UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

CAROLINE WANDERLEY ESPINOLA

ANÁLISE COMPUTACIONAL DA VOZ COMO UMA FERRAMENTA DE AUXÍLIO DIAGNÓSTICO DE TRANSTORNOS MENTAIS

Recife
2021

CAROLINE WANDERLEY ESPINOLA

ANÁLISE COMPUTACIONAL DA VOZ COMO UMA FERRAMENTA DE AUXÍLIO DIAGNÓSTICO DE TRANSTORNOS MENTAIS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Engenharia Biomédica.

Área de concentração: Computação Biomédica.

Orientador: Wellington Pinheiro dos Santos

Recife

2021

Catalogação na fonte
Bibliotecário Gabriel Luz, CRB-4 / 2222

E77a Espinola, Caroline Wanderley.
Análise computacional da voz como uma ferramenta de auxílio diagnóstico de transtornos mentais / Caroline Wanderley Espinola – Recife, 2021.
146 f.: figs., quads., tabs., abrev. e siglas.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Biomédica, 2021.
Inclui referências e anexos.

1. Engenharia Biomédica. 2. Transtornos mentais. 3. Diagnóstico. 4. Voz. 5. Parâmetros acústicos. 6. Aprendizado de máquina. I. Santos, Wellington Pinheiro dos (Orientador). II. Título.

UFPE

610.28 CDD (22. ed.) BCTG / 2021 - 113

CAROLINE WANDERLEY ESPINOLA

ANÁLISE COMPUTACIONAL DA VOZ COMO UMA FERRAMENTA DE AUXÍLIO DIAGNÓSTICO DE TRANSTORNOS MENTAIS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Engenharia Biomédica.

Aprovada em: 04/05/2021

BANCA EXAMINADORA

Prof. Dr. Wellington Pinheiro dos Santos (Orientador)

Universidade Federal de Pernambuco

Prof. Dr. Ricardo Emmanuel de Souza (Examinador Interno)

Universidade Federal de Pernambuco

Prof. Dr. Antonio Medeiros Peregrino da Silva (Examinador Externo)

Universidade de Pernambuco

Dedico este trabalho ao meu esposo Antonio Jr., meu eterno companheiro na jornada da vida. Dedico também à minha avó Cacilda (in memoriam), que foi um exemplo de acolhimento e generosidade para todos ao seu redor e uma grande patrocinadora das minhas atividades acadêmicas.

AGRADECIMENTOS

Primeiramente, a Deus, pela vida e pela capacidade física e mental para finalizar este trabalho em um momento tão difícil para a humanidade.

Ao meu amor, Antonio Jr., que me encorajou a navegar por águas desconhecidas para além da Medicina. Por todo seu acolhimento, compaixão e imenso amor nos momentos difíceis deste projeto. Por toda sua alegria e vibração a cada pequena conquista minha. Por sua incrível habilidade de adicionar cores às paletas minha da vida.

Aos meus amados pais, Alba e Gilberto, por seu apoio irrestrito a este e a todos os meus projetos. Que apesar dos inúmeros obstáculos, fomentaram e patrocinaram os meus estudos em Recife. Por todo o amor recebido por meios de cuidados e por todos seus ensinamentos, sabedoria e valores éticos que me guiaram pelas vidas pessoal e profissional na Medicina.

Ao meu orientador, Prof. Wellington, que aos poucos se tornou uma referência pessoal e acadêmica para mim. Por ter acreditado na minha ideia e me dado uma oportunidade de estudar um pouco desta área tão fascinante, e também por sua disponibilidade, sua dedicação à docência e por compartilhar seus conhecimentos de maneira tão generosa. Esta experiência com o senhor foi fundamental para meu amadurecimento na carreira acadêmica.

À minha tia e segunda mãe Ana, exemplo de mulher corajosa e determinada, que nunca mediu esforços para me apoiar em todos os meus projetos e aventuras, acadêmicos ou não. Que apesar de fisicamente longe, consegue sempre fazer-se presente nos momentos importantes da minha vida com seus conselhos e seu imenso afeto materno.

À minha querida sogra D. Sônia, que assim como para uma filha, sempre me dedicou todo o carinho e cuidado. Agradeço por ter me incentivado e acolhido no difícil momento do início deste projeto.

Aos meus filhos de pelo Amy e Fusquinha, por seu amor incondicional e por me ensinar a cada dia a tentar ser uma pessoa melhor.

Aos familiares, pelo apoio e torcida, em particular meus primos Leonardo e Artur, que gentilmente compartilharam suas experiências tão prolíficas na área acadêmica. Agradeço a Léo por seus ensinamentos na área de Fonoaudiologia e por fornecer materiais essenciais à minha pesquisa, e a Artur por dividir conhecimentos em pesquisa e em ciência de dados e por vibrar com minhas conquistas.

Aos amigos e colegas de mestrado, especialmente minha amiga Ainoã, a quem tive a sorte de conhecer e de me acompanhar no desenvolvimento deste projeto, e que me apoiou de diversas maneiras para o sucesso deste trabalho.

Aos colegas do Grupo de Pesquisas em Computação Biomédica do LCB, em especial Juliana Gomes, que me ajudou em diversos momentos, desde a elaboração de códigos à escrita de artigos científicos e gentilmente compartilhou comigo conhecimentos da Engenharia Biomédica.

Aos colegas do HUP e do HC-UFPE, sem os quais a realização deste trabalho não seria possível. Agradeço particularmente ao Dr. Ezron, pelo acolhimento no Ambulatório de Psiquiatria (HC), e aos psiquiatras e residentes de psiquiatria que generosamente me ajudaram na árdua etapa da coleta de dados. Também gostaria de agradecer às minhas colegas de plantão Elisa e Luciana, por toda compreensão e torcida pelo sucesso do meu projeto.

Aos membros da banca, os professores Ricardo Emmanuel e Antonio Peregrino, por terem aceitado o convite e por sua contribuição para o engrandecimento deste trabalho. Agradeço especialmente ao Prof. Antonio, que desde a faculdade se tornou para mim uma grande referência de professor, psiquiatra e ser humano e a quem nutro profundo carinho e admiração.

Por último, aos voluntários desta pesquisa, principalmente os pacientes dos serviços de Psiquiatria do HC e do HUP, os quais foram a motivação para a idealização deste projeto e sem os quais este não seria possível. A eles minha sincera gratidão.

Conheça todas as teorias, domine todas as técnicas, mas ao tocar uma alma humana, seja apenas outra alma humana.

Carl Jung

RESUMO

A psiquiatria é uma especialidade médica que ainda carece de marcadores e exames objetivos em sua rotina, levando a uma grande necessidade pelo desenvolvimento de tais parâmetros nessa área. Nesse contexto, diversos estudos têm demonstrado a existência de mudanças nas propriedades acústicas da voz em vários transtornos mentais, como depressão maior, transtorno bipolar e esquizofrenia, sugerindo que tais alterações possam ser indicadores da presença de determinado transtorno. O presente trabalho avaliou o uso de parâmetros vocais como biomarcadores que podem auxiliar o diagnóstico psiquiátrico. Foram utilizados métodos computacionais e de aprendizado de máquina para a extração de parâmetros acústicos e para a construção de ferramentas automatizadas de apoio diagnóstico de quatro transtornos mentais: depressão maior, esquizofrenia, transtorno bipolar e transtorno de ansiedade generalizada. Para tanto, foi construída uma base de dados própria em ambientes naturalísticos, a qual foi utilizada nos experimentos de testes e validação. Foram realizados dois conjuntos de experimentos computacionais independentes de classificação com algoritmos supervisionados de aprendizado de máquina, o primeiro com o balanceamento padrão do software Weka (*ClassBalancer*) e o segundo com o método SMOTE. O *framework* desenvolvido neste trabalho forneceu acurácias classificatórias gerais de 79,23% com o modelo SVM para o primeiro conjunto de experimentos, e de 81,45% com *Random Forest* para o segundo. Esses resultados reforçam a robustez do emprego de atributos acústicos para a detecção de transtornos mentais com base em modelos de aprendizado de máquina.

Palavras-chave: Transtornos mentais. Diagnóstico. Voz. Parâmetros acústicos. Aprendizado de máquina.

ABSTRACT

Psychiatry is a medical specialty that still lacks the use of objective markers and exams in its routine, which leads to a great need for the development of such parameters in this area. In this context, several studies have demonstrated changes in vocal acoustic properties in various mental disorders, such as major depression, bipolar disorder and schizophrenia, which ultimately suggests that such alterations might be indicators of a certain disorder. The current work assessed the use of vocal features as biomarkers that may assist psychiatric diagnosis. Computational and machine learning techniques were applied to vocal feature extraction and to the development of automatic tools of auxiliary diagnosis of four mental disorders: major depressive disorder, schizophrenia, bipolar disorder and generalized anxiety disorder. To this aim, a new database was created in naturalistic environments and was further applied to test and validation experiments. Two independent groups of classification experiments were conducted, the first one using Weka software's standard balancing method (*ClassBalancer*), and the second one using SMOTE technique. The framework developed in this work provided overall classification accuracies of 79.23% with SVM model for the former experiments and 81.45% with Random Forest for the latter. These findings underscore the strength of the use of acoustic features for the detection of mental disorders based on machine learning models.

Keywords: Mental disorders. Diagnosis. Voice. Acoustic features. Machine learning.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

A/D	Analógico-digital
AUC	*Area Under Curve*
BPRS	*Brief Psychiatric Rating Scale*
C	Parâmetro de complexidade
DFT	*Discrete Fourier Transform*
DSM-5	Manual Diagnóstico e Estatístico dos Transtornos Mentais (5ª Edição)
DWT	*Discrete Wavelet Transform*
EAM	Escala de Avaliação de Mania
EEG	Eletroencefalografia
F0	Frequência fundamental
F1	Primeiro formante
F2	Segundo formante
F3	Terceiro formante
FFT	*Fast Fourier Transform*
GAD-7	*Generalized Anxiety Disorder Scale (7-item)*
GMM	*Gaussian Mixture Models*
GPS	*Global Positioning System*
HAM-D	Escala de Depressão de Hamilton
HC	Hospital das Clínicas
HNR	*Harmonics-to-Noise Ratio*
HUP	Hospital Ulysses Pernambucano
IA	Inteligência Artificial
IC	Intervalo de Confiança
LPC	*Linear Predictive Coding*
MFCC	*Mel Frequency Cepstral Coefficient*
ML	*Machine Learning*
MLP	*Multilayer Perceptron*
NSA-16	*Negative Symptom Assessment (16-item)*
OR	*Odds Ratio*
PCM	*Pulse Code Modulation*
PSD	*Power Spectral Density*
PSO	*Particle Swarm Optimization*

PUK	*Pearson Universal Kernel*
QV	Qualidade Vocal
RBF	*Radial Basis Function*
RF	*Random Forest*
RMS	*Root Mean Square*
RNA	Redes Neurais Artificiais
ROC	*Receiver Operating Characteristic*
SLT	*Statistical Learning Theory*
SMOTE	*Synthetic Minority Oversampling Technique*
SNA	Sistema Nervoso Autônomo
SNS	Sistema Nervoso Simpático
SRM	*Structural Risk Minimization*
SRQ-20	*Self-Reporting Questionnaire (20-item)*
SVM	*Support Vector Machine*
TAG	Transtorno de Ansiedade Generalizada
TALE	Termo de Assentimento Livre e Esclarecido
TAS	Transtorno de Ansiedade Social
TB	Transtorno Bipolar
TCLE	Termo de Consentimento Livre e Esclarecido
TDM	Transtorno Depressivo Maior
TEO	*Teager Energy Operator*
TEPT	Transtorno de Estresse Pós-Traumático
UFPE	Universidade Federal de Pernambuco
VC	Validação Cruzada
YLDs	*Years Lived with Disability*
YMRS	*Young Mania Rating Scale*
ZCR	*Zero Crossing Rate*

SUMÁRIO

1 INTRODUÇÃO

Este capítulo versa sobre as principais motivações para o desenvolvimento deste trabalho, apresentando os problemas do diagnóstico dos transtornos mentais na área da psiquiatria e a exploração do potencial da aprendizagem de máquina como uma ferramenta automatizada de auxílio diagnóstico, tendo a voz como um potencial biomarcador de transtornos mentais.

1.1 MOTIVAÇÃO E JUSTIFICATIVA

Até os dias atuais, a avaliação, o diagnóstico e a decisão terapêutica na psiquiatria são pautados no relato do subjetivo do paciente e no julgamento clínico do especialista, tornando esses processos sujeitos a vieses de memória e de subjetividade, tanto do paciente quanto do profissional (MUNDT *et al.*, 2007; FAURHOLT-JEPSEN *et al.*, 2016; JIANG *et al.*, 2018). Enquanto diversas especialidades médicas dispõem de marcadores diagnósticos objetivos de doença e de sua gravidade, como testes bioquímicos na cardiologia (VITTORINI; CLERICO, 2008) e marcadores de volume tumoral na oncologia (SCHALPER *et al.*, 2015), a psiquiatria ainda sofre com a falta de métodos objetivos na sua rotina (BEDI *et al.*, 2015). Um segundo desafio encontrado na prática clínica psiquiátrica é a falta de acesso aos serviços de saúde mental, sendo o número de profissionais insuficiente para atender às demandas da população, principalmente em países subdesenvolvidos (HIRSCHTRITT; INSEL, 2018; TOROUS; CERRATO; HALAMKA, 2019). Por último, o estigma associado à doença mental, devido ao medo do rótulo social de “doente mental”, e a demora dos pacientes em buscar assistência contribuem para a subutilização dos serviços de saúde mental e prejudicam a aderência ao tratamento, levando ao aumento da morbimortalidade desses transtornos (FARLEY-TOOMBS, 2012; HIRSCHTRITT; INSEL, 2018).

Uma estratégia promissora para tentar solucionar esses problemas é a psiquiatria computacional. Utilizando técnicas de aprendizado de máquina ou *machine learning* (ML), essa área realiza uma combinação de métodos computacionais com técnicas estatísticas, de forma a avançar-se a compreensão, a definição prognóstica e o tratamento dos transtornos mentais (HUYS; MAIA; FRANK, 2016; BZDOK; MEYER-LINDENBERG, 2018). O aprendizado de máquina é uma subárea da inteligência artificial caracterizada por uma gama de algoritmos capazes de detectar padrões e gerar predições confiáveis em sistemas, como as diferentes regiões cerebrais (DWYER; FALKAI; KOUTSOULERIS, 2018). Essa tecnologia já se

mostrou útil na elaboração de ferramentas de triagem ou de apoio diagnóstico de doenças de diferentes áreas médicas, como esclerose múltipla (COMMOWICK *et al.*, 2018), doença de Alzheimer (DOS SANTOS *et al.*, 2009; QU; YUAN; LIU, 2009; BHAGYA SHREE; SHESHADRI, 2014) e câncer de mama (HAZRA; MANDAL; GUPTA, 2016; DE SANTANA *et al.*, 2018). Além disso, a capacidade de generalização dessas técnicas para problemas no nível individual tem o potencial de oferecer aplicações clínicas no futuro, com o desenvolvimento de modelos preditivos para a elaboração de tratamentos personalizados, em vez da tradicional abordagem psicofarmacológica baseada em tentativa e erro (PETZSCHNER *et al.*, 2017; BZDOK; MEYER-LINDENBERG, 2018).

As necessidades apontadas na prática clínica psiquiátrica levam à busca de novos marcadores objetivos. Nesse contexto, dada a natureza psicofisiológica da produção vocal, a voz se destaca por ser um dos atributos não verbais mais importantes para informar sobre o estado afetivo e as funções cognitivas e psicomotoras de um pessoa (LAUKKA *et al.*, 2008; VAN PUYVELDE *et al.*, 2018). Uma vez que o sinal acústico da fala costuma ser parametrizado em atributos objetivos para sua investigação (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010), métodos computacionais tornaram-se ideais para a análise da fala na saúde mental, tanto dos seus aspectos linguísticos (MAAS *et al.*, 2011), quanto paralinguísticos (LARSEN *et al.*, 2015; BONE *et al.*, 2016), sendo este o foco deste trabalho. Mais especificamente, essas técnicas têm demonstrado resultados bastante promissores para a detecção e monitoramento de diversos transtornos, como depressão maior, esquizofrenia, autismo, transtorno bipolar, transtornos de ansiedade e até de comportamentos suicidas (SHARDA *et al.*, 2010; LARSEN *et al.*, 2015; LIU *et al.*, 2015; FAURHOLT-JEPSEN *et al.*, 2016; WEEKS *et al.*, 2016; TAHIR *et al.*, 2019). Por último, o emprego da análise computacional da voz na psiquiatria, associada ou não ao uso de outras ferramentas (e.g., tecnologias digitais móveis), pode ser uma solução para aumentar o acesso aos serviços de saúde mental (TOROUS; CERRATO; HALAMKA, 2019), engajar os pacientes em seus tratamentos e, ainda, diminuir consideravelmente o estigma social relacionado à doença mental (RAUSEO-RICUPERO; TOROUS, 2021).

1.2 OBJETIVOS

Diante do acima exposto, o presente trabalho propõe desenvolver uma ferramenta de automação, baseada em aprendizado de máquina, para auxiliar os profissionais no diagnóstico

de quatro transtornos mentais (transtorno depressivo maior, esquizofrenia, transtorno bipolar e transtorno de ansiedade generalizada), por meio do uso da análise das características da voz.

Como objetivos específicos, têm-se:

a) Revisar os conceitos relacionados aos transtornos mentais abordados neste trabalho (transtorno depressivo maior, esquizofrenia, transtorno bipolar e transtorno de ansiedade generalizada), bem como de padrões de alterações vocais nesses transtornos;
b) Revisar os conceitos de aprendizado de máquina, com foco em algoritmos supervisionados para classificação de padrões, os quais têm o objetivo de realizar predições sobre um determinado domínio a partir de informações prévias (SINGH; THAKUR; SHARMA, 2016; OSISANWO *et al.*, 2017);
c) Revisar os conceitos referentes ao processamento digital de sinais e extração de atributos acústicos;
d) Desenvolver um sistema de extração de atributos acústicos da fala para a representação paramétrica da voz de pacientes reais portadores de um dos transtornos acima, e de indivíduos saudáveis;
e) Construir uma plataforma de aprendizado de máquina para a classificação de atributos acústicos da fala dentre os transtornos mentais supracitados, com a habilidade de diferenciar cada um deles dos outros transtornos e de indivíduos saudáveis;
f) Desenvolver uma solução para auxiliar o diagnóstico e/ou a triagem dos transtornos abordados neste trabalho.

1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está estruturado da seguinte forma: no capítulo “Fundamentação teórica”, são introduzidos os conceitos teóricos necessários para a compreensão deste trabalho, e estes foram organizados em três grandes áreas: (1) conceitos de acústica vocal e dos parâmetros acústicos e os motivos para sua escolha como uma ferramenta de auxílio diagnóstico; (2) conceitos fundamentais dos transtornos mentais abordados neste trabalho, com ênfase em seus aspectos de comunicação oral; e (3) definições referentes às técnicas adotadas para a elaboração da solução deste problema de pesquisa, como processamento digital de sinais e algoritmos de aprendizado de máquina para classificação. Em seguida, no capítulo “Trabalhos relacionados”, é realizada uma revisão sobre artigos que utilizaram padrões de acústica vocal na área da

Psiquiatria, com foco naqueles sobre a detecção ou a avaliação da gravidade dos transtornos mentais aqui abordados ou relacionados a este trabalho. No capítulo “Resultados e discussão”, são mostrados os resultados experimentais obtidos, seguidos por uma análise qualitativa e quantitativa destes e das limitações desta pesquisa. Por fim, no capítulo de “Conclusão”, será realizada uma análise sobre a contribuição científica deste trabalho, como também sobre as perspectivas para estudos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo é composto por três seções. Na primeira, serão introduzidos os conceitos de acústica vocal relevantes para justificar seu uso como uma ferramenta auxiliar para a detecção de transtornos mentais, seguidos por definições sobre os parâmetros acústicos e, ainda, por uma breve síntese sobre o uso da acústica vocal para o reconhecimento de emoções. Na segunda seção, serão apresentados conceitos fundamentais dos transtornos mentais abordados nesta pesquisa – transtorno depressivo maior (TDM), transtorno bipolar (TB), esquizofrenia e transtorno de ansiedade generalizada (TAG) – enfatizando seus aspectos paralinguísticos. Devido à falta de estudos sobre correlatos acústicos no TAG, foi necessário estender a revisão para outros transtornos de ansiedade ou de reação ao estresse, como o transtorno de ansiedade social e o transtorno de estresse pós-traumático. A terceira e última seção abordará as ferramentas utilizadas para a resolução do problema desta pesquisa.

2.1 A VOZ

A voz é um sinal acústico originado pelos movimentos dos órgãos da fala e cujo objetivo fundamental é a comunicação (KENT; READ, 2015). Origina-se no trato vocal na laringe, estrutura anatômica em formato de tubo alongado localizada no pescoço e no interior da qual se encontram as pregas vocais (BEHLAU; PONTES; MORETTI, 2017). A energia para a produção da voz é proveniente do ar que sai dos pulmões, o qual, ao passar entre as pregas vocais, coloca-as em vibração (BEHLAU; PONTES; MORETTI, 2017). Esse fenômeno origina uma onda acústica analógica que, em seguida, sofrerá modulação pelos demais órgãos fonadores, como língua, lábios, mandíbula e véu palatino (RABINER; SCHAFER, 2007; KENT; READ, 2015).

De acordo com a teoria acústica da produção da fala, a onda acústica é o resultado de um sistema quase invariante no tempo em resposta a uma excitação que pode ser um ruído aleatório, um pulso de onda *quasi*-periódico ou uma mistura de ambos (SCHAFER; RABINER, 1970). Esse processo resulta em um sinal não estacionário (YADAV; JAIN; BHARGAV, 2015). A onda acústica representa, em última análise, a pressão do ar em função do tempo e, com auxílio de um software, pode fornecer uma visualização direta dos sons da fala captados por um microfone (BOERSMA, 2013).

Em relação às suas propriedades físicas, a voz é um sinal que compreende uma largura de banda acima de 10 KHz, possui uma extensão dinâmica de energia de 60 dB e sofre variações

significativas em intervalos curtos de tempo, inferiores a dez milissegundos (KENT; READ, 2015). Em alguns casos, como em frequências fundamentais muito baixas e em componentes fonêmicos de alta frequência, a largura de banda dos sinais da fala pode variar desde abaixo de 50 Hz até a faixa de 16-20 KHz (RABINER; SCHAFER, 2007; BEHRMAN, 2018). Entretanto, a maior parte da energia acústica do sinal vozeado se concentra nas frequências mais baixas, abaixo de 10 KHz, sofrendo decaimento médio de energia de 12 dB por oitava (KENT; READ, 2015; BEHRMAN, 2018).

Os sons da fala podem ser gerados de diferentes formas. Com base na fonte de excitação, estes podem ser classificados em vozeados, não vozeados e mistos (KADAMBE *et al.*, 1993) Os sons vozeados são periódicos ou semiperiódicos por natureza e resultam da excitação do trato vocal provocada pelos pulsos de pressão do ar, acarretando abertura e fechamento *quasi*-periódicos do orifício glótico (KADAMBE *et al.*, 1993; RABINER; SCHAFER, 2007). Exemplos destes são os sons vocálicos, como /i/ ,/e/, /a/, /ʌ/, /ɛ/, /æ/, /y/ e /u / (MATHEWS; MILLER; DAVID, 1961; POLS; VAN DER KAMP; PLOMP, 1969). Os sons não vozeados, por outro lado, são produzidos pela passagem de ar através de uma constrição em algum trecho do trato vocal (RABINER; SCHAFER, 2007), gerando sons de alta frequência semelhantes a ruídos (KADAMBE *et al.*, 1993). Os fonemas fricativos /s/, /ʃ/, /f/ e /θ/, por exemplo, são sons não vozeados (NARAYANAN; ALWAN, 2000). O termo fricativo indica consoantes que são produzidas com energia significativa decorrente de ruído (KENT; READ, 2015).

Os sons da fala são sinais complexos que contêm uma grande quantidade de informações (MUDA; BEGAM; ELAMVAZUTHI, 2010). Assim sendo, um dos desafios para a sua análise é identificar e obter suas informações mais significativas (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Para lidar com esse problema, foram desenvolvidas técnicas de extração de parâmetros acústicos. Estas fornecem uma representação compacta e paramétrica da onda acústica a uma taxa consideravelmente menor de dados, facilitando o processamento automatizado e a análise desses sinais (HASAN *et al.*, 2004; MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). A seguir será apresentada uma revisão sobre diferentes classes de parâmetros acústicos e seus principais representantes para o problema de pesquisa atual.

2.1.1 Parâmetros acústicos

Atualmente existem dezenas de parâmetros acústicos vocais heterogêneos, organizados em diferentes grupos de acordo com suas propriedades estruturais e semânticas. Entretanto, ainda não existe uma taxonomia bem definida para a classificação desses atributos

(MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Na literatura revisada, os atributos acústicos são frequentemente classificados em cinco grupos: (1) prosódicos; (2) espectrais, (3) fonéticos ou de qualidade vocal (QV); (4) cepstrais e (5) glóticos (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; HÖNIG *et al.*, 2014; CUMMINS *et al.*, 2015).

Para alguns autores, os parâmetros cepstrais não constituem um grupo distinto, sendo classificados como espectrais (ZHOU *et al.*, 2009; HÖNIG *et al.*, 2014; CUMMINS *et al.*, 2015). Entretanto, devido ao fato de os parâmetros cepstrais fornecerem uma representação acústica única que não se encontra nem no domínio do tempo nem do domínio da frequência (OPPENHEIM; SCHAFER, 2004), nesta revisão foi adotada a visão dos autores Mitrović, Zeppelzauer e Breiteneder (2010) e Lowell *et al.* (2012), a qual os classifica como um grupo de atributos à parte.

2.1.1.1 Parâmetros prosódicos

Os atributos prosódicos modelam os movimentos de relaxamento e tensão no trato vocal (HÖNIG *et al.*, 2014). Por meio desses, é possível obter informações valiosas sobre o estilo pessoal da fala e a entonação de um falante (DEHAK; DUMOUCHEL; KENNY, 2007). Pertencem a esse grupo a frequência fundamental (F0 ou *pitch*), os atributos de duração, de energia e de ritmo da fala (DEHAK; DUMOUCHEL; KENNY, 2007; LUGGER; YANG, 2007; HÖNIG *et al.*, 2014).

A frequência fundamental é definida como a medida da taxa de vibração das pregas vocais durante a fonação (BEHRMAN, 2018). Consiste na frequência mais baixa da tessitura vocal de um falante e constitui o determinante primário da percepção auditiva do *pitch* vocal (WEEKS *et al.*, 2016; BEHRMAN, 2018). A F0 resulta do comprimento e das características biodinâmicas das pregas vocais e da integração destas com a pressão subglótica, sendo calculada pelo número de ciclos glóticos por segundo (BEHLAU, 2001).

Os valores da F0 sofrem enorme influência do sexo e da idade do falante. Para indivíduos adultos, as faixas de frequência situam-se entre 80 a 150 Hz no sexo masculino, e entre 150 a 250 Hz no feminino (BEHLAU, 2001). Em crianças os valores da F0 são maiores que os de adultos; por exemplo, na idade de cinco anos, os valores médios da F0 são 240 Hz em meninos e de 243 Hz em meninas (BEHRMAN, 2018). Com o desenvolvimento, essa diferença entre os sexos se acentua, sendo que em meninos de 10 anos a F0 média é de 220 Hz, enquanto em meninas de 11 anos essa medida se situa em torno de 238 Hz (BEHRMAN, 2018).

Durante o processo de extração de atributos, a F0 de um falante pode ser estimada pela taxa de cruzamentos (*Zero Crossing Rate*, ZCR) de um sinal (BEHLAU, 2001). Esse parâmetro permite a obtenção da frequência dominante (i.e., a F0) de um sinal no domínio do tempo (KEDEM, 1986), sendo bastante popular em uma grande variedade de aplicações em áudio, como análise da fala, classificação de gêneros musicais e detecção de sons ambientais (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010).

Os parâmetros de duração avaliam durações silábicas, assim como a duração de pausas silenciosas entre as sílabas (HÖNIG *et al.*, 2014). Esses atributos são considerados robustos para tarefas de reconhecimento de fala em ambientes com altos níveis de ruído (CHIEN; HUANG, 2003). Os parâmetros relacionados ao ritmo, por sua vez, calculam a duração de outras estruturas linguísticas durante a fala, como pseudossílabas e intervalos vocálicos (TIMOSHENKO *et al.*, 2007; HÖNIG *et al.*, 2014). Foram utilizados com sucesso para a detecção de idiomas com base em suas diferentes configurações rítmicas (TIMOSHENKO *et al.*, 2007).

Por último, a energia de um sinal pode ser conceituada como o quadrado da amplitude da onda acústica. A potência, por sua vez, corresponde à energia transmitida por unidade de tempo (segundos), sendo definida como o quadrado médio de um sinal (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Como consequência, frequentemente a raiz quadrada da potência (*Root Mean Square*, RMS) é extraída para representar o comportamento da energia de um sinal acústico (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Também podem ser extraídos atributos que descrevem as propriedades estatísticas do contorno de energia de um sinal, como os valores máximo, médio, mínimo e mediano de curvas ascendentes e descentes e suas respectivas durações (VERVERIDIS; KOTROPOULOS; PITAS, 2004).

Como visto acima, os atributos prosódicos, como *pitch*, volume e ritmo, são utilizados para extrair os padrões de entonação de um falante (SUDHKAR; ANIL, 2015). Há vários anos têm sido considerados capazes de identificar o estado emocional de um falante (FRICK, 1985), por se correlacionarem com a informação afetiva contida na fala (SUDHKAR; ANIL, 2015). Devido a isso, foram fundamentais para o posterior desenvolvimento de sistemas automatizados de reconhecimentos de emoções, propiciando o aprimoramento de interfaces homem-máquina (VERVERIDIS; KOTROPOULOS; PITAS, 2004; LUGGER; YANG, 2007; ZHOU *et al.*, 2009; AMARAKEERTHI *et al.*, 2013).

2.1.1.2 Parâmetros espectrais

Os parâmetros espectrais são derivados da distribuição espectral da energia acústica e fornecem uma análise da onda acústica no domínio da frequência (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; LOWELL *et al.*, 2012). Atributos espectrais comumente utilizados são a densidade espectral de energia (*Power Spectral Density*, PSD) e os formantes (ZHOU *et al.*, 2009; HÖNIG *et al.*, 2014; CUMMINS *et al.*, 2015). Para alguns autores, também pertencem a essa classe os coeficientes cepstrais em frequência mel (*Mel Frequency Cepstral Coefficients*, MFCCs) e os coeficientes cepstrais de predição linear (*Linear Prediction Cepstral Coefficients*, LPCCs) (UTANE; NALBALWAR, 2013; JIANG *et al.*, 2019). Esta revisão, contudo, adota a visão de diversos autores que classificam os parâmetros cepstrais como um grupo distinto, conforme mencionado anteriormente.

Os formantes correspondem às frequências naturais de ressonância do trato vocal para a produção de determinados sons e indicam diferentes conformações do trato vocal durante a articulação de sons vocálicos (BEHRMAN, 2018; COMPTON *et al.*, 2018). A frequência do primeiro formante (F1) corresponde à abertura da boca, com inferiorização da mandíbula e consequente diminuição da altura da língua (COMPTON *et al.*, 2018). O segundo formante (F2) é majoritariamente determinado pelo formato da porção posterior da língua (BEHRMAN, 2018), também sendo influenciado pelo formato arredondado dos lábios (COMPTON *et al.*, 2018). A frequência do terceiro formante (F3), por outro lado, depende da posição da ponta da língua para a produção de determinadas vogais e semivogais (BEHRMAN, 2018).

Os parâmetros espectrais têm sido utilizados com êxito em várias aplicações relacionadas à fala, como sistemas de reconhecimento de fala e de falantes (KOOLAGUDI; RAO, 2012) e para o reconhecimento de emoções, sendo considerados, por alguns autores, superiores a outras classes de atributos para a identificação de falantes (UTANE; NALBALWAR, 2013). Também possuem a vantagem de modelar o espectro dos sinais da fala em uma imagem (espectrograma) e, a partir desta, extrair informações sobre as emoções, com base em parâmetros de imagem (JIANG *et al.*, 2019).

A PSD é outro atributo espectral bastante utilizado em tarefas de reconhecimento de emoções desencadeadas por imagens ou músicas, por exemplo (JATUPAIBOON; PANNGUM; ISRASENA, 2013; BHATTI *et al.*, 2016). Entretanto, suas aplicações se concentram em outros tipos de sinais biológicos, como a eletroencefalografia (EEG) e não serão aqui abordados.

2.1.1.3 Parâmetros de qualidade vocal

Os atributos de qualidade vocal ou fonéticos descrevem as propriedades da fonte glótica do sinal (LUGGER; YANG, 2007; KÄCHELE *et al.*, 2014). Essa classe de parâmetros mede o grau de irregularidade da fonação, obtido por meio de informações referentes a qualidades laríngeas, como soprosidade, crepitação, laringalização e aspereza (CUMMINS *et al.*, 2015). O termo qualidade vocal pode ser definido como o atributo perceptual que descreve o som da voz para além de seu *pitch* e volume (BEHRMAN, 2018). Para Behlau (2001), "é nossa avaliação perceptiva principal e relaciona-se à impressão total criada por uma voz" (p. 91).

Atributos de qualidade vocal comumente utilizados são as medidas de perturbação da frequência fundamental, denominados *jitter* e *shimmer* (CUMMINS *et al.*, 2015). Estes correspondem, respectivamente, a variações a curto prazo da frequência fundamental e da amplitude da onda acústica, e são calculados pela comparação entre ciclos glóticos sucessivos (BEHLAU, 2001; GIDDENS *et al.*, 2013; BEHRMAN, 2018; SPAZZAPAN *et al.*, 2019). O *jitter* correlaciona-se com aspereza vocal, enquanto o *shimmer* com a presença de rouquidão e soprosidade (BEHLAU, 2001). Desta forma, ambos são frequentemente utilizados para caracterizar a voz em condições normais e patológicas em diferentes faixas etárias (SPAZZAPAN *et al.*, 2019).

Os valores do *jitter* podem estar alterados em condições onde há pobre controle sobre a fonação (HÖNIG *et al.*, 2014), como nas disfonias neurológicas (BEHLAU, 2001) e em estados depressivos (QUATIERI; MALYSKA, 2012). As medidas do *shimmer*, por sua vez, podem sofrer alterações em situações de edema nas pregas vocais, na presença de lesões de massa na região laríngea (BEHLAU, 2001), como também na depressão maior (QUATIERI; MALYSKA, 2012; ALGHOWINEM *et al.*, 2013; HÖNIG *et al.*, 2014).

Outros descritores de qualidade vocal são a proporção harmônico-ruído (*Harmonics-to-Noise Ratio*, HNR), a harmonicidade espectral e o *tilt* espectral (D'ALESSANDRO; DOVAL, 2003; HÖNIG *et al.*, 2014; CUMMINS *et al.*, 2015). A HNR é uma medida da quantidade de ruído aditivo no sinal de voz (FERRAND, 2002), indicando sua periodicidade (BOERSMA, 1993) e o grau de rouquidão ou rugosidade (*hoarseness*) presente nesse sinal (YUMOTO; GOULD; BAER, 1982). Seus valores, com frequência, estão diminuídos na presença de fonação rouca e/ou soprosa (HÖNIG *et al.*, 2014). Já o *tilt* espectral é definido pela proporção entre as intensidades do primeiro e do segundo harmônicos (CAMPBELL; BECKMAN, 1997), refletindo as diferenças entre as bandas de frequências mais altas e mais baixas (KAKOUROS; RÄSÄNEN; ALKU, 2018). Esse parâmetro têm sido utilizado com sucesso para a identificação

de valências afetivas distintas em sistemas de classificação de emoções (LISCOMBE; VENDITTI; HIRSCHBERG, 2003). Por último, a harmonicidade espectral avalia a integridade dos harmônicos em um segmento vocal, fornecendo uma análise sistemática e detalhada da estrutura harmônica (YU; WANG, 2004). Assim como o parâmetro anterior, a harmonicidade espectral também tem se mostrado útil na detecção de valências afetivas no contexto de reconhecimento de emoções (SCHULLER *et al.*, 2012).

2.1.1.4 Parâmetros cepstrais

Este grupo é composto principalmente pelos MFCCs (HASAN *et al.*, 2004). Introduzido por Bogert, Healy e Tukey (1963), o termo "cepstrum" é um anagrama da palavra *spectrum* e pode ser definido como o espectro da amplitude logarítmica do espectro de energia (NOLL, 1967). Os MFCCs começaram a ser utilizados no início dos anos 1980 em aplicações de reconhecimento de fala, sendo posteriormente adotados para a identificação de falantes (KINNUNEN; LI, 2010). Caracterizam a energia do sinal acústico em bandas críticas de frequência de acordo a audição humana e, para isso, utilizam a escala em frequência mel, composta por um banco de filtros linear e logarítmico (HASAN *et al.*, 2004; BEDOYA-JARAMILLO *et al.*, 2012)

O cálculo dos coeficientes cepstrais é realizado com a ajuda desse banco de filtros, seguida pela compressão logarítmica do espectro de energia e pela transformada discreta do cosseno sobre a escala não linear de frequência mel (KINNUNEN; LI, 2010; HIBARE; VIBHUTE, 2014). Os MFCCs são parâmetros muito populares no processamento de áudio e da fala (KINNUNEN; LI, 2010). Também fornecem uma boa representação das propriedades espectrais de um sinal em determinada janela (TIWARI, 2010) e são adequados para calcular uma aproximação do envelope do espectro (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010), definido como o formato do espectro de potência de um sinal acústico (WARREN; JENNINGS; GRIFFITHS, 2005).

Deve-se salientar que os MFCCs apresentam uma relação física com os mecanismos de produção da fala (BEDOYA-JARAMILLO *et al.*, 2012) e são capazes de refletir mudanças no trato vocal (TAGUCHI *et al.*, 2018). Possivelmente devido a essas propriedades, são atributos amplamente utilizados em sistemas de reconhecimento de voz (HASAN *et al.*, 2004; MUDA; BEGAM; ELAMVAZUTHI, 2010) e no reconhecimento de emoções (LEE *et al.*, 2004; ZHOU *et al.*, 2009; MEUDT; SCHWENKER, 2012; DEMIRCAN; KAHRAMANLI, 2014). Além disso, alguns estudos têm demonstrado a utilidade dos MFCCs para a detecção de transtornos

mentais, como ansiedade na infância (MCGINNIS *et al.*, 2019), depressão maior (REJAIBI *et al.*, 2019) e esquizofrenia (ZHANG *et al.*, 2016).

2.1.1.5 Parâmetros glóticos

Os atributos glóticos se relacionam à fonte do sinal de voz, denominadas em inglês de *source features*. Estes captam informações relacionadas à fonte da produção vocal, originada pelo fluxo de ar dos pulmões através da glote (CUMMINS *et al.*, 2015). Os atributos glóticos têm se mostrado valiosos para a identificação de valências afetivas pela voz (SUN; MOORE, 2011), como também se mostraram capazes de detectar transtornos depressivos (MOORE *et al.*, 2008; LOW *et al.*, 2011; SCHERER *et al.*, 2013). Exemplos de parâmetros glóticos no domínio do tempo são os quocientes de abertura, de amplitude, de fechamento, de amplitude normalizada, de quase abertura, entre outros (SUN; MOORE, 2011).

2.1.2 Reconhecimento de emoções por meio da voz

A produção vocal resulta de um conjunto de características anatômicas, fisiológicas, de personalidade e da cultura onde o indivíduo está inserido (BEHLAU; PONTES; MORETTI, 2017). O som produzido na laringe depende de um controle cerebral refinado, por meio da inervação dos nervos laríngeos, e da integridade do aparelho fonador (BEHLAU, 2001; BEHLAU; PONTES; MORETTI, 2017), sendo o resultado de interações complexas entre diversas funções cognitivas e o sistema musculoesquelético (LARSEN *et al.*, 2015). Nesse contexto, destaca-se a influência da formação psíquica do indivíduo sobre a produção da voz, a qual constitui uma forte extensão de sua personalidade ou, ainda, uma "manifestação com base psicológica" (BEHLAU, 2001).

Durante a comunicação oral, a voz transmite várias informações sobre um indivíduo, como seu gênero, idade e estado afetivo (BELIN *et al.*, 2000; AGUS *et al.*, 2010; RAMIREZ; BALTRUŠAITIS; MORENCY, 2011). Nesse contexto, diversos estudos evidenciaram que alterações cognitivas ou fisiológicas, provocadas por flutuações no estado afetivo, mesmo quando discretas, podem produzir alterações acústicas perceptíveis (CUMMINS *et al.*, 2015; LARSEN *et al.*, 2015; VAN PUYVELDE *et al.*, 2018). Assim sendo, ao longo de décadas, diversos trabalhos têm demonstrado que as emoções podem ser identificadas por meio da acústica vocal (FRICK, 1985; LUGGER; YANG, 2007; KÄCHELE *et al.*, 2014). Investigar e

determinar essas alterações são objetivos da área de reconhecimento de emoções, um importante braço de pesquisa da computação afetiva (KÄCHELE *et al.*, 2014).

Em tarefas de classificação de emoções, todos os grupos de parâmetros vocais se mostraram úteis (BACHOROWSKI; OWREN, 1995; NWE; FOO; DE SILVA, 2003; LEE *et al.*, 2004; ZHOU *et al.*, 2009; MEUDT; SCHWENKER, 2012; TAHON; DEGOTTEX; DEVILLERS, 2012; KÄCHELE *et al.*, 2014). Por exemplo, Lugger e Yang (2007) demonstraram que os parâmetros prosódicos e os de qualidade vocal têm a capacidade de reconhecer emoções, particularmente raiva, tristeza e ansiedade. Similarmente, nos trabalhos de Nwe, Foo, e De Silva (2003) e de Qin e Zhang (2011) os atributos espectrais permitiram a identificação de emoções como raiva, nojo, medo, alegria, tristeza e surpresa.

No contexto das dimensões afetivas, os autores Elamir, Al-Atabany e Eldosoky (2019) demonstraram em seu trabalho que os parâmetros de Hjorth, frequentemente utilizados para avaliar a complexidade de sinais de EEG, também podem ser eficazes para o reconhecimento automatizado de emoções relacionadas à valência e à ativação. De forma semelhante, Tahon, Degottex e Devillers (2012) relataram que os atributos glóticos, em conjunto com os de QV, se mostraram promissores para a detecção de valências afetivas em modelos independentes do falante (*speaker-independent*). Por último, Sun e Moore (2011) relataram o desempenho equivalente ou mesmo superior dos atributos glóticos e do operador de energia de Teager (TEO) para a identificação das quatro dimensões afetivas em comparação com outras classes de atributos (espectrais e prosódicos). O TEO é um parâmetro acústico não linear da fala utilizado para estimar a energia de um sinal, calculado como o produto da amplitude pela frequência (KAMBLE; PATIL, 2019). Esse atributo possui variadas aplicações relacionadas à fala e, na área de reconhecimento de emoções, tem sido empregado na identificação de emoções relacionadas ao estresse (BANDELA; KUMAR, 2017). O Quadro 1 abaixo exibe uma síntese dos correlatos acústicos de algumas dessas emoções.

Quadro 1 – Correlatos acústicos de diferentes emoções positivas e negativas

Emoção	Dados acústicos
Raiva	Aumento de F0 Maior variabilidade de F0, com contornos descendentes Aumento do nível médio de energia do espectro Aumento de energia nas regiões altas Maior velocidade de fala

Continua

Cont. Quadro 1

Medo	Aumento de F0 Maior variabilidade de F0 Aumento de energia nas regiões altas Maior velocidade de fala
Tristeza	Diminuição de F0 Menor variabilidade de F0 Contornos descendentes de F0 Redução do nível médio de energia do espectro Menor velocidade de fala Pouca energia nas regiões altas
Alegria	Aumento de F0 Maior variabilidade de F0 Aumento do nível médio de energia Energia nas regiões de maior frequência Maior velocidade de fala

Fonte: Adaptado de Pittam e Scherer (1993) apud Behlau (2001, p. 156).

