos saudáveis, Microfones: AC - acústico cardióide, SP - smartphone.

3.3 Métricas de Avaliação

Para analisar o desempenho do classificador, cinco medidas bem conhecidas foram empregadas: Taxa do Erro (EER), Área Sobre a Curva (AUC), Precisão, Especificidade e Sensibilidade, cujas definições são brevemente recapituladas abaixo.

Detection Error Trade-off (DET) é uma curva que apresenta pontos de operação relacionados à razão identificada entre FNR (False Negative Rate) e FPR (False Positive Rate). FNR representa as decisões incorretas para aceitar como PD um sujeito que de fato não tem PD, e FPR representa uma rejeição errada do grupo de controle representado por indivíduos sadios (HC). O classificador com melhor desempenho apresenta uma curva DET mais acentuada próxima a origem no gráfico (GARCIA-PERERA et al., 2012). O EER é obtido quando ($FNR = FPR$) no ponto de interseção entre a linha diagonal em torno da curva DET produzida pelo classificador. Quanto mais próximo de zero for o EER, melhor será o desempenho do classificador.

A curva ROC (Receiver Operating Characteristic) é uma ferramenta fundamental para avaliação de testes diagnósticos. Em uma curva ROC, a taxa positiva real (TPR) é representada graficamente como uma função do FPR para diferentes valores de corte de um parâmetro. Cada ponto na curva ROC representa um par sensibilidade / especificidade correspondente a um limiar de decisão específico.

Área sob a curva (AUC) é uma medida de quão bem um classificador pode distinguir entre dois grupos de diagnóstico (PD/HC). Portanto, quanto mais próxima a curva ROC estiver do canto superior esquerdo da plotagem, maior será a precisão geral do teste (ZWEIG; CAMPBELL, 1993).

Acurácia (Acc) é definida como a razão entre o número de padrões PD e HC corretamente classificados e o número total de padrões PD e HC:

$$Acurácia = \frac{(\text{PD} + \text{HC}) \text{ amostras classificadas corretamente}}{(\text{PD} + \text{HC}) \text{ amostras}} \quad (3.1)$$

Especificidade (Spe) é a razão entre o número de padrões de HC corretamente classificados de uma determinada classe e o número total de padrões de HC disponíveis para aquela classe, incluindo aqueles que foram classificados incorretamente:

$$Especificidade = \frac{\text{verdadeiro negativos}}{\text{falso positivos} + \text{verdadeiros negativos}} \quad (3.2)$$

Sensibilidade (Sen) é a razão entre o número de padrões de PD corretamente classificados de uma determinada classe e o número total de padrões de PD disponíveis para essa classe, incluindo aqueles que foram classificados incorretamente:

$$Sensibilidade = \frac{\text{verdadeiro positivos}}{\text{verdadeiro positivos} + \text{falso negativos}} \quad (3.3)$$

4 Resultados e Discussão

Nesta seção apresentamos os resultados alcançados pelos métodos de aprendizagem de máquina em conjunto com os dados obtidos por técnicas de extração de características. Com base nos dados apresentados nas Tabelas 4 à 6, selecionamos a melhor configuração de cada classificador. Os números correspondentes aos melhores resultados de cada coluna da tabela são indicados em negrito e são mostrados nas Tabelas 8 à 14. A Tabela 15 mostra os resultados finais.

Seguindo-se a metodologia aplicada para o reconhecimento de PD, no início do Passo 6, mostrado no Capítulo 3, selecionamos os melhores resultados de cada extrator de característica e classificador e mostramos nas Tabelas 4 até 6. Assim, a fim de obter um equilíbrio entre os casos falso positivos e os casos falso-negativos, consideramos a escolha mais adequada. No segundo estágio, os resultados de cada extrator de características e classificador são selecionados por acurácia e são mostrados em Tabelas 8 à 14. A intenção é obter um classificador que apresente a melhor acurácia (Acc). Observamos que, na maioria dos casos, a seleção por acurácia é caracterizada por um maior equilíbrio em especificidade (Spe) e sensibilidade (Sen). Geralmente, selecionando desta maneira, podemos obter o melhor desempenho de classificação em termos de AUC e EER. Os melhores resultados são destacados em estilo de fonte em negrito. Destes, escolhemos o que obteve o melhor desempenho, como mostra a Tabela 15.

Obtive-se resultados satisfatórios quando comparados com a abordagem de Vaiciukynas (VAICIUKYNAS et al., 2017). Os resultados comparados referem-se aos dados obtidos sem combinar diferentes tipos de áudio. Em nossa abordagem, os áudios foram classificados individualmente.

Ao analisar a Tabela 15, nossa abordagem selecionou dois classificadores com melhor desempenho. Na modalidade AC, o uso do tipo de áudio Fonação (P), do extrator YA e do classificador (K1) permitiu obter uma acurácia de 94,55% e EER de 19,01%. Vaiciukynas (VAICIUKYNAS et al., 2017) usou esses extratores com o classificador Random Forest e obteve um EER de 20,30% com (ED) extractor. Na modalidade (SP) destacamos o tipo de áudio (P) e extrator (KT) e o classificador (K1) com 92,94% de acurácia e EER 14,15%. Na modalidade (AC) e (SP), nossa abordagem obteve o melhor desempenho. Nossa abordagem é simplificada e eficiente porque não usamos fusão dos tipos de áudio. Apresentou vantagens porque obteve bom desempenho com classificador do tipo de áudio individual (P). Usando a fusão de todos os conjuntos de recursos, a abordagem de Vaiciukynas obteve EER de 19,27% para com microfone (AC) e 23,00% com o microfone (SP). Nossa abordagem obteve desempenho satisfatório.

Além do mais pode-se perceber que a tarefa de fonação mostrou-se a mais promissora para a detecção da doença (DP). Esta é uma característica interessante porque a tarefa de fonação sustentada é mais prática de ser realizada durante os exames de voz.

Os resultados obtidos reforçam a afirmação de Duffy (2013), que menciona sobre as características da fala e a sua utilidade em estabelecer um diagnóstico de patologias relacionadas a voz, neste caso a doença de Parkinson.

É importante considerar que existe a problemática do sistema de classificação abranger uma única língua. Para realizar o treinamento de algoritmos de aprendizado em uma linguagem específica, precisaríamos de novos bancos de dados de fala nessa linguagem. O desenvolvimento de um sistema multilíngue de classificação de doenças ainda é um grande desafio.

Os dados analisados possuem uma dimensão relativamente grande. Para redução dimensional, usamos *t-distributed Stochastic Neighbor Embedding* (t-SNE). É uma ferramenta para visualizar dados de alta dimensão. Neste caso, aplicamos o t-SNE para visualizar as amostras obtidas após extração com YA e KT marcados em negrito na Tabela 15. Pode-se ver na Figura 9 a visualização resultante do t-SNE, em que apresenta as amostras de PD e HC. Observe que usamos o t-SNE apenas para visualização.

Ter dados desbalanceados é muito comum em geral, mas neste caso tivemos mais amostras com a doença do que amostras de controle saudáveis. Como um conjunto de dados desbalanceado influenciará o modelo de predição para a classe mais comum, não vemos um problema significativo com o conjunto de dados sendo desequilibrado a esse respeito. Utilizamos 75% das amostras para treinamento e 25% para o teste usando a técnica de *hold-out*. Observe que não usamos validação cruzada por causa da grande dimensão de dados, o que levaria mais tempo para o treinamento.

A matriz de confusão dos resultados de classificação usando métodos de melhor desempenho é mostrada na Tabela 16. Observou-se que na modalidade AC, o número de falsos positivos e falsos negativos foi apresentado de forma equilibrada e isso poderia ser uma característica atrativa. Enquanto no modo SP, o número de falsos positivos foi relativamente maior. Isto significa que um maior número de indivíduos saudáveis foi confundido com indivíduos com DP. No entanto, o número de falsos negativos foi menor. A probabilidade de um indivíduo com DP ser diagnosticado como saudável é menor.

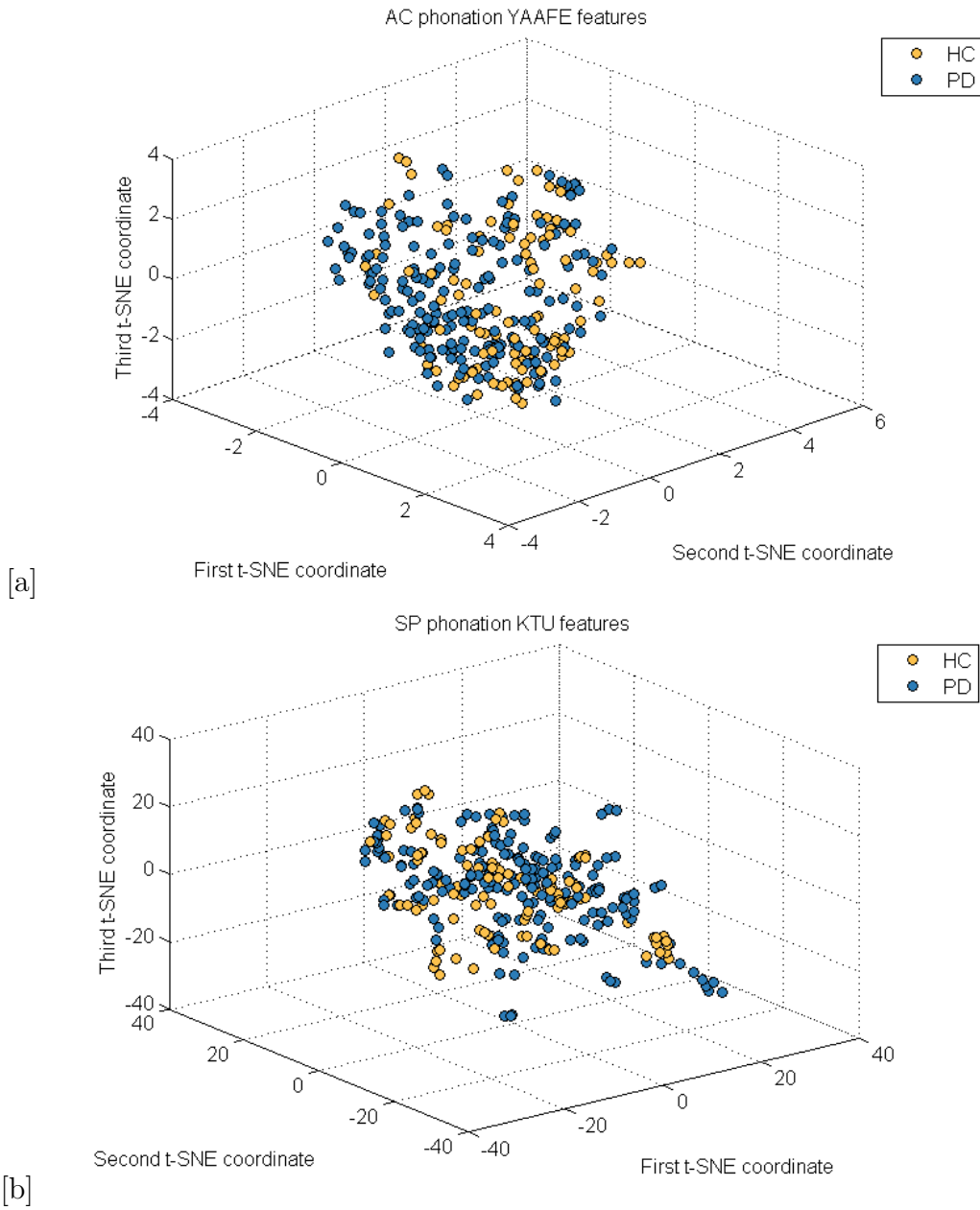


Figura 9 – Visualização de dados pelo t-SNE.: (a) AC - acústico cardióide, fonação e extrator YA; e (b) SP-smartphone, fonação e extrator KT. Gravação de: PD - Pacientes com doença de Parkinson e HC - indivíduo saudável.

