

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE

**MODELOS DE APRENDIZADO DE MÁQUINA PARA A PREDIÇÃO
DO ESTÁGIO DE PRONTIDÃO PARA TERAPIA DE VOZ**

MARIA JÚLIA GALINDO SOARES

João Pessoa – PB
2025

MARIA JÚLIA GALINDO SOARES

**MODELOS DE APRENDIZADO DE MÁQUINA PARA A PREDIÇÃO
DO ESTÁGIO DE PRONTIDÃO PARA TERAPIA DE VOZ**

Dissertação apresentada à banca examinadora do Programa de Pós-Graduação em Modelos de Decisão e Saúde, do Centro de Ciências Exatas e da Natureza, da Universidade Federal da Paraíba, como requisito para o Mestrado em Modelos de Decisão e Saúde.

Linha de Pesquisa: Modelos em Saúde.

Orientadores: Prof^a. Dr^a. Anna Alice Figueiredo de Almeida Queiroz e Prof^a. Dr^a. Ana Hermínia Andrade e Silva.

Catálogo na publicação
Seção de Catalogação e Classificação

S676m Soares, Maria Júlia Galindo.

Modelos de aprendizado de máquina para a predição do estágio de prontidão para terapia de voz / Maria Júlia Galindo Soares. - João Pessoa, 2025.

131 f. : il.

Orientação: Anna Alice Figueiredo de Almeida, Ana Hermínia Andrade e Silva.

Dissertação (Mestrado) - UFPB/CCEN.

1. Fonoaudiologia. 2. Modelos de aprendizado de máquina - Disfonia. 3. Terapia de voz. I. Almeida, Anna Alice Figueiredo de. II. Silva, Ana Hermínia Andrade e. III. Título.

UFPB/BC

CDU 616.89-008.434.5(043)

Ata da Sessão Pública da Defesa de
Dissertação de Mestrado de **MARIA JÚLIA
GALINDO SOARES**, candidata ao Título
de Mestre em Modelos de Decisão e Saúde,
realizada em **25 de agosto de 2025**.

Aos vinte e cinco dias do mês de agosto do ano de dois mil e vinte e cinco, às 09h00min, no Laboratório Integrado de Estudos da Voz (LIEV), instalou-se a banca examinadora de dissertação de Mestrado da aluna MARIA JÚLIA GALINDO SOARES. A banca examinadora foi composta pelos professores: Dra. LETÍCIA CALDAS TEIXEIRA, UFMG, examinadora externa, Dr. MARCELO RODRIGO PORTELA FERREIRA, UFPB, e Dr. LEONARDO WANDERLEY LOPES, UFPB, examinadores internos, Dra. ANA HERMINIA ANDRADE E SILVA, UFPB, como orientadora e Dra. ANNA ALICE F. DE ALMEIDA QUEIROZ, UFPB, como orientadora e presidente da banca examinadora. Dando início aos trabalhos, a presidente da banca cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra a candidata para que se fizesse, oralmente, a exposição do trabalho de dissertação intitulado “MODELOS PARA A PREDIÇÃO DO ESTÁGIO DE PRONTIDÃO PARA TERAPIA DE VOZ”. Concluída a apresentação a candidata foi arguida pela Banca Examinadora que emitiu o parecer “**APROVADA**”. Sendo assim, após a candidata proceder às devidas correções exigidas pela Banca Examinadora no prazo máximo de **30 dias** e depositar as cópias finais da dissertação, deverá a Coordenação expedir uma declaração de conclusão do Programa e a Universidade Federal da Paraíba deverá expedir o respectivo diploma de Mestre em Modelos de Decisão e Saúde, na forma da lei. E, para constar, o Prof. Hemílio Fernandes Campos Coêlho, Sr, Coordenador do Programa de Pós-Graduação em Modelos de Decisão e Saúde, lavrou a presente Ata, que vai assinada por ele e pelos demais membros da Banca Examinadora. João Pessoa, 25 de agosto de 2025.

Prof. Anna Alice F. de Almeida Queiroz _____
Orientadora e Presidente

Profa. Ana Hermínia Andrade e Silva _____
Orientadora

Prof. Marcelo Rodrigo Portela Ferreira _____
Examinador Interno

Prof. Leonardo Wanderley Lopes _____
Examinador Interno

Profa. Letícia Caldas Teixeira _____
Examinadora Externa à Instituição



Documento assinado digitalmente

LETICIA CALDAS TEIXEIRA

Data: 26/08/2025 15:08:52-0300

Verifique em <https://validar.iti.gov.br>

Emitido em 25/08/2025

ATA Nº 12/2025 - PPGMDS (11.01.14.54)
(Nº do Documento: 12)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 27/08/2025 15:41)
MARCELO RODRIGO PORTELA FERREIRA
PROFESSOR DO MAGISTERIO SUPERIOR
1584539

(Assinado digitalmente em 18/09/2025 12:48)
LEONARDO WANDERLEY LOPES
PROFESSOR TITULAR-LIVRE MAG SUPERIOR
2634755

(Assinado digitalmente em 27/08/2025 11:06)
ANNA ALICE FIGUEIREDO DE ALMEIDA QUEIROZ
CHEFE DE DEPARTAMENTO
1668545

(Assinado digitalmente em 27/08/2025 12:14)
ANA HERMINIA ANDRADE E SILVA
PROFESSOR DO MAGISTERIO SUPERIOR
1147621

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número:
12, ano: 2025, documento (espécie): ATA, data de emissão: 27/08/2025 e o código de verificação: e69b9e5cda

Dedico aos meus pais, Jocineide Galindo de Almeida Soares e José Maria Soares, que sempre incentivaram e apoiaram meus estudos. Minhas conquistas também são deles, que me ensinaram o que era ser amada e me permitiram alcançar altos voos.

AGRADECIMENTOS

À Deus, que me guiou durante toda a minha jornada, abençoando meus caminhos e fazendo com que fosse possível essa dissertação.

À Nossa Senhora de Aparecida por interceder por mim diariamente.

À Santa Rita de Cássia, padroeira das causas impossíveis, que nas vezes que pensei que não ia conseguir, me lembrou que para Deus nada é impossível.

Aos meus pais, Jocineide e José Maria, que me ensinaram que eu poderia vencer através dos estudos e apoiaram os meus sonhos.

Ao meu esposo, Geraldo Paulino, que esteve ao meu lado durante esse processo, sempre me motivou e me deu o suporte necessário.

Aos meus demais familiares, que acreditaram na minha capacidade e torceram para que obtivesse sucesso.

Aos meus queridos amigos, que me escutaram, incentivaram e deram forças nos momentos que precisei.

As minhas orientadoras, Anna Alice e Ana Hermínia, por acreditarem em mim, me ajudarem a ampliar meu potencial e me tornar uma pesquisadora melhor.

Aos membros da banca, Letícia, Leonardo e Marcelo, pelo tempo dedicado e contribuições realizadas para melhora da pesquisa.

Aos meus colegas de turma do PPGMDS, pelos momentos compartilhados.

Aos professores do PPGMDS, pelos conhecimentos transmitidos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo essencial apoio financeiro.

"No final, tudo dá certo. Se não deu certo, é porque ainda não chegou ao fim."

(Fernando Sabino)

RESUMO

A avaliação da voz é multidimensional por abordar diversos aspectos, como os principais métodos de avaliação: o julgamento perceptivo-auditivo (JPA), a avaliação acústica, a avaliação laringológica e a autoavaliação. Esta última contempla a perspectiva do paciente e proporciona informações da vivência sobre sintomas, impactos na qualidade de vida e aspectos cognitivos comportamentais relativos à voz. O estágio de prontidão para terapia de voz surge nesse contexto e reflete a disposição do paciente em adotar mudanças que contribuam para o seu comportamento vocal. Identificar o estágio do indivíduo permite uma intervenção mais direcionada e eficaz. Ademais, a utilização de modelos de aprendizado de máquina (AM) para a predição dos estágios de prontidão representa uma inovação para agregar a tomada de decisão clínica na área de avaliação e tratamento de disfonia. Desta forma, o objetivo desta pesquisa é avaliar e comparar o desempenho de modelos de AM para predição do estágio de prontidão para terapia fonoaudiológica de voz. Trata-se de um estudo observacional, analítico e quantitativo, de delineamento transversal, conduzido a partir da análise retrospectiva de dados clínicos, com aplicação de modelos preditivos de AM para predição do estágio de prontidão. As fontes de dados para alimentar os modelos foram obtidas da avaliação fonoaudiológica e da avaliação laringológica iniciais de pacientes, de ambos os sexos, que buscam atendimento no Laboratório Integrado de Estudos da Voz, da Universidade Federal da Paraíba. Foram extraídos os dados do Protocolo de Anamnese e Avaliação Vocal e a Escala University of Rhode Island Change Assessment - Voice Validated, do Índice de Desvantagem Vocal, das Escalas de Qualidade de Vida em Voz, Sintomas Vocais e de Desconforto do Trato Vocal e dos dados do JPA, da acústica e do diagnóstico laríngeo. A amostra foi composta por 236 indivíduos, divididos em dois grupos dos estágios de prontidão: contemplação e manutenção. Foi realizada a análise estatística descritiva das variáveis quantitativas e qualitativas e foram testados algoritmos de AM supervisionado: regressão logística, k-nearest neighbors (KNN), naive bayes, árvore de decisão e floresta aleatória. O desempenho dos modelos foi avaliado por meio de acurácia, sensibilidade, especificidade, entre outras medidas. Os resultados indicam que os modelos KNN e Floresta Aleatória apresentaram melhor desempenho nos dados da anamnese. Para os dados do JPA, diagnóstico laríngeo e classificação da disfonia, os modelos Regressão Logística e Naive Bayes demonstraram desempenho balanceado entre sensibilidade e especificidade. Nos dados dos instrumentos de autoavaliação e das medidas acústicas, os modelos apresentaram desempenho inferior. Nos modelos com dados combinados das diferentes dimensões da avaliação vocal, o KNN se destacou, apresentando alta acurácia. Embora ainda apresentem limitações em seu poder preditivo, esses achados sugerem que os modelos de AM possuem potencial para contribuir na identificação do estágio de prontidão, possibilitando estratégias de intervenção mais direcionadas na terapia de voz.

Descritores: Modelo Transteórico; Voz; Disfonia; Fonoaudiologia; Aprendizado de Máquina.

ABSTRACT

Voice assessment is multidimensional as it addresses various aspects, including the main evaluation methods: perceptual-auditory judgment (PAJ), acoustic assessment, laryngological evaluation, and self-assessment. The latter encompasses the patient's perspective and provides information on the experience of symptoms, impacts on quality of life, and cognitive-behavioral aspects related to voice. The stage of readiness for voice therapy emerges in this context and reflects the patient's willingness to adopt changes that contribute to their vocal behavior. Identifying the individual's stage allows for a more targeted and effective intervention. Moreover, the use of machine learning (ML) models to predict stages of readiness represents an innovation to support clinical decision-making in the evaluation and treatment of dysphonia. The objective of this research is to evaluate and compare the performance of ML models for predicting the stage of readiness for speech-language voice therapy. This is an observational, analytical, and quantitative study with a cross-sectional design, conducted through the retrospective analysis of clinical data, applying predictive ML models to predict the stage of readiness. The data sources for the models were obtained from the initial speech-language and laryngological assessments of patients of both sexes seeking care at the Integrated Voice Studies Laboratory of the Federal University of Paraíba. Data were extracted from the Anamnesis and Vocal Assessment Protocol, the University of Rhode Island Change Assessment – Voice Validated Scale, the Vocal Handicap Index, the Voice-Related Quality of Life Scales, Vocal Symptoms and Vocal Tract Discomfort Scales, and from PAJ, acoustic measures, and laryngological diagnosis. The sample comprised 236 individuals, divided into two readiness stage groups: contemplation and maintenance. Descriptive statistical analysis was performed for quantitative and qualitative variables, and supervised ML algorithms were tested, including logistic regression, k-nearest neighbors (KNN), naive Bayes, decision tree, and random forest. Model performance was evaluated through accuracy, sensitivity, specificity, among other measures. Results indicate that the KNN and Random Forest models performed best with the anamnesis data. For PAJ data, laryngological diagnosis, and dysphonia classification, the Logistic Regression and Naive Bayes models demonstrated balanced performance between sensitivity and specificity. For self-assessment instruments and acoustic measures, the models showed lower performance. In models combining data from different dimensions of voice assessment, KNN stood out, presenting high accuracy. Although still limited in predictive power, these findings suggest that ML models have the potential to contribute to identifying the stage of readiness, enabling more targeted intervention strategies in voice therapy.

Keywords: Transtheoretical Model; Voice; Dysphonia; Speech Therapy; Machine Learning.

LISTA DE FIGURAS

Figura 1 - Funcionamento do algoritmo KNN.	29
Figura 2 - Floresta Aleatória.	33
Figura 3 - Matriz de Gráficos de Correlação: dados da Anamnese.	47
Figura 4 - Efeito Marginal do Modelo de Regressão Logística: Dados da Anamnese.	52
Figura 5 - Número de Vizinhos do KNN com Base na Acurácia: Dados da Anamnese.	53
Figura 6 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados da Anamnese.	54
Figura 7 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia: Dados da Anamnese.	55
Figura 8 - Árvore de Decisão com os Dados da Anamnese.	56
Figura 9 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados da Anamnese.	57
Figura 10 - Erro OOB no Modelo Floresta Aleatória: Dados da Anamnese.	58
Figura 11 - Valores de AUC dos Modelos: Dados da Anamnese.	61
Figura 12 - Efeito Marginal do Modelo de Regressão Logística: Dados dos Instrumentos de Autoavaliação.	63
Figura 13 - Número de Vizinhos do KNN com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.	64
Figura 14 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.	65
Figura 15 - Árvore de Decisão com Base nos Dados dos Instrumentos de Autoavaliação.	67
Figura 16 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.	68
Figura 17 - Erro OOB no Modelo Floresta Aleatória: Dados dos Instrumentos de Autoavaliação.	69
Figura 18 - Valores de AUC dos Modelos: Dados dos Instrumentos de Autoavaliação.	72
Figura 19 - Efeito Marginal do Modelo de Regressão Logística: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	73
Figura 20 - Número de Vizinhos do KNN com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	74
Figura 21 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	75
Figura 22 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	76
Figura 23 - Árvore de Decisão de Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	77
Figura 24 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	78
Figura 25 - Erro OOB do Modelo Floresta Aleatória: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.	79
Figura 26 - Valores de AUC dos Modelos: Dados clínicos JPA, Diagnóstico Laríngeo e Classificação da Disfonia.	82
Figura 27 - Efeito Marginal do Modelo de Regressão Logística: Medidas Acústicas.	84
Figura 28 - Número de Vizinhos do KNN com Base na Acurácia: Medidas Acústicas.	85
Figura 29 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Medidas Acústicas.	86
Figura 30 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia: Medidas Acústicas.	87

Figura 31 – Árvore de Decisão das Medidas Acústicas.....	88
Figura 32 - Hiperparâmetros do Floresta Aleatória com Base na Acurácia: Medidas Acústicas.	89
Figura 33 – Erro OOB do Modelo de Floresta Aleatória: Medidas acústicas.	90
Figura 34 - Valores de AUC dos Modelos: Medidas Acústicas.	93
Figura 35 - Efeitos Marginais do Modelo de Regressão Logística: Dados Multidimensionais da Voz.	95
Figura 36 - Número de Vizinhos do KNN com Base na Acurácia: Dados Multidimensionais da Voz.	96
Figura 37 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados Multidimensionais da Voz.	97
Figura 38 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia: Dados Multidimensionais da Voz.	98
Figura 39 – Árvore de Decisão dos Dados Multidimensionais da Voz.	99
Figura 40 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Multidimensionais da Voz.	100
Figura 41 - Erro OOB do Modelo Floresta Aleatória: Dados Multidimensionais da Voz.....	101
Figura 42 – Valores de AUC dos Modelos: Dados Multidimensionais da Voz.	103

LISTA DE TABELAS

Tabela 1 - Comparação das variáveis da anamnese com o estágio de prontidão.	42
Tabela 2 - Comparação de instrumentos de autoavaliação com o estágio de prontidão.....	47
Tabela 3 - Comparação das variáveis referentes ao JPA, Diagnóstico Laríngeo e Classificação da Disfonia com estágio de prontidão.	48
Tabela 4 - Matriz de confusão OOB: dados da Anamnese.	58
Tabela 5 – Matriz de Confusão da Floresta Aleatória: Dados da Anamnese.	59
Tabela 6 – Desempenho dos Modelos para Classificação do Estágio de Prontidão com Dados da Anamnese.	59
Tabela 7 - Matriz de Confusão OOB: Dados dos Instrumentos de Autoavaliação.	69
Tabela 8 – Matriz de Confusão da Floresta Aleatória: Dados dos Instrumentos de Autoavaliação.	70
Tabela 9 - Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados dos Instrumentos de Autoavaliação.	70
Tabela 10 – Matriz de Confusão OOB: dados clínicos do JPA, Diagnóstico Laríngeo e Disfonia.....	79
Tabela 11 – Matriz de Confusão da Floresta Aleatória: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.....	80
Tabela 12 – Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados Clínicos do JPA, Diagnóstico Laríngeo e Classificação da Disfonia para classificação.	80
Tabela 13 – Matriz de Confusão OOB: Medidas Acústicas.....	89
Tabela 14 – Matriz de Confusão da Floresta Aleatória: Medidas Acústicas.....	90
Tabela 15 – Desempenho dos Modelos Para Classificação do Estágio de Prontidão com Dados das Medidas Acústicas.	91
Tabela 16 – Matriz de Confusão OOB: Dados Multidimensionais da Voz.....	100
Tabela 17 – Matriz de Confusão da Floresta Aleatória: Dados Multidimensionais da Voz.	101
Tabela 18 - Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados Multidimensionais da Voz.	102

SUMÁRIO

1. INTRODUÇÃO	13
2. OBJETIVOS	16
2.1 Objetivo Principal	16
2.2 Objetivos Específicos	16
3. FUNDAMENTAÇÃO TEÓRICA.....	17
3.1 Avaliação Multidimensional da Voz	17
3.2 Estágio de Prontidão para a terapia de Voz	21
3.3 Aprendizado de Máquina (AM)	24
3.3.1 Regressão Logística	26
3.3.2 K-Nearest Neighbors (KNN).....	28
3.3.3 Naive Bayes.....	30
3.3.4 Árvore de Decisão	31
3.3.5 Floresta Aleatória	32
4. MÉTODOS.....	34
4.1. Delineamento do Estudo	34
4.2 População e Amostra	34
4.3 Instrumentos de Coleta de Dados	34
4.3.1 Protocolo de Anamnese e Avaliação Vocal (PAAV)	34
4.3.2 URICA-VV.....	36
4.4 Fonte de Dados.....	37
4.5 Análise de dados.....	39
4.6 Considerações éticas	41
5. RESULTADOS E DISCUSSÃO	41
5.1 Análise Descritiva.....	41
5.2 Modelos de AM – Dados da Anamnese.....	51
5.3 Modelos de AM - Dados dos Instrumentos de Autoavaliação	62
5.4 Modelos de AM – Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia	73
5.5 Modelos de AM – Dados das Medidas Acústicas	83
5.6 Modelos de AM – Dados Multidimensionais da Voz	94
6. CONCLUSÃO.....	106
REFERÊNCIAS.....	107
ANEXO 1.....	116
ANEXO 2.....	119
ANEXO 3.....	120
APÊNDICE	125

1. INTRODUÇÃO

A voz humana é um instrumento complexo e essencial na comunicação interpessoal. Desempenha um papel crucial na expressão de emoções, intenções e informações. A qualidade da voz pode ser afetada por diversos fatores, que resulta em distúrbios vocais, que podem impactar significativamente a vida social, profissional e emocional dos indivíduos (Behlau, 2001). A disfonia é um distúrbio comumente observado e de origens variadas, entre elas abuso vocal, lesões nas pregas vocais ou condições médicas subjacentes; seus sintomas, como rouquidão, esforço ao falar e perda de voz, podem comprometer a saúde vocal, o bem-estar emocional e a vida social do indivíduo (Constantini; Ribeiro; Behlau, 2022).

A avaliação da voz é multidimensional por abordar diversos aspectos, como o julgamento perceptivo-auditivo (JPA) conduzido por profissionais treinados, que avalia a qualidade vocal do paciente (Patel *et al.*, 2018; Roy *et al.*, 2013; Carding *et al.*, 2009; Dejonckere *et al.*, 2001). Dentre as demais dimensões que estão focadas na perspectiva do clínico, pode-se destacar a avaliação instrumental, como a avaliação acústica qualitativa e quantitativa para extração de medidas, e a avaliação laringológica que oferece uma visualização direta das estruturas da laringe para trazer informações de possíveis alterações funcionais e/ou estruturais (Payten *et al.*, 2022; Almeida *et al.*, 2021; Lopes *et al.*, 2021).

Somado às anteriores, elenca-se a dimensão que contempla a perspectiva do paciente, quem vivencia o problema de voz, a autoavaliação. Os instrumentos de autoavaliação são respondidos pelo paciente e proporcionam informações da vivência sobre sintomas, impactos na qualidade de vida e aspectos cognitivos comportamentais relativos à voz (Behlau; Madazio; Yamasaki, 2023; Behlau *et al.*, 2022; Almeida *et al.*, 2024; Almeida *et al.*, 2021). Deste modo, essa abordagem multidimensional auxilia a compreender profundamente o comportamento vocal, além de permitir diagnósticos precisos para um planejamento de tratamento eficaz e personalizado.

O estágio de prontidão para terapia de voz surge dentro desse contexto de aspectos cognitivos comportamentais a serem avaliados. Emerge como um conceito central, que reflete a disposição do paciente em adotar mudanças que contribuam para o seu comportamento vocal. O estágio de prontidão pode ser explicado com base no Modelo Transteórico (MTT) (Prochaska, DiClemente, 1982). É uma teoria

amplamente utilizada para compreender como as pessoas mudam comportamentos inadequados, que inclui aqueles relacionados à saúde. O modelo propõe que a mudança de comportamento ocorra em estágios sequenciais, cada um representa um nível diferente de prontidão para mudança comportamental. Os estágios descritos no MTT são pré-contemplação, contemplação, preparação, ação e manutenção (Prochaska, DiClemente, 1982).

É essencial promover mudanças comportamentais na terapia de voz, pois as disfonias podem ter origem ou serem agravadas por comportamentos vocais inadequados, assim é necessária a participação ativa do paciente para obter resultados eficazes, pois a prontidão não se limita apenas ao seguimento das orientações dos profissionais de saúde, mas também inclui a compreensão e aceitação do diagnóstico, como o comprometimento com o tratamento, de modo que avaliar as características individuais que influenciam a prontidão permite uma intervenção mais direcionada e eficaz, a fim de minimizar as falhas no tratamento e melhorar o prognóstico a longo prazo (Costa *et al.*, 2017).

Nesse cenário, a University of Rhode Island Change Assessment - Voice Validated (URICA-VV), adaptada (Teixeira *et al.*, 2013) e validada para o português brasileiro (Aguiar *et al.*, 2021), surge como um instrumento promissor para avaliar o estágio de prontidão dos pacientes em modificar comportamentos vocais prejudiciais e aderir ao tratamento recomendado. Desse modo, a classificação adequada do estágio de prontidão para a mudança pode otimizar a intervenção terapêutica.

Para otimizar ainda mais essa classificação, utilização de modelos de aprendizado de máquina (AM) para a predição dos estágios de prontidão representa uma inovação metodológica para facilitar a tomada de decisão clínica na área de avaliação e terapia de voz, e contribui para o desenvolvimento teórico e prático. Recentes avanços no campo do AM oferecem novas perspectivas em seu uso na área da saúde, pois essas técnicas podem permitir uma análise mais objetiva e quantitativa dos dados, com vistas a explorar padrões complexos que podem não ser facilmente discerníveis por métodos tradicionais (Habebeh; Gohel, 2021).

Na área de voz, o AM possibilitou o desenvolvimento de modelos em diferentes contextos clínicos e populacionais demonstrando-se eficazes tanto em tarefas de classificação binária quanto multiclasse, como na identificação da doença de Parkinson com base na qualidade vocal (Suppa *et al.*, 2022), na detecção do tremor vocal e sua resposta ao tratamento farmacológico em pacientes com tremor essencial

(Suppa *et al.*, 2021), na classificação de distúrbios neurológicos da voz (Yagnavajjula *et al.*, 2024), para distinguir disartria hipocinética da presbifonia (Byeon, 2021), na voz como um biomarcador para detectar depressão menor e maior (Shin *et al.*, 2021), na identificação de parâmetros relevantes para distúrbios funcionais da voz (Schlegel *et al.*, 2020) e na classificação de vozes disfônicas e não disfônicas (Leite; Moraes; Lopes, 2025).

Desta forma, utilizar métodos de AM para predizer o estágio de prontidão possibilita tornar a avaliação mais rápida e aplicar em larga escala, de forma que beneficiará um número maior de pacientes. Além disso, mais pesquisas focadas em compreender os fatores que influenciam para mudança de comportamento vocal podem contribuir na escolha de estratégias para a terapia de voz, promover maior prontidão e ampliar as chances do sucesso terapêutico.