A área de reconhecimento de emoções pela voz possui diversas aplicações. Entre estes, destacam-se as interfaces homem-máquina, a identificação de falantes, a avaliação de doenças neurológicas e o desenvolvimento de ferramentas para o apoio diagnóstico de transtornos mentais (BEDOYA-JARAMILLO *et al.*, 2012). Sendo o último o propósito desta pesquisa, na seção seguinte serão definidos e contextualizados os transtornos mentais abordados neste trabalho. Ainda nessa seção, os conceitos sobre os parâmetros de acústica vocal serão apresentados, enquanto sua utilização para a identificação de transtornos mentais será exposta no próximo capítulo (“Trabalhos relacionados”). Na terceira e última seção deste capítulo, será apresentada uma revisão sobre as técnicas de processamento digital de sinais acústicos e os algoritmos computacionais de classificação de padrões adotados nesta pesquisa para a identificação dos transtornos mentais com base em atributos acústicos.

2.2 TRANSTORNOS MENTAIS E SEUS PADRÕES DE FALA

Desde 1938, inúmeros trabalhos têm evidenciado alterações da comunicação oral nos transtornos mentais, inicialmente estudados nos transtornos de humor (NEWMAN; MATHER, 1938). Dalgalarrondo (2019) define o humor como o tônus afetivo de um indivíduo, ou seja, “o estado emocional basal e difuso em que se encontra uma pessoa em determinado momento” (p. 280). Os transtornos de humor, também denominados transtornos afetivos, compreendem um

grupo de doenças psiquiátricas que apresentam como característica principal os distúrbios do humor (AMERICAN PSYCHIATRIC ASSOCIATION, 2000), sendo representados principalmente pelos transtornos depressivos e pelo transtorno bipolar (SADOCK; SADOCK; RUIZ, 2017).

Sabendo-se que mudanças no estado de humor podem se manifestar na voz (MUNDT *et al.*, 2012; HIGUCHI *et al.*, 2019), é natural supor que estas também se manifestem nos transtornos de humor. De fato, vários estudos associam tanto o transtorno depressivo maior quanto o transtorno bipolar a alterações de parâmetros vocais (CUMMINS *et al.*, 2011; MUNDT *et al.*, 2012; VANELLO *et al.*, 2012; FAURHOLT-JEPSEN *et al.*, 2016; MAXHUNI *et al.*, 2016). Para além dos transtornos de humor, alterações vocais também foram encontradas em outras categorias nosológicas, como os transtornos de ansiedade, a esquizofrenia e os transtornos por uso de substâncias (BEHLAU, 2001; ALPERT *et al.*, 2002; WEEKS *et al.*, 2012). Os trabalhos relacionados a alterações acústicas nos transtornos mentais relacionados a este trabalho serão discutidos no capítulo seguinte.

Para identificar o impacto dessas alterações sobre a voz, todas as classes de parâmetros acústicos são consideradas relevantes. Por exemplo, existem evidências robustas de alterações em diversos atributos vocais em indivíduos deprimidos, como o *jitter*, o *shimmer* e parâmetros de energia e de intensidade (ALGHOWINEM *et al.*, 2013; HÖNIG *et al.*, 2014). Entre estes, o *jitter* é particularmente importante para o reconhecimento de estados afetivos, devido à sua capacidade de identificar mudanças rápidas e temporárias na voz decorrentes de desregulação do sistema nervoso autônomo (SNA), com consequente influência sobre o tônus muscular e controle articulatório (OZDAS *et al.*, 2004; MAXHUNI *et al.*, 2016). Outro exemplo são MFCCs, os quais têm sido utilizados como atributo tanto para a detecção de depressão (CUMMINS *et al.*, 2011, 2014; JIANG *et al.*, 2018) como para a avaliação da gravidade dos seus sintomas (STURIM *et al.*, 2011).

Considerando a importância da análise da voz e suas alterações nos transtornos mentais, os recentes avanços computacionais podem torná-la uma ferramenta promissora para a avaliação objetiva dos sintomas psiquiátricos (KENT; READ, 2015; TAHIR *et al.*, 2019). Nas duas subseções seguintes, será apresentada a fundamentação teórica sobre os padrões de acústica vocal nos transtornos mentais relacionados a este trabalho, primeiramente nos transtornos de humor (transtorno depressivo maior e transtorno bipolar), seguidos por estudos referentes à esquizofrenia e, por último, os transtornos de ansiedade.

2.2.1 Transtorno depressivo maior

O transtorno depressivo maior (TDM) ou depressão maior é um transtorno mental bastante comum, afetando mundialmente mais de 300 milhões de pessoas em todas as faixas etárias (WORLD HEALTH ORGANIZATION, 2018a). Seu quadro clínico é caracterizado por humor triste ou irritável e/ou anedonia, associado a alterações cognitivas e psicomotoras (retardo ou agitação psicomotora, dificuldades na tomada de decisão, pensamentos de culpa ou inutilidade, pensamentos de morte); sinais e sintomas neurovegetativos (distúrbios do sono, alterações do apetite ou do peso corporal), que causam sofrimento e/ou prejuízo funcional significativo e podem levar ao suicídio (AMERICAN PSYCHIATRIC ASSOCIATION, 2013; SADOCK; SADOCK; RUIZ, 2017). Dessa maneira, o TDM está associado a uma alta morbimortalidade e compromete múltiplos domínios da vida do indivíduo, como as esferas profissional, acadêmica e social (WEINBERGER *et al.*, 2017; WORLD HEALTH ORGANIZATION, 2017). Entretanto, apesar do imenso impacto sócio-ocupacional e de investimentos existentes em intervenções terapêuticas, o tratamento desse transtorno ainda costuma ser tardio e inacessível à maioria dos pacientes (WEINBERGER *et al.*, 2017).

Atualmente, a prevalência global média da depressão maior é estimada em 4,4% e varia de acordo com idade, sexo e região do planeta (WORLD HEALTH ORGANIZATION, 2017). Seu pico de prevalência ocorre na faixa etária mais idosa, afetando cerca de 7,5% das mulheres e 5,5% dos homens com idade entre 55 e 74 anos (WORLD HEALTH ORGANIZATION, 2017). Além disso, nos últimos anos a depressão tem se tornado mais comum nos extremos de idade, porém com uma velocidade maior na população mais jovem (WEINBERGER *et al.*, 2017). Por exemplo, em adolescentes foi registrado um rápido aumento da prevalência recentemente, saltando de 3,9% em 2013 para 5,8% em 2017 (BARROS *et al.*, 2017; WORLD HEALTH ORGANIZATION, 2017).

No Brasil, a prevalência da depressão está atualmente estimada em 5,8%, sendo a quinta maior do mundo (WORLD HEALTH ORGANIZATION, 2017), enquanto sua prevalência durante a vida pode chegar a até 16,8% (MIGUEL; GENTIL; GATTAZ, 2011). Em outras palavras, uma em cada seis pessoas desenvolverá depressão pelo menos uma vez na vida no nosso país. Contudo, são necessários estudos mais recentes para avaliar a atual prevalência de depressão durante a vida no Brasil.

Sobre suas características sociodemográficas, o transtorno depressivo acomete cerca de duas mulheres para cada homem e é mais comum nos continentes africano e americano (OTTE *et al.*, 2016; WORLD HEALTH ORGANIZATION, 2017). A média de idade para início desse

transtorno situa-se em torno dos 40 anos, com metade dos pacientes apresentando início dos sintomas entre os 20 e 50 anos (SADOCK; SADOCK; RUIZ, 2017).

As consequências da depressão em termos de perda de saúde são imensas. Em 2017, foi a terceira maior causa mundial de incapacitação no sexo feminino e a quinta no sexo masculino, em termos de anos vividos com incapacitação (*Years Lived with Disability*, YLDs) (GBD 2017 DISEASE AND INJURY INCIDENCE AND PREVALENCE COLLABORATORS, 2018). Considerados medidas do *burden* de uma doença, os YLDs correspondem ao número de anos de saúde perdidos em decorrência de determinada doença ou transtorno (US BURDEN OF DISEASE COLLABORATORS, 2013). Ao final de 2017, a depressão foi a maior causa de incapacitação entre os transtornos mentais, responsável por um total de 43 milhões de YLDs (GBD 2017 DISEASE AND INJURY INCIDENCE AND PREVALENCE COLLABORATORS, 2018).

O TDM também é o maior fator de risco isolado para o suicídio (WEINBERGER *et al.*, 2017), estimando-se que cerca de metade destes estejam relacionados à ocorrência de depressão (CUMMINS *et al.*, 2015). Anualmente, o suicídio leva à perda de cerca de 800.000 vidas e é a segunda causa de morte entre jovens de 15 a 29 anos (WORLD HEALTH ORGANIZATION, 2017). Além disso, a prevalência de suicídio se concentra em regiões subdesenvolvidas, havendo 79% dos casos no ano de 2016 ocorrido em países de baixa e média renda (WORLD HEALTH ORGANIZATION, 2018b).

A depressão também está associada a hábitos de vida prejudiciais. Em um estudo com brasileiros, Barros *et al.* (2017) encontraram maior prevalência de hábitos não saudáveis entre portadores de depressão, como tabagismo, sedentarismo, uso abusivo de álcool, consumo de alimentos gordurosos e refrigerantes. Esses hábitos de vida, em suma, poderiam justificar o aumento da mortalidade por outras causas entre portadores de depressão, como as doenças cardiovasculares (VAN DER KOOY *et al.*, 2007).

Mudanças no estado afetivo são comuns na depressão e podem afetar os mecanismos de produção da fala (CUMMINS *et al.*, 2015). Dentre os sintomas depressivos, o retardo psicomotor e o comprometimento cognitivo, sintomas comuns e precocemente encontrados na depressão, podem se apresentar como alterações na fala (HASHIM *et al.*, 2016). De fato, alterações nos aspectos paralinguísticos da fala de indivíduos deprimidos foram relatadas por clínicos várias décadas atrás (DARBY; HOLLIEN, 1977; ALPERT; POUGET; SILVA, 2001). Para ilustrar, há um século, Kraepelin (1921) caracterizou o discurso depressivo como lento, hesitante, monótono, com baixo volume, monossilábico, podendo chegar até o mutismo.

Diversos relatos se seguiram, os quais consistentemente descreveram a fala do deprimido como monótona, entediante e sem energia (JIANG *et al.*, 2018).

Mais recentemente, com o avanço das ferramentas computacionais de análise acústica, alterações quantitativas do discurso também foram descritas. Em conformidade com descrições anteriores sobre a fala de indivíduos deprimidos, foram relatadas diminuição da variabilidade do *pitch* vocal (MUNDT *et al.*, 2007), aumento do número de pausas (MUNDT *et al.*, 2012), diminuição da velocidade do discurso (CANNIZZARO *et al.*, 2004; FAURHOLT-JEPSEN *et al.*, 2016) e da velocidade de articulação (SCHERER *et al.*, 2013) e redução da intensidade ou volume da fala (HÖNIG *et al.*, 2014). Considerando esses achados, a análise das características acústicas da voz poderia, então se tornar um marcador objetivo para a identificação do transtorno depressivo maior (JIANG *et al.*, 2018), aumentando a precisão diagnóstica e auxiliando a redução do alto impacto socioeconômico associado a esse transtorno (CUMMINS *et al.*, 2015).

2.2.2 Transtorno bipolar

O transtorno bipolar (TB) é uma condição crônica que envolve perturbações graves do humor, alterações imunológicas e fisiológicas, déficits neuropsicológicos e comprometimento do funcionamento em diversas áreas da vida do indivíduo (SADOCK; SADOCK; RUIZ, 2017; ROWLAND; MARWAHA, 2018). Seu quadro clínico é caracterizado por episódios de elevação persistente do humor para expansibilidade ou irritabilidade e aumento da energia, acompanhados por aumento da atividade psicomotora, alterações cognitivas (aceleração do pensamento, distratibilidade) e neurovegetativas, como redução da necessidade de sono (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). O TB possui prevalência estimada em torno de 2,4% e costuma iniciar-se na faixa etária de 20 a 30 anos (ROWLAND; MARWAHA, 2018), podendo variar desde a infância até 50 anos ou mais, em casos raros (SADOCK; SADOCK; RUIZ, 2017). Também está associado a altas taxas de mortalidade prematura, tanto por causas naturais, como por causas externas, em particular o suicídio (HAYES *et al.*, 2015).

Em relação ao *burden* da doença, o transtorno bipolar também se relaciona a prejuízos significativos nos funcionamentos profissional, familiar e social que se estendem para além das fases agudas da doença (SANCHEZ-MORENO *et al.*, 2009). Parte substancial desse *burden* está relacionada ao comportamento suicida, estimando-se que 32 a 36% dos portadores do TB apresentarão pelo menos uma tentativa de suicídio durante a vida (NOVICK; SWARTZ;

FRANK, 2010). Além disso, cerca 3,4 até 14% de todos os suicídios ocorrem com pacientes bipolares (SCHAFFER *et al.*, 2015).

De acordo com a quinta edição do Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5) da Associação Americana de Psiquiatria (APA), os episódios de perturbação do humor podem ser graves ao ponto de causar prejuízos acentuados no funcionamento sócio-ocupacional do indivíduo ou necessitar de hospitalização para evitar danos ao paciente ou a terceiros. Nesse caso, são denominados mania; quando de menor gravidade, são definidos como hipomania. Com base nesses episódios, existem duas subcategorias diagnósticas oficiais do transtorno bipolar: tipo I, definido pela de existência pelo menos um episódio de mania durante a vida do indivíduo; e o tipo II, caracterizado pela ocorrência de pelo menos um episódio hipomaníaco. A presença de episódios depressivos é frequente tanto no tipo I como no tipo II, porém não é necessária para o diagnóstico de TB em nenhum dos tipos (AMERICAN PSYCHIATRIC ASSOCIATION, 2013).

Assim como na depressão, as alterações de fala são frequentemente encontradas no transtorno bipolar. Para a fase maníaca, Sadock, Sadock e Ruiz (2017) descrevem que a intensidade destas alterações guarda uma relação direta com a gravidade do episódio. De acordo com esses autores, a fala do indivíduo tende a se tornar mais alta, mais rápida e de difícil interpretação, passando a apresentar piadas, rimas, trocadilhos, jogo de palavras e irrelevância. Com o aumento da gravidade do episódio, pode ocorrer afrouxamento das associações ideativas, com fuga de ideias e neologismos. Nos casos mais graves, a fala pode ser tornar completamente incoerente, tornando-se indistinguível daquela de um indivíduo com esquizofrenia.

Outro exemplo da importância das alterações no discurso no transtorno bipolar é a presença de itens sobre alterações na fala na Escala de Avaliação de Mania de Young (*Young Mania Rating Scale*, YMRS), como velocidade da fala e nível de incoerência do discurso. Considerada padrão-ouro para a avaliação da gravidade do episódio de mania, a YMRS é a escala psicométrica mais utilizada em estudos sobre o transtorno bipolar (YOUNG *et al.*, 1978; VILELA *et al.*, 2005; MUAREMI *et al.*, 2014).

Por último, variações no *pitch* parecem estar correlacionadas com mudanças de humor em pacientes bipolares (MAXHUNI *et al.*, 2016), assim como um aumento da atividade do discurso pode ser um sinal de virada de humor para mania ou hipomania (FAURHOLT-JEPSEN *et al.*, 2016). Todos esses achados demonstram a importância de aspectos relativos à produção da fala para a avaliação da presença e da gravidade dos sintomas maníacos ou hipomaníacos no transtorno bipolar.

2.2.3 Esquizofrenia

A esquizofrenia é um grupo heterogêneo de transtornos psicóticos graves com diferentes etiologias, apresentações clínicas e respostas ao tratamento (SADOCK; SADOCK; RUIZ, 2017). Seus sinais e sintomas variam e incluem alterações perceptivas, cognitivas, afetivas e comportamentais (SADOCK; SADOCK; RUIZ, 2017). Esse transtorno possui prevalência estimada em torno de 0,48% a 1%; costuma ter início entre o final da adolescência e o início da idade adulta e geralmente continua por toda a vida (FREEDMAN, 2003; SIMEONE *et al.*, 2015), com sintomas graves que afetam profundamente o funcionamento pessoal, profissional e psicossocial (GREEN, 2006; BUCHANAN, 2007; RABINOWITZ *et al.*, 2012; MILLIER *et al.*, 2014; CHARLSON *et al.*, 2018).

Apesar da prevalência relativamente baixa, o *burden* da doença na esquizofrenia é substancial e está associado a baixas taxas de remissão sintomatológica e de recuperação funcional e à redução da expectativa de vida (CHARLSON *et al.*, 2018). Não obstante a redução da mortalidade por suicídio em determinados países (TANSKANEN; TIIHONEN; TAIPALE, 2018), a esquizofrenia ainda está associada a alto risco de comportamento suicida, com taxas de suicídio em torno de 5% (HOR; TAYLOR, 2010) e cerca de 20-30% dos pacientes apresentando pelo menos uma tentativa de suicídio durante a vida (RADOMSKY *et al.*, 1999; AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Entretanto, a maior parcela da mortalidade em excesso na esquizofrenia se deve a comorbidades médicas, especialmente doenças cardiovasculares, diabetes tipo II, doenças respiratórias e alguns tipos de câncer (BUSHE; TAYLOR; HAUKKA, 2010; CHARLSON *et al.*, 2018). Essas altas taxas de mortalidade podem ser explicadas, em parte, por hábitos de vida não saudáveis, distúrbios metabólicos associados à doença e ao tratamento, com consequente aumento da incidência de doenças cardiovasculares (MILLIER *et al.*, 2014).

O quadro clínico da esquizofrenia é heterogêneo e envolve a presença de sintomas psicóticos ditos “positivos” como delírios e alucinações, sintomas de desorganização do pensamento (evidenciada pelo discurso) e do comportamento motor e sintomas “negativos”, como avolia, diminuição da expressividade emocional, alogia, anedonia ou perda da sensação de prazer e retraimento social (FREEDMAN, 2003; AMERICAN PSYCHIATRIC ASSOCIATION, 2013).

Entre os sintomas positivos, os delírios são juízos patologicamente falsos, ou seja, são crenças fixas impossíveis de modificação pela experiência objetiva, irremovíveis e irrefutáveis apesar de evidências contraditórias (JASPERS, 1946; AMERICAN PSYCHIATRIC

ASSOCIATION, 2013; DALGALARRONDO, 2019). Já as alucinações são definidas pela percepção clara e definida de um objeto (e.g., voz, imagem) sem a presença deste, i.e., sem o respectivo estímulo sensorial (TELLES-CORREIA; MOREIRA; GONÇALVES, 2015; DALGALARRONDO, 2019). De acordo com a natureza do canal sensorial envolvido, as alucinações são classificadas em visuais, auditivas, táteis, olfativas, gustativas, somáticas (ou cenestésicas) e cinestésicas ou de movimento (DALGALARRONDO, 2019). De todas estas, o tipo mais comum na esquizofrenia são as alucinações auditivas, seguidas pelas visuais (SADOCK; SADOCK; RUIZ, 2017; DALGALARRONDO, 2019).

Os sintomas negativos da esquizofrenia são definidos como a ausência ou diminuição de comportamentos ou funções normais (BUCHANAN, 2007). Costumam ser persistentes e causam maior impacto sobre o funcionamento do que os sintomas positivos (BUCHANAN, 2007; RABINOWITZ *et al.*, 2012). Devido a isso, esses sintomas são responsáveis por grande parcela da morbidade crônica e desfecho funcional desfavorável (BUCHANAN, 2007; RABINOWITZ *et al.*, 2012) e costumam persistir por toda a vida, estimando-se que apenas 13,5% dos pacientes consigam alcançar recuperação clínica e social (CHARLSON *et al.*, 2018).

As disfunções cognitivas também são sintomas centrais da esquizofrenia e tendem a surgir antes do aparecimento de sintomas psicóticos (FREEDMAN, 2003; KEEFE; HARVEY, 2012). Compreendem prejuízos em diversos domínios cognitivos, como memória de trabalho, atenção/vigilância, memórias verbal e visual, resolução de problemas, funcionamento executivo, velocidade de processamento e cognição social (KEEFE; HARVEY, 2012; MILLIER *et al.*, 2014). O comprometimento cognitivo está presente na maioria dos pacientes esquizofrênicos e é considerado um preditor confiável de pobre desempenho laborativo e social (FREEDMAN, 2003; GREEN, 2006; MILLIER *et al.*, 2014).

Desde as primeiras descrições sobre a esquizofrenia, as anormalidades de voz e linguagem foram incluídas entre características principais desse transtorno, sendo frequentemente associadas a sintomas negativos e a prejuízos na sociabilidade (PAROLA *et al.*, 2020). Essas alterações incluem pobreza de discurso, descarrilamento, tangencialidade, discurso desorganizado, neologismo, incoerência, mutismo, perseveração, ecolalia, bloqueio do pensamento e prosódia inapropriada ou aprosódia (ELITE *et al.*, 2014; CHAKRABORTY *et al.*, 2018a; MAC-KAY; JEREZ; PESENTI, 2018). Considerada um sintoma negativo da esquizofrenia (COVINGTON *et al.*, 2012), a aprosódia consiste na diminuição da ênfase vocal, na redução da fluência e da inflexão da fala e em déficits de compreensão de prosódia, como, por exemplo, a dificuldade de reconhecer padrões de entonação vocal (ALPERT; ANDERSON, 1977; ALPERT *et al.*, 2000; ELITE *et al.*, 2014).

Em suma, os sintomas negativos são a consequência de perturbações nos processos cognitivos subjacentes e contribuem para os déficits de comunicação frequentemente vistos na esquizofrenia. Todos esses sintomas poderiam, portanto, auxiliar na identificação desse transtorno tomando como base a análise das suas características de comunicação oral.

2.2.4 Transtornos de ansiedade

Os transtornos de ansiedade formam uma categoria diagnóstica de transtornos heterogêneos que compartilham respostas comportamentais disfuncionais relacionadas ao medo excessivo e à ansiedade (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Abrangem o transtorno de ansiedade generalizada, o transtorno de ansiedade social, o transtorno de pânico, as fobias específicas, entre outros (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Quando considerado como um grupo, são os transtornos mentais mais prevalentes (GBD 2017 DISEASE AND INJURY INCIDENCE AND PREVALENCE COLLABORATORS, 2018), podendo atingir uma prevalência durante a vida de até 33,7%. Além disso, os transtornos de ansiedade também estão associados a uma considerável incapacidade funcional e a elevados custos de saúde (BANDELOW; MICHAELIS, 2015).

De acordo com um levantamento realizado pela Organização Mundial de Saúde (OMS), a proporção da população mundial com algum transtorno de ansiedade apresentou aumento entre os anos de 2005 e 2015, sendo atualmente estimada em 3,6%. Assim como na depressão, esses transtornos são mais comuns em mulheres do que em homens (4,6% versus 2,6%, respectivamente, em uma escala global). Ainda segundo esse estudo, o Brasil foi considerado o líder mundial em transtornos de ansiedade, com uma prevalência de 9,3%, correspondendo a mais de 18 milhões de pessoas acometidas (WORLD HEALTH ORGANIZATION, 2017).

Os transtornos de ansiedade também estão associados a *burden* significativo, com 27 milhões de YLDs no ano de 2017 (GBD 2017 DISEASE AND INJURY INCIDENCE AND PREVALENCE COLLABORATORS, 2018). Entretanto, apesar de serem mais prevalentes que a depressão, estão associados a menor *burden*, provavelmente por estarem associados a um nível menor de incapacitação (WORLD HEALTH ORGANIZATION, 2017). Ainda assim, são a sexta maior causa mundial de perda de saúde não fatal (WORLD HEALTH ORGANIZATION, 2017).

Neste capítulo de revisão, foram abordados dois transtornos de ansiedade: o transtorno de ansiedade generalizada e o transtorno de ansiedade social. Também foi incluído o transtorno de estresse pós-traumático, por compartilhar diversas características com essa categoria

diagnóstica. Primeiramente, devido à sua alta prevalência e ao seu caráter crônico e persistente, o transtorno de ansiedade generalizada (TAG) foi contemplado no desenvolvimento da ferramenta de apoio diagnóstico com base em atributos vocais deste trabalho. Contudo, como não foram encontrados trabalhos relacionados sobre os correlatos acústicos do TAG, com o objetivo de enriquecer a revisão bibliográfica deste trabalho, foram incluídos nesta revisão os dois transtornos acima por possuírem estudos sobre padrões de acústica vocal. Os demais transtornos de ansiedade fogem do escopo desta revisão e, portanto, não serão abordados.

O TAG é um transtorno frequente, crônico e debilitante, apresentando prevalência durante a vida de até 8% a 9% (AMERICAN PSYCHIATRIC ASSOCIATION, 2013; SADOCK; SADOCK; RUIZ, 2017). Caracteriza-se por ansiedade ou preocupações constantes com diversas atividades, associadas a sintomas físicos, como dificuldade de concentração, perturbações do sono, fatigabilidade, tensão muscular, inquietude e irritabilidade. Esses sintomas são intensos ao ponto de causar sofrimento significativo ou de interferir negativamente no funcionamento do indivíduo (HANS-ULRICH WITTCHEN, 2002; AMERICAN PSYCHIATRIC ASSOCIATION, 2013).

Estudos sobre a evolução do TAG sugerem que este é um transtorno crônico, com curso flutuante e pouquíssimas remissões completas (HANS-ULRICH WITTCHEN, 2002). Está associado a grave comprometimento funcional, em particular dos relacionamentos afetivos, a altos níveis de absenteísmo e de visitas a serviços de saúde (ROWA *et al.*, 2017). Adicionalmente, quando comparados a indivíduos não ansiosos, portadores de TAG relatam qualidade de vida significativamente pior e menores índices de satisfação com diversas áreas da vida, como trabalho, saúde, autoestima e relacionamentos sociais (ROWA *et al.*, 2017).

O transtorno de ansiedade social (TAS) ou fobia social apresenta como característica central o medo ou a ansiedade acentuada e persistente em interações sociais, situações de desempenho, ser observado e o medo de se comportar de maneira constrangedora ou humilhante (FEHM *et al.*, 2008; AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Sua prevalência em 12 meses se situa entre 0,5 e 2%, sendo o sexo feminino mais afetado, com uma razão de chances (*Odds Ratio*, OR) de 1,5 a 2,2 (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Costuma apresentar altas taxas de comorbidades com outros transtornos mentais, como depressão e outros transtorno de ansiedade (FEHM *et al.*, 2008).

Exemplos de situações sociais incluem reuniões, contato com pessoas desconhecidas, apresentações orais, alimentar-se em público ou falar em público (SADOCK; SADOCK; RUIZ, 2017). O indivíduo com TAS age de forma a evitar tais situações sociais ou as suporta com intenso medo ou ansiedade, de forma que o sofrimento e/ou os prejuízos decorrentes dessa

condição interferem substancialmente no funcionamento social, profissional ou em outras áreas importantes de sua vida (AMERICAN PSYCHIATRIC ASSOCIATION, 2013).

O TAS está associado a comprometimentos que vão além da esfera social. Portadores desse transtorno relatam pior qualidade de vida em vários aspectos, como saúde mental, queixas clínicas e satisfação reduzida nos domínios laboral, financeiro e familiar (FEHM *et al.*, 2008). Essa redução da qualidade de vida se torna ainda mais acentuada no caso de comorbidades com outros transtornos mentais, sendo observado um aumento do *burden* adicional à medida que aumenta o número de transtornos comórbidos (WATSON; SWAN; NATHAN, 2011).

O transtorno de estresse pós-traumático (TEPT) é caracterizado por aumento do estresse ou ansiedade após exposição a evento traumático, seguido pelo surgimento de sintomas relacionados, como sofrimento psicológico intenso, reações de evitação a estímulos relacionados ao trauma, sonhos angustiantes, lembranças intrusivas, hipervigilância e *flashbacks* (AMERICAN PSYCHIATRIC ASSOCIATION, 2013; SADOCK; SADOCK; RUIZ, 2017). Possui prevalência de 8,3% durante a vida e de 4,7% em 12 meses, sendo mais comum no sexo feminino (KILPATRICK *et al.*, 2013). Atualmente classificado no grupo de transtornos relacionados a trauma e estressores, o TEPT era anteriormente considerado um transtorno de ansiedade por guardar uma relação sintomatológica íntima com esses transtornos (AMERICAN PSYCHIATRIC ASSOCIATION, 2000, 2013).

O TEPT é um transtorno mental grave associado a grande comprometimento funcional, cronicidade, redução da qualidade de vida, problemas de saúde física, dificuldades interpessoais e mortalidade em excesso (KESSLER, 2000; KIEFER *et al.*, 2020). A grande maioria dos pacientes com TEPT apresenta alguma comorbidade psiquiátrica que pode, inclusive, precedê-lo, como os transtornos de humor ou de ansiedade, em particular o transtorno de pânico (PERKONIGG *et al.*, 2000). Além disso, o TEPT também pode ser um importante fator de risco para o desenvolvimento de diversos transtornos mentais, principalmente TAG, agorafobia, transtornos depressivos e somatoformes e transtorno por uso de substâncias (PERKONIGG *et al.*, 2000), assim como também está associado a alto risco de tentativas de suicídio (KESSLER, 2000).

Do ponto de vista da produção vocal, existe uma relação estreita entre estresse e a vocalização. A fonação é um processo psicofisiológico composto pela integração entre o sistema nervoso central, o sistema nervoso periférico, os sistemas cardiorrespiratório e musculoesquelético. Envolve, nesse processo, diversas estruturas cerebrais corticais e subcorticais, nervos cranianos e espinhais e dezenas de músculos (VAN PUYVELDE *et al.*, 2018). Além de ser uma ferramenta para a comunicação, a voz também integra o aparato

humano de resposta ao estresse, sofrendo uma complexa regulação pelos sistemas nervosos simpático (SNS) e parassimpático (SNP) (VAN PUYVELDE *et al.*, 2018). Portanto, as mudanças no corpo decorrentes da reação ao estresse tendem a provocar alterações na função vocal, com consequente influência sobre o comportamento dos parâmetros acústicos (HOLMQVIST *et al.*, 2013).

O estresse é um termo genérico com várias definições. Do ponto de vista biológico, refere-se a um fenômeno aversivo que tende a desequilibrar a homeostase do organismo, provocando consequências adaptativas fisiológicas, emocionais, cognitivas e comportamentais (GIDDENS *et al.*, 2013). O estresse ainda pode ser definido como uma resposta não específica do corpo a qualquer demanda, a qual pode ser física e/ou mental, desencadeada por circunstâncias ambientais tanto internas quanto externas, como frio, calor, dor, isolamento (VAN PUYVELDE *et al.*, 2018). O SNS é o componente do SNA primariamente envolvido na resposta ao estresse, a qual provoca alterações cardiológicas, autonômicas, neuroendócrinas, imunológicas e psicológicas (HOLMQVIST *et al.*, 2013), com aumento da frequência cardíaca, da pressão arterial e da condutância da pele, liberação de cortisol, broncodilatação, entre outras respostas. Portanto, observa-se que a ativação simpática se traduz tanto em respostas centrais quanto periféricas (GIDDENS *et al.*, 2013).

Diversos estudos mostram que a ansiedade está relacionada ao aumento da tensão muscular (HOLMQVIST *et al.*, 2013). Na perspectiva da fonação, a hiperatividade do sistema nervoso simpático em resposta ao estresse tende a provocar aumento da tensão muscular, com pobre regulação da atividade muscular laríngea e consequentes alterações nos padrões vocais (ANDREA *et al.*, 2017). Outras consequências da ativação simpática são o aumento da frequência cardíaca e a broncodilatação. Devido à sua influência sobre a pressão subglótica, a frequência cardíaca é responsável, em parte, por perturbações da F0 (*jitter*) e da intensidade vocal (*shimmer*), sendo que a F0 aumenta linearmente com o aumento da pressão subglótica (GIDDENS *et al.*, 2013). Além disso, para Özseven *et al.* (2018), o aumento da pressão subglótica causado por estados ansiosos provocaria diminuição da vocalização de vogais. De maneira complementar, Almeida, Behlau e Leite (2011) descrevem que outras anormalidades nos atributos vocais estariam diretamente relacionados à gravidade dos sintomas ansiosos, como desequilíbrio da ressonância vocal, prejuízo na modulação e na articulação e alterações de expressões faciais.

Com o intuito de complementar e sintetizar o conteúdo desta seção, o Quadro 2 abaixo relaciona as principais características comunicativas e disfonias observadas nos transtornos mentais abordados nesta revisão.

Quadro 2 – Características da comunicação oral dos transtornos mentais abordados neste trabalho

Transtorno	Aspectos fonoarticulatórios
Depressão	Voz grave, qualidade fluida ou soprosa, às vezes rouca e basal; modulação restrita, monotom, entonação descendente, hipofonia, falta de volume e projeção; velocidade lenta, com pausas e projeções; demora de mudança nos turnos de falantes.
Transtorno Bipolar (mania/hipomania)	Voz clara e viva; ênfase marcada; modulações amplas; pausas rítmicas, longas e interpretativas, ressonância oral ou faríngea, com articulação marcada e vigorosa; sintaxe rica, mudanças imediatas nos turnos de falantes.
Esquizofrenia	Ajustes vocais hipocinéticos ou hipercinéticos; qualidade vocal infantilizada, com registro de cabeça, risos frequentes, grunhidos, gargalhadas e grimaças; mutismo, ecolalia e ecopraxia. Pode haver também voz rouca, monótona, qualidade destimbrada e hipofonia; alterações articulatórias complexas com nasalidade, imprecisão, distorções e substituições de sons; rupturas no discurso, disfluência; fala desorganizada, incoerente, jargão incompreensível, linguagem confusa, descarrilamento, neologismos ou uso diferente de palavras conhecidas, preseverações.
Transtorno de Ansiedade Generalizada	Voz aguda; intensidade e velocidade da fala elevadas; pode haver síndrome de tensão muscular e movimentos paradoxais das pregas vocais

Fonte: Adaptado de Behlau (2001, p. 81).

2.3 FERRAMENTAS

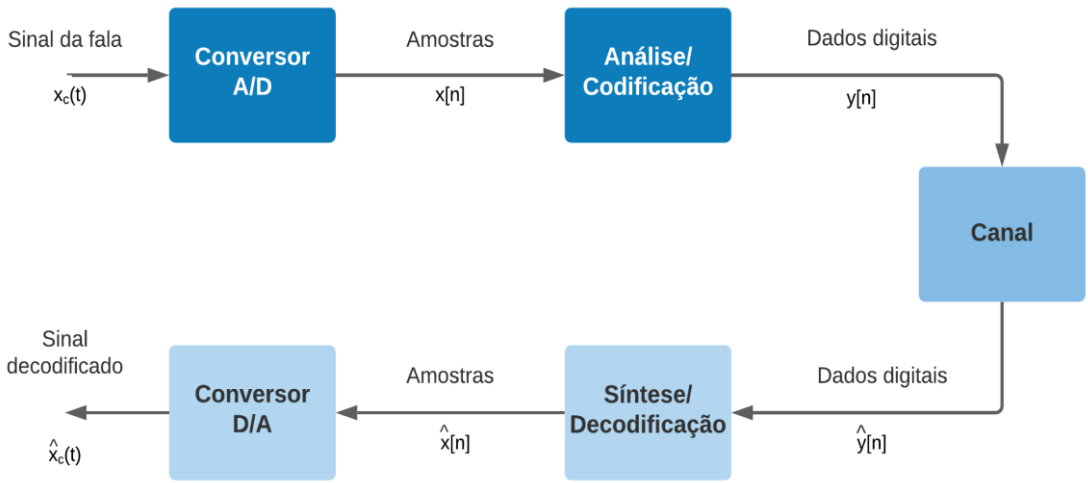
Esta seção apresenta uma revisão de conceitos referentes às ferramentas utilizadas para a solução do problema de pesquisa, o qual consiste na detecção dos quatro transtornos mentais supracitados por meio de atributos acústicos e modelos de aprendizado de máquina para classificação de padrões. Serão apresentados conceitos referentes ao processamento digital dos sinais acústicos da fala e de técnicas de inteligência artificial, mais especificamente sobre diferentes modelos de aprendizado de máquina para classificação.

2.3.1 Processamento digital de sinais acústicos da fala

O processamento digital dos sinais da fala é composto por diversas operações, destacando-se entre estas a extração dos atributos que servirão de base para o processo posterior de classificação (HIBARE; VIBHUTE, 2014). A primeira etapa desse processo envolve a conversão da onda acústica analógica em sinais digitais ou conversão analógico-digital (A/D), composta pelas subetapas de filtragem, amostragem, quantização e codificação (RABINER; SCHAFER, 2007; BAERT; THEUNISSEN; VERGULT, 2013). Em determinadas situações, (e.g., reprodução de músicas), um processo inverso realiza a decodificação do sinal digital armazenado de volta a um sinal analógico (conversão D/A), obtendo-se novamente a onda acústica original (BAERT; THEUNISSEN; VERGULT, 2013; LONLA; MBIHI; NNEME, 2017). O decodificador que realiza a conversão D/A é, por vezes, chamado de sintetizador, uma vez que este reconstitui a onda analógica da fala a partir de dados que podem não possuir nenhuma relação direta com a onda original (RABINER; SCHAFER, 2007).

Neste trabalho, os sinais digitalizados da fala serão submetidos às operações de extração de atributos acústicos e classificação por algoritmos computacionais. Portanto, as etapas de decodificação e a conversão D/A não serão utilizadas nem detalhadas aqui. Abaixo a Figura 1 exibe um sistema básico de processamento digital de sinais.

Figura 1 – Diagrama de fluxo de um sistema básico de conversão de sinais



A Figura 1 ilustra um diagrama de fluxo de um sistema de conversão A/D, onde a parte superior à esquerda exibe um conversor A/D, que converte o sinal analógico da fala $x_c(t)$ em uma representação amostrada com valores discretos x[n]. Algoritmos computacionais analisam esse sinal e produzem um novo sinal digital y[n], o qual pode

ser armazenado (ŷ[n]) ou transmitido por um canal de comunicação digital. A depender da aplicação, um processo de análise inversa decodifica o sinal ŷ[n], gerando uma sequência de amostras digitais (x̂[n]) que, por sua vez, são convertidas de volta ao sinal analógico original x̂c(t) (conversão D/A), tornando-os audíveis para humanos. Obs.: neste trabalho as etapas de decodificação e conversão D/A não são necessárias, pois nele o sinal acústico é analisado em seu formato digital. **Fonte:** Adaptado de Rabiner e Schafer (2007, p. 21).

Após o processo de amostragem (detalhado a seguir), os sinais da fala podem ser manipulados de variadas formas por diferentes técnicas de processamento digital de sinais, gerando diversas aplicações, como transmissão e armazenamento digital de dados, elaboração sintética da fala, aprimoramento da qualidade do sinal da fala e uso em métodos assistivos (RABINER; SCHAFER, 2007; HIBARE; VIBHUTE, 2014). Outras importantes aplicações, já citadas na seção anterior, incluem o reconhecimento automático da fala (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010), a identificação de falantes (HASAN *et al.*, 2004; MATĚJKA *et al.*, 2016; NAGRANI; CHUNG; ZISSERMAN, 2017) e o reconhecimento de emoções (BEDOYA-JARAMILLO *et al.*, 2012; MEUDT; SCHWENKER, 2012; PORIA; CHATURVEDI; CAMBRIA, 2016).

2.3.1.1 Aquisição e digitalização dos sinais vocais

O primeiro passo para um bom processamento dos sinais da fala consiste na instrumentação, com escolha de um gravador adequado para a aquisição da fala (BEHRMAN, 2018). Esta etapa se inicia com a captura dos sons analógicos da fala por um ou mais microfones, que são sensores que realizam a transdução da onda acústica em sinais elétricos (BEHRMAN, 2018). Em seguida, o processo de filtragem pode ser utilizado para modelar o espectro do sinal, enfatizando determinadas frequências e reduzindo ruídos (PUPIN, 2011; NATH, 2012). Esses ruídos podem ter sido gerados durante a aquisição do sinal, durante o processo de quantização e digitalização ou, ainda, por problemas na transmissão (PUPIN, 2011).

Com base nos componentes de frequência do sinal que são mantidos ou atenuados (BEHRMAN, 2018), os filtros ideais são classificados em cinco grupos: (1) passa-baixa, (2) passa-alta, (3) passa-banda, (4) rejeita-banda, e (5) passa-tudo (NATH, 2012). Seus nomes se referem às propriedades seletivas de cada filtro para determinadas frequências (NATH, 2012). Logo, um filtro passa-baixa permite a passagem de frequências abaixo de sua frequência de corte e atenua frequências acima desta, enquanto o filtro passa-alta faz o oposto (PACTITIS, 2018). Já um filtro passa-banda permite a passagem de um intervalo de frequência, eliminando os demais valores (PUPIN, 2011). Filtros rejeita-banda, por sua vez, rejeitam frequências dentro

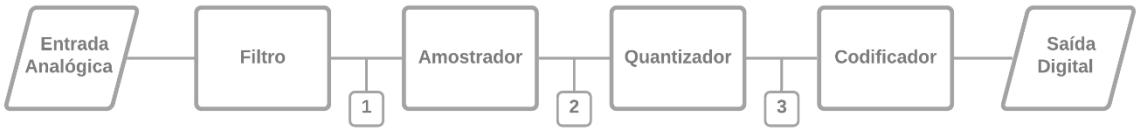
de determinada banda e permitem a passagem de componentes fora desta (PACTITIS, 2018), sendo projetados para eliminar interferências indesejadas (FATHELBAB, 2011). Por último, os filtros passa-tudo permitem a passagem de todo o espectro de frequências, porém produzem defasagens ao longo do espectro sem variar a amplitude do sinal de entrada (METIN; PAL; CICEKOGLU, 2011; NATH, 2012).

Em qualquer aplicação de processamento digital de sinais, a etapa seguinte envolve a amostragem do sinal analógico e a quantização das amostras obtidas em formato digital, processo conhecido como conversão A/D (RABINER; SCHAFER, 2007). Essa representação digital dos sinais da voz é muito relevante, pois permite sua análise com todo o poder computacional atual (KENT; READ, 2015). Dessa forma, a conversão de sinais de analógico para digital (e vice-versa) constitui uma etapa crítica do processamento digital de sinais de áudio (BAERT; THEUNISSEN; VERGULT, 2013).

O Teorema da Amostragem estabelece que, sob certas condições, um sinal de tempo contínuo pode ser exatamente representado por um sinal de tempo discreto, produzindo amostras uniformemente espaçadas no tempo, desde que a taxa de amostragem seja suficientemente grande (OPPENHEIM; WILLSKY, 2010; PUPIN, 2011). Para tal operação, técnicas de *framing* são aplicadas de forma que um sinal de áudio contínuo no tempo é enquadrado em N amostras (SAINI; MEHRA, 2015). Após a conversão A/D, obtém-se uma sequência de números do tipo $x[n] = x_c(nT)$, onde T é o período de amostragem; a frequência de amostragem será, portanto, $f_s = 1/T$ (RABINER; SCHAFER, 2007). O sinal analógico original limitado em banda será representado, portanto, unicamente por suas amostras (OPPENHEIM; WILLSKY, 2010).

Em seguida, a quantização realiza a conversão de um sinal discreto com valores contínuos em um sinal com valores discretos e binários (BAERT; THEUNISSEN; VERGULT, 2013). Por exemplo, um quantizador com resolução de bits igual a $\boldsymbol{B}$, apresentará $\mathbf{2}^{\boldsymbol{B}}$ níveis de resolução do sinal (RABINER; SCHAFER, 2007). Essa representação do sinal de áudio por amostras quantizadas binárias recebe o nome de modulação por código de pulso (*Pulse Code Modulation*, PCM), uma vez que números binários podem ser utilizados em transmissões de amplitude de pulso do tipo *on/off* (RABINER; SCHAFER, 2007). Um sistema hipotético de conversão A/D está esquematizado a seguir na Figura 2.

Figura 2 – Diagrama de blocos de um sistema de conversão analógico-digital (A/D)



A Figura 2 exibe um diagrama de blocos de um sistema básico de conversão A/D, no qual um sinal analógico capturado em sua entrada de dados e submetido à filtragem de frequências indesejadas e, em seguida, aos processos de amostragem, quantização e codificação do sinal, até a obtenção de um sinal digital na saída de dados. Os números representam o sinal de áudio em suas diferentes configurações, onde: (1) sinal analógico (contínuo no tempo, valores contínuos); (2) sinal amostrado (discreto no tempo, valores contínuos); e (3) sinal digital (discreto no tempo, valores discretos). **Fonte:** Baert, Theunissen e Vergult (2013, p. 28, tradução nossa).