Tabela 4 – Passo 1a - Escolha do extrator e classificador com menor EER (%),

Extrator Classif,		Modalidade canal AC				Modalidade canal SP				Extrator Classif,		Modalidade canal AC				Modalidade canal SP			
		P	S	U	V	P	S	U	V			P	S	U	V	P	S	U	V
AV1	K1	29,60	41,65	46,16	44,50	30,20	46,49	41,50	50,74	K1	29,55	46,63	46,77	43,14	31,20	46,38	45,66	50,98	
	K3	34,83	39,79	44,14	41,90	36,44	41,09	48,19	41,75	K3	37,46	42,07	43,54	37,77	35,50	45,52	45,73	34,48	
	K5	35,75	41,03	45,03	37,32	37,28	41,95	42,38	39,96	K5	36,38	38,21	38,06	36,16	38,17	37,62	44,12	37,21	
	M	52,59	51,97	50,68	52,32	51,51	53,89	53,66	55,06	M	52,69	51,79	57,09	53,02	53,48	57,92	54,49	55,09	
	OB	41,39	45,83	42,80	38,97	42,53	46,21	39,26	46,23	OB	41,33	46,36	40,01	38,01	37,85	45,88	39,20	47,77	
	OC	37,22	47,61	37,32	43,52	39,41	46,90	43,28	46,36	OC	33,44	49,27	33,49	49,36	37,43	46,97	37,43	46,07	
	OCS	59,13	46,50	58,26	46,73	57,32	58,77	56,92	48,66	AV2	OCS	57,84	58,70	52,74	53,60	48,70	52,65	54,94	53,01
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	30,92	38,36	38,24	41,37	28,91	50,81	43,57	47,87	OMN	31,00	41,65	43,18	49,34	30,12	48,05	41,33	49,34	
	OSCS	32,79	55,38	38,79	47,09	38,76	44,55	46,23	49,37	OSCS	30,67	50,93	42,31	46,78	36,07	46,86	45,94	53,19	
	OSC	57,22	66,88	59,51	52,74	55,29	53,32	59,17	54,24	OSC	57,53	59,08	61,92	52,06	56,49	55,87	61,78	50,72	
	SL	30,46	37,54	33,46	42,38	27,48	36,45	40,39	38,79	SL	31,55	38,17	39,17	38,80	25,73	41,23	41,67	41,11	
	SP	32,19	40,25	32,65	37,21	30,56	30,56	37,32	31,03	SP	31,95	35,50	40,13	32,88	28,56	31,31	40,14	31,09	
	SR	30,13	34,77	39,82	17,28	26,92	31,96	27,44	31,04	SR	29,74	36,91	36,59	30,79	25,35	22,75	35,22	26,29	
EL	K1	20,25	46,07	44,3	44,29	21,82	44,86	45,91	35,90	K1	25,20	44,10	46,55	39,56	23,86	45,37	44,65	50,68	
	K3	24,66	42,28	38,90	41,31	28,77	40,36	36,27	40,30	K3	34,29	40,22	32,53	37,32	31,37	31,77	44,04	43,28	
	K5	25,85	38,21	43,60	38,76	32,03	33,13	37,32	36,39	K5	36,88	36,08	38,04	35,75	33,99	34,48	38,97	38,25	
	M	26,75	55,34	56,83	56,27	55,31	55,16	55,38	53,78	M	43,87	47,97	47,22	49,24	45,59	49,89	46,27	50,65	
	OB	28,99	42,76	44,29	48,33	36,86	39,91	50,68	46,51	OB	39,66	46,11	41,72	49,16	40,32	44,88	40,04	42,19	
	OC	30,22	36,52	44,74	46,07	34,71	43,90	36,90	46,07	OC	33,33	38,50	44,22	48,53	36,71	47,00	43,89	47,66	
	OCS	30,50	59,83	43,65	49,77	45,67	47,05	48,82	53,59	EMI	OCS	46,82	54,55	45,83	58,86	51,93	54,38	56,52	51,74
	OE	32,38	46,95	41,06	43,49	27,19	48,33	37,70	43,20	OE	29,93	45,34	41,86	48,42	31,02	38,80	42,44	44,58	
	OMS	35,25	62,73	62,73	62,73	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	51,78	41,86	41,98	42,16	25,75	38,94	43,35	44,40	OMN	27,99	49,80	42,04	46,64	26,75	42,36	46,5	45,96	
	OSCS	56,09	41,07	45,31	45,27	32,88	51,78	46,52	54,91	OSCS	32,58	40,13	46,91	49,30	37,40	50,13	47,82	42,45	
	OSC	57,56	52,94	39,01	51,64	53,32	57,11	53,41	44,11	OSC	57,54	54,63	61,45	51,70	51,36	60,7	61,01	55,23	
	SL	59,49	39,32	42,49	39,73	25,55	38,17	40,44	41,61	SL	32,92	32,67	43,51	36,00	30,14	38,69	35,15	41,37	
	SP	60,39	29,29	34,71	34,47	27,08	38,21	36,52	31,42	SP	29,15	34,77	39,65	34,76	27,99	28,8	33,35	25,97	
SR	62,62	32,82	21,83	34,31	25,28	34,36	36,77	26,65	SR	23,36	34,14	28,60	27,65	21,76	26,81	33,76	27,04		
EM2	K1	30,10	48,71	43,64	42,10	28,59	43,45	46,70	43,93	K1	21,29	48,68	42,21	46,02	23,51	44,14	46,45	49,39	
	K3	39,30	45,35	39,40	39,90	33,74	39,46	48,20	44,82	K3	30,62	50,99	39,48	46,81	30,32	41,09	41,69	46,20	
	K5	39,37	39,98	36,77	39,30	36,63	44,12	43,57	36,73	K5	31,40	46,59	38,90	48,58	33,10	44,21	35,19	44,58	
	M	50,29	55,97	54,06	54,10	49,16	54,78	56,81	54,18	M	45,98	53,01	50,12	55,93	55,5	53,25	57,81	56,66	
	OB	39,81	49,21	43,30	44,20	41,20	48,12	40,00	45,04	OB	34,64	49,10	45,61	47,17	38,26	45,04	38,47	47,69	
	OC	33,47	47,61	46,77	39,40	38,36	51,25	41,65	46,96	OC	35,02	47,12	37,84	50,36	34,88	46,71	39,58	43,93	
	OCS	51,13	47,33	54,00	58,40	51,44	61,55	56,12	55,52	ED	OCS	53,97	55,01	52,09	47,09	50,12	45,25	55,41	48,04
	OE	29,99	50,05	38,24	43,50	32,58	49,13	46,45	44,22	OE	27,36	50,57	43,29	50,36	31,40	45,42	45,33	43,86	
	OMS	62,62	62,73	62,73	62,70	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	30,40	45,81	40,45	43,50	27,73	45,05	49,14	41,51	OMN	25,66	50,12	42,45	49,25	29,10	41,31	42,81	43,45	
	OSCS	35,89	52,23	43,70	45,90	39,62	58,50	43,22	45,79	OSCS	32,45	44,99	40,88	45,48	33,24	47,82	39,47	41,65	
	OSC	60,24	47,30	60,21	64,10	56,25	56,50	51,74	45,01	OSC	55,69	55,71	48,71	57,48	53,11	50,42	59,3	54,91	
	SL	35,36	43,42	42,28	40,60	31,49	37,33	45,02	44,18	SL	24,10	36,40	34,22	36,54	26,58	42,82	42,08	44,66	
	SP	32,86	29,73	37,29	36,90	32,26	26,55	36,87	34,50	SP	25,12	39,91	35,50	36,45	29,06	33,88	40,15	41,67	
SR	27,16	34,56	39,46	37,80	27,20	30,46	36,73	35,41	SR	15,40	26,67	37,56	36,87	21,43	38,63	37,62	33,01		

Nota: Microfones: AC-cardióide acústico, SP-smartphone, Modalidade: P-fonação, S-fala, V-dublado, U-mudo,

Tabela 5 – Passo 1b - Escolha do extrator e classificador com menor EER (%),

Extrator Classif,		Modalidade canal AC				Modalidade canal SP				Extrator Classif,		Modalidade canal AC				Modalidade canal SP			
		P	S	U	V	P	S	U	V			P	S	U	V	P	S	U	V
IS1	K1	27,72	49,58	44,53	54,82	22,55	48,91	41,30	32,38	K1	29,10	44,01	41,23	38,52	27,08	45,36	49,58	46,58	
	K3	38,08	35,54	42,66	45,14	27,36	38,95	41,06	30,46	K3	37,59	43,39	39,27	42,04	33,76	38,88	49,48	43,35	
	K5	39,92	38,46	41,84	34,40	32,45	26,60	41,45	28,05	K5	43,94	36,48	44,41	40,65	37,44	44,14	46,25	39,92	
	M	42,24	46,71	42,79	50,75	39,66	50,43	44,23	48,21	M	53,31	53,82	54,62	50,20	52,06	54,29	54,20	56,98	
	OB	41,04	47,87	52,35	46,83	40,38	51,35	40,22	44,94	OB	39,33	41,15	44,07	42,51	38,58	47,59	33,56	38,97	
	OC	39,82	50,78	43,21	52,27	40,93	48,94	43,41	51,30	OC	36,41	43,75	41,55	43,57	35,04	44,49	39,32	41,23	
	OCS	45,69	42,90	48,84	39,79	59,01	47,06	51,17	50,89	IS2	OCS	55,50	45,31	58,04	50,50	57,04	37,93	41,95	50,85
	OE	33,82	51,90	40,92	41,17	27,08	39,68	47,45	40,41	OE	33,83	49,87	40,68	36,29	31,57	46,55	47,06	43,90	
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	28,58	51,74	38,36	41,61	25,26	48,71	44,12	43,87	OMN	31,05	48,68	40,39	46,63	30,30	44,98	42,99	40,86	
	OSCS	38,24	49,97	47,31	45,63	36,81	41,68	39,82	47,05	OSCS	37,35	47,44	40,88	41,75	34,47	47,56	43,41	41,76	
	OSC	57,70	50,34	47,64	60,09	55,11	52,60	57,69	57,00	OSC	56,68	54,79	53,87	67,01	59,64	54,57	59,49	51,74	
	SL	30,94	40,07	33,95	38,61	29,81	37,11	44,30	28,01	SL	32,33	43,14	36,45	39,10	28,13	35,36	43,00	39,79	
SP	33,11	33,77	38,62	37,51	31,01	26,38	42,28	37,43	SP	28,39	38,49	36,10	38,00	29,74	33,10	37,79	34,87		
SR	30,21	29,63	27,64	38,24	29,41	30,43	32,25	35,75	SR	23,19	36,59	37,89	33,20	23,63	36,10	20,37	34,59		
IS3	K1	25,10	50,17	41,91	44,53	25,01	46,07	50,94	46,69	K1	28,59	41,90	44,49	43,68	31,89	36,10	48,44	47,44	
	K3	36,30	44,95	43,22	44,86	32,93	34,58	50,16	45,60	K3	37,53	41,27	39,26	41,65	39,12	36,59	42,86	45,05	
	K5	42,16	45,22	40,30	36,87	38,58	40,03	45,04	36,09	K5	40,24	37,84	43,12	40,61	41,13	35,32	40,27	37,25	
	M	53,21	53,73	56,62	52,67	49,90	48,33	51,97	56,98	M	56,57	51,78	49,81	50,75	54,93	55,43	53,10	57,13	
	OB	36,68	49,48	49,82	46,81	38,69	47,44	36,70	44,57	OB	39,55	37,91	45,34	41,80	41,89	43,33	38,44	44,24	
	OC	34,90	53,97	46,50	42,88	34,46	48,00	42,28	44,66	OC	34,97	39,20	45,94	51,90	39,52	46,33	36,72	43,54	
	OCS	52,63	47,60	54,34	54,25	48,61	48,35	46,50	48,84	IS4	OCS	50,51	55,07	54,76	60,31	46,11	49,06	52,04	62,42
	OE	32,12	53,30	44,06	44,62	31,21	49,29	48,44	47,75	OE	36,42	41,10	38,24	42,72	33,24	38,94	49,83	44,04	
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	31,10	48,64	43,03	44,70	30,05	45,44	47,56	46,88	OMN	29,92	41,96	36,72	42,61	29,10	38,63	48,16	43,09	
	OSCS	35,06	49,63	48,92	41,43	34,97	48,12	38,84	46,32	OSCS	36,29	45,21	46,09	45,52	37,68	45,64	44,49	46,71	
	OSC	57,23	57,97	51,70	58,64	57,37	51,14	48,30	54,78	OSC	57,88	56,04	60,88	57,56	58,09	56,99	52,99	46,44	
	SL	31,17	44,70	37,89	38,29	28,53	35,28	43,22	38,25	SL	30,10	37,79	35,47	42,28	26,36	34,29	37,04	40,05	
SP	28,29	39,75	36,00	40,00	27,45	38,53	39,53	31,01	SP	29,10	34,27	35,97	28,15	22,01	27,08	36,87	41,96		
SR	21,53	31,72	34,33	31,84	21,10	28,45	27,31	32,99	SR	27,10	22,61	35,18	36,32	27,26	29,16	38,48	32,10		
IS5	K1	25,18	46,36	39,73	42,86	31,25	42,43	45,64	42,60	K1	29,10	43,71	44,66	43,35	29,01	42,07	43,18	45,99	
	K3	32,85	42,95	35,75	41,68	34,50	42,49	47,76	39,53	K3	30,08	44,41	37,80	36,45	32,50	46,28	47,46	38,46	
	K5	35,47	41,22	39,40	39,22	37,18	41,21	45,63	41,65	K5	36,35	40,95	33,18	35,92	33,05	39,66	45,47	37,51	
	M	54,13	54,95	53,89	49,49	57,47	55,13	54,48	54,01	M	57,83	55,77	56,28	55,65	58,13	49,32	50,93	52,77	
	OB	39,99	38,97	45,19	38,76	41,39	47,37	47,42	43,14	OB	37,89	37,35	48,20	39,98	39,14	39,24	41,43	44,79	
	OC	38,26	44,70	43,57	39,22	39,61	47,64	41,31	43,89	OC	37,94	42,21	45,80	44,05	36,28	51,78	44,07	37,63	
	OCS	56,41	51,53	54,43	54,42	50,31	51,25	52,20	54,95	IS6	OCS	53,29	60,27	57,71	52,04	48,72	51,02	49,53	53,79
	OE	34,35	50,55	42,69	44,54	31,35	44,67	39,41	45,06	OE	34,63	48,81	41,21	41,80	33,52	41,17	43,09	44,55	
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73	
	OMN	28,98	47,05	36,6	46,63	30,39	38,73	40,97	42,58	OMN	25,01	48,56	41,65	40,74	30,10	42,91	41,98	45,07	
	OSCS	32,61	42,98	43,18	41,20	36,77	47,01	39,73	44,99	OSCS	36,18	43,76	43,18	46,18	39,38	49,87	41,20	43,70	
	OSC	56,84	56,12	54,93	52,25	25,37	58,64	54,63	53,08	OSC	58,56	60,21	66,53	54,85	54,32	50,10	58,50	51,64	
	SL	33,32	34,07	30,99	39,48	25,36	33,55	31,27	38,26	SL	32,45	40,91	28,02	43,07	29,68	35,36	35,34	42,88	
SP	29,41	36,91	37,61	37,34	29,54	32,49	31,61	36,23	SP	29,91	30,99	28,36	38,63	28,37	30,10	35,63	32,93		
SR	26,38	34,71	32,60	31,87	22,41	35,22	31,38	40,70	SR	25,90	30,43	26,10	33,16	25,83	27,26	29,02	33,87		
Nota: Microfones: AC-cardiófone acústico. SP-smartphone. Modalidade: P-fonção. S-fala. V-dublado. U-mudo.																			