Assim, esta pesquisa se propõe a responder as seguintes investigações: Qual é a classificação do estágio de prontidão para mudanças comportamentais de pacientes com disfonia? As variáveis da avaliação multidimensional da voz possuem poder de predição do estágio de prontidão para terapia de voz? Quais modelos de aprendizagem de máquina apresentam melhor desempenho na predição dos estágios de prontidão para a terapia de voz?

Solucionar esses questionamentos possibilitará fornecer dados que possam melhorar a prática clínica fonoaudiológica na terapia de voz por meio de intervenção individualizada e personalizada, que possa beneficiar os pacientes com disfonia, pois uma predição assertiva do estágio de prontidão poderá contribuir para o profissional decidir a abordagem terapêutica, além de contribuir para a ampliação do conhecimento científico na área.

2. OBJETIVOS

2.1 Objetivo Principal:

- Avaliar e comparar o desempenho de modelos de aprendizado de máquina na predição do estágio de prontidão para terapia fonoaudiológica de voz.

2.2 Objetivos Específicos:

- Verificar a classificação do estágio de prontidão de pacientes com disfonia;
- Investigar se há associação dos dados da avaliação multidimensional da voz com o estágio de prontidão para terapia de voz de pacientes que buscam por atendimento;
- Identificar as variáveis que possuem poder de predição do estágio de prontidão.

3. FUNDAMENTAÇÃO TEÓRICA

3.1 Avaliação Multidimensional da Voz

A voz é uma ferramenta essencial de comunicação humana, permite a expressão de opiniões e conhecimentos, e contribui para o sucesso profissional e pessoal, por isso profissionais que dependem da voz são diretamente impactados por qualquer alteração vocal. A produção da voz é um ato fisiológico da laringe, envolvendo a vibração das pregas vocais pelo ar dos pulmões, seguida pela ressonância e articulação nas cavidades nasal e oral (Silva, 2021). De acordo com Constantini, Ribeiro e Behlau (2022): “a voz é um fenômeno complexo e multidimensional que é objeto de interesse interdisciplinar”.

Quando ocorre uma dificuldade na produção normal da voz é atribuído o nome de disfonia. Além de prejudicar a comunicação, problemas vocais podem impactar negativamente diversos aspectos da qualidade de vida de uma pessoa. Em casos que a disfonia é causada por comportamentos vocais inadequados, é classificada como a disfonia comportamental. Por outro lado, quando não há envolvimento de comportamentos vocais no seu desenvolvimento, sendo decorrente de lesões nos músculos ou nervos que controlam a voz, é denominada disfonia orgânica (Behlau, 2018).

A equipe multidisciplinar que contribui para o diagnóstico da disfonia inclui uma combinação de profissionais da saúde com os fonoaudiólogos e os otorrinolaringologistas. Enquanto os fonoaudiólogos realizam o julgamento perceptivo-auditivo (JPA) para descrever a alteração vocal, por sua vez os otorrinolaringologistas realizam exame laríngeo para verificar presença ou ausência de fisiopatologia como nódulos ou pólipos nas pregas vocais (Payten *et al.*, 2022).

Na avaliação da voz humana por ser multidimensional, além da perspectiva do clínico, envolve também a do paciente. Deste modo, os instrumentos de autoavaliação da voz abrangem aspectos dimensionais da doença, como incapacidade e desvantagem vocal, e possuem robustez estatística que os tornam úteis tanto na prática clínica quanto na pesquisa (Capucho, 2017).

Desta forma, na autoavaliação são coletadas informações detalhadas sobre suas queixas vocais e o impacto desses problemas em sua vida social, profissional, diária e emocional. Então, a partir dos instrumentos, é adquirida uma visão única e

essencial que não pode ser obtida por meio de outros métodos clínicos tradicionais (Behlau; Madazio; Yamasaki, 2023).

Diversos protocolos de autoavaliação vocal foram adaptados e validados para o português brasileiro, inclui instrumentos como o Questionário de Qualidade de Vida em Voz (QVV), Índice de Desvantagem Vocal (IDV), Questionário de Performance Vocal (QPV), Protocolo de Participação e Atividades Vocais (PPAV), URICA-VV, Questionário de Qualidade de Vida em Voz Pediátrico (QVV-P), Escala de Sintomas Vocais (ESV), e Protocolo de Estratégias de Enfrentamento na Disfonia (PEED-15). Considerando a diversidade cultural e sociodemográfica do país, foram publicados em periódicos e utilizados em pesquisas observacionais e de intervenção para estudar diferentes aspectos vocais em diversas populações e regiões do Brasil (Almeida *et al.*, 2021).

Ademais, o que é percebido como um desvio vocal intenso para um profissional da voz pode ser considerado um desvio de voz leve para quem não depende da voz para trabalhar. Assim, o instrumento de autoavaliação pode capturar uma dimensão diferente da função vocal que o JPA e o exame laríngeo não observam. Recomenda-se que o fonoaudiólogo use todas as informações obtidas com os instrumentos de forma complementar durante a avaliação, o tratamento e a alta dos pacientes, pois são úteis, válidos e confiáveis para diferenciar indivíduos com problemas vocais de pessoas saudáveis (Behlau *et al.*, 2016).

Outrossim, existe uma correlação moderada entre o JPA e os protocolos de autoavaliação, destaca-se a importância de verificar essa associação em vez de inferi-la. Desta forma, os instrumentos de autoavaliação vocal fornecem informações valiosas que as demais avaliações não podem determinar. Sendo recomendado o uso combinado de instrumentos em casos difíceis, mas não devem ser usados isoladamente para diagnósticos (Behlau, 2017).

Quanto ao JPA da voz é um método qualitativo que envolve a avaliação da qualidade vocal de um indivíduo através da escuta atenta e sistemática por parte de um avaliador treinado. Esse processo permite que o profissional identifique características vocais que podem estar associadas a alterações vocais, além de reconhecer aspectos importantes da voz que podem necessitar de terapia de voz. Assim, proporciona uma base para um planejamento terapêutico eficaz (Lee *et al.*, 2021).

Para uma análise mais precisa, são frequentemente utilizados protocolos padronizados, como o GRBAS (Grade, Roughness, Breathiness, Asthenia, Strain), que ajuda a sistematizar a avaliação das qualidades vocais com base em uma escala específica (Bodt; Muyts; Heyning, 1997). E a Escala Analógica Visual (EAV), que consiste em uma linha horizontal de 100 milímetros, onde um dos extremos representa a ausência de desvio vocal e o outro extremo representa o grau máximo do desvio vocal, é realizada a marcação em algum ponto ao longo dessa linha, que indica o grau do desvio vocal (Kempster *et al.*, 2009).

Ainda não há um consenso sobre um programa específico de treinamento para o JPA que possa ser considerado o padrão de referência. Contudo, as estratégias mais comuns incluem o uso de âncoras auditivas, feedback durante o treinamento para informar sobre respostas corretas e erradas, avaliação de parâmetros vocais, principalmente sopro e aspereza, e treinos de até duas horas de duração (Paz *et al.* 2023).

Em relação a análise acústica, é importante para gerar um diagnóstico preciso e monitorar o tratamento oferecido ao paciente, e permite caracterizar o sinal vocal de forma quantitativa ou qualitativa, fornecendo estimativas sobre os aspectos fisiológicos envolvidos na produção vocal (Lopes; Vieira; Behlau, 2022). Do ponto de vista clínico, espera-se que cada medida acústica seja sensível a variações aerodinâmicas e biomecânicas da produção vocal e relacionada aos parâmetros do JPA, além disso que ajude os fonoaudiólogos a entenderem a qualidade vocal do paciente (Paiva *et al.*, 2024).

Quanto a análise acústica quantitativa, existem as medidas lineares, não lineares e cepstrais. As medidas lineares são extraídas com base no modelo de produção vocal de fonte-filtro, diferem das não-lineares que se baseiam em conceitos caóticos de produção vocal, focam nas relações biomecânicas e aerodinâmicas para a extração de características. Já as medidas cepstrais destacam os harmônicos da F0 e sua relação com o ruído da emissão (Florêncio *et al.* 2021).

As medidas acústicas lineares e não lineares variam conforme a intensidade vocal. As medidas lineares mostraram aumentos consistentes na regularidade do sinal com o aumento da intensidade vocal e maior irregularidade com a diminuição da intensidade, enquanto as medidas não lineares apresentaram variações significativas dependendo da intensidade vocal (Florêncio *et al.*, 2021).

Também existe uma correlação entre medidas acústicas lineares e perceptivoauditivas na quantificação do desvio vocal, onde o Glottal to Noise Excitation (GNE) mostrou-se eficaz para diagnosticar vozes saudáveis e alteradas, além de identificar a presença de soprosidade e classificar seu grau na emissão vocal. O shimmer identificou presença ou ausência de rugosidade vocal, e o desvio padrão de F0 diferenciou vozes com e sem tensão, contudo a média de F0 foi útil para classificar a intensidade de tensão fonatória (Lopes; Cavalcante; Costa, 2014).

Em outro estudo sobre a relação do desvio vocal com as medidas acústicas Cepstral Peak Prominence-Smoothed (CPPS) e o declínio espectral, que se tratam de medidas cepstrais, foi constatado que existe associação entre a presença de desvio vocal, o grau geral do desvio vocal (GG), a qualidade vocal predominante e o CPPS. Desta forma, os valores do CPPS foram capazes de distinguir entre vozes rugosas versus soprosas, rugosas versus tensas e soprosas versus tensas. Porém, o declínio espectral foi útil apenas para distinguir vozes soprosas versus tensas (Lopes *et al.*, 2019).

Os exames laríngeos, como a laringoscopia e a estroboscopia, são essenciais para a avaliação detalhada da laringe, pois nenhum diagnóstico definitivo deve ser feito sem a visualização da laringe. A laringoscopia, que pode ser indireta (com o uso de um instrumento óptico) ou direta (com um laringoscópio), permite a visualização direta das pregas vocais e da laringe, sendo útil para diagnósticos (Sulica, 2013).

A estroboscopia, por sua vez, é a técnica que oferece imagens da movimentação das pregas vocais, utiliza-se luz estroboscópica para criar uma ilusão contínua e lenta do movimento da mucosa. Isso permite uma análise detalhada do padrão vibratório das pregas vocais, sendo fundamental para identificar irregularidades como nódulos ou cistos, sendo eficaz na maioria dos casos de disfonia (Sulica, 2013).

Além disso, atualmente também são realizados a videolaringoscopia e videoestroboscopia. A videolaringoscopia utiliza um laringoscópio com câmera de vídeo para transmitir imagens da laringe em tempo real para um monitor, proporciona uma visualização detalhada e mais confortável para o paciente. Já a videoestroboscopia integra uma câmera de vídeo ao estroboscópio, permite a captura contínua e em tempo real das vibrações das pregas vocais. Essa técnica oferece uma visualização mais nítida e detalhada em comparação aos métodos convencionais (Yilmaz *et al.*, 2023).

3.2. Estágio de Prontidão para a terapia de Voz

O Modelo Transteórico (MTT) propõe que a alteração de um comportamento não ocorre por acaso, mas sim através de um processo. Segundo este modelo, diferentes indivíduos se encontram em diferentes estágios de prontidão para a mudança. Durante esse processo, as pessoas atravessam cinco estágios distintos: pré-contemplação, contemplação, preparação, ação e manutenção (Prochaska; DiClemente, 1982).

O MTT classifica os indivíduos nos estágios de prontidão, mas reconhece a complexidade e a natureza não linear da mudança de comportamento, pois os indivíduos podem avançar e retroceder entre os estágios de mudança, reflete a realidade comum de recaídas e dificuldades na manutenção de novos comportamentos (Prochaska *et al.*, 1993).

No estágio de pré-contemplação, os indivíduos não estão considerando ativamente a necessidade de mudança e podem não estar plenamente conscientes dos impactos negativos de seus comportamentos. Na fase seguinte, contemplação, há um reconhecimento dos problemas associados ao comportamento e uma ponderação sobre a possibilidade de mudança, embora ainda não tenham sido tomadas ações concretas para isso (Prochaska; Norcross; DiClemente, 2013).

À medida que avançam para o estágio de preparação, os indivíduos estão preparados para agir em um futuro próximo e começam a implementar medidas pequenas em direção à mudança. No estágio de ação, ocorre a implementação ativa de comportamentos novos e saudáveis, através de um compromisso ativo com a mudança e a adoção de estratégias para modificar o comportamento. Finalmente, o estágio de manutenção visa consolidar as mudanças alcançadas durante a ação, com foco na prevenção de recaídas e na sustentação do novo comportamento a longo prazo (Prochaska; Norcross; DiClemente, 2013).

Deste modo, os dois primeiros estágios são caracterizados por ambivalência quanto à mudança, enquanto os três últimos envolvem um comprometimento maior com a mudança. Sugere-se que as intervenções comportamentais devem ser adaptadas ao estágio do paciente para aumentar as chances de mudança comportamental e a eficácia das intervenções (Sutton *et al.*, 2003).

Na saúde, diversas pesquisas foram realizadas utilizando os estágios de prontidão, o que inclui pesquisas para verificar o estágio de prontidão para o manejo de doenças crônicas (Vásquez; Garrido, 2022), de sobrepeso e obesidade (Silva;

Oliveira; Velasquez, 2020; Soares, 2022), de alteração auditivas (Laplante-Lévesque *et al.*, 2015), da diabetes (Silva; Luz, 2022), do tabagismo (Barbosa *et al.*, 2020), da ansiedade (Dozois *et al.* 2004) e do consumo de bebidas alcoólicas (Soares; Vargas, 2020; Lima *et al.*, 2022) .

Ademais, o estágio de prontidão pode ser utilizado para avaliar se determinada medida adotada contribui para mudança de hábitos dos paciente, em um estudo foi verificada a efetividade de uma tecnologia educativa que contribuiu para avanço do estágio de prontidão para mudança comportamental em adultos hipertensão e excesso ponderal (Santiago *et al.*, 2021), enquanto em outro estudo com em indivíduos com obesidade verificou-se que a intervenção realizada foi eficaz na evolução dos estágios de prontidão para a mudança de comportamento alimentar e de atividade física (Luz *et al.*, 2021).

Além das várias áreas da saúde já citadas onde os estágios de prontidão são utilizados, também é relevante na terapia de voz. Com base no MTT, foi desenvolvida a escala University of Rhode Island Change Assessment (URICA), e adaptada para a área da voz (Teixeira *et al.*, 2013; Aguiar *et al.*, 2021), para verificar o estágio de prontidão na terapia de voz, intitulada de URICA-Voice Validated (URICA-VV). Desde então, o instrumento passou a ser bastante empregado na literatura em pessoas que buscam a terapia de voz.

A URICA, desenvolvida por McConnaughy, Prochaska e Velicer em 1983, é um questionário de autopreenchimento para medir estágios motivacionais de mudança comportamental (pré-contemplação, contemplação, ação e manutenção) em adultos, inicialmente não especificando o tipo de comportamento-problema. Embora concebida para medir cinco estágios, a preparação foi excluída após avaliação. A versão original possui 32 perguntas, com 8 itens para cada estágio, enquanto a versão reduzida contém 24 perguntas, com 6 itens para cada estágio.

O instrumento elaborado na pesquisa de Teixeira e colaboradores (2013) possui 32 itens e foi baseado na Teoria Clássica do Teste (TCT) que trata de um modelo estatístico psicométrico onde o nível do atributo é determinado pelo somatório das respostas dos itens de um teste. Os itens são agrupados em quatro conjuntos de oito afirmações, cada um correspondendo aos estágios de mudança: Pré-contemplação, Contemplação, Ação e Manutenção.

O estudo de adaptação brasileira da escala URICA para a área de voz, foi transversal e contou com uma amostra com 66 pacientes com disfonia, dos quais 57,6

estavam no estágio de contemplação, 30,3% no de pré-contemplação e 12,1% no de ação. Devido a maioria dos pacientes com disfonia em tratamento está no estágio de contemplação, pode limitar os resultados da terapia, e não houve relação entre as variáveis idade, tipo de disfonia, escolaridade e número de sessões de fonoterapia e os estágios de prontidão (Teixeira *et al.*, 2013).

Posteriormente, foi realizado outro estudo que utilizou a análise fatorial confirmatória para verificar a acurácia dos componentes presentes na escala URICA-VV, além de propor a aplicação da Teoria de Resposta ao Item (TRI) por ser considerada uma análise estatística mais robusta que a TCT e permite diferentes pesos para diferentes itens, reconhecendo que alguns itens podem ser mais informativos do que outros sobre a habilidade do indivíduo (Aguiar *et al.*, 2021).

Após a análise fatorial e a aplicação da TRI, a URICA-VV passou de 32 para 25 questões, além de possuir dois domínios: contemplação e manutenção. Isso ocorreu porque apenas 3 dos 8 itens do estágio de pré-contemplação e 3 dos 8 itens do estágio de ação apresentaram uma correlação significativa dentro de seus respectivos domínios. Enquanto as cargas fatoriais associadas aos estágios de contemplação e manutenção mostraram uma correlação adequada com seus fatores correspondentes (Aguiar *et al.*, 2021; Aguiar *et al.*, 2023).

A URICA-VV, em sua versão validada pela TRI, apresentou relação com outras variáveis da avaliação multidimensional da voz, foi observada validade discriminante entre o escore total e os sintomas vocais auditivos, assim como entre o estágio de contemplação e os sintomas vocais auditivos, sensoriais e totais. Também mostrou validade discriminante do escore total e contemplação com fatores de risco organizacionais e totais, e do escore manutenção com fatores de risco organizacionais. No JPA, apresentou validade discriminante com o grau geral (Silva, 2022).

Em relação ao tipo de disfonia, verificou-se que os indivíduos com disfonia organofuncional apresentam maior adesão para a terapia de voz. Além disso, foi observado que quanto maior o impacto emocional da disfonia, como sentimentos de vergonha, baixa autoestima, preocupação e insatisfação, melhor é a adesão do paciente a terapia de voz. Deste modo, verificou-se a importância da escala URICA-VV no início e ao longo do tratamento para desenvolver estratégias motivacionais adaptadas ao estágio de prontidão de cada paciente, o que pode ter um impacto positivo significativo no progresso terapêutico (Góes *et al.*, 2016).

Um estudo investigou a eficácia da terapia de grupo e utilizou a escala URICA-VV para avaliar o estágio de prontidão de 49 pacientes com disfonia comportamental. A maioria dos participantes, principalmente mulheres com lesão membranosa da prega vocal e não profissionais da voz, estava no estágio de Contemplação tanto antes quanto depois da terapia. Entretanto, houve mudanças significativas nos estágios de prontidão, a maioria dos pacientes apresentou redução e poucos avançaram (Costa *et al.*, 2017).

Quanto aos pacientes com disfonia orgânica, em um estudo com indivíduos com doença de Parkinson (DP), foi analisado se assistir a vídeos curtos sobre sintomas de voz e tratamento na doença de Parkinson influência no estágio de prontidão para mudança. Foi descoberto que a prontidão para mudança aumentou significativamente após assistir aos vídeos, sugerindo que assistir a vídeos ajudou esses indivíduos a se aproximarem da melhoria ativa de sua voz e na mudança de comportamento vocal. Isso ressalta a importância de explorar e integrar tecnologias de saúde sem fio em futuras abordagens terapêuticas, para melhorar os resultados de tratamento e a qualidade de vida desses pacientes. Além de medir o estágio de prontidão, pode ser útil para verificar a eficácia das abordagens utilizadas nos pacientes (Kopf; Graetzer; Hun, 2015).

3.3 Aprendizado de Máquina (AM)

A aprendizagem de máquina trata-se de uma área da Inteligência Artificial (IA) que tem por objetivo desenvolver técnicas computacionais sobre o aprendizado, além de construir sistemas capazes de obter conhecimento de maneira automática. Com os sistemas de AM é possível tomar decisões com base em experiências anteriores (Rezende, 2003).

Os avanços recentes em estatística têm focado no desenvolvimento de técnicas automatizadas mais poderosas para modelagem preditiva, como regressão e classificação, dentro do aprendizado de máquina estatístico. Diferentemente dos métodos estatísticos clássicos, essas abordagens são orientadas por dados, evitando impor estruturas predeterminadas aos dados. Um exemplo simples é o método dos K-Vizinhos Mais Próximos, que classifica uma observação baseado na classificação de observações similares (Bruce; Bruce, 2019).

Existem três tipos principais de AM: Supervisionado, Não Supervisionado e por Reforço. No Aprendizado Supervisionado, o algoritmo aprende a partir de exemplos

rotulados, em que cada exemplo é descrito por um vetor de valores (atributos) e um rótulo associado, que pode ser qualitativo ou quantitativo. O objetivo é construir um modelo capaz de prever corretamente o rótulo de novos exemplos não rotulados (Ludermir, 2021).

Esse método é amplamente utilizado e inclui problemas de classificação e de regressão. Problemas de classificação têm saídas discretas, como a nacionalidade de um usuário, enquanto problemas de regressão têm saídas contínuas, como a previsão do preço de uma casa baseada em atributos como área e número de quartos. Alguns dos algoritmos mais comuns de aprendizado supervisionado incluem Regressão Logística, K-Vizinhos Mais Próximos, Naive Bayes, Árvore de Decisão e Floresta Aleatória (Alvim, 2019). Esses modelos foram abordados neste trabalho.

Os dados relevantes são coletados e preparados, abrange históricos, características individuais, variáveis independentes e o desfecho a ser previsto. Com base nesses dados, escolhe-se um modelo adequado de aprendizado de máquina, os dados são divididos em conjuntos de treinamento e teste. No conjunto de treinamento, o modelo aprende padrões que relacionam variáveis independentes ao desfecho. Após o treinamento, com o conjunto de teste é avaliada a capacidade de predição do modelo em novos dados. Ademais, ajustes e otimizações são realizados conforme necessário para melhorar a precisão do modelo (Alloghani *et al.*, 2020).

No Aprendizado Não Supervisionado, os exemplos são fornecidos sem rótulos, e o algoritmo busca explorar a estrutura subjacente dos dados, o que pode envolver agrupar observações semelhantes, reduzir a dimensionalidade, estimar distribuições de probabilidade ou identificar padrões de associação (Ludermir, 2021). Entre os algoritmos mais comuns destacam-se os de agrupamento, incluindo K-Means e agrupamento hierárquico; os de visualização e redução de dimensionalidade, como a análise de componentes principais; e os de regras de associação, como Apriori e Eclat (Alvim, 2019).

Já no Aprendizado por Reforço, o algoritmo recebe um sinal de reforço, como recompensa ou punição, em vez da resposta correta. O algoritmo então ajusta suas hipóteses com base nesse feedback, sendo bastante utilizado em jogos e robótica. A escolha dos algoritmos apropriados para o problema é essencial, bem como a definição de seus parâmetros. Após o treinamento, é necessário avaliar a precisão do modelo e garantir que ele continue a funcionar bem com dados novos e em mudança (Ludermir, 2021).

Para aplicar o Aprendizado de Máquina de maneira eficaz, é necessário ter um bom conjunto de dados, que frequentemente precisa ser atualizado e aprimorado. Dados desequilibrados limitam o desempenho e a precisão dos classificadores. Existem abordagens amplas para resolver problemas de desequilíbrio na classificação de dados, como pré-processamento de métodos, abordagens centradas em algoritmos e abordagens híbridas (Kaur, 2020).

Além disso, as técnicas de modelagem enfrentam o desafio do sobreajuste. Por exemplo, ao incorporar muitas variáveis em uma equação de regressão, o modelo pode realizar previsões imprecisas. Contudo, em muitas técnicas estatísticas, é possível evitar esse sobreajuste ao selecionar criteriosamente as variáveis preditoras que são usadas para prever ou explicar o valor de uma variável de interesse, assim garante que o modelo seja capaz fazer previsões mais precisas em novos dados não utilizados durante o treinamento (Bruce; Bruce, 2019).

O aprendizado de máquina na área da saúde já demonstrou progressos significativos, ao melhorar a capacidade em fornecer tratamentos mais eficazes com maior qualidade e precisão. Os Registros Eletrônicos de Saúde aprimoraram a organização de dados e contribuem para o uso de previsão, permitindo decisões clínicas mais informadas. Da mesma forma, em exames de imagens, o aprendizado de máquina demonstrou grandes vantagens em precisão e diagnóstico (Habeheh; Gohel, 2021).

Ademais, a aprendizagem profunda na saúde facilita a análise de dados biomédicos complexos e heterogêneos, com modelos mais precisos e eficientes para lidar com grandes volumes de dados não estruturados, embora a interpretabilidade dos resultados ainda seja um desafio. Assim, é essencial desenvolver arquiteturas que não só garantam precisão, mas também sejam compreensíveis e aplicáveis aos profissionais de saúde, com o objetivo de melhorar a prática clínica e os resultados para os pacientes (Miotto *et al.*, 2018).

3.3.1 Regressão Logística

A Regressão Logística é um método utilizado para prever a probabilidade de uma classe binária com base em variáveis independentes (Cox, 1958). Ela utiliza uma função logística para modelar a relação entre as variáveis explicativas e a resposta categórica, que serve como a base matemática que transforma os valores previstos para que se ajustem ao intervalo $[0, 1]$. Para transformar probabilidades em uma

decisão binária, define-se um limiar (geralmente 50%), acima do qual o evento é previsto como 1 e abaixo como 0. Essa transformação garante que o modelo seja adequado para problemas de classificação binária (Lantz, 2023).