De acordo com o teorema de Nyquist-Shannon, a frequência de amostragem para a aquisição de dados deve ser, no mínimo, maior que o dobro da frequência mais alta presente no sinal analógico original, sendo comumente conhecida como taxa de Nyquist (OPPENHEIM; WILLSKY, 2010; BEHRMAN, 2018). Caso um sinal original contenha frequências superiores à taxa de Nyquist, o efeito de *aliasing* pode potencialmente ocorrer. Decorrente de subamostragem, este fenômeno resulta em uma representação digital de frequência menor que o sinal original, causando distorção e perda irreversível de informação da onda acústica original (NASIRI; WANG, 2017; BEHRMAN, 2018). O efeito de *aliasing* pode, portanto, afetar gravemente a qualidade do áudio por corromper os dados digitalmente representados (ESQUEDA; BILBAO; VÄLIMÄKI, 2016).

Como exemplo da taxa de Nyquist, para representar adequadamente um sinal acústico com até 10 KHz, é necessária uma taxa de amostragem pelo menos acima de 20 KHz; enquanto isso, um sinal de 16 KHz exige uma frequência de amostragem mínima maior que 32 KHz, e assim sucessivamente (BEHRMAN, 2018). A frequência padrão utilizada em discos compactos (*Compact Discs*, CDs) é de 44,1 KHz, podendo ser utilizada para a digitalização de sinais cuja frequência máxima é de 22.050 Hz. Como a frequência dos sinais acústicos da voz se estendem tipicamente até, no máximo, 20 KHz, a frequência de CD pode ser considerada adequada para a representação desses sinais (BAERT; THEUNISSEN; VERGULT, 2013).

2.3.1.2 Extração de atributos acústicos

A extração de atributos acústicos da fala corresponde à obtenção de informações relevantes do sinal acústico a fim de obter uma descrição mais compacta e representativa, facilitando o processamento computacional de dados (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Em problemas de classificação como o da pesquisa atual, a extração

de atributos acústicos tem como objetivo reduzir a dimensionalidade do vetor de dados de entrada sem perder as características mais discriminantes do sinal (GAIKWAD; GAWALI; YANNAWAR, 2010). Em outras palavras, a extração de atributos busca uma transformação ótima dos dados acústicos em um vetor de atributos que será utilizado como entrada para um modelo de aprendizado de máquina para a classificação de determinado áudio (STORCHEUS; ROSTAMIZADEH; KUMAR, 2015).

A seleção das técnicas de extração de atributos acústicos é considerada determinante para o processamento de sinais acústicos da fala, uma vez que exerce grande influência sobre a acurácia de um sistema de classificação (HIBARE; VIBHUTE, 2014). Exemplos de importantes técnicas de extração de atributos são a transformada rápida de Fourier (*Fast Fourier Transform*, FFT), a Transformada Discreta de Wavelet (*Discrete Wavelet Transform*, DWT), a codificação preditiva linear (*Linear Predictive Coding*, LPC) e os MFCCs (HIBARE; VIBHUTE, 2014).

As transformadas são funções matemáticas que convertem valores númericos de um domínio para outro (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Por exemplo, a Transformada de Fourier converte os dados do domínio do tempo no domínio espectral, revelando a distribuição de frequências de determinado sinal (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; PUPIN, 2011). Esta função matemática possui inúmeras aplicações que abrangem desde as telecomunicações e o reconhecimento de fala à meteorologia e à arqueologia (HIBARE; VIBHUTE, 2014).

Similarmente, a transformada de Wavelet também fornece uma representação do tempo-frequência, porém com maior resolução que as Transformadas de Fourier (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). A Transformada Discreta do Cosseno, por sua vez, é principalmente utilizada para a conversão do domínio espectral para o cepstral, sendo, portanto, empregada para a obtenção dos MFCCs (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; HIBARE; VIBHUTE, 2014). Como o processamento digital de sinais exige que seu domínio seja reduzido a um determinado intervalo (i.e., discretizado), para essa aplicação é necessária a utilização das versões discretas das transformadas em tempo contínuo, como a Transformada Discreta de Fourier (DFT) e a DWT (PUPIN, 2011; HIBARE; VIBHUTE, 2014).

Devido às propriedades de duração finita das transformadas discretas, antes de sua utilização é necessário aplicar uma função janela ao sinal acústico (KINNUNEN; LI, 2010). Esta operação consiste na multiplicação dos valores em um intervalo definido pelo peso da função janela, resultando em valores diferentes de zero dentro de determinado intervalo, ao

passo que todos os valores fora desse intervalo são iguais a zero (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; SAINI; MEHRA, 2015). A função janela é também importante para minimizar descontinuidades no sinal acústico enquadrado, assim como para calcular a largura do lobo central da janela, onde estará a maior parte do sinal janelado (SAINI; MEHRA, 2015). Além disso, viabiliza a aplicação de filtros digitais de resposta finita (FIR) e permite análises de desempenho espectral (APARNA; CHITHRA, 2017).

As funções de janelamento podem ser divididas em duas categorias: fixas e ajustáveis. No processamento digital dos sinais da fala, as janelas fixas mais utilizadas são as janelas Hamming, Hann, Blackman e a janela retangular, sendo esta a mais simples de todas as funções janelas (SAINI; MEHRA, 2015; APARNA; CHITHRA, 2017). Entre as janelas ajustáveis, cita-se como exemplo a janela Kaiser (APARNA; CHITHRA, 2017). Na literatura existem divergências quanto à importância do tipo de janela para o processamento digital de sinais. Enquanto alguns autores afirmam que essa escolha, na prática, não é considerada crítica para o resultado do processamento (KINNUNEN; LI, 2010), outros defendem que determinados formatos de janela seriam mais adequados para o processamento digital de sinais acústicos da fala (PODDER *et al.*, 2014) e de outros sinais (DATAR; JAIN; SHARMA, 2009; RAJPUT; BHADAURIA, 2012), e para o desenvolvimento de filtros (CHAKRABORTY, 2013).

Uma técnica fundamental para a extração de atributos acústicos é a Transformada Rápida de Fourier (FFT) (LIU *et al.*, 2019), que consiste em uma implementação mais rápida e em tempo real da DFT, tornando mais eficiente o processamento computacional dos sinais (HIBARE; VIBHUTE, 2014; LIU *et al.*, 2019). Mais recentemente, a DWT tem sido considerada bastante atraente para o processamento digital de sinais vocais, uma vez que seu processamento de áudio se assemelha ao de um ouvido humano, além de ser adequada para processar sinais não estacionários como o da fala (HIBARE; VIBHUTE, 2014). Desenvolvida para superar as limitações da Transformada de Fourier, a DWT fornece uma representação simultânea do sinal nos domínios do tempo e da frequência, proporcionando, desta forma, uma análise multiescala e multirresolucional dos sinais (HIBARE; VIBHUTE, 2014; YADAV; JAIN; BHARGAV, 2015). Entretanto, apesar das vantagens da DWT, a DFT ainda é considerada a transformada discreta mais importante para o processamento de sinais (APARNA; CHITHRA, 2017).

Outro método popular para a análise de componentes espectrais de um sinal é a LPC. Essa técnica consiste na premissa básica de que, em uma série temporal, cada amostra pode ser aproximada por uma combinação linear de amostras precedentes (SPRATLING, 2017). No processamento de sinais da fala, a LPC é utilizada para estimar parâmetros acústicos básicos,

como a identificação de formantes e a função de transferência do trato vocal (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010; BEHRMAN, 2018). Outras aplicações abrangem a interpolação e restauração de sinais e a redução de ruídos (SPRATLING, 2017). Por exemplo, a LPC é o método de escolha para analisar amostras de voz digitalizadas a baixos *bitrates* com maior qualidade que outras técnicas (HIBARE; VIBHUTE, 2014). Também tem sido empregada com sucesso em sistemas biométricos de identificação de pessoas pela voz (HIBARE; VIBHUTE, 2014).

Por último, os MFCCs são outra abordagem amplamente utilizada para a extração de atributos acústicos vocais. Suas propriedades já foram discutidas na seção anterior, na subseção referente aos parâmetros cepstrais.

Os atributos vocais, uma vez extraídos do sinal acústico digitalizado, estão prontos para serem utilizados como dados de entrada em sistemas de classificação. Nesta pesquisa, foram adotadas técnicas de classificação de padrões baseadas em algoritmos computacionais de aprendizado de máquina, definidos a seguir.

2.3.2 Modelos de aprendizado de máquina

Considera-se que o termo inteligência artificial (IA) foi utilizado pela primeira vez em 1956 (BRUNETTE; FLEMMER; FLEMMER, 2009). A IA pode ser definida como a compreensão e a construção de agentes inteligentes que empregam uma abordagem racional e lógica para a resolução de problemas, a fim de alcançar o melhor resultado possível (RUSSELL; NORVIG, 2016). Embora não seja um conceito recente, neste século a IA se tornou um dos mais importantes campos de pesquisa em diversas áreas, como engenharia, ciência, medicina e economia (OKE, 2008).

A IA é considerada uma área universal e relevante a qualquer tarefa intelectual (RUSSELL; NORVIG, 2016). Dessa forma, possui um número crescente de aplicações que envolvem desde carros autônomos ao diagnóstico de doenças (HENGSTLER; ENKEL; DUELLI, 2016; NICHOLS; HERBERT CHAN; BAKER, 2019). Nesse contexto, observa-se que a IA é uma área abrangente, integrada por diversas subáreas, como aprendizado de máquina, robótica, visão computacional e processamento de linguagem natural (RUSSELL; NORVIG, 2016).

De fundamental relevância para esta pesquisa, o aprendizado de máquina ou *machine learning* (ML) é um ramo da IA dedicado a fornecer aos computadores a habilidade de aprendizado sem instruções explícitas (BZDOK; MEYER-LINDENBERG, 2018), de

adaptação a novas circunstâncias e de detecção e extrapolação de padrões (RUSSELL; NORVIG, 2016). Pretende, em última análise, capacitar os computadores a modificar ou adaptar suas ações de forma a aumentar sua acurácia (MARSLAND, 2015). Para alcançar esse objetivo, o campo de ML utiliza conhecimentos multidisciplinares provenientes das neurociências, biologia, física, estatística e matemática (MARSLAND, 2015).

No contexto de ML, conceitua-se o aprendizado como a capacidade de um agente melhorar sua performance em tarefas futuras a partir observações sobre o problema (RUSSELL; NORVIG, 2016). De acordo com o tipo de aprendizado utilizado para a resolução de problemas, os algoritmos de ML podem ser divididos em três subgrupos principais:

- Aprendizado supervisionado: envolve o desenvolvimento de algoritmos capazes de identificar padrões de generalização sobre determinado domínio, a partir do fornecimento de instâncias externas com as respostas corretas. Com base nesses dados prévios, o algoritmo busca realizar predições corretas para futuras instâncias (SINGH; THAKUR; SHARMA, 2016; OSISANWO *et al.*, 2017; NICHOLS; HERBERT CHAN; BAKER, 2019).
- Aprendizado não supervisionado: neste as respostas corretas não são fornecidas. Mesmo sem nenhum feedback externo, o algoritmo busca aprender semelhanças entre os dados de entrada; aqueles com similaridades são agrupados em uma mesma categoria (MARSLAND, 2015; RUSSELL; NORVIG, 2016).
- Aprendizado por reforço: situa-se entre o aprendizado supervisionado e o não supervisionado. O algoritmo é informado se uma resposta é incorreta, mas não como corrigi-la. Dessa forma, ele precisa explorar diferentes possibilidades até encontrar a maneira de alcançar a resposta correta (MARSLAND, 2015).

De acordo com a natureza da resposta procurada (ou variável dependente), os métodos de aprendizado supervisionado podem ser utilizados para problemas de classificação e de regressão (HAYKIN, 2009; RUSSELL; NORVIG, 2016). A regressão envolve a predição de uma variável numérica contínua, como peso, altura e temperatura (NICHOLS; HERBERT CHAN; BAKER, 2019). A classificação, por outro lado, é a predição de uma variável qualitativa (i.e., discreta); por exemplo, definir se a imagem de determinado animal corresponde a um gato ou um cachorro (MARSLAND, 2015; NICHOLS; HERBERT CHAN; BAKER, 2019). Dessa forma, técnicas de ML para classificação podem ser aplicadas em diversos problemas, como classificação de textos e imagens, detecção de fraudes, filtragem de spam etc. (SINGH; THAKUR; SHARMA, 2016). O problema desta pesquisa envolve a detecção de

determinados transtornos mentais, ou seja, de diagnóstico. Estamos, portanto, diante de um problema de classificação.

Os algoritmos supervisionados de ML para classificação têm o objetivo de categorizar dados a partir de informações prévias (SINGH; THAKUR; SHARMA, 2016). Consistem em uma das técnicas mais aplicadas e bem estudadas para a resolução de problemas por sistemas inteligentes (OSISANWO *et al.*, 2017). Devido a isso esses algoritmos foram selecionados na elaboração da ferramenta de resolução do problema desta pesquisa. Os outros tipos de aprendizado fogem do escopo desta revisão e não serão abordados.

O processo para a resolução de problemas utilizando ML é composto por diversas etapas. A primeira delas envolve a coleta de dados para construção de um conjunto de dados (*dataset*), o qual, nesta pesquisa, corresponde a arquivos de áudio contendo diálogos provenientes de consultas psiquiátricas. Devido à escolha por algoritmos de aprendizado supervisionado para esta pesquisa, os dados coletados precisam estar devidamente rotulados para posterior treinamento dos algoritmos de ML. Nesse caso, a cada participante foi atribuído seu diagnóstico correto fornecido por um especialista.

Após a coleta, os dados precisam ser preparados para a etapa seguinte, de extração e seleção de atributos. Nesse momento, é importante certificar-se de que a quantidade de dados coletados é suficientes para o problema em questão, e se os dados estão limpos, ou seja, sem erros, ruídos excessivos ou dados faltantes (DUDA; HART; STORK, 2001; HAYKIN, 2009; MARSLAND, 2015).

A etapa de extração de atributos foi abordada na seção anterior. A depender da quantidade de parâmetros extraídos e do custo computacional envolvido, pode ser necessária a etapa de seleção de atributos, que visa simplificar os modelos computacionais por meio da redução de sua dimensionalidade (MARSLAND, 2015; RUSSELL; NORVIG, 2016). Como seu nome sugere, esta etapa consiste na identificação e seleção dos atributos mais representativos para o problema avaliado, descartando aqueles que parecem irrelevantes (MARSLAND, 2015; RUSSELL; NORVIG, 2016). Ambas as etapas de extração e seleção de atributos invariavelmente exigem conhecimento prévio sobre o problema e sobre a base de dados (DUDA; HART; STORK, 2001; MARSLAND, 2015).

A etapa seguinte envolve a seleção de um ou mais algoritmos apropriados para a modelagem do problema a ser resolvido, seguida pelo ajuste (*tuning*) de seus parâmetros, o qual frequentemente é realizado manualmente ou por experimentação para a identificação dos valores mais apropriados (MARSLAND, 2015). O ajuste dos parâmetros de um algoritmo é crucial para que este atinja a melhor performance possível e pode fazer a diferença entre um

desempenho medíocre e um excepcional (HAYKIN, 2009; HUTTER; SCHMIDT-THIEME; LÜCKE, 2015).

Em problemas de classificação, é importante estimar o desempenho de um classificador por meio de sua verdadeira taxa de erros (KIM, 2009). Para isso, o processo de classificação precisa ser realizado em duas fases: treinamento e teste, exigindo a separação dos exemplos da base de dados em dois conjuntos (NICHOLS; HERBERT CHAN; BAKER, 2019). Essa segmentação do *dataset* é fundamental, pois uma maneira confiável de avaliar a performance de um classificador é testar suas predições em dados novos, ausentes no conjunto de treinamento (VABALAS *et al.*, 2019). Na fase de treinamento, utiliza-se um algoritmo de classificação no conjunto de dados de treinamento, gerando um modelo que busca generalizar sua predição para dados desconhecidos (MARSLAND, 2015; SINGH; THAKUR; SHARMA, 2016). Na fase de teste, o modelo construído é utilizado na base de dados de teste rotulada para, então, ser validado de acordo com seu desempenho (SINGH; THAKUR; SHARMA, 2016).

Os métodos de validação de modelos de ML também são essenciais para avaliar e mitigar a possibilidade de *overfitting* (VABALAS *et al.*, 2019). Esse fenômeno acontece quando se tenta modelar muito bem um conjunto de treinamento, fazendo com que o algoritmo memorize a base de dados de treinamento, aprendendo os ruídos nela presentes, em vez de desenvolver regras gerais de predição (DIETTERICH, 1995; MARSLAND, 2015). Sua probabilidade de ocorrência aumenta quanto maior for o número de atributos e diminui com o aumento do conjunto de treinamento (RUSSELL; NORVIG, 2016). O o*verfitting* é considerado um problema, pois compromete a capacidade de generalização de um algoritmo para dados novos (HAYKIN, 2009; VABALAS *et al.*, 2019). Diante disso, a validação do desempenho de um modelo de ML assume uma importância ainda maior em *datasets* pequenos (VABALAS *et al.*, 2019).

O fenômeno de *underfitting*, oposto ao *overfitting*, ocorre quando se constroem modelos de ML excessivamente simples, que não conseguem capturar adequadamente os padrões presentes em uma base de dados. Entretanto, ao passo que o *overfitting* é um problema de difícil abordagem, o *underfitting* pode ser facilmente solucionado por meio da utilização de modelos de aprendizado computacional de maior complexidade (VABALAS *et al.*, 2019).

Existem diferentes técnicas de validação de um classificador de ML, podendo-se citar entre os mais utilizados o método *hold-out* e a validação cruzada. Na validação por *hold-out*, também denominada *train-test split*, uma parte da base de dados é aleatoriamente separada somente para validação, correspondendo geralmente a um terço da base de dados. O classificador é, então, construído apenas com os dados restantes (dois terços da base), e sua

acurácia será estimada pelo seu desempenho de classificação no conjunto de testes (KIM, 2009; VABALAS *et al.*, 2019).

A validação cruzada (VC) é uma abordagem tradicional em bases de dados pequenas, pois, ao contrário da técnica anterior, não reserva uma porção significativa da base de dados somente para o teste, permitindo que a base de dados seja inteiramente disponibilizada para o treinamento do modelo (KIM, 2009; VABALAS *et al.*, 2019). Entre as técnicas de VC, o método aleatório é bastante comum e consiste na segmentação aleatória da base de dados, destinando-se comumente 70% para o treinamento do modelo e 30% para os testes. Para configurações de validação cruzada do tipo *k-folds*, esse processo de segmentação, treino e teste é repetido *k* vezes (com *k* variando entre cinco e dez), sendo a acurácia final do modelo a média ponderada das acurácias obtidas em cada repetição (FERDINANDY *et al.*, 2020). O método *k-fold* utiliza os dados de maneira econômica, permitindo que o mesmo exemplo seja utilizado ora para treinamento, ora para validação. Dessa forma, é adequado em bases de dados com tamanho muito limitado, frequentemente encontrados em problemas na área médica devido às dificuldades inerentes à construção de bases de dados com pacientes (VABALAS *et al.*, 2019).

A seguir será apresentada uma revisão sobre os modelos de aprendizado supervisionado selecionados para implementação na ferramenta de solução do problema desta pesquisa. Tal escolha residiu na relevância e na ampla utilização desses modelos pela literatura em problemas de classificação.

2.3.2.1 Redes Neurais Artificiais e Multilayer Perceptron

As redes neurais artificiais (RNAs) são modelos computacionais inspirados no funcionamento do cérebro humano durante tarefas de resolução de problemas (PAL; MITRA, 1992), sendo também conhecidas como conexionismo ou computação neural (RUSSELL; NORVIG, 2016). As RNAs utilizam funções lineares e não lineares para o aprendizado de padrões para, então, buscar uma generalização de hipóteses para dados desconhecidos (TAUD; MAS, 2017). Seu desenvolvimento se iniciou graças ao trabalho pioneiro de McCulloch e Pitts (1943), o qual introduziu o primeiro modelo matemático de um neurônio (PARK; LEK, 2016; RUSSELL; NORVIG, 2016), surgindo com este a ideia de redes neurais como algoritmos computacionais (HAYKIN, 2009).

Do ponto de vista estrutural, as RNAs são compostas por neurônios, também denominados unidades (ou nós), conectados entre si (RUSSELL; NORVIG, 2016). As ligações (*links*) entre os neurônios possuem um peso que determina a importância da conexão sináptica,

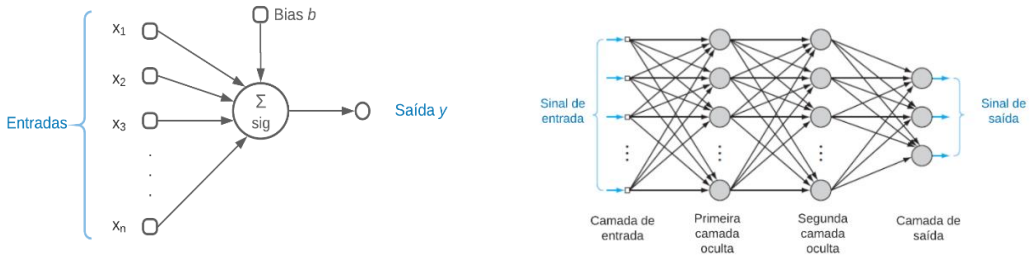
assim como são responsáveis pela propagação da ativação entre os nós (RUSSELL; NORVIG, 2016). O neurônio artificial consiste em uma unidade com limiar de ativação binária, a qual calcula a soma ponderada seus valores de entrada multiplicados por fatores conhecidos como pesos sinápticos (MARSLAND, 2015; RUSSELL; NORVIG, 2016). Caso o valor obtido ultrapasse o limiar de ativação, o neurônio dispara (i.e., é ativado), produzindo uma saída de valor igual a 1 (hum); caso contrário, o neurônio não dispara, resultando em uma saída de valor zero (MARSLAND, 2015). Esse processo de tomada de decisão quanto à ativação ou não de um neurônio em resposta a um valor de entrada é denominado função de ativação (MARSLAND, 2015). Apesar de sua inspiração biológica, as RNAs possuem poucas semelhanças com o cérebro humano, que é consideravelmente mais complexo (PARK; LEK, 2016). Entretanto, ambos compartilham duas características: (1) são redes altamente conectadas; e (2) as conexões entre os neurônios determinam a função de todo o circuito (PARK; LEK, 2016).

O segundo marco histórico na evolução das RNAs foi alcançado por Rosenblatt (1958), com o desenvolvimento do *perceptron*. Considerado o primeiro modelo de aprendizado supervisionado, o *perceptron* é a forma mais simples de rede neural, composta por um único neurônio com pesos e *bias* ajustáveis, utilizada para a classificação de padrões ditos linearmente separáveis (HAYKIN, 2009), ou seja, aqueles que podem ser separados por uma linha (TAUD; MAS, 2017). O *bias* mede o erro do modelo em comparação com o resultado considerado "verdadeiro" (NICHOLS; HERBERT CHAN; BAKER, 2019) e consiste em um valor externo aplicado à função de ativação (HAYKIN, 2009; MARSLAND, 2015).

O algoritmo *Multilayer Perceptron* (MLP) é o tipo mais popular de RNA (PARK; LEK, 2016). Enquanto o *perceptron* é basicamente uma rede neural de camada única, o MLP consiste em uma rede neural de aprendizado supervisionado composta por uma ou mais camadas ocultas altamente conectadas (HAYKIN, 2009; PARK; LEK, 2016; TAUD; MAS, 2017). Assim sendo, a estrutura de uma rede MLP pode ser dividida em três tipos de camadas de neurônios artificiais: entrada, camadas ocultas e saída (PARK; LEK, 2016).

A rede MLP é uma rede neural de tipo *feedforward*, na qual a informação flui de maneira unidirecional da camada entrada à de saída, passando pelas camadas intermediárias (TAUD; MAS, 2017). Em sua arquitetura, todos os neurônios de uma camada estão conectados aos neurônios das camadas adjacentes, e cada conexão tem seu próprio peso (PARK; LEK, 2016; TAUD; MAS, 2017). As estruturas de um *perceptron* e de uma rede MLP estão esquematizadas na Figura 3 abaixo:

Figura 3 – Esquema de grafos de arquiteturas de diferentes redes neurais artificiais



(a) (b)

Na Figura 3 estão esquematizados dois exemplos de diferentes arquiteturas de RNA, onde a imagem (a) exibe uma rede *perceptron* com *n* entradas e um único neurônio, com viés (*bias*) *b*. Nesta, o *perceptron* recebe os sinais de entrada, multiplica-os por seus pesos sinápticos e soma o valor do *bias*. Caso ultrapassem o limiar de ativação, uma função de ativação sigmoide é executada, gerando a saída *y*. **Fonte:** Elaborado pela autora (2021). A imagem (b) ilustra uma rede MLP composta por duas camadas ocultas com *n* sinais de entrada e três sinais de saída. Os neurônios da camada de entrada recebem o sinal de entrada e propagam o sinal de ativação para a primeira camada oculta através de suas conexões com os neurônios adjacentes. Esse sinal é então propagado para a segunda camada oculta até os neurônios da camada de saída, que fornecem a resultado final da rede MLP. **Fonte:** Adaptado de Haykin (2009).

Um método de treinamento bastante utilizado para as redes MLP é denominado *back-propagation* ou propagação retrógrada (TAUD; MAS, 2017). Neste, o treinamento se divide em duas etapas:

1. Na fase anterógrada (*forward propagation*), o sinal de entrada é propagado por cada camada através da rede até a camada de saída, que fornece o resultado computacional da rede. Um sinal de erro é, então, produzido por meio da comparação entre o resultado obtido e a resposta desejada (HAYKIN, 2009; PARK; LEK, 2016).

2. Na fase retrógada (*back propagation*), o erro obtido é transmitido retrogradamente (retropropagado) da camada de saída através das camadas intermediárias, com ajustes sucessivos dos pesos das conexões sinápticas e consequente geração de um novo resultado pela camada de saída (HAYKIN, 2009; PARK; LEK, 2016).

Devido à sua ampla utilização na literatura para soluções de reconhecimento de padrões, o algoritmo MLP com aprendizado em propagação retrógada foi escolhido entre os tipos de RNAs para os experimentos computacionais deste trabalho de pesquisa.

2.3.2.2 Máquinas de vetor de suporte

Desenvolvidas por Cortes e Vapnik (1995), as máquinas de vetor de suporte (*Support Vector Machines*, SVMs) formam uma família de algoritmos de aprendizado supervisionado

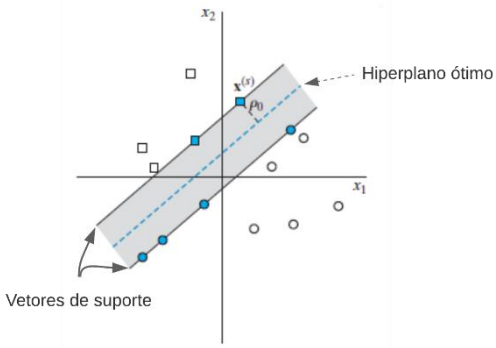
bastante popular (DATTA; DAS, 2015; RUSSELL; NORVIG, 2016). São compostas por redes *feedforward* propostas originalmente para a solução de problemas de classificação binários, que admitem apenas dois valores possíveis como resposta, em contextos de reconhecimento de padrões (SCHÖLKOPF; SMOLA, 2002; HAYKIN, 2009; DATTA; DAS, 2015; RUSSELL; NORVIG, 2016). Segundo Russell e Norvig (2016), o uso das SVMs também é indicado como método inicial em situações onde não há conhecimento especializado prévio sobre o problema.

Os modelos SVM são considerados a primeira aplicação da teoria do aprendizado estatístico (*Statistical Learning Theory*, SLT) e do método de minimização estrutural de riscos (*Structural Risk Minimization*, SRM) (HAYKIN, 2009; VAPNIK; IZMAILOV, 2019). A SLT permite a formulação estatística da teoria do aprendizado de máquina para maximizar a acurácia preditiva dos modelos de ML (CHAPMAN; WEISS; DUBERSTEIN, 2016), enquanto o SRM é um método de inferência estatística universalmente utilizado para a minimização do erro empírico (VAPNIK; IZMAILOV, 2019). As SVMs são úteis tanto em problemas de regressão não lineares quanto de classificação, contudo, em problemas complexos de classificação de padrões, esses algoritmos atingem sua maior relevância (HAYKIN, 2009).

Os classificadores de SVMs utilizam como princípio básico a construção de uma fronteira de decisão com a máxima margem de separação entre as classes, denominada hiperplano ótimo (SCHÖLKOPF; SMOLA, 2002; HAYKIN, 2009; DATTA; DAS, 2015). Por meio de funções matemáticas conhecidas como *kernels*, esses algoritmos projetam dados de treinamento não lineares em um espaço de maior dimensão, construindo um hiperplano linear com margem máxima de separação nesse novo espaço e equidistante às classes (HEARST *et al.*, 1998). Essas propriedades permitem que dados não linearmente separáveis em sua dimensão original sejam mais facilmente separados em um espaço com maior dimensão e contribuem para a grande capacidade de generalização desses algoritmos (SCHÖLKOPF; SMOLA, 2002; RUSSELL; NORVIG, 2016; NANDA *et al.*, 2018).

A construção do hiperplano de separação ótimo se baseia nos chamados vetores de suporte, que dão nome a essa classe de algoritmos. Esses vetores consistem em um subconjunto de pontos de treinamento localizados nos extremos das distribuições das classes, ou seja, mais próximos à fronteira de decisão (HAYKIN, 2009; MARSLAND, 2015; NANDA *et al.*, 2018). São esses pontos que determinarão o posicionamento do hiperplano de separação, enquanto as posições dos demais pontos são consideradas irrelevantes (SCHÖLKOPF; SMOLA, 2002; HAYKIN, 2009). Abaixo a Figura 4 fornece uma ilustração dos vetores de suporte e o hiperplano de separação ótimo.

Figura 4 – Representação de um hiperplano de separação ótimo em um padrão linearmente separável



Na Figura 4, um hiperplano de separação ótimo, indicado pela linha tracejada azul, é traçado sobre um padrão linearmente separável composto por duas classes de dados (circunferências e quadrados). Os exemplos mais extremos de cada classe (cor azul) formam os vetores de suporte, equidistantes ao hiperplano ótimo. **Fonte:** Adaptado de Haykin (2009, p.270).

Um dos desafios encontrados em problemas de classificação é a dispersão dos vetores de entrada no espaço original, dificultando sua separação linear (NANDA *et al.*, 2018). Para lidar com esse obstáculo, os algoritmos SVM desenvolveram as funções *kernel*, que transformam o espaço original dos dados em um espaço de maior dimensão (ou *feature space*) por meio de uma função de transformação não linear (SCHÖLKOPF; SMOLA, 2002; YU; KIM, 2012; NANDA *et al.*, 2018). Quando projetadas de volta ao espaço original, essas fronteiras lineares podem se tornar não lineares e até mesmo bastante sinuosas (RUSSELL; NORVIG, 2016). Também conhecida como *kernel trick*, essa operação possibilita que fronteiras de decisão lineares sejam eficientemente encontradas em espaços multidimensionais (RUSSELL; NORVIG, 2016), facilitando a separação dos dados em suas diferentes classes (SCHÖLKOPF; SMOLA, 2002; NANDA *et al.*, 2018).

Os algoritmos SVM dispõem de diferentes funções *kernel*, citando-se, entre as mais populares, os *kernels* linear, *Radial Basis Function* (RBF) e polinomial (DREWNIK; PASTERNAK-WINIARSKI, 2017). Outros exemplos são os *kernels* sigmoide (MARSLAND, 2015), normalizado (VAPNIK; IZMAILOV, 2019) e o Pearson *Universal Kernel* VII, PUK (KREMIC; SUBASI, 2016). Cada função *kernel* possui seus próprios parâmetros que precisam ser otimizados de forma a obter o melhor desempenho para cada problema em particular (NANDA *et al.*, 2018).

No contexto de reconhecimento de padrões de acústica vocal em transtornos neuropsiquiátricos, praticamente todos os *kernels* demonstraram utilidade, como os *kernels* RBF (JIANG *et al.*, 2017; LAHMIRI; SHMUEL, 2019), polinomial (BENBA; JILBAB;

HAMMOUCH, 2016), PUK (ESPINOLA *et al.*, 2020a) e linear (BENBA *et al.*, 2015; MCGINNIS *et al.*, 2019). Dessa maneira, nesta pesquisa serão utilizadas as diferentes funções *kernels* com seus respectivos ajustes de parâmetros que possuam respaldo na literatura para a resolução do problema de pesquisa.

2.3.2.3 Árvores de decisão

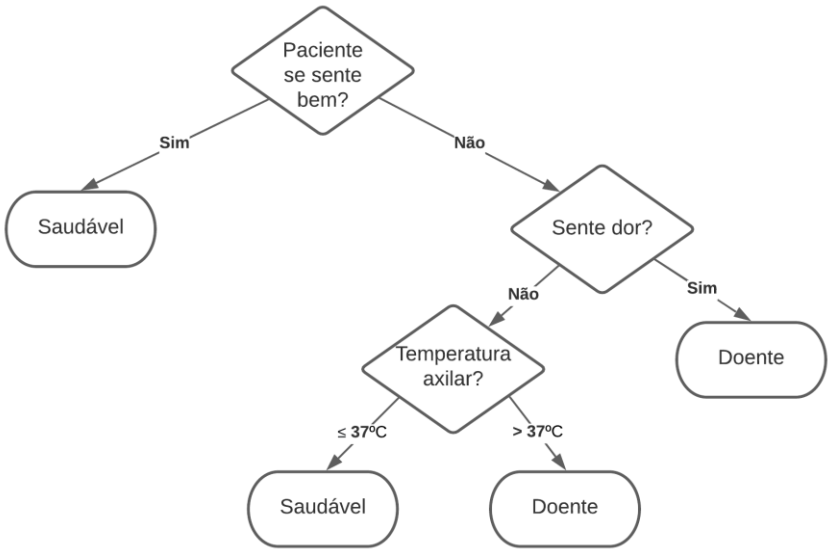
As árvores de decisão são modelos hierárquicos que realizam uma combinação lógica de um conjunto de regras simples por meio de uma busca heurística para segmentar o espaço de predição (KOTSIANTIS, 2013; MARSLAND, 2015; SRIVIDYA; MOHANAVALLI; BHALAJI, 2018). Utilizadas tanto em problemas de regressão quanto de classificação, sua arquitetura se baseia no conceito de árvore binária, que consiste em uma função que recebe como entrada um vetor de atributos e fornece uma tomada de decisão como valor único de saída (TIMOFEEV, 2004; MARSLAND, 2015; RUSSELL; NORVIG, 2016). A classificação realizada pela árvore binária é booleana, onde um atributo nominal é comparado a um valor limite, sendo possíveis dois valores de resposta; cada valor de entrada será classificado como verdadeiro ou falso (KOTSIANTIS, 2013; RUSSELL; NORVIG, 2016).

Os modelos de árvores de decisão são considerados um dos algoritmos mais simples e, ao mesmo tempo, mais poderosos e utilizados de aprendizado de máquina (MARSLAND, 2015; RUSSELL; NORVIG, 2016). Consistem em ferramentas estatísticas robustas para classificação, predição, interpretação e manipulação de dados, com potenciais aplicações em diversas áreas (ALI *et al.*, 2012; SONG; LU, 2015). Entre as vantagens desses modelos, destacam-se seu baixo custo computacional e uma estrutura lógica de fácil compreensão e interpretação (KOTSIANTIS, 2013; MARSLAND, 2015). Além disso, as árvores de decisão são modelos não paramétricos que apresentam bom desempenho em amostras com distribuição assimétrica (*skewed*) e que apresentem dados faltantes ou *outliers* (SONG; LU, 2015).

O processo de classificação de uma árvore de decisão se estrutura conforme o seguinte fluxo: iniciando-se na base ou raiz da árvore, cada atributo é avaliado isoladamente e, por meio de um conjunto de decisões subsequentes e mutualmente excludentes, progride até as folhas da árvore ou os nós finais, que representam o resultado final da combinação decisões (MARSLAND, 2015; SONG; LU, 2015). Suas regras lógicas são simples e podem ser transformadas em estruturas condicionais de decisão (*if-else*), tornando-as adequadas para o emprego de técnicas essenciais de estruturação lógica do aprendizado de máquina, como os sistemas de indução de regras (GRZYMALA-BUSSE, 2009; MARSLAND, 2015). A Figura 5

abaixo ilustra o funcionamento de uma árvore de decisão simples em um problema de classificação.

Figura 5 – Representação esquemática de uma árvore de decisão binária



A Figura 5 representa o esquema de uma árvore de decisão binária simples para o diagnóstico de um paciente hipotético baseada em perguntas subjetivas e na aferição objetiva da temperatura axilar para a decisão diagnóstica entre saudável e doente. **Fonte:** Adaptado de Monard e Baranauskas (2003).

Entre suas aplicações, as árvores de decisão podem ser utilizadas para a resolução de problemas em diversas áreas, como classificação de textos e comparação de dados estatísticos (ALI *et al.*, 2012). Na área médica, a simplicidade conceitual e a capacidade de automatização desses modelos se traduzem em efetividade e confiabilidade para o processo de tomada de decisão clínica (PODGORELEC *et al.*, 2002). Essas características têm propiciado o uso potencial das árvores de decisão no auxílio diagnóstico de doenças cardiovasculares (SHOUMAN; TURNER; STOCKER, 2011) e câncer (ELOUEDI *et al.*, 2014), na estratificação de risco para desenvolvimento de transtornos mentais (SONG; LU, 2015) ou do risco de suicídio em pacientes psiquiátricos (MANN *et al.*, 2008) e na descoberta de novos medicamentos (LANGDON; BARRETT; BUXTON, 2002).

2.3.2.4 Random Forest

Proposto por Breiman (2001), *Random Forest* (RF) é um algoritmo de *ensemble learning* extremamente poderoso e popular para classificação e regressão (QI, 2012; BIAU;

SCORNET, 2016). Consiste em um conjunto (*ensemble*) de árvores de decisão compostas por vetores de variáveis aleatórias cujo resultado se baseia na moda de suas predições individuais (BREIMAN, 2001; ÖZÇIFT, 2011; CUTLER; CUTLER; STEVENS, 2012; BIAU; SCORNET, 2016). A RF se caracteriza por ser um método não paramétrico, eficiente, versátil e de fácil interpretação, que fornece alta acurácia preditiva para diversos tipos de dados, com a vantagem única de apresentar excelente desempenho em conjunto de dados de alta dimensionalidade e tamanho amostral reduzido (QI, 2012; BIAU; SCORNET, 2016).

Diversas outras características tornam a RF um método generalista extremamente bem sucedido atualmente (HOWARD; BOWLES, 2012, apud BIAU; SCORNET, 2016). Do ponto de vista computacional, seus modelos apresentam treinamento relativamente rápido, possuem poucos parâmetros para ajustar e podem ser utilizados para problemas multidimensionais, podendo ser facilmente implementados em paralelo e utilizados em métodos de aprendizado não supervisionado (CUTLER; CUTLER; STEVENS, 2012; BIAU; SCORNET, 2016). Adicionalmente, possuem propriedades valiosas para a análise estatística, como medidas da importância de variáveis, imputação de dados faltantes, detecção de *outliers* e visualização de dados (CUTLER; CUTLER; STEVENS, 2012). Tais características justificam a popularidade e a excelente performance dos modelos de RF em uma grande variedade de problemas de predição em diversas áreas (CUTLER; CUTLER; STEVENS, 2012; BIAU; SCORNET, 2016).

Assim como as árvores de decisão, técnicas de RF têm sido aplicadas com sucesso em uma vasta diversidade de problemas, desde o reconhecimento tridimensional de objetos à ecologia e à bioinformática (QI, 2012; BIAU; SCORNET, 2016). Na área de pesquisa médica, também possuem várias aplicações potenciais, como na classificação de imagens médicas (DÉSIR *et al.*, 2012), no auxílio no diagnóstico oncológico (FAN *et al.*, 2011; RAZAK; YUSOF; RAUS, 2016) e cardiológico (ÖZÇIFT, 2011) e na avaliação do efeito de tratamentos (WAGER; ATHEY, 2018). Especificamente na saúde mental, escopo da aplicação desta pesquisa, esses modelos têm apresentado resultados promissores em áreas como psicologia (ZHAO *et al.*, 2010; SCOTT *et al.*, 2013), neuropsicologia e cognição (BYEON, 2015), *mindfulness* (SAUER *et al.*, 2015) e para o diagnóstico e planejamento terapêutico de transtornos mentais (ABOU-WARDA *et al.*, 2017; ESPINOLA *et al.*, 2020a).

2.3.2.5 Redes bayesianas

As redes bayesianas são modelos gráficos de aprendizado probabilístico que utilizam a inferência bayesiana para o cálculo de probabilidades (PANAGIOTAKOPOULOS *et al.*, 2010;

SINGH; THAKUR; SHARMA, 2016; NAKRA; DUHAN, 2019). Com base no Teorema de Bayes, que fornece inferências estatísticas para o cálculo de probabilidades condicionais, e na teoria dos grafos, esses algoritmos buscam representar as relações de probabilidades em um conjunto de variáveis (PANAGIOTAKOPOULOS *et al.*, 2010; EFRON, 2013; MARSLAND, 2015).

A estrutura das redes bayesianas é formada pela combinação de um grafo direcionado acíclico sobre um conjunto de variáveis aleatórias e uma tabela com suas respectivas probabilidades condicionais (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997; BOUCKAERT, 2008; MARSLAND, 2015). Nesses modelos, a relação estrutural entre as variáveis é determinada pelo conhecimento prévio sobre o domínio, o qual pode ser alcançado pelo treinamento ou pela opinião de um especialista (PANAGIOTAKOPOULOS *et al.*, 2010; SINGH; THAKUR; SHARMA, 2016). Esses modelos representam, então, as distribuições de probabilidades conjuntas dessas variáveis (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997; BOUCKAERT, 2008).

O problema de aprendizado das redes bayesianas consiste em calcular as probabilidades de cada hipótese em um conjunto de dados e fornecer predições baseadas nesses resultados (RUSSELL; NORVIG, 2016). Também pode ser informalmente definido da seguinte forma: dado um conjunto de treinamento $\mathbf{U} = \{x_1, x_2, ..., x_n\}$, $n \geq 1$, com instâncias de x, deve-se encontrar uma rede B que melhor se adapta a **U** (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997). Com base em todas as hipóteses ponderadas por suas respectivas probabilidades, as redes bayesianas realizam, então, suas predições, de forma que o aprendizado se resume a um processo de inferência probabilística de determinado resultado (RUSSELL; NORVIG, 2016).

As redes bayesianas são consideradas uma ferramenta de aprendizado poderosa, oferecendo soluções para problemas como ruídos, *overfitting* e predições otimizadas (RUSSELL; NORVIG, 2016). São bastante úteis em problemas de classificação que envolvem predição ou tomada de decisão (NAKRA; DUHAN, 2019) e possuem aplicações em diversas áreas do conhecimento; entre elas, podem-se citar os programas para predições da ocorrência de crimes (OATLEY; EWART, 2003) e as ferramentas para detecção de COVID-19 baseada em imagens radiográficas (ABRAHAM; NAIR, 2020). Na saúde mental, tem-se mostrado uma técnica bastante promissora para a detecção de depressão maior com base em análise de sentimentos (WANG *et al.*, 2013), no auxílio do tratamento de transtornos de ansiedade (PANAGIOTAKOPOULOS *et al.*, 2010), na detecção de demência com base em imagens de ressonância magnética funcional de crânio (CHEN; HERSKOVITS, 2007), em recomendações comportamentais para promoção de bem-estar (CHEN *et al.*, 2017), entre outras aplicações.