Nota: Microfones: AC-cardióide acústico, SP-smartphone, Modalidade: P-fonação, S-fala, V-dublado, U-mudo,

Tabela 6 – Passo 1c - Escolha do extrator e classificador com menor EER (%),

Extrator Classif,		Modalidade canal AC				Modalidade canal SP				Extrator Classif,		Modalidade canal AC				Modalidade canal SP			
		P	S	U	V	P	S	U	V			P	S	U	V	P	S	U	V
IS7	K1	25,14	43,86	43,28	42,95	29,99	39,28	41,94	41,64	JA	K1	29,04	48,04	44,67	47,34	29,91	44,66	36,30	48,00
	K3	35,98	37,50	38,16	44,53	32,72	39,28	39,86	40,41		K3	30,28	45,44	43,87	45,50	31,30	47,00	41,65	48,71
	K5	36,71	37,02	37,03	42,66	36,22	34,59	41,09	42,35		K5	29,73	41,23	46,63	41,31	35,23	37,04	40,22	36,07
	M	58,44	53,33	51,74	53,61	54,85	57,08	50,46	54,84		M	52,20	54,16	50,87	55,64	50,28	51,98	53,75	58,01
	OB	39,97	42,75	44,19	44,17	40,46	40,81	39,68	36,07		OB	38,23	45,17	49,52	44,22	39,05	53,04	48,26	42,63
	OC	38,54	50,94	45,50	44,86	36,79	44,72	32,28	40,54		OC	37,78	52,81	48,58	49,61	38,97	45,60	41,09	43,33
	OCS	52,11	50,61	49,55	59,05	54,60	53,56	48,86	52,88		OCS	55,48	47,84	53,32	55,42	56,87	52,20	53,43	41,54
	OE	33,75	45,15	42,69	45,65	35,00	38,52	43,37	43,66		OE	32,17	49,10	45,68	53,86	34,61	50,86	41,64	52,70
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73		OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73
	OMN	28,87	44,41	40,10	33,29	30,18	42,83	42,71	45,50		OMN	35,18	51,33	44,56	56,18	35,12	47,05	46,07	49,87
	OSCS	38,95	44,19	48,20	41,69	37,36	47,94	36,15	45,50		OSCS	34,04	52,71	47,05	54,52	35,25	46,33	43,94	44,70
	OSC	60,13	68,32	67,92	53,51	53,74	55,36	55,08	50,45		OSC	54,93	54,05	44,95	52,58	37,37	50,31	47,14	53,83
	SL	32,15	35,87	34,45	40,74	28,47	34,13	40,39	41,86		SL	26,73	30,41	28,58	41,33	29,71	38,53	42,19	43,70
	SP	26,84	31,15	36,60	34,97	27,20	33,94	31,67	41,10		SP	30,50	40,86	37,05	45,64	31,58	37,23	41,24	45,44
	SR	27,07	34,25	34,20	25,30	23,55	29,78	35,22	28,28		SR	23,53	32,59	31,87	30,31	24,67	28,10	34,13	35,41
KT	K1	19,38	45,42	47,18	38,63	14,16	47,44	49,30	53,56	MP	K1	19,95	49,71	48,57	45,50	21,48	47,60	30,84	58,22
	K3	24,46	39,40	42,95	34,11	26,68	44,59	39,02	48,66		K3	20,99	37,39	43,96	43,47	25,73	42,12	43,90	51,25
	K5	28,52	38,08	37,67	31,42	28,81	43,44	42,63	35,06		K5	17,95	44,58	34,59	50,49	23,84	45,40	48,58	44,95
	M	47,19	52,16	48,17	52,67	50,79	51,47	52,26	56,37		M	35,84	50,74	40,98	47,74	41,18	52,33	45,63	52,48
	OB	36,75	54,03	50,94	52,14	37,21	55,11	45,32	49,52		OB	40,22	46,25	47,77	45,27	39,18	47,12	44,91	45,81
	OC	38,63	47,06	53,87	51,00	36,45	50,07	38,63	45,92		OC	37,10	48,16	34,74	48,54	37,78	50,28	45,21	44,59
	OCS	46,70	33,96	54,74	42,83	45,44	56,80	47,83	51,11		OCS	57,90	57,13	57,04	46,63	53,31	46,48	49,77	59,94
	OE	22,67	41,77	46,80	41,64	24,47	46,84	50,36	49,36		OE	21,66	52,54	51,43	49,32	26,01	48,37	39,79	52,76
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73		OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73
	OMN	18,95	43,30	44,01	38,26	22,69	52,05	44,84	53,08		OMN	24,93	51,94	45,04	50,57	25,87	42,42	40,28	44,91
	OSCS	34,16	44,22	48,21	52,19	37,51	45,95	40,63	51,14		OSCS	31,89	51,86	45,83	49,69	35,48	50,75	46,41	48,83
	OSC	47,12	46,72	49,44	58,18	48,56	58,64	50,45	51,25		OSC	60,30	59,76	55,22	42,92	55,59	49,40	56,06	46,69
	SL	25,59	36,22	37,30	37,04	30,48	40,22	31,39	40,28		SL	23,93	34,50	37,70	38,34	31,06	43,18	45,41	40,87
	SP	26,87	25,62	39,06	32,25	28,73	19,64	31,38	38,95		SP	28,11	39,40	38,01	40,54	31,09	35,90	37,02	36,55
	SR	18,98	29,37	33,79	30,50	17,21	26,19	32,42	33,20		SR	21,03	39,02	33,47	31,22	26,06	32,07	34,39	38,67
TS	K1	29,33	46,21	46,08	40,27	26,07	39,09	38,01	40,19	YA	K1	19,10	43,28	38,05	42,38	20,47	45,25	48,16	43,75
	K3	30,84	39,91	38,57	38,69	32,38	40,15	40,58	31,18		K3	27,35	40,86	37,08	40,41	26,52	40,04	39,91	42,71
	K5	33,42	39,27	33,07	31,35	37,37	34,41	36,20	30,45		K5	31,14	40,51	46,49	38,36	29,63	38,88	42,56	38,01
	M	35,08	47,33	41,35	47,92	35,32	45,79	44,66	45,41		M	49,06	52,07	53,30	53,65	52,18	55,34	53,57	52,06
	OB	44,48	36,39	42,80	53,75	43,60	38,54	43,01	53,76		OB	36,90	45,78	39,62	45,03	36,38	43,57	38,97	44,10
	OC	40,29	43,88	47,76	45,98	43,02	40,61	47,83	48,65		OC	32,05	48,24	43,94	46,16	33,32	48,10	40,98	47,02
	OCS	53,15	61,11	56,95	43,99	48,55	33,74	57,91	60,32		OCS	57,88	57,65	50,56	50,11	50,19	51,30	37,10	38,63
	OE	32,43	41,79	45,36	42,26	31,14	38,17	46,15	47,44		OE	26,17	48,07	38,19	48,67	25,90	46,43	47,36	45,64
	OMS	62,62	62,73	62,73	62,73	62,73	62,73	62,73	62,73		OMS	62,62	62,73	62,38	62,73	62,73	62,73	62,73	62,73
	OMN	31,89	42,17	41,90	51,04	32,28	41,31	41,31	47,66		OMN	26,49	45,52	39,71	46,50	23,46	44,00	39,91	50,20
	OSCS	36,05	42,44	49,71	49,70	38,05	39,17	52,71	50,88		OSCS	31,56	48,09	40,63	44,71	28,93	48,43	41,98	52,50
	OSC	53,64	69,13	40,66	49,42	53,80	59,86	47,95	43,21		OSC	61,15	69,35	59,68	57,49	49,84	56,42	54,93	51,99
	SL	31,86	36,39	31,90	39,71	32,82	33,85	40,75	31,98		SL	22,24	36,23	36,01	41,00	30,93	37,96	36,99	40,77
	SP	28,61	40,28	35,87	38,06	30,90	38,88	35,00	39,17		SP	24,43	32,49	32,25	34,83	21,93	31,78	35,85	34,22
	SR	26,06	37,96	35,02	21,81	26,97	41,75	33,82	44,21		SR	5,29	36,46	33,74	39,17	9,46	31,14	33,29	32,91

Nota: Microfones: AC-cardióide acústico, SP-smartphone, Modalidade: P-fonação, S-fala, V-dublado, U-mudo,

Tabela 7 – Passo 2a - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
(P) Fonação	AV1	K1	85,47±3,69	85,84±3,73	85,84±3,73	76,26	29,6
		M	66,47±5,34	56,60±7,71	57,13±5,65	46,54	52,59
		OMN	83,54±3,38	84,16±2,38	84,16±2,38	74,67	30,92
		SR	73,01±6,27	79,01±3,90	79,01±3,90	75,63	30,13
	AV2	K1	86,72±3,19	86,53±3,04	86,53±3,04	76,33	29,55
		M	67,44±7,88	57,33±9,98	57,33±9,98	46,41	52,69
		OMN	87,54±3,08	86,24±2,40	86,24±2,40	74,58	31,0
		SR	74,37±5,69	79,90±3,43	79,90±3,43	76,10	29,74
	EL	K1	91,04±4,07	90,99±3,29	90,99±3,29	81,97	20,25
		M	58,86±4,65	46,53±6,88	46,53±6,88	39,97	26,75
		OB	71,48±5,95	76,93±4,22	76,93±4,22	72,43	28,99
		SR	83,31±7,72	88,22±4,34	88,22±4,34	86,67	32,82
	EM1	K1	90,52±2,63	90,50±3,65	90,50±3,66	81,36	25,2
		M	75,20±3,51	70,89±4,06	70,89±4,07	58,14	43,87
		OMN	89,41±2,94	88,71±2,80	88,71±2,81	78,17	27,99
		SR	82,54±5,35	86,83±4,64	86,83±4,65	83,40	23,36
	EM2	K1	83,93±4,30	83,96±3,15	83,96±3,16	73,74	30,1
		M	67,39±5,15	60,59±4,05	60,59±4,05	49,61	50,29
		OE	86,02±5,27	85,94±4,31	85,94±4,31	75,79	29,99
		SR	64,02±5,54	75,54±3,48	75,54±3,49	79,14	27,16
	ED	K1	90,01±4,92	91,58±3,54	91,58±3,54	85,60	21,29
		M	70,24±7,94	66,63±6,70	66,63±6,71	55,35	45,98
		OMN	86,66±2,74	88,22±2,36	88,22±2,37	80,85	25,66
		SR	81,38±4,78	88,32±3,50	88,32±3,51	91,32	15,4
	IS1	K1	84,21±7,54	86,04±5,09	86,04±5,10	78,48	27,72
		M	66,82±4,70	68,51±4,28	68,51±4,29	60,29	42,24
		OMN	82,15±3,84	84,55±3,01	84,55±3,02	77,48	28,58
		SR	71,76±5,23	78,32±2,48	78,32±2,49	75,53	30,21
	IS2	K1	84,53±3,97	85,05±3,76	85,05±3,77	75,55	29,1
		M	67,27±2,59	56,34±5,32	56,34±5,33	45,58	53,31
		OMN	86,47±3,85	85,64±3,35	85,64±3,36	74,52	31,05
		SR	74,97±3,52	82,77±2,17	82,77±2,18	83,59	23,19
	IS3	K1	85,48±4,76	87,43±1,88	87,43±1,89	80,42	25,1
		M	69,07±5,16	57,13±7,40	57,13±7,41	45,72	53,21
		OMN	87,46±2,30	85,64±3,16	85,64±3,17	73,34	31,1
		SR	79,11±6,66	85,54±5,12	85,54±5,13	85,36	21,53
(S) Fala	AV1	K5	46,79±7,12	61,47±4,82	61,47±4,83	61,88	41,02
		M	60,94±11,20	56,18±9,06	56,18±9,07	47,37	51,96
		OMN	53,58±4,95	65,59±6,31	65,59±6,32	65,35	38,35
		SR	61,01±12,03	70,88±6,23	70,88±6,24	69,92	34,76
	AV2	K5	52,66±5,45	65,29±4,88	65,29±4,89	65,54	38,20
		M	62,02±6,60	56,76±4,16	56,76±4,17	47,61	51,79
		OMN	57,96±4,17	65,29±5,70	65,29±5,71	61,07	41,64
		SP	58,69±9,11	69,41±7,34	69,41±7,35	68,99	35,50
	EL	K5	49,75±6,65	64,12±7,41	64,12±7,42	65,54	38,20
		M	52,93±11,37	49,12±12,68	49,12±12,69	42,90	55,33
		OC	53,30±8,27	66,47±6,33	66,47±6,34	67,70	36,51
		SP	57,70±10,97	71,76±6,19	71,76±6,20	76,63	29,29
	EM1	K5	50,01±5,17	65,29±5,22	65,29±5,23	68,26	36,08
		M	71,44±7,19	65,00±8,96	65,00±8,97	52,70	47,97
		OC	64,48±9,23	70,29±8,46	70,29±8,47	65,16	38,50
		SL	61,27±6,95	72,06±2,71	72,06±2,72	72,54	32,66
	EM2	K5	50,02±8,95	63,24±7,11	63,24±7,12	63,25	39,98
		M	61,90±9,20	50,29±5,33	50,29±5,34	42,06	55,96
		OMN	53,34±9,16	60,29±7,35	60,29±7,36	55,58	45,80
		SP	44,87±8,13	65,59±4,16	65,59±4,17	76,11	29,73
	ED	K5	45,68±6,23	57,35±5,30	57,35±5,31	54,53	46,59
		M	60,90±9,21	54,71±9,50	54,71±9,51	45,98	53,01
		OSCS	60,95±11,03	63,82±8,82	63,82±8,83	56,66	44,98
		SR	55,81±12,94	71,76±6,73	71,76±6,74	79,70	26,66
	IS1	K3	56,16±9,84	68,24±7,75	68,24±7,76	68,95	35,54
		M	66,57±8,27	64,41±6,50	64,41±6,51	54,38	46,70
		OCS	45,97±9,42	54,41±7,58	54,41±7,59	50,03	49,97
		SR	59,16±4,49	72,35±4,40	72,35±4,41	76,23	29,63
	IS2	K5	46,02±9,36	63,53±5,61	63,53±5,62	67,76	36,47
		M	57,88±13,00	52,65±11,11	52,65±11,12	44,91	53,82
		OB	49,54±5,07	62,35±4,88	62,35±4,89	61,72	41,15
		SR	40,89±8,26	61,76±4,92	61,76±4,93	67,61	36,59
	IS3	K3	42,97±4,90	57,94±6,17	57,94±6,18	56,72	44,94
		M	65,23±7,11	55,00±6,58	55,00±6,59	45,02	53,73
		OCS	37,52±3,62	55,59±3,59	55,59±3,60	53,19	47,60
		SR	38,34±5,24	62,65±4,16	62,65±4,17	73,71	31,71