No entanto, para problemas de classificação, a regressão logística calcula a probabilidade de a resposta assumir o valor 1, dado o vetor de variáveis independentes. Para isso, aplica-se a função logística à equação linear, o que garante que a saída do modelo fique restrita a esse intervalo, tornando-a apropriada para representar uma probabilidade (Molnar, 2022). A equação de regressão logística é:

$$P(Y^{(i)} = 1) = \text{logistic}(x^{(i)T}\beta) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_p x_p^{(i)}))},$$

Onde:

- $P(Y^{(i)} = 1)$ é probabilidade de que a variável dependente Y assumo o valor 1 na i -ésima observação;
- $\text{logistic}(x^{(i)T}\beta)$ é a função logística que transforma a combinação linear dos preditores para um valor de probabilidade entre 0 e 1;
- \exp é a função exponencial do número de Euler (aproximadamente 2.7183);
- β_0 é o intercepto do modelo, representa o valor predito quando todas as variáveis independentes são iguais a zero;
- $\beta_1 \dots \beta_p$ são coeficientes de inclinação, representam a influência de cada variável independente no valor predito;
- $x_1^{(i)} \dots x_p^{(i)}$ são valores das variáveis independentes para a i -ésima observação.

Seu ponto forte reside na interpretabilidade dos coeficientes, que indicam o impacto de cada variável sobre o resultado. Na regressão logística, esses coeficientes são interpretados de forma multiplicativa em relação às chances (*odds*) de ocorrência de um evento, que mostram como um aumento unitário em uma variável preditora altera as chances do evento ocorrer (Molnar, 2022).

Desta forma, é bastante eficaz em problemas em que os dados seguem uma relação linear aproximadamente, mas pode apresentar limitações em situações de alta multicolinearidade ou em dados não lineares. Apesar disso, surge como ferramenta prática para resolver problemas de classificação binária, sendo assim especialmente útil em situações com conjuntos de dados limitados (Conway; White, 2012).

3.3.2 K-Nearest Neighbors (KNN)

O algoritmo K-Nearest Neighbors (KNN) é um método de classificação em que a categoria de um dado desconhecido é determinada com base nas categorias dos K eventos mais próximos no espaço dos atributos (Fix; Hodges, 1951). Esses vizinhos mais próximos são identificados utilizando a distância euclidiana, uma métrica que calcula a diferença geométrica entre os vetores dos atributos do dado não rotulado e os dados conhecidos. A classificação final é atribuída à classe mais frequente entre os vizinhos identificados. A escolha do valor de K é crucial, pois valores muito baixos tornam o modelo sensível a outliers, ou seja, dados atípicos que podem distorcer os resultados. Por outro lado, valores muito altos podem aumentar o tempo de processamento e diluir a influência de vizinhos relevantes (Lopes, 2020). Observa-se a forma da distância euclidiana a seguir:

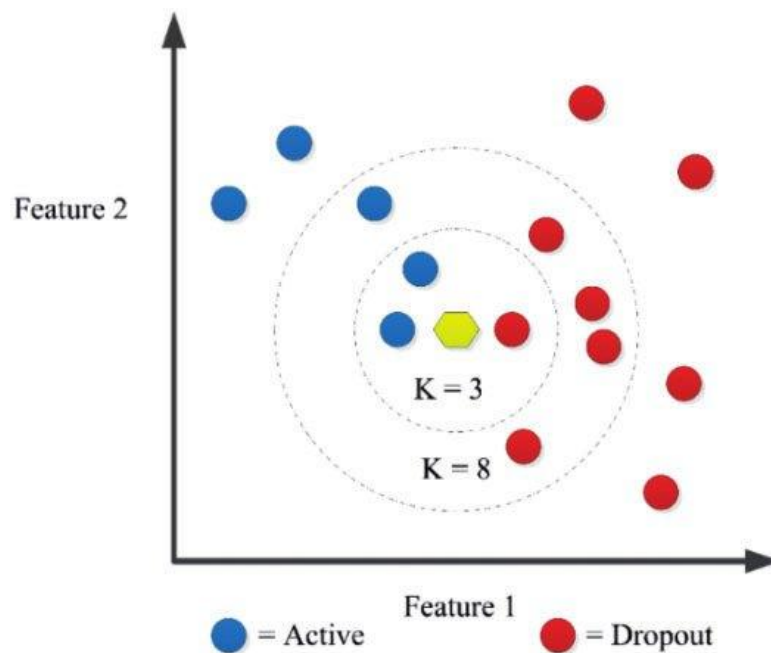
$$d = \sqrt{\sum_{k=1}^n (p_{ik} - p_{jk})^2},$$

Onde:

- d é a distância euclidiana entre as observações i e j ;
- n é o número total de variáveis utilizadas para calcular a distância;
- k é o índice que representa cada variável, variando de 1 até n ;
- p_{ik} é o valor da k -ésima variável para a observação i ;
- p_{jk} é o valor da k -ésima variável para a observação j ;
- $\sum_{k=1}^n (p_{ik} - p_{jk})^2$ é a soma dos quadrados das diferenças entre os valores correspondentes das duas observações em cada uma das n variáveis.

Depois de selecionados os K menores distâncias, verifica-se a classe de cada um desses dados mais próximos, contando a quantidade de vezes que cada classe aparece. Por fim, o algoritmo atribui ao novo dado a classe que mais se repetiu entre esses vizinhos mais próximos, realizando assim a classificação. A Figura 1 a seguir ilustra o funcionamento do KNN.

Figura 1 - Funcionamento do algoritmo KNN.



Fonte: Xing; Du, 2018.

Na Figura 1, o ponto amarelo representa uma nova observação que ainda não foi classificada, enquanto os pontos já classificados estão divididos em duas categorias: os azuis representam a classe *Active* e os vermelhos representam a classe *Dropout*. Para classificar esse novo ponto, o algoritmo calcula a distância dele em relação a todos os demais pontos do conjunto. Em seguida, seleciona os k vizinhos mais próximos — por exemplo, $k = 3$ ou $k = 8$, como mostrado nos dois círculos. No caso de $k = 3$, o ponto amarelo possui 2 vizinhos azuis e 1 vizinho vermelho dentro do círculo, portanto, será classificado como *Active*, já que essa é a classe predominante entre os mais próximos. Por outro lado, se $k = 8$, há uma maioria de pontos vermelhos (5 vermelhos contra 3 azuis), e o ponto seria classificado como *Dropout*.

Ademais, neste algoritmo são utilizadas métricas de similaridade para tarefas como recomendar itens a usuários de websites, sendo intuitivo e fácil de compreender, com seu conceito básico inspirado na forma como as pessoas fazem recomendações. A abordagem mais simples, o 1-nearest neighbor, sugere itens baseando-se no mais próximo dos que um usuário já gosta. O KNN é versátil e considerado bastante útil em problemas de classificação, como separar pontos em duas categorias (Conway; White, 2012).

3.3.3 Naive Bayes

O algoritmo Naive Bayes baseia-se no teorema de Bayes, trata-se de um método simples e eficaz para classificação, muito popular especialmente para classificação de texto, como filtragem de spam, devido à sua eficiência computacional e robustez, mesmo com conjuntos de dados grandes. Apesar de sua força e rapidez, o algoritmo faz a suposição “ingênua” de que todas as características (features) são independentes e igualmente importantes, o que raramente ocorre na prática. Mesmo assim, o Naive Bayes costuma apresentar bom desempenho mesmo quando essa suposição é violada (Lantz, 2023). Sejam A e B eventos pertencentes ao espaço amostral de um experimento e P a função de probabilidade, observe a equação do Teorema de Bayes abaixo:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$

Onde:

- $P(A|B)$ é a probabilidade de A dado B ;
- $P(B|A)$ é a probabilidade de B dado A ;
- $P(A)$ é a probabilidade de A ocorrer;
- $P(B)$ é a probabilidade de B ocorrer.

Inicialmente é realizada a coleta das probabilidades condicionais de cada atributo dado uma determinada classe, são organizadas em uma tabela de verossimilhança. Em seguida, calcula-se a probabilidade conjunta de observar aquele conjunto específico de atributos em uma nova instância. Entretanto, esse cálculo pode ser computacionalmente complexo, pois exige considerar todas as possíveis combinações de atributos. Para simplificar esse processo, assume-se a independência condicional entre os atributos, o que permite multiplicar as probabilidades condicionais individuais. Com isso, estima-se a probabilidade de cada classe dada a instância observada, e a classe com maior probabilidade é atribuída como resultado da classificação (Mitchell, 1997).

Esse processo torna viável a aplicação do algoritmo mesmo em cenários com muitas características, ao reduzir drasticamente o custo computacional. As vantagens incluem simplicidade, velocidade, boa performance com dados ruidosos e poucos exemplos de treinamento, além da facilidade de obtenção das probabilidades de predição. Entre as limitações, estão a dependência da suposição de independência

das variáveis e a menor confiabilidade das probabilidades estimadas em comparação com as classes previstas.

3.3.4 Árvore de Decisão

O conceito de Árvores de Decisão tem origem no trabalho de Leo Breiman e colegas, formalizado no CART (Classification and Regression Trees) em 1984 (Breiman et al., 1984). É um modelo preditivo inspirado na forma como os humanos tomam decisões. Uma de suas principais vantagens é apresentar a informação visualmente, de uma forma fácil de entender e interpretar. As árvores podem ser usadas para problemas de Classificação ou Regressão.

Uma árvore de decisão usa amostras das características dos dados para criar regras de decisão no formato de árvore, isto é, reduz os dados em um conjunto de regras que podem ser usadas para uma decisão. Elas possibilitam a seleção automática de variáveis para compor suas estruturas: cada nó interno representa uma decisão sobre um atributo que determina como os dados estão particionados pelos seus nós filhos. Para classificar um novo exemplo, basta testar os valores dos atributos na árvore e percorrê-la até se atingir um nó folha (classe predita) (Escovedo; Koshiyama, 2021).

Na interpretação, a previsão de um modelo de árvore é feita rastreando o caminho desde o nó raiz até um nó terminal, com decisões baseadas nos limites definidos pelos splits (pontos de divisão). A importância das variáveis é medida pela contribuição na redução da variância ou do índice de Gini ao longo da árvore, sendo este um critério de impureza que quantifica a probabilidade de um elemento, escolhido aleatoriamente de um nó, ser classificado incorretamente. Além disso, é possível decompor o caminho de decisão para analisar a contribuição individual de cada variável na previsão final, mesmo que algumas variáveis sejam usadas mais de uma vez ou não sejam usadas. Assim, para obter a previsão final, percorre-se o caminho da instância na árvore, somando as contribuições de cada split no trajeto, conforme a fórmula (Molnar, 2022):

$$\hat{f}(x) = \bar{y} + \sum_{d=1}^D \text{split.contrib}(d, x) = \bar{y} + \sum_{j=1}^P \text{feat.contrib}(j, x),$$

Onde:

- $\hat{f}(x)$ é o valor predito pelo modelo para a entrada x ;
- \bar{y} é a média da variável resposta no conjunto de dados de treinamento;
- D é o número total de divisões (splits) nas árvores do modelo;

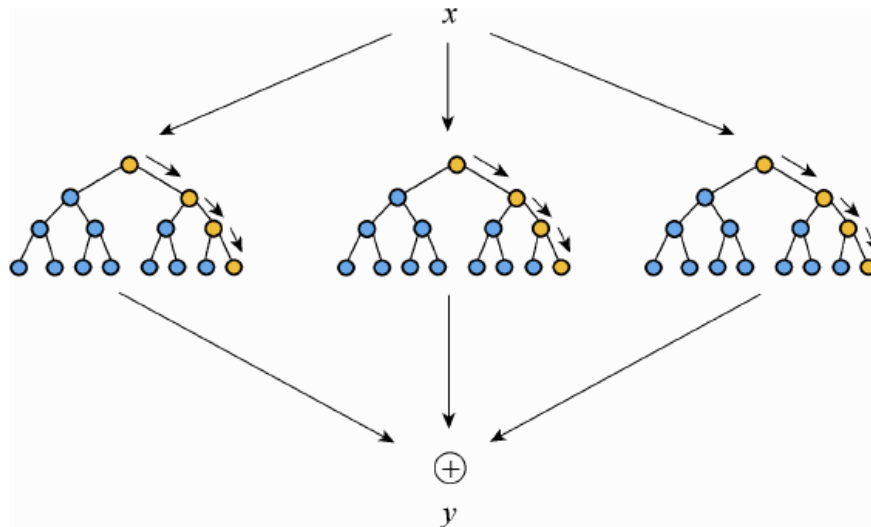
- $\text{split.contrib}(d, x)$ é a contribuição da d -ésima divisão (split) para a predição do ponto x ;
- P é o número total de variáveis (features) usadas no modelo.
- $\text{feat.contrib}(j, x)$ é a contribuição da j -ésima variável (feature) para a predição do ponto x .

Entre as vantagens, as árvores de decisão são intuitivas, facilmente visualizadas e permitem explicações baseadas em cenários contrafactuais. São úteis para capturar relações não lineares e interações entre variáveis. Contudo, podem ser instáveis devido à dependência do conjunto de treino e árvores muito profundas tornam-se difíceis de interpretar (Molnar, 2022).

3.3.5 Floresta Aleatória

O algoritmo Floresta Aleatória é uma extensão das árvores de decisão, baseada em ensembles, uma técnica que combina os resultados de vários modelos, o que ajuda a criar resultados mais precisos para chegar à sua conclusão (Breiman, 2001). É combinado o poder de múltiplas árvores de decisão, treinadas com amostras aleatórias do conjunto de dados e subconjuntos de variáveis, para melhorar a precisão e reduzir o risco de overfitting, que ocorre quando o modelo se ajusta tão bem aos dados de treinamento que perde a capacidade de generalizar para novos dados. Adicionalmente, essa técnica é eficaz em dados com alta dimensionalidade e fornece uma estimativa de importância das variáveis, sendo amplamente utilizada em problemas de classificação e regressão.

A partir do conjunto de dados original, são construídas múltiplas árvores de decisão, cada uma usando uma amostra aleatória dos dados (com reposição) e, a cada nó de decisão, considerando apenas um subconjunto aleatório das variáveis disponíveis para definir o melhor corte. Essa combinação de amostragem e seleção aleatória de variáveis promove diversidade entre as árvores. Para fazer uma predição, cada árvore “vota” em uma classe, e o modelo final escolhe a classe que receber mais votos, o que melhora a capacidade de generalização. A Floresta Aleatória é popular por sua versatilidade, robustez e boa performance, mesmo em conjuntos de dados grandes e com muitas variáveis, em que outros métodos podem falhar (Lanz, 2023).

Figura 2 - Floresta Aleatória.

Fonte: Deng *et al.*, 2019.

Na Figura 2, observamos três árvores de decisão individuais que, juntas, formam uma Floresta Aleatória. Isso aumenta a precisão do modelo, pois analisa os resultados de muitas árvores de decisão diferentes e encontra uma média (Deng *et al.*, 2019).

4. MÉTODOS

4.1. Delineamento do Estudo

Esta pesquisa adota uma abordagem quantitativa, de natureza observacional e transversal, com delineamento transversal.

4.2 População e Amostra

A população foi composta por pacientes, de ambos os sexos, que buscam atendimento no Laboratório Integrado de Estudos da Voz (LIEV), da Universidade Federal da Paraíba (UFPB), Campus I, João Pessoa - PB. Para participar do estudo, o indivíduo deveria estar de acordo com os seguintes critérios de elegibilidade:

- Possuir idade acima de 18 anos;
- Não ter realizado terapia de voz anteriormente;
- Apresentar queixa vocal por meio do autorrelato;
- Ter respondido a todos os itens da URICA-VV;
- Ausência de limitação neurológica, cognitiva e genética que pudesse afetar a comunicação e as respostas aos instrumentos.

Assim, a amostra foi por conveniência, por isso a seleção dos sujeitos foi baseada em quem estava disposto a participar no momento da coleta de dados e que estivesse dentro dos critérios pré-estabelecidos para essa pesquisa.

4.3 Instrumentos de Coleta de Dados

Foram utilizados instrumentos para obter as informações necessárias para atingir os objetivos do estudo. Por isso, os instrumentos escolhidos foram: o Protocolo de Anamnese e Avaliação Vocal (PAAV) e a Escala URICA-VV, além da coleta de amostras vocais.

4.3.1 Protocolo de Anamnese e Avaliação Vocal (PAAV)

Foi utilizado o Protocolo de Anamnese e Avaliação Vocal (PAAV) (Anexo 1), desenvolvido no Laboratório Integrado de Estudos da Voz (LIEV) do Departamento de Fonoaudiologia da UFPB (Almeida et al, 2021). Este protocolo, elaborado por especialistas na área de voz, constitui um instrumento de avaliação que aborda tanto dados pessoais quanto vocais. Ele foi adotado para as atividades de pesquisa e

assistência vinculadas ao laboratório, conforme descrito por Almeida *et al.* (2015), Almeida *et al.* (2021).

O PAAV é dividido em sete partes: identificação pessoal; queixa e duração; tratamentos prévios para disfonia; investigação complementar; antecedentes familiares; motivação para terapia; e outras observações. No entanto, nesta pesquisa, foram utilizadas apenas as três primeiras partes. A exclusão das demais seções do protocolo foi motivada pelo foco clínico do instrumento, não sendo essencial para alcançar os objetivos deste estudo.

A primeira parte do PAAV visa coletar informações pessoais do paciente, inclui nome, idade, data de nascimento, sexo, estado civil, nível de instrução, endereço e contato. Também são registradas informações relacionadas à ocupação, como profissão, período de trabalho e carga horária. Na segunda parte são coletados dados sobre a queixa e duração, divide-se em quatro tópicos: motivo da consulta/duração, história pregressa da disfonia, sintomas vocais e fatores de risco. A terceira parte diz respeito à realização anterior de tratamento para a disfonia.

Ao abordar os sintomas vocais, o PAAV lista 24, distribuídos de forma equitativa entre auditivos e sensoriais. Os pacientes indicam se estão ou não presentes esses sintomas, e então é feita uma soma simples para determinar o total de sintomas relatados. Quanto maior for esse total, maior será a sintomatologia de disfonia (Vital *et al.*, 2016).

Os sintomas vocais auditivos incluem rouquidão, voz muda, projeção da voz, mudança da voz ao longo do dia, voz monótona, dificuldade com notas agudas, falar baixo, ar na voz, instabilidade vocal, dificuldade com notas graves, falhas na voz e perda constante da voz. Os sintomas vocais sensoriais incluem fadiga vocal, sensação de bolo na garganta, tensão, presença de muco, desconforto, garganta seca, pigarro, gosto ácido, esforço ao falar, dor na garganta, tosse e odinofagia (dor ao engolir) (Almeida *et al.*, 2021).

Enquanto os fatores de risco vocais são subdivididos em três sendo 5 fatores de risco organizacionais, 9 fatores ambientais e 21 fatores de risco pessoais. Para obtenção do resultado é realizado um somatório simples, tanto para todos os fatores de risco em conjunto quanto separadamente para os fatores organizacionais, ambientais e pessoais. Quanto mais elevado o número de fatores de risco identificados, maior é a possibilidade de a disfonia estar relacionada ao comportamento vocal (Almeida *et al.*, 2015).

Os fatores de risco são categorizados em organizacionais, ambientais e pessoais. Os fatores de risco organizacionais incluem a jornada de trabalho, o número de ouvintes, o acúmulo de tarefas, o tempo de uso da voz e a demanda vocal. Os fatores de risco ambientais incluem ruído, umidade, ergonomia, acústica, poluição, estresse, distância, poeira/mofo e equipamentos inadequados (Almeida *et al.*, 2021).

Os fatores de risco pessoais incluem fumo, falar muito, uso de telefone, falar em ambientes ruidosos, gritar, vida social intensa, automedicação, consumo de álcool, falar alto, falar com esforço, falar em público, torcer, tossir, falta de repouso, uso de drogas, falar rápido, falar em pitch (altura) inadequado, imitar vozes, cantar, hidratação inadequada e alimentação inadequada (Almeida *et al.*, 2021).

Desta forma, esse protocolo foi escolhido para a pesquisa por distinguir os sintomas vocais entre auditivos e sensoriais, além de distinguir os fatores de risco vocais entre organizacionais, ambientais e pessoais, que permite uma investigação separada e a observação de possíveis diferenças no estágio de prontidão para a terapia de voz.

4.3.2 URICA-VV

A Escala URICA-VV (Anexo 2) é um instrumento de autoavaliação que avalia o estágio de prontidão do paciente em mudar seu comportamento vocal a partir da terapia de voz. Foi traduzida e adaptada para área de voz (Teixeira *et al.*, 2013) e validada para o português brasileiro a partir da Teoria de Resposta ao Item (TRI); envolve 25 itens divididos em dois estágios de prontidão: contemplação, com 12 itens, e manutenção, com 13 itens (Aguilar *et al.*, 2021).

Neste instrumento, os itens que fazem parte do estágio de “Contemplação” são: 1, 2, 3, 6, 9, 10, 11, 13, 15, 17, 19 e 20, sendo o item 15 o mais representativo; e os itens que fazem parte do estágio de “Manutenção” são 4, 5, 7, 8, 12, 14, 16, 18, 21, 22, 23, 24 e 25, onde o item 14 é o mais representativo. A chave de resposta é a partir de uma escala Likert de 5 pontos para a marcação das respostas, oferecendo uma medida gradativa da disposição do indivíduo em relação à motivação para as mudanças vocais necessárias.

No estágio de contemplação, o indivíduo reconhece o problema vocal e considera mudanças, mas ainda não tomou iniciativas. Já no estágio de manutenção, ele está em ação e busca manter comportamentos saudáveis para a voz, com a finalidade de avaliar a continuidade das ações adotadas (Prochaska; Norcross;

DiClemente, 2013). O cálculo e interpretação dos resultados pode ser realizado a partir de algoritmo desenvolvido por meio da TRI, realizada pelo software SelfVox, desenvolvido pelo LIEV, com um ponto de corte estabelecido em -0,236, esse valor é obtido pelo teta do instrumento. Valores abaixo desse ponto indicam estágio de contemplação, enquanto valores acima sugerem que o indivíduo está no estágio de manutenção em relação a terapia de voz.

4.4 Fonte de Dados

Os dados utilizados foram obtidos da avaliação fonoaudiológica e da avaliação otorrinolaringológica dos pacientes que procuram o LIEV por problemas de voz. Antes de iniciar a coleta de dados, todos os pacientes atendidos assinaram um Termo de Autorização de Uso de Imagem e Dados, garantindo que as informações coletadas durante o atendimento possam ser utilizadas para fins científicos e acadêmicos, assegurando privacidade e confidencialidade.

Em seguida, cada paciente foi orientado a preencher o PAAV e a escala URICA-VV. Ademais, responderam também instrumentos de autoavaliação baseados na Teoria da Resposta ao Item (TRI), incluindo o Índice de Desvantagem Vocal (IDV-10-TRI), a Escala de Qualidade de Vida em Voz (QVV-TRI), a Escala de Sintomas Vocais (ESV-TRI) e a Escala de Desconforto do Trato Vocal (EDTV-TRI). O IDV-10-TRI investiga o impacto da voz na qualidade de vida, considerando dimensões funcionais, físicas e emocionais. A QVV-TRI avalia o quanto a voz interfere nas esferas emocional, física e social do indivíduo. A ESV-TRI identifica sintomas relacionados ao uso da voz autorreferidos, enquanto a EDTV-TRI examina desconfortos e sensações físicas no trato vocal, considerando tanto a frequência quanto a intensidade desses sintomas. Para fins de análise, as respostas foram agrupadas em duas categorias: “sim”, indicando presença de disfonia, e “não”, indicando sua ausência.

Após o preenchimento dos instrumentos de autoavaliação, os indivíduos foram encaminhados para a coleta das amostras vocais, onde foram posicionados dentro de uma cabine acústica e ficaram a uma distância de 10 centímetros de um microfone profissional dinâmico, montado em um tripé. Durante o procedimento, foram solicitados a emitir a vogal [ε] de forma sustentada e realizar contagem de 0 a 10, tanto em seu volume habitual quanto pelo tempo máximo que conseguem. As gravações das amostras foram realizadas em um ambiente com ruído inferior a 50 dB

de pressão sonora e através do software Fono View 4, com uma taxa de amostragem de 44100 Hz.

As vozes coletadas foram avaliadas por juízes fonoaudiólogos especialistas em voz, com experiência no JPA da voz. Para essa pesquisa foi utilizada Escala de Desvio Vocal (EDV), com foco no parâmetro grau geral, onde foram classificadas as vozes com pontuação até 35,5 como dentro da variabilidade normal da qualidade vocal, enquanto pontuações a partir de 35,6 indicaram qualidade vocal desviada (Yamasaki et al, 2017).