2.3.2.6 Naïve Bayes

Naïve Bayes (NB) é um modelo de aprendizado estatístico que prevê a probabilidade máxima de uma variável pertencer a determinada classe dentro de uma cadeia de probabilidades condicionais (DAI *et al.*, 2007; RUSSELL; NORVIG, 2016; SRIVIDYA; MOHANAVALLI; BHALAJI, 2018; CHO *et al.*, 2019). É um modelo de classificação amplamente utilizado, devido à sua simplicidade, eficácia, eficiência e robustez (JIANG *et al.*, 2016; ARAR; AYAN, 2017) e, assim como as redes bayesianas, também se baseia na aplicação do Teorema de Bayes (PARSANIA; JANI; BHALODIYA, 2014; VEMBANDASAMY; SASIPRIYA; DEEPA, 2015; NAKRA; DUHAN, 2019). Seu nome *naïve* (ingênuo, em inglês) advém do pressuposto de independência, o qual presume que, dada uma classificação, os elementos de um vetor de atributos são mutualmente independentes (MARSLAND, 2015; ALONSO *et al.*, 2018). Como frequentemente existe um grau de dependência entre as variáveis em problemas reais, a "ingenuidade" do modelos reside em considerar que os valores de diferentes atributos não possuem relação de dependência entre si (MARSLAND, 2015; CHO *et al.*, 2019).

Em problemas de alta dimensionalidade, ou seja, com grande número de atributos, pode ocorrer um fenômeno denominado "maldição da dimensionalidade" (MARSLAND, 2015). Este se refere ao aumento exponencial do número necessário de dados de treinamento à medida que a quantidade de dimensões (ou variáveis) de um problema aumenta, tornando o custo computacional do algoritmo proibitivo em determinadas situações (KUO; SLOAN, 2005; VERLEYSEN; FRANÇOIS, 2005). Entretanto, o pressuposto de independência torna o modelo NB mais simples e permite que as probabilidades condicionais sejam facilmente calculadas, reduzindo o impacto da maldição da dimensionalidade e, consequentemente, o custo computacional (MARSLAND, 2015; CHO *et al.*, 2019). Nesse caso, cabe ao usuário do modelo decidir quais atributos são dependentes de outros para, então, ser realizado o cálculo das probabilidades condicionais (CHO *et al.*, 2019).

Apesar da simplificação de um algoritmo computacional significar, em alguns casos, fracasso nas tarefas de classificação, na prática o modelo NB pode apresentar resultados surpreendentemente positivos (CHO *et al.*, 2019). Em certos domínios, seu desempenho é comparável ou até superior a outros métodos de classificação mais sofisticados e será determinado estritamente pelas características presentes na base de dados (MARSLAND, 2015; VEMBANDASAMY; SASIPRIYA; DEEPA, 2015; CHO *et al.*, 2019). Na área da neuropsiquiatria, por exemplo, Bhagya Shree e Sheshadri (2014) relataram superioridade do NB frente a outros a algoritmos, como árvores de decisão e RF, para o diagnóstico da doença

de Alzheimer (DA). De maneira complementar, Qu Yuan e Liu (2009) reportaram bons resultados para a transição de comprometimento cognitivo leve para a DA a partir de dados de neuroimagem. Por último, NB foi superior a SVMs para o diagnóstico de demência em portadores da doença de Parkinson (MORALES *et al.*, 2013).

Uma característica que torna o modelo NB bastante atraente para aplicações na área médica é sua facilidade de implementação, exigindo poucos ajustes de parâmetros (VEMBANDASAMY; SASIPRIYA; DEEPA, 2015). Além dos exemplos acima citados, tem sido utilizado com sucesso em sistemas de predição de doenças cardíacas (SRINIVAS; RANI; GOVRDHAN, 2010; PATTEKARI, S.A.; PARVEEN, 2012; NAHAR *et al.*, 2013) e dermatológicas (MANJUSHA; SANKARANARAYANAN; SEENA, 2014), e na detecção de câncer de mama (KHARYA; AGRAWAL; SONI, 2014; HAZRA; MANDAL; GUPTA, 2016). Importantes aplicações em outras áreas envolvem classificação de textos (DAI *et al.*, 2007; JIANG *et al.*, 2016; XU, 2018) e imagens (LIU; GUO; LEE, 2011), predição de problemas de software (ARAR; AYAN, 2017) e detecção de fraudes financeiras (NGAI *et al.*, 2011; YEE *et al.*, 2018). Todos esses exemplos ilustram a relevância e a flexibilidade dessa técnica para a resolução dos mais variados problemas de classificação.

3 TRABALHOS RELACIONADOS

Considerando os objetivos desta pesquisa, este capítulo propõe uma revisão crítica sobre trabalhos que utilizem padrões de acústica vocal na área da Psiquiatria. Especificamente, foram incluídas nesta revisão pesquisas com foco na detecção e na avaliação de determinados transtornos mentais, assim como na distinção entre diferentes transtornos mentais por meio de atributos vocais. Entretanto, com algumas poucas exceções, a grande maioria dos estudos com fins diagnósticos envolveu somente classificação binária (saudável versus doente), ou seja, na ferramenta de detecção foi incluído apenas um transtorno mental. Não foi encontrado nenhum estudo com o objetivo de distinguir um número maior de transtornos mentais, fato este que ratifica o ineditismo deste trabalho.

Os trabalhos apresentados neste capítulo estão organizados em seções de acordo com o transtorno mental abordado. Visto que nenhum trabalho relacionado ao TAG foi encontrado, foi necessário ampliar o escopo desta revisão para estudos de acústica vocal em outros transtornos de ansiedade como o transtorno de ansiedade social (TAS). Apesar de não ser mais considerado pelo DSM-5 como um transtorno de ansiedade, foram incluídos trabalhos relacionados ao TEPT devido a suas semelhanças com esses transtornos.

3.1 TRANSTORNO DEPRESSIVO MAIOR

Nas últimas duas décadas, um número crescente de trabalhos tem identificado alterações em parâmetros vocais em pacientes com depressão. Por exemplo, Scherer *et al.* (2013) investigaram o seu poder de detecção de depressão e TEPT por meio de atributos derivados da fonte glótica do sinal. Utilizando classificadores SVM, esses autores reportaram uma acurácia de 75% para a detecção de depressão e, a partir das alterações encontradas, eles sugerem que os pacientes deprimidos tendem a apresentar uma qualidade vocal mais tensa que indivíduos saudáveis, a qual seria um indicador de estresse psicológico. Os resultados deste trabalho sobre o TEPT serão comentados na seção referente aos transtornos de ansiedade.

Valendo-se de outra classe de atributos, Taguchi *et al.* (2018) investigaram o comportamento dos coeficientes cepstrais para a classificação de pacientes deprimidos. Após avaliarem 12 dimensões de MFCCs, esses autores relataram que os valores do segundo coeficiente do MFCC (MFCC 2), que fornece detalhes espectrais mais refinados (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010), foram significativamente mais altos em pacientes deprimidos, tanto em grupos separados por gênero como por idade. Com base apenas no MFCC

2, a acurácia reportada foi de 81,9% (Sens.: 77,8%; Espec.: 86,1%) para a discriminação entre pacientes e controles. Também foram observadas diferenças espectrais entre os grupos, com baixa energia na faixa de frequência de 2000 a 3000 Hz no grupo dos deprimidos.

Cohn *et al.* (2009), por sua vez, compararam o desempenho de parâmetros prosódicos combinados com atributos de expressões faciais para a detecção de depressão utilizando regressão logística. Entre os atributos prosódicos, foram avaliadas a variabilidade da F0 e a latência para resposta. A análise das imagens faciais foi realizada automaticamente pelos modelos *Facial Action Coding System* (FACS) e *Active Appearance Modeling* (AAM). O FACS é um sistema utilizado para mensurar movimentos faciais com as características anatômicas dessa região (EKMAN; ROSENBERG, 1997), enquanto o AAM consiste em um *framework* com modelo estatístico para a interpretação de imagens faciais e sequências de imagens (EDWARDS; COOTES; TAYLOR, 1998). Os modelos com atributos vocais e expressões faciais foram treinados separadamente, com acurácias reportadas de 79-88% para expressões faciais, e de 79% para os atributos prosódicos. Como próximo passo, esses autores sugerem a fusão multimodal desses atributos para a criação de uma ferramenta mais poderosa para a detecção de depressão.

Hönig *et al.* (2014) investigaram o impacto do gênero e da seleção de atributos sobre a identificação automatizada da depressão. Nesse trabalho, foram comparados os desempenhos de três classes de parâmetros (prosódicos, espectrais e de qualidade vocal) para a detecção de depressão em amostras separadas por gênero. Os melhores resultados foram alcançados com a combinação das três classes de parâmetros, seguida pelos de qualidade vocal isoladamente. Em concordância com os achados de Low *et al.* (2011) (detalhados abaixo), esses autores relataram uma correlação discretamente maior entre depressão em homens, sugerindo que os sintomas da depressão devem provocar alterações vocais mais perceptíveis em homens do que em mulheres. Também foi relatada correlação negativa entre três atributos de qualidade vocal e escores de depressão: *raw shimmer* (ρ = - 0,46 em homens; ρ = - 0,31 em mulheres), harmonicidade espectral (ρ = - 0,32 em homens; ρ = - 0,33 em mulheres) e *tilt* espectral (ρ = - 0,14 em homens; ρ = - 0,18 em mulheres). As alterações desses atributos indicam a presença de fonação mais irregular, com maior soprosidade no grupo dos deprimidos.

Similarmente, Jiang *et al.* (2017) também observaram diferenças no desempenho de classificadores quanto ao gênero, com resultados superiores em homens. Com base em uma amostra de 170 indivíduos, eles investigaram o poder discriminatório dos algoritmos SVM, *Gaussian Mixture Models* (GMM) e *k-nearest neighbors* (kNN) para a detecção de depressão. Os algoritmos de SVM forneceram os melhores resultados, observando-se também nesse

trabalho performance classificatória superior no gênero masculino, com acurácia de 80,30% em homens (S: 75,00%; E: 85,29%), e de 75,96% em mulheres (S: 77,36%; E: 74,51%).

Por outro lado, o estudo de Alghowinem *et al.* (2012) encontrou resultados divergentes quanto ao gênero, com maior detecção de depressão em mulheres. Nesse trabalho, atributos espectrais, de qualidade vocal, cepstrais e prosódicos foram extraídos e utilizados para treinamento de um modelo GMM para a classificação de depressão. Esses autores reportaram que a identificação de depressão foi superior no gênero feminino para a maioria dos atributos analisados, com *recall* médio ponderado (*Weighted Average Recall*, WAR) de 71%, enquanto o gênero masculino obteve um WAR médio de apenas 55%, e modelos mistos forneceram WAR médio de 64%. Justificando seus resultados, os autores adotam a hipótese de que as mulheres tenderiam a amplificar suas respostas afetivas, o que facilitaria a identificação de depressão nesse gênero.

Mais recentemente, Higuchi *et al.* (2018) não encontraram diferenças entre os gêneros para a diferenciação de amostras de áudio de participantes entre depressão maior, transtorno bipolar e controles saudáveis. Nesse estudo foram analisados a F0, cinco atributos de MFCC e o centroide espectral utilizando-se regressão logística politômica. Além de o desempenho do classificador não ter sido, aparentemente, influenciado pelas diferenças acústicas quanto ao gênero, a acurácia relatada foi de 90,79%, a maior entre os estudos revisados para este trabalho. Entretanto, o reduzido tamanho amostral desse estudo, de 44 participantes para as três classes, sendo apenas oito para o TB, constitui uma importante limitação para os seus resultados.

Os parâmetros acústicos também têm sido estudados para a detecção de sintomas depressivos em adolescentes. Ooi, Lech e Allen (2013) analisaram o uso de atributos prosódicos, espectrais e glóticos combinados com o TEO para a predição de sintomas iniciais de depressão nessa faixa etária. reportando acurácia de 73% (S: 79%; E: 67%). Conforme mencionado no capítulo anterior, o TEO consiste em um conjunto de ferramentas desenvolvidas para o processamento não linear da fala, sendo capazes de capturar a amplitude e as modulações das frequência de ressonância geradas pela passagem do fluxo de ar pelo trato vocal (MARAGOS; KAISER; QUATIERI, 1993; CUMMINS *et al.*, 2015).

Analogamente, Low *et al.* (2011) associaram parâmetros cepstrais ao mesmo conjunto de atributos do trabalho de Ooi, Lech e Allen (2013) para a detecção de depressão em uma amostra maior de adolescentes. Por meio dos classificadores SVM e GMM, esses autores relataram diferenças significativas de desempenho dos classificadores quanto ao gênero, novamente com maior poder de predição de depressão para o gênero masculino (acurácia: 81-87%) do que o feminino (acurácia: 72-79%).

Outra importante aplicação dos padrões de acústica vocal é a avaliação da gravidade dos episódios depressivos. Em um estudo piloto, Cannizzaro *et al.* (2004) investigaram a relação entre os parâmetros de velocidade da fala, proporção do tempo de pausa e variabilidade do *pitch* e os escores de gravidade da Escala de Depressão de Hamilton (HAM-D). Seus resultados mostraram uma discreta correlação negativa, porém significativa, entre a velocidade do discurso e escores de gravidade ($r = - 0,089$, $p = 0,0076$), indicando que, à medida que a gravidade da depressão aumenta, a velocidade do discurso diminui. Também foi encontrada forte correlação negativa entre a variabilidade do *pitch* vocal e os escores da Escala HAM-D, próximo à significância estatística ($r = - 0,74$, $p = 0,0581$). Esses achados são coerentes com os inúmeros relatos de que os deprimidos tendem a falar mais lentamente e de maneira mais monótona, sendo um provável resultado da interferência do retardo psicomotor na produção vocal (CUMMINS *et al.*, 2015).

Similarmente, Hashim *et al.* (2017) avaliaram o desempenho de atributos acústicos para a predição de escores da Escala HAM-D e do Inventário de Depressão de Beck (*Beck Depression Inventory*, BDI). Separando-se os modelos preditivos por gênero, foram analisados os atributos espectrais, cepstrais e prosódicos extraídos de tarefas de leitura. Considerando-se uma margem de erro de três pontos do escore real, os modelos preditivos alcançaram alta acurácia para a pontuação da Escala HAM-D, com desempenho superior e menor variabilidade dos escores no sexo masculino (90,48% versus 87,88%). Enquanto isso, na escala BDI, as taxas de acerto ficaram abaixo de 50%. Apesar dessas diferenças, foi observado que a amostra masculina apresentou pontuação significativamente maior na Escala HAM-D e, portanto, maior gravidade da depressão. Esse fato limita eventuais interpretações desse estudo em relação a um maior impacto da depressão sobre a produção vocal em homens, visto que estes apresentavam sintomas depressivos consideravelmente mais graves que a amostra feminina.

Além de avaliar a gravidade da depressão, os parâmetros acústicos também têm sido estudados como indicadores de resposta ao tratamento antidepressivo. Há mais de quarenta anos, Darby e Hollien (1977) relataram que mudanças na fala de pacientes deprimidos após tratamento foram subjetivamente percebidas, com melhora na prosódia e articulação traduzindo-se em recuperação da vitalidade após tratamento. Três décadas depois, Mundt *et al.* (2007) analisaram medidas acústicas de pacientes deprimidos semanalmente durante um período de seis semanas. Esses autores observaram que os pacientes que obtiveram resposta ao tratamento, indicada pela redução de pelo menos 50% na Escala HAM-D, apresentaram aumento significativo da variabilidade de F2 e da velocidade da fala, e diminuição significativa do número de pausas em comparação com os pacientes que não responderam ao tratamento.

Esses achados foram corroborados em um trabalho subsequente dos mesmos autores (MUNDT *et al.*, 2012), o qual apontou correlação significativa entre atributos acústicos relacionados ao número, duração, proporção e variabilidade de pausas, assim como correlação inversa entre a velocidade da fala e a gravidade dos sintomas depressivos.

Utilizando a base de dados do trabalho de Mundt *et al.* (2007), Quatieri e Malyska (2012) relataram que alterações em determinados parâmetros de qualidade vocal estavam diretamente relacionadas à gravidade da depressão. Enquanto o s*himmer* e o *jitter* apresentaram correlação positiva, a HNR apresentou correlação negativa com as pontuações na escala HAM-D, indicando altos níveis de aspiração durante a fonação. Esses autores sugerem que esses achados são consequências do retardo psicomotor, causando aumento da turbulência do fluxo de ar pela glote.

Em um estudo longitudinal com pacientes deprimidos em tratamento, Yang *et al.* (2013) avaliaram os atributos prosódicos da F0, de duração e do número de pausas como indicadores de resposta ao tratamento. Seus resultados evidenciaram que a diminuição na gravidade da depressão, indicada pela redução do escore na Escala HAM-D, foi significativamente associada a reduções da média e a variabilidade da duração de pausas no mesmo participante (*within-subject*). Por outro lado, F0 não foi associada à melhora dos sintomas depressivos tanto em análises entre participantes (*between-subjects*) como no mesmo indivíduo. Esse achado é sustentado por resultados conflitantes sobre a correlação entre alterações de F0 e a gravidade da depressão (CUMMINS *et al.*, 2015).

Os atributos acústicos também podem se tornar uma ferramenta valiosa na identificação do risco de suicídio. Ozdas *et al.* (2004) investigaram o uso do *jitter* e da curva do espectro de fluxo glótico (*Glottal Flow Spectral Slope*) para a diferenciação entre pacientes deprimidos, controles, e indivíduos em risco iminente de suicídio. Os valores do *jitter* foram significativamente diferentes somente entre controles e pacientes em risco de suicídio, sendo maiores no último. Por outro lado, a curva do espectro de fluxo glótico possibilitou a discriminação entre os três grupos de maneira significativa ($p < 0,05$). Em experimentos de classificação binária, a combinação desses dois atributos atingiu acurácia de 85% para a diferenciação entre controles e pacientes em risco de suicídio, de 90% entre deprimidos e controles, e de 75% entre pacientes em risco de suicídio e deprimidos. Dessa forma, os atributos vocais poderiam auxiliar de maneira objetiva os clínicos na difícil tarefa de estratificação de risco de suicídio. Contudo, como não há dados sobre os transtornos mentais presentes no grupo em risco de suicídio, não é certo se as diferenças encontradas seriam indicadores de risco de suicídio ou se refletiriam o impacto de outros transtornos mentais sobre os atributos avaliados.

Algumas características do desenho de estudo parecem exercer grande influência no poder de predição de depressão dos classificadores. A primeira delas é o tipo do discurso utilizado para a extração de atributos acústicos. Mitra e Shriberg (2015) compararam as taxas de acerto dos classificadores entre tarefas de leitura e discurso espontâneo e observaram que este levou a menores taxas de erro que aquele. Esses autores levantaram a hipótese de que, durante tarefas de leitura, os indivíduos deprimidos consigam suprimir seu estado afetivo devido à natureza irrelevante do conteúdo lido e/ou à concentração na leitura, dificultando a identificação da depressão. O mesmo achado também foi relatado no trabalho Alghowinem *et al.* (2013a), que sugeriram que a maior variabilidade acústica do discurso espontâneo aumentaria a detecção da depressão. Adicionalmente, Jiang *et al.* (2017) reportaram desempenho superior para o discurso espontâneo frente a taferas de leitura, com melhor resultado para entrevistas no sexo feminino, e para descrição de figuras no sexo masculino.

Um segundo fator que pode interferir no desempenho do classificador, produzindo resultados diferentes quanto ao gênero, é a natureza dos atributos acústicos avaliados. Em outro estudo de Alghowinem *et al.* (2012), foram encontradas diferenças significativas entre os gêneros para os modelos de detecção de depressão a depender do atributo acústico considerado. Nesse estudo, os melhores atributos para a classificação de depressão no gênero feminino foram volume, *shimmer* e energia; para o gênero masculino, apenas os parâmetros de energia forneceram resultados superiores. Já em modelos mistos, os melhores atributos foram volume e novamente os atributos de energia. Em todos os três modelos, o parâmetro HNR apresentou os piores resultados para a identificação de depressão nesse estudo.

Outro exemplo dessa relação entre atributo e gênero foi relatado por Scherer *et al.* (2013). Nesse estudo, foi observado que os parâmetros de qualidade vocal, com exceção do atributo de pico da curva (*Peak Slope*), eram menos dependentes do gênero que a F0 para a classificação de depressão. Além disso, esses autores relataram diferenças no desempenho do classificador automatizado quanto à polaridade afetiva do discurso. Trechos de áudio contendo todas as polaridades (positiva, neutra e negativa) forneceram os melhores resultados, enquanto aqueles com polaridade negativa renderam o pior desempenho. Esse achado está possivelmente relacionado ao fato de que na depressão ocorre redução acentuada da reatividade emocional a estímulos positivos (BYLSMA; MORRIS; ROTTENBERG, 2008), facilitando, desta forma, a distinção entre indivíduos saudáveis e pacientes deprimidos.

3.2 TRANSTORNO BIPOLAR

Ao contrário da depressão, existe uma quantidade limitada de trabalhos utilizando parâmetros acústicos no transtorno bipolar, a maior parte deles com foco na identificação de estados afetivos, havendo um número ainda menor de estudos com a aplicação da acústica com fins diagnósticos.

Um dos poucos estudos que visaram à detecção do TB foi realizado por Higuchi *et al.* (2018). Com o objetivo de diferenciar pacientes bipolares, deprimidos unipolares e controles saudáveis com base na análise vocal, esses autores analisaram mais de 6000 atributos extraídos pelo software OpenSMILE®, programa composto por ferramentas de extração automática de parâmetros vocais e de música (EYBEN; WÖLLMER; SCHULLER, 2010). Após a seleção de atributos, esses autores relataram diferenças significativas entre os valores dos MFCCs, do envelope de F0 e do centroide espectral entre pacientes bipolares, deprimidos unipolares e controles saudáveis. Seus resultados apontaram acurácia geral de 90,79%, com 85,71% de acerto para o TB. Este trabalho também se destaca por empregar parâmetros vocais para a detecção de mais de um transtorno mental, porém não fornece detalhes sobre os episódios afetivos dos participantes bipolares no momento do estudo.

Em um estudo subsequente, Higuchi *et al.* (2019) utilizaram o mesmo conjunto de atributos acústicos para a classificação entre os tipos I e II do transtorno bipolar e indivíduos saudáveis. Utilizando regressão logística politômica, sua acurácia geral no conjunto de testes foi de 66,7% e de 96,7% no conjunto de treino, o que levanta a possibilidade de *overfitting*. Além disso, o modelo não conseguiu distinguir com precisão os pacientes bipolares tipo I e tipo II, talvez pelas semelhanças acústicas entre os dois grupos. Apesar dessas limitações, a relevância desse trabalho reside não só na tentativa de detecção do TB, como também em ter considerado suas diferentes nuances diagnósticas ao incluir dois subtipos com apresentações clínicas e prognósticos distintos.

Alguns trabalhos investigaram o uso de parâmetros vocais para a detecção de estados afetivos no TB. Em um estudo preliminar, Vanello *et al.* (2012) analisaram a F0 média e a variabilidade da F0 e do *jitter* de seis pacientes bipolares durante tarefas de leitura e testes de apercepção temática em diferentes fases da doença (eutimia, hipomania, depressão). Apesar da amostra reduzida e de seus resultados não serem consistentes nas duas tarefas, os autores observaram variações individuais significativas, com aumento da F0 média na hipomania em comparação com eutimia, e aumento do *jitter* tanto na depressão quanto na hipomania versus eutimia.

Em outro estudo preliminar, Karam *et al.* (2014) gravaram o áudio de conversas reais de seis pacientes bipolares por telefones celulares durante um ano, a fim de monitorar os estados afetivos do TB no longo prazo. Após a extração dos atributos F0, ZCR, MFCCs, energia e amplitude, modelos de SVM foram testados com e sem seleção de atributos para a classificação de estados afetivos do TB. O modelo com seleção de atributos forneceu resultados discretamente melhores, com área sob a curva (*Area Under Curve*, AUC) ROC (*Receiver Operating Characteristic*) de 0,63 ± 0,04 para a detecção de hipomania, e de 0,64 ± 0,16 para depressão, enquanto o modelo sem seleção de atributos alcançou AUC de 0,61 ± 0,24 e 0,59 ± 0,11, respectivamente. Entretanto, novamente o reduzido tamanho amostral desse estudo limita a capacidade de generalização de seus resultados.

A partir de atributos acústicos extraídos de chamadas telefônicas de pacientes bipolares durante um período de seis a 12 meses, Gideon, Provost e McInnis (2016) investigaram a relevância das etapas de coleta e do pré-processamento de dados sobre o desempenho de modelos preditivos. Para isso, foram utilizados dois modelos de aparelho celular para a coleta de dados, e diferentes procedimentos foram testados na etapa de pré-processamento, como *declipping*, normalização dos atributos de áudio e segmentação com remoção de trechos silenciosos. Na etapa de classificação, algoritmos SVM foram utilizados para a detecção de estados depressivos e maníacos. De todos os modelos testados, os autores relataram uma performance significativamente superior para aquele sem segmentação, com AUC de 0,74 ± 0,24 para detecção de mania e de 0,77 ± 0,15 para depressão.

Faurholt-Jepsen *et al.* (2016) também utilizaram atributos acústicos extraídos de conversas telefônicas para monitorar a atividade da doença no TB. Nesse estudo, os parâmetros vocais foram analisados tanto isoladamente como combinados à autoavaliação do humor e a dados telefônicos sobre interações sociais (número de chamadas telefônicas e de mensagens de texto) e atividade motora por meio de dados de acelerometria e sistema de posicionamento global (*Global Positioning System*, GPS) por 12 semanas. De todos os modelos testados, o melhor desempenho foi alcançado com base apenas em parâmetros vocais, com acurácia de 68% para detecção de estados depressivos (AUC = 0,78) e de 74% para estados maníacos ou mistos (AUC = 0,89). Por outro lado, modelos preditivos mistos, compostos pela combinação de parâmetros vocais com outros tipos de atributos, não trouxeram benefícios adicionais. Dentre estes, a combinação de dados vocais com dados gerados automaticamente por celulares e autoavaliação do humor alcançou acurácia de 73-77% para estados maníacos ou com sintomas mistos, e de 63-66% em estados depressivos.

De maneira análoga, Maxhuni *et al.* (2016) coletaram áudio de chamadas telefônicas, dados gerados automaticamente pelo celular e questionários de autoavaliação diária de cinco pacientes bipolares durante atividades cotidianas pelo mesmo período de 12 semanas. Foram analisados atributos prosódicos e espectrais, dados sobre interações sociais (número e duração de ligações telefônicas, número e tamanho de mensagens de texto) e atividade motora por meio de dados de acelerometria e GPS. Os autores relataram que o número de pausas longas foi diretamente associado à transição de eutimia para um episódio depressivo, enquanto a tendência oposta foi observada para virada maníaca. Adicionalmente, modelos individuais com diferentes conjuntos de atributos foram testados para a predição de recorrência de episódios do TB, com acurácias variando entre 62 a 85% individualmente para cada paciente. Poucos modelos de classificação foram elaborados com dados de todos os pacientes. Dentre estes, a combinação de atributos espectrais da voz com dados de acelerometria forneceu o melhor resultado, com acurácia de 79,84%. Entretanto, esse estudo não detalha sobre a natureza dos episódios identificados (se depressivo, maníaco, com sintomas mistos etc.), além de suas conclusões serem limitadas pelo pequeno número de participantes.

O uso de expressões faciais também foi estudado no monitoramento do transtorno bipolar. Ringeval *et al.* (2018) combinaram atributos de áudio e vídeo para a classificação de episódios afetivos do TB (mania, hipomania e remissão) em uma amostra de pacientes bipolares hospitalizados em decorrência de um episódio maníaco. Os descritores de áudio foram compostos por atributos espectrais, cepstrais, prosódicos e de qualidade vocal; entre os descritores de vídeo, foram incluídos os parâmetros de aparência e informação geométrica. Para a classificação foram testados modelos de aprendizado supervisionados, semi-supervisionados e não supervisionados. No conjunto de testes, o melhor desempenho foi alcançado pelo modelo de aprendizagem supervisionada, com *recall* médio não-ponderado (*Unweighted Average Recall*, UAR) de 57,41% para detecção de uma das três classes. Entretanto, devido à performance superior do modelo de aprendizagem profunda no conjunto de treino (UAR: 63,49%), os autores enfatizam a relevância de abordagens não supervisionadas como uma alternativa para a representação de dados vocais e visuais no TB.

3.3 ESQUIZOFRENIA

As anormalidades do discurso são um elemento central da esquizofrenia desde as primeiras descrições desse transtorno. Nesse contexto, diversos trabalhos foram publicados na área, em sua maioria sobre alterações de linguagem (ELVEVÅG *et al.*, 2010; BEDI *et al.*, 2015;

CHAKRABORTY *et al.*, 2018a; KAYI *et al.*, 2018; TOVAR *et al.*, 2019). Entretanto, ainda existem relativamente poucos estudos dedicados aos aspectos paralinguísticos na esquizofrenia, como, por exemplo, os parâmetros de acústica vocal (CHAKRABORTY *et al.*, 2018b; TAHIR *et al.*, 2019; PAROLA *et al.*, 2020).

Quando comparados com indivíduos saudáveis, portadores de esquizofrenia tendem a apresentar discurso alentecido, redução na variabilidade do *pitch* vocal, aumento significativo no número de pausas e diminuição na variabilidade do tempo silábico. Essas características foram relatadas por Martínez-Sánchez *et al.* (2015), por meio de uma análise acústica semi-automática do *pitch* (F0) durante uma tarefa de leitura emocionalmente neutra. Utilizando algoritmos de processamento de sinais, esses autores reportaram uma acurácia de 93,8% para a detecção de esquizofrenia. Além disso, também observaram diferenças marcantes entre os grupos, especialmente lentificação do discurso, baixa intensidade (ou volume) e aumento do número de pausas no grupo com esquizofrenia.

Analogamente, Rapcan *et al.* (2010) compararam o *pitch* e atributos temporais e de energia de 39 esquizofrênicos e 18 controles durante a leitura de um texto emocionalmente neutro. Seus resultados demonstraram diferenças significativas entre os grupos, sendo observado aumento do número e da duração de pausas e da variação relativa de energia. Por outro lado, ao contrário dos resultados do trabalho acima, não foram relatadas diferenças quanto às variações relativas do *pitch*. Entretanto, a falta de controle do nível educacional entre os grupos em um estudo utilizando tarefa de leitura representa uma limitação importante aos seus achados, uma vez que diferentes níveis educacionais podem se traduzir em diferenças na velocidade de leitura e de fluência entre os dois grupos.

Assim como na depressão, a análise acústica da voz também tem sido empregada para a avaliação da gravidade de sintomas negativos da esquizofrenia. Nesse contexto, Compton *et al.* (2018) compararam áudios de pacientes esquizofrênicos com aprosódia, pacientes sem aprosódia e controles saudáveis em relação à variabilidade (medida pelo desvio-padrão) da F0, do primeiro (F1) e do segundo (F2) formantes e da intensidade. De acordo com seus resultados, os pacientes com aprosódia apresentaram menor variabilidade de F0 e de F2 que os controles. Além disso, a variabilidade da intensidade permitiu diferenciar os dois grupos de pacientes, com menores valores para o grupo com aprosódia em relação ao grupo sem aprosódia e aos controles. Diante desses achados, os autores sugerem que a percepção da falta de inflexão relacionada à aprosódia seria multifatorial e originada por alterações em múltiplos componentes acústicos.

De maneira semelhante, Covington *et al.* (2012) avaliaram os valores da F0, de F1 e F2 em gravações de entrevistas com pacientes em primeiro episódio de transtorno do espectro da esquizofrenia. Com o objetivo de investigar os movimentos de língua como possíveis indicadores da gravidade de sintomas negativos, seu estudo apontou que a variabilidade de F2, uma medida de posição anteroposterior da língua, apresentou correlação negativa com os seguintes itens da Escala para Avaliação da Síndrome Positiva e Negativa (*Positive and Negative Syndrome* Scale, PANSS): gravidade de sintomas negativos ($r = - 0,446$, $p = 0,03$); retraimento emocional ($r = - 0,423$, $p = 0,04$); e falta de espontaneidade e fluência ($r = - 0,523$, $p = 0,007$). Entretanto, seu tamanho amostral contendo apenas 25 minutos de gravações constitui uma limitação dos resultados desse trabalho.

Tahir *et al.* (2019), por sua vez, averiguaram o uso de atributos conversacionais e prosódicos como métricas objetivas de sintomas negativos da esquizofrenia. Durante entrevistas de pacientes por psicólogos, foram avaliados parâmetros relacionados a duração do discurso, turnos de conversação, interrupções, interjeições, F0, formantes F1, F2 e F3, MFCC e medidas de amplitude. Em seguida, diferentes algoritmos foram testados (SVM, MLP, RF e *bagging*) para a classificação entre indivíduos saudáveis e pacientes esquizofrênicos. Dentre estes, MLP obteve o melhor desempenho, acurácia de 81,3%, com as maiores diferenças observadas para o ritmo de discurso e a entropia de frequência e de volume. Além disso, os autores observaram que certos atributos conversacionais, como falha na interrupção, sobreposição de falas, silêncio mútuo, lacunas no discurso e latência de resposta, apresentaram correlação direta com sintomas negativos. Enquanto isso, os atributos de turnos naturais, interjeições, interrupções, percentagem do discurso e duração de turnos apresentaram correlação inversa com sintomas negativos de esquizofrenia. Como já esperado, os autores concluíram que os pacientes com esquizofrenia tendem a falar de maneira mais monótona, mais lenta e com menor variabilidade no volume em comparação com indivíduos saudáveis.

Com base em entrevistas semiestruturadas de pacientes esquizofrênicos, Chakraborty *et al.* (2018a) extraíram parâmetros conversacionais para a predição de avaliações subjetivas de clínicos na escala de sintomas negativos *Negative Symptom Assessment 16-item* (NSA-16). Utilizando o modelo SVM linear, suas acurácias variaram de 64% para a predição do item conteúdo do discurso empobrecido, até 82% para o item latência de resposta prolongada. Entretanto, não foram fornecidos detalhes sobre a gravidade dos sintomas negativos dos pacientes da amostra analisada, como também não foi incluído neste estudo um grupo controle para a comparação dos resultados.

Em um trabalho seguinte com pacientes esquizofrênicos e controles saudáveis, Chakraborty *et al.* (2018b) utilizaram o mesmo conjunto de parâmetros acústicos do seu estudo anterior associado a movimentos corporais, novamente para a predição da avaliações da escala NSA-16, por meio de entrevistas semiestruturadas. Com base nesses conjuntos de atributos, esses autores também propuseram a classificação entre saudáveis e esquizofrênicos. Diversos modelos de ML e métodos de seleção de atributos foram testados, destacando-se a predição dos itens da NSA-16 sobre gestos expressivos reduzidos, quantidade do discurso restrita, latência de resposta prolongada e conteúdo do discurso empobrecido, todos estes com acurácia preditiva acima de 80%. Para a classificação, as acurácias relatadas foram de 79,49% para atributos acústicos e de 86,36% para a associação destes a movimentos corporais. Esse incremento no desempenho classificatório provavelmente advém do fato de os gestos comunicarem o estado afetivo, auxiliando a distinção entre indivíduos saudáveis e pacientes esquizofrênicos.

Em uma metanálise por artigos relacionados a padrões de acústica vocal na esquizofrenia, Parola *et al.* (2020) compararam estudos com três desenhos distintos: avaliações qualitativas, análises quantitativas univariáveis e estudos de ML multivariáveis. Dentre esses, os autores afirmam que os quatro estudos com ML forneceram resultados superiores, com acurácias entre 76,5% e 87,5% para discriminação entre esquizofrenia e controles, demonstrando, assim, serem métodos mais promissores. Diferenças significativas entre os grupos foram relatadas, semelhantes àquelas descritas no trabalho de Tahir *et al.* (2019), com redução da velocidade do discurso e da proporção do tempo de fala, e aumento do número de pausas no grupo de pacientes, sendo estas diretamente associadas a afeto plano e alogia. Adicionalmente, essa metanálise identificou que os estudos com produção dialógica alcançaram o maior tamanho de efeito, seguidos por estudos com produção monológica livre e, por último, aqueles com produção vocal restrita. De certa forma, esses achados se assemelham àqueles anteriormente relatados sobre a natureza da tarefa para obtenção do áudio e a acurácia de classificadores para identificação de depressão maior.

3.4 TRANSTORNOS DE ANSIEDADE

O estresse, tanto agudo quanto crônico, pode interferir negativamente na produção vocal mesmo em indivíduos saudáveis (VAN PUYVELDE *et al.*, 2018). Para ilustrar, há décadas o trabalho de Cook (1969) identificou uma correlação positiva entre ansiedade aguda e alterações na voz e na fala, como vocalização de sons incompreensíveis, repetições supérfluas, tartamudez (gagueira) e sentenças incompletas. Além disso, alterações na velocidade do discurso foram

encontradas em ambos os estados ansiosos (agudo e crônico). Adicionalmente, no estudo de Pope *et al.* (1970), a ansiedade apresentou correlação positiva com velocidade da fala, e negativa com número de pausas silenciosas. Contudo, não são fornecidos detalhes sobre o caráter agudo ou crônico dos estados ansiosos.

Em relação ao comportamento dos parâmetros de acústica vocal em estados de ansiedade, Dietrich e Abbott (2012) relataram medidas reduzidas da F0 e da intensidade vocal em mulheres saudáveis submetidas a estressor social (falar em público). De maneira inversa, Giddens *et al.* (2013) relataram que em diferentes estudos com pilotos de aviação submetidos a tarefas de sobrecarga mental, houve aumento significativo de F0, da intensidade vocal, da frequência cardíaca e da velocidade de fala.

Uma vez que indivíduos saudáveis tendem a apresentar alterações da fala em reação ao estresse, é natural supor que estas também estejam presentes em estados ansiosos patológicos, como os transtornos de ansiedade. Corroborando essa hipótese, Iverach *et al.* (2009) mostraram alta associação entre tartamudez e transtornos de ansiedade, com razão de chances (OR) de 16-34 para TAS, de quatro para TAG e de seis para transtorno de pânico. Achados semelhantes foram relatados por Nerrière *et al.* (2009), que observaram uma associação entre distúrbios da fala e transtornos mentais em uma amostra de professores, com OR de 1,4 (IC 95%: 1,0-1,8) para TAG, e OR de 1,6 (IC 95%: 1,3-2,0) para depressão. Apesar de a associação não poder ser interpretada em termos de causalidade, esses autores sugerem a investigação de transtornos mentais nessa população com queixas vocais.

Alterações de parâmetros vocais também têm sido consistentemente relatadas no transtorno de ansiedade social. Por exemplo, Laukka *et al.* (2008) compararam os valores da F0 de pacientes com TAS antes e após farmacoterapia. De acordo com seus resultados, a redução nos níveis de ansiedade dos pacientes após o tratamento foi acompanhada por uma diminuição correspondente dos valores da F0 média e máxima, de componentes de alta frequência no espectro de energia e da proporção do número de pausas silenciosas em tarefas de falar em público. Adicionalmente, esses autores relataram que a diminuição dos níveis de ansiedade dos participantes também foi qualitativamente perceptível para ouvintes leigos.

Analogamente, Weeks *et al.* (2012) compararam os valores da F0 de indivíduos portadores de TAS e de controles saudáveis durante uma tarefa de interação social e reportaram uma forte correlação positiva entre o aumento da F0 média e a gravidade dos sintomas fóbico-sociais para o subtipo não generalizado em homens ($r = 0,72$, $p = 0,002$), porém não em mulheres ($r = 0,02$, $p = 0,92$). Já para o subtipo generalizado, a correlação foi positiva para ambos os gêneros ($p < 0,001$).

Entretanto, em um estudo posterior com portadores de TAS do subtipo generalizado, os mesmos autores identificaram tal correlação apenas em homens ($p = 0,039$), sugerindo que a F0 poderia ser um indicador de comportamento submisso e de TAS especificamente no sexo masculino (WEEKS *et al.*, 2016). Por fim, deve-se ressaltar que, em ambos os estudos, as amostras foram compostas predominantemente por indivíduos caucasianos, fato que pode limitar a generalização de seus resultados para outros grupos étnicos.

Por último, o citado trabalho de Scherer *et al.* (2013) investigou atributos glóticos como indicadores de TEPT e de depressão maior durante tarefas de interação com um humano virtual. Utilizando um classificador SVM, seus resultados mostraram que o desempenho para a classificação do TEPT variou com a polaridade afetiva dos discursos analisados, com acurácia de 52,38% para trechos envolvendo polaridade emocional negativa, e de 72,09% para trechos com polaridade neutra. Nesse estudo não foram encontradas diferenças de acurácia entre os gêneros para depressão nem para TEPT.

4 MATERIAIS E MÉTODOS

Este capítulo detalha os métodos adotados para a elaboração da ferramenta de solução do problema de pesquisa e engloba os seguintes procedimentos: (1) seleção dos participantes; (2) coleta de dados; (3) pré-processamento com edição das amostras; (4) extração de atributos acústicos; (5) balanceamento de classes; (6) classificação com modelos de aprendizado de máquina; e (7) avaliação de desempenho dos modelos.

4.1 PROTOCOLO DE COLETA DE DADOS

Para este trabalho foram selecionados participantes saudáveis (controles) e pacientes do Hospital Psiquiátrico Ulysses Pernambucano (HUP) e do Hospital das Clínicas da Universidade Federal de Pernambuco (HC-UFPE). Foram incluídos apenas indivíduos portadores de um dos transtornos contemplados neste trabalho: (1) transtorno depressivo maior; (2) esquizofrenia; (3) transtorno bipolar; e (4) transtorno de ansiedade generalizada. Esses quatro grupos serão, a partir de agora, coletivamente denominados de “grupos-transtorno”.

Para os grupos-transtorno foram adotados os seguintes critérios de inclusão:

- Idade acima de 18 anos;
- Diagnóstico anterior de um dos transtornos acima estabelecido por psiquiatra independente, de acordo com os critérios do DSM-5;
- Para o grupo Depressão Maior, escore superior a sete pontos na Escala de Depressão de Hamilton (17 itens) (HAM-D 17), descrita no Anexo C (HAMILTON, 1960);
- Para o grupo TAG, escore igual ou superior a cinco pontos na versão em português da escala *Generalized Anxiety Disorder-7* (GAD-7), conforme Anexo D (SPITZER *et al.*, 2006; SOUSA *et al*., 2015; JORDAN; SHEDDEN-MORA; LÖWE, 2017);
- Ausência de comorbidade com qualquer outro transtorno contemplado neste estudo.

Os participantes do grupo controle foram selecionados em locais diversos, de acordo com o interesse em participar desta pesquisa. Os critérios de inclusão para esse grupo foram: (1) idade acima de 18 anos; e (2) ausência de transtorno mental atual, confirmada pela pontuação igual ou inferior a seis pontos no *Self-Reporting Questionnaire* (SRQ-20), questionário validado para o rastreamento de transtornos mentais comuns (vide Anexo E) (GONÇALVES; STEIN; KAPCZINSKI, 2008; SANTOS *et al.*, 2010; VAN DER WESTHUIZEN *et al.*, 2016).

Foram adotados os seguintes critérios de exclusão para todos os grupos:

- Presença de doença neurológica ou outra condição que interfira na produção vocal;
- Indivíduos transgêneros, uma vez que a inclusão destes impossibilitaria qualquer análise dos resultados por gênero;
- Uso profissional da voz (p. ex.: cantores, locutores).

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (CEP) do HC-UFPE, sob o Parecer nº 3.565.104, e a coleta de dados foi somente iniciada após tal aprovação. Todos os participantes forneceram consentimento formal, por meio do Termo de Consentimento Livre e Esclarecido (TCLE). Em casos de pacientes legalmente incapazes, estes manifestaram sua concordância por meio do Termo de Assentimento Livre e Esclarecido (TALE), enquanto um familiar responsável consentiu por meio do TCLE. Neste estudo, todos os pacientes internados em enfermarias psiquiátricas foram considerados como incapazes, ainda que temporariamente, cabendo-lhes, portanto, o preenchimento do TALE e a um familiar responsável o TCLE.

Após a inclusão dos participantes no estudo, foi realizada a gravação de consulta psiquiátrica de rotina com médico assistente não participante do estudo, sem cortes ou interrupções, exceto por solicitação do participante. Ao final da consulta, a pesquisadora principal solicitou aos profissionais o preenchimento da escala psicométrica correspondente, de acordo com a sintomatologia atual do paciente, exceto os grupos controle e TAG, por se basearem em escalas autoaplicáveis. Por fim, as gravações do grupo controle foram realizadas pela pesquisadora principal.