Tabela 8 – Passo 2b - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
(P) Fonação	IS4	K1	85,49±3,75	86,34±3,98	86,34±3,99	77,44	28,59
		M	59,04±8,23	48,42±7,73	48,42±7,74	41,27	56,57
		OMN	84,64±5,05	85,25±4,34	85,25±4,35	75,49	29,92
		SR	71,66±6,62	78,81±3,65	78,81±3,66	76,99	27,1
	IS5	K1	79,55±4,55	84,55±2,97	84,55±2,98	81,38	25,18
		M	59,18±12,28	52,48±6,32	52,48±6,33	44,49	54,13
		OMN	86,58±3,37	86,73±2,97	86,73±2,98	77,00	28,98
		SR	66,87±6,40	77,33±4,02	77,33±4,03	80,03	26,38
	IS6	K1	81,42±2,84	83,47±4,79	83,47±4,80	75,70	29,1
		M	58,23±7,23	45,84±8,33	45,84±8,34	39,60	57,83
		OMN	85,02±4,41	86,83±3,86	86,83±3,87	79,44	25,01
		SR	68,41±5,10	78,32±3,14	78,32±3,15	80,57	25,9
	IS7	K1	80,62±6,25	85,15±3,67	85,15±3,68	81,43	25,14
		M	58,74±9,51	44,75±6,16	44,75±6,17	38,81	58,44
		OMN	82,40±4,49	84,55±3,04	84,55±3,05	77,14	28,87
		SP	68,81±5,25	77,92±3,96	77,92±3,97	79,50	26,84
	JA	K1	81,81±4,60	84,16±3,45	84,16±3,46	76,93	29,04
		M	63,46±7,93	56,63±14,91	56,63±14,92	47,06	52,2
		OE	75,71±4,62	79,41±3,69	79,41±3,70	73,14	32,17
		SR	77,18±5,24	83,86±3,19	83,86±3,20	83,22	25,53
	KT	K1	92,41±4,16	93,47±2,84	93,47±2,85	87,56	19,38
		M	69,01±6,25	64,85±6,52	64,85±6,53	53,74	47,19
		OMN	86,09±3,63	90,10±2,61	90,10±2,62	87,99	18,95
		SR	83,77±5,69	88,81±2,73	88,81±2,74	87,95	18,98
	MP	K5	66,16±4,06	79,41±2,75	79,41±2,76	88,97	17,95
		M	76,33±3,53	77,62±4,06	77,62±4,07	68,57	35,84
		OE	82,98±4,46	87,62±2,47	87,62±2,48	85,22	21,66
		SR	80,47±4,93	86,44±3,25	86,44±3,26	85,88	21,33
	TS	K1	81,88±4,66	84,06±4,25	84,06±4,26	76,58	29,33
		M	71,87±4,52	75,84±3,37	75,84±3,38	69,52	35,08
		OMN	80,17±5,55	81,88±4,58	81,88±4,59	73,49	31,89
		SR	72,23±6,23	80,30±3,99	80,30±3,99	80,40	26,06
	YA	K1	94,26±3,96	94,55±3,32	94,55±3,33	87,84	19,01
		M	73,73±4,18	64,65±6,29	64,65±6,30	51,24	49,06
		OE	92,34±3,12	91,09±2,80	91,09±2,81	80,27	26,17
		SR	88,81±4,39	93,66±2,43	93,66±2,44	98,32	5,29
(S) Fala	IS4	K5	54,28±8,41	66,18±6,20	66,18±6,21	66,01	37,84
		M	60,19±10,97	56,18±15,19	56,18±15,20	47,62	51,78
		OB	50,82±5,78	64,71±5,09	64,71±5,10	65,93	37,90
		SR	55,16±10,77	72,65±4,75	72,65±4,76	84,21	22,61
	IS5	K5	50,46±7,61	62,65±5,43	62,65±5,44	61,64	41,21
		M	52,87±10,13	49,71±8,86	49,71±8,87	43,41	54,94
		OB	59,25±6,37	67,65±7,08	67,65±7,09	64,56	38,97
		SL	63,33±4,49	72,35±5,76	72,35±5,77	70,80	34,06
	IS6	K5	60,28±8,41	66,76±5,88	66,76±5,89	61,98	40,94
		M	50,39±17,32	47,94±15,22	47,94±15,23	42,32	55,76
		OB	60,97±7,06	69,41±4,95	69,41±4,96	66,64	37,35
		SR	53,77±10,09	69,41±6,73	69,41±6,74	75,26	30,48
	IS7	K5	49,69±6,81	64,71±4,36	64,71±4,37	67,07	37,01
		M	51,72±12,62	51,76±11,70	51,76±11,71	45,56	53,32
		OB	54,03±5,42	62,94±6,19	62,94±6,20	59,62	42,75
		SP	55,61±7,76	70,00±4,32	70,00±4,33	74,39	31,15
	JA	K5	46,25±4,47	61,18±3,90	61,18±3,91	61,62	41,22
		M	55,35±9,86	51,17±7,80	51,17±7,81	44,46	54,16
		OB	49,99±5,58	59,71±4,56	59,71±4,57	56,43	45,16
		SL	61,00±4,89	72,94±3,42	72,94±3,43	75,29	30,40
	KT	K5	60,65±11,51	68,82±6,85	68,82±6,86	65,70	38,08
		M	63,54±11,27	56,76±8,00	56,76±8,01	47,12	52,15
		OCS	42,45±6,24	63,24±3,28	63,24±3,29	70,93	33,96
		SP	53,98±12,91	71,18±5,54	71,18±5,55	80,89	25,62
	MP	K3	47,15±3,20	63,53±2,69	63,53±2,70	66,59	37,39
		M	56,94±12,35	56,47±11,22	56,47±11,23	49,01	50,73
		OB	44,54±6,44	57,35±5,46	57,35±5,47	54,99	46,24
		SL	46,12±8,63	64,41±3,07	64,41±3,08	70,26	34,50
	TS	K5	52,72±1,85	64,71±0,00	64,71±0,01	64,17	39,27
		M	63,72±7,09	62,65±7,89	62,65±7,90	53,55	47,32
		OB	48,93±7,11	64,71±4,92	64,71±4,93	67,87	36,39
		SL	55,84±4,48	67,65±2,63	67,65±2,64	67,87	36,39
	YA	K5	54,61±5,99	64,71±6,57	64,71±6,58	62,56	40,50
		M	63,16±11,25	56,76±9,11	56,76±9,12	47,24	52,06
		OMN	53,50±5,53	60,59±7,57	60,59±7,58	55,95	45,52
		SP	59,28±7,02	71,18±5,39	71,18±5,40	72,75	32,49

Tabela 9 – Passo 2c - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
(U) Mudo	AV1	K3	52,85±8,87	61,47±6,76	61,47±6,77	57,79	44,14
		M	66,28±14,30	59,71±13,02	59,71±13,03	49,09	50,67
		OC	54,98±10,14	66,76±7,78	66,76±7,79	66,68	37,32
		SP	58,20±12,77	70,59±6,16	70,59±6,17	72,56	32,64
	AV2	K5	51,74±6,21	65,00±4,99	65,00±4,100	65,74	38,15
		M	52,84±15,22	46,18±12,20	46,18±12,21	40,58	57,08
		OC	61,49±6,34	71,76±5,29	71,76±5,30	71,52	33,48
		SR	51,31±12,70	65,59±2,29	65,59±2,30	67,62	36,58
	EL	K3	57,79±9,04	67,06±5,39	67,06±5,40	64,65	38,89
		M	48,13±10,03	45,88±10,78	45,88±10,79	40,93	56,82
		OSC	42,46±5,82	61,18±6,14	61,18±6,15	64,50	39,01
		SR	41,24±8,15	65,88±1,95	65,88±1,96	85,05	21,82
	EM1	K3	62,35±7,71	72,65±3,23	72,65±3,24	72,70	32,53
		M	61,88±10,86	62,06±8,56	62,06±8,57	53,70	47,22
		OB	52,13±8,52	62,94±7,80	62,94±7,81	60,97	41,72
		SR	54,79±13,32	70,59±6,57	70,59±6,58	77,45	28,60
	EM2	K5	52,22±5,99	65,88±4,59	65,88±4,60	67,39	36,76
		M	62,15±12,39	53,53±7,97	53,53±7,98	44,59	54,05
		OE	54,12±7,14	65,88±5,12	65,88±5,13	65,50	38,23
		SP	46,23±9,35	63,24±3,01	63,24±3,02	66,72	36,76
	ED	K5	57,79±12,28	67,06±6,28	67,06±6,28	64,65	38,89
		M	56,50±9,90	57,06±14,23	57,06±14,24	49,84	50,11
		OC	54,28±7,34	66,18±6,06	66,18±6,07	66,01	37,84
		SL	68,94±8,61	75,00±5,30	75,00±5,31	70,60	34,22
	IS1	K5	47,38±3,79	61,18±2,56	61,18±2,57	60,81	41,84
		M	71,68±6,55	70,29±6,23	70,29±6,24	59,56	42,79
		OMN	68,09±3,56	72,06±4,78	72,06±4,79	65,34	38,36
		SR	44,81±11,32	66,18±4,21	66,18±4,22	78,58	27,63
	IS2	K3	51,72±4,76	64,71±3,72	64,71±3,73	64,17	39,27
		M	58,00±13,18	51,47±11,10	51,47±11,11	43,85	54,61
		OMN	60,06±10,92	67,06±7,87	67,06±7,88	62,71	40,39
		SP	57,46±9,98	68,53±7,08	68,53±7,09	68,24	36,09
	IS3	K5	48,94±4,07	62,65±4,93	62,65±4,94	62,83	40,30
		M	58,51±11,10	48,24±13,03	48,24±13,04	41,20	56,62
		OMN	55,33±11,84	63,24±7,91	63,24±7,92	59,26	43,02
		SR	48,11±12,02	65,29±3,90	65,29±3,91	70,47	34,32
(V) Dublado	AV1	K5	50,07±5,52	64,71±6,02	64,71±6,03	66,68	37,32
		M	61,54±12,76	55,88±7,78	55,88±7,79	46,90	52,32
		OB	59,25±8,24	67,65±5,88	67,65±5,89	64,56	38,97
		SR	39,72±7,72	65,88±1,95	65,88±1,96	89,60	17,28
	AV2	K5	45,64±9,22	63,53±6,47	63,53±6,48	68,17	36,15
		M	59,98±11,25	54,41±12,97	54,41±12,98	45,97	53,02
		OB	55,20±7,43	66,47±6,05	66,47±6,06	65,79	38,01
		SR	41,25±6,17	63,82±2,64	63,82±2,65	74,83	30,78
	EL	K5	50,50±9,73	64,12±8,40	64,12±8,41	64,83	38,75
		M	49,91±10,62	47,06±6,70	47,06±6,71	41,66	56,26
		OMN	63,09±8,62	67,06±5,39	67,06±5,40	60,39	42,16
		SR	56,10±7,01	68,82±4,20	68,82±4,21	70,50	34,30
	EM1	K5	49,63±9,50	65,29±4,51	65,29±4,52	68,68	35,75
		M	72,31±12,02	63,82±12,41	63,82±12,42	51,00	49,24
		OMN	46,97±11,08	57,65±8,12	57,65±8,13	54,47	46,63
		SR	49,88±10,70	68,53±6,17	68,53±6,18	78,57	27,64
	EM2	K5	52,72±9,24	64,71±8,31	64,71±8,32	64,17	39,27
		M	54,97±12,71	51,47±11,63	51,47±11,64	44,54	54,09
		OC	58,55±7,32	67,06±7,97	67,06±7,98	63,98	39,41
		SP	50,23±9,27	65,00±4,64	65,00±4,65	67,26	36,86
	ED	K1	55,56±10,01	60,88±8,72	60,88±8,73	55,30	46,01
		M	60,61±9,51	50,00±9,84	50,00±9,85	42,11	55,93
		OSCS	55,88±7,38	61,47±6,50	61,47±6,51	56,02	45,47
		SP	51,84±11,44	65,88±6,19	65,88±6,20	67,79	36,45
	IS1	K5	55,56±1,88	68,53±1,34	68,53±1,35	70,38	34,40
		M	56,02±6,41	56,18±5,17	56,18±5,18	49,00	50,74
		OCS	54,93±65,29	65,29±7,05	65,29±7,06	63,49	39,79
		SP	55,90±5,83	67,06±5,06	67,06±5,07	66,44	37,60
	IS2	K1	54,50±8,75	65,88±6,05	65,88±6,06	65,14	38,51
		M	66,22±9,14	60,29±9,59	60,29±9,60	49,73	50,20
		OE	54,38±7,05	67,06±5,54	67,06±5,55	68,00	36,28
		SR	41,69±10,09	63,24±3,01	63,24±3,02	71,88	33,19
	IS3	K5	48,77±12,05	64,41±7,49	64,41±7,50	67,26	36,86
		M	61,22±10,24	55,29±10,58	55,29±10,59	46,43	52,67
		OSCS	54,67±5,21	64,12±5,70	64,12±5,71	61,35	41,42
		SR	44,39±10,89	64,71±2,94	64,71±2,95	73,56	31,84