Para a análise acústica, foi utilizado o software gratuito VoxMore para extração das medidas, foram as medidas da frequência fundamental (F0), incluindo média, desvio padrão (DP), primeiro quartil (Q1), mediana, terceiro quartil (Q3), mínimo, máximo e coeficiente de variação (CV); medidas do período, como média, desvio padrão e logaritmo natural do desvio padrão (lnDP); medidas de perturbação do período da F0 (jitter), como jitter local, jitter local absoluto, jitter RAP (Relative Average Perturbation), jitter PPQ5 (Pitch Perturbation Quotient) e jitter DDP (Difference of Differences of Periods); e medidas de perturbação de amplitude da F0 (shimmer), incluindo shimmer local, shimmer em decibéis (dB), shimmer APQ3 (Amplitude Perturbation Quotient de 3 períodos), shimmer APQ5, shimmer APQ11 e shimmer DDA (Difference of Differences of Amplitudes).

Além disso, foram extraídas medidas espectrais, como a proporção harmônico-ruído (HNR), a proporção harmônico-ruído de Dejonckere (Hfno), a diferença entre a amplitude do primeiro e do segundo harmônicos (H1-H2), o declínio espectral e o tilt espectral; medidas de ruído glotal (GNE), incluindo GNE1000, GNE2000 e GNE3000; e medidas no domínio cepstral, com destaque para o CPPS (Cepstral Peak Prominence Smoothed), considerando média, desvio padrão, Q1, mediana, Q3, mínimo, máximo, CV e frequência fundamental do CPPS (F0_CPPS). Também foram consideradas medidas de intensidade (dB), correlação média, voice breaks (quebras de vozeamento) e unvoiced frames (quadros não vozeados).

Além do mais, os pacientes foram solicitados a fornecer o laudo do exame laríngeo para documentação e categorização. Os voluntários foram agrupados de acordo com o diagnóstico laríngeo em cinco categorias: sem lesão laríngea, lesão na porção membranosa da prega vocal, distúrbios neurológicos da voz, fenda glótica sem causa orgânica ou neurológica e distúrbio de voz secundário à refluxo gastroesofágico, bem como disfonia com e sem lesão e disfonia orgânica.

4.5 Análise de dados

Os dados foram organizados em uma planilha eletrônica através do Microsoft Office Excel. A análise estatística foi conduzida com o uso do software estatístico livre R, versão 4.4.3 (R Core Team, 2025). Inicialmente, foi realizada a análise descritiva dos dados, para as variáveis quantitativas foram exploradas medidas centrais e de dispersão, como valor mínimo, valor máximo, média, mediana, desvio padrão e intervalo interquartilico, e as variáveis qualitativas foram analisadas por meio de frequência simples e percentual.

Ademais, para a análise comparativa entre os grupos dos estágios de prontidão contemplação e manutenção, foram selecionados testes estatísticos com base na natureza das variáveis envolvidas. Para variáveis quantitativas, quando as suposições de normalidade foram atendidas foi utilizado o teste paramétrico t de Student para comparar médias dos dois grupos independentes. Quando os dados não seguiram uma distribuição normal, foi aplicado o teste não paramétrico de Wilcoxon. Para variáveis qualitativas, foi utilizado o teste qui-quadrado e o teste exato de Fisher (Bussab e Morettin, 2023). Foi adotado o nível de significância de 5%.

Além disso, para medir a força e a direção da relação linear entre duas variáveis quantitativas, foi gerada uma matriz de gráficos de correlação de Spearman. Essa matriz fornece uma visão geral das correlações entre as variáveis, facilita a identificação de padrões e relações lineares entre elas. No coeficiente de correlação linear de Spearman (r_s) considera-se que $r_s = \pm 1$ indica uma relação linear perfeita; $r_s = \pm 0,70$ indica uma relação linear forte; $r_s = \pm 0,50$ indica uma relação linear moderada; $r_s = \pm 0,30$ indica uma relação linear fraca; e $r_s = 0$ indica ausência de relação linear (Rumsey, 2019). Variáveis altamente correlacionadas não foram incluídas nos modelos de classificação. Além disso, também foi utilizado o método stepwise do tipo *both* para realizar a seleção de variáveis, que combina os procedimentos de entrada e remoção automática de variáveis (Draper; Smith, 1998).

Adicionalmente, foram aplicados modelos de aprendizagem de máquina de classificação, para predição do desfecho de interesse, contudo a escolha do modelo mais adequado dependerá da distribuição dos dados. Para preparar os dados para a modelagem preditiva foram divididos em dois conjuntos, o treinamento com 70% e o teste com 30% dos dados. Depois, realizou-se a imputação dos dados ausentes, foi utilizado o método dos K-vizinhos mais próximos (KNN), técnica baseada em aprendizado de máquina que utiliza os dados disponíveis para identificar padrões e, a

partir disso, estimar os valores faltantes com base nas observações mais semelhantes no conjunto de dados (Costa, 2018).

Devido ao desbalanceamento das classes do banco de dados, foi aplicado a técnica de oversampling no treinamento para criar novas amostras da classe minoritária e igualar sua proporção à da classe majoritária. O oversampling consiste em aumentar artificialmente a quantidade de exemplos da classe minoritária, seja por replicação direta ou por geração de novos exemplos sintéticos, com o objetivo de equilibrar a distribuição das classes e permitir que o modelo aprenda melhor os padrões da classe menos representada (Altalhan; Algarni; Monia, 2025). Dessa maneira, o modelo pode aprender de forma mais eficaz sobre a classe minoritária e melhorar seu desempenho (Chawla *et al.*, 2002). Ademais, para realizar a predição do estágio de prontidão foram usados os seguintes modelos de aprendizado de máquina supervisionados: Regressão logística, KNN, naive bayes, árvore de decisão e floresta aleatória (Apêndice).

Foi aplicada a técnica de validação cruzada com 10 repartições (10-fold cross-validation) durante o processo de treinamento dos modelos. Esse procedimento consiste em dividir aleatoriamente o conjunto de dados em 10 subconjuntos aproximadamente do mesmo tamanho, de modo que, em cada iteração, nove subconjuntos foram usados para treinar o modelo e o décimo para testá-lo, alternando-se até que cada subconjunto fosse utilizado uma vez como teste (Hastie; Tibshirani; Friedman, 2009). A acurácia obtida nas iterações foi utilizada como métrica para a escolha dos melhores hiperparâmetros, com objetivo de melhorar a capacidade preditiva dos algoritmos.

A avaliação do desempenho de modelo foi realizada através de métricas. Na matriz de confusão, uma tabela que compara os valores reais das classes com os valores preditos pelo modelo. Na matriz, a diagonal principal contém as classificações corretas, enquanto os valores que estão fora dessa diagonal representam erros de classificação. Com base nela, são calculadas métricas como acurácia, precisão e sensibilidade (Lopes, 2020).

A acurácia mede a proporção de classificações corretas entre todas as predições feitas pelo modelo. A precisão avalia a capacidade do modelo de evitar falsos positivos, indicando o quão bem ele classificou corretamente uma determinada classe. Já a sensibilidade indica a capacidade do modelo de identificar corretamente

os verdadeiros positivos, mostrando o quão bem ele reconhece uma classe específica entre todos os exemplos reais dessa classe (Lopes, 2020).

Enquanto, a especificidade mensura a proporção de verdadeiros negativos entre todos os indivíduos realmente negativos. O valor preditivo positivo (VPP) representa a proporção de predições positivas que realmente correspondem a casos positivos, enquanto o valor preditivo negativo (VPN) indica a proporção de predições negativas que, de fato, correspondem a casos negativos. A acurácia balanceada, por sua vez, corresponde à média entre sensibilidade e especificidade, sendo especialmente útil em cenários com classes desbalanceadas, pois evita que o desempenho seja inflado artificialmente pela classe majoritária. E o coeficiente Kappa de Cohen avalia o grau de concordância entre as classificações do modelo e os valores reais (Vilela Junior *et al.*, 2022).

A Taxa de Informação Nula (NIR) refere-se à acurácia que seria obtida ao sempre prever a classe majoritária, servindo como comparação básica. O p-valor [Acurácia > NIR] testa se o desempenho do modelo supera estatisticamente essa acurácia nula. O Teste de McNemar é utilizado para comparar a precisão preditiva de dois modelos, bem como para avaliar, em um único modelo de classificação, se há diferença significativa entre o número de falsos positivos e falsos negativos; em ambos os casos, um p-valor baixo ($p < 0,05$) indica diferença estatisticamente significativa (McNemar, 1947). O ponto de corte utilizado nos modelos foi definido com base no critério de Youden, que busca maximizar a soma da sensibilidade e da especificidade, e assim identifica o limiar que proporciona o melhor equilíbrio (Youden, 1950).

4.6 Considerações éticas

O projeto de pesquisa foi analisado e aprovado pelo Comitê de Ética em Pesquisas do Centro de Ciências da Saúde da Universidade Federal da Paraíba – CEP/CCS/UFPB sob o número de parecer 4.866.871/2021 (Anexo 3).

5. RESULTADOS E DISCUSSÃO

5.1 Análise Descritiva

Na análise descritiva realizada, a amostra foi composta por 236 indivíduos, constatou-se que os grupos dos estágios de prontidão contemplação e manutenção, contém 18 (7,63%) e 218 (92,37%) indivíduos, respectivamente. Deste modo,

observa-se um desbalanceamento no tamanho das amostras entre os grupos de contemplação e manutenção, com o grupo de contemplação sendo consideravelmente menor.

A variável idade apresentou $p < 0,0001$, o que indica não seguir uma distribuição normal. O mesmo ocorreu com as variáveis Número Total de Sintomas Vocais ($N^{\circ}\text{Total.SV}$; $p = 0,0325$), Número de Sintomas Vocais Auditivos ($N^{\circ}\text{SVA}$; $p = 0,0008$), Número de Sintomas Vocais Sensoriais ($N^{\circ}\text{SVS}$; $p < 0,0001$), Número de Fatores de Risco Organizacionais ($N^{\circ}\text{FRO}$; $p < 0,0001$), Número de Fatores de Risco Ambientais ($N^{\circ}\text{FRA}$; $p < 0,0001$) e Número de Fatores de Risco Pessoais ($N^{\circ}\text{FRP}$; $p = 0,0133$), cujos p-valores foram todos inferiores a 0,05. A única variável que apresentou distribuição normal foi o Número Total de Fatores de Risco ($N^{\circ}\text{Total.FR}$; $p = 0,09846$), com $p > 0,05$. Como a maioria das variáveis não apresentou distribuição normal, optou-se por utilizar testes não paramétricos para comparação entre os grupos.

A Tabela 1 apresenta estatísticas descritivas do Banco de Dados (BD) da anamnese, que inclui variáveis demográficas, socioeconômicas, queixas, sintomas vocais e fatores de risco.

Tabela 1 - Comparação das variáveis da anamnese com o estágio de prontidão.

Variáveis	Total, n = 236	Estágio de prontidão contemplação, n = 18	Estágio de prontidão manutenção, n = 218	Valor p^2
Idade				0,018*
Média				
(Desvio Padrão)	38,39 (16,30)	29,56 (9,43)	39,11 (16,54)	
Mediana (AIQ)	34,50 (24,00)	27,50 (10,00)	36,00 (24,00)	
Min	18	19	18	
Max	88	57	88	
Sexo				0,4
Masculino	94 (40%)	9 (50%)	85 (39%)	
Feminino	142 (60%)	9 (50%)	133 (61%)	
Estado Civil				0,3
Solteiro	122 (54%)	13 (72%)	109 (52%)	
Casado	83 (36%)	3 (17%)	80 (38%)	
Divorciado	16 (7,0%)	2 (11%)	14 (6,7%)	

Viúvo	5 (2,2%)	0 (0%)	5 (2,4%)	
União Estável	2 (0,9%)	0 (0%)	2 (1,0%)	
Desconhecido	8	0	8	
Grau de Instrução				0,5
Fundamental				
Incompleto	8 (3,6%)	0 (0%)	8 (3,9%)	
Fundamental				
Completo	39 (18%)	1 (5,9%)	38 (19%)	
Médio Incompleto	42 (19%)	3 (18%)	39 (19%)	
Médio Completo	50 (23%)	3 (18%)	47 (23%)	
Superior				
Incompleto	46 (21%)	7 (41%)	39 (19%)	
Superior				
Completo	33 (15%)	3 (18%)	30 (15%)	
Pós-Graduado	3 (1,4%)	0 (0%)	3 (1,5%)	
Desconhecido	15	1	14	
Profissional da Voz				0,6
Não	158 (67%)	13 (72%)	145 (67%)	
Sim	78 (33%)	5 (28%)	73 (33%)	
Queixa				0,07*
Sintomas				
Auditivos	81 (34%)	6 (33%)	75 (34%)	
Sintomas				
Proprioceptivos	14 (5,9%)	1 (5,6%)	13 (6,0%)	
Sintomas Mistos	111 (47%)	5 (28%)	106 (49%)	
Aprimoramento				
Vocal	19 (8,1%)	4 (22%)	15 (6,9%)	
Encaminhamento				
Médico	6 (2,5%)	1 (5,6%)	5 (2,3%)	
Patologia laríngea	5 (2,1%)	1 (5,6%)	4 (1,8%)	
Nº Total SV				0,003*
Média (Desvio				
Padrão)	11,77 (4,98)	8,00 (5,39)	12,08 (4,82)	
Mediana (AIQ)	11,00 (8,00)	7,00 (7,50)	12,00 (8,00)	

Min	0	0	1	
Max	24	17	24	
Nº SVA				0,009*
Média (Desvio				
Padrão)	6,32 (2,64)	4,61 (2,68)	6,46 (2,60)	
Mediana (AIQ)	6,00 (4,00)	4,50 (3,00)	7,00 (3,75)	
Min	0	0	0	
Max	12	9	12	
Nº SVS				0,005*
Média				
(Desvio Padrão)	5,39 (3,05)	3,39 (3,09)	5,56 (2,99)	
Mediana (AIQ)	5,00 (5,00)	2,50 (5,50)	5,00 (5,00)	
Min	0	0	0	
Max	12	9	12	
Nº Total FR				0,9
Média				
(Desvio Padrão)	13,08 (5,86)	13,00 (4,43)	13,09 (5,97)	
Mediana (AIQ)	13,00 (8,00)	12,00 (4,75)	14,00 (9,50)	
Min	0	6	0	
Max	29	23	29	
Nº FRO				0,5
Média				
(Desvio Padrão)	2,20 (1,77)	2,44 (1,65)	2,18 (1,78)	
Mediana (AIQ)	2,00 (4,00)	2,00 (3,00)	2,00 (4,00)	
Min	0	0	0	
Max	5	5	5	
Nº FRA				0,3
Média				
(Desvio Padrão)	2,98 (2,29)	2,33 (1,85)	3,03 (2,32)	
Mediana (AIQ)	3,00 (4,00)	2,00 (2,75)	3,00 (4,00)	
Min	0	0	0	
Max	9	6	9	
Nº FRP				0,5
Média				
(Desvio Padrão)	7,88 (3,24)	8,61 (2,99)	7,82 (3,26)	

Mediana (AIQ)	8,00 (4,00)	7,50 (5,50)	8,00 (4,00)
Min	0	5	0
Max	15	14	15

Legenda: ¹Mediana (Q1 – Q3); n (%), ²Teste de soma de postos de Wilcoxon; Teste qui-quadrado de independência; Teste exato de Fisher. *p < 0.05. **Fonte:** Dados da Pesquisa, 2025.

Na Tabela 1, a idade média geral foi de 38,39 anos, sendo significativamente menor no grupo em contemplação (29,56 anos) em comparação ao grupo em manutenção (39,11 anos). Quanto ao sexo, observou-se predominância do feminino (60%), com distribuição semelhante entre os estágios, não havendo diferença estatística. Em relação ao estado civil, a maior parte era solteira (54%), seguida de casados (36%), com distribuição semelhante entre os grupos. No grau de instrução, os níveis mais frequentes foram ensino médio completo (23%) e superior incompleto (21%), com distribuição semelhante entre os grupos. A maioria dos participantes não se identificava como profissional da voz (67%), também sem diferença significativa entre os grupos.

No que se refere à queixa vocal, os sintomas mistos foram os mais frequentes (47%), seguidos por sintomas auditivos (34%). Embora não tenha alcançado significância estatística, observou-se tendência para maior proporção de participantes com queixa relacionada ao aprimoramento vocal no grupo contemplação (22%) em comparação ao grupo manutenção (6,9%). Observou-se diferença significativa nos sintomas vocais, grupo contemplação apresentou menor N° total de SV (média de 8,00) em relação ao grupo manutenção (12,08). O mesmo padrão foi observado no número de SVA e de SVS. No número total de FR, não houve diferença entre os grupos (p = 0,9), assim como para o N° de FRO, FRA e FRP.

Esses dados sugerem que pessoas mais velhas tendem a apresentar maior percepção dos prejuízos causados a voz, conseqüentemente, maior motivação para modificações comportamentais. Pode estar relacionado à maturação do córtex pré-frontal que ocorre por volta dos 25 anos e é responsável pelas funções executivas que envolvem o controle comportamental (Arain *et al.*, 2013), com essa maturação ocorre o aumento do controle dos impulsos, do estabelecimento de metas e do desenvolvimento de estratégias eficazes.

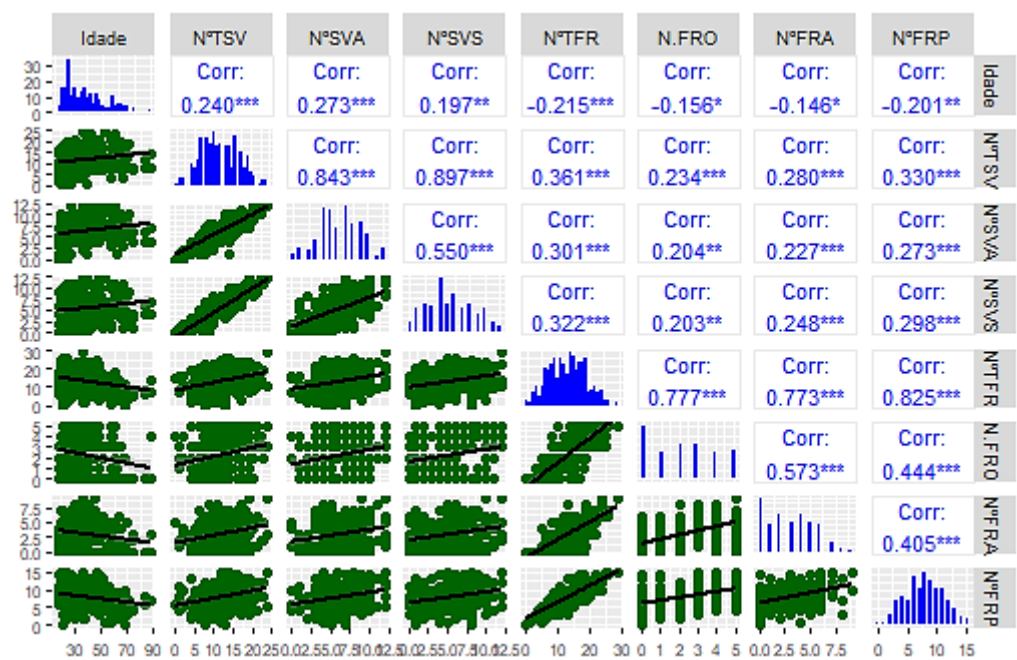
Em relação ao número de sintomas vocais, sejam auditivos ou sensoriais, quanto maior for esse número, maior pode ser a percepção do problema de voz e maior o impacto na vida do indivíduo, o que pode motivar a buscar por terapia (Deary *et al.*, 2003). Assim, o maior número de sintomas pode favorecer a mudança do indivíduo do estágio de contemplação para o estágio de manutenção.

Ademais, o número de fatores de risco é semelhante nos dois estágios. Como os participantes estavam iniciando a terapia, é provável que ainda não tinham recebido orientações específicas sobre os riscos. Muitos desses fatores, especialmente os relacionados ao ambiente de trabalho, são difíceis de modificar, e a falta de conscientização pode ter contribuído para a semelhança entre os grupos, sugerindo que o entendimento sobre os riscos se desenvolve durante o processo terapêutico (Ghirardi; Constantini, 2022).

Além disso, as variáveis Estado Civil e Grau de Instrução apresentaram distribuição desigual em relação à variável resposta Estágio de Prontidão, caracterizando um caso de separação completa (Agresti, 2002), uma vez que não houve indivíduos no estágio de contemplação que fossem viúvos, estivessem em união estável ou que tivessem ensino fundamental incompleto ou pós-graduação. Embora esse padrão possa sugerir uma associação relevante entre essas variáveis e o estágio de prontidão, ele inviabiliza a estimação adequada dos coeficientes, pois os algoritmos de estimação tendem a atribuir valores infinitos para representar essa separação perfeita, o que compromete a convergência do modelo e resulta em estimativas instáveis e erros-padrão excessivamente elevados. Dessa forma, tais variáveis não foram incluídas nos modelos para não comprometerem a acurácia das predições.

Realizou-se também a matriz de correlação de Spearman entre as variáveis explicativas contínuas, que pode ser observada na Figura 3.

Figura 3 - Matriz de Gráficos de Correlação: dados da Anamnese.



Fonte: Dados da Pesquisa, 2025.

Observa-se na Figura 3 que o Número Total de SV apresentou correlação positiva forte com o Número de SVA e com o Número de SVS. Além de uma correlação positiva forte do Número Total de FR com FRO, FRA e FRP. Desta forma, para os modelos foram considerados apenas o Número Total de SV e Número Total de FR, pois como são variáveis fortemente correlacionadas poderia haver sobreposição para o modelo.

A Tabela 2 apresenta a comparação de proporção de participantes que tiveram os resultados relacionados ao Protocolo do Índice de Desvantagem Vocal (IDV-10-TRI), à Escala de Qualidade de Vida em Voz (QVV-TRI), à Escala de Sintomas Vocais (ESV-TRI) e à Escala de Desconforto do Trato Vocal (EDTV-TRI). A resposta 'sim' indica que o participante apresentou resultado compatível com presença de disfonia segundo o respectivo instrumento, enquanto 'não' indica a ausência.

Tabela 2 - Comparação de instrumentos de autoavaliação com o estágio de prontidão.

Variáveis	Total, n = 236	Estágio de prontidão contemplação, n = 18	Estágio de prontidão manutenção, n = 218	Valor p ²
IDV10-TRI				0,023*
Não	61 (26%)	9 (50%)	52 (24%)	
Sim	175 (74%)	9 (50%)	166 (76%)	

QVV-TRI				0,2
Não	21 (8,9%)	3 (17%)	18 (8,3%)	
Sim	214 (91%)	15 (83%)	199 (92%)	
Desconhecido	1	0	1	
ESV-TRI				0,027*
Não	43 (18%)	7 (39%)	36 (17%)	
Sim	193 (82%)	11 (61%)	182 (83%)	
EDTV-TRI				0,5
Não	33 (22%)	4 (31%)	29 (22%)	
Sim	114 (78%)	9 (69%)	105 (78%)	
Desconhecido	89	5	84	

Legenda: ¹n (%), ²Teste exato de Fisher. *p < 0.05. **Fonte:** Dados da Pesquisa, 2025.

Verifica-se na Tabela 2 que o IDV10-TRI e a ESV-TRI, apresentaram diferença estatisticamente significativa entre os estágios, com um maior número de pessoas com indicativo de disfonia a partir destes instrumentos no estágio de Manutenção. Esses achados sugerem que uma maior percepção do impacto na funcionalidade vocal e na presença de sintomas leva os indivíduos a reconhecerem a necessidade de mudança. Assim, essa conscientização motiva os pacientes a iniciar ou consolidar estratégias para modificar seus comportamentos vocais. Por outro lado, não foram observadas diferenças estatisticamente significativas entre os estágios quanto à QVV-TRI e à EDTV-TRI.

Na Tabela 3 apresenta estatísticas descritivas de outros dados referentes à avaliação multidimensional da voz, como o JPA, o diagnóstico laríngeo e a classificação da disfonia.

Tabela 3 - Comparação das variáveis referentes ao JPA, Diagnóstico Laríngeo e Classificação da Disfonia com estágio de prontidão.

Variáveis	Total, n = 236	Estágio de prontidão contemplação, n = 18	Estágio de prontidão manutenção, n = 218	Valor p ²
Alteração do JPA				0,010
Sim	61 (27%)	10 (56%)	51 (25%)	
Não	162 (73%)	8 (44%)	154 (75%)	
Desconhecido	13	0	13	

Classificação do JPA				<0,001
Normal	61 (27%)	10 (56%)	51 (25%)	
Leve	79 (35%)	8 (44%)	71 (35%)	
Moderado	79 (35%)	0 (0%)	79 (39%)	
Intenso	4 (1,8%)	0 (0%)	4 (2,0%)	
Desconhecido	13	0	13	
Diagnóstico Laríngeo				0,683
Sem lesão laríngea	96 (43%)	9 (53%)	87 (43%)	
Lesão na porção membranosa das PPVV	53 (24%)	2 (12%)	51 (25%)	
Distúrbios neurológicos da voz	8 (3,6%)	0 (0%)	8 (3,9%)	
Fenda glótica sem causa orgânica ou neurológica	33 (15%)	3 (18%)	30 (15%)	
Distúrbio de voz secundário à refluxo gastroesofágico	31 (14%)	3 (18%)	28 (14%)	
Desconhecido	15	1	14	
Alteração Laríngea				0,437
Sim	110 (50%)	10 (59%)	100 (49%)	
Não	111 (50%)	7 (41%)	104 (51%)	
Desconhecido	15	1	14	
Disfonia				0,287
Comportamental com lesão	74 (35%)	3 (19%)	71 (36%)	
Comportamental sem lesão	119 (56%)	11 (69%)	108 (55%)	
Orgânica	18 (8,5%)	2 (13%)	16 (8,2%)	

Legenda: ¹n (%), ²Teste exato de Fisher. **Fonte:** Dados da Pesquisa, 2025.