4.1.1 Seleção dos participantes

Para este trabalho foram selecionados 78 participantes de ambos os gêneros, alocados em um dos cinco grupos acima descritos. Mesmo com diagnóstico previamente estabelecido, antes da inclusão de cada participante foi utilizada uma escala psicométrica validada específica para cada diagnóstico para fins de validação e estratificação da gravidade da doença. A distribuição do número de participantes por grupo, assim como a escala psicométrica adotada para cada um destes estão descritas abaixo e esquematizadas na Tabela 1 a seguir:

- Grupo controle: inclui 12 participantes saudáveis (cinco mulheres);
- Depressão maior: composto por 28 pacientes (23 mulheres) com transtorno depressivo maior;

- Esquizofrenia: possui 20 pacientes (oito mulheres) com diagnóstico de esquizofrenia, cujos sintomas foram avaliados pela Escala Breve de Avaliação Psiquiátrica (*Brief Psychiatric Rating Scale*, BPRS) (vide Anexo A), um dos instrumentos mais amplamente utilizados para a avaliação da gravidade dos sintomas da esquizofrenia (OVERALL; GORHAM, 1962; LEUCHT *et al.*, 2005);
- Transtorno bipolar: inclui 14 portadores (11 mulheres) com transtorno bipolar em episódio atual de mania ou hipomania, com sintomatologia avaliada pela Escala de Avaliação de Mania (EAM) (vide Anexo B), versão em português da *Young Mania Rating Scale* (YMRS) (YOUNG *et al.*, 1978; VILELA *et al.*, 2005);
- Transtorno de ansiedade generalizada: composto por quatro pacientes (três mulheres) com diagnóstico de TAG.

A Tabela 1 abaixo exibe o número de participantes, as características demográficas (idade, gênero) e os escores médios das escalas utilizadas para cada grupo da amostra.

Tabela 1 – Características demográficas e escores médios das escalas psicométricas da amostra de dados

Grupo	**Número de participantes**	**Idade (anos) (DP)**	**Escala psicométrica**	**Escore médio (DP)**
Controle	12 (5 ♀)	29,2 (± 12,4)	SRQ-20	3,00 pontos (± 1,86)
Transtorno depressivo maior	28 (23 ♀)	42,0 (± 12,4)	HAM-D 17	19,32 pontos (± 7,36)
Esquizofrenia	20 (8 ♀)	36,0 (± 11,3)	BPRS	45,16 pontos (± 11,25)
Transtorno bipolar	14 (11 ♀)	40,5 (± 8,0)	EAM	23,00 pontos (± 11,95)
Transtorno de ansiedade generalizada	4 (3 ♀)	25,8 (± 8,5)	GAD-7	13,75 pontos (± 2,22)

Abreviações: BPRS: *Brief Psychiatric Rating Scale*; DP: desvio-padrão; EAM: Escala de Avaliação de Mania; GAD-7: *Generalized Anxiety Disorder-7 Scale*; HAM-D 17: Escala de Depressão de Hamilton; SRQ-20: *Self-Reporting Questionnaire*. **Fonte:** Elaborado pela autora (2021).

Por meio da Tabela 1 acima, observa-se heterogeneidade entre os grupos quanto ao número de participantes e às características sociodemográficas, com tendência à forte predominância do gênero feminino em três grupos (Depressão, TB e TAG). Por exemplo, o grupo Depressão possui a maior amostra, sendo composto majoritariamente por mulheres e por pessoas mais velhas (média: 42,0 anos); enquanto isso, o grupo TAG apresentou a menor média

de idade (média: 25,8 anos) e o menor número de participantes. Por fim, os grupos controle e esquizofrenia foram compostos, em sua maioria, por indivíduos do gênero masculino.

Sobre os valores encontrados das escalas psicométricas, no grupo controle foi encontrada uma pontuação média de três pontos. Para o grupo depressão, como o critério de elegibilidade considerou uma pontuação acima de sete pontos na escala de HAM-D 17, foram incluídos nesta pesquisa pacientes com quadro depressivo desde leve até grave. Nesse grupo, o escore médio encontrado de 19,32 pontos na escala HAM-D 17 indica a presença de transtorno depressivo moderado (ZIMMERMAN *et al.*, 2013). Já para o grupo esquizofrenia, participantes com diagnóstico prévio deste transtorno foram selecionados independentemente de sua pontuação na escala BPRS, de forma a evitar a exclusão de pacientes com sintomas leves. Foi obtido nesse grupo um valor médio de 45,16 pontos na escala BPRS, correspondendo a sintomatologia de moderada intensidade (LEUCHT *et al.*, 2005). Similarmente, pacientes do grupo transtorno bipolar também foram incluídos independentemente de sua pontuação na escala EAM, desde que diagnosticados com episódio atual maníaco ou hipomaníaco. O valor médio obtido na escala EAM foi de 23,00 pontos, indicando a presença de quadro maníaco grave (LUKASIEWICZ *et al.*, 2013). Infelizmente, durante a coleta de dados, só foi possível selecionar quatro pacientes com TAG, cujo escore médio de 13,75 pontos na escala GAD-7 corresponde a doença moderada (SPITZER *et al.*, 2006).

4.2 COLETA DE DADOS ACÚSTICOS

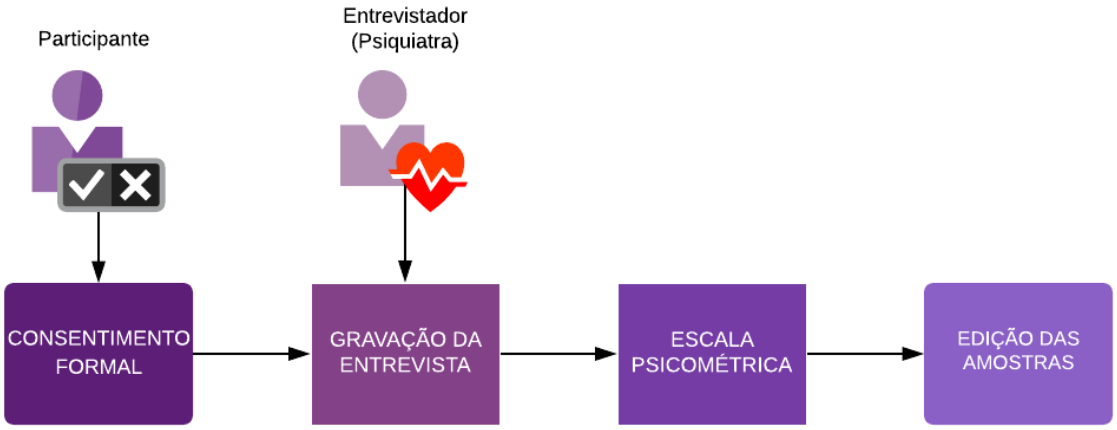
Todas as gravações de áudio foram realizadas com um gravador digital de 16 bits da marca Tascam®, modelo DR-05. Foram obtidas com as seguintes especificações técnicas: formato WAV; taxa de amostragem de 44,1 KHz; sem compressões (para evitar perda de dados); e sem a utilização de filtros. Não foi estipulada duração mínima ou máxima para as gravações, de forma que os participantes poderiam falar livremente durante sua consulta de rotina com psiquiatra assistente.

Para os participantes dos grupos-transtorno, a coleta de dados foi realizada em três locais: (1) no Ambulatório de Psiquiatria do HC/UFPE; (2) em enfermarias do HC/UFPE; e (3) na enfermaria do HUP, todos na cidade do Recife. Em relação aos voluntários do grupo controle, a coleta de dados foi realizada em ambientes variados, que consistiram em salas de aula ou de reunião, escritórios, residências e laboratórios, priorizando-se locais com níveis de ruído semelhantes àqueles dos grupos-transtorno. Todas as gravações foram obtidas em ambiente naturalístico, ou seja, durante uma consulta de rotina com psiquiatra em ambulatório

ou enfermaria hospitalar para os participantes dos grupos-transtorno, e nos locais supracitados para o grupo controle. Os indivíduos selecionados integraram este estudo de maneira voluntária, não recebendo nenhuma compensação financeira por sua participação.

Conforme mencionado, após cada entrevista, o profissional responsável por esta deveria preencher a escala psicométrica adequada ao grupo do participante, para avaliar a elegibilidade (no caso do grupo depressão) e a gravidade dos sintomas. Exceções são os grupos controle e TAG, pois uma vez que o questionário SRQ-20 e a escala GAD-7 são autoaplicáveis, os participantes desses grupos foram solicitados a respondê-los antes ou após sua entrevista, respectivamente. O processo de aquisição das amostras de áudio para os grupos-transtorno está esquematizado abaixo na Figura 6. Todos os diálogos foram inteiramente gravados e, portanto, a fala do entrevistador e de eventuais terceiras pessoas, como acompanhantes dos pacientes e estudantes também foram gravadas e precisaram ser posteriormente removidas. O tempo total de gravação para todos os grupos corresponde a 980,3 minutos (16,3 horas).

Figura 6 – Diagrama da coleta de dados para os grupos-transtorno



A Figura 6 esquematiza o processo de coleta de dados para os pacientes dos grupos Depressão Maior, Transtorno Bipolar e Esquizofrenia. Após o consentimento formal do participante (e de seu familiar, se necessário), o psiquiatra assistente (entrevistador) dá início à consulta de rotina e, após o término desta, preenche a escala psicométrica adequada ao diagnóstico do paciente entrevistado. Para o grupo TAG, o participante responde a escala GAD-7 após o término de sua consulta psiquiátrica, uma vez que esta é autoaplicável. Os participantes do grupo controle foram entrevistados pela autora deste estudo após responderem o questionário SRQ-20 (vide critérios de inclusão do grupo controle). Após a verificação dos critérios de elegibilidade, todas as amostras seguem para o processo de edição, utilizando-se o programa Audacity®. **Fonte:** Elaborado pela autora (2021).

4.3 EDIÇÃO DAS AMOSTRAS DE ÁUDIO

Nessa etapa foi utilizado o programa editor de áudio Audacity® para a remoção da fala do entrevistador e de quaisquer outras pessoas, restando, ao final desta, apenas o discurso do

participante. Todo o processo de edição foi realizado manualmente pelos pesquisadores do grupo de pesquisas em Computação Biomédica da UFPE. Ao término desse processo, o tempo de gravação das amostras totalizou 591 minutos (9,85 horas) para os cinco grupos, divididos da seguinte forma: 100,7 minutos no grupo controle; 222,6 minutos no grupo depressão; 125,7 minutos no grupo esquizofrenia; 102 minutos no grupo TB; e 40 minutos no grupo TAG. As informações detalhadas sobre as durações das gravações após a edição para cada grupo encontram-se na Tabela 2 abaixo.

Tabela 2 – Tempo total de gravação e duração média das gravações após edição das amostras de áudio

Grupo	**Número de participantes**	**Tempo total de gravação após edição**	**Duração média de gravação (DP)**
Controle	12	6039s (100,7 min)	503,3s (8,4 min) ± 159,0s
Transtorno depressivo maior	28	13355s (222,6 min)	477,0s (≅ 8,0 min) ± 203,0s
Esquizofrenia	20	7541s (125,7 min)	377,1s (6,3 min) ± 270,4s
Transtorno bipolar	14	6122s (102,0 min)	437,3s (7,3 min) ± 253,9s
Transtorno de ansiedade generalizada	4	2401s (40,0 min)	600,3s (10,0 min) ± 194,8s

Abreviação: DP: desvio-padrão. **Fonte:** Elaborado pela autora (2021).

A Tabela 2 acima nos mostra diferenças consideráveis nas durações médias de fala entre os diferentes grupos. Os participantes com maior duração de fala foram os do grupo TAG, seguidos pelos do grupo controle, ainda que o reduzido tamanho amostral no primeiro impossibilite análises mais aprofundadas. No outro extremo, o grupo com menor duração das entrevistas foi o de pacientes com esquizofrenia, podendo estar relacionado a importantes manifestações da doença, como o desinteresse social e a alogia.

4.4 EXTRAÇÃO DE ATRIBUTOS ACÚSTICOS

Após a etapa de edição, as amostras editadas foram submetidas ao processo de extração de atributos acústicos em busca da melhor representação desses sinais. Para isso, foram desenvolvidos algoritmos próprios para o processamento de sinais pelo Grupo de Pesquisas em Computação Biomédica da UFPE, utilizando-se o programa *open-source* GNU Octave®. Para

o janelamento, foram adotadas janelas de formato retangular, com duração de 10 segundos e sobreposição de 50%. A decisão por essas propriedades de janelamento se baseou em um estudo piloto realizado pela autora que comparou os desempenhos dos classificadores de aprendizado de máquina com diferentes tamanhos de janela (10 s, 5 s, 1 s, 50 ms) e sobreposições (10%, 25% e 50%), obtendo melhores resultados com as configurações acima descritas. Outro estudo prévio nosso também demonstrou melhores desempenhos com a sobreposição de janelas de 50% (ESPINOLA *et al.*, 2020a). Por último, optou-se por utilizar os áudios "crus" (*raw*), ou seja, sem a utilização de filtros, pois entende-se que esse procedimento traria uma representação mais fidedigna dos sinais de áudio. Consequentemente, ruídos de fundo também foram capturados; entretanto, acredita-se que estes não interferiram significativamente nos sinais de interesse por apresentarem um padrão espectral homogêneo.

Em seguida, foram extraídos 33 atributos para a obtenção de uma representação paramétrica dos sinais acústicos da fala. São estes: média; variância; desvio-padrão; *skewness*; comprimento de onda (*Wavelength*, WL); *kurtosis*; taxa de cruzamentos (*Zero Crossing Rate*, ZCR); variações do sinal da curva (*slope sign changes*, SSC); valor absoluto médio (*Mean Absolute Value*, MAV); detector logarítmico (*Logarithm Detector*, LOGD); raiz quadrada média (*Root Mean Square*, RMS); variações médias de amplitude (*Average Amplitude Changes*, AAC); desvio diferencial absoluto (*Difference Absolute Desviation*, DASDV); valor absoluto integrado (*Integrated Absolute Value*, IAV); *kernel* logarítmico médio (*Mean Logarithm Kernel*, MLOGK); integral quadrada simples (*Simple Square Integral*, SSI); valor absoluto médio (*Mean Absolute Value*, MAV); terceiro, quarto e quinto momentos; amplitude máxima; razão do espectro de potência (*Power Spectrum Ratio*, PSR); pico de frequência (*Peak Frequency*, PKF); potência média (*Mean Power*, MNP); frequência média (*Mean Frequency*, MNF); frequência mediana (*Median Frequency*, MDF); potência total (*Total Power*, TP); variância da frequência central (*Variance of Central Frequency*, VCF); primeiro, segundo e terceiro momentos espectrais; e atividade, mobilidade e complexidade do parâmetro de Hjorth.

A seleção inicial dos atributos acima se baseia na ampla expertise do Grupo de Pesquisas em Computação Biomédica do Laboratório de Computação Biomédica (LCB), o qual possui um *framework* consolidado na literatura para a extração de atributos de sinais biológicos e imagens médicas, como, por exemplo, eletroencefalografia (DA SILVA JUNIOR *et al.*, 2019; OLIVEIRA *et al.*, 2020; SILVA *et al.*, 2021), ressonância magnética (DOS SANTOS *et al.*, 2008, 2009), mamografia (AZEVEDO *et al.*, 2015; CRUZ; CRUZ; SANTOS, 2018) e termografia (DE SANTANA *et al.*, 2018; RODRIGUES *et al.*, 2019) As fórmulas dos atributos utilizados neste trabalhos estão exibidas no Quadro 3 abaixo.

Quadro 3 – Equações dos parâmetros extraídos

Parâmetro	Equação	Parâmetro	Equação
Média (μ)	$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$	**Comprimento de Onda**	$WL = \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Variância	$var = \frac{1}{N-1}\sum_{n=1}^{N} (x_n - \mu)^2$	**Taxa de Cruzamentos**	$ZCR = \sum_{n=1}^{N-1} [sgn(x_n \times x_{n+1}) \cap \lvert x_n - x_{n+1} \rvert \geq limiar]$ $sgn(x) = \begin{cases} 1, if\ x \geq limiar \\ 0, \quad caso\ contrário \end{cases}$
Desvio-padrão (σ)	$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N} \lvert x_n - \mu \rvert^2}$	**Variações do Sinal da Curva**	$SSC = \sum_{n=1}^{N-1} [f(x_n - x_{n-1}) \times (x_n - x_{n+1})]]$ $f(x) = \begin{cases} 1, if\ x \geq limiar \\ 0, \quad caso\ contrário \end{cases}$
Raiz Quadrada Média	$RMS = \sqrt{\frac{\sum_{n=1}^{N} (x_n)^2}{N}}$	**Atividade do parâmetro Hjorth**	$Hjorth_{ativ} = \frac{1}{N-1}\sum_{n=1}^{N} (x_n - \mu)^2$
Variações Médias de Amplitude	$AAC = \frac{1}{N}\left(\sum_{n=1}^{N} \left\lvert \frac{dx(t)}{dt} \right\rvert\right)$	**Mobilidade do parâmetro Hjorth**	$Hjorth_{mobilidade} = \sqrt{\frac{var\left(\frac{dx(t)}{dt}\right)}{var(x(t))}}$
Desvio Diferencial Absoluto	$DASDV = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(\frac{dx(t)}{dt}\right)^2}$	**Complexidade do parâmetro Hjorth**	$Hjorth_{complexidade} = \frac{Hjorth_{mobilidade}\left(\frac{dx(t)}{dt}\right)}{Hjorth_{mobilidade}(x(t))}$
Valor Absoluto Integrado	$\mathrm{IAV} = \sum_{n=1}^{N} x_n$	**Frequência Média**	$MNF = \frac{\sum_{j=1}^{M} f_j P_j}{\sum_{j=1}^{M} P_j}$ Onde f_j,P_j são as frequências e energia do espectro, respectivamente, e M é o comprimento das frequências
Detector Logarítmico	$LOGD = e^{(\frac{1}{N}\sum_{n=1}^{N} \log(\lvert x_n \rvert))}$	**Frequência Mediana**	$MDF = \frac{1}{2}\sum_{j=1}^{M} P_j$
Integral Quadrada Simples	$\mathrm{SSI} = \sum_{n=1}^{N} {x_n}^2$	**Potência Média**	$MNP = \sum_{j=1}^{M} \frac{P_j}{M}$

Continua

Cont. Quadro 3

Valor Absoluto Médio	$\mathrm{MAV} = \frac{1}{N}\sum_{n=1}^{N}\lvert x_n\rvert$	**Pico de Frequência**	$PKF = \max(P_j)$
***Kernel* Logarítmico Médio**	$\mathrm{MLOGK} = \frac{1}{N}\left\lvert\sum_{n=1}^{N} x_n\right\rvert$	**Razão do Espectro de Potência**	$PSR = \frac{PKF}{\sum_{j=1}^{M} P_j}$
Skewness (s)	$s = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^3}{\sigma^3}$	**Energia Total**	$TP = \sum_{j=1}^{M} P_j$
Kurtosis	$\mathrm{kurtosis} = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^4}{\sigma^4}$	**Primeiro Momento Espectral**	$SM1 = \sum_{j=1}^{M} f_j P_j$
Amplitude Máxima	$MAX = \max(x_n)$	**Segundo Momento Espectral**	$SM2 = \sum_{j=1}^{M} {f_j}^2 P_j$
Terceiro Momento	$M3 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^3\right\rvert$	**Terceiro Momento Espectral**	$SM3 = \sum_{j=1}^{M} {f_j}^3 P_j$
Quarto Momento	$M4 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^4\right\rvert$	**Variância da Frequência Central**	$VCF = \frac{SM2}{TP} - \left(\frac{SM1}{TP}\right)^2$
Quinto Momento	$M5 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^5\right\rvert$		

Abreviações: AAV, variações médias de amplitude; DASDV, desvio diferencial absoluto; IAV, valor absoluto integrado; LOGD, detector logarítmico; MAV, valor absoluto médio; MAX, amplitude máxima; MDF, frequência mediana; MLOGK, kernel logarítmico médio; MNF, frequência média; MNP, potência média; M3, terceiro momento; M4, quarto momento; M5, quinto momento; PKF, pico de frequência; PSR, razão do espectro de potência; RMS, raiz quadrada média; s, *skewness*; SM1, primeiro momento espectral; SM2, segundo momento espectral; SM3, terceiro momento espectral; SSC, variações do sinal da curva; SSI, integral quadrada simples; TP, potência total; VAR, variância; VCF, variância da frequência central; WL, comprimento de onda; ZCR, taxa de cruzamentos; μ, média; σ, desvio-padrão. **Fonte:** Elaborado pela autora (2021).

4.5 SELEÇÃO DE ATRIBUTOS

Após a extração dos 33 parâmetros acústicos, foi realizado um estudo piloto para investigar o impacto da seleção de atributos sobre o desempenho dos classificadores e a redução do custo computacional. Para isso, foi utilizado o método *Particle Swarm Optimization* (PSO) de seleção de atributos para a redução da dimensionalidade em problemas de classificação (XUE *et al.*, 2012). Desenvolvidos por Eberhart e Kennedy (1995), a técnica PSO consiste em um conjunto de algoritmos de otimização inspirados no comportamento social de animais, como peixes e aves, compartilhando elementos em comum com os algoritmos genéticos e a

programação evolucionária (EBERHART; KENNEDY, 1995; KENNEDY; EBERHART, 1995; ESPINOLA *et al.*, 2020a). A utilização dos algoritmos PSO selecionou 12 atributos como os mais representativos da base de dados, os que se encontram listados no Quadro 4 abaixo.

Quadro 4 – Atributos selecionados pelo método PSO

Atributos selecionados por PSO	
Taxa de cruzamentos (ZCR)	Complexidade do parâmetro Hjorth
Variações médias de amplitude	Valor médio absoluto (MAV)
Kurtosis	Amplitude máxima
Terceiro momento	Quarto momento
Pico de frequência	Razão do espectro de potência
Potência média	Potência total

Abreviação: PSO, *Particle Swarm Optimization*. **Fonte:** Elaborado pela autora (2021).

Os atributos acima selecionados foram utilizados em experimentos de um estudo piloto prévio. Apesar de terem proporcionado uma redução significativa do tempo necessário para os experimentos, essa seleção de atributos provocou uma piora considerável dos desempenhos de praticamente todos os classificadores. Devido a isso, os dois conjuntos de experimentos foram realizados com todos os 33 atributos extraídos.

4.6 BALANCEAMENTO DE CLASSES

A etapa de balanceamento entre as classes diagnósticas é fundamental para evitar o favorecimento do aprendizado dos classificadores para classes com maior representatividade, que, nesta base de dados, são as classes depressão e esquizofrenia, em detrimento das classes minoritárias. Caso não fosse solucionado, o desbalanceamento entre as classes traria importantes consequências ao processo de aprendizado, pois os modelos gerados pelos classificadores tenderiam a classificar a maioria das amostras como pertencentes às classes majoritárias, com baixo poder preditivo sobre as classes minoritárias. Isso, por conseguinte, comprometeria a performance e a capacidade de generalização dos classificadores (BERMEJO; GÁMEZ; PUERTA, 2011; BLAGUS; LUSA, 2013).

Nessa etapa, foram testadas duas formas de balanceamento de classes, cada uma originando um conjunto independente de experimentos. No primeiro, foi utilizado o *ClassBalancer*, um balanceamento padrão do programa Weka® (FRANK, 2019). O método *ClassBalancer* consiste em um filtro simples aplicado automaticamente pelo programa que

adiciona pesos às instâncias, de forma que cada classe de instâncias terá o mesmo peso, enquanto a soma total de pesos das instâncias na base de dados permanece inalterada (FRANK, 2019). Dessa forma, as classes minoritárias, aqui principalmente o grupo TAG, adquirem mais peso que as majoritárias, evitando que sejam ignoradas pelos classificadores.

Para a segunda rodada de experimentos, foi adotada uma técnica mais sofisticada de balanceamento baseada em métodos de *resampling*, que consistem na reamostragem da base de dados por meio da remoção de seus elementos ou adição de exemplos artificiais (BURNAEV; EROFEEV; PAPANOV, 2015; LI; VASCONCELOS, 2019). Neste trabalho, foi utilizado um método de *resampling* bastante conhecido, denominado *Synthetic Minority Oversampling Technique* (SMOTE). Desenvolvido por Chawla *et al.* (2002), essa técnica corresponde à combinação de *oversampling* das classes minoritárias por meio da criação de instâncias sintéticas e de *undersampling* das classes majoritárias (CHAWLA *et al.*, 2002; BERMEJO; GÁMEZ; PUERTA, 2011). O resultado da aplicação do método SMOTE é a redução dos *bias* de representação do classificador para as classes mais representativas, com o consequente aumento da acurácia para a detecção de classes minoritárias (BERMEJO; GÁMEZ; PUERTA, 2011).

4.7 CLASSIFICAÇÃO

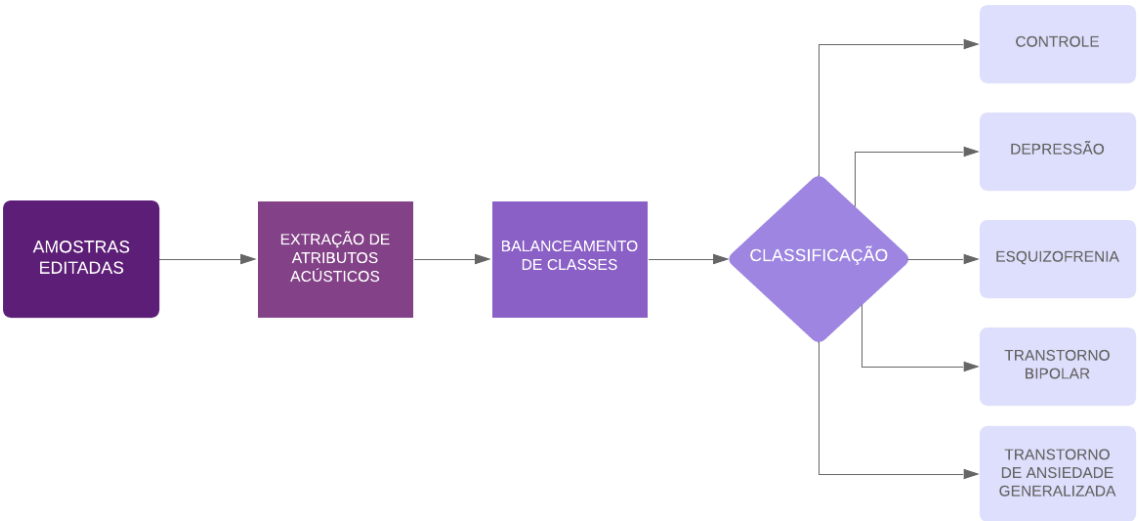
Nessa etapa, as amostras balanceadas foram submetidas a experimentos computacionais com modelos de classificação no programa Weka®, um ambiente aberto de desenvolvimento em inteligência artificial (THE UNIVERSITY OF WAIKATO, 2020). Para a realização dos experimentos computacionais, foram selecionados os seguintes algoritmos de aprendizado supervisionado para classificação: MLP, *Random Forest*, árvores de decisão (algoritmo J48) SVM (*kernels* polinomial, PUK e RBF), redes bayesianas e NB. Para cada modelo, foram testadas as seguintes configurações:

- MLP:
    - Padrão;
    - Número de neurônios na camada oculta: 20, 50, 100, 200, 300, 400 ou 500 (todos com taxa de aprendizado de 0,1 e momentum de 0,2);
- RF:
    - 10, 20 50 e 100 iterações;
- Arvores de decisão (J48):
    - Padrão;

- SVM polinomial:
  - Grau (expoente): 1º (linear), 2º e 3º;
  - Parâmetro de complexidade (*C*): 0,01, 0,1, 1,0, 10 e 100;
- SVM PUK:
  - Parâmetro de complexidade (*C*): 0,01, 0,1, 1,0, 10 e 100;
- SVM RBF:
  - Hiperparâmetro *gamma (g)*: 0,01;
  - Parâmetro de complexidade (*C*): 0,01, 0,1, 1,0, 10 e 100;
- Redes Bayesianas:
  - Padrão;
- NB:
  - Padrão.

Tanto no primeiro quanto no segundo conjunto de experimentos, foram realizadas 30 iterações para cada teste, com o método de validação cruzada com 10 *folds.* Na Figura 7 a seguir estão esquematizadas as etapas da solução proposta neste trabalho.

Figura 7 – Diagrama da solução proposta neste trabalho



A Figura 7 exibe um diagrama de fluxo com as etapas envolvidas na ferramenta de solução do problema desta pesquisa. Após finalizado o processo da edição das amostras de áudio, as amostras editadas são submetidas a algoritmos de processamento de sinais para a extração dos atributos acústicos desejados para a representação desses sinais a uma menor taxa de informações. Em seguida, é aplicado um método de balanceamento de classes a depender do conjunto de experimentos (*ClassBalancer* ou SMOTE), para evitar viés de seleção dos classificadores para as classes com maior número de amostras de áudio. Por último, na etapa de classificação, as amostras compõem os dados de entrada de modelos supervisionados de aprendizado de máquina (MLP, SVM, RF, árvores de decisão, NB e redes bayesianas) para a classificação entre os cinco grupos diagnósticos: controle,

depressão maior, esquizofrenia, transtorno bipolar e transtorno de ansiedade generalizada. **Fonte:** Elaborado pela autora (2021).

O diagrama exposto na Figura 7 acima exibe, de maneira resumida, todas as etapas adotadas no processo de resolução do problema de pesquisa. É importante ressaltar-se que foram utilizados neste trabalho somente softwares gratuitos e, se possível, *open-source*, de forma a aumentar-se a reprodutibilidade desta pesquisa. Além disso, todos os softwares adotados (Audacity®, Octave® e Weka®) possuem uma interface amigável ao usuário (*user-friendly*), estão disponíveis fora de laboratórios de pesquisa, e podem ser utilizados por computadores pessoais, dispensando a necessidade de servidores ou outras máquinas de alto custo.

4.8 MÉTRICAS DE DESEMPENHO

Para a avaliação do desempenho dos classificadores nos experimentos, as seguintes métricas estatísticas foram adotadas: (1) acurácia; (2) sensibilidade; (3) especificidade; e (4) índice kappa, o qual é um coeficiente que avalia o grau de concordância entre dois avaliadores independentes (VIERA; GARRETT, 2005). Neste trabalho, tal concordância corresponde ao grau de conformidade entre os resultados dos diferentes *folds* da fase de validação cruzada dos experimentos.

5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados dos experimentos computacionais realizados para a classificação de depressão maior, esquizofrenia, transtorno bipolar, transtorno de ansiedade generalizada e grupo controle por meio da implementação de modelos supervisionados de aprendizado de máquina.

Neste trabalho, a base de dados foi submetida a dois tipos de balanceamento de classes (*ClassBalancer* e SMOTE), detalhados no capítulo anterior, originando dois conjuntos de experimentos independentes. A etapa de balanceamento é crucial para evitar vieses de aprendizado do modelo de aprendizado de máquina em favorecimento das classes com maior representatividade, as quais, nesta base de dados, são as classes Depressão Maior e Esquizofrenia. No primeiro experimento, foi aplicado o método *ClassBalancer* do Weka® para o balanceamento de classes, descrito no capítulo anterior. Ao final desse processo, foram geradas 6979 instâncias, correspondendo a 1395,8 instâncias por classe. Em seguida, foram realizados os experimentos de classificação com os modelos de ML e seus respectivos ajustes de hiperparâmetros, também detalhados no capítulo anterior. Para cada experimento, foi adotado o método de 30 iterações. Logo, a Tabela 3 abaixo exibe os valores médios de desempenho para cada classificador, com precisão de quatro casas decimais, para as seguintes métricas e seus desvios-padrões: acurácia; índice kappa; sensibilidade; e especificidade. Todos os resultados foram validados com o método de validação cruzada com 10 *folds*.

Tabela 3 – Desempenhos médios dos modelos computacionais para classificação após o balanceamento *ClassBalancer* do Weka®

Classificador	**Configuração**	**Acurácia (%) (DP)**	**Índice Kappa (DP)**	**Sensibilidade (DP)**	**Especificidade (DP)**
MLP	Padrão (L:0,3)	58,4296 (2,9541)	0,4703 (0,3160)	0,5542 (0,1046)	0,8693 (0,0481)
	H: 20; L:0,1	61,5628 (2,5244)	0,5072 (0,0277)	0,6107 (0,0919)	0,8719 (0,0418)
	H: 50; L:0,1	64,3091 (2,5694)	0,5410 (0,0287)	0,6599 (0,0893)	0,8777 (0,0386)
	H: 100; L:0,1	64,9649 (2,7501)	0,5490 (0,0311)	0,6672 (0,0858)	0,8778 (0,0393)

Continua

Cont. Tabela 3

MLP	H: 200; L:0,1	65,4111 (2,4264)	0,5542 (0,0278)	0,6751 (0,0834)	0,8800 (0,0375)
	H: 300; L:0,1	65,3599 (2,4315)	0,5537 (0,0279)	0,6768 (0,0814)	0,8786 (0,03656)
	H:400; L:0,1	65,7120 (2,3141)	0,5579 (0,0263)	0,6801 (0,0808)	0,8793 (0,0363)
	H: 500; L:0,1	65,5495 (2,3593)	0,5560 (0,0267)	0,6834 (0,0777)	0,8771 (0,0370)
Árvores de Decisão (J48)	Padrão	62,0280 (1,7354)	0,4932 (0,0231)	0,5219 (0,0460)	0,9008 (0,0138)
Random Forest	Padrão (100 árvores)	76,4293 (1,4698)	0,6883 (0,0192)	0,7540 (0,0388)	0,9234 (0,0108)
	10 árvores	71,1301 (1,6322)	0,6183 (0,0214)	0,7197 (0,0419)	0,8913 (0,0131)
	20 árvores	73,6920 (1,4840)	0,6520 (0,0195)	0,7346 (0,0406)	0,9081 (0,0119)
	50 árvores	75,6766 (1,5570)	0,6784 (0,0204)	0,7481 (0,0413)	0,9193 (0,0112)
SVM Polinomial	Exp, = 1; C = 0,01	42,6872 (1,9887)	0,2836 (0,0248)	0,3052 (0,0537)	0,9038 (0,0275)
	Exp, = 1; C = 0,1	45,9231 (1,9547)	0,3240 (0,0244)	0,4321 (0,0417)	0,8235 (0,0202)
	Exp, = 1; C = 1,0	41,1907 (1,5948)	0,2915 (0,0183)	0,4315 (0,0382)	0,7970 (0,0159)
	Exp, = 1; C = 10	46,6457 (1,7393)	0,3411 (0,0204)	0,4479 (0,0407)	0,8131 (0,0160)
	Exp, = 1; C = 100	51,2146 (1,6908)	0,3894 (0,0203)	0,4979 (0,0428)	0,8213 (0,0143)
	Exp, = 2; C = 0,01	44,1254 (1,8700)	0,3016 (0,0234)	0,4226 (0,0458)	0,8210 (0,0309)
	Exp, = 2; C = 0,1	48,1692 (2,0428)	0,3521 (0,0255)	0,4431 (0,0434)	0,8290 (0,0202)
	Exp, = 2; C = 1,0	47,5197 (1,6959)	0,3594 (0,0196)	0,4950 (0,0444)	0,8370 (0,0146)
	Exp, = 2; C = 10	57,3177 (1,7360)	0,4630 (0,0207)	0,5756 (0,0457)	0,8688 (0,0140)
	Exp, = 2; C = 100	64,7347 (1,6923)	0,5489 (0,0207)	0,6397 (0,0460)	0,8761 (0,0127)

Continua

Cont. Tabela 3

SVM Polinomial	Exp, = 3; *C* = 0,01	45,4884 (1,9450)	0,3186 (0,0243)	0,4798 (0,0444)	0,7854 (0,0303)
	Exp, = 3; *C* = 0,1	51,7789 (1,9082)	0,3972 (0,0238)	0,4683 (0,0432)	0,8460 (0,0175)
	Exp, = 3; *C* = 1,0	52,2405 (1,6834)	0,4106 (0,0196)	0,5488 (0,0474)	0,8564 (0,0144)
	Exp, = 3; *C* = 10	61,5045 (1,6317)	0,5120 (0,0196)	0,5957 (0,0439)	0,8808 (0,0130)
	Exp, = 3; *C* = 100	69,2229 (1,6194)	0,6037 (0,0201)	0,6854 (0,0427)	0,8881 (0,0131)
SVM PUK	*C* = 0,01	45,8580 (1,9818)	0,3232 (0,0248)	0,3942 (0,0403)	0,8520 (0,0179)
	C = 0,1	55,2260 (2,1235)	0,4403 (0,0265)	0,4148 (0,0419)	0,9064 (0,0125)
	C = 1,0	62,5940 (1,6309)	0,5251 (0,0199)	0,6339 (0,0433)	0,8889 (0,0132)
	C = 10	73,4284 (1,4993)	0,6549 (0,0191)	0,7367 (0,0403)	0,9063 (0,0113)
	C = 100	**79,2286** (1,3304)	**0,7262** (0,0174)	**0,7790** (0,0378)	**0,9231** (0,0111)
SVM RBF (*g* = 0,01)	*C* = 0,01	20,1364 (1,3288)	0,0036 (0,0166)	0,2610 (0,4391)	0,7357 (0,4379)
	C = 0,1	29,8434 (2,1062)	0,1248 (0,0263)	0,3751 (0,4686)	0,6763 (0,4117)
	C = 1,0	33,3333 (1,4028)	0,2098 (0,0162)	0,3913 (0,0395)	0,8148 (0,0195)
	C = 10	38,3751 (1,5596)	0,2597 (0,0178)	0,4381 (0,0414)	0,7883 (0,0174)
	C = 100	43,7155 (1,6017)	0,3139 (0,0185)	0,4381 (0,0402)	0,8081 (0,0162)
Redes Bayesianas	Padrão	41,7462 (1,7138)	0,2823 (0,0199)	0,4080 (0,0412)	0,8017 (0,0191)
Naïve Bayes	Padrão	36,7521 (1,5392)	0,2438 (0,0173)	0,4641 (0,0404)	0,7500 (0,0198)

Abreviações *C*: parâmetro de complexidade; Exp.: expoente; *g*: parâmetro *gamma*; L: taxa de aprendizado (*learning rate*); MLP: *Multilayer Perceptron*; PUK: *Pearson Universal VII Kernel*; RBF: *Radial Basis Function*; RRSE: *Root Relative Squared Error*; SVM: *Support Vector Machines*. **Fonte:** Elaborado pela autora (2021).

Por meio da análise da Tabela 3 acima, é possível observar a presença de diferenças significativas entre os desempenhos dos classificadores, de acordo com o modelo e a

configuração de hiperparâmetros utilizada. Em geral, percebe-se uma tendência de resultados superiores a partir das configurações mais complexas dos algoritmos testados. No primeiro conjunto de experimentos, o algoritmo SVM PUK com parâmetro de complexidade C = 100 obteve a melhor performance, com acurácia média igual a 79,2286% (± 1,3304), sensibilidade e especificidade médias de 0,7790 (± 0,0378) e 0,9231 (± 0,0111), respectivamente. Seu índice kappa médio obtido foi de 0,7262 (± 0,0174), correspondendo a um nível de concordância substancial entre as diferentes iterações do experimento (VIERA; GARRETT, 2005). O algoritmo SVM PUK proporcionou, portanto, alto poder preditivo e alta concordância entre as diferentes etapas dos experimentos de validação (*folds*). Além disso, a presença de baixa variância indica que o modelo apresenta baixo risco de *overfitting* com essa configuração (VABALAS *et al.*, 2019).

Um desempenho próximo ao acima descrito foi atingido pelo modelo *Random Forest* em sua configuração padrão, que corresponde à arquitetura com 100 árvores (ou iterações). Sua acurácia média obtida foi de 76,4293% (± 1,4698), com sensibilidade e especificidades médias de 0,7540 (± 0,0388) e 0,9234 (± 0,0108), respectivamente. Também foi observado um nível de concordância substancial entre as iterações desse experimento, evidenciado pelo coeficiente kappa médio acima de 0,6 (0,6883 ± 0,0192) com baixa variância (VIERA; GARRETT, 2005), demonstrando menor risco de *overfitting* que o modelo SVM PUK nesse experimento.

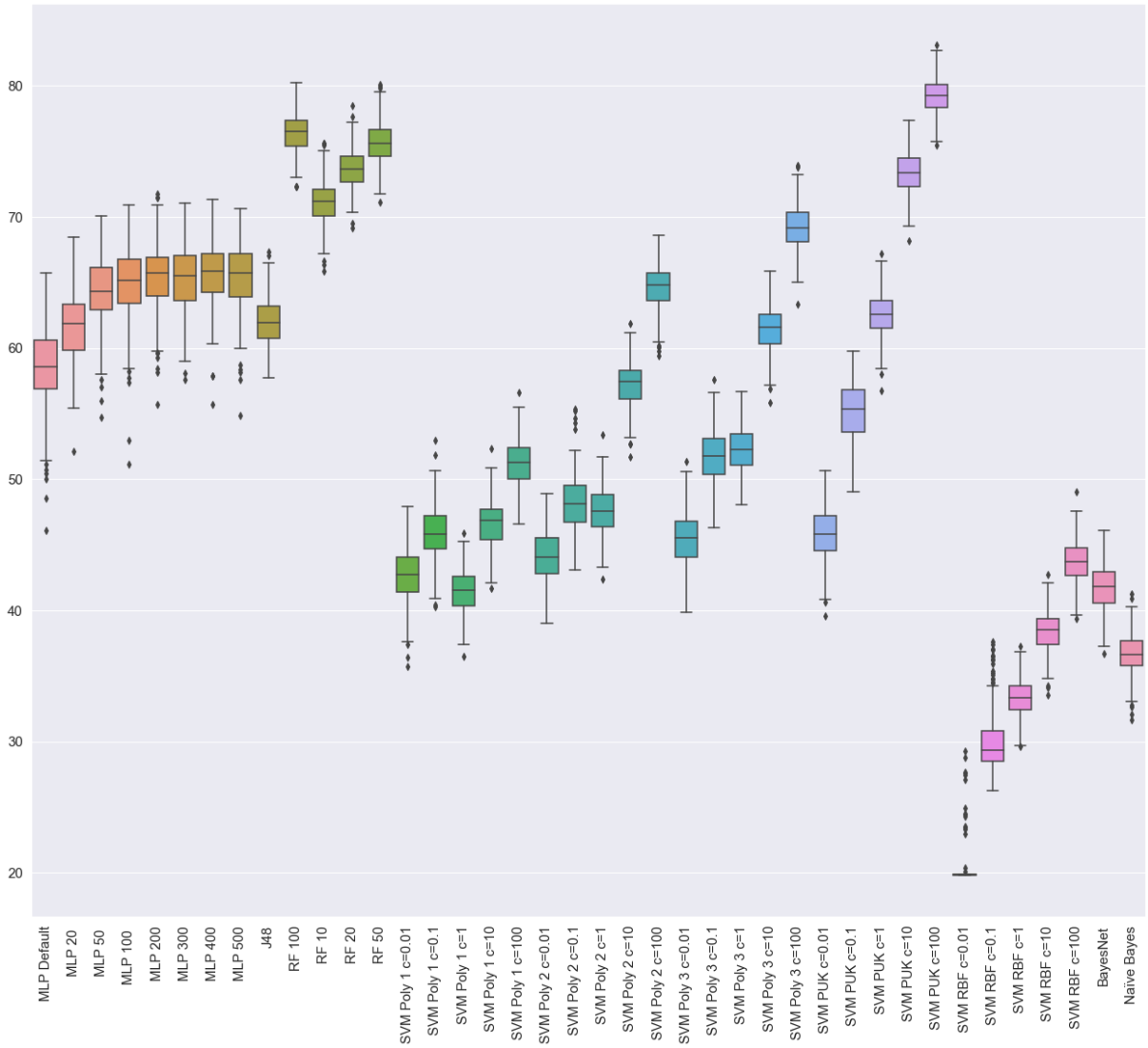
Por outro lado, o modelo SVM com *kernel* RBF obteve o pior desempenho entre todos os algoritmos testados, com acurácias médias de 20,1% e 29,8% para as configurações C = 0,01 e C = 0,1, respectivamente, resultados próximos àqueles obtidos pelo acaso. Além disso, seus desempenhos não foram consistentes, pois há grande variância na sensibilidade e especificidade médias desse modelo, e seu índice kappa médio demonstra baixa concordância entre as iterações (VIERA; GARRETT, 2005). Uma provável explicação para esses pobres resultados é o fato de o protocolo de experimentos não ter incluído a variação o parâmetro *gamma*, mantendo-o fixo em seu valor padrão (g = 0,01), comprometendo a otimização das fronteiras de decisão desse modelo para a base de dados desta pesquisa.