Tabela 10 – Passo 2d - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
(U) Mudo	IS4	K3	46,51±5,21	62,35±5,22	62,35±5,23	64,18	39,26
		M	59,04±15,38	58,24±14,33	58,24±14,34	50,25	49,81
		OMN	58,21±12,42	68,53±7,67	68,53±7,68	67,45	36,72
		SR	60,85±12,21	70,59±4,55	70,59±4,56	69,41	35,17
	IS5	K3	53,08±12,20	66,76±6,71	66,76±6,72	68,68	35,75
		M	58,26±12,12	52,65±9,96	52,65±9,97	44,81	53,89
		OMN	61,29±7,75	70,00±5,85	70,00±5,86	67,60	36,60
		SL	69,14±7,53	76,76±7,25	76,76±7,26	74,59	30,98
	IS6	K5	54,96±11,51	68,82±7,80	68,82±7,81	71,90	33,18
		M	53,26±15,27	47,65±8,99	47,65±8,100	41,64	56,28
		OE	57,21±8,62	65,29±3,67	65,29±3,68	61,65	41,20
		SR	65,57±7,75	76,47±5,88	76,47±5,89	78,83	27,42
	IS7	K5	51,15±7,28	65,29±5,70	65,29±5,71	67,05	37,03
		M	56,14±16,78	55,00±15,28	55,00±15,29	47,68	51,73
		OMN	58,23±11,66	66,47±9,41	66,47±9,42	63,09	40,10
		SR	62,25±12,11	71,76±6,33	71,76±6,34	70,63	34,20
	JA	K3	47,88±7,73	60,00±7,57	60,00±7,58	58,14	43,87
		M	62,83±11,85	58,24±8,99	58,24±8,100	48,83	50,87
		OMN	47,56±8,96	59,41±7,97	59,41±7,98	57,23	44,55
		SL	48,64±5,73	67,65±4,74	67,65±4,75	77,47	28,58
	KT	K5	44,24±9,42	62,35±7,41	62,35±7,42	66,23	37,67
		M	64,53±7,96	62,06±7,83	62,06±7,84	52,44	48,16
		OMN	53,39±5,15	61,76±5,88	61,76±5,89	57,96	44,01
		SR	48,28±10,82	65,59±3,23	65,59±3,24	71,15	33,78
	MP	K5	52,49±1,02	67,06±4,32	67,06±4,33	70,14	34,59
		M	65,73±7,95	69,12±4,60	69,12±4,61	61,95	40,97
		OC	58,48±5,83	69,71±6,03	69,71±6,04	69,95	34,74
		SR	55,88±6,66	69,12±4,00	69,12±4,01	71,54	33,47
	TS	K5	50,97±5,28	67,06±2,88	67,06±2,89	72,05	33,06
		M	65,03±7,53	68,53±6,71	68,53±6,72	61,46	41,34
		OSC	43,28±5,52	60,59±4,59	60,59±4,60	62,36	40,65
		SL	64,13±11,14	73,82±6,09	73,82±6,10	73,49	31,89
	YA	K3	59,32±8,87	68,18±7,32	68,18±7,33	66,98	37,08
		M	56,56±12,39	52,73±13,08	52,73±13,09	45,60	53,29
		OE	61,28±5,54	68,48±4,32	68,48±4,33	65,57	38,18
		SP	61,75±10,95	71,82±6,91	71,82±6,92	73,05	32,24
(V) Dublado	IS4	K5	44,19±5,13	60,88±6,31	60,88±6,32	62,43	40,60
		M	56,02±14,87	56,18±13,31	56,18±13,32	49,00	50,74
		OB	50,68±9,64	62,35±7,53	62,35±7,54	60,87	41,79
		SP	39,36±4,80	63,82±2,95	63,82±2,96	77,98	28,14
	IS5	K5	45,59±5,75	62,06±4,25	62,06±4,26	64,23	39,22
		M	63,03±8,18	60,00±7,69	60,00±7,70	50,68	49,48
		OB	50,50±8,05	64,12±6,55	64,12±6,56	64,83	38,75
		SR	45,84±11,12	65,29±5,06	65,29±5,07	73,51	31,87
	IS6	K5	42,88±7,71	62,65±6,84	62,65±6,85	68,46	35,92
		M	55,20±9,12	49,12±10,52	49,12±10,53	42,48	55,65
		OB	50,02±7,68	63,24±4,96	63,24±4,97	63,25	39,98
		SR	46,98±7,47	65,29±3,90	65,29±3,91	71,93	33,15
	IS7	K5	49,82±10,23	61,47±5,64	61,47±5,65	59,73	42,66
		M	54,53±11,84	52,06±6,03	52,06±6,04	45,19	53,60
		OMN	56,96±9,89	69,71±5,88	69,71±5,89	71,77	33,28
		SR	44,75±8,72	66,76±1,88	66,76±1,89	81,26	25,29
	JA	K5	55,21±5,38	64,41±7,25	64,41±7,26	61,51	41,31
		M	50,93±11,95	48,24±8,44	48,24±8,45	42,49	55,64
		OB	50,47±9,18	60,59±5,91	60,59±5,92	57,68	44,22
		SR	38,12±4,42	62,94±4,40	62,94±4,41	75,42	30,30
	KT	K5	57,06±7,33	70,59±4,92	70,59±4,93	74,07	31,41
		M	61,22±5,47	55,29±7,75	55,29±7,76	46,43	52,67
		OMN	57,57±5,77	67,35±6,36	67,35±6,37	65,47	38,25
		SR	50,70±10,21	67,94±7,49	67,94±7,50	75,18	30,49
	MP	K3	66,28±7,25	67,35±5,00	67,35±5,01	58,68	43,46
		M	56,81±9,50	59,71±8,32	59,71±8,33	53,01	47,73
		OSC	42,47±6,06	59,12±5,49	59,12±5,50	59,40	42,91
		SR	46,00±9,78	65,59±5,26	65,59±5,27	74,31	31,22
	TS	K5	49,46±9,99	67,06±8,40	67,06±8,41	74,15	31,35
		M	59,02±8,86	60,29±6,60	60,29±6,61	52,77	47,92
		OE	45,39±5,12	60,29±3,77	60,29±3,78	60,27	42,25
		SR	39,24±4,94	65,00±0,88	65,00±0,89	85,07	21,80
	YA	K5	53,58±7,81	65,59±4,93	65,59±4,94	65,35	38,35
		M	58,04±14,52	52,94±11,98	52,94±11,99	45,13	53,65
		OSCS	57,28±6,88	62,65±5,10	62,65±5,11	57,04	44,70
		SP	57,94±12,40	69,41±5,76	69,41±5,77	69,84	34,83

Tabela 11 – Passo 2e - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
Canal SP	AV1	K1	84,01±4,20	84,80±3,04	84,80±3,05	75,54	30,20
		M	72,78±6,61	60,98±5,78	60,98±5,79	47,99	51,50
		OMN	84,67±3,27	85,78±3,04	85,78±3,05	77,09	28,90
		SR	76,67±7,45	82,45±4,00	82,45±4,01	79,41	26,91
	AV2	K1	85,38±3,18	85,00±2,77	85,00±2,78	74,34	31,19
		M	66,72±5,73	55,88±7,42	55,88±7,43	45,36	53,47
		OMN	85,81±3,19	85,78±3,80	85,78±3,81	75,64	30,11
		SR	81,05±3,71	85,39±2,45	85,39±2,46	81,20	25,34
	EL	K1	90,04±2,95	91,47±2,23	91,47±2,24	85,06	21,81
		M	53,67±9,36	49,31±6,56	49,31±6,57	42,93	55,31
		OMN	90,54±3,66	90,29±3,14	90,29±3,15	80,75	25,74
		SR	82,84±2,89	86,37±2,64	86,37±2,65	81,27	25,28
	EM1	K1	87,89±4,96	89,61±4,20	89,61±4,21	82,86	23,86
		M	74,59±2,54	68,92±3,97	68,92±3,98	55,87	45,58
		OMN	84,70±3,46	86,76±3,72	86,76±3,73	79,60	26,75
		SR	74,51±4,88	83,14±4,29	83,14±4,30	85,12	21,75
	(P) Fonação EM2	K1	81,78±4,79	84,41±2,07	84,41±2,08	77,46	28,58
		M	72,06±3,03	63,82±3,23	63,82±3,24	51,11	49,16
		OMN	88,28±4,69	88,24±4,78	88,24±4,79	78,47	27,72
		SR	66,17±8,40	76,86±4,43	76,86±4,44	79,09	27,19
	ED	K1	86,56±5,06	89,02±3,24	89,02±3,25	83,24	23,51
		M	61,83±6,70	51,08±7,52	51,08±7,53	42,68	55,50
		OMN	86,24±3,87	85,88±3,82	85,88±3,83	75,34	30,36
		SR	77,98±1,68	85,10±1,37	85,10±1,38	85,47	21,42
	IS1	K1	84,85±5,09	88,43±3,33	88,43±3,34	84,27	22,55
		M	71,32±5,80	72,65±3,98	72,65±3,99	63,66	39,66
		OMN	83,56±3,02	86,76±2,60	86,76±2,61	81,30	25,26
		SR	77,25±81,67	81,67±3,42	81,67±3,43	76,49	29,41
	IS2	K1	84,59±4,98	82,22±6,93	82,22±6,94	79,23	27,7
		M	69,08±7,27	58,82±9,31	58,82±9,32	47,25	52,05
		OMN	84,85±3,48	85,20±2,71	85,20±2,72	75,43	30,29
		SR	76,01±6,45	83,33±4,72	83,33±4,73	83,11	23,63
	IS3	K1	85,95±5,23	88,14±4,64	88,14±4,65	81,58	25,1
		M	69,66±9,12	61,96±7,29	61,96±7,30	50,13	49,90
		OMN	84,42±2,05	85,10±2,26	85,10±2,27	75,73	30,4
		SR	78,79±6,18	85,20±4,89	85,20±4,90	84,33	22,50
	AV1	K3	54,83±7,38	64,41±4,45	64,41±4,46	61,80	41,08
		M	64,15±7,87	54,41±10,20	54,41±10,21	44,82	53,88
		OSCS	63,65±6,93	65,29±7,64	65,29±7,65	57,24	44,55
		SP	46,16±8,39	65,88±2,35	65,88±2,36	75,11	30,55
	AV2	K5	50,45±7,99	64,71±6,02	64,71±6,03	66,30	37,61
		M	53,17±11,91	44,71±6,80	44,71±6,81	39,49	57,92
		OB	63,76±5,73	64,12±5,54	64,12±5,55	55,49	45,87
		SR	41,62±7,05	65,88±3,52	65,88±3,53	84,06	22,74
	EL	K5	45,52±3,46	64,71±2,63	64,71±2,64	71,96	33,13
		M	50,12±7,01	48,82±6,33	48,82±6,34	43,12	55,16
		OMN	55,79±7,51	66,18±5,46	66,18±5,47	64,59	38,64
		SR	53,57±10,10	67,65±6,57	67,65±6,58	70,43	34,35
	EM1	K3	54,91±9,49	69,41±1,44	69,41±1,45	73,64	31,76
		M	62,71±7,19	59,41±5,85	59,41±5,86	50,14	49,88
		OE	49,05±7,18	63,53±6,47	63,53±6,48	64,78	38,80
		SR	54,73±11,98	71,18±5,99	71,18±5,100	79,53	26,81
	(S) Fala EM2	K3	56,55±10,76	66,18±8,34	66,18±8,35	63,92	39,45
		M	57,84±7,99	51,18±12,06	51,18±12,07	43,64	54,77
		OMN	52,37±9,64	60,59±7,57	60,59±7,58	56,58	45,04
		SP	62,66±16,11	75,29±8,44	75,29±8,45	79,84	26,54
	ED	K3	56,29±11,32	65,00±8,15	65,00±8,16	61,80	41,08
		M	54,85±12,77	52,65±8,67	52,65±8,68	45,67	53,24
		OMN	55,21±8,81	64,41±5,64	64,41±5,65	61,51	41,31
		SP	53,73±12,26	67,94±4,64	67,94±4,65	71,03	33,87
	IS1	K5	62,12±5,06	75,00±3,01	75,00±3,02	79,78	26,59
		M	58,56±10,14	57,35±9,50	57,35±9,51	49,42	50,42
		OE	44,51±3,60	61,47±3,07	61,47±3,08	63,64	39,67
		SP	53,28±5,12	70,59±2,63	70,59±2,64	80,02	26,38
	IS2	K3	59,79±5,83	67,94±3,34	67,94±3,35	64,67	33,88
		M	62,37±10,12	53,24±10,22	53,24±10,23	44,29	54,28
		OCS	43,70±6,31	62,06±4,04	62,06±4,05	65,90	37,92
		SP	64,73±8,38	73,53±3,94	73,53±3,95	72,00	33,09
	IS3	K3	59,55±9,26	70,29±5,64	70,29±5,65	70,16	34,57
		M	63,45±9,76	61,47±7,49	61,47±7,50	52,22	48,33
		OMN	59,71±8,53	62,94±6,60	62,94±6,61	56,07	45,43
		SR	49,18±14,16	67,94±6,23	67,94±6,24	77,63	28,45