Na Tabela 3, observou-se que o estágio Contemplação apresenta uma proporção maior de indivíduos classificados em variabilidade normal da qualidade vocal enquanto no estágio de Manutenção, predomina a voz classificada em moderada, indicando uma percepção mais acentuada da alteração vocal nos indivíduos neste estágio de prontidão. Dessa forma, como alterações da qualidade de voz podem prejudicar a comunicação e relações sociais (Santos *et al.*, 2016), pode assim levar os indivíduos a buscarem mudanças para melhora da saúde vocal. Dessa forma, sugere que pessoas com voz maior intensidade do desvio vocal percebem mais o impacto da disfonia, o que pode ser um fator determinante para o avanço nos estágios de prontidão.

Além disso, na Classificação do JPA não são encontrados indivíduos classificados em moderado e intenso que pertençam ao estágio de prontidão contemplação, e no Diagnóstico Laríngeo nenhum indivíduo em contemplação apresenta distúrbios neurológicos da voz. Isso indica que são casos de separação completa, como essas variáveis apresentam distribuição desigual em relação variável resposta (Agresti, 2002), não foram incluídas no modelo. No entanto, a classificação do JPA se relaciona à presença de alteração no JPA, assim como o diagnóstico laríngeo se relaciona com a presença ou ausência de alteração laríngea, sendo utilizadas nos modelos as variáveis Alteração da JPA, Alteração Laríngea e Classificação da Disfonia.

Ademais, também foram criados modelos utilizando medidas acústicas que abrangem diferentes domínios da produção vocal. Aplicou-se o teste de Shapiro-Wilk a todas as variáveis das medidas acústicas. Os resultados indicam que a maioria das variáveis apresentou p-valores menores que 0,05, o que sugere rejeição da hipótese nula de normalidade para essas variáveis. Entre os exemplos mais expressivos estão variáveis como F0media ($p < 0,0001$), F0desvio ($p < 0,0001$), Correlacao.media ($p < 0,0001$), VoiceBreaks ($p < 0,0001$) e Jitter.local ($p < 0,0001$), todas com distribuição significativamente diferente da normal. Apenas três variáveis apresentaram p-valores acima do limiar de significância de 0,05, não rejeitando a hipótese de normalidade: HNRD ($p = 0,6906$), Hfno ($p = 0,1213$) e CPPSmedia ($p = 0,1927$).

Além disso, foi realizada uma análise de correlação de Spearman entre as variáveis acústicas, com o objetivo de identificar colinearidades e reduzir redundâncias que pudessem impactar o desempenho do modelo. Foram excluídas as variáveis que apresentaram correlação forte entre si ($\geq 0,07$), priorizou aquelas com

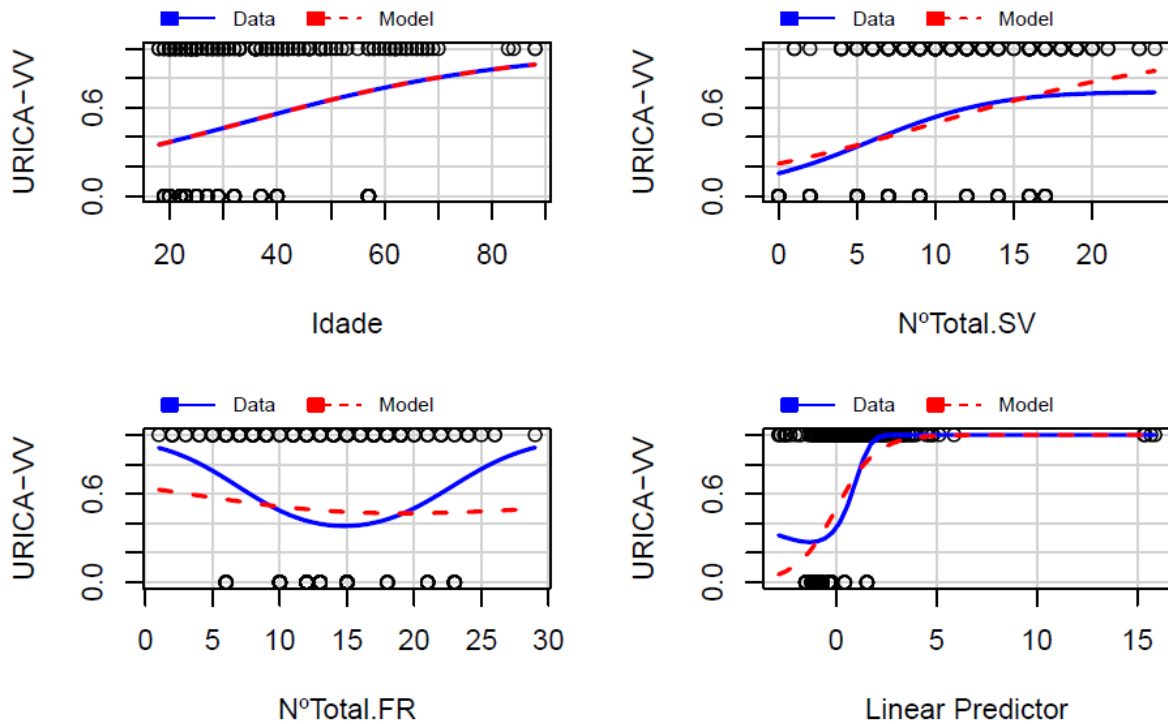
maior relevância estatística e plausibilidade clínica na predição do estágio de prontidão para a terapia vocal. As variáveis selecionadas após essa etapa foram: F0media, F0desvio, Intensidade.db, Período, PSD, VoiceBreaks, UnvoicedFrames, Shimmer.local, HNR, Hfno, H1H2, Declínio, Tilt, GNE1000 e CPPSdesvio.

Ademais, foi utilizada a técnica de seleção de variáveis stepwise, com direção do tipo both, para o modelo final foram selecionadas as variáveis Intensidade.db, Período, Shimmer.local, Tilt e CPPSdesvio, as quais apresentaram associação estatisticamente significativa com a variável resposta e contribuíram para melhorar o desempenho do modelo.

5.2 Modelos de AM – Dados da Anamnese

Os primeiros modelos apresentados usaram o banco de dados (BD) da anamnese (sexo, idade, queixa, profissional da voz, número de sintomas vocais e número de fatores de risco). Essas variáveis foram empregadas como preditoras na construção de modelos de aprendizado de máquina supervisionado. Inicialmente, foi desenvolvido o modelo de regressão logística. Na Figura 4, é possível avaliar a relação entre as variáveis preditoras e a variável resposta foi adequadamente modelado na regressão logística. O efeito marginal mostra a resposta prevista do modelo logístico em função do preditor, mantendo as demais variáveis constantes. A linha azul representa a tendência dos dados reais, enquanto a linha vermelha tracejada indica a predição feita pelo modelo.

Figura 4 - Efeito Marginal do Modelo de Regressão Logística: Dados da Anamnese.



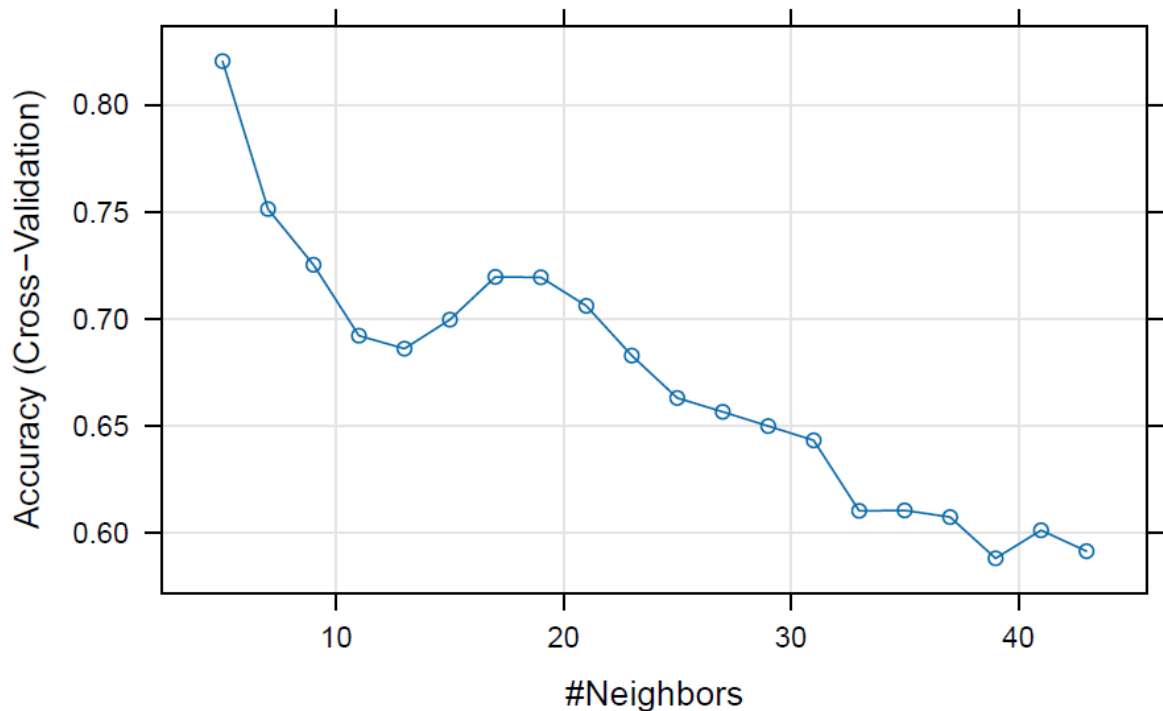
Fonte: Dados da Pesquisa, 2025.

Na Figura 4, observa-se que as variáveis idade e N° Total de SV apresentam relações crescentes com a resposta, o que indica bom ajuste do modelo. Entretanto, na variável N° Total de FR, a relação é mais complexa e não linear. Além disso, o gráfico do preditor linear mostra a curva sigmoide esperada da regressão logística, com bom alinhamento entre os valores observados e os preditos, indica um bom ajuste geral do modelo.

Quanto ao desempenho geral do modelo de regressão logística, a área sob a curva ROC (AUC) foi de 0,6031. Esse valor indica que o modelo possui uma capacidade moderada de discriminar entre os estágios da prontidão. Embora superior a um classificador aleatório ($AUC = 0,5$), esse resultado sugere que o modelo ainda apresenta limitações. Além disso, o ponto de corte escolhido a partir do critério de Youden foi de 0,710.

Depois, foi utilizado o algoritmo k-Nearest Neighbors (KNN). Para selecionar o valor ideal do hiperparâmetro k (número de vizinhos), realizou-se uma validação cruzada com 10 partições (10-fold cross-validation). A Figura 5 apresenta a acurácia obtida para diferentes valores do k. Através desta visualização, pode-se identificar o valor de k que amplia a precisão do modelo em dados não vistos.

Figura 5 - Número de Vizinhos do KNN com Base na Acurácia: Dados da Anamnese.



Fonte: Dados da Pesquisa, 2025.

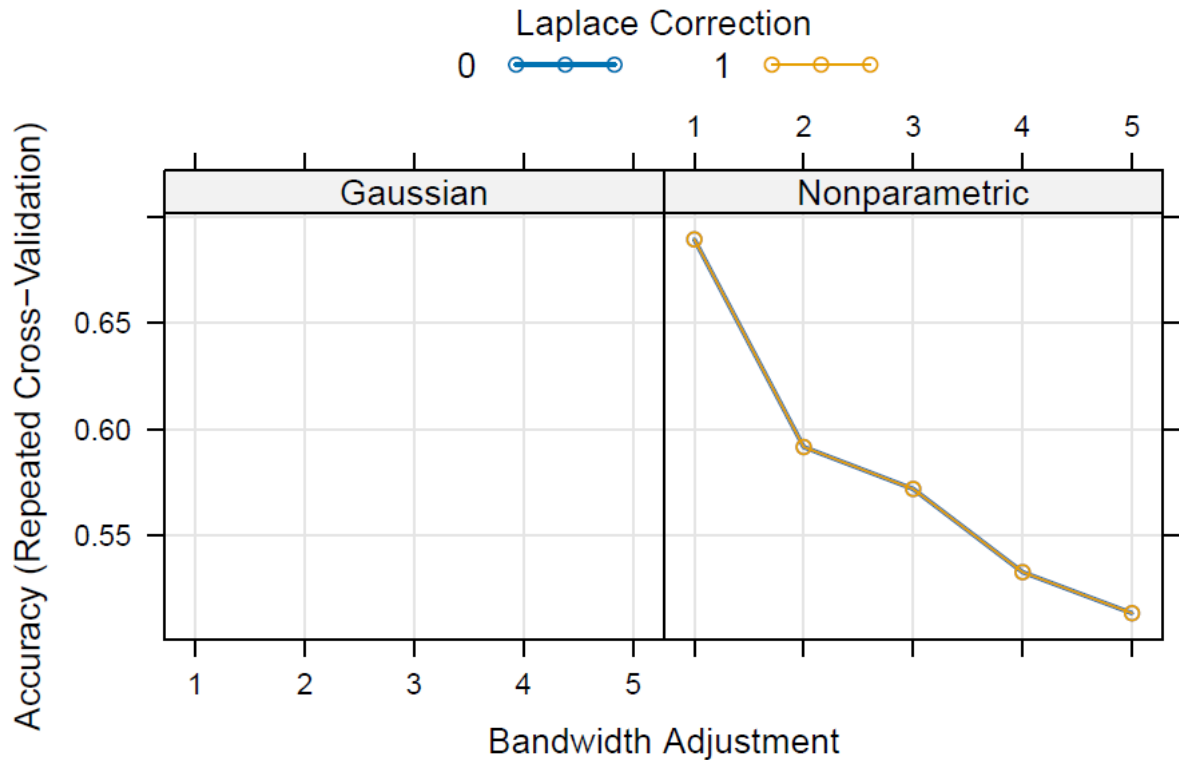
A Figura 5 indica que o valor ideal de k para este modelo KNN é 5, pois é nesse ponto que o modelo alcança a maior acurácia (aproximadamente 82%) durante o processo de validação cruzada. Utilizar um k maior resultaria em um modelo com menor poder preditivo (menor acurácia). Portanto, $k = 5$ foi escolhido como o hiperparâmetro ideal para este modelo KNN.

Outrossim, o modelo KNN apresentou AUC de 0,5923, indicando uma capacidade moderada, porém ligeiramente inferior a da regressão logística, de discriminar entre os indivíduos nos diferentes estágios de prontidão. Seguindo o critério de Youden, o ponto de corte escolhido foi de 0,024, para classificar os indivíduos no estágio de prontidão.

No modelo Naive Bayes, a acurácia foi utilizada como critério para selecionar os hiperparâmetros fL , $usekernel$ e $adjust$. O parâmetro fL define o grau de suavização de Laplace, sendo usado para evitar probabilidades nulas, especialmente em variáveis categóricas. O parâmetro $usekernel$ indica se será utilizada a estimativa de densidade por kernel, permite uma modelagem não paramétrica e mais flexível da distribuição das variáveis contínuas. Já o parâmetro $adjust$ regula a largura da função

kernel utilizada na estimativa de densidade, influencia diretamente o grau de suavização aplicado (John; Langley, 1995).

Figura 6 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados da Anamnese.

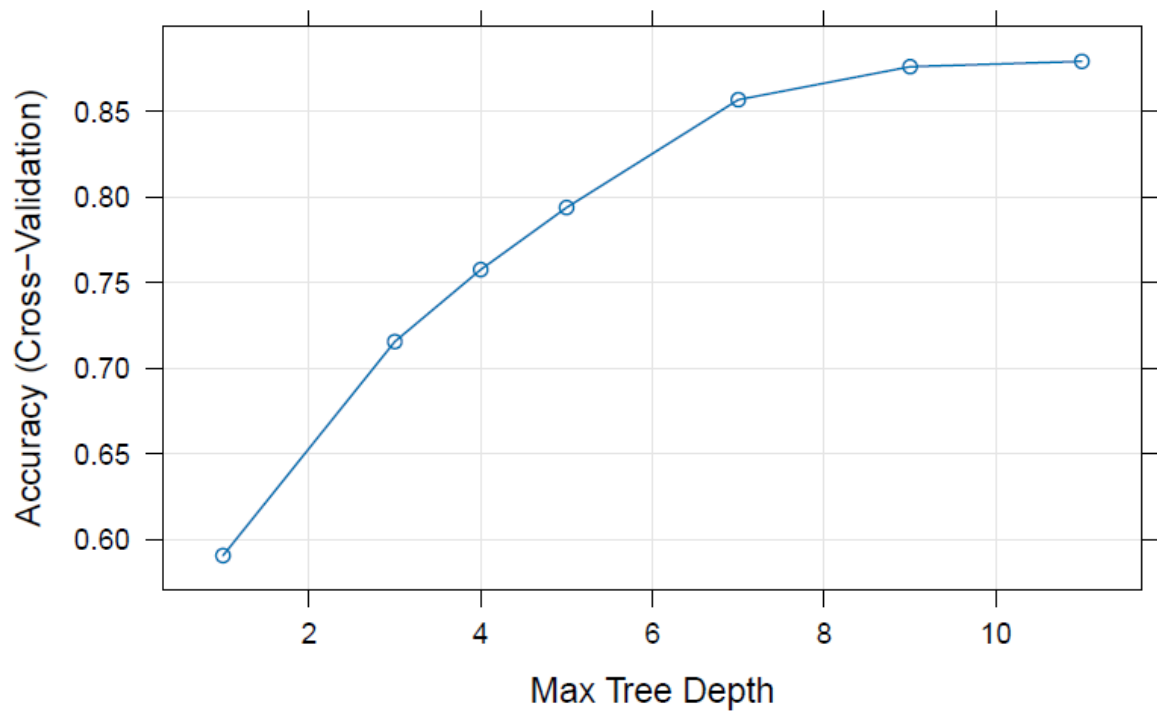


Fonte: Dados da Pesquisa, 2025.

Com base na Figura 6, os hiperparâmetros adotados para o modelo Naive Bayes foram: $fL = 0$, $usekernel = TRUE$ e $adjust = 1$. O valor $fL = 0$ define que não será aplicada suavização aos dados de treinamento, ou seja, o modelo utilizará os dados como estão. O hiperparâmetro $usekernel = TRUE$ ativa a estimativa de densidade por kernel, permite modelar a distribuição das variáveis de forma mais flexível e não paramétrica. Já o $adjust = 1$ indica o uso do valor padrão de ajuste da largura da função kernel.

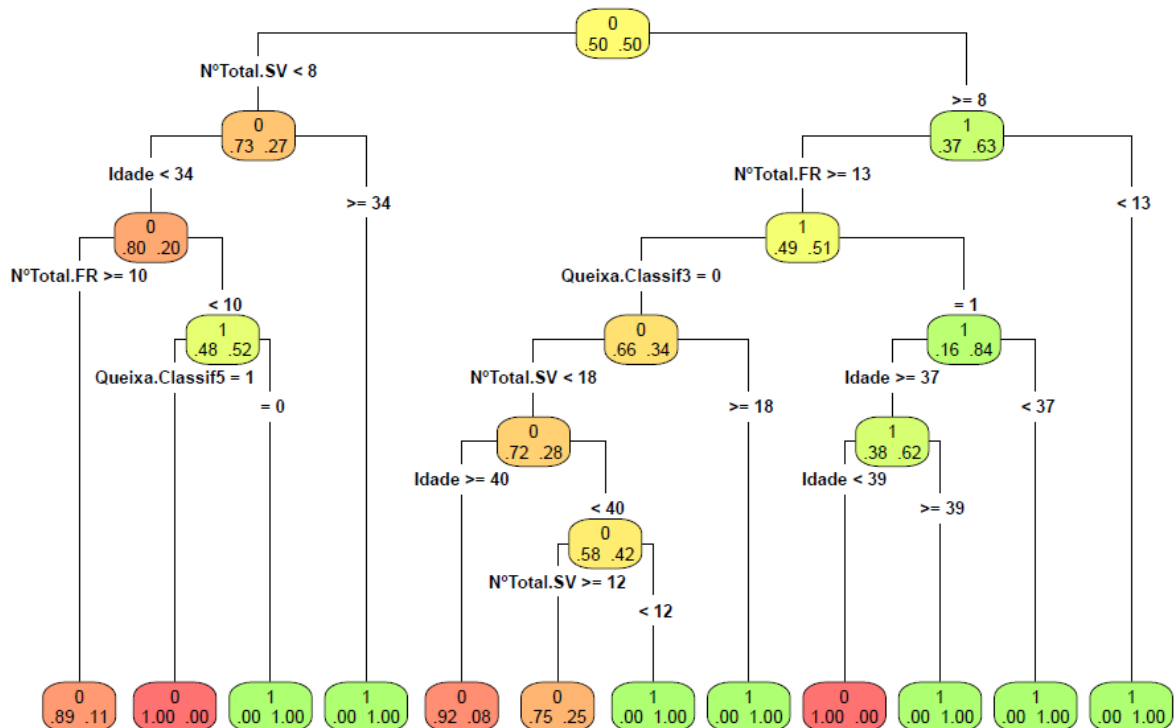
Quanto a árvore de decisão, a Figura 7 mostra como a acurácia do modelo varia com a profundidade máxima permitida para suas árvores. A profundidade máxima para a árvore escolhida para o modelo foi 11 por apresentar a maior acurácia, esse parâmetro controla o quanto a árvore pode crescer.

Figura 7 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia:
Dados da Anamnese.



Fonte: Dados da Pesquisa, 2025.

A Figura 8 apresenta a árvore de decisão gerada pelo modelo, que ilustra as regras utilizadas para classificar as observações com base nas variáveis preditoras.

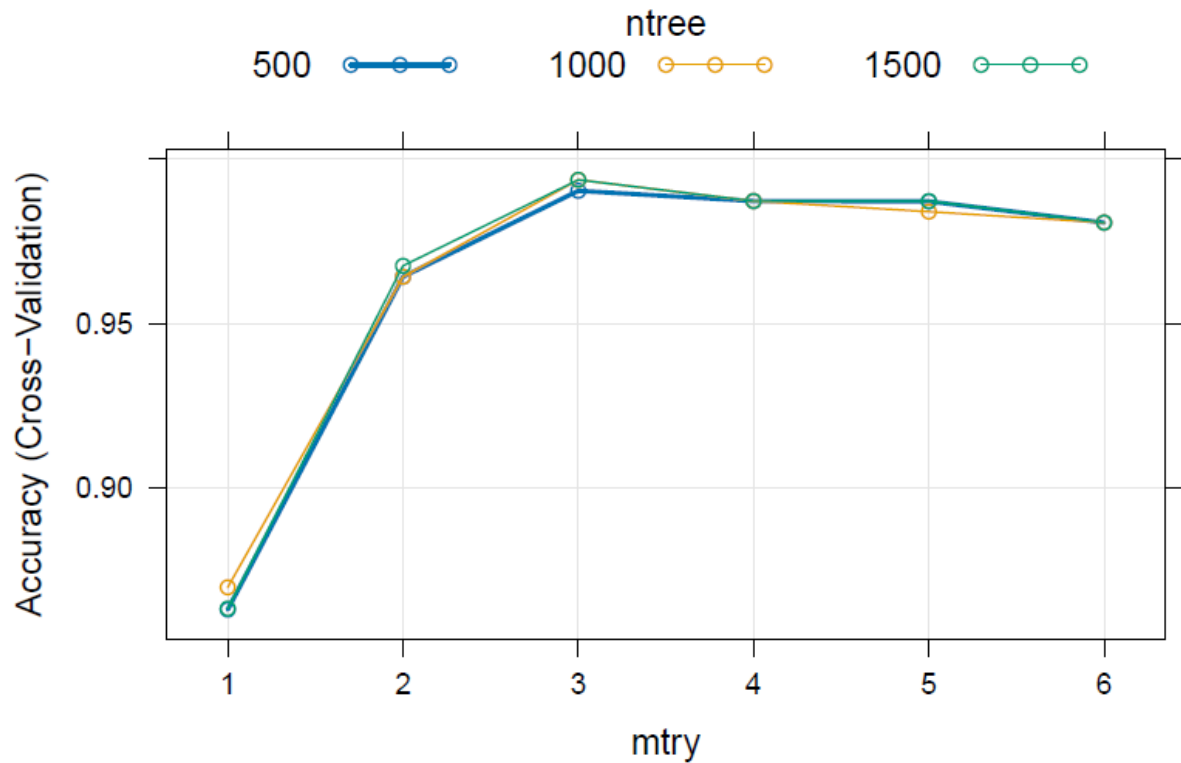
Figura 8 - Árvore de Decisão com os Dados da Anamnese.

Fonte: Dados da Pesquisa, 2025.