Ainda em relação aos modelos SVM, observa-se também grande variação do desempenho dos classificadores com base no *kernel* utilizado e em suas respectivas configurações. Por exemplo, as acurácias dos *kernels* polinomiais variaram amplamente, desde 41% (exp. = 1; C = 1,0) até 69%. (exp. = 3; C = 100). Considerando a natureza desses experimentos com o envolvimento de cinco classes, ainda assim esses resultados foram significativamente superiores ao acaso, que equivale a uma probabilidade de 20% de acerto. Por fim, percebe-se novamente que as configurações mais complexas desses algoritmos

forneceram os melhores resultados, provavelmente por proporcionarem uma modelagem mais adequada a este complexo problema de classificação com múltiplas classes.

A Figura 8 abaixo exibe, de maneira mais detalhada, uma análise gráfica comparativa entre os desempenhos dos classificadores em suas diversas configurações, utilizando a distribuição de suas acurácias em *boxplots* para as 30 iterações do experimento.

Figura 8 – Gráfico de *boxplots* da distribuição das acurácias dos classificadores no conjunto de experimentos com o balanceamento *ClassBalancer*



A Figura 8 exibe um gráfico de *boxplots* onde o eixo horizontal corresponde às configurações dos classificadores, e o eixo vertical às acurácias. Nota: *C*, parâmetro de complexidade; J48, Árvores de decisão do modelo J48; MLP, *Multilayer Perceptron*; Poly, *kernel* polinomial; PUK, *Pearson Universal VII Kernel*; RBF, *Radial Basis Function* RF, *Random Forest*; SVM, *Support Vector Machines*. Os números nos modelos MLP e RF correspondem, respectivamente, à quantidade de neurônios na camada intermediária e ao número de iterações. **Fonte:** Elaborado pela autora (2021).

O gráfico na Figura 8 acima nos mostra mais nitidamente a ampla dispersão dos valores de acurácia entre os diversos classificadores, com suas configurações mais complexas exibindo maiores acurácias e menor dispersão de seus resultados. Uma exceção é observada para o MLP com arquitetura de 500 neurônios na camada intermediária, que obteve desempenho inferior ao da maioria das configurações mais simples desse modelo.

Em relação aos algoritmos SVM, os modelos com o maior valor do parâmetro *C* de complexidade (*C* = 100) apresentaram tendência à menor dispersão dos resultados em comparação àqueles com o valor mínimo desse parâmetro (*C* = 0,01). Esse comportamento já era esperado, uma vez que o parâmetro *C* determina uma maior penalidade para erros, tornando a fronteira de decisão mais curva e rígida (*hard margin*), porém às custas de um maior risco de *overfitting* (BEN-HUR; WESTON, 2010; PRADHAN, 2012). Já para os *kernels* polinomiais, o aumento do grau da função *kernel* também promoveu uma melhor modelagem do algoritmo à base de dados. Isso pode ser constatado pelas acurácias superiores dos *kernels* de 3º grau, seguidos, pelos de 2º e, em último lugar, os de 1º grau, para os mesmos valores de *C*. Por fim, as performances dos *kernels* gaussianos (RBF) parecem ter sido prejudicadas pela manutenção do hiperparâmetro *gamma* em um valor baixo e fixo; isso provavelmente tornou esses modelos excessivamente simples, não permitindo uma modelagem otimizada aos dados, com consequente *underfitting* (BEN-HUR; WESTON, 2010).

A Tabela 4 abaixo exibe a matriz de confusão da melhor configuração para o primeiro conjunto de experimentos (SVM PUK, *C* = 100) em um único experimento realizado como exemplo.

Tabela 4 – Matriz de confusão para o modelo computacional de maior performance (SVM PUK; *C* = 100) com o balanceamento *ClassBalancer* do Weka®

	Classificado como controle	**Classificado como TDM**	**Classificado como SCZ**	**Classificado como TB**	**Classificado como TAG**
Controle	**1075,06 (77,02%)**	109,29 (7,83%)	83,15 (5,96%)	41,58 (2,98%)	86,72 (6,21%)
Depressão Maior (TDM)	151,74 (10,87%)	**1082,8 (77,58%)**	90,41 (6,48%)	34,89 (2,50%)	35,95 (2,58%)
Esquizofrenia (SCZ)	73,41 (5,26%)	110,12 (7,89%)	**1103,09 (79,03%)**	80,94 (5,80%)	28,24 (2,02%)

Continua

Cont. Tabela 4

Transtorno Bipolar (TB)	39,35 (2,82%)	56,71 (4,06%)	105,32 (7,55%)	**1181,68 (84,66%)**	12,73 (0,91%)
Transtorno de Ansiedade Generalizada	138,11 (9,89%)	85,22 (6,11%)	38,2 (2,74%)	32,32 (2,32%)	**1101,95 (78,95%)**

Para cada célula, o valor superior corresponde ao número total de instâncias; o valor inferior exibe o valor relativo de cada instância classificada em determinada categoria. Abreviações: *C* = parâmetro de complexidade; SCZ, esquizofrenia; TAG, Transtorno de Ansiedade Generalizada; TB, Transtorno Bipolar; TDM, Transtorno Depressivo Maior. Fonte: Elaborado pela autora (2021).

A matriz de confusão na Tabela 4 acima nos mostra que o desempenho do modelo SVM PUK foi semelhante para as classes Controle e Depressão Maior, com acurácias de 77,02% e 77,58%, respectivamente. Outrossim, as taxas de discriminação entre Esquizofrenia e TAG também foram equivalentes, com as respectivas acurácias de 79,03% e 78,95%. Dentre as cinco classes, a maior performance classificatória foi alcançada para o Transtorno Bipolar, classificando corretamente 84,66% das instâncias. Por outro lado, altas taxas de confusão foram observadas entre os seguintes grupos: Controle e Depressão Maior, com 10,87% das amostras de indivíduos deprimidos classificadas como controles saudáveis; e TAG e Controle, no qual 9,89% das amostras de pacientes com TAG foram classificados como controles.

Por meio da análise da matriz de confusão, observa-se que o desempenho para a classificação das amostras entre depressão, esquizofrenia, TAG e controle apresentou somente uma discreta variação de 2,01%. Entretanto, as taxas de discriminação para o TB foram consideravelmente mais elevadas. Essa diferença possivelmente se deve ao fato de os pacientes desse grupo apresentarem sintomas mais graves e, portanto, alterações mais intensas nos atributos vocais. Enquanto isso, os pacientes dos outros grupos de transtornos apresentavam quadro clínico de moderada gravidade e, dessa forma, alterações mais sutis nos parâmetros acústicos. Outra hipótese é que as alterações nos atributos acústicos encontradas em pacientes bipolares, como a maior variabilidade do *pitch* e o aumento da intensidade/volume, seriam "únicas", ou seja, exclusivas desse transtorno e consequentemente, mais facilmente distinguíveis. Por outro lado, certos grupos diagnósticos compartilham algumas características acústicas; por exemplo, a redução na variabilidade do *pitch* é uma alteração frequente tanto na depressão quanto na esquizofrenia e poderia ser um fator de confusão para os classificadores automatizados.

As maiores taxas de confusão foram observadas entre as classes depressão e controle, e acredita-se que esse erro de classificação se deve à inclusão de indivíduos com quadro depressivo leve no grupo depressão. Admitindo-se que sintomas depressivos leves produzam

apenas alterações sutis nos parâmetros vocais, o reconhecimento desses quadros se tornaria mais difícil. Altas taxas de confusão também foram observadas entre as classes Controle e TAG, porém ligeiramente menores. Nossa hipótese para esse achado é que os sintomas de ansiedade generalizada não provocariam tantas alterações nos atributos vocais extraídos quanto os outros transtornos abordados neste trabalho. Entretanto, deve-se ressaltar que o reduzido tamanho amostral deste grupo constitui uma importante limitação para essa análise.

Como visto no capítulo anterior, o *ClassBalancer* é um método de balanceamento bastante simples. Considerando a complexidade desta base de dados – multiclasses e com forte desbalanceamento entre elas – é plausível que esse método não forneça o balanceamento de classes desejado. Para contornar esse obstáculo, foi realizada uma segunda rodada de experimentos com os mesmos classificadores, porém com o método consagrado de *resampling* denominado SMOTE. Devido ao considerável custo computacional, durante a segunda rodada de experimentos com o método SMOTE, foi aplicada uma técnica de *resampling* para a redução proporcional da base de dados para 25% de seu tamanho original. Tal abordagem é considerada válida, pois mantém o mesmo comportamento estatístico da base original enquanto reduz o seu tamanho, diminuindo o tempo necessário para a conclusão dos experimentos sem interferir nos seus resultados (YILDIRIM; ÖZDOĞAN; WATSON, 2016). Esse processo gerou um total aproximado de 3300 instâncias, igualmente distribuídas entre as cinco classes. Abaixo a Tabela 5 exibe de maneira detalhada os valores médios de desempenho desses experimentos, após 30 iterações e utilizando-se as mesmas métricas dos experimentos anteriores. Similarmente, o método de validação empregado também foi o de validação cruzada com 10 *folds*.

Tabela 5 – Desempenhos médios dos modelos computacionais para classificação após balanceamento de classes pelo método SMOTE com *resampling* da base de dados para 25% do tamanho original

Classificador	**Configuração**	**Acurácia (%) (DP)**	**Índice Kappa (DP)**	**Sensibilidade (DP)**	**Especificidade (DP)**
MLP	Padrão (L:0,3)	62,8720 (3,0108)	0,5359 (0,0376)	0,5723 (0,1016)	0,8855 (0,0451)
MLP	H: 20; L:0,1	64,6809 (2,9180)	0,5585 (0,0365)	0,6063 (0,0964)	0,8914 (0,0381)
	H: 50; L:0,1	66,8442 (2,9514)	0,5855 (0,0369)	0,6600 (0,0974)	0,8959 (0,0432)
	H: 100; L:0,1	67,5019 (2,7194)	0,5938 (0,0340)	0,6655 (0,0959)	0,9009 (0,0389)

Continua

Cont. Tabela 5

MLP	H: 200; L:0,1	67,8234 (2,8555)	0,5978 (0,0357)	0,6791 (0,0993)	0,8985 (0,0412)
	H: 300; L:0,1	67,9750 (2,7310)	0,5997 (0,0341)	0,6743 (0,0990)	0,9007 (0,0380)
	H:400; L:0,1	67,7738 (2,7396)	0,5972 (0,0343)	0,6731 (0,0931)	0,8984 (0,0372)
	H: 500; L:0,1	68,0528 (2,7464)	0,6001 (0,0343)	0,6786 (0,0981)	0,8981 (0,0387)
Árvores de Decisão (J48)	Padrão	68,9035 (2,7802)	0,6113 (0,0348)	0,6560 (0,0614)	0,9149 (0,0177)
Random Forest	Padrão (100 árvores)	**81,4538** (2,1476)	**0,7682** (0,0268)	**0,8203** (0,0439)	**0,9408** (0,0150)
	10 árvores	77,0606 (2,2431)	0,7133 (0,0280)	0,7930 (0,0496)	0,9168 (0,0174)
	20 árvores	79,3713 (2,2093)	0,7421 (0,0276)	0,8102 (0,0461)	0,9291 (0,0168)
	50 árvores	80,8597 (2,1320)	0,7607 (0,0267)	0,8162 (0,0442)	0,9379 (0,0148)
SVM Polinomial	Exp, = 1; *C* = 0,01	37,1261 (3,2020)	0,2141 (0,0400)	0,4368 (0,2870)	0,7394 (0,2175)
	Exp, = 1; *C* = 0,1	44,5346 (2,4194)	0,3067 (0,0302)	0,3961 (0,0575)	0,8321 (0,0248)
	Exp, = 1; *C* = 1,0	48,66 (2,4792)	0,3582 (0,0310)	0,3954 (0,0573)	0,8480 (0,0226)
	Exp, = 1; *C* = 10	52,7471 (2,5691)	0,4093 (0,0321)	0,4145 (0,0574)	0,8547 (0,0224)
	Exp, = 1; *C* = 100	55,3803 (2,5437)	0,4422 (0,0318)	0,4954 (0,0548)	0,8389 (0,0242)
	Exp, = 2; *C* = 0,01	42,8059 (2,3920)	0,2851 (0,0299)	0,3852 (0,0648)	0,8375 (0,0404)
SVM Polinomial	Exp, = 2; *C* = 0,1	46,3747 (2,6981)	0,3297 (0,0337)	0,4297 (0,0577)	0,8122 (0,0289)
	Exp, = 2; *C* = 1,0	57,7165 (2,4270)	0,4340 (0,0303)	0,5055 (0,0557)	0,8519 (0,0210)
	Exp, = 2; *C* = 10	61,9347 (2,3146)	0,5242 (0,0289)	0,5658 (0,0613)	0,8822 (0,0191)
	Exp, = 2; *C* = 100	67,8881 (2,4292)	0,5986 (0,0304)	0,6279 (0,0573)	0,8943 (0,0180)

Continua

Cont. Tabela 5

SVM Polinomial	Exp, = 3; *C* = 0,01	43,8405 (2,6028)	0,2980 (0,0325)	0,4631 (0,0596)	0,7751 (0,0476)
	Exp, = 3; *C* = 0,1	50,8900 (2,4751)	0,3861 (0,0309)	0,4796 (0,0591)	0,8358 (0,0243)
	Exp, = 3; *C* = 1,0	58,5519 (2,3210)	0,4819 (0,0290)	0,5565 (0,0571)	0,8693 (0,0203)
	Exp, = 3; *C* = 10	65,6955 (2,3435)	0,5712 (0,0293)	0,5972 (0,0620)	0,8943 (0,0186)
	Exp, = 3; *C* = 100	71,1202 (2,5662)	0,6390 (0,0321)	0,6788 (0,0572)	0,8999 (0,0178)
SVM PUK	*C* = 0,01	39,4301 (2,7127)	0,2428 (0,0339)	0,4264 (0,2426)	0,7511 (0,1645)
	C = 0,1	52,0822 (2,4942)	0,4010 (0,0312)	0,4154 (0,0566)	0,8910 (0,0195)
	C = 1,0	66,6028 (2,5253)	0,5825 (0,0316)	0,6414 (0,0587)	0,8945 (0,0191)
	C = 10	76,7018 (2,3001)	0,7075 (0,0288)	0,7821 (0,0500)	0,9116 (0,0186)
	C = 100	**80,6920** (2,2825)	**0,7586** (0,0285)	**0,8078** (0,0476)	**0,9326** (0,0153)
SVM RBF (*g* = 0,01)	*C* = 0,01	31,7967 (2,0374)	0,1475 (0,0256)	0,3198 (0,0507)	0,8872 (0,0208)
	C = 0,1	31,1806 (3,1058)	0,1398 (0,0389)	0,4141 (0,2353)	0,7800 (0,2167)
	C = 1,0	42,1099 (2,3150)	0,2764 (0,0289)	0,3157 (0,0569)	0,8899 (0,0262)
	C = 10	45,3884 (2,5369)	0,3174 (0,0317)	0,4120 (0,0576)	0,8188 (0,0257)
SVM RBF (*g* = 0,01)	*C* = 100	50,9042 (2,4131)	0,3863 (0,0302)	0,3960 (0,0551)	0,8621 (0,0212)
Redes Bayesianas	Padrão	50,2637 (2,4448)	0,3783 (0,0306)	0,4204 (0,0529)	0,8534 (0,0242)
Naïve Bayes	Padrão	45,6481 (2,5103)	0,3206 (0,0314)	0,4913 (0,0582)	0,7810 (0,0281)

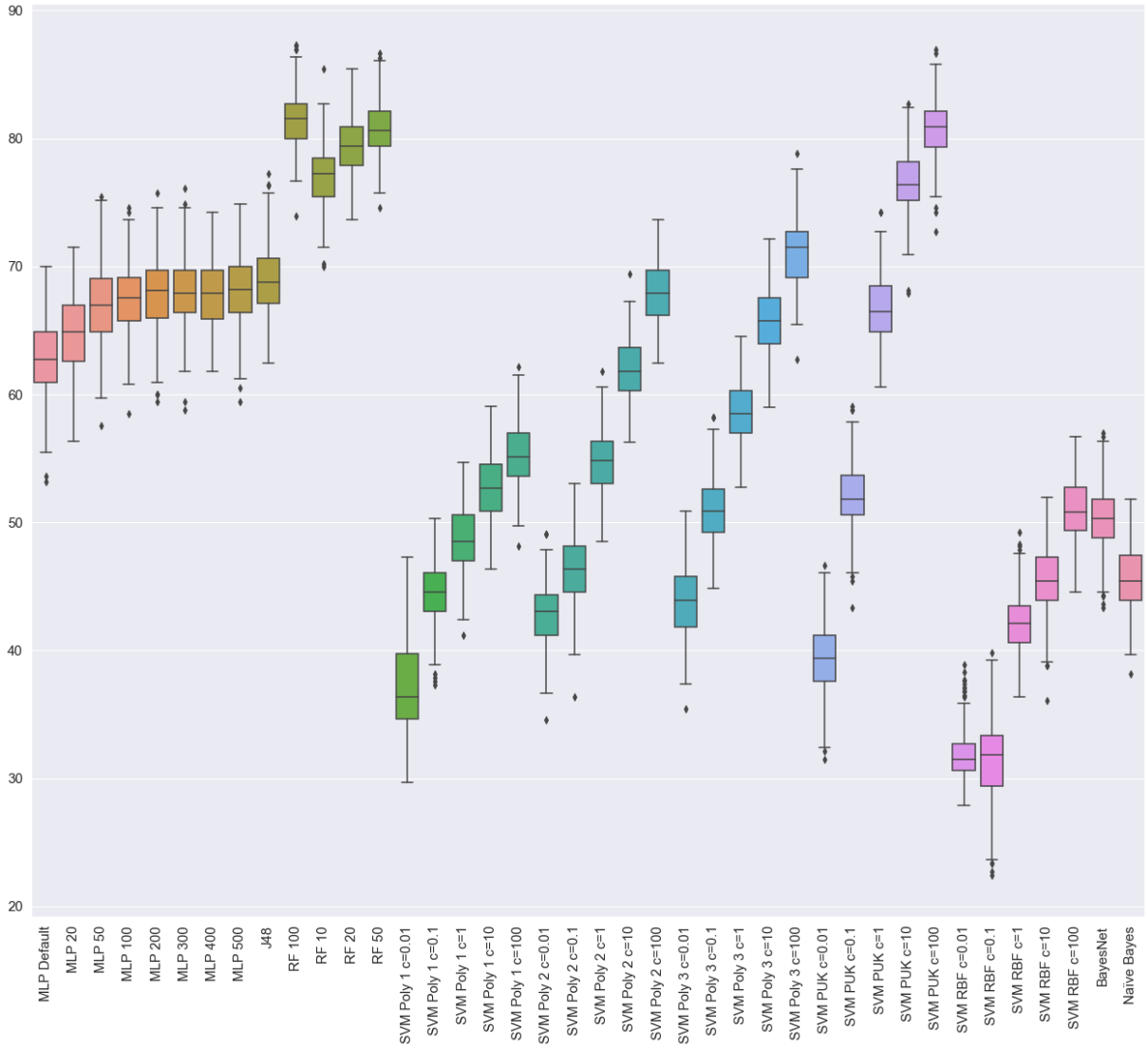
Abreviações*: C*: parâmetro de complexidade; Exp.: expoente; *g*: parâmetro *gamma*; L: taxa de aprendizado (*learning rate*); MLP, *Multilayer Perceptron*; PUK, *Pearson Universal VII Kernel*; RBF, *Radial Basis Function*; RMSE, *Root Mean Squared Error*; SMOTE, *Synthetic Minority Oversampling Technique*; SVM, *Support Vector Machines.* Fonte: Elaborado pela autora (2021).

Os resultados do segundo conjunto de experimentos realizados com o balanceamento SMOTE na Tabela 5 acima mostram uma tendência de aumento global dos desempenhos dos classificadores MLP, RF, SVM RBF e árvores de decisão (J48) para todas as configurações testadas. Em relação aos outros modelos SVM, o comportamento dos classificadores divergiu: os *kernels* polinomiais e PUK exibiram piora da performance para menores valores do parâmetro de complexidade (C = 0,01 e C = 0,1); enquanto isso, para maiores valores desse parâmetro, houve um incremento em seus desempenhos em comparação com o conjunto de experimentos anterior. Apesar da melhora global dos resultados do SVM RBF *kernel*, estes ainda se mantiveram aquém daqueles dos outros algoritmos SVM, demonstrando não ser um modelo adequado com o valor de *gamma* adotado para esta base de dados, independentemente do balanceamento de classes utilizado.

Nos experimentos com o método SMOTE, também houve novamente ampla variação dos desempenhos dos classificadores, porém em uma faixa mais estreita que no primeiro conjunto de experimentos (31%-81% versus 20%-79%, respectivamente). Os modelos SVM PUK e RF continuaram a apresentar desempenhos de classificação superiores aos demais para a base de dados desta pesquisa, além de um discreto incremento em sua performance em relação ao primeiro conjunto de experimentos. Conforme observado na Tabela 5 acima, o modelo RF com configuração padrão (100 iterações) obteve os melhores desempenhos, com valores médios de acurácia iguais a 81,4538% (± 2,1476%), sensibilidade de 0,8203 (± 0,0439) e especificidade de 0,9408 (± 0,0150). Por último, seu índice kappa médio foi de 0,7682 (± 0,0268), indicando concordância substancial entre as diferentes iterações (VIERA; GARRETT, 2005).

O modelo SVM PUK demonstrou ser o segundo melhor, novamente com a configuração de C = 100, alcançando os seguintes valores médios para as métricas estatísticas avaliadas: acurácia igual a 80,6920% (± 2,2825%); sensibilidade e especificidade de 0,8078 (± 0,0476) 0,9326 (± 0,0153), respectivamente; e índice kappa igual a 0,7586 (± 0,0285). Como os modelos SVM PUK e RF obtiveram os melhores desempenhos em ambos os conjuntos de experimentos, pode-se deduzir que estes forneceram uma modelagem mais consistente dos dados por meio dos atributos acústicos extraídos, independentemente do balanceamento de classes empregado. A visualização da Figura 9 abaixo permite uma análise comparativa mais detalhada com base nos gráficos de *boxplots* da distribuição das acurácias entre as iterações desse conjunto de experimentos.

Figura 9 – Gráfico de *boxplots* da distribuição das acurácias dos classificadores no conjunto de experimentos com o balanceamento SMOTE



A Figura 9 exibe um gráfico de *boxplots* onde o eixo horizontal corresponde às configurações dos classificadores, enquanto o eixo vertical às acurácias. Nota: *C*, parâmetro de complexidade; J48, Árvores de decisão do modelo J48; MLP, *Multilayer Perceptron*; Poly, *kernel* polinomial; PUK, *Pearson Universal VII Kernel*; RBF, *Radial Basis Function* RF, *Random Forest*; SVM, *Support Vector Machines*. Os números nos modelos MLP e RF correspondem, respectivamente, à quantidade de neurônios na camada intermediária e ao número de iterações. **Fonte:** Elaborado pela autora (2021).

Por meio do gráfico de *boxplots* da Figura 9 acima, observa-se um aumento global das acurácias para a maioria dos classificadores, com três configurações ultrapassando 80% (RF com 50 e 100 iterações e SVM PUK com C = 100), enquanto no experimento anterior esse valor de desempenho não foi alcançado por nenhum modelo. Também se evidencia o mesmo padrão de desempenho dos classificadores quanto à complexidade de suas configurações, com modelos mais complexos apresentando valores superiores de acurácia e com menor dispersão de seus resultados. Uma exceção é o modelo MLP, no qual, ao contrário do conjunto anterior

de experimentos, o aumento do número de neurônios na camada intermediária acima de 100 neurônios não trouxe benefícios significativos à performance do classificador.

A Tabela 6 abaixo expõe a matriz de confusão de uma iteração do experimento com o balanceamento SMOTE para a configuração com os melhores resultados (RF com 100 iterações).

Tabela 6 – Matriz de confusão para o modelo computacional com melhor performance (*Random Forest*) após balanceamento pelo método SMOTE, com *resampling* da base de dados para 25% do tamanho original

	Classificado como controle	**Classificado como TDM**	**Classificado como SCZ**	**Classificado como TB**	**Classificado como TAG**
Controle	**546 (82,73%)**	31 (4,70%)	42 (6,36%)	7 (1,06%)	34 (5,15%)
Depressão Maior (TDM)	79 (11,97%)	**461 (69,85%)**	65 (9,85%)	23 (3,48%)	32 (6,94%)
Esquizofrenia (SCZ)	44 (6,67%)	40 (6,06%)	**525 (79,55%)**	36 (5,45%)	14 (2,12%)
Transtorno Bipolar (TB)	13 (1,97%)	9 (1,36%)	53 (8,03%)	**576 (87,27%)**	9 (1,36%)
Transtorno de Ansiedade Generalizada	21 (3,18%)	14 (2,12%)	9 (1,36%)	6 (0,91%)	**610 (92,42%)**

Para cada célula, o valor superior corresponde ao número total de instâncias; o valor inferior exibe o valor relativo de cada instância classificada em determinada categoria. Abreviações: SCZ, esquizofrenia; SMOTE, *Synthetic Minority Oversampling Technique*; TAG, Transtorno de Ansiedade Generalizada; TB, Transtorno Bipolar; TDM, Transtorno Depressivo Maior. **Fonte:** Elaborado pela autora (2021).

A matriz de confusão da Tabela 6 acima nos mostra um considerável poder discriminatório do modelo RF para as cinco classes, porém com variações significativas entre estas, com acurácias oscilando desde abaixo de 70% até acima de 90%. Seu melhor desempenho foi alcançado nos exemplos da classe TAG, com 92,42% de acurácia. Por outro lado, sua pior performance foi observada na classe Depressão Maior, cujas instâncias só foram corretamente identificadas em 69,85% dos casos, com uma queda considerável em relação ao primeiro conjunto de experimentos. Desempenhos intermediários, medidos por meio da acurácia, foram atingidos para as classes Esquizofrenia (79,55%), Controle (82,73%) e Transtorno Bipolar (87,57%), todos ainda assim em patamares equivalentes ou superiores àqueles previamente relatados na literatura.

Por meio da análise da matriz de confusão acima, ainda é possível observar que as maiores taxas de erros residiram na diferenciação entre as classes Depressão Maior,

Esquizofrenia e Controle para amostras de indivíduos deprimidos. Nesse experimento, o modelo RF classificou incorretamente 11,97% das instâncias de depressão como controle e 9,85% como esquizofrenia. Assim como no primeiro conjunto de experimentos, uma provável explicação para o primeiro resultado foi a inclusão nesta pesquisa de indivíduos com quadros leves de depressão, fato este que deve ter dificultado a distinção entre estes e os indivíduos saudáveis. Quanto ao segundo achado, tanto a depressão maior quanto a esquizofrenia compartilham sintomas afetivos e volitivos com manifestações em sua expressão vocal, como apatia, desinteresse, discurso monótono e hipofonia, conforme discorrido nos experimentos anteriores (BEHLAU, 2001; SADOCK; SADOCK; RUIZ, 2017).

Considera-se que um maior número de categorias aumenta a complexidade do sistema de classificação, diminuindo, consequentemente, a performance dos classificadores automatizados. Entretanto, ainda assim nossos resultados são superiores à maioria dos estudos já publicados, os quais foram realizados com classificação simples binária. Este fato demonstra a robustez da solução proposta neste trabalho para a identificação de transtornos mentais simultaneamente. Até onde se sabe, a relevância deste trabalho é singular e inédita, uma vez que nenhum estudo prévio utilizou classificadores automatizados para a classificação simultânea de diversos transtornos mentais, assim como nenhum estudo abarcou essa diversidade de transtornos anteriormente. O alto poder discriminatório demonstrado pela ferramenta de solução elaborada nesta pesquisa favorece a utilização de modelos de aprendizado de máquina como instrumentos de auxílio diagnóstico e até de triagem para diferentes transtornos mentais.

6 CONCLUSÃO

Neste trabalho, foram avaliados atributos de acústica vocal como biomarcadores de auxílio diagnóstico de quatro transtorno mentais – depressão maior, transtorno bipolar, esquizofrenia e transtorno de ansiedade generalizada – utilizando classificadores de aprendizado de máquina. Para tanto, foram elaborados um protocolo de coleta de dados com foco em ambientes naturalísticos e um *framework* para a extração de atributos acústicos da fala e de classificação de padrões, baseando-se na vasta expertise do Grupo de Pesquisas em Computação Biomédica da UFPE e em referências bibliográficas relevantes da literatura.

6.1 PRINCIPAIS CONTRIBUIÇÕES

Neste trabalho, foi desenvolvido um método de pesquisa completamente novo para a classificação de padrões de acústica vocal em diferentes transtornos mentais. Primeiramente, a coleta e o pré-processamento de dados priorizaram a aquisição de dados em consultas psiquiátricas e a utilização de softwares gratuitos, permitindo aplicações no mundo real e facilitando a replicação deste estudo. Foi construída uma base de dados acústicos gerados em consultas reais de pacientes com um dos quatro transtornos mentais acima descritos e de indivíduos saudáveis, totalizando mais de 16 horas pré-edição e quase 10 horas após a edição. Essa base de dados poderá ser ampliada e integrada com bases de outros transtornos mentais para utilização futura por estudos sobre biomarcadores vocais para a detecção desses transtornos, assim como para o monitoramento longitudinal de pacientes psiquiátricos e da gravidade de seus sintomas.

Adicionalmente, a construção de uma ferramenta de detecção simultânea de quatro transtornos mentais de categorias diagnósticas distintas é inédita na literatura e favorece o desenvolvimento de futuras plataformas de triagem baseadas em biomarcadores vocais. O ineditismo deste trabalho também está presente no conjunto único de atributos selecionados para a representação dos sinais acústicos da fala utilizando a expertise do grupo de pesquisas em processamento de sinais. Por fim, a realização de experimentos exaustivos comparando diferentes métodos de balanceamento e configurações dos classificadores permitiu a identificação dos melhores modelos para as características da base de dados desta pesquisa. Por exemplo, em cada conjunto de experimentos as maiores taxas de acerto por transtorno foram: transtorno bipolar, com balanceamento *ClassBalancer* e classificador SVM; e TAG, pelo método de balanceamento SMOTE e o classificador RF. Os bons resultados classificatórios

obtidos neste trabalho para as cinco classes tornaram-se ainda melhores em situações de classificação binária, como, por exemplo, a acurácia de 87,6% para a detecção de depressão maior (ESPINOLA *et al.*, 2020b) e de 91,8% para esquizofrenia (ESPINOLA *et al.*, 2020a), ambas superiores àquelas da maioria dos estudos aqui relatados.

6.2 PUBLICAÇÕES GERADAS

Este trabalho gerou, até o momento de sua conclusão, as seguintes publicações:

- *Detection of major depressive disorder using vocal acoustic analysis and machine learning – an exploratory study. Res Biomed Eng. 2020, doi:10,1007/s42600-020-00100-9;*
- *Vocal acoustic analysis and machine learning for the identification of schizophrenia. Res Biomed Eng. 2020, doi:10,1007/s42600-020-00097-1;*
- Detecção de transtornos mentais por meio da análise computacional da voz. XXXVIII Congresso Brasileiro de Psiquiatria, 2021, Porto Alegre (em processo de impressão).

6.3 DIFICULDADES APRESENTADAS E LIMITAÇÕES

Durante a realização deste trabalho, foram encontrados diversos desafios, a maior parte destes relacionada à coleta de dados em ambientes naturalísticos. Primeiramente, por este trabalho contemplar cinco classes, foi necessário calcular um tamanho amostral inicial consideravelmente maior que o de estudos semelhantes envolvendo um número menor de classes (RAPCAN *et al.*, 2010; ALGHOWINEM *et al.*, 2013; SCHERER *et al.*, 2013; HIGUCHI *et al.*, 2018, 2019; TAHIR *et al.*, 2019). Adicionalmente, como todas as coletas foram padronizadas com a utilização do mesmo modelo de gravador para evitar diferenças acústicas entre as gravações devido a diferenças de hardware entre aparelhos distintos, havia somente um gravador disponível para as coletas. Esta situação, somada ao fato de as gravações geralmente serem longas por contemplarem a duração de consultas inteiras, limitou significativamente a velocidade da coleta de dados. Além disso, coletar dados em enfermarias psiquiátricas de pacientes portadores de transtornos mentais graves em fase aguda apresentou-se bastante difícil pela própria natureza de vários sintomas psiquiátricos (e.g., desconfiança, hostilidade, beligerância, agitação psicomotora). Também foi observada dificuldade de obter o consentimento de familiares de alguns pacientes internados, possivelmente devido à

desconfiança daqueles sobre os objetivos da pesquisa. Por último, e não menos importante, o advento da pandemia de COVID-19 encerrou precocemente a etapa de coleta de dados em cerca de oito meses antes do prazo inicial.

Outra importante dificuldade diz respeito à etapa de pré-processamento com a edição dos áudios. Por se tratar de conteúdo sensível proveniente de consultas psiquiátricas reais, mais de 97% das edições foram realizadas pela própria autora, ficando o restante com outro membro do LCB, o qual não obteve acesso a elementos identificadores. Adicionalmente, devido às gravações serem frequentemente longas, o processo de edição apresentou-se ainda mais laborioso.

Dentre as limitações deste trabalho, observa-se a falta de controle para possíveis variáveis de confusão entre as classes, ocasionada pela dificuldade em recrutar participantes. Por exemplo, variáveis demográficas que podem influenciar as propriedades acústicas da fala, como idade, gênero, raça/etnia, e nível educacional, não foram controladas e podem ter interferido nos resultados obtidos. Outra limitação é o reduzido tamanho amostral das classes Transtorno Bipolar (14 indivíduos) e, principalmente, TAG, com apenas quatro participantes. Outrossim, diferenças quanto aos ambientes de coleta entre o grupo controle (locais variados) e os demais (ambientes hospitalares), pode ter interferido nos níveis de ruídos de fundo e, consequentemente, nos desempenhos dos classificadores. Por fim, fatores como tabagismo, que sabidamente afeta as características da voz, e farmacoterapia também não foram controlados e podem limitar nossos achados.

6.4 TRABALHOS FUTUROS

Os bons resultados gerados por este trabalho impulsionam a realização de novos estudos para lidar com as limitações supracitadas, por exemplo, por meio da inclusão de amostras maiores com perfil sociodemográfico equivalente e do controle de possíveis variáveis de confusão (e.g., tabagismo, farmacoterapia) para validação das análises estatísticas. Outras importantes perspectivas futuras são a investigação de outros atributos acústicos e a inclusão de atributos de expressões faciais para o aprimoramento da ferramenta de detecção. De maneira complementar, uma interessante linha de investigação também seria avaliar o possível impacto do tabagismo e de outras substâncias psicoativas sobre os padrões de acústica vocal nos diferentes transtornos mentais.

Em relação às aplicações deste trabalho, a construção de uma plataforma para uso em aplicativos móveis tornaria a ferramenta desta pesquisa mais prática e acessível e, dessa forma,

também já está sendo considerada pelo nosso grupo de pesquisas. O desenvolvimento de um aplicativo mobile para telefones celulares permitiria seu uso sob demanda por profissionais de saúde mental para fins de triagem e até para o diagnóstico diferencial de transtornos mentais. Esse aplicativo pode, inclusive, vir a ser utilizado como uma ferramenta de suporte para profissionais não especialistas e em locais longínquos sem acesso a serviços de saúde mental. Dessa maneira, tal ferramenta móvel poderia contribuir significativamente para a solução do grave problema de acesso às redes de saúde mental, especialmente nas populações mais vulneráveis e desassistidas.

REFERÊNCIAS

ABOU-WARDA, H.; BELAL, N. A.; EL-SONBATY, Y.; DARWISH, S. A random forest model for mental disorders diagnostic systems. In: HASSANIEN, A.; SHAALAN, K.; GABER, T.; AZAR, A.; TOLBA, M. (Ed.). **Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016. AISI 2016. Advances in Intelligent Systems and Computing**. Cham: Springer, 2017. 533p. 670–680.

ABRAHAM, B.; NAIR, M. S. Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. **Biocybernetics and Biomedical Engineering**, v. 40, n. 4, p. 1436–1445, 2020. Disponível em: <https://doi.org/10.1016/j.bbe.2020.08.005>.

AGUS, T. R.; SUIED, C.; THORPE, S. J.; PRESSNITZER, D. Characteristics of human voice processing. **ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems**, p. 509–512, 2010.

ALGHOWINEM, S.; GOECKE, R.; WAGNER, M.; EPPS, J. Detecting Depression: A Comparison Between Spontaneous and Read Speech. **IEEE**, p. 7547–7551, 2013.

ALGHOWINEM, S.; GOECKE, R.; WAGNER, M.; EPPS, J.; BREAKSPEAR, M.; PARKER, G. From joyous to clinically depressed: Mood detection using spontaneous speech. In: Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25, **Anais**...2012.

ALI, J.; KHAN, R.; AHMAD, N.; MAQSOOD, I. Random Forests and Decision Trees. **International Journal of Computer Science Issues**, v. 9, n. 5, p. 272–278, 2012.

ALMEIDA, A. A.; BEHLAU, M.; LEITE, J. R. Correlação entre ansiedade e performance comunicativa. **Rev. Soc. Bras. Fonoaudiol.**, v. 16, n. 4, p. 384–389, 2011.

ALONSO, S. G.; TORRE-DÍEZ, I. De; HAMRIOUI, S.; LÓPEZ-CORONADO, M.; BARRENO, D. C.; NOZALEDA, L. M.; FRANCO, M. Data Mining Algorithms and Techniques in Mental Health: A Systematic Review. **Journal of Medical Systems**, v. 42, n. 161, 2018.

ALPERT, M.; ANDERSON, L. T. Imagery mediation of vocal emphasis in flat affect. **Archives of General Psychiatry**, v. 34, n. 2, p. 208–212, 1977.

ALPERT, M.; POUGET, E. R.; SILVA, R. R. Reflections of depression in acoustic measures of the patient's speech. **Journal of Affective Disorders**, v. 66, p. 59–69, 2001.

ALPERT, M.; ROSENBERG, S. D.; POUGET, E. R.; SHAW, R. J. Prosody and lexical accuracy in flat affect schizophrenia. **Psychiatry Research**, v. 97, p. 107–118, 2000.

ALPERT, M.; SHAW, R. J.; POUGET, E. R.; LIM, K. O. A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia. **Journal of Psychiatric Research**, v. 36, n. 5, p. 347–353, 2002.

AMARAKEERTHI, S.; MORIKAWA, C.; NWE, T. L.; DE SILVA, L. C.; COHEN, M. Cascaded subband energy-based emotion classification. In: IEEJ Transactions on Electronics,

Information and Systems, 1, **Anais**...2013.

AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and Statistical Manual of Mental Disorder - Text Revision - DSM-IV-TR**. 4th. ed. [s.l: s.n.]v. 1

AMERICAN PSYCHIATRIC ASSOCIATION. **DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais**. 5. ed. Porto Alegre: Artmed, 2013.

ANDREA, M.; DIAS, Ó.; ANDREA, M.; FIGUEIRA, M. L. Functional Voice Disorders: The Importance of the Psychologist in Clinical Voice Assessment. **Journal of Voice**, v. 31, n. 4, p. 507.e13-507.e22, 2017. Disponível em: <http://dx.doi.org/10.1016/j.jvoice.2016.10.013>.

APARNA, R.; CHITHRA, P. L. Role of Windowing Techniques in Speech Signal Processing For Enhanced Signal Cryptography. In: **Advanced Engineering Research and Applications**. [s.l: s.n.]p. 446–458.

ARAR, Ö. F.; AYAN, K. A feature dependent Naive Bayes approach and its application to the software defect prediction problem. **Applied Soft Computing Journal**, v. 59, p. 197–209, 2017. Disponível em: <http://dx.doi.org/10.1016/j.asoc.2017.05.043>.

AZEVEDO, W. W.; LIMA, S. M.; FERNANDES, I. M.; ROCHA, A. D.; CORDEIRO, F. R.; DA SILVA-FILHO, A. G.; DOS SANTOS, W. P. Fuzzy morphological extreme learning machines to detect and classify masses in mammograms. In: 2015 IEEE international conference on fuzzy systems (fuzz-IEEE), August, **Anais**...2015.

BACHOROWSKI, J. A.; OWREN, M. J. Vocal expression of emotion: Acoustic Properties of Speech Are Associated With Emotional Intensity and Context. **Psychological Science**, v. 6, n. 4, p. 219–224, 1995.

BAERT, L.; THEUNISSEN, L.; VERGULT, G. (ed.). **Digital Audio and Compact Disc Technology**. [s.l.] Newnes, 2013.

BANDELA, S. R.; KUMAR, T. K. Stressed Speech Emotion Recognition using feature fusion of Teager Energy Operator and MFCC. In: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), July, **Anais**...IEEE, 2017.

BANDELOW, B.; MICHAELIS, S. Epidemiology of anxiety disorders in the 21st century. **Dialogues in Clinical Neuroscience**, v. 17, n. 3, p. 327–335, 2015.

BARROS, M. B. A.; LIMA, M. G.; AZEVEDO, R. C. S.; MEDINA, L. B. P.; LOPES, C. S.; MENEZES, P. R.; MALTA, D. C. Depression and health behaviors in Brazilian adults – PNS 2013. **Revista de Saúde Pública**, v. 51, p. 1–9, 2017.

BEDI, G.; CARRILLO, F.; CECCHI, G. A.; SLEZAK, D. F.; SIGMAN, M.; MOTA, N. B.; RIBEIRO, S.; JAVITT, D. C.; COPELLI, M.; CORCORAN, C. M. Automated analysis of free speech predicts psychosis onset in high-risk youths. **Nature Partner Journals**, 2015. Disponível em: <http://dx.doi.org/10.1038/npjschz.2015.30>.

BEDOYA-JARAMILLO, S.; BELALCAZAR-BOLAÑOS, E.; VILLA-CAÑAS, T.; OROZCO-ARROYAVE, J. R.; ARIAS-LONDOÑO, J. D.; VARGAS-BONILLA, J. F. Automatic emotion detection in speech using mel frequency cesptral coefficients. In: 2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA), Antioquia. **Anais**... Antioquia: 2012.

BEHLAU, M. **Voz: O Livro do Especialista**. Rio de Janeiro: Revinter, 2001.

BEHLAU, M.; PONTES, P.; MORETTI, F. **Higiene Vocal: Cuidando da Voz**. 5. ed. ed. Rio de Janeiro: Revinter, 2017.

BEHRMAN, A. **Speech and Voice Science**. Third ed. San Diego, CA: Plural Publishing, 2018.

BELIN, P.; ZATORRE, R. J.; LAFALLIE, P.; AHAD, P.; PIKE, B. Voice-selective areas in human auditory cortex. **Nature**, v. 403, n. 6767, p. 309–312, 2000.

BEN-HUR, A.; WESTON, J. A user's guide to support vector machines. In: **Data mining techniques for the life sciences**. [s.l.] Humana Press, 2010. p. 223–239.

BENBA, A.; JILBAB, A.; HAMMOUCH, A. Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. **International Journal of Speech Technology**, v. 19, n. 3, p. 449–456, 2016.

BENBA, A.; JILBAB, A.; HAMMOUCH, A.; SANDABAD, S. Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In: Proceedings of 2015 International Conference on Electrical and Information Technologies, ICEIT 2015, **Anais**...2015.

BERMEJO, P.; GÁMEZ, J. A.; PUERTA, J. M. Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. **Expert Systems with Applications**, v. 38, p. 2072–2080, 2011. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2010.07.146>.