Tabela 12 – Passo 2f - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
Canal SP	(P) Fonação	K1	84,21±4,13	84,02±3,87	84,02±3,88	73,49	31,89
		M	61,55±15,37	51,96±12,26	51,96±12,27	43,44	54,92
		OMN	85,23±5,04	85,20±3,73	85,20±3,74	74,96	30,69
		SP	80,04±5,41	85,39±3,46	85,39±3,47	82,82	23,89
		K1	81,66±5,39	83,04±4,25	83,04±4,26	74,27	31,25
		M	56,81±7,07	46,27±8,31	46,27±8,32	40,08	57,47
		OMN	85,70±3,75	85,59±3,59	85,59±3,60	75,31	30,39
		SR	74,71±2,90	83,04±2,87	83,04±2,88	84,42	22,41
		K1	83,88±4,69	84,80±4,02	84,80±4,03	75,70	30,7
		M	49,23±14,04	43,73±8,76	43,73±8,77	39,22	58,12
		OMN	85,02±2,93	84,80±3,40	84,80±3,41	74,32	31,21
		SR	73,80±3,46	81,37±2,80	81,37±2,81	80,66	25,83
		K1	82,68±4,69	84,22±2,78	84,22±2,79	75,80	29,99
		M	59,69±10,55	51,57±11,21	51,57±11,22	43,55	54,84
		OMN	84,37±4,93	85,00±4,90	85,00±4,91	75,56	30,18
		SR	72,83±6,52	81,67±4,27	81,67±4,28	83,19	23,55
		K1	81,47±3,86	83,63±3,89	83,63±3,90	75,89	29,90
		M	52,16±4,15	55,59±8,61	55,59±8,62	49,62	50,28
		OE	78,65±4,63	79,61±3,63	79,61±3,64	70,12	34,60
		SR	79,46±5,99	84,80±4,66	84,80±4,67	81,96	24,67
		K1	89,21±3,62	92,94±2,54	92,94±2,55	92,40	14,15
		M	65,87±3,57	59,41±4,62	59,41±4,63	48,94	50,78
		OMN	87,67±4,06	89,90±3,72	89,90±3,73	84,12	22,69
		SR	82,43±3,45	88,63±2,32	88,63±2,33	89,67	17,20
		K1	85,12±3,88	88,92±3,48	88,92±3,49	85,41	21,48
		M	67,62±3,79	69,80±3,03	69,80±3,04	61,68	41,18
		OMN	86,23±6,11	87,94±4,36	87,94±4,37	80,61	25,87
		SR	76,57±6,48	82,75±3,82	82,75±3,83	80,40	26,06
		K1	85,27±4,37	87,35±3,54	87,35±3,55	80,38	26,07
		M	71,23±4,52	75,49±2,66	75,49±2,67	69,22	35,32
		OE	79,07±3,82	81,76±2,67	81,76±2,68	74,41	31,13
		SR	72,40±8,03	80,20±3,91	80,20±3,92	79,36	26,96
		K1	91,26±3,59	92,55±1,76	92,55±1,77	86,46	20,47
		M	65,07±4,39	57,25±4,62	57,25±4,63	47,10	52,17
		OMN	88,72±2,79	90,20±2,44	90,20±2,45	83,29	23,45
		SR	83,09±4,66	90,29±2,85	90,29±2,86	95,99	9,45
		K5	51,25±8,19	66,18±6,86	66,18±6,87	69,22	35,32
		M	50,71±8,34	48,53±7,23	48,53±7,24	42,77	55,42
		OMN	53,96±8,60	65,59±5,10	65,59±5,11	64,99	38,63
		SP	63,04±8,61	75,29±4,77	75,29±4,78	79,22	27,08
		K5	57,21±7,07	65,29±5,54	65,29±5,55	61,65	41,20
		M	56,60±8,46	50,29±8,04	50,29±8,05	43,17	55,13
		OMN	56,87±11,83	66,76±6,96	66,76±6,97	64,87	38,73
		SP	59,28±7,46	71,18±3,42	71,18±3,43	72,75	32,49
		K5	54,02±8,68	65,00±6,89	65,00±6,90	63,67	39,65
		M	60,82±11,18	59,41±12,87	59,41±12,88	50,90	49,31
		OB	63,62±6,41	69,41±6,05	69,41±6,06	64,21	39,24
		SR	67,73±9,26	77,65±5,91	77,65±5,92	79,02	27,25
		K5	52,49±3,46	67,06±4,88	67,06±4,89	70,14	34,59
		M	49,11±11,09	45,59±10,20	45,59±10,21	40,60	57,07
		OE	54,50±8,72	65,88±6,33	65,88±6,34	65,14	38,51
		SR	62,24±8,41	73,82±5,64	73,82±5,65	76,05	29,78
		K5	48,23±8,00	64,12±4,88	64,12±4,89	67,04	37,03
		M	50,63±6,30	53,24±6,63	53,24±6,64	47,36	51,97
		OC	42,27±4,54	57,35±4,21	57,35±4,22	55,86	45,59
		SR	50,80±10,53	68,82±4,40	68,82±4,41	78,04	28,10
		K5	44,91±5,49	59,41±6,55	59,41±6,56	58,70	43,44
		M	68,39±6,72	59,41±5,06	59,41±5,07	48,04	51,46
		OSCS	59,39±12,06	62,35±11,06	62,35±11,07	55,39	45,94
		SP	50,94±6,91	71,18±4,51	71,18±4,52	87,30	19,63
		K3	44,09±5,07	60,00±3,52	60,00±3,53	60,44	42,12
		M	59,71±8,17	55,29±4,88	55,29±4,89	46,89	52,33
		OMN	51,81±9,04	62,35±7,64	62,35±7,65	60,06	42,41
		SR	45,30±11,20	65,00±5,00	65,00±5,01	73,27	32,07
		K5	63,71±5,62	72,35±2,69	72,35±2,70	70,37	34,41
		M	64,30±7,74	64,41±5,17	64,41±5,18	55,59	45,79
		OCS	45,36±6,95	64,41±9,61	64,41±9,62	71,21	33,74
		SL	56,26±10,80	69,12±4,78	69,12±4,79	71,07	33,84
		K5	52,88±7,83	65,00±6,09	65,00±6,10	64,67	38,88
		M	55,52±11,00	49,71±7,37	49,71±7,38	42,90	55,33
		OB	52,92±6,47	63,82±7,21	63,82±7,22	58,54	43,56
		SR	62,83±7,38	73,53±6,16	73,53±6,17	74,40	31,14

Tabela 13 – Passo 2g - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
Canal SP	(U) Mudo	K1	63,95±5,64	67,94±4,25	67,94±4,26	61,26	41,49
		M	65,77±9,71	55,29±9,99	55,29±9,100	45,13	53,65
		OB	57,63±6,08	66,76±4,16	66,76±4,17	64,19	39,25
		SR	43,35±11,44	65,59±2,64	65,59±2,65	78,81	27,43
		K5	55,25±6,98	62,35±6,80	62,35±6,81	57,81	44,12
		M	57,24±17,12	51,47±12,56	51,47±12,57	44,02	54,48
		OC	54,44±8,15	66,47±8,84	66,47±8,85	66,54	37,43
		SR	46,88±11,22	64,41±1,58	64,41±1,59	69,35	35,22
		K3	50,93±3,97	65,59±4,75	65,59±4,76	68,03	36,26
		M	50,33±14,20	48,53±12,42	48,53±12,43	42,86	55,38
		OC	61,67±9,77	70,00±5,99	70,00±5,100	67,22	36,90
		SP	53,30±11,70	66,47±3,52	66,47±3,53	67,70	36,51
	EM1	K5	59,25±7,41	67,65±5,42	67,65±5,43	64,56	38,97
		M	76,72±4,99	69,12±8,13	69,12±8,14	54,96	46,27
		OB	52,40±9,36	64,12±5,39	64,12±5,40	63,17	40,03
		SP	62,57±14,59	72,35±10,20	72,35±10,21	71,69	33,35
	EM2	K5	57,92±6,46	63,82±6,84	63,82±6,85	58,54	43,56
		M	58,73±11,63	47,94±10,77	47,94±10,78	40,94	56,81
		OB	47,65±7,09	62,35±4,88	62,35±4,89	63,22	40,01
		SR	39,97±6,22	61,47±5,80	61,47±5,81	67,44	36,72
	ED	K5	51,79±8,87	66,47±5,29	66,47±5,30	69,38	35,19
		M	57,98±7,87	45,88±16,06	45,88±16,07	39,63	57,81
		OB	51,58±9,66	64,71±8,42	64,71±8,43	65,20	38,47
		SR	55,36±6,19	66,76±3,49	66,76±3,50	66,30	38,47
	IS1	K3	63,20±7,06	67,94±4,04	67,94±4,05	61,84	41,05
		M	69,26±10,30	67,94±6,63	67,94±6,64	57,66	44,23
		OSCS	59,84±6,94	67,35±5,94	67,35±5,95	63,45	39,82
		SR	57,82±12,46	70,59±5,26	70,59±5,27	73,05	32,24
	IS2	K5	60,15±10,08	62,35±5,39	62,35±5,40	54,98	46,25
		M	62,91±10,58	53,53±12,04	53,53±12,05	44,40	54,20
		OB	60,95±6,40	71,47±3,73	71,47±3,74	71,43	33,55
		SR	41,94±6,83	66,47±2,99	66,47±2,100	86,55	20,37
	IS3	K5	55,66±4,25	61,76±4,92	61,76±4,93	56,60	45,03
		M	59,11±16,84	55,59±15,58	55,59±15,59	47,36	51,97
		OB	56,22±6,83	67,65±5,88	67,65±5,89	67,47	36,70
		SR	41,89±6,81	65,00±2,05	65,00±2,06	78,96	27,31
	AV1	K5	57,31±7,54	66,18±5,61	66,18±5,62	63,27	39,96
		M	58,44±9,39	50,88±10,27	50,88±10,28	43,26	55,06
		OB	53,56±7,91	60,00±6,19	60,00±6,20	55,01	46,23
		SP	48,54±12,00	66,76±3,95	66,76±3,96	74,54	31,03
	AV2	K3	51,03±10,85	66,47±7,10	66,47±7,11	70,28	34,48
		M	57,52±7,65	50,59±6,28	50,59±6,29	43,22	55,09
		OC	52,26±10,83	59,71±8,32	59,71±8,33	55,23	46,07
		SR	40,60±9,95	64,71±3,48	64,71±3,49	80,13	26,29
	EL	K1	49,09±7,48	65,00±4,64	65,00±4,65	68,49	35,90
		M	56,59±10,68	52,35±10,08	52,35±10,09	44,96	53,77
		OE	50,04±11,42	61,18±7,53	61,18±7,54	59,03	43,20
		SR	45,51±9,06	66,76±3,23	66,76±3,24	79,72	26,65
	EM1	K5	48,29±8,06	63,53±4,77	63,53±4,78	65,48	38,25
		M	65,36±9,62	59,41±8,80	59,41±8,81	49,13	50,64
		OB	55,27±6,79	63,82±4,37	63,82±4,38	60,36	42,18
		SP	41,14±7,92	65,00±4,45	65,00±4,46	80,50	25,97
	EM2	K5	47,85±6,94	64,12±4,88	64,12±4,89	67,44	36,72
		M	59,78±12,27	52,65±10,55	52,65±10,56	44,44	54,17
		OMN	51,76±7,95	62,94±7,91	62,94±7,92	61,25	41,50
		SP	46,12±9,12	64,41±3,82	64,41±3,83	70,26	34,50
	ED	K5	44,43±3,03	58,53±2,44	58,53±2,45	57,20	44,58
		M	63,32±8,36	49,41±5,39	49,41±5,40	41,14	56,66
		OSCS	57,96±9,27	65,29±6,93	65,29±6,94	61,07	41,64
		SR	43,15±9,31	63,82±4,37	63,82±4,38	72,12	33,00
	IS1	K5	45,73±4,16	66,47±4,20	66,47±4,21	78,09	28,05
		M	60,70±9,18	60,59±6,60	60,59±6,61	52,38	48,21
		OE	56,61±4,32	65,59±1,34	65,59±1,35	62,69	40,40
		SL	64,88±6,19	75,88±3,90	75,88±3,91	78,15	28,00
	IS2	K5	45,81±7,22	61,76±5,58	61,76±5,59	63,32	39,92
		M	53,38±11,28	46,47±10,17	46,47±10,18	40,73	56,97
		OB	59,25±9,89	67,65±7,32	67,65±7,33	64,56	38,97
		SR	52,49±10,61	67,06±6,28	67,06±6,29	70,14	34,59
	IS3	K5	47,09±9,14	64,12±5,99	64,12±5,10	68,26	36,08
		M	50,03±13,47	45,88±9,31	45,88±9,32	40,72	56,98
		OB	59,27±12,68	63,53±8,54	63,53±8,55	57,22	44,57
		SP	52,05±9,42	67,65±4,36	67,65±4,37	72,35	32,82