A primeira divisão da árvore ocorre com base na variável N° Total de SV, o que indica sua importância como preditora. Em seguida, a árvore se ramifica por outras variáveis relevantes, como idade, N° Total de FR e queixa. O modelo de árvore de decisão apresentou AUC de 0,4077, indica baixa capacidade discriminativa do modelo. O ponto de corte utilizado foi 0,1, o modelo não foi capaz de prever adequadamente o estágio de prontidão Contemplação o que também é evidenciado pela baixa AUC.

Por fim, ainda com os dados da anamnese foi criado um modelo de Floresta Aleatória, a Figura 9 apresenta a variação da acurácia para diferentes combinações de hiperparâmetros, assim permite visualizar qual configuração proporcionou o melhor desempenho para este modelo.

Figura 9 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados da Anamnese.

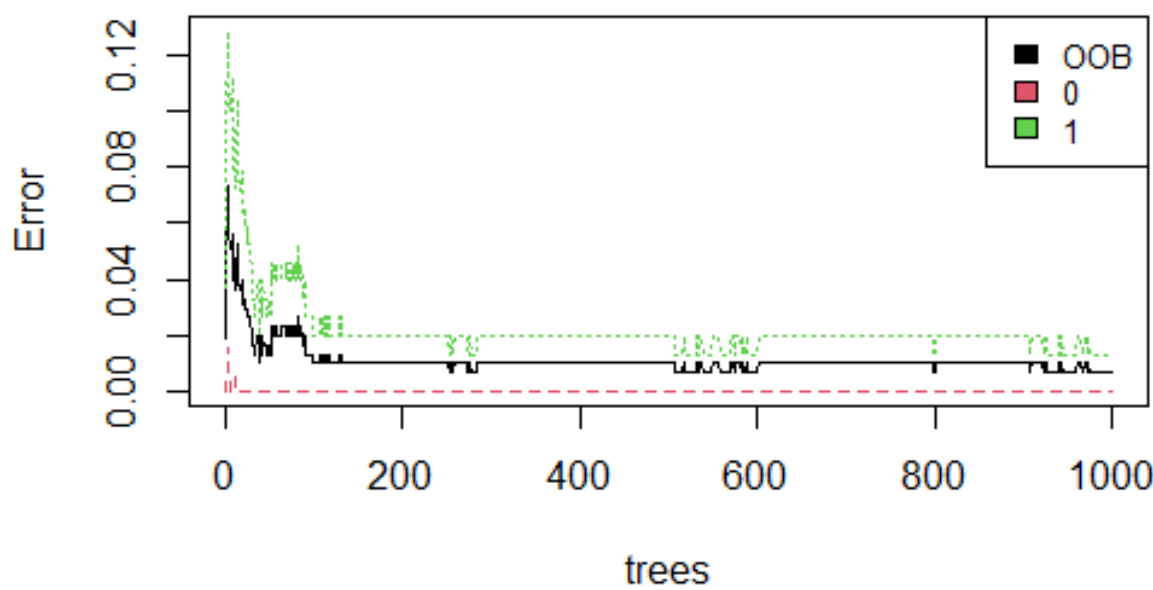


Fonte: Dados da Pesquisa, 2025.

Com base no observado na Figura 9, os hiperparâmetros adotados para o modelo foram $mtry = 3$, que determina o número de variáveis consideradas em cada divisão da árvore, e $ntree = 1000$, que corresponde ao número total de árvores construídas no modelo de Floresta Aleatória.

O erro OOB é uma estimativa do erro de generalização calculado durante o treinamento, faz uso dos dados que não foram usados na construção de cada árvore no modelo de floresta aleatória. Na Figura 10, pode-se observar a visualização do erro OOB.

Figura 10 - Erro OOB no Modelo Floresta Aleatória: Dados da Anamnese.



Fonte: Dados da Pesquisa, 2025.

De acordo com a Figura 10, as taxas de erro OOB são inicialmente altas, mas conforme o número de árvores aumenta, as taxas de erro diminuem e, em seguida, se estabilizam em valores baixos com o erro OOB de 1.31%. Isso indica que o modelo atinge um desempenho robusto e generalizável, sendo ligeiramente melhor na classificação do Estágio de Prontidão Manutenção.

Essa matriz a seguir mostra como o modelo classificou as amostras OOB, separadas por estágio:

Tabela 4 - Matriz de confusão OOB: dados da Anamnese.

	Contemplação	Manutenção	Erro da classe
Contemplação	153	0	0.000 (0%)
Manutenção	4	149	0.026 (2,6%)

Fonte: Dados da Pesquisa, 2025.

A matriz de confusão OOB mostrou acurácia quase perfeita, com todos os acertos no estágio de contemplação e apenas quatro erros no estágio de manutenção, dessa forma indica alta capacidade de classificação. Apesar desses resultados

promissores, é importante ressaltar que a validação em dados de teste é essencial para verificar a capacidade de generalização do modelo.

O valor da AUC obtido foi de 0,8262, indicando um bom poder discriminativo do modelo. O melhor comparado aos demais modelos. Foi identificado o ponto de corte com o critério de Youden de 0,853. Com esse novo ponto de corte, foi realizada a classificação dos casos, permitindo a construção da matriz de confusão a seguir:

Tabela 5 – Matriz de Confusão da Floresta Aleatória: Dados da Anamnese.

	Contemplação	Manutenção
Contemplação	5	22
Manutenção	0	43

Fonte: Dados da Pesquisa, 2025.

Na matriz de confusão com os dados de teste, o modelo apresentou desempenho perfeito na classificação de indivíduos no estágio de contemplação, sem ocorrência de falsos positivos. No entanto, teve dificuldade em identificar corretamente os casos do estágio de manutenção, cometendo 22 falsos negativos. Esse desempenho sugere limitação na capacidade de generalização do modelo e evidencia a importância da validação externa para uma avaliação mais precisa.

Tabela 6 – Desempenho dos Modelos para Classificação do Estágio de Prontidão com Dados da Anamnese.

	Regressão Logística	KNN	Naive Bayes	Árvore de Decisão	Floresta Aleatória
Acurácia	0,6429	0,9286	0,7286	0,8429	0,6857
Acurácia Balanceada	0,7154	0,6846	0,3923	0,4538	0,8308
Kappa	0,1379	0,4068	-0,1176	-0,0845	0,2183
Sensibilidade	0,6308	0,9692	0,7846	0,9077	0,6615
Especificidade	0,8000	0,4000	0,0000	0,0000	1,0000
Valor Preditivo Positivo	0,9762	0,9545	0,9107	0,9219	1,0000

Valor Preditivo					
Negativo	0,1429	0,5000	0,0000	0,0000	0,1852
Taxa de Detecção	0,5857	0,9000	0,7286	0,8429	0,6143
Prevalência					
Estimada	0,6000	0,9429	0,8000	0,9143	0,6143
Taxa de Informação	0,9286	0,9286	0,9286	0,9286	0,9286
Nula					
P-Valor [Acurácia > Taxa de Informação Nula]					
	1,0000	0,6160	1,0000	0,9962	1,0000
P-Valor do Teste de McNemar					
	1,083e-05	1,000	0,0665	1,0000	7,562e-06

Fonte: Dados da Pesquisa, 2025.

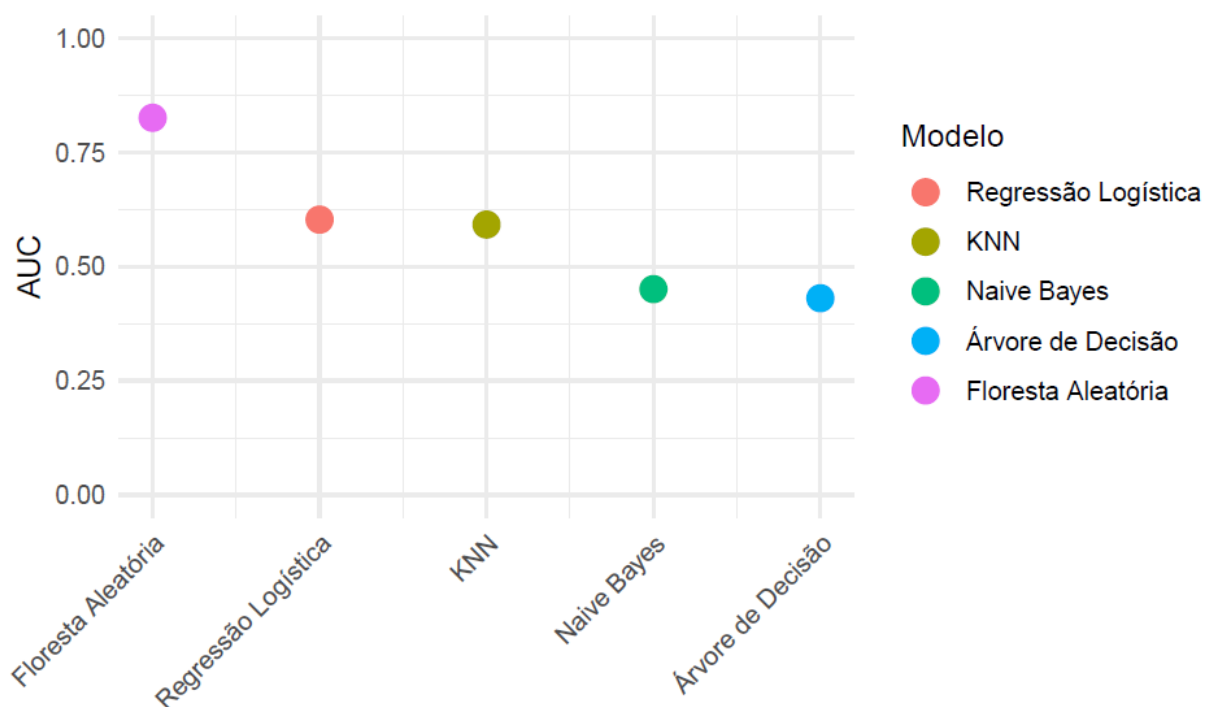
A Tabela 6 apresenta uma visão abrangente das métricas de desempenho dos modelos. Os modelos avaliados devem apresentar acurácia superior a 92,37%, valor equivalente ao desempenho de um classificador que atribui todas as observações à classe majoritária (estágio de manutenção). Superar esse valor indica que o modelo é capaz de identificar padrões relevantes nos dados, classificando corretamente observações que não pertencem à classe majoritária.

Entre os modelos testados, apenas o K-NN demonstrou efetividade, apresentando desempenho acima (0,9286). Em relação à sensibilidade, o KNN apresentou um valor alto (0,9692), indicando que é eficaz na detecção do estágio de manutenção. No entanto, sua especificidade foi baixa (0,4000), o que sugere uma maior propensão a classificar erroneamente o estágio de contemplação. Sugere que os dados da anamnese podem contribuir para identificar o estágio em que o paciente se encontra, permitindo estratégias de intervenção mais direcionadas desde o início da terapia de voz.

A Floresta Aleatória apresentou acurácia balanceada mais alta (0,8308), além de especificidade perfeita (1,0000) e valor preditivo positivo também perfeito (1,0000), o que demonstra excelente desempenho na identificação de contemplação e precisão nas classificações de manutenção. Considerando o conjunto de métricas e a necessidade de classificar adequadamente ambos os estágios nos dados da anamnese, a floresta aleatória é um modelo mais indicado.

A Regressão Logística também se destacou pela alta especificidade (0,8000) e pelo valor preditivo positivo elevado (0,9762), embora sua sensibilidade tenha sido mais baixa (0,6308), o que indica maior dificuldade em identificar corretamente os casos no estágio de manutenção. Enquanto os modelos Naive Bayes e Árvore de Decisão apresentaram especificidade nula (0,0000), o que indica falha total na identificação dos casos de contemplação. Isso também é refletido nos valores negativos ou muito baixos do Kappa e da acurácia balanceada, sugere que apesar de apresentarem acurácias absolutas aparentemente altas, seu desempenho prático é comprometido.

Figura 11 – Valores de AUC dos Modelos: Dados da Anamnese.



Fonte: Dados da Pesquisa, 2025.

Com base nos valores de AUC obtidos (Figura 11), observa-se que a Floresta Aleatória apresentou o melhor desempenho, com AUC de 0,8262, indicando maior capacidade de discriminar corretamente os diferentes estágios de prontidão. Em seguida, aparecem a Regressão Logística e o KNN, com valores de 0,6031 e 0,5923, respectivamente, sugerindo desempenho moderado. Já os modelos Naive Bayes (0,4508) e Árvore de Decisão (0,4308) apresentaram valores abaixo de 0,5, o que indica desempenho inferior, próximo ou abaixo do nível esperado pelo acaso.

Esses resultados sugerem que a escolha do modelo influencia de forma relevante a capacidade preditiva no contexto avaliado, pois diferentes algoritmos apresentam desempenhos variados na classificação do estágio de prontidão. A Floresta Aleatória, conseguiu capturar padrões mais complexos nos dados e resultou em maior discriminabilidade, refletida pelo valor mais alto de AUC. Por outro lado, modelos como Naive Bayes e Árvore de Decisão, que apresentaram desempenho inferior, podem não ser adequados para lidar com as características específicas dos dados da anamnese.

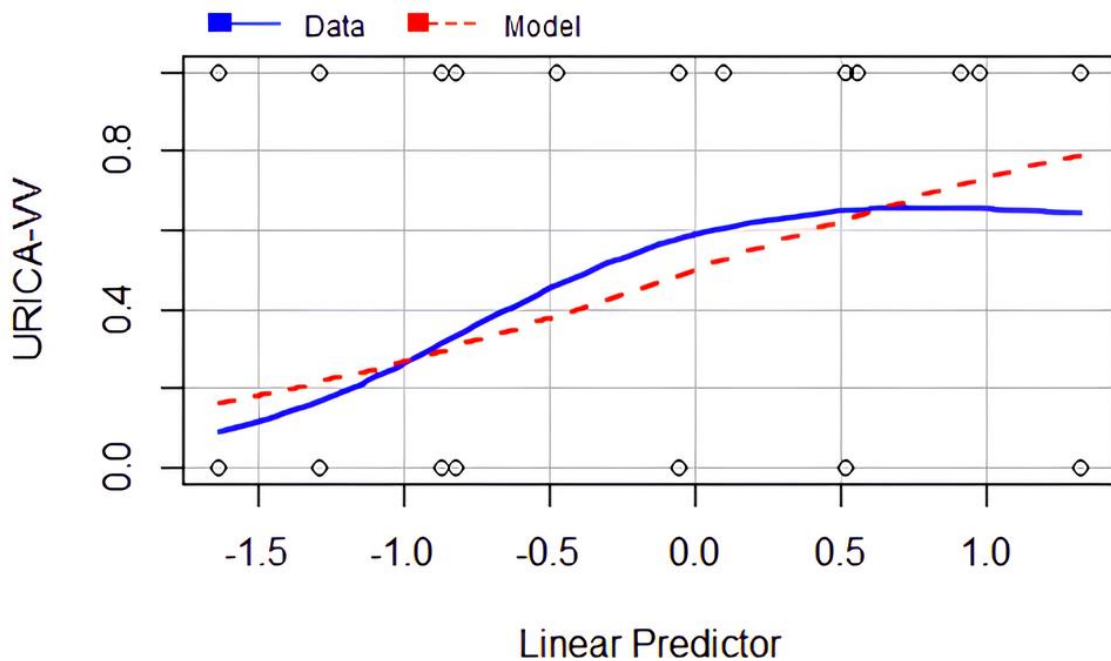
Na área da voz, o modelo de Floresta Aleatória já se mostrou capaz de identificar com alta precisão e sensibilidade diferentes distúrbios vocais, incluindo a detecção precoce de alterações acústicas associadas a doenças neurológicas e funcionais (Byeon, 2021). Esses achados evidenciam o potencial da aplicação do modelo preditivo de Floresta Aleatória, para auxiliar fonoaudiólogos na tomada de decisão e no direcionamento de condutas preventivas e terapêuticas mais assertivas.

Quanto às variáveis analisadas, a literatura demonstra que fatores sociodemográficos como sexo e idade influenciam a prontidão para mudança em contextos de saúde (Pedersen *et al.*, 2009; Bulc; Svab; Godycki-Cwirko, 2014; Rocha, 2020). Ademais, a exposição a fatores de riscos intensifica a gravidade dos sintomas vocais, (Martins *et al.*, 2024), o que pode impactar no avanço do estágio de prontidão para mudança, pois indivíduos com queixas possuem um maior comprometimento emocional e vocal (Costa *et al.*, 2013), e ao aumentar a percepção das limitações impostas pela condição, pode também funcionar como motivador para realizar as mudanças necessárias do comportamento vocal. Assim, destaca-se a necessidade de considerar essas particularidades dos indivíduos nas intervenções.

5.3 Modelos de AM - Dados dos Instrumentos de Autoavaliação

Foram desenvolvidos modelos de regressão logística, KNN, Naive Bayes, Árvore de Decisão e Floresta aleatória com variáveis extraídas dos instrumentos de autoavaliação IDV10-TRI, QVV-TRI, ESV-TRI e EDTV-TRI, as quais foram empregadas como preditoras na construção dos modelos. Inicialmente, foi ajustado um modelo de regressão logística. O gráfico do efeito marginal do modelo (Figura 12) permitiu avaliar se a relação entre as variáveis preditoras derivadas dos instrumentos e a variável resposta foi adequadamente modelada.

Figura 12 – Efeito Marginal do Modelo de Regressão Logística: Dados dos Instrumentos de Autoavaliação.

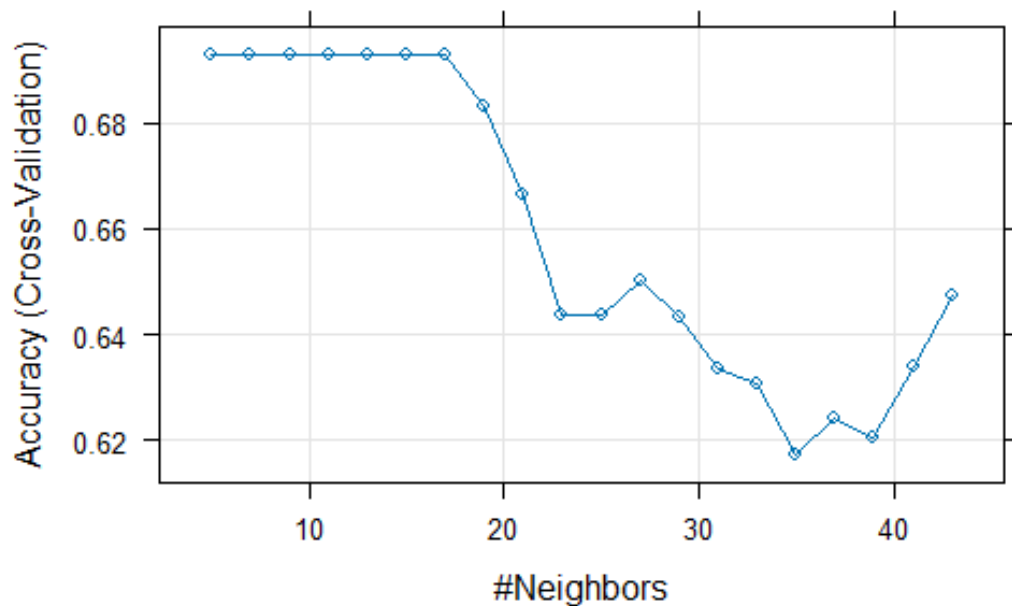


Fonte: Dados da Pesquisa, 2025.

Observa-se na Figura 12, que o modelo segue de forma geral o padrão esperado, com uma curva em S, típica da regressão logística. No entanto, há pequenas discrepâncias, pois o modelo tende a subestimar as probabilidades nos valores mais baixos do preditor linear e a superestimar nos valores mais altos. Ainda assim, o ajuste pode ser considerado razoável, indicando que o modelo representa adequadamente a relação entre as variáveis preditoras e a variável resposta. Além disso, a AUC foi de aproximadamente 0,46, isso sugere que o modelo tem um desempenho inferior ao acaso na distinção do estágio de prontidão, pois uma AUC de 0,5 indicaria desempenho equivalente ao acaso. O ponto de corte selecionado para este modelo foi de 0,195.

Em seguida foi criado o modelo KNN. Na Figura 13, no eixo horizontal (x), os diferentes valores de k testados, enquanto no eixo vertical (y), tem a acurácia correspondente a cada valor de k, obtida durante o processo de validação cruzada.

Figura 13 - Número de Vizinhos do KNN com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.

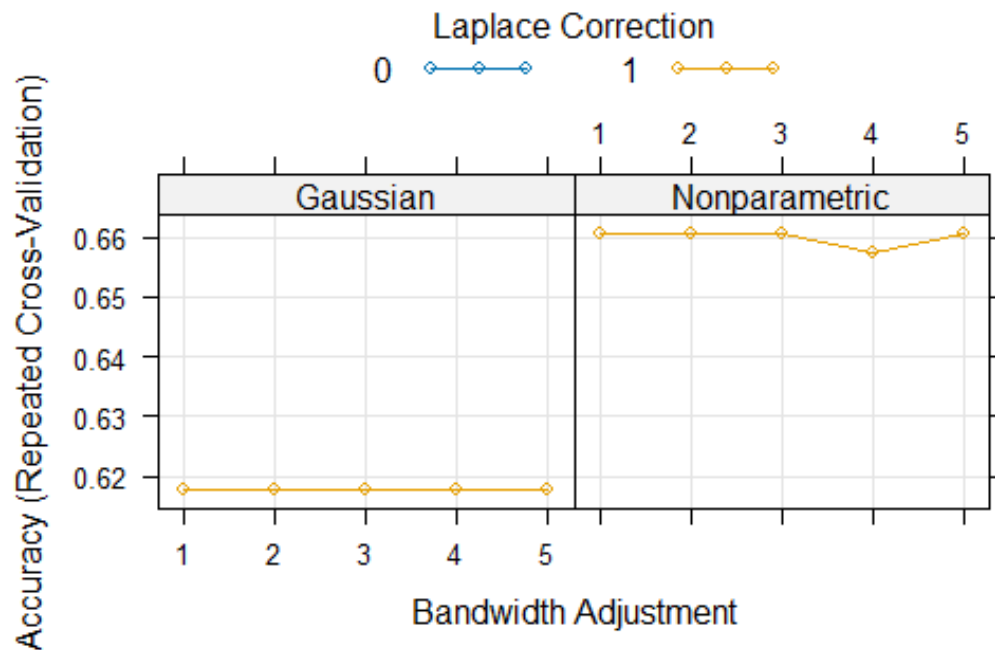


Fonte: Dados da Pesquisa, 2025.

A Figura 13 mostra que a acurácia variou com os diferentes valores de k , atingiu seu ponto máximo quando $k = 17$. Esse ponto é considerado o melhor valor de k , pois é onde o modelo obteve o melhor desempenho preditivo de acordo com a acurácia. Assim, entre os valores testados, foi considerado para o modelo os 17 vizinhos mais próximos. O modelo KNN apresentou AUC de 0,5831, indicando 58,31% de chances de prever corretamente, esse desempenho é levemente melhor do que o acaso na classificação do estágio de prontidão. Além disso, foi definido o ponto de corte de 0,239, aplicou-se esse ponto de corte para classificar os dados de teste.

No modelo Naive Bayes, também foi utilizada a validação cruzada com 10 repartições. Na Figura 14 a acurácia foi utilizada como critério para selecionar os hiperparâmetros fL , $usekernel$ e $adjust$. A definição adequada desses hiperparâmetros é essencial para melhorar o desempenho do modelo, pois busca equilibrar viés e variância para que consiga generalizar melhor com novos dados.

Figura 14 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.



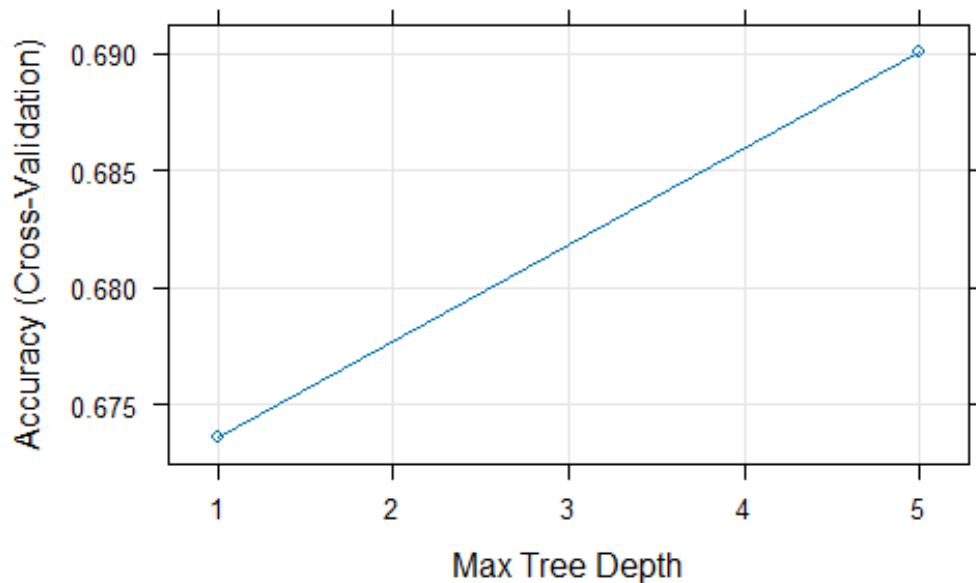
Fonte: Dados da Pesquisa, 2025.