BHAGYA SHREE, S. R.; SHESHADRI, H. S. An initial investigation in the diagnosis of Alzheimer ' s disease using various classification techniques. In: 2014 IEEE International Conference on Computational Intelligence and Computing Research, **Anais**...2014.

BHATTI, A. M.; MAJID, M.; ANWAR, S. M.; KHAN, B. Human emotion recognition and analysis in response to audio music using brain signals. **Computers in Human Behavior**, v. 65, p. 267–275, 2016. Disponível em: <http://dx.doi.org/10.1016/j.chb.2016.08.029>.

BIAU, G.; SCORNET, E. A random forest guided tour. **TEST**, v. 25, p. 197–227, 2016. BLAGUS, R.; LUSA, L. SMOTE for high-dimensional class-imbalanced data. **BMC Bioinformatics**, v. 14, n. 106, 2013.

BOERSMA, P. Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound. In: Proceedings of the Institute of Phonetic Sciences, **Anais**...1993. Disponível em: <http://isip.lzu.edu.cn/Members/sunny/speech-

processing/papers-on-audio-speech-language-processing/formant-extraction/Proceedings_1993.pdf>.

BOERSMA, P. Acoustic analysis. In: PODESVA, R.; SHARMA, D. (Ed.). **Research Methods in Linguistics**. [s.l.] Cambridge University Press, 2013. p. 375–398.

BOGERT, B. P.; HEALY, M. J. R.; TUKEY, J. W. The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking. In: Proceedings of the Symposium on Time Series Analysis, **Anais**...1963.

BONE, D.; GIBSON, J.; CHASPARI, T.; CAN, D.; NARAYANAN, S. Speech and Language Processing for Mental Health Research and Care. **50th Asilomar Conference on Signals, Systems and Computers**, p. 831–835, 2016.

BOUCKAERT, R. R. Bayesian Network Classifiers in Weka for Version 3-5-7. **Artificial Intelligence Tools**, v. 11, n. 3, p. 369–387, 2008.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BRUNETTE, E. S.; FLEMMER, R. C.; FLEMMER, C. L. A Review of Artificial Intelligence. In: 2009 4th International Conference on Autonomous Robots and Agents, Wellington. **Anais**... Wellington: 2009.

BUCHANAN, R. W. Persistent negative symptoms in schizophrenia: An overview. **Schizophrenia Bulletin**, v. 33, n. 4, p. 1013–1022, 2007.

BURNAEV, E.; EROFEEV, P.; PAPANOV, A. Influence of resampling on accuracy of imbalanced classification. In: Eighth international conference on machine vision (ICMV 2015), Barcelona, Spain. **Anais**... Barcelona, Spain: 2015.

BUSHE, C. J.; TAYLOR, M.; HAUKKA, J. Mortality in schizophrenia: a measurable clinical endpoint. **Journal of Psychopharmacology**, v. 24, n. 4 Suppl, p. 17–25, 2010.

BYEON, H. A Prediction Model for Mild Cognitive Impairment Using Random Forests. **International Journal of Advanced Computer Science and Applications**, v. 6, n. 12, p. 8–12, 2015.

BYLSMA, L. M.; MORRIS, B. H.; ROTTENBERG, J. A meta-analysis of emotional reactivity in major depressive disorder. **Clinical Psychology Review**, v. 28, p. 676–691, 2008.

BZDOK, D.; MEYER-LINDENBERG, A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. **Biological Psychiatry: Cognitive Neuroscience and Neuroimaging**, v. 3, p. 223–230, 2018. Disponível em: <https://doi.org/10.1016/j.bpsc.2017.11.007>.

CAMPBELL, N.; BECKMAN, M. Stress, Prominence, and Spectral Tilt. In: Intonation: Theory, Models, and Applications, Athens. **Anais**... Athens: 1997.

CANNIZZARO, M.; HAREL, B.; REILLY, N.; CHAPPELL, P.; SNYDER, P. J. Voice

acoustical measurement of the severity of major depression. **Brain and Cognition**, v. 56, p. 30–35, 2004.

CHAKRABORTY, D.; XU, S.; YANG, Z.; HAN, Y.; CHUA, V.; TAHIR, Y.; DAUWELS, J.; THALMANN, N. M.; TAN, B.; LEE, J. Prediction of Negative Symptoms of Schizophrenia from Objective Linguistic, Acoustic and Non-verbal Conversational Cues. **IEEE 2018 International Conference on Cyberworlds Prediction**, p. 280–283, 2018a.

CHAKRABORTY, D.; YANG, Z.; TAHIR, Y.; MASZCZYK, T.; DAUWELS, J.; THALMANN, N.; ZHENG, J.; MANIAM, Y.; AMIRAH, N.; TAN, B. L.; LEE, J. Prediction of negative symptoms of schizophrenia from emotion related low-level speech signals. **IEEE**, p. 6024–6028, 2018b.

CHAKRABORTY, S. Advantages of Blackman Window over Hamming Window Method for designing FIR Filter. **International Journal of Computer Science & Engineering Technology**, v. 4, n. 8, p. 1181–1189, 2013.

CHAPMAN, B. P.; WEISS, A.; DUBERSTEIN, P. R. Statistical learning theory for high dimensional prediction: Application to criterion-keyed scale development. **Psychological Methods**, v. 21, n. 4, p. 603–620, 2016.

CHARLSON, F. J.; FERRARI, A. J.; SANTOMAURO, D. F.; DIMINIC, S.; STOCKINGS, E.; SCOTT, J. G.; MCGRATH, J. J.; WHITEFORD, H. A. Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. **Schizophrenia Bulletin**, v. 44, n. 6, p. 1195–1203, 2018.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002.

CHEN, R.; HERSKOVITS, E. H. Clinical Diagnosis Based on Bayesian Classification of Functional Magnetic-Resonance Data. **Neuroinformatics**, v. 5, p. 178–188, 2007.

CHEN, Y.; YANN, M. L.; DAVOUDI, H.; CHOI, J.; AN, A.; MEI, Z. Contrast Pattern Based Collaborative Behavior Recommendation for Life Improvement. In: **Advances in Knowledge Discovery and Data Mining. PAKDD 2017. Lecture Notes in Computer Science**. [s.l: s.n.]10235p. 106–118.

CHIEN, J.-T.; HUANG, C.-H. Bayesian Learning of Speech Duration Models. In: IEEE Transactions on Speech and Audio Processing, 6, **Anais**...2003.

CHO, G.; YIM, J.; CHOI, Y.; KO, J.; LEE, S. H. Review of machine learning algorithms for diagnosing mental illness. **Psychiatry Investigation**, v. 16, n. 4, p. 262–269, 2019.

COHN, J. F.; KRUEZ, T. S.; MATTHEWS, I.; YANG, Y.; NGUYEN, M. H.; PADILLA, M. T.; ZHOU, F.; DE LA TORRE, F. Detecting depression from facial actions and vocal prosody. **Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009**, n. October, 2009.

COMMOWICK, O.; ISTACE, A.; KAIN, M.; LAURENT, B.; LERAY, F.; SIMON, M.; KERBRAT, A. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. **Scientific Reports**, v. 8, n. 13650, p. 1–17, 2018.

COMPTON, M. T.; LUNDEN, A.; CLEARY, S. D.; PAUSELLI, L.; ALOLAYAN, Y.; HALPERN, B.; BROUSSARD, B.; CRISA, A.; CAPULONG, L.; MARIA, P.; BERNARDINI, F.; COVINGTON, M. A. The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. **Schizophrenia Research**, p. 1–8, 2018.

COOK, M. Anxiety, Speech Disturbances and Speech Rate. **Br. J. Soc. Cin. Psychol.**, v. 8, p. 13–21, 1969.

CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 20, p. 273–297, 1995.

COVINGTON, M. A.; LUNDEN, S. L. A.; CRISTOFARO, S. L.; WAN, C. R.; BAILEY, C. T.; BROUSSARD, B.; FOGARTY, R.; JOHNSON, S.; ZHANG, S.; COMPTON, M. T. Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. **Schizophrenia Research**, v. 142, p. 93–95, 2012.

CRUZ, T.; CRUZ, T.; SANTOS, W. Detection and classification of lesions in mammographies using neural networks and morphological wavelets. **IEEE Latin America Transactions**, v. 16, n. 3, p. 926–932, 2018.

CUMMINS, N.; EPPS, J.; BREAKSPEAR, M.; GOECKE, R. An Investigation of Depressed Speech Detection: Features and Normalization. In: **Anais**...2011.

CUMMINS, N.; EPPS, J.; SETHU, V.; KRAJEWSKI, J. Variability compensation in small data: Oversampled extraction of i-vectors for the classification of depressed speech. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, p. 970–974, 2014.

CUMMINS, N.; SCHERER, S.; KRAJEWSKI, J.; SCHNIEDER, S.; EPPS, J.; QUATIERI, T. F. A review of depression and suicide risk assessment using speech analysis. **Speech Communication**, v. 71, n. April, p. 10–49, 2015. Disponível em: <http://dx.doi.org/10.1016/j.specom.2015.03.004>.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. In: ZHANG, C.; MA, Y. (Ed.). **Ensemble machine learning: methods and applications**. [s.l.] Springer Science & Business Media, 2012. p. 157–175.

D'ALESSANDRO, C.; DOVAL, B. Voice quality modification for emotional speech synthesis. In: EUROSPEECH 2003, **Anais**...2003.

DA SILVA JUNIOR, M.; DE FREITAS, R. C.; DOS SANTOS, W. P.; DA SILVA, W. W. A., RODRIGUES, M. C. A., CONDE, E. F. Q. Exploratory study of the effect of binaural beat stimulation on the EEG activity pattern in resting state using artificial neural networks. **Cognitive Systems Research**, v. 54, p. 1–20, 2019.

DAI, W.; XUE, G.-R.; YANG, Q.; YU, Y. Transferring Naive Bayes Classifiers for Text Classification. In: AAAI'07: Proceedings of the 22nd national conference on Artificial intelligence, **Anais**...2007.

DALGALARRONDO, P. **Psicopatologia e Semiologia dos Transtornos Mentais**. 3. ed. Porto Alegre: Artmed, 2019.

DARBY, J. K.; HOLLIEN, H. Vocal and Speech Patterns of Depressive Patiens. **Folia Phoniatrica**, v. 29, p. 279–291, 1977.

DATAR, A.; JAIN, A.; SHARMA, P. C. Performance of Blackman window family in M-channel Cosine Modulated Filter Bank for ECG signals. In: 2009 International Multimedia, Signal Processing and Communication Technologies, Aligarth. **Anais**... Aligarth: 2009.

DATTA, S.; DAS, S. Near-Bayesian Support Vector Machines for imbalanced data classification with equal or unequal misclassification costs. **Neural Networks**, v. 70, p. 39–52, 2015. Disponível em: <http://dx.doi.org/10.1016/j.neunet.2015.06.005>.

DE SANTANA, M. A.; PEREIRA, J. M. S.; DA SILVA, F. L.; DE LIMA, N. M.; DE SOUSA, F. N.; DE ARRUDA, G. M. S., ... DOS SANTOS, W. P. Breast cancer diagnosis based on mammary thermography and extreme learning machines. **Research on Biomedical Engineering**, v. 34, n. 1, p. 45–53, 2018.

DEHAK, N.; DUMOUCHEL, P.; KENNY, P. Modeling prosodic features with joint factor analysis for speaker verification. In: IEEE Transactions on Audio, Speech and Language Processing, 7, **Anais**...2007.

DEMIRCAN, S.; KAHRAMANLI, H. Feature Extraction from Speech Data for Emotion Recognition. **Journal of Advances in Computer Networks**, v. 2, n. 1, p. 28–30, 2014.

DÉSIR, C.; BERNARD, S.; PETITJEAN, C.; HEUTTE, L. A Random Forest Based Approach for One Class Classification in Medical Imaging. In: WANG, F.; SHEN, D.; YAN, P.; SUZUKI, K. (Ed.). **Machine Learning in Medical Imaging. MLMI 2012. Lecture Notes in Computer Science**. Berlin, Heidelberg: Springer, 2012. p. 250–257.

DIETRICH, M.; ABBOTT, K. V. Vocal function in Introverts and Extraverts During a Psychological Stress Reactivity Protocol. **Journal of Speech, Language, and Hearing Research**, v. 55, n. 3, p. 973–987, 2012.

DIETTERICH, T. Overfitting and Undercomputing in Machine Learning. **ACM Computing Surveys**, v. 27, n. 3, p. 326–327, 1995.

DOS SANTOS, W. P.; DE ASSIS, F. M.; DE SOUZA, R. E.; MENDES, P. B.; DE SOUZA MONTEIRO, H. S., ALVES, H. D. A Dialectical Method to Classify Alzheimer's Magnetic Resonance Images. **Evolutionary Computation**, p. 473, 2009.

DOS SANTOS, W. P.; DE ASSIS, F. M.; DE SOUZA, R. E.; SANTOS, D.; FILHO, P. B. Evaluation of Alzheimer's disease by analysis of MR images using Objective Dialectical Classifiers as an alternative to ADC maps. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, **Anais**...2008.

DREWNIK, M.; PASTERNAK-WINIARSKI, Z. SVM Kernel Configuration and Optimization for the Handwritten Digit Recognition. In: SAEED, K.; HOMENDA, W.; CHAKI, R. (Ed.). **Computer Information Systems and Industrial Management. CISIM 2017. Lecture Notes in Computer Science**. [s.l.] Springer, 2017. p. 87–98.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. 2nd. ed. New York: Wiley-Interscience, 2001.

DWYER, D. B.; FALKAI, P.; KOUTSOULERIS, N. Machine Learning Approaches for Clinical Psychology and Psychiatry. **Annual Review of Clinical Psychology**, p. 1–28, 2018.

EBERHART, R.; KENNEDY, J. A new optimizer using particle swarm theory. In: MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan. **Anais**... Nagoya, Japan: 1995.

EDWARDS, G. J.; COOTES, T. E.; TAYLOR, C. J. Face recognition using active appearance models. In: European Conference on Computer Vision, Berlin, Heidelberg. **Anais**... Berlin, Heidelberg: Springer, 1998.

EFRON, B. Bayes' Theorem in the 21st Century. **Science**, v. 340, n. 6137, p. 1177–1178, 2013.

EKMAN, P.; ROSENBERG, E. L. **What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)**. New York: Oxford University Press, 1997.

ELAMIR, M. M.; AL-ATABANY, W.; ELDOSOKY, M. A. Emotion recognition via physiological signals using higher order crossing and Hjorth parameter. **Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences**, v. 5, n. 2, p. 839–846, 2019.

ELITE, A.; PEDRÃO, L. J.; ZAMBERLAN-AMORIM, N. E.; CARVALHO, A. M. P.; BÁRBARO, A. M. Comportamento comunicativo de indivíduos com esquizofrenia. **Rev. CEFAC**, v. 16, n. 4, p. 1283–1293, 2014.

ELOUEDI, H.; MELIANI, W.; ELOUEDI, Z.; BEN AMOR, N. A hybrid approach based on decision trees and clustering for breast cancer classification. In: 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), Tunis. **Anais**... Tunis: 2014.

ELVEVÅG, B.; FOLTZ, P. W.; ROSENSTEIN, M.; DELISI, L. E. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. **J Neurolinguistics**, v. 23, n. 3, p. 270–284, 2010.

ESPINOLA, C. W.; GOMES, J. C.; PEREIRA, J. M. S.; DOS SANTOS, W. P. Vocal acoustic analysis and machine learning for the identification of schizophrenia. **Research on Biomedical Engineering**, 2020a.

ESPINOLA, C. W.; GOMES, J. C.; PEREIRA, J. M. S.; DOS SANTOS, W. P. Detection of major depressive disorder using vocal acoustic analysis and machine learning—an

exploratory study. **Research on Biomedical Engineering**, 2020b.

ESQUEDA, F.; BILBAO, S.; VÄLIMÄKI, V. Aliasing Reduction in Clipped Signals. In: IEEE Transactions on Signal Processing, 64, **Anais**...2016.

EYBEN, F.; WÖLLMER, M.; SCHULLER, B. openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor. In: Proceedings of the 18th ACM international conference on Multimedia, Firenza. **Anais**... Firenza: 2010.

FAN, Y.; MURPHY, T. B.; BYRNE, J. C.; BRENNAN, L.; FITZPATRICK, J. M.; WATSON, R. W. G. Applying random forests to identify biomarker panels in serum 2D-DIGE data for the detection and staging of prostate cancer. **Journal of Proteome Research**, v. 10, n. 3, p. 1361–1373, 2011.

FARLEY-TOOMBS, C. The stigma of a psychiatric diagnosis: prevalence, implications and nursing interventions in clinical care settings. **Critical Care Nursing Clinics of North America**, v. 24, n. 1, p. 149–156, 2012. Disponível em: <http://dx.doi.org/10.1016/j.ccell.2012.01.009>.

FATHELBAB, W. M. Two Novel Classes of Band-Reject Filters Realizing Broad Upper Pass Bandwidth — Synthesis and Design. In: IEEE Transactions on Microwave Theory and Techniques, 2, **Anais**...2011.

FAURHOLT-JEPSEN, M.; BUSK, J.; FROST, M.; VINBERG, M.; CHRISTENSEN, E. M.; WINTHER, O.; BARDRAM, J. E.; KESSING, L. V. Voice analysis as an objective state marker in bipolar disorder. **Transl Psychiatry**, v. 6, n. 7, p. e856-8, 2016. Disponível em: <http://dx.doi.org/10.1038/tp.2016.123>.

FEHM, L.; BEESDO, K.; JACOBI, F.; FIEDLER, A. Social anxiety disorder above and below the diagnostic threshold: prevalence, comorbidity and impairment in the general population. **Social Psychiatry and Psychiatric Epidemiology**, v. 43, p. 257–265, 2008.

FERDINANDY, B.; GERENCSÉR, L.; CORRIERI, L.; PEREZ, P.; ÚJVÁRY, D.; CSIZMADIA, G.; MIKLÓSI, Á. Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures. **PLoS ONE**, v. 15, n. 7, p. e0236092, 2020.

FERRAND, C. T. Harmonics-to-noise ratio: An index of vocal aging. **Journal of Voice**, v. 16, n. 4, p. 480–487, 2002.

FRANK, E. **Oversampling and Undersampling**. Disponível em: <https://waikato.github.io/weka-blog/posts/2019-01-30-sampling/>. Acesso em: 15 jun. 2020.

FREEDMAN, R. Schizophrenia. **The New England Journal of Medicine**, v. 349, n. 18, p. 1738–1749, 2003.

FRICK, R. W. Communicating Emotion: The Role of Prosodic Features. **Psychological Bulletin**, v. 97, n. 3, p. 412–429, 1985.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian Network Classifiers. **Machine**

Learning, v. 29, p. 131–163, 1997.

GAIKWAD, S. K.; GAWALI, B. W.; YANNAWAR, P. A Review on Speech Recognition Technique. **International Journal of Computer Applications**, v. 10, n. 3, p. 16–24, 2010.

GBD 2017 DISEASE AND INJURY INCIDENCE AND PREVALENCE COLLABORATORS. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. **The Lancet**, v. 392, n. 10159, p. 1789–1858, 2018.

GIDDENS, C. L.; BARRON, K. W.; BYRD-CRAVEN, J.; CLARK, K. F.; WINTER, A. S. Vocal Indices of Stress: A Review. **Journal of Voice**, v. 27, n. 3, p. 390.e21-390.e29, 2013. Disponível em: <http://dx.doi.org/10.1016/j.jvoice.2012.12.010>.

GIDEON, J.; PROVOST, E. M.; MCINNIS, M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. **Proc IEEE Int Conf Acoust Speech Signal Process**, v. 2016 Mar, p. 2359–2363, 2016.

GONÇALVES, D. M.; STEIN, A. T.; KAPCZINSKI, F. Avaliação de desempenho do Self-Reporting Questionnaire como instrumento de rastreamento psiquiátrico: Um estudo comparativo com o Structured Clinical Interview for DSM-IV-TR. **Cadernos de Saude Publica**, v. 24, n. 2, p. 380–390, 2008.

GREEN, M. F. Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. **The Journal of Clinical Psychiatry**, v. 67, n. Suppl 9, p. 3–8, 2006.

GRZYMALA-BUSSE, J. W. Rule Induction. In: MAIMON, O.; ROKACH, L. (Ed.). **Data Mining and Knowledge Discovery Handbook**. Boston, MA: Springer, 2009.

HAMILTON, M. A RATING SCALE FOR DEPRESSION. **J. Neurol. Neurosurg. Psychiat.**, v. 23, p. 56–62, 1960.

HANS-ULRICH WITTCHEN. Generalized anxiety disorder: Prevalence, burden, and cost to society. **Depression and Anxiety**, v. 16, p. 162–171, 2002.

HASAN, R.; JAMIL, M.; RABBANI, G.; RAHMAN, S. Speaker Identification Using Mel Frequency Cepstral Coefficients. **3rd International Conference on Electrical & Computer Engineering ICECE 2004**, n. December, p. 565–568, 2004.

HASHIM, N. W.; WILKES, M.; SALOMON, R.; MEGGS, J.; FRANCE, D. J. Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. **Journal of Voice**, v. 31, n. 2, p. 256.e1-256.e6, 2017.

HAYES, J. F.; MILES, J.; WALTERS, K.; KING, M.; OSBORN, D. P. J. A systematic review and meta-analysis of premature mortality in bipolar affective disorder. **Acta Psychiatrica Scandinavica**, v. 131, p. 417–425, 2015.

HAYKIN, S. **Neural Networks and Learning Machines**. Third ed. Hamilton: Pearson Prentice Hall, 2009.

HAZRA, A.; MANDAL, S. K.; GUPTA, A. Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms. **International Journal of Computer Applications**, v. 145, n. 2, p. 39–45, 2016.

HEARST, M. A.; DUMAIS, S. T.; OSUNA, E.; PLATT, J.; SCHOLKOPF, B. Support vector machines. **IEEE Intelligent Systems and their Applications**, v. 13, n. 4, p. 18–28, 1998.

HENGSTLER, M.; ENKEL, E.; DUELLI, S. Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices. **Technological Forecasting and Social Change**, v. 105, p. 105–120, 2016. Disponível em: <http://dx.doi.org/10.1016/j.techfore.2015.12.014>.

HIBARE, R.; VIBHUTE, A. Feature Extraction Techniques in Speech Processing: A Survey. In: International Journal of Computer Applications, 5, **Anais**...2014.

HIGUCHI, M.; NAKAMURA, M.; SHINOHARA, S.; OMIYA, Y.; TAKANO, T.; TODA, H.; SAITO, T.; YOSHINO, A.; MITSUYOSHI, S.; TOKUNO, S. Discrimination of Bipolar Disorders Using Voice. **MindCare**, v. 1, p. 199–207, 2019. Disponível em: <http://dx.doi.org/10.1007/978-3-030-25872-6_16>.

HIGUCHI, M.; TOKUNO, S.; NAKAMURA, M.; SHINOHARA, S. Classification of bipolar disorder, major depressive disorder, and healthy state using voice. **Asian Journal of Pharmaceutical and Clinical Research**, v. 11, n. 3, p. 89–93, 2018.

HIRSCHTRITT, M.; INSEL, T. Digital Technologies in Psychiatry: Present and Future. **Focus**, v. 16, n. 3, p. 251–258, 2018.

HOLMQVIST, S.; SANTTILA, P.; LINDSTRÖM, E.; SALA, E.; SIMBERG, S. The Association Between Possible Stress Markers and Vocal Symptoms. **Journal of Voice**, v. 27, n. 6, p. 787.e1-787.e10, 2013.

HÖNIG, F.; BATLINER, A.; NÖTH, E.; SCHNIEDER, S.; KRAJEWSKI, J. Automatic modelling of depressed speech: Relevant features and relevance of gender. **Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH**, n. 444, p. 1248–1252, 2014.

HOR, K.; TAYLOR, M. Suicide and schizophrenia: a systematic review of rates and risk factors. **Journal of Psychopharmacology**, v. 24, n. 11, p. 81–90, 2010.

HUTTER, F.; SCHMIDT-THIEME, L.; LÜCKE, J. Beyond Manual Tuning of Hyperparameters. **KI - Künstliche Intelligenz**, v. 29, n. 4, p. 329–337, 2015.

HUYS, Q. J. M.; MAIA, T. V; FRANK, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. **Nature Neuroscience**, v. 19, n. 3, p. 404–413, 2016.

IVERACH, L.; O'BRIAN, S.; JONES, M.; BLOCK, S.; LINCOLN, M.; HARRISON, E.; HEWAT, S.; MENZIES, R. G.; PACKMAN, A.; ONSLOW, M. Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. **Journal of Anxiety Disorders**, v. 23, p. 928–934, 2009.

JASPERS, K. **Allgemeine Psychopathologie**. 4. ed. Berlin and Heildelberg: Springer-Verlag Berlin Heildelberg, 1946.

JATUPAIBOON, N.; PAN-NGUM, S.; ISRASENA, P. Real-time EEG-based happiness detection system. **The Scientific World Journal**, v. 2013, 2013.

JIANG, H.; HU, B.; LIU, Z.; WANG, G.; ZHANG, L.; LI, X.; KANG, H. Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features. **Computational and Mathematical Methods in Medicine**, v. 2018, 2018.

JIANG, H.; HU, B.; LIU, Z.; YAN, L.; WANG, T.; LIU, F.; KANG, H.; LI, X. Investigation of different speech types and emotions for detecting depression using different classifiers. **Speech Communication**, v. 90, p. 39–46, 2017.

JIANG, L.; LI, C.; WANG, S.; ZHANG, L. Deep feature weighting for naive Bayes and its application to text classification. **Engineering Applications of Artificial Intelligence**, v. 52, p. 26–39, 2016. Disponível em: <http://dx.doi.org/10.1016/j.engappai.2016.02.002>.

JIANG, P.; FU, H.; TAO, H.; LEI, P.; ZHAO, L. I. Parallelized Convolutional Recurrent Neural Network With Spectral Features for Speech Emotion Recognition. **IEEE Access**, v. 7, p. 90368–90377, 2019.

JORDAN, P.; SHEDDEN-MORA, M. C.; LÖWE, B. Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. **PLoS ONE**, v. 12, n. 8, p. 1–14, 2017.

KÄCHELE, M.; ZHARKOV, D.; MEUDT, S.; SCHWENKER, F. Prosodic, Spectral and Voice Quality Feature Selection Using a Long-Term Stopping Criterion for Audio-Based Emotion Recognition. In: International Conference on Pattern Recognition, **Anais**...2014.

KADAMBE, S.; SRINIVASAN, P.; TELFER, B.; SZU, H. Representation and classification of unvoiced sounds using adaptive wavelets. In: Proc. SPIE 1961, Visual Information Processing II, **Anais**...1993.

KAKOUROS, S.; RÄSÄNEN, O.; ALKU, P. Comparison of spectral tilt measures for sentence prominence in speech—Effects of dimensionality and adverse noise conditions. **Speech Communication**, v. 103, p. 11–26, 2018.

KAMBLE, M.; PATIL, H. Analysis of Reverberation via Teager Energy Features for Replay Spoof Speech Detection. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May, **Anais**...2019.

KARAM, Z. N.; PROVOST, E. M.; SINGH, S.; MONTGOMERY, J.; ARCHER, C.; HARRINGTON, G.; MCINNIS, M. G. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. **2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)**, p. 4858–4862, 2014.

KAYI, E. S.; DIAB, M.; PAUSELLI, L.; COMPTON, M.; COPPERSMITH, G. Predictive Linguistic Features of Schizophrenia. p. 1–15, 2018.

KEDEM, B. Spectral Analysis and Discrimination by Zero-Crossings. In: Proceedings of the IEEE, 11, **Anais**...1986.

KEEFE, R. S. E.; HARVEY, P. D. Cognitive Impairment in Schizophrenia. **Handbook of Experimental Pharmacology**, v. 213, p. 11–37, 2012. Disponível em: <http://link.springer.com/10.1007/978-3-642-25758-2>.

KENNEDY, J.; EBERHART, R. Particle Swarm Optimisation. In: Proceedings of the IEEE international conference on neural networks, **Anais**...1995.

KENT, R. D.; READ, C. **ANÁLISE ACÚSTICA DA FALA**. São Paulo: Cortez, 2015.
KESSLER, R. C. Posttraumatic Stress Disorder: The Burden to the Individual and to Society. **The Journal of Clinical Psychiatry**, v. 61, n. suppl 5, p. 4–12, 2000.

KHARYA, S.; AGRAWAL, S.; SONI, S. Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer. **International Journal of Computer Applications**, v. 92, n. 10, p. 26–31, 2014.

KIEFER, R.; CHELMINSKI, I.; DALRYMPLE, K.; ZIMMERMAN, M. Principal Diagnoses in Psychiatric Outpatients with Posttraumatic Stress Disorder: Implications for Screening Recommendations. **Journal of Nervous and Mental Disease**, v. 208, n. 4, p. 283–287, 2020.

KILPATRICK, D. G.; RESNICK, H. S.; MILANAK, M. E.; MILLER, M. W.; KEYES, K. M.; FRIEDMAN, M. J. National Estimates of Exposure to Traumatic Events and PTSD Prevalence Using DSM-IV and DSM-5 Criteria. **Journal of Traumatic Stress**, v. 26, n. 5, p. 537–547, 2013.

KIM, J. H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. **Computational Statistics and Data Analysis**, v. 53, n. 11, p. 3735–3745, 2009. Disponível em: <http://dx.doi.org/10.1016/j.csda.2009.04.009>.

KINNUNEN, T.; LI, H. An overview of text-independent speaker recognition: from features to supervectors. **Speech Communication**, v. 52, n. 1, p. 12–40, 2010.

KOOLAGUDI, S. G.; RAO, K. S. Emotion recognition from speech: a review. **International Journal of Speech Technology**, v. 15, n. 2, p. 99–117, 2012.

KOTSIANTIS, S. B. Decision trees: A recent overview. **Artificial Intelligence Review**, v. 39, n. 4, p. 261–283, 2013.

KRAEPELIN, E. **Manic-Depressive Insanity and Paranoia**. Edinburgh: Livingstone, 1921.

KREMIC, E.; SUBASI, A. Performance of random forest and SVM in face recognition. **International Arab Journal of Information Technology**, v. 13, n. 2, p. 287–293, 2016.

KUO, F. Y.; SLOAN, I. H. Lifting the Curse of Dimensionality. **Notices of the AMS**, v. 52, n. 11, p. 1320–1328, 2005.

LAHMIRI, S.; SHMUEL, A. Detection of Parkinson’s disease based on voice patterns

ranking and optimized support vector machine. **Biomedical Signal Processing and Control**, v. 49, p. 427–433, 2019. Disponível em: <https://doi.org/10.1016/j.bspc.2018.08.029>.

LANGDON, W. B.; BARRETT, S. J.; BUXTON, B. F. Combining decision trees and neural networks for drug discovery. In: European Conference on Genetic Programming. Springer, Berlin, Heidelberg. **Anais**... Berlin, Heidelberg: Springer, 2002.

LARSEN, M. E.; CUMMINS, N.; BOONSTRA, T. W.; O'DEA, B.; TIGHE, J.; NICHOLAS, J.; SHAND, F.; EPPS, J.; CHRISTENSEN, H. The use of technology in Suicide Prevention. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, **Anais**...2015.

LAUKKA, P.; LINNMAN, C.; ÅHS, F.; PISSIOTA, A.; FRANS, Ö.; FARIA, V.; MICHELGÅARD, Å.; APPEL, L.; FREDRIKSON, M.; FURMARK, T. In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech. **J Nonverbal Behav**, v. 32, p. 195–214, 2008.

LEE, C. M.; YILDIRIM, S.; BULUT, M.; KAZEMZADEH, A.; BUSSO, C.; DENG, Z.; LEE, S.; NARAYANAN, S. Emotion Recognition based on Phoneme Classes. In: Eighth International Conference on Spoken Language Processing, **Anais**...2004.

LEUCHT, S.; KANE, J. M.; KISSLING, W.; HAMANN, J.; ETSCHEL, E.; ENGEL, R. Clinical implications of Brief Psychiatric Rating Scale scores. **British Journal of Psychiatry**, v. 187, n. OCT., p. 366–371, 2005.

LI, Y.; VASCONCELOS, N. REPAIR: Removing representation bias by dataset resampling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, **Anais**...2019.

LISCOMBE, J.; VENDITTI, J.; HIRSCHBERG, J. Classifying Subject Ratings of Emotional Speech Using Acoustic Features. In: Eurospeech, Geneva. **Anais**... Geneva: 2003.

LIU, W.; LIAO, Q.; QIAO, F.; XIA, W.; WANG, C. Approximate designs for fast Fourier transform (FFT) with application to speech recognition. In: IEEE Transactions on Circuits and Systems I: Regular Papers, **Anais**...IEEE, 2019.

LIU, Y.; GUO, J.; LEE, J. Halftone Image Classification Using LMS Algorithm and Naive Bayes. **IEEE Transactions on Image Processing**, v. 20, n. 10, p. 2837–2847, 2011.

LIU, Z.; HU, B.; YAN, L.; WANG, T.; LIU, F.; LI, X.; KANG, H. Detection of Depression in Speech. In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), **Anais**...2015.

LONLA, B. M.; MBIHI, J.; NNEME, L. N. FPGA-Based Multichannel Digital Duty-Cycle Modulation and Application to Simultaneous Generation of Analog Signals. **Journal of Electronic Design Technology**, v. 8, n. 1, p. 23–35, 2017.

LOW, L. S. A.; MADDAGE, N. C.; LECH, M.; SHEEBER, L. B.; ALLEN, N. B. Detection of clinical depression in adolescents' speech during family interactions. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 3 PART 1, p. 574–586, 2011.

LOWELL, S. Y.; KELLEY, R. T.; AWAN, S. N.; COLTON, R. H.; CHAN, N. H. Spectral- and Cepstral-Based Acoustic Features of Dysphonic, Strained Voice Quality. **Annals of Otology, Rhinology and Laryngology**, v. 121, n. 8, p. 539–548, 2012.

LUGGER, M.; YANG, B. The relevance of voice quality features in speaker independent emotion recognition. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI. **Anais**... Honolulu, HI: 2007.

LUKASIEWICZ, M.; GERARD, S.; BESNARD, A.; FALISSARD, B.; PERRIN, E.; SAPIN, H.; TOHEN, M.; REED, C.; AZORIN, J.-M.; GROUP, T. E. S. Young Mania Rating Scale: how to interpret the numbers?Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort. **International Journal of Methods in Psychiatric Research**, v. 22, n. 1, p. 46–58, 2013. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/mpr.247/abstract>.

MAAS, A. L.; DALY, R. E.; PHAM, P. T.; HUANG, D.; NG, A. Y.; POTTS, C. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, **Anais**...2011.
MAC-KAY, A.; JEREZ, I.; PESENTI, P. Speech-language intervention in schizophrenia: an integrative review. **Rev. CEFAC**, v. 20, n. 2, p. 238–246, 2018.

MANJUSHA, K. K.; SANKARANARAYANAN, K.; SEENA, P. Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification. **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 4, n. 1, p. 864–868, 2014.

MANN, J. J.; ELLIS, S. P.; WATERNAUX, C. M.; LIU, X.; OQUENDO, M. A.; MALONE, K. M.; BRODSKY, B. S.; HAAS, G. L.; CURRIER, D. Classification trees distinguish suicide attempters in major psychiatric disorders: A model of clinical decision making. **Journal of Clinical Psychiatry**, v. 69, n. 1, p. 23–31, 2008.

MARAGOS, P.; KAISER, J. F.; QUATIERI, T. F. Maragos, Kaiser, Quatieri - - Energy separation in signal modulations with application to speech analysis.pdf. In: IEEE Transactions on Signal Processing, **Anais**...1993.

MARSLAND, S. **Machine Learning: An Algorithmic Perspective**. Second ed. [s.l.] CRC Press, 2015.

MARTÍNEZ-SÁNCHEZ, F.; MUELA-MARTÍNEZ, J. A.; CORTÉS-SOTO, P.; JOSÉ, J.; MEILÁN, G.; ANTONIO, J.; FERRÁNDIZ, V.; CAPARRÓS, A. E.; MARÍA, I.; VALVERDE, P. Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? **The Spanish Journal of Psychology**, v. 18, n. 86, p. 1–9, 2015.

MATĚJKA, P.; GLEMBEK, O.; NOVOTNÝ, O.; PLCHOT, O.; GRÉZL, F.; BURGET, L.; CERNOCKÝ, J. H. Analysis of DNN approaches to speaker identification. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), **Anais**...2016.

MATHEWS, M. V.; MILLER, J. E.; DAVID, E. E. Pitch Synchronous Analysis of Voiced Sounds. **The Journal of the Acoustical Society of America**, v. 33, n. 2, p. 179–186, 1961.

MAXHUNI, A.; MUÑOZ-MELÉNDEZ, A.; OSMANI, V.; PEREZ, H.; MAYORA, O.; MORALES, E. F. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. **Pervasive and Mobile Computing**, v. 31, n. 1, p. 50–66, 2016. Disponível em: <http://dx.doi.org/10.1016/j.pmcj.2016.01.008>.

MCCULLOCH, W. S.; PITTS, W. A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY. **Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943.

MCGINNIS, E. W.; ANDERAU, S. P.; HRUSCHAK, J.; GURCHIEK, R. D.; LOPEZ-DURAN, N. L.; FITZGERALD, K.; ROSENBLUM, K. L.; MUZIK, M.; MCGINNIS, R. S. Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. **IEEE**, p. 1–8, 2019.

METIN, B.; PAL, K.; CICEKOGLU, O. All-pass filters using DDCC- and MOSFET-based electronic resistor. **International Journal of Circuit Theory and Applications**, v. 39, n. 8, p. 881–891, 2011.

MEUDT, S.; SCHWENKER, F. On instance selection in audio based emotion recognition. In: IAPR Workshop on Artificial Neural Networks in Pattern Recognition, **Anais**...Springer Berlin Heidelberg, 2012.

MIGUEL, E. C.; GENTIL, V.; GATTAZ, W. F. **Clínica Psiquiátrica**. Barueri: Manole, 2011.

MILLIER, A.; SCHMIDT, U.; ANGERMEYER, M. C.; CHAUHAN, D.; MURTHY, V.; TOUMI, M.; CADI-SOUSSI, N. Humanistic burden in schizophrenia: A literature review. **Journal of Psychiatric Research**, v. 54, n. 1, p. 85–93, 2014. Disponível em: <http://dx.doi.org/10.1016/j.jpsychires.2014.03.021>.

MITRA, V.; SHRIBERG, E. Effects of Feature Type, Learning Algorithm and Speaking Style for Depression Detection from Speech. **IEEE**, p. 4774–4778, 2015.

MITROVIĆ, D.; ZEPPELZAUER, M.; BREITENEDER, C. Features for Content-Based Audio Retrieval. In: **Advances in Computers**. [s.l: s.n.]78p. 71–150.

MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. In: **Sistemas Inteligentes-Fundamentos e Aplicações**. [s.l: s.n.]p. 115–139.

MOORE, E.; CLEMENTS, M. A.; PEIFER, J. W.; WEISSER, L. Critical Analysis of the Impact of Glottal Features in the Classification of Clinical Depression in Speech. **IEEE Transactions on Biomedical Engineering**, v. 55, n. 1, p. 96–107, 2008.

MORALES, D. A.; VIVES-GILABERT, Y.; GÓMEZ-ANSÓN, B.; BENGOETXEA, E.; LARRAÑAGA, P.; BIELZA, C.; PAGONABARRAGA, J.; KULISEVSKY, J.; CORCUERA-SOLANO, I.; DELFINO, M. Predicting dementia development in Parkinson's disease using Bayesian network classifiers. **Psychiatry Research - Neuroimaging**, v. 213, n. 2, p. 92–98, 2013.

MUAREMI, A.; GRAVENHORST, F.; GRÜNERBL, A.; ARNRICH, B.; TRÖSTER, G.

Assessing Bipolar Episodes Using Speech Cues Derived from Phone Calls. In: International Symposium on Pervasive Computing Paradigms for Mental Health (MindCare), **Anais**...2014.

MUDA, L.; BEGAM, M.; ELAMVAZUTHI, I. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. **Journal Of Computing**, v. 2, n. 3, p. 138–143, 2010. Disponível em: <http://arxiv.org/abs/1003.4083>.

MUNDT, J. C.; SNYDER, P. J.; CANNIZZARO, M. S.; CHAPPIE, K.; GERALTS, D. S. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. **Journal of Neurolinguistics**, v. 20, p. 50–64, 2007.

MUNDT, J. C.; VOGEL, A. P.; FELTNER, D. E.; LENDERKING, W. R. Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. **Biological Psychiatry**, v. 72, n. 7, p. 580–587, 2012.

NAGRANI, A.; CHUNG, J. S.; ZISSERMAN, A. VoxCeleb: A large-scale speaker identification dataset. **Telephony**, v. 3, p. 33–39, 2017.

NAHAR, J.; IMAM, T.; TICKLE, K. S.; CHEN, Y. P. P. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. **Expert Systems with Applications**, v. 40, n. 1, p. 96–104, 2013. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2012.07.032>.

NAKRA, A.; DUHAN, M. Comparative Analysis of Bayes Net Classifier, Naive Bayes Classifier and Combination of both Classifiers using WEKA. **International Journal of Information Technology and Computer Science**, v. 11, n. 3, p. 38–45, 2019.

NANDA, M. A.; SEMINAR, K. B.; NANDIKA, D.; MADDU, A. A comparison study of kernel functions in the support vector machine and its application for termite detection. **Information (Switzerland)**, v. 9, n. 1, p. 5, 2018.

NARAYANAN, S.; ALWAN, A. Noise source models for fricative consonants. In: IEEE Transactions on Speech and Audio Processing, 3, **Anais**...2000.

NASIRI, R. M.; WANG, Z. Perceptual aliasing factors and the impact of frame rate on video quality. In: 2017 IEEE International Conference on Image Processing (ICIP), **Anais**...2017.

NATH, M. Review of Filter Techniques. **International Journal of Engineering Trends and Technology**, v. 3, n. 3, p. 415–421, 2012.

NERRIÈRE, E.; VERCAMBRE, M. N.; GILBERT, F.; KOVESS-MASFÉTY, V. Voice disorders and mental health in teachers: A cross-sectional nationwide study. **BMC Public Health**, v. 9, n. 370, 2009.

NEWMAN, S.; MATHER, V. G. Analysis of spoken language of patients with affective disorders. **American Journal of Psychiatry**, v. 94, p. 913–942, 1938.

NGAI, E. W. T.; HU, Y.; WONG, Y. H.; CHEN, Y.; SUN, X. The application of data mining

techniques in financial fraud detection: A classification framework and an academic review of literature. **Decision Support Systems**, v. 50, n. 3, p. 559–569, 2011. Disponível em: <http://dx.doi.org/10.1016/j.dss.2010.08.006>.

NICHOLS, J. A.; HERBERT CHAN, H. W.; BAKER, M. A. B. Machine learning: applications of artificial intelligence to imaging and diagnosis. **Biophysical Reviews**, v. 11, p. 111–118, 2019.

NOLL, A. M. Cepstrum Pitch Determination. **The Journal of the Acoustical Society of America**, v. 41, n. 2, p. 293–309, 1967.

NOVICK, D. M.; SWARTZ, H. A.; FRANK, E. Suicide attempts in bipolar I and bipolar II disorder: a review and meta-analysis of the evidence. **Bipolar Disord.**, v. 12, n. 1, p. 1–9, 2010.

NWE, T. L.; FOO, S. W.; DE SILVA, L. C. Speech emotion recognition using hidden Markov models. **Speech Communication**, v. 41, n. 4, p. 603–623, 2003.

OATLEY, G. C.; EWART, B. W. Crimes analysis software: 'pins in maps', clustering and Bayes net prediction. **Expert Systems with Applications**, v. 25, n. 4, p. 569–588, 2003.
OKE, S. A. A Literature Review on Artificial Intelligence. **International Journal of Information and Management Sciences**, v. 19, n. 4, p. 535–570, 2008.

OLIVEIRA, A. P. S. de; SANTANA, M. A. de; ANDRADE, M. K. S.; GOMES, J. C.; RODRIGUES, M. C. A.; SANTOS, W. P. dos. Early diagnosis of Parkinson's disease using EEG, machine learning and partial directed coherence. **Research on Biomedical Engineering**, v. 36, n. 311–331, 2020.