(V) Dublado

Tabela 14 – Passo 2h - Escolha do extrator e classificador com maior acurácia (%),

Modalidade	Extrator	Classificador	Spe (%)	Sen (%)	Acc (%)	AUC (%)	EER (%)
(U) Mudo	IS4	K5	54,23±10,79	64,71±6,96	64,71±6,97	62,87	40,27
		M	61,28±12,24	54,71±11,02	54,71±11,03	45,87	53,09
		OC	50,77±11,12	65,29±8,80	65,29±8,81	67,44	36,72
		SP	48,77±10,89	64,41±4,82	64,41±4,83	67,26	36,86
	IS5	K5	50,21±9,44	59,41±10,50	59,41±10,51	55,82	45,62
		M	47,79±10,28	49,41±9,09	49,41±9,10	44,03	54,48
		OE	58,55±8,24	67,06±5,06	67,06±5,07	63,98	39,41
		SL	61,76±9,64	72,94±6,28	72,94±6,29	74,26	31,26
	IS6	K1	60,08±8,40	65,00±5,94	65,00±5,95	59,06	43,17
		M	57,70±12,55	56,47±13,33	56,47±13,34	48,75	50,93
		OSCS	64,11±10,37	68,24±7,18	68,24±7,19	61,65	41,20
		SR	47,02±11,35	66,76±5,58	66,76±5,59	76,96	29,01
	IS7	K3	57,85±7,74	66,47±7,34	66,47±7,35	63,40	39,85
		M	65,52±12,89	59,71±11,47	59,71±11,48	49,38	50,46
		OC	54,75±9,58	69,12±5,61	69,12±5,62	73,01	32,28
		SP	50,37±9,64	67,35±3,07	67,35±3,08	73,76	31,66
	JA	K1	63,44±12,47	71,18±9,46	71,18±9,47	67,98	36,30
		M	55,29±12,39	52,06±13,73	52,06±13,74	45,01	53,74
		OC	56,29±8,53	65,00±6,63	65,00±6,64	61,80	41,08
		SR	47,20±7,91	65,00±5,49	65,00±5,50	70,73	34,12
	KT	K3	50,88±4,34	64,12±5,54	64,12±5,55	64,49	39,02
		M	64,84±9,70	57,06±10,37	57,06±10,38	46,99	52,25
		OC	57,41±11,95	67,06±8,19	67,06±8,20	65,00	38,63
		SP	64,83±11,35	74,41±6,31	74,41±6,32	74,11	31,38
	MP	K1	61,38±3,30	72,94±3,90	72,94±3,91	74,77	30,84
		M	63,92±4,68	64,41±7,13	64,41±7,14	55,81	45,63
		OE	54,93±7,22	65,29±5,70	65,29±5,71	63,49	39,79
		SR	58,10±9,77	69,71±4,16	69,71±4,17	70,39	34,39
	TS	K5	52,92±1,06	66,47±1,44	66,47±1,45	68,11	36,19
		M	66,02±6,64	66,18±6,34	66,18±6,35	57,09	44,66
		OMN	56,67±3,72	65,00±2,44	65,00±2,45	61,51	41,31
		SR	65,49±6,51	73,53±5,58	73,53±5,59	71,11	33,81
	YA	K3	54,39±7,35	65,00±4,04	65,00±4,05	63,34	39,90
		M	64,47±13,00	55,00±11,00	55,00±11,01	45,23	53,57
		OCS	46,77±5,43	63,53±4,59	63,53±4,60	66,97	37,09
		SR	66,72±11,12	74,41±6,31	74,41±6,32	71,77	33,28
(V) Dublado	IS4	K5	53,52±8,03	66,18±6,60	66,18±6,61	66,77	37,25
		M	53,22±8,76	46,18±5,10	46,18±5,11	40,53	57,12
		OMN	57,70±5,28	64,12±4,70	64,12±4,71	59,18	43,08
		SR	55,82±13,20	69,71±7,21	69,71±7,22	73,24	32,09
	AV2	K3	53,10±10,18	64,71±5,73	64,71±5,74	63,83	39,52
		M	52,00±12,10	50,88±11,24	50,88±11,25	44,65	54,01
		OMN	57,48±7,10	64,41±4,64	64,41±4,65	59,84	42,58
		SP	48,01±12,70	64,41±3,82	64,41±3,83	68,07	36,23
	IS6	K5	55,90±5,95	67,06±2,88	67,06±2,89	66,44	37,50
		M	55,71±12,34	53,53±9,18	53,53±9,19	46,30	52,76
		OC	51,90±11,78	65,29±9,09	65,29±9,10	66,28	37,63
		SP	49,51±10,12	66,47±6,05	66,47±6,06	72,21	32,93
	IS7	K3	52,24±9,04	63,82±4,93	63,82±4,94	62,68	40,41
		M	58,22±12,99	51,18±12,49	51,18±12,50	43,55	54,84
		OB	48,55±8,56	64,71±7,89	64,71±7,90	68,28	36,06
		SR	43,19±7,11	65,29±2,20	65,29±2,21	77,83	28,27
	JA	K5	48,55±8,75	64,71±7,44	64,71±7,45	68,28	36,06
		M	43,82±9,45	43,53±10,00	43,53±10,01	39,37	58,00
		OCS	36,79±8,03	59,12±5,80	59,12±5,81	61,21	41,54
		SR	46,34±11,91	64,12±3,90	64,12±3,91	69,11	34,30
	KT	K5	45,96±7,29	64,12±5,70	64,12±5,71	69,55	35,06
		M	56,61±13,96	48,24±11,17	48,24±11,18	41,53	56,36
		OC	51,89±12,84	59,71±9,21	59,71±9,22	55,42	45,92
		SR	41,69±7,18	63,24±1,97	63,24±1,98	71,88	33,19
	MP	K5	44,26±4,25	58,24±4,88	58,24±4,89	56,72	44,94
		M	51,98±11,03	52,94±6,57	52,94±6,58	46,70	52,47
		OC	55,44±6,08	62,06±4,64	62,06±4,65	57,20	44,58
		SP	49,85±11,50	65,00±5,49	65,00±5,50	67,66	36,55
	TS	K5	44,71±5,55	65,29±2,20	65,29±2,21	75,24	30,45
		M	61,17±9,52	63,53±6,19	63,53±6,20	56,10	45,41
		OSC	41,93±2,93	58,82±4,15	58,82±4,16	59,01	43,21
		SL	51,29±10,72	67,65±6,02	67,65±6,03	73,38	31,98
	YA	K5	55,20±6,86	66,47±5,29	66,47±5,30	65,79	38,01
		M	59,49±8,92	55,59±7,60	55,59±7,61	47,25	52,06
		OCS	53,96±9,08	65,59±7,78	65,59±7,79	64,99	38,63
		SR	62,73±10,96	72,65±5,26	72,65±5,27	72,23	32,91

Tabela 15 – Passo 3 - Resultado final e comparação com outra pesquisa,

Modalidade	Áudio	Extrator	Classif,	Nesta abordagem					E, Vaiciukynas et, al,		
				Acc (%)	AUC (%)	EER (%)	Tempo treino (s)	Tempo teste (s)	Extrator	Classif,	EER (%)
AC	P	YA	K1	94,55±3,33	87,84	19,01	0,002±0,005	0,034±0,0339	ED		20,78
	S	IS4	SR	72,65±4,76	84,21	22,61	2,900±0,035	0,008±0,007	ED	RF	20,30
	U	IS6	SR	76,47±5,89	78,83	27,42	4,557±0,292	0,008±0,007	ED		24,52
	V	EM1	SR	68,53±6,18	78,57	27,64	0,656±0,012	0,001±0,000	IS7		24,89
SP	P	KT	K1	92,94±2,55	92,40	14,15	0,002±0,004	0,025±0,024	TS		29,02
	S	KT	SP	71,18±4,52	87,30	19,63	7,720±0,330	0,0001±0,0000	JA	RF	26,12
	U	MP	K1	72,94±3,91	74,77	30,84	0,001±0,000	0,003±0,003	IS4		28,36
	V	EM1	SP	65,00±4,46	80,50	25,97	6,094±0,151	0,001±0,000	YA		25,57

Tabela 16 – Matriz de confusão de resultados de classificação usando métodos de melhor desempenho: (a) AC obtido com P/YA e classificador K1; (b) SP obtido com P/KT classificador K1.

AC Performance					SP Performance				
		Classe Predita					Classe Predita		
		HC	PD				HC	PD	
Classe atual	HC	33.80	2.20	Actual class	HC	30.50	5.50		
	PD	3.30	61.70		PD	1.70	64.30		

Notes: PD - Paciente com doença de Parkinson; HC - Indivíduo saudável.

5 Conclusões, Contribuições e Trabalhos Futuros

Apresentamos uma abordagem viável para a detecção da doença de Parkinson (DP) usando técnicas de extração de características e aprendizagem de máquina. Aplicamos dezoito extratores de características, quatro classificadores no conjunto de dados de voz recentemente aquisitado e avaliamos os resultados. Os resultados mostraram que não há diferença significativa entre os microfones acústico cardióide (AC) e do smartphone (SP), apoiando assim a ideia de que a DP pode ser detectada processando no smartphone. O áudio de Fonação (P) forneceu resultados significativamente melhores do que os de Fala (S), Dublado (V) e Mudo (U).

A melhor combinação extrator e classificador foi o YAAFE (YA) definido com o classificador de k-vizinho mais próximo (K1) ao usar a modalidade de fonação (P) do microfone de áudio cardióide (AC). A segunda melhor combinação foi o extrator KTU (KT) com K1 ao usar a fonação (P) modalidade do microfone smartphone (SP). As acurácias obtidas foram de 94,55% e 92,94%, AUC de 87,84% e 92,40%, EER de 19,01% e 14,15%, respectivamente. Nossa abordagem obteve resultados diferentes da abordagem de Vaiciukynas (VAICIUKYNAS et al., 2017) e provou ser relativamente melhor. O desempenho entre as modalidades AC e SP mostrou-se coerente.