Conforme observa-se na Figura 14, os parâmetros finais adotados para o modelo foram $fl = 0$, que controla o ajuste de Laplace; $usekernel = TRUE$, indica o uso de uma função kernel para suavizar as estimativas; e $adjust = 1$, que ajusta a largura do kernel. Essa configuração buscou melhorar o desempenho preditivo, como evidenciado pelo valor máximo de acurácia obtido durante o treinamento.

O modelo Naive Bayes apresentou AUC de 0,4508, indicando um desempenho inferior ao acaso na classificação do estágio de prontidão para a mudança. O ponto de corte selecionado foi de 0,126 e foi aplicado para classificar os dados de teste. Contudo, o baixo valor da AUC sugere que o modelo Naive Bayes pode não ser eficaz para prever novos dados.

A Figura 15 apresenta a acurácia do modelo de árvore de decisão em função da profundidade máxima ($maxdepth$), obtido a partir da acurácia média nas 10 repartições da validação. Por meio dessa visualização, é possível observar o comportamento do modelo conforme o aumento da complexidade, auxiliando na escolha do ponto ótimo de corte para o parâmetro profundidade máxima.

Figura 15 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia:
Dados dos Instrumentos de Autoavaliação.

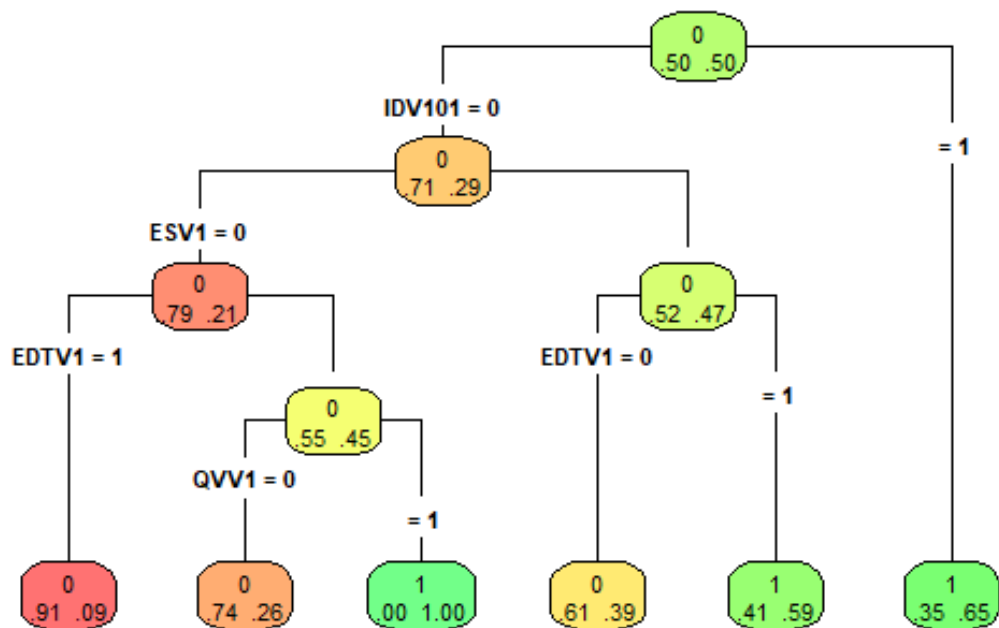


Fonte: Dados da Pesquisa, 2025.

Desta forma, o parâmetro final adotado para o modelo de árvore de decisão foi $\text{maxdepth} = 5$, que limita a profundidade máxima da árvore, em busca de um equilíbrio entre complexidade do modelo e capacidade preditiva, e de evitar o ajuste excessivo aos dados de treino, o que comprometeria o desempenho em novos dados.

Além da avaliação numérica do desempenho preditivo, o modelo de árvore de decisão permite a visualização gráfica de sua estrutura, o que contribui para a interpretação clínica dos resultados. A Figura 16 apresenta o diagrama gerado a partir do modelo final, com profundidade máxima de cinco níveis, conforme estabelecido na etapa de ajuste.

Figura 15 – Árvore de Decisão com Base nos Dados dos Instrumentos de Autoavaliação.

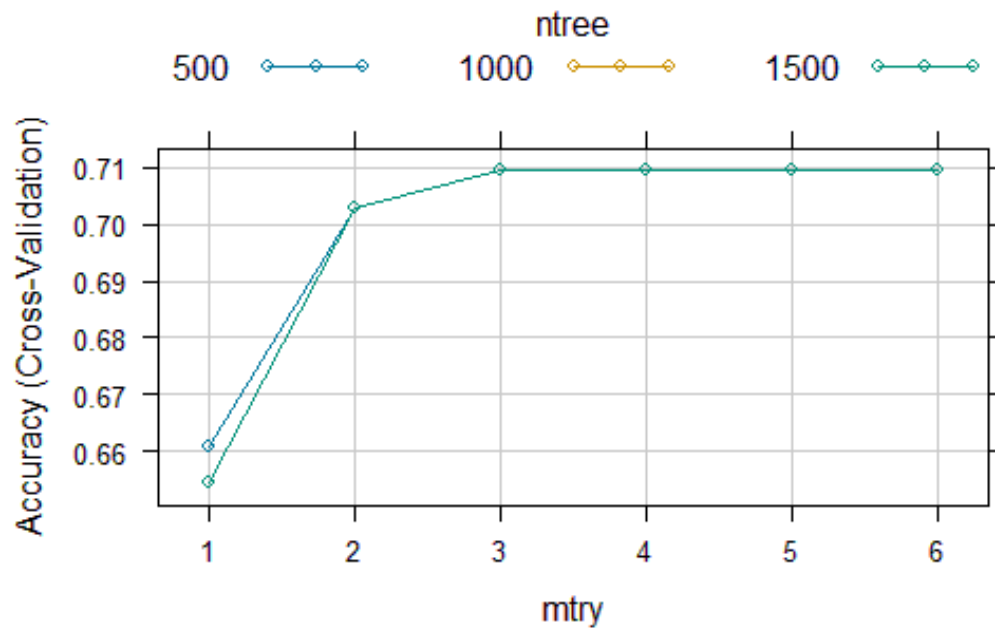


Fonte: Dados da Pesquisa, 2025.

Este diagrama de árvore de decisão mostra como diferentes variáveis são usadas para classificar observações. As variáveis mais relevantes para a divisão dos nós encontram-se nas camadas superiores da árvore, pois indica maior importância na definição do desfecho. Deste modo, a variável mais importante é o IDV10, que realiza a primeira divisão; quando alterado classifica como pertencentes ao estágio de manutenção, ou seja, sua alteração auxilia a prever que o indivíduo está nesse estágio de prontidão para a terapia de voz. Enquanto as outras variáveis como ESV, EDTV e QVV refinam a previsão. Assim, o gráfico ilustra claramente as regras de decisão do modelo. O modelo de Árvore de Decisão apresentou AUC de 0,5662, que indica um desempenho um pouco superior ao acaso na classificação do estágio de prontidão para a terapia de voz. O ponto de corte utilizado para fazer as previsões com o conjunto teste foi de 0,174.

No modelo de Floresta Aleatória, a acurácia foi utilizada como critério para selecionar o modelo ótimo, como pode-se observar na Figura 16, para o modelo foi escolhido os parâmetros que apresentaram o maior valor dessa métrica. Os parâmetros finais adotados foram $mtry = 3$, que determina o número de variáveis consideradas em cada divisão da árvore, e $ntree = 500$, que define a quantidade de árvores construídas no conjunto.

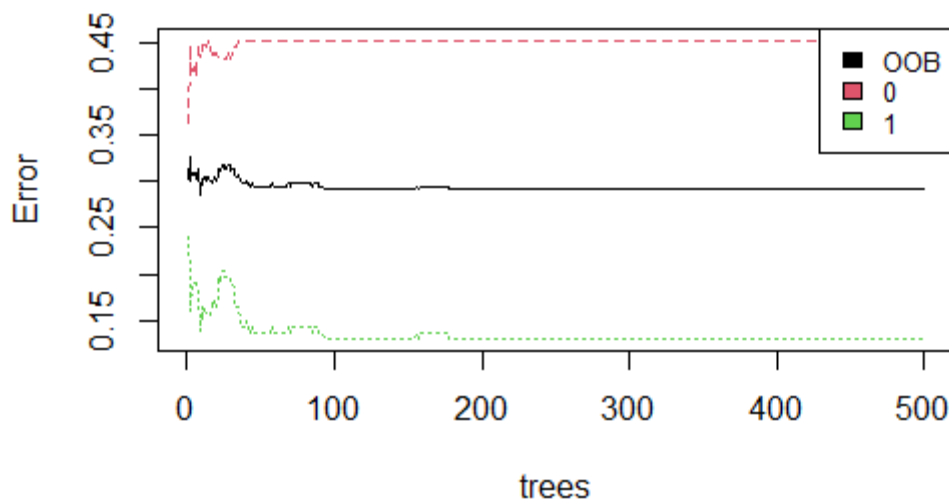
Figura 16 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados dos Instrumentos de Autoavaliação.



Fonte: Dados da Pesquisa, 2025.

A validação interna do modelo de Floresta Aleatória foi realizada por meio da estimativa do erro Out-Of-Bag (OOB), que utiliza apenas os dados não empregados na construção de cada árvore para avaliar o desempenho preditivo. Nesse caso, o erro OOB foi de aproximadamente 29,08%, como pode ser observado na Figura 17, isso indica que o modelo errou 29,08% das vezes ao classificar essas observações, mesmo sem ter sido exposto a elas durante o treinamento.

Figura 17 – Erro OOB no Modelo Floresta Aleatória: Dados dos Instrumentos de Autoavaliação.



Fonte: Dados da Pesquisa, 2025.

A matriz de confusão a seguir mostra como o modelo classificou as amostras OOB, separadas por estágio:

Tabela 7 - Matriz de Confusão OOB: Dados dos Instrumentos de Autoavaliação.

	Contemplação	Manutenção	Erro da classe
Contemplação	84	20	0.451 (45,1%)
Manutenção	69	133	0.131 (13,1%)

Fonte: Dados da Pesquisa, 2025.

Constata-se que o modelo teve desempenho significativamente melhor para classificar o estágio de manutenção, enquanto quase metade dos que estavam no estágio contemplação foram classificados de forma equivocada. Ademais, o valor da AUC obtido foi de 0,5862, indicando um desempenho moderado na classificação do estágio de prontidão para a terapia de voz. Foi identificado o ponto de corte de 0,858. Com esse ponto de corte, foi realizada a classificação com o conjunto de teste, que permite a construção da matriz de confusão a seguir:

Tabela 8 – Matriz de Confusão da Floresta Aleatória: Dados dos Instrumentos de Autoavaliação.

	Contemplação	Manutenção
Contemplação	2	12
Manutenção	3	53

Fonte: Dados da Pesquisa, 2025.

Na matriz de confusão com os dados de teste (Tabela 8), o modelo apresentou desempenho excelente na classificação de indivíduos no estágio de manutenção. No entanto, teve dificuldade em identificar corretamente os casos do estágio de contemplação, semelhante aos resultados achados na validação interna. Esse desequilíbrio de desempenho pode indicar que mesmo com oversampling a presença de desbalanceamento ainda interferiu nas predições ou ainda que as variáveis preditoras utilizadas não foram suficientemente discriminativas.

Tabela 9 - Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados dos Instrumentos de Autoavaliação.

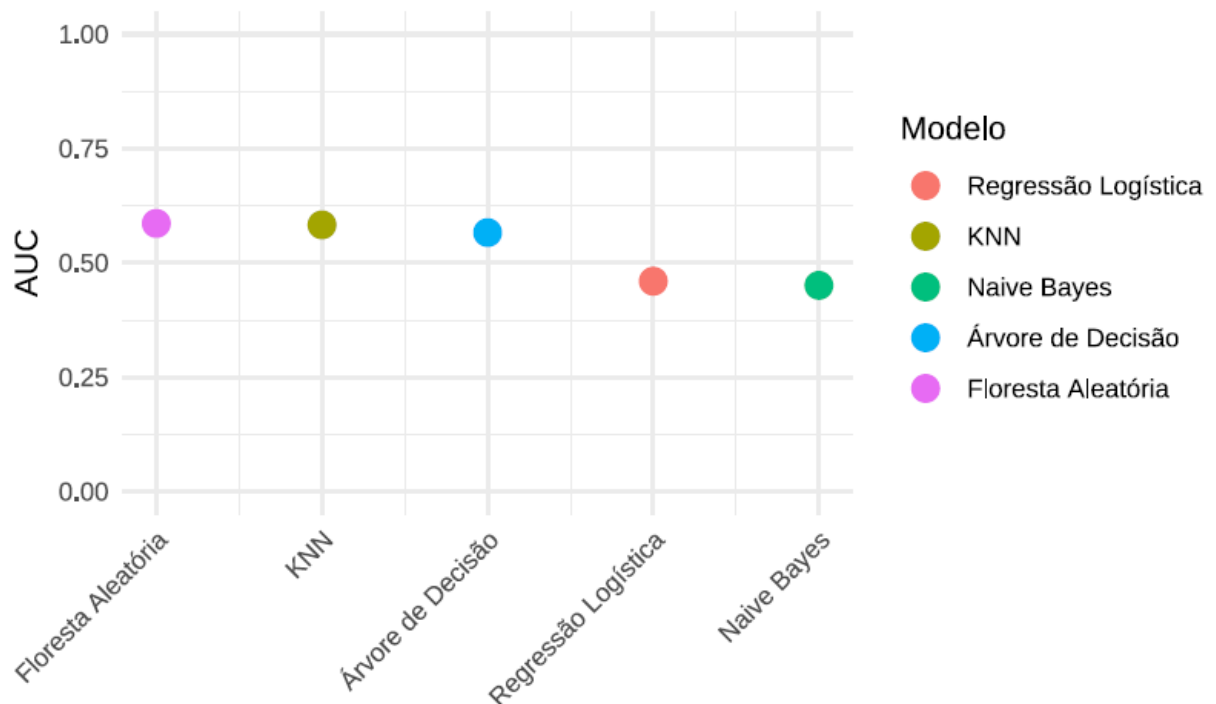
	Regressão Logística	KNN	Naive Bayes	Árvore De Decisão	Floresta Aleatória
Acurácia	0,9143	0,9143	0,9143	0,9143	0,7857
Acurácia Balanceada	0,5846	0,5846	0,5846	0,5846	0,6077
Kappa	0,2075	0,2075	0,2075	0,2075	0,1176
Sensibilidade	0,9692	0,9692	0,9692	0,9692	0,8154
Especificidade	0,2000	0,2000	0,2000	0,2000	0,4000
Valor Preditivo Positivo	0,9403	0,9403	0,9403	0,9403	0,9464
Valor Preditivo Negativo	0,3333	0,3333	0,3333	0,3333	0,1429
Taxa de Detecção	0,9000	0,9000	0,9000	0,9000	0,8000
Prevalência Estimada	0,9571	0,9571	0,9571	0,9571	0,6077

		Regressão Logística	KNN	Naive Bayes	Árvore De Decisão	Floresta Aleatória
Taxa de Informação Nula (NIR)		0,9286	0,9286	0,9286	0,9286	0,9286
p-valor [Acurácia > NIR]		0,7678	0,7678	0,7678	0,7678	0,9997
p-valor do Teste de McNemar		0,6831	0,6831	0,6831	0,6831	0,0389

Fonte: Dados da Pesquisa, 2025.

Em relação à capacidade dos modelos com variáveis dos instrumentos de autoavaliação em identificar corretamente os indivíduos no estágio de contemplação, observa-se que a especificidade foi limitada. Conforme apresentado na Tabela 9, a maioria dos modelos apresentou especificidade próxima a 2.000, indicando baixa performance na correta classificação dos casos deste estágio. Apenas o modelo de Floresta Aleatória alcançou desempenho superior, com especificidade de 4.000, evidenciando maior sensibilidade à detecção da classe menos frequente. Esses achados sugerem que, de forma geral, os modelos não foram eficazes em prever o estágio de prontidão contemplação com base em instrumentos de autorrelato, com a Floresta Aleatória se destacando como o algoritmo de melhor desempenho relativo, ainda que com limitações.

Figura 18 - Valores de AUC dos Modelos: Dados dos Instrumentos de Autoavaliação.



Fonte: Dados da Pesquisa, 2025.

Os valores de AUC obtidos para os diferentes modelos indicam variações no desempenho preditivo na classificação do estágio de prontidão. O modelo de Floresta Aleatória apresentou o maior valor de AUC (0,5862), seguido de perto pelo KNN (0,5831) e pela Árvore de Decisão (0,5662), sugerindo desempenho ligeiramente superior ao acaso, mas ainda modesto. A Regressão Logística (0,4599) e o Naive Bayes (0,4508) apresentaram AUC inferiores a 0,5, o que indica baixa capacidade discriminatória.

Esses resultados sugerem que, apesar de alguns algoritmos apresentarem desempenho um pouco melhor, nenhum modelo atingiu níveis elevados de predição, apontando para a complexidade da tarefa e possíveis limitações nas variáveis utilizadas para prever o estágio de prontidão. Adicionalmente, o desbalanceamento dos dados, com prevalência maior do estágio manutenção, também favorece a tendência dos modelos em classificar erroneamente os casos menos frequentes, reduzindo a especificidade.

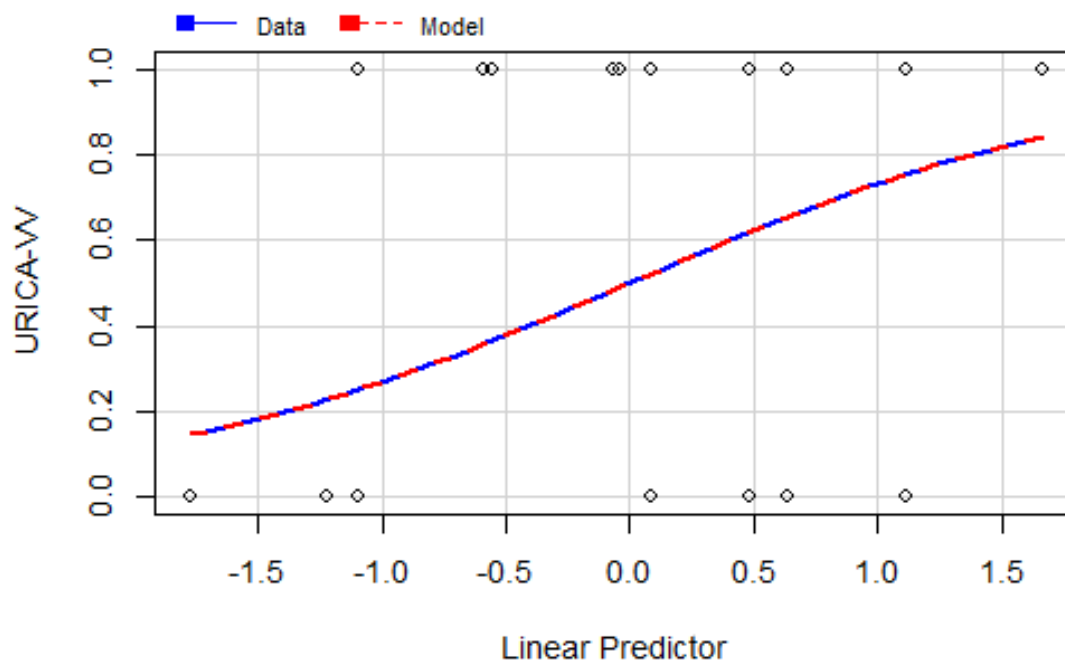
Contudo, o estágio de contemplação caracteriza-se pela ambivalência do indivíduo, com consciência do problema, porém sem decisão definitiva para a mudança, o que pode torna sua identificação um desafio para modelos preditivos que

dependem de padrões claros entre classes, especialmente na ausência de dados comportamentais complementares. Apesar da Floresta Aleatória apresentar desempenho relativo superior, ela ainda não alcançou níveis satisfatórios de especificidade, demonstrando as limitações impostas pela qualidade dos dados. Assim, recomenda-se a inclusão de variáveis comportamentais, psicossociais e dados clínicos objetivos para aprimorar a predição.

5.4 Modelos de AM – Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia

Foram desenvolvidos modelos com as variáveis clínicas relativas à presença de alteração do JPA, alteração laríngea e classificação da disfonia. O primeiro modelo com esses dados, foi a regressão logística. Na Figura 19, observa-se o gráfico do efeito marginal deste modelo, a sobreposição quase perfeita das duas linhas no gráfico indica que o modelo se ajusta muito bem aos dados, mostrando um bom poder preditivo para a variável URICA-VV.

Figura 19 - Efeito Marginal do Modelo de Regressão Logística: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.



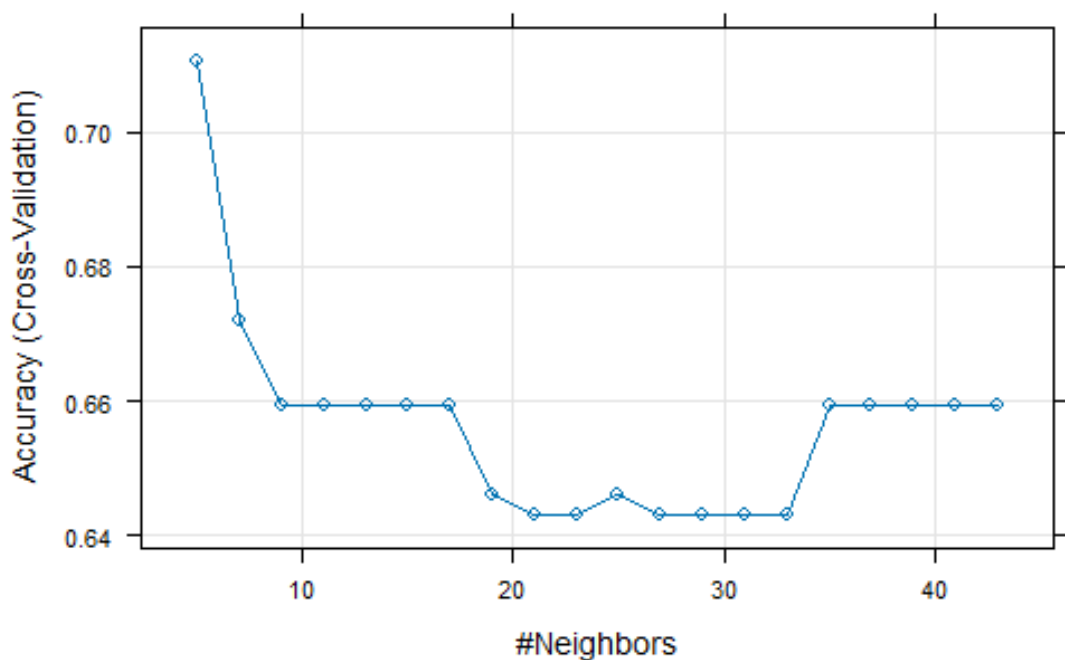
Fonte: Dados da Pesquisa, 2025.

No modelo de regressão logística, o valor da AUC obtido foi de 0,6815, indicando uma capacidade moderada de discriminação entre as classes. Com o

critério de Youden, foi identificado o ponto de corte de 0,486. Com esse limiar, foi gerada uma nova classificação para verificar o impacto do ajuste no desempenho do modelo.

Após isso, foi desenvolvido um modelo KNN, através da validação cruzada foi possível identificar o melhor valor de K, a acurácia foi utilizada como critério para a seleção do mais adequado, conforme observado na Figura 20. Entre as diferentes configurações testadas, o modelo com melhor acurácia foi aquele com $k=5$, valor adotado para construção do modelo. Esse resultado indica que, ao considerar os cinco vizinhos mais próximos para a tomada de decisão, o modelo obteve a melhor taxa de acerto na classificação dos dados.

Figura 20 - Número de Vizinhos do KNN com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.



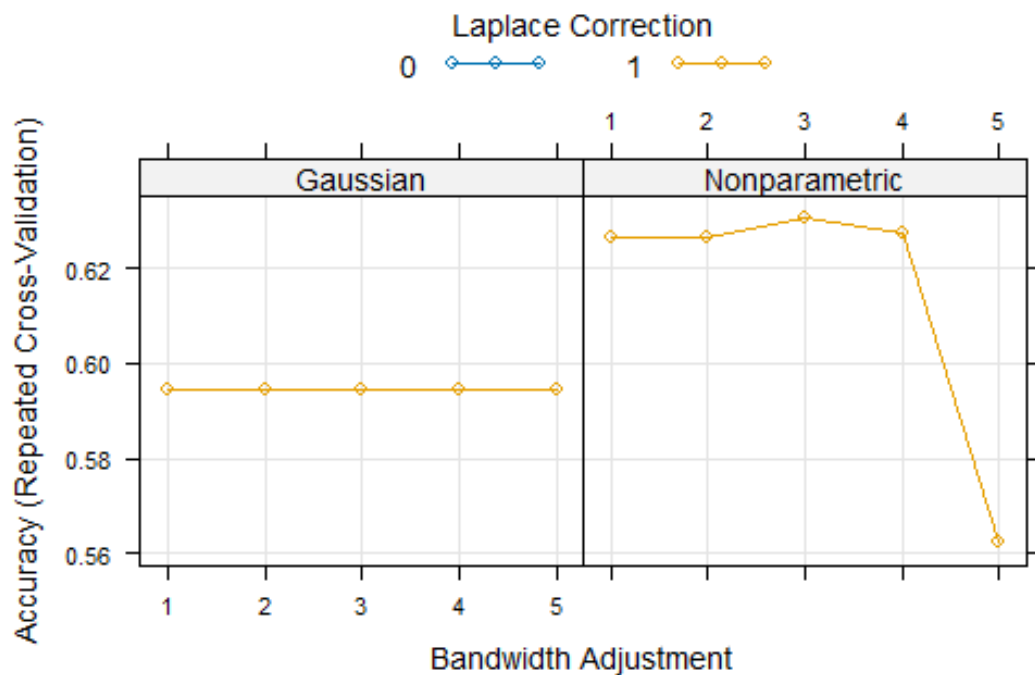
Fonte: Dados da Pesquisa, 2025.