OOI, K. E. B.; LECH, M.; ALLEN, N. B. Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents. **IEEE Transactions on Biomedical Engineering**, v. 60, n. 2, p. 497–506, 2013.

OPPENHEIM, A. V.; SCHAFER, R. W. From frequency to quefrency: A history of the cepstrum. **IEEE Signal Processing Magazine**, v. 21, n. 5, p. 95–106, 2004.

OPPENHEIM, A. V.; WILLSKY, A. S. **SINAIS E SISTEMAS**. 2. Edição ed. São Paulo: Pearson Prentice Hall, 2010.

OSISANWO, F. Y.; AKINSOLA, J. E. T.; AWODELE, O.; HINMIKAIYE, J. O.; OLAKANMI, O.; AKINJOBI, J. Supervised Machine Learning Algorithms: Classification and Comparison. **International Journal of Computer Trends and Technology (IJCTT)**, v. 48, n. 3, p. 128–138, 2017.

OTTE, C.; GOLD, S. M.; PENNINX, B. W.; PARIANTE, C. M.; ETKIN, A.; FAVA, M.; MOHR, D. C.; SCHATZBERG, A. F. Major depressive disorder. **Nature Reviews Disease Primers**, v. 2, n. 16065, 2016. Disponível em: <http://dx.doi.org/10.1038/nrdp.2016.65>.

OVERALL, J. E.; GORHAM, D. R. THE BRIEF PSYCHIATRIC RATING SCALE. **Psychological Reports**, v. 10, p. 799–812, 1962.

ÖZÇIFT, A. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. **Computers in Biology and Medicine**, v. 41, n. 5, p. 265–271, 2011.

OZDAS, A.; SHIAVI, R. G.; SILVERMAN, S. E.; SILVERMAN, M. K.; WILKES, D. M. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. **IEEE Transactions on Biomedical Engineering**, v. 51, n. 9, p. 1530–1540, 2004.

ÖZSEVEN, T.; DÜGENCI, M.; DORUK, A.; KAHRAMAN, H. I. Voice Traces of Anxiety: Acoustic Parameters Affected by Anxiety Disorder. **Archives of Acoustics**, v. 43, n. 4, p. 625–636, 2018.

PACTITIS, S. A. **Active Filters: Theory and Design**. [s.l.] CRC Press, 2018.

PAL, S. K.; MITRA, S. Multilayer Perceptron, Fuzzy Sets, Classification. **IEEE Transactions on Neural Networks**, v. 3, n. 5, p. 683–697, 1992.

PANAGIOTAKOPOULOS, T. C.; LYRAS, D. P.; LIVADITIS, M.; SGARBAS, K. N.; ANASTASSOPOULOS, G. C.; LYMBEROPOULOS, D. K. A Contextual Data Mining Approach Toward Assisting the Treatment of Anxiety Disorders. **IEEE Transactions on Information Technology in Biomedicine**, v. 14, n. 3, p. 567–581, 2010.

PARK, Y. S.; LEK, S. Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. In: **Developments in Environmental Modelling**. [s.l.] Elsevier, 2016. 28p. 123–140.

PAROLA, A.; SIMONSEN, A.; BLIKSTED, V.; FUSAROLI, R. Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. **Schizophrenia Research**, v. 216, p. 24–40, 2020. Disponível em: <https://doi.org/10.1016/j.schres.2019.11.031>.

PARSANIA, V. S.; JANI, N. N.; BHALODIYA, N. H. Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis. **International Journal of Darshan Institute on Engineering Research & Emerging Technologies**, v. 3, n. 1, 2014.

PATTEKARI, S.A.; PARVEEN, A. Prediction system for heart disease using Naïve Bayes. **International Journal of Advanced Computer and Mathematical Sciences**, v. 3, n. 3, p. 290–294, 2012.

PERKONIGG, A.; KESSLER, R. C.; STORZ, S.; WITTCHEN, H. U. Traumatic events and post-traumatic stress disorder in the community: prevalence, risk factors and comorbidity. **Acta Psychiatrica Scandinavica**, v. 101, n. 1, p. 46–59, 2000.

PETZSCHNER, F. H.; WEBER, L. A. E.; GARD, T.; STEPHAN, K. E. Review Computational Psychosomatics and Computational Psychiatry : Toward a Joint Framework for Differential Diagnosis. **Biological Psychiatry**, p. 1–10, 2017. Disponível em: <http://dx.doi.org/10.1016/j.biopsych.2017.05.012>.

PODDER, P.; KHAN, T. Z.; KHAN, M. H.; RAHMAN, M. M. Comparative Performance

Analysis of Hamming, Hanning and Blackman Window. **International Journal of Computer Applications**, v. 96, n. 18, 2014.

PODGORELEC, V.; KOKOL, P.; STIGLIC, B.; ROZMAN, I. Decision Trees: An Overview and Their Use in Medicine. **Journal of Medical Systems**, v. 26, n. 5, p. 445–463, 2002.

POLS, L. C. W.; VAN DER KAMP, L. J. T.; PLOMP, R. Perceptual and Physical Space of Vowel Sounds. **The Journal of the Acoustical Society of America**, v. 46, n. 2B, p. 458–467, 1969.

POPE, B.; BLASS, T.; SIEGMAN, A. W.; RAHER, J. Anxiety and depression in speech. **Journal of Consulting and Clinical Psychology**, v. 35, n. 1, p. 128–133, 1970.

PORIA, S.; CHATURVEDI, I.; CAMBRIA, E. Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis. **IEEE Computer Society**, 2016.

PRADHAN, A. Online support vector machine: A survey. **International Journal of Emerging Technology and Advanced Engineering**, v. 2, n. 8, p. 82–85, 2012.

PUPIN, J. R. **Introdução às Séries e Transformadas de Fourier e Aplicações no Processamento de Sinais e Imagens**. São Carlos: Universidade Federal de São Carlos, 2011.

QI, Y. Random Forest for Bioinformatics. In: ZHANG, C.; MA, Y. (Ed.). **Ensemble Machine Learning**. Boston, MA: Springer, 2012. p. 307–323.

QIN, Y.; ZHANG, X. HMM-based speaker emotional recognition technology for speech signal. In: Advanced Materials Research, **Anais**...2011.

QU, X.; YUAN, B.; LIU, W. A predictive model for identifying possible MCI to AD Conversions in the ADNI database. In: 2009 2nd International Symposium on Knowledge Acquisition and Modeling, KAM 2009, **Anais**...2009.

QUATIERI, T. F.; MALYSKA, N. Vocal-source biomarkers for depression: A link to psychomotor activity. In: 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012, Portland. **Anais**... Portland: 2012.

RABINER, L. R.; SCHAFER, R. W. Introduction to Digital Speech Processing. **Foundations and Trends in Signal Processing**, v. 1, n. 1–2, p. 1–194, 2007.

RABINOWITZ, J.; LEVINE, S. Z.; GARIBALDI, G.; BUGARSKI-KIROLA, D.; BERARDO, C. G.; KAPUR, S. Negative symptoms have greater impact on functioning than positive symptoms in schizophrenia: Analysis of CATIE data. **Schizophrenia Research**, v. 137, p. 147–150, 2012. Disponível em: <http://dx.doi.org/10.1016/j.schres.2012.01.015>.

RADOMSKY, E. D.; HAAS, G. L.; MANN, J. J.; SWEENEY, J. A. Suicidal behavior in patients with schizophrenia and other psychotic disorders. **American Journal of Psychiatry**, v. 156, n. 10, p. 1590–1595, 1999.

RAJPUT, S. S.; BHADAURIA, S. S. Comparison of Band-stop FIR Filter using Modified Hamming Window and Other Window functions and Its Application in Filtering a Mutitone

Signal. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, v. 1, n. 8, p. 325–328, 2012.

RAMIREZ, G. A.; BALTRUŠAITIS, T.; MORENCY, L. P. Modeling latent discriminative dynamic of multi-dimensional affective signals. In: D'MELLO, S.; GRAESSER, A.; SCHULLER, B.; MARTIN, J. (Ed.). **Lecture Notes in Computer Science**. Berlin, Heidelberg: Springer, 2011. 6975p. 396–406.

RAPCAN, V.; D'ARCY, S.; YEAP, S.; AFZAL, N.; THAKORE, J.; REILLY, R. B. Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. **Medical Engineering and Physics**, v. 32, p. 1074–1079, 2010. Disponível em: <http://dx.doi.org/10.1016/j.medengphy.2010.07.013>.

RAUSEO-RICUPERO, N.; TOROUS, J. Technology Enabled Clinical Care (TECC): Protocol for a Prospective Longitudinal Cohort Study of Smartphone-Augmented Mental Health Treatment. **JMIR Research Protocols**, v. 10, n. 1, p. e23771, 2021.

RAZAK, E.; YUSOF, F.; RAUS, R. A. Classification of miRNA Expression Data Using Random Forests for Cancer Diagnosis. In: 2016 International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur. **Anais**... Kuala Lumpur: 2016.

REJAIBI, E.; KOMATY, A.; MERIAUDEAU, F.; AGREBI, S.; OTHMANI, A. MFCC-based recurrent neural network for automatic clinical depression recognition and assessment from speech. 2019.

RINGEVAL, F.; VALSTAR, M.; COWIE, R.; SCHMITT, M.; CUMMINS, N.; LALANNE, D.; MICHAUD, A.; SALAH, A. A. AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition. **AVEC'18**, p. 3–13, 2018.

RODRIGUES, A. L.; DE SANTANA, M. A.; AZEVEDO, W. W.; BEZERRA, R. S.; BARBOSA, V. A.; DE LIMA, R. C.; DOS SANTOS, W. P. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. **Research on Biomedical Engineering**, v. 35, n. 3, p. 213–222, 2019.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958.

ROWA, K.; WAECHTER, S.; HOOD, H. K.; ANTONY, H. M. M. Generalized Anxiety Disorder. In: CRAIGHEAD, W. E.; MIKLOWITZ, D. J.; CRAIGHEAD, L. W. (Ed.). **Psychopathology: History, Diagnosis, and Empirical Foundations**. Third ed. San Francisco, CA: John Wiley & Sons, 2017. p. 206.

ROWLAND, T. A.; MARWAHA, S. Epidemiology and risk factors for bipolar disorder. **Therapeutic Advances in Psychopharmacology**, v. 8, n. 9, p. 251–269, 2018.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Third ed. Harlow: Pearson Education, 2016.

SADOCK, B.; SADOCK, V.; RUIZ, P. **Compêndio de Psiquiatria: Ciência do Comportamento e Psiquiatria Clínica**. 11. ed. Porto Alegre: Artmed, 2017.

SAINI, J.; MEHRA, R. Power Spectral Density Analysis of Speech Signal using Window Techniques. **International Journal of Computer Applications**, v. 131, n. 14, p. 33–36, 2015.

SANCHEZ-MORENO, J.; MARTINEZ-ARAN, A.; TABARÉS-SEISDEDOS, R.; TORRENT, C.; VIETA, E.; AYUSO-MATEOS, J. L. Functioning and disability in bipolar disorder: An extensive review. **Psychotherapy and Psychosomatics**, v. 78, n. 5, p. 285–297, 2009.

SANTOS, K. O. B.; ARAÚJO, T. M.; PINHO, P. S.; SILVA, A. C. C. Avaliação de um Instrumento de Mensuração de Morbidade Psíquica. **Revista Baiana de Saúde Pública**, v. 34, n. 3, p. 544–560, 2010.

SAUER, S.; LEMKE, J.; ZINN, W.; BUETTNER, R.; KOHLS, N. Mindful in a random forest: Assessing the validity of mindfulness items using random forests methods. **Personality and Individual Differences**, v. 81, p. 117–123, 2015. Disponível em: <http://dx.doi.org/10.1016/j.paid.2014.09.011>.

SCHAFER, R. W.; RABINER, L. R. System for Automatic Formant Analysis of Voiced Speech. **The Journal of the Acoustical Society of America**, v. 47, n. 2B, p. 634–648, 1970.

SCHAFFER, A.; ISOMETSÄ, E. T.; TONDO, L.; MORENO, D. H.; SINYOR, M.; LARS VEDEL, K.; TURECKI, G.; WEIZMAN, A.; AZORIN, J. M.; HA, K.; REIS, C.; CASSIDY, F.; GOLDSTEIN, T.; RIHMER, Z.; BEAUTRAIS, A.; CHOU, Y. H.; DIAZGRANADOS, N.; LEVITT, A. J.; ZARATE, C. A.; YATHAM, L. Epidemiology, neurobiology and pharmacological interventions related to suicide deaths and suicide attempts in bipolar disorder: Part I of a report of the International Society for Bipolar Disorders Task Force on Suicide in Bipolar Disorder. **Australian and New Zealand Journal of Psychiatry**, v. 49, n. 9, p. 785–802, 2015.

SCHALPER, K. A.; BROWN, J.; CARVAJAL-HAUSDORF, D.; MCLAUGHLIN, J.; VELCHETI, V.; SYRIGOS, K. N.; HERBST, R. S.; RIMM, D. L. Objective Measurement and Clinical Significance of TILs in Non–Small Cell Lung Cancer. **JNCI: Journal of the National Cancer Institute**, v. 107, n. 3, p. dju435, 2015.

SCHERER, S.; STRATOU, G.; GRATCH, J.; MORENCY, L. P. Investigating voice quality as a speaker-independent indicator of depression and PTSD. **Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH**, n. August, p. 847–851, 2013.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond**. Cambridge, MA: The MIT Press, 2002.

SCHULLER, B.; HANTKE, S.; WENINGER, F.; HAN, W.; ZHANG, Z.; NARAYANAN, S. AUTOMATIC RECOGNITION OF EMOTION EVOKED BY GENERAL SOUND EVENTS Bj¨orn. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), **Anais**...2012.

SCOTT, S. B.; WHITEHEAD, B. R.; BERGEMAN, C. S.; PITZER, L. Combinations of stressors in midlife: Examining role and domain stressors using regression trees and random

forests. **Journals of Gerontology - Series B Psychological Sciences and Social Sciences**, v. 68, n. 3, p. 464–475, 2013.

SHARDA, M.; SUBHADRA, T. P.; SAHAY, S.; NAGARAJA, C.; SINGH, L.; MISHRA, R.; SEN, A.; SINGHAL, N.; ERICKSON, D.; SINGH, N. C. Sounds of melody—Pitch patterns of speech in autism. **Neuroscience Letters**, v. 478, n. 1, p. 42–45, 2010.
SHOUMAN, M.; TURNER, T.; STOCKER, R. Using decision tree for diagnosing heart disease patients. In: Proceedings of the Ninth Australasian Data Mining Conference, **Anais**...2011.

SILVA, V. M. B. da; VIANA, W. F.; GOMES, B. de L. X.; CAVALCANTI, R. S.; SOUZA, J. G. de; PEREIRA, J. M. S.; SANTOS, W. P. dos. Diagnóstico Precoce da Doença de Parkinson a Partir de Sinais Eletroencefalográficos e Inteligência Artificial. In: IV Simpósio de Inovação em Engenharia Biomédica - SABIO 2020, Recife, Brazil. **Anais**... Recife, Brazil: 2021.

SIMEONE, J. C.; WARD, A. J.; ROTELLA, P.; COLLINS, J.; WINDISCH, R. An evaluation of variation in published estimates of schizophrenia prevalence from 1990-2013: A systematic literature review. **BMC Psychiatry**, v. 15, n. 193, p. 1–14, 2015.

SINGH, A.; THAKUR, N.; SHARMA, A. A Review of Supervised Machine Learning Algorithms. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi. **Anais**... New Delhi: Bharati Vidyapeeth, New Delhi as the Organizer of INDIACom - 2016, 2016.

SONG, Y.; LU, Y. Decision tree methods: applications for classification and prediction. **Shanghai Archives of Psychiatry**, v. 27, n. 2, p. 130–135, 2015. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/>.

SOUSA, T. V.; VIVEIROS, V.; CHAI, M. V.; VICENTE, F. L.; JESUS, G.; CARNOT, M. J.; GORDO, A. C.; FERREIRA, P. L. Reliability and validity of the Portuguese version of the Generalized Anxiety Disorder (GAD-7) scale. **Health and Quality of Life Outcomes**, v. 13, n. 50, 2015. Disponível em: <???>.

SPAZZAPAN, E. A.; MARINO, V. C. de C.; CARDOSO, V. M.; BERTI, L. C.; FABRON, E. M. G. Características acústicas da voz em diferentes ciclos da vida: revisão integrativa da literatura. **Revista CEFAC**, v. 21, n. 3, 2019.

SPITZER, R. L.; KROENKE, K.; WILLIAMS, J. B. W.; LÖWE, B. A Brief Measure for Assessing Generalized Anxiety Disorder. **Archives of Internal Medicine**, v. 166, n. 10, p. 1092–1097, 2006.

SPRATLING, M. W. A review of predictive coding algorithms. **Brain and Cognition**, v. 112, p. 92–97, 2017. Disponível em: <http://dx.doi.org/10.1016/j.bandc.2015.11.003>.
SRINIVAS, K.; RANI, B. K.; GOVRDHAN, A. Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. **International Journal on Computer Science and Engineering (IJCSE)**, v. 2, n. 2, p. 250–255, 2010.

SRIVIDYA, M.; MOHANAVALLI, S.; BHALAJI, N. Behavioral Modeling for Mental Health using Machine Learning Algorithms. **Journal of Medical Systems**, v. 42, n. 88, 2018.

STORCHEUS, D.; ROSTAMIZADEH, A.; KUMAR, S. A Survey of Modern Questions and Challenges in Feature Extraction. In: Feature Extraction: Modern Questions and Challenges, **Anais**...2015.

STURIM, D.; TORRES-CARRASQUILLO, P.; QUATIERI, T. F.; MALYSKA, N.; MCCREE, A. Automatic detection of depression in speech using Gaussian mixture modeling with factor analysis. **Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH**, p. 2981–2984, 2011.

SUDHKAR, R. S.; ANIL, M. C. Analysis of Speech Features for Emotion Detection: A Review. In: 2015 International Conference on Computing Communication Control and Automation Analysis, Pune. **Anais**... Pune: 2015.

SUN, R.; MOORE, E. Investigating Glottal Parameters and Teager Energy Operators in Emotion Recognition. (S. D'Mello, A. Graesser, B. Schuller, J. Martin, Eds.) In: International Conference on Affective Computing and Intelligent Interaction, **Anais**...Springer Berlin Heidelberg, 2011.

TAGUCHI, T.; TACHIKAWA, H.; NEMOTO, K.; SUZUKI, M.; NAGANO, T.; TACHIBANA, R.; NISHIMURA, M.; ARAI, T. Major depressive disorder discrimination using vocal acoustic features. **Journal of Affective Disorders**, v. 225, p. 214–220, 2018. Disponível em: <http://dx.doi.org/10.1016/j.jad.2017.08.038>.

TAHIR, Y.; YANG, Z.; ID, D. C.; THALMANN, N.; THALMANN, D.; MANIAM, Y.; AMIRAH, N.; TAN, L.; LEE, J.; KEONG, C.; DAUWELS, J. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. **PLOS ONE**, p. 1–17, 2019.

TAHON, M.; DEGOTTEX, G.; DEVILLERS, L. Usual voice quality features and glottal features for emotional valence detection. In: Proceedings of the 6th International Conference on Speech Prosody, **Anais**...2012.

TANSKANEN, A.; TIIHONEN, J.; TAIPALE, H. Mortality in schizophrenia: 30-year nationwide follow-up study. **Acta Psychiatrica Scandinavica**, v. 138, n. 6, p. 492–499, 2018.

TAUD, H.; MAS, J. Multilayer Perceptron (MLP). In: CAMACHO OLMEDO, M.; PAEGELOW, M.; MAS, J.; ESCOBAR, F. (Ed.). **Geomatic Approaches for Modeling Land Change Scenarios. Lecture Notes in Geoinformation and Cartography.** [s.l.] Springer International Publishing, 2017. p. 451–455.

TELLES-CORREIA, D.; MOREIRA, A. L.; GONÇALVES, J. S. Hallucinations and related concepts - their conceptual background. **Frontiers in Psychology**, v. 6, n. 991, 2015.

TIMOFEEV, R. **Classification and regression trees (CART) theory and applications**. Berlin: Humboldt University, 2004.

TIMOSHENKO, E.; HARALD, H.; AG, S.; TECHNOLOGY, C. Using Speech Rhythm for Acoustic Language Identification. In: Eighth Annual Conference of the International Speech Communication Association, Antwerp. **Anais**... Antwerp: 2007.

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

TOROUS, J.; CERRATO, P.; HALAMKA, J. Targeting depressive symptoms with technology. **mHealth**, v. 5, p. 19–19, 2019.

TOVAR, A.; FUENTES-CLARAMONTE, P.; SOLER-VIDAL, J.; RAMIRO-SOUSA, N.; RODRIGUEZ-MARTINEZ, A.; SARRI-CLOSA, C.; SARRÓ, S.; LARRUBIA, J.; ANDRÉS-BERGARECHE, H.; MIGUEL-CESMA, M. C.; PABLO, P.; SALVADOR, R.; POMAROL-CLOTET, E.; HINZEN, W. The linguistic signature of hallucinated voice talk in schizophrenia. **Schizophrenia Research**, v. 206, p. 111–117, 2019. Disponível em: <https://doi.org/10.1016/j.schres.2018.12.004>.

US BURDEN OF DISEASE COLLABORATORS. The State of US health, 1990-2010: Burden of diseases, injuries, and risk factors. **JAMA - Journal of the American Medical Association**, v. 310, n. 6, p. 591–608, 2013.

UTANE, A. S.; NALBALWAR, S. L. Emotion Recognition through Speech. **International Journal of Applied Information Systems (IJAIS)**, p. 5–8, 2013.

VABALAS, A.; GOWEN, E.; POLIAKOFF, E.; CASSON, A. J. Machine learning algorithm validation with a limited sample size. **PLoS ONE**, v. 14, n. 11, p. e0224365, 2019. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0224365>.

VAN DER KOOY, K.; VAN HOUT, H.; MARWIJK, H.; MARTEN, H.; STEHOUWER, C.; BEEKMAN, A. Depression and the risk for cardiovascular diseases: Systematic review and meta analysis. **International Journal of Geriatric Psychiatry**, v. 22, n. 7, p. 613–626, 2007.

VAN DER WESTHUIZEN, C.; WYATT, G.; WILLIAMS, J. K.; STEIN, D. J.; SORSDAHL, K. Validation of the Self Reporting Questionnaire 20-Item (SRQ-20) for Use in a Low- and Middle-Income Country Emergency Centre Setting. **International Journal of Mental Health and Addiction**, v. 14, n. 1, p. 37–48, 2016.

VAN PUYVELDE, M.; NEYT, X.; MCGLONE, F.; PATTYN, N. Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance. **Frontiers in Psychology**, v. 9, n. NOV, p. 1–25, 2018.

VANELLO, N.; GUIDI, A.; GENTILI, C.; WERNER, S.; BERTSCHY, G.; VALENZA, G.; LANAT, A.; SCILINGO, E. P. Speech analysis for mood state characterization in bipolar patients. In: 34th Annual International Conference of the IEEE EMBS, **Anais**...2012.

VAPNIK, V.; IZMAILOV, R. Rethinking statistical learning theory: learning using statistical invariants. **Machine Learning**, v. 108, n. 3, p. 381–423, 2019. Disponível em: <https://doi.org/10.1007/s10994-018-5742-0>.

VEMBANDASAMY, K.; SASIPRIYA, R.; DEEPA, E. Heart Diseases Detection Using Naive Bayes Algorithm. **International Journal of Innovative Science, Engineering & Technology**, v. 2, n. 9, p. 441–444, 2015.

VERLEYSEN, M.; FRANÇOIS, D. The Curse of Dimensionality in Data Mining and Time

Series Prediction. In: Analysis work-conference on artificial neural networks, **Anais**...Springer-Verlag Berlin Heildelberg, 2005. Disponível em: <http://www.springerlink.com/index/n65tna6vwt3b1pw6.pdf>.

VERVERIDIS, D.; KOTROPOULOS, C.; PITAS, I. Automatic emotional speech classification. In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing., **Anais**...2004.

VIERA, A. J.; GARRETT, J. M. Understanding interobserver agreement: the kappa statistic. **Family Medicine**, v. 37, n. 5, p. 360–363, 2005. Disponível em: <http://www1.cs.columbia.edu/~julia/courses/CS6998/Interrater_agreement.Kappa_statistic.pdf>.

VILELA, J. A. A.; CRIPPA, J. A. S.; DEL-BEN, C. M.; LOUREIRO, S. R. Reliability and validity of a Portuguese version of the Young Mania Rating Scale. **Brazilian Journal of Medical and Biological Research**, v. 38, n. 9, p. 1429–1439, 2005.

VITTORINI, S.; CLERICO, A. Cardiovascular biomarkers: Increasing impact of laboratory medicine in cardiology practice. **Clinical Chemistry and Laboratory Medicine**, v. 46, n. 6, p. 748–763, 2008.

WAGER, S.; ATHEY, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. **Journal of the American Statistical Association**, v. 113, n. 523, p. 1228–1242, 2018. Disponível em: <https://doi.org/10.1080/01621459.2017.1319839>.

WAIKATO., T. M. L. G. at the U. of. **WEKA: The workbench for machine learning**. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 1 mar. 2021.

WANG, X.; ZHANG, C.; JI, Y.; SUN, L.; WU, L. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. In: **Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science**. Berlin, Heidelberg: Springer, 2013. p. 201–213.

WARREN, J. D.; JENNINGS, A. R.; GRIFFITHS, T. D. Analysis of the spectral envelope of sounds by the human brain. **NeuroImage**, v. 24, n. 4, p. 1052–1057, 2005.

WATSON, H. J.; SWAN, A.; NATHAN, P. R. Psychiatric diagnosis and quality of life: The additional burden of psychiatric comorbidity. **Comprehensive Psychiatry**, v. 52, n. 3, p. 265–272, 2011. Disponível em: <http://dx.doi.org/10.1016/j.comppsych.2010.07.006>.

WEEKS, J. W.; LEE, C.; REILLY, A. R.; HOWELL, A. N.; FRANCE, C.; KOWALSKY, J. M.; BUSH, A. Journal of Anxiety Disorders " The Sound of Fear ": Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. **Journal of Anxiety Disorders**, v. 26, n. 8, p. 811–822, 2012. Disponível em: <http://dx.doi.org/10.1016/j.janxdis.2012.07.005>.

WEEKS, J. W.; SRIVASTAV, A.; HOWELL, A. N.; MENATTI, A. R. "Speaking More than Words": Classifying Men with Social Anxiety Disorder via Vocal Acoustic Analyses of Diagnostic Interviews. **J Psychopathol Behav Assess**, v. 38, p. 30–41, 2016.

WEINBERGER, A. H.; GBEDEMAH, M.; MARTINEZ, A. M.; NASH, D.; GALEA, S.; GOODWIN, R. D. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. **Psychological Medicine**, p. 1–10, 2017.

WORLD HEALTH ORGANIZATION. **Depression and Other Common Mental Disorders: Global Health Estimates**, 2017.

WORLD HEALTH ORGANIZATION. **Depression**, 2018. Disponível em: <https://www.who.int/en/news-room/fact-sheets/detail/depression>. Acesso em: 11 nov. 2019a.

WORLD HEALTH ORGANIZATION. **Suicide**, 2018. Disponível em: <https://www.who.int/en/news-room/fact-sheets/detail/suicide>. Acesso em: 20 jan. 2021b.

XU, S. Bayesian Naïve Bayes classifiers to text classification. **Journal of Information Science**, v. 44, n. 1, p. 48–59, 2018.

XUE, B.; ZHANG, M.; MEMBER, S.; BROWNE, W. N. Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach. In: Ieee Transactions on Cybernetics, **Anais**...2012.

YADAV, V. K.; JAIN, A.; BHARGAV, L. Analysis and Comparison of Audio Compression Using Discrete Wavelet Transform. **International Journal of Advanced Research in Computer and Communication Engineering**, v. 4, n. 1, p. 310–313, 2015.

YANG, Y.; FAIRBAIRN, C.; COHN, J. F.; MEMBER, A. Detecting Depression Severity from Vocal Prosody. **IEEE Transactions on Affective Computing**, v. 4, n. 2, p. 142–150, 2013.

YEE, O. S.; SAGADEVAN, S.; HASHIMAH, N.; MALIN, A. H. Credit Card Fraud Detection Using Machine Learning As Data Mining Technique. **Journal of Telecommunication, Electronic and Computer Engineering (JTEC)**, v. 10, n. 1–4, p. 23–27, 2018.

YILDIRIM, A. A.; ÖZDOĞAN, C.; WATSON, D. Parallel data reduction techniques for big datasets. In: **Big Data: Concepts, Methodologies, Tools, and Applications**. [s.l.] IGI Global, 2016. p. 734–756.

YOUNG, R. C.; BIGGS, J. T.; ZIEGLER, V. E.; MEYER, D. A. A Rating Scale for Mania. **British Journal of Psychiatry**, v. 133, p. 429–435, 1978.

YU, A.; WANG, H. New Harmonicity Measures for Pitch Estimation and Voice Activity Detection. In: Eighth International Conference on Spoken Language Processing, **Anais**...2004.

YU, H.; KIM, S. SVM Tutorial-Classification, Regression and Ranking. In: **Handbook of Natural Computing**. [s.l: s.n.]1p. 479–506.

YUMOTO, E.; GOULD, J.; BAER, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. **The Journal of the Acoustical Society of America**, v. 71, n. 6, p. 1544–1550,

1982.

ZHANG, J.; PAN, Z.; GUI, C.; ZHU, J.; CUI, D. Clinical investigation of speech signal features among patients with schizophrenia. **Shanghai Archives of Psychiatry**, v. 28, n. 2, p. 95–102, 2016.

ZHAO, C.; ZHENG, C.; ZHAO, M.; LIU, J. Physiological Assessment of Driving Mental Fatigue Using Wavelet Packet Energy and Random Forests. **American Journal of Biomedical Sciences**, v. 2, n. 3, p. 262–274, 2010.

ZHOU, Y.; SUN, Y.; ZHANG, J.; YAN, Y. Speech emotion recognition using both spectral and prosodic features. In: International Conference on Information Engineering and Computer Science, **Anais**...2009.

ZIMMERMAN, M.; MARTINEZ, J. H.; YOUNG, D.; CHELMINSKI, I.; DALRYMPLE, K. Severity classification on the Hamilton depression rating scale. **Journal of Affective Disorders**, v. 150, n. 2, p. 384–388, 2013. Disponível em: <http://dx.doi.org/10.1016/j.jad.2013.04.028>.

ANEXO A – ESCALA BREVE DE AVALIAÇÃO PSIQUIÁTRICA (BPRS)

Nome:		
Idade:	Sexo:	Data:
Entrevistador:		

Favor atribuir o escore que melhor descreve a condição do paciente.

Escores: 0 = não avaliado, 1 = ausente, 2 = muito discreto, 3 = discreto, 4 = moderado, 5 = moderadamente grave, 6 = grave, 7 = extremamente grave

ITEM	ESCORE
1. Preocupação somática	
2. Ansiedade	
3. Retraimento afetivo	
4. Desorganização conceitual	
5. Sentimentos de culpa	
6. Tensão	
7. Maneirismos e postura	
8. Ideias de grandeza	
9. Humor depressivo	
10. Hostilidade	
11. Desconfiança	
12. Comportamento alucinatório (alucinações)	
13. Retardamento psicomotor/motor	
14. Falta de cooperação com a entrevista	
15. Alteração de conteúdo do pensamento (delírios)	
16. Afeto embotado	
17. Excitação	
18. Desorientação	
Escore Total:	

ANEXO B – ESCALA DE AVALIAÇÃO DE MANIA (EAM)

Nome:		
Idade:	Sexo:	Data:
Entrevistador:		

1. Humor e afeto elevados
(0) Ausência de elevação do humor ou afeto
(1) Humor ou afeto discreta ou possivelmente aumentados, quando questionado
(2) Relato subjetivo de elevação clara do humor; mostra-se otimista, autoconfiante, alegre; afeto apropriado ao conteúdo do pensamento
(3) Afeto elevado ou inapropriado ao conteúdo do pensamento; jocoso
(4) Eufórico; risos inadequados; cantando
(X) Não avaliado

2. Atividade motora - energia aumentada
(0) Ausente
(1) Relato subjetivo de aumento da energia ou atividade motora
(2) Apresenta-se animado ou com gestos aumentados
(3) Energia excessiva; às vezes hiperativo; inquieto (mas pode ser acalmado)
(4) Excitação motora; hiperatividade contínua (não pode ser acalmado)
(X) Não avaliado

3. Interesse sexual
(0) Normal; sem aumento
(1) Discreta ou possivelmente aumentado
(2) Descreve aumento subjetivo, quando questionado
(3) Conteúdo sexual espontâneo; discurso centrado em questões sexuais; auto-relato de hipersexualidade
(4) Relato confirmado ou observação direta de comportamento explicitamente sexualizado, pelo entrevistador ou outras pessoas
(X) Não avaliado

4. Sono
(0) Não relata diminuição do sono
(1) Dorme menos que a quantidade normal, cerca de 1 hora a menos do que o seu habitual
(2) Dorme menos que a quantidade normal, mais que 1 hora a menos do que o seu habitual
(3) Relata diminuição da necessidade de sono
(4) Nega necessidade de sono
(X) Não avaliado

Continua

Cont. Anexo B

5. Irritabilidade
(0) Ausente
(2) Subjetivamente aumentada
(4) Irritável em alguns momentos durante a entrevista; episódios recentes (nas últimas 24 horas) de ira ou irritação na enfermaria
(6) Irritável durante a maior parte da entrevista; ríspido e lacônico o tempo todo
(8) Hostil; não cooperativo; entrevista impossível
(X) Não avaliado

6. Fala (velocidade e quantidade)
(0) Sem aumento
(2) Percebe-se mais falante do que o seu habitual
(4) Aumento da velocidade ou quantidade da fala em alguns momentos; verborréico, às vezes (com solicitação, consegue-se interromper a fala)
(6) Quantidade e velocidade constantemente aumentadas; dificuldade para ser interrompido (não atende a solicitações; fala junto com o entrevistador)
(8) Fala pressionada, ininterruptível, contínua (ignora a solicitação do entrevistador)
(X) Não avaliado

7. Linguagem
(0) Sem alterações
(1) Circunstancial; pensamentos rápidos
(2) Perde objetivos do pensamento; muda de assuntos freqüentemente; pensamentos muito acelerados
(3) Fuga de idéias; tangencialidade; dificuldade para acompanhar o pensamento; ecolalia consonante
(4) Incoerência; comunicação impossível
(X) Não avaliado

8. Conteúdo
(0) Normal
(2) Novos interesses e planos compatíveis com a condição sócio-cultural do paciente, mas questionáveis
(4) Projetos especiais totalmente incompatíveis com a condição sócio-econômica do paciente; hiper-religioso
(6) Ideias supervalorizadas
(8) Delírios
(X) Não avaliado

9. Comportamento disruptivo agressivo
(0) Ausente, cooperativo
(2) Sarcástico; barulhento, às vezes, desconfiado

Continua

Cont. Anexo B

(4) Ameaça o entrevistador; gritando; entrevista dificultada
(6) Agressivo; destrutivo; entrevista impossível
(X) Não avaliado

10. Aparência
(0) Arrumado e vestido apropriadamente
(1) Descuidado minimamente; adornos ou roupas minimamente inadequados ou exagerados
(2) Precariamente asseado; despenteado moderadamente; vestido com exagero
(3) Desgrenhado; vestido parcialmente; maquiagem extravagante
(4) Completamente descuidado; com muitos adornos e adereços; roupas bizarras
(X) Não avaliado

11. Insight (discernimento)
(0) Insight presente: espontaneamente refere estar doente e concorda com a necessidade de tratamento
(1) Insight duvidoso: com argumentação, admite possível doença e necessidade de tratamento
(2) Insight prejudicado: espontaneamente admite alteração comportamental, mas não a relaciona com a doença, ou discorda da necessidade de tratamento
(3) Insight ausente: com argumentação, admite de forma vaga alteração comportamental, mas não a relaciona com a doença e discorda da necessidade de tratamento
(4) Insight ausente: nega a doença, qualquer alteração comportamental e necessidade de tratamento
(X) Não avaliado

Escore Total: _____

ANEXO C – ESCALA DE DEPRESSÃO DE HAMILTON (HAM-D 17)

Nome:		
Idade:	Sexo:	Data:
Entrevistador:		

1	**HUMOR DEPRIMIDO** 0. Ausente 1. Sentimentos relatados apenas ao ser perguntado 2. Sentimentos relatados espontaneamente, com palavras 3. Comunica os sentimentos com expressão facial, postura, voz e tendência ao choro 4. Sentimentos deduzidos da comunicação verbal e não verbal do paciente
2	**SENTIMENTOS DE CULPA** 0. Ausentes 1. Autorrecriminação; sente que decepcionou os outros 2. Ideias de culpa ou ruminação sobre erros passados ou más ações 3. A doença atual é um castigo. Delírio de culpa 4. Ouve vozes de acusação ou denúncia e/ou tem alucinações visuais ameaçadoras
3	**SUICÍDIO** 0. Ausente 1. Sente que a vida não vale a pena 2. Desejaria estar morto; pensa na possibilidade de sua morte 3. Ideias ou gestos suicidas 4. Tentativa de suicídio (qualquer tentativa séria)
4	**INSÔNIA INICIAL** 0. Sem dificuldade 1. Tem alguma dificuldade ocasional, isto é, mais de meia hora 2. Queixa de dificuldade para conciliar todas as noites
5	**INSÔNIA INTERMEDIÁRIA** 0. Sem dificuldade 1. Queixa-se de inquietude e perturbação durante a noite 2. Acorda à noite; qualquer saída da cama (exceto para urinar)
6	**INSÔNIA TARDIA** 0. Sem dificuldade 1. Acorda de madrugada, mas volta a dormir 2. Incapaz de voltar a conciliar o sono ao deixar a cama

Continua

Cont. Anexo C

7	**TRABALHOS E ATIVIDADES** 0. Sem dificuldade 1. Pensamento/sentimento de incapacidade, fadiga, fraqueza relacionada às atividades; trabalho ou passatempos 2. Perda de interesse por atividades (passatempos, trabalho) – quer diretamente relatada pelo paciente, ou indiretamente, por desatenção, indecisão e vacilação (sente que precisa se esforçar para o trabalho ou atividades). 3. Diminuição do tempo gasto em atividades ou queda da produtividade. No hospital, marcar 3 se o paciente passa menos de 3h em atividades externas (passatempos ou trabalho hospitalar) 4. Parou de trabalhar devido à doença atual. No hospital, marcar 4 se o paciente não se ocupar de outras atividades além de pequenas tarefas do leito, ou for incapaz de realizá-las sem auxílio
8	**RETARDO** 0. Pensamento e fala normais 1. Leve retardo durante a entrevista 2. Retardo óbvio à entrevista 3. Estupor completo
9	**AGITAÇÃO** 0. Nenhuma 1. Brinca com as mãos ou com os cabelos, etc. 2. Torce as mãos, rói as unhas, puxa os cabelos, morde os lábios
10	**ANSIEDADE PSÍQUICA** 0. Sem ansiedade 1. Tensão e irritabilidade subjetivas 2. Preocupação com trivialidades 3. Atitude apreensiva aparente no rosto ou fala 4. Medos expressos sem serem inquiridos
11	**ANSIEDADE SOMÁTICA** (sintomas fisiológicos de ansiedade: boca seca, flatulência, indigestão, diarreia, cólicas, eructações; palpitações, cefaleia, hiperventilação, suspiros, sudorese, frequência urinária**)** 0. Ausente 1. Leve 2. Moderada 3. Grave 4. Incapacitante

Continua

Cont. Anexo C

12	**SINTOMAS SOMÁTICOS GASTROINTESTINAIS** 0. Nenhum 1. Perda do apetite, mas alimenta-se voluntariamente; sensações de peso no abdome 2. Dificuldade de comer se não insistirem. Solicita ou exige laxativos ou medicações para os intestinos ou para sintomas digestivos
13	**SINTOMAS SOMÁTICOS EM GERAL** 0. Nenhum 1. Peso nos membros, costas ou cabeça. Dores nas costas, cefaleia, mialgia. Perda de energia e cansaço 2. Qualquer sintoma bem caracterizado e nítido, marcar 2
14	**SINTOMAS GENITAIS (perda da libido, sintomas menstruais)** 0. Ausentes 1. Leves distúrbios menstruais 2. Intensos
15	**HIPOCONDRIA** 0. Ausente 1. Auto-observação aumentada (com relação ao corpo) 2. Preocupação com a saúde 3. Queixas frequentes, pedidos de ajuda, etc. 4. Ideias delirantes hipocondríacas
16	**PERDA DE PESO** (Marcar A ou B; A – pela história; B – pela avaliação semanal do psiquiatra responsável) **A.** 0. Sem perda de peso 1. Provável perda de peso da doença atual 2. Perda de peso definida **B.** 0. Menos de 0,5kg de perda por semana 1. Mais de 0,5kg de perda por semana 2. Mais de 1kg de perda por semana
17	**CONSCIÊNCIA DA DOENÇA** 0. Reconhece que está deprimido e doente 1. Reconhece a doença, mas atribui a causa à má alimentação, ao clima, ao excesso de trabalho, a vírus, necessidade de repouso 2. Nega estar doente

Escore total: _____

ANEXO D – ESCALA DE TRANSTORNO DE ANSIEDADE GENERALIZADA (GAD-7)

Nome:		
Idade:	Sexo:	Data:
Entrevistador:		

Durante as últimas 02 semanas, com que frequência você foi incomodado (a) pelos problemas abaixo? (Marque sua resposta com "X").

	Nenhuma vez	**Vários dias**	**Mais da metade dos dias**	**Quase todos os dias**
1. Sentir-se nervoso, ansioso ou muito tenso	(0)	(1)	(2)	(3)
2. Não ser capaz de impedir ou de controlar as preocupações	(0)	(1)	(2)	(3)
3. Preocupar-se muito com diversas coisas	(0)	(1)	(2)	(3)
4. Dificuldade para relaxar	(0)	(1)	(2)	(3)
5. Ficar tão agitado/a que se torna difícil permanecer sentado/a	(0)	(1)	(2)	(3)
6. Ficar facilmente aborrecido/a ou irritado/a	(0)	(1)	(2)	(3)
7. Sentir medo como se algo horrível fosse acontecer	(0)	(1)	(2)	(3)

Escore total: _____

ANEXO E – SELF-REPORTING QUESTIONNAIRE (SRQ-20)

Teste que avalia o sofrimento mental. Por favor, leia estas instruções antes de preencher as questões abaixo. É muito importante que todos que estão preenchendo o questionário sigam as mesmas instruções.

Instruções:

Estas questões são relacionadas a certas dores e problemas que podem ter lhe incomodado nos últimos 30 dias. Se você acha que a questão se aplica a você e você teve o problema descrito nos últimos 30 dias responda SIM. Por outro lado, se a questão não se aplica a você, e você não teve o problema nos últimos 30 dias, responda NÃO.

OBS: Lembre-se de que o diagnóstico definitivo só pode ser fornecido por um profissional.

PERGUNTAS	RESPOSTAS	
	SIM	NÃO
1- Você tem dores de cabeça frequentes?		
2- Tem falta de apetite?		
3- Dorme mal?		
4- Assusta-se com facilidade?		
5- Tem tremores nas mãos?		
6- Sente-se nervoso(a), tenso(a) ou preocupado(a)?		
7- Tem má digestão?		
8- Tem dificuldades de pensar com clareza?		
9- Tem se sentido triste ultimamente?		
10- Tem chorado mais do que costume?		
11- Encontra dificuldades para realizar com satisfação suas atividades diárias?		
12- Tem dificuldades para tomar decisões?		
13- Tem dificuldades no serviço (seu trabalho é penoso, lhe causa sofrimento?)		
14- É incapaz de desempenhar um papel útil em sua vida?		
15- Tem perdido o interesse pelas coisas?		

Continua

Cont. Anexo E

16- Você se sente uma pessoa inútil, sem préstimo?		
17- Tem tido ideia de acabar com a vida?		
18- Sente-se cansado(a) o tempo todo?		
19- Você se cansa com facilidade?		
20- Têm sensações desagradáveis no estômago?		

Nome: _____

Escore: _____ Data da Avaliação: _____