Nossa principal contribuição é o uso de outros classificadores para a detecção de DP e a investigação de uma abordagem simplificada, sem o uso da fusão de áudio, técnica esta empregada por (VAICIUKYNAS et al., 2017). Com base nos resultados apresentados, podemos afirmar que a tarefa de fonação (P) é a mais adequada para detectar a doença. Além de ser aquela que permitiu obter o melhor desempenho nas modalidades de microfones AC e SP. Demonstramos que é possível desenvolver aplicativos móveis para detectar PD usando smartphones sem muita perda de desempenho.

Em trabalhos futuros, aplicaremos a validação cruzada para resolver o problema do conjunto de dados desbalanceado. Outro trabalho futuro incluirá testes e otimização de extratores de características por meio da poda de atributos que contribuem pouco para a detecção de PD. Investigar a detecção de PD aplicando a transformação dimensional usando a técnica *t-distributed Stochastic Neighbor Embedding* (t-SNE) combinada com métodos de aprendizado de máquina. Poderá também ser avaliado o uso de espectrograma. Além disso, proporemos uma nova ferramenta de extração de características para detecção de PD.

6 Declaração ética

O protocolo do estudo relacionado a esse conjunto de dados foi aprovado pelo Comitê Regional de Bioética da Universidade de Kaunas (P2-24 / 2013). Consentimentos informados por escrito foram obtidos dos participantes do estudo, os identificadores dos pacientes foram removidos para garantir o anonimato.

Referências

BARBAGALLO, G. et al. The placebo effect on resting tremor in parkinson's disease: an electrophysiological study. *Parkinsonism & Related Disorders*, v. 52, p. 17 – 23, 2018. ISSN 1353-8020. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1353802018301275>>. Citado na página 24.

BERNTHAL, J. E.; BANKSON, N. W.; FLIPSEN, P. *Articulation and phonological disorders: Speech sound disorders in children*. [S.l.]: Pearson Boston, MA, 2009. Citado 2 vezes nas páginas 31 e 32.

BOGDANOV, D. et al. Essentia: an audio analysis library for music information retrieval. In: *International Society for Music Information Retrieval Conference (ISMIR'13)*. Curitiba, Brazil: [s.n.], 2013. p. 493–498. Disponível em: <<http://hdl.handle.net/10230/32252>>. Citado 2 vezes nas páginas 36 e 39.

CERNAK, M. et al. Characterisation of voice quality of parkinson's disease using differential phonological posterior features. *Computer Speech & Language*, Elsevier BV, v. 46, p. 196–208, nov 2017. Citado na página 26.

CRYSANDT, H.; TUMMARELLO, G.; PIAZZA, F. Mpeg-7 encoding and processing: Mpeg7audioenc+ mpeg7audiodb. In: *3rd MUSICNETWORK Open Workshop: MPEG AHG on Music Notation Requirements. Munich, Germany*. [S.l.: s.n.], 2004. p. 13–14. Citado 2 vezes nas páginas 37 e 39.

DIAS, A. E.; LIMONGI, J. C. P. Tratamento dos distúrbios da voz na doença de parkinson, o método lee silverman. *Arq Neuropsiquiatr*, SciELO Brasil, v. 61, n. 1, p. 61–6, 2003. Citado na página 33.

DUFFY, J. R. *Motor Speech disorders-E-Book: Substrates, differential diagnosis, and management*. [S.l.]: Elsevier Health Sciences, 2013. Citado 5 vezes nas páginas 32, 33, 34, 35 e 50.

EYBEN, F. et al. Recent developments in opensmile, the munich open-source multimedia feature extractor. In: *Proceedings of the 21st ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2013. (MM '13), p. 835–838. ISBN 978-1-4503-2404-5. Citado 2 vezes nas páginas 36 e 39.

FEARNLEY, J. M.; LEES, A. J. Ageing and parkinson's disease: substantia nigra regional selectivity. *Brain*, v. 114, n. 5, p. 2283–2301, 1991. Citado na página 25.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithms for computing k-nearest neighbors. *IEEE Trans. Computers*, v. 24, n. 7, p. 750–753, 1975. Citado na página 43.

GARCIA-MONTES, J.-R.; BORONAT-GARCIA, A.; DRUCKER-COLÍN, R. Pharmacological strategies for parkinson's disease. *Health*, v. 4, n. 11A, p. 1153–1166, 2012. Citado na página 24.

GARCIA-PERERA, L. P. et al. Optimization of the det curve in speaker verification. In: *2012 IEEE Spoken Language Technology Workshop (SLT)*. [S.l.: s.n.], 2012. p. 318–323. Citado na página 47.

GELZINIS, A.; VERIKAS, A.; BACAUSKIENE, M. Automated speech analysis applied to laryngeal disease categorization. *Computer Methods and Programs in Biomedicine*, v. 91, n. 1, p. 36 – 47, 2008. ISSN 0169-2607. Citado 2 vezes nas páginas 37 e 39.

HAREL, B.; CANNIZZARO, M.; SNYDER, P. J. Variability in fundamental frequency during speech in prodromal and incipient parkinsons disease: A longitudinal case study. *Brain and Cognition*, Elsevier BV, v. 56, n. 1, p. 24–29, oct 2004. Citado na página 25.

HAYKIN, S. O. *Neural Networks and Learning Machines*. [S.l.]: Pearson Prentice Hall, 2008. Citado 2 vezes nas páginas 40 e 41.

HAYKIN, S. S. *Redes Neurais: Princípios e Práticas*. Second. [S.l.]: Bookman Companhia ED., 2002. Citado na página 39.

HAYKIN, S. S. *Neural networks and learning machines*. Third. [S.l.]: Pearson Education, 2009. Citado na página 43.

HO, A. K. et al. Speech impairment in a large sample of patients with parkinson's disease. *Behavioural Neurology*, Hindawi Limited, v. 11, n. 3, p. 131–137, 1999. Citado na página 25.

HO, T. K. *Nearest neighbors in random subspaces*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. 640-648 p. Citado na página 40.

KE, Q. et al. Big data analytics enabled by feature extraction based on partial independence. *Neurocomputing*, Elsevier BV, v. 288, p. 3–10, may 2018. Citado na página 26.

LAURAITIS, A.; MASKELIŪNAS, R.; DAMAŠEVIČIUS, R. ANN and fuzzy logic based model to evaluate huntington disease symptoms. *Journal of Healthcare Engineering*, Hindawi Limited, v. 2018, p. 1–10, 2018. Citado na página 25.

LITTLE, M. et al. Suitability of dysphonia measurements for telemonitoring of parkinsons disease. *IEEE Transactions on Biomedical Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 56, n. 4, p. 1015–1022, apr 2009. Citado na página 26.

LOGEMANN, J. A. et al. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. *Journal of Speech and Hearing Disorders*, American Speech Language Hearing Association, v. 43, n. 1, p. 47, feb 1978. Citado na página 25.

LUDERMIR, A. C. P. de C. A. P. B. T. B. *Redes Neurais Artificiais: Teoria e aplicações*. First. [S.l.]: LTC Editora, 2007. Citado na página 39.

MATHIEU, B. et al. Yaafe, an easy to use and efficient audio feature extraction software. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference*. Utrecht, The Netherlands: [s.n.], 2010. p. 441–446. <<http://ismir2010.ismir.net/proceedings/ismir2010-75.pdf>>. Citado 2 vezes nas páginas 38 e 39.

- MCCULLOCH, W. S.; PITTS, W. *Neurocomputing: Foundations of research*. First. [S.l.]: MIT Press, 1988. Citado na página 39.
- MCENNIS, D.; MCKAY, C.; FUJINAGA, I. jaudio: Additions and improvements. In: *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*. ISMIR. Victoria British Columbia, Canada: University of Victoria.: [s.n.], 2006. p. 385–386. Disponível em: <<http://github.com/dmcennis/jAudioGIT>> Citado 2 vezes nas páginas 37 e 39.
- OROZCO-ARROYAVE, J. R. et al. NeuroSpeech: An open-source software for parkinsons speech analysis. *Digital Signal Processing*, Elsevier BV, v. 77, p. 207–221, jun 2018. Citado na página 26.
- PAPA, J. P.; aO, A. X. F.; SUZUKI, C. T. N. Supervised pattern classification based on optimum-path forest. *Int. J. Imaging Syst. Technol.*, John Wiley & Sons, Inc., New York, NY, USA, v. 19, n. 2, p. 120–131, jun. 2009. ISSN 0899-9457. Citado na página 41.
- PAPA, J. P. et al. Designed of robust pattern classifiers based on optimum-path forest. In: *Mathematical Morphology and its Applications to Signal and Image Processing (ISMM)*. [S.l.]: MCT/INPE, 2007. p. 337–348. Citado na página 43.
- PARKINSON, J. An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*, v. 14, n. 2, p. 223–236, 1817. PMID: 11983801. Citado na página 23.
- PEREIRA, C. R. et al. A new computer vision-based approach to aid the diagnosis of parkinson’s disease. *Computer Methods and Programs in Biomedicine*, v. 136, p. 79 – 88, 2016. ISSN 0169-2607. Citado na página 24.
- PEREIRA, J. C. et al. Residual signal auto-correlation to evaluate speech in parkinson’s disease patients. *Arquivos de Neuro-Psiquiatria*, FapUNIFESP (SciELO), v. 64, n. 4, p. 912–915, dec 2006. Citado na página 26.
- RUSZ, J. et al. Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson’s disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 26, n. 8, p. 1495–1507, aug 2018. Citado na página 26.
- SAKAR, B. E. et al. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, Institute of Electrical and Electronics Engineers (IEEE), v. 17, n. 4, p. 828–834, jul 2013. Citado na página 26.
- SCHULLER, B. et al. Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, IEEE, v. 1, n. 2, p. 119–131, 2010. Citado na página 36.
- SHULMAN, L. et al. Non-recognition of depression and other non-motor symptoms in parkinson’s disease. *Parkinsonism & Related Disorders*, v. 8, n. 3, p. 193 – 197, 2002. ISSN 1353-8020. Citado na página 24.
- SOUZA, J. W. M. de et al. A new approach to diagnose parkinson’s disease using a structural cooccurrence matrix for a similarity analysis. *Computational Intelligence and Neuroscience*, v. 18, p. 8, 2018. Citado na página 24.

TSANAS, A. et al. Novel speech signal processing algorithms for high-accuracy classification of parkinsons disease. *IEEE Transactions on Biomedical Engineering*, v. 59, n. 5, p. 1264–1271, 2012. Citado na página 26.

TSANAS, A. et al. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinsons disease symptom severity. *Journal of The Royal Society Interface*, The Royal Society, v. 8, n. 59, p. 842–855, nov 2010. Citado 2 vezes nas páginas 38 e 39.

UPADHYA, S. S.; CHEERAN, A. N. PERFORMANCE COMPARISON OF REGRESSION TECHNIQUES IN PREDICTING PARKINSON DISEASE SEVERITY SCORE USING SPEECH FEATURES. *Biomedical Engineering: Applications, Basis and Communications*, National Taiwan University, v. 30, n. 04, p. 1850025, aug 2018. Citado na página 26.

VAICIUKYNAS, E. et al. Detecting parkinson's disease from sustained phonation and speech signals. *PLOS ONE*, Public Library of Science, v. 12, n. 10, p. 1–16, 10 2017. Disponível em: <10.1371/journal.pone.0185613>. Citado 5 vezes nas páginas 27, 45, 46, 49 e 65.

VAPNIK, V. N. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995. Citado na página 42.

VAPNIK, V. N. *Statistical learning theory*. [S.l.]: John Wiley & Sons, Inc., 1998. Citado na página 43.

VAPNIK, V. N.; CHERVONENKIS, A. Y. *On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities*. Cham: Springer International Publishing, 2015. 11-30 p. Citado 2 vezes nas páginas 15 e 42.

VERIKAS, A. et al. Data dependent random forest applied to screening for laryngeal disorders through analysis of sustained phonation: Acoustic versus contact microphone. *Medical Engineering & Physics*, v. 37, n. 2, p. 210 – 218, 2015. ISSN 1350-4533. Citado na página 26.

WHO. *Neurological Disorders: Public Health Challenges*. [S.l.]: World Health Organization, 2006. Citado na página 25.

WOŹNIAK, M.; POŁAP, D. Hybrid neuro-heuristic methodology for simulation and control of dynamic systems over time interval. *Neural Networks*, Elsevier BV, v. 93, p. 45–56, sep 2017. Citado na página 26.

WOŹNIAK, M.; POŁAP, D. Adaptive neuro-heuristic hybrid model for fruit peel defects detection. *Neural Networks*, Elsevier BV, v. 98, p. 16–33, feb 2018. Citado na página 26.

YADAV, H. P.; LI, Y. The development of treatment for parkinson's disease. *Advances in Parkinson's Disease*, v. 4, n. 3, p. 59–78, 2015. Citado na página 25.

ZHAO, S. et al. Automatic detection of expressed emotion in parkinson's disease. *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 4813–4817, 2014. Citado na página 26.

ZWEIG, M. H.; CAMPBELL, G. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, Clinical Chemistry, v. 39, n. 4, p. 561–577, 1993. Citado na página 48.