Além disso, no KNN o valor da AUC obtido foi de 0,4923, indicando um desempenho inferior ao esperado, com capacidade de discriminação próxima ao acaso. O ponto de corte utilizado para as novas predições foi de 0,307, que proporcionou a melhor combinação entre sensibilidade e especificidade.

Para o ajuste do modelo Naive Bayes, também foi realizada uma validação cruzada com o objetivo de identificar a combinação de hiperparâmetros que proporcionasse o melhor desempenho na classificação. Na Figura 21, é possível

observar que a melhor acurácia foi alcançada com a combinação dos parâmetros $fL = 0$, $usekernel = TRUE$ e $adjust = 3$. Esses valores indicam que o modelo utilizou estimativa de densidade com kernel para a função de probabilidade, sem aplicação de suavização adicional via parâmetro fL , e com um fator de ajuste ampliado na largura de banda do kernel, para melhorar o desempenho na classificação.

Figura 21 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laringeo e Disfonia.

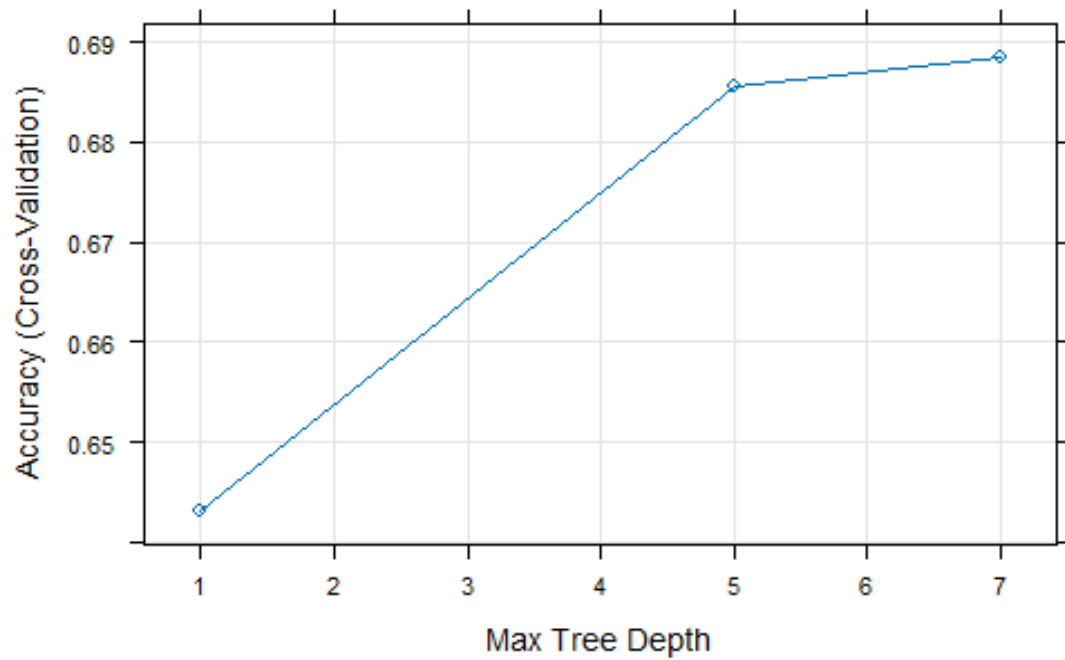


Fonte: Dados da Pesquisa, 2025.

No modelo Naive Bayes, também foi realizada a análise da AUC com o valor de 0,6231, que indica uma capacidade moderada de discriminar o estágio de prontidão. A partir da curva, identificou o ponto de corte de 0,556 que foi utilizado para gerar novas predições com os dados de teste.

Enquanto, na Árvore de Decisão, foi identificado que o modelo com melhor desempenho apresentava como parâmetro o valor de $maxdepth = 7$. Isso indica que a profundidade máxima da árvore de decisão foi limitada a sete níveis. A Figura 22 apresenta a variação da acurácia em função do parâmetro profundidade máxima, e evidencia o ponto de melhor performance adotado como modelo final.

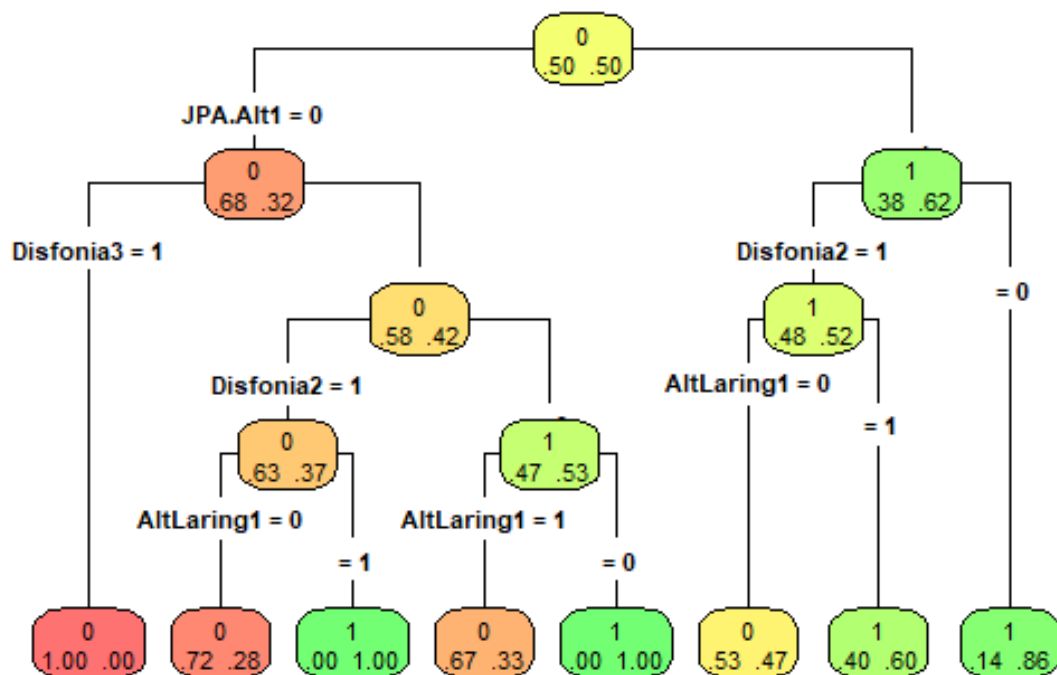
Figura 22 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia:
Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.



Fonte: Dados da Pesquisa, 2025.

A árvore (Figura 23) mostra que a alteração do JPA foi a variável mais relevante, aparecendo no topo da estrutura, as outras variáveis como alteração laríngeo e classificação da disfonia entram na sequência para realizar a previsão, e resulta em combinações com diferentes percentuais da classificação final.

Figura 23 – Árvore de Decisão de Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.

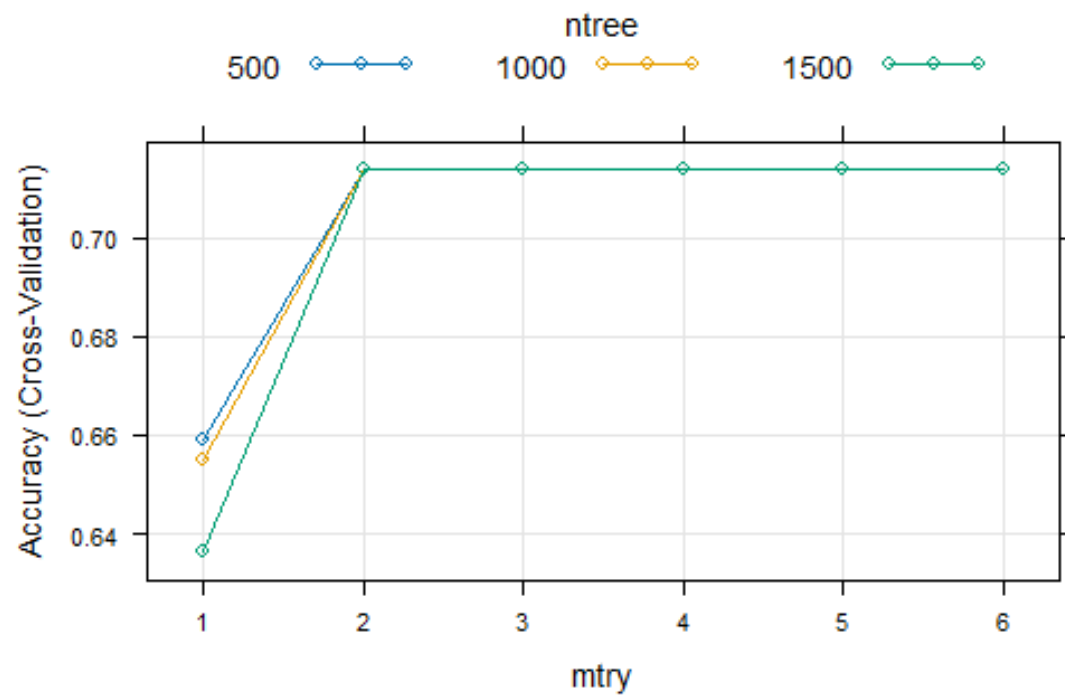


Fonte: Dados da Pesquisa, 2025.

Quanto a curva ROC, foi utilizada para avaliar o desempenho do modelo de árvore de decisão, resultando em uma AUC de 0,4646, o que indica desempenho abaixo do esperado. Foi identificado o ponto de corte ideal de 0,307, que foi aplicado para realizar as predições no conjunto teste.

No modelo de Floresta Aleatória, foi analisada a acurácia na validação cruzada em função dos parâmetros mtry e ntree. O modelo final selecionado utilizou mtry = 2, então duas variáveis foram consideradas em cada divisão, e ntree = 500, ou seja, a floresta foi composta por 500 árvores. Na Figura 24, mostra como a acurácia variou com diferentes valores de mtry e ntree.

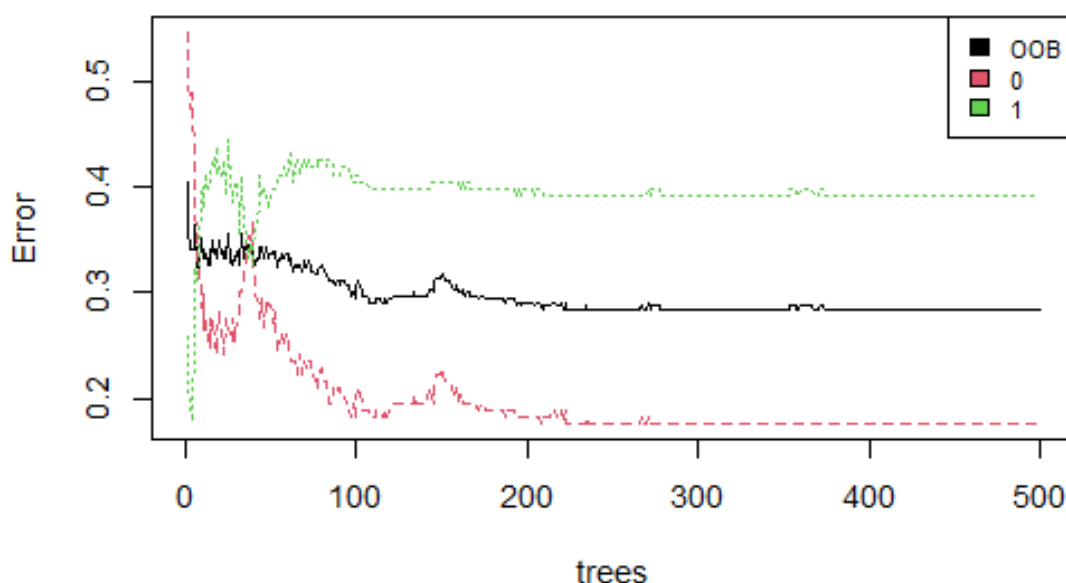
Figura 24 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.



Fonte: Dados da Pesquisa, 2025.

Na floresta aleatória, a estimativa de erro OOB foi de 28,76%, indicando que aproximadamente um quarto das observações foram classificadas incorretamente pelo modelo durante a validação interna, como constatado na Figura 25.

Figura 25 – Erro OOB do Modelo Floresta Aleatória: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.



Fonte: Dados da Pesquisa, 2025.

A matriz de confusão, gerada com as amostras OOB, mostrou que o modelo teve maior dificuldade em classificar corretamente os casos de manutenção, com uma taxa de erro de aproximadamente 39,86%, enquanto contemplação apresentou desempenho superior, com erro de 17,64%. A matriz de confusão a seguir mostra como o modelo classificou as amostras OOB:

Tabela 10 – Matriz de Confusão OOB: dados clínicos do JPA, Diagnóstico Laríngeo e Disfonia.

	Contemplação	Manutenção	Erro da classe
Contemplação	126	27	0.1764 (17,64%)
Manutenção	61	92	0.3986 (39,86%)

Fonte: Dados da Pesquisa, 2025.

Esses resultados sugerem que teve melhor desempenho na identificação do estágio contemplação. Além disso, a AUC obtida foi de 0,6292, o que representa um desempenho moderado na classificação do estágio de prontidão para a terapia de

voz. Identificou-se o ponto de corte ideal em 0,867. Com base nesse limiar, foi realizada a classificação dos dados do conjunto de teste, resultando na matriz de confusão apresentada a seguir:

Tabela 11 – Matriz de Confusão da Floresta Aleatória: Dados Clínicos do JPA, Diagnóstico Laríngeo e Disfonia.

	Contemplação	Manutenção
Contemplação	5	44
Manutenção	0	21

Fonte: Dados da Pesquisa, 2025.

Os resultados com o conjunto de teste também mostraram melhor predição do estágio contemplação, assim como as predições obtidas com as amostras OOB. Embora o desempenho geral do modelo não tenha sido excelente, a validação externa confirmou os achados da validação interna, o que reforça a consistência dos resultados obtidos.

Tabela 12 – Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados Clínicos do JPA, Diagnóstico Laríngeo e Classificação da Disfonia para classificação.

	Regressão Logística	KNN	Naive Bayes	Árvore de Decisão	Floresta Aleatória
Acurácia	0,7571	0,8571	0,7000	0,8857	0,3714
Acurácia Balanceada	0,6846	0,5538	0,6538	0,5692	0,6615
Kappa	0,1678	0,0909	0,1198	0,1385	0,0638
Sensibilidade	0,7692	0,9077	0,7077	0,9385	0,3231
Especificidade	0,6000	0,2000	0,6000	0,2000	1,0000
Valor Preditivo Positivo	0,9615	0,9365	0,9583	0,9385	1,0000
Valor Preditivo Negativo	0,1667	0,1429	0,1364	0,2000	0,1020
Prevalência	0,9286	0,9286	0,9286	0,9286	0,9286
Taxa de Detecção	0,7143	0,8429	0,6571	0,8714	0,3000
Prevalência Estimada	0,7429	0,9000	0,6857	0,9286	0,3000

	Regressão Logística	KNN	Naive Bayes	Árvore de Decisão	Floresta Aleatória
Taxa de Informação Nula	0,9286	0,9286	0,9286	0,9286	0,9286
P-Valor [Acurácia > Taxa de Inf. Nula]	0,999999	0,9894	1,000000	0,9391	1,0000
P-Valor Teste de McNemar	0,003609	0,7518	0,0004803	1,0000	9,022e-11

Fonte: Dados da Pesquisa, 2025.

Os resultados indicam que a Árvore de Decisão apresentou um bom desempenho, com a maior acurácia (0,8857) e sensibilidade de 0,9385, o que demonstra ótima capacidade para detectar o estágio manutenção, embora tenha baixa especificidade de 0,2000, indicando dificuldade em reconhecer o estágio contemplação. O KNN apresentou resultados semelhantes em acurácia (0,8571), sensibilidade (0,9077) e especificidade (0,2000).

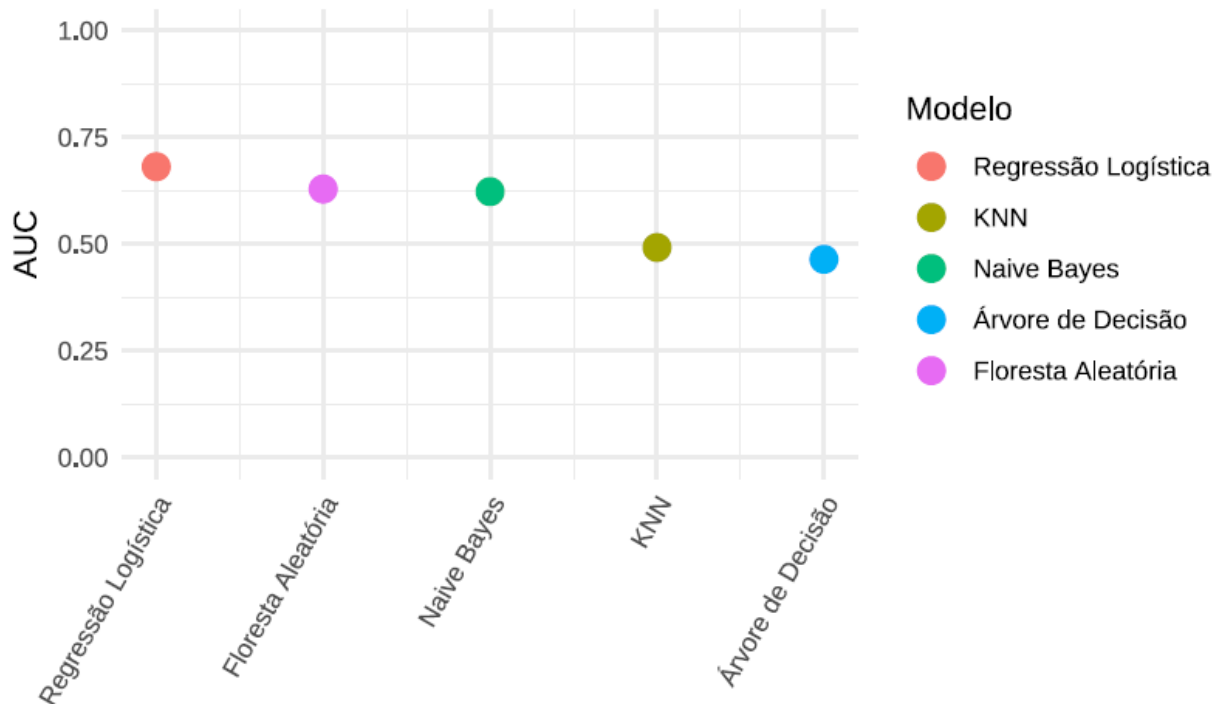
A Regressão Logística e o Naive Bayes mostraram desempenho mais equilibrado entre os estágios, com acurácia balanceada superior, sendo 0,6846 para a Regressão Logística e 0,6538 para o Naive Bayes, além de especificidade razoável de 0,6000 e sensibilidade de 0,7692 para regressão e 0,7077 para naive bayes. Por outro lado, a Floresta Aleatória apresentou especificidade perfeita (1,0000) e alto valor preditivo positivo, o que indica excelente capacidade para identificar corretamente o estágio contemplação. Contudo, teve baixa sensibilidade (0,3231) e acurácia baixa (0,3714), demonstrando dificuldade significativa para detectar o estágio manutenção.

Além disso, os valores de valor preditivo positivo são elevados em todos os modelos, especialmente na Floresta Aleatória, o que indica que quando o modelo prevê o estágio de manutenção, a chance de estar correto é alta. Porém, os valores de valor preditivo negativo são baixos, apontando dificuldade em afirmar corretamente os casos no estágio de contemplação.

Em suma, a Regressão Logística destaca-se pelo melhor equilíbrio entre sensibilidade e especificidade, acurácia balanceada e valores estatísticos, tornando-a mais adequada para predição nesse conjunto de dados. O KNN a Árvore de Decisão, apesar de maior acurácia, sofre com baixa especificidade e acurácia

balanceada, comprometendo a identificação da classe minoritária. O Naive Bayes também é um candidato viável, apresentando desempenho intermediário e equilíbrio razoável. Já a Floresta Aleatória, apesar da especificidade perfeita, apresenta baixa sensibilidade e acurácia inferior, sugerindo que não é o modelo ideal.

Figura 26 – Valores de AUC dos Modelos: Dados clínicos JPA, Diagnóstico Laríngeo e Classificação da Disfonia.



Fonte: Dados da Pesquisa, 2025.

Os valores de AUC obtidos mostram que a Regressão Logística apresentou o melhor desempenho (0,6815), seguida pela Floresta Aleatória (0,6292) e pelo Naive Bayes (0,6231), indicando moderada capacidade de discriminação entre os estágios de prontidão. Por outro lado, o KNN (0,4923) e a Árvore de Decisão (0,4646) tiveram desempenho inferior a 0,5, sugerindo baixa capacidade discriminatória e resultados próximos ao acaso. Esses achados indicam que, embora alguns modelos tenham alcançado desempenho moderado (Regressão Logística e Naive Bayes), a tarefa de prever o estágio de prontidão continua desafiadora, possivelmente devido à complexidade do constructo e à necessidade de variáveis adicionais que capturem melhor as diferenças entre os grupos.

Em outros estudos que verificaram o tipo de disfonia, em casos de disfonia comportamental, observou-se predominância do estágio de contemplação, com

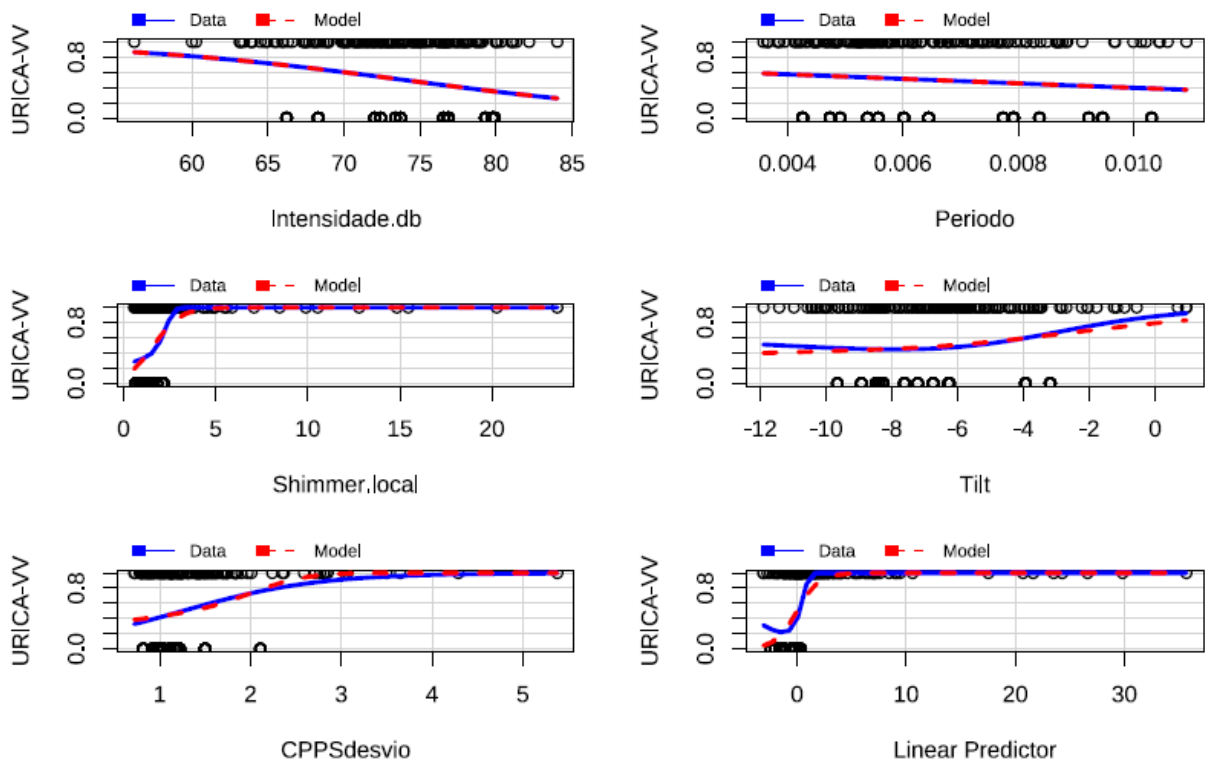
mudanças limitadas ao longo da terapia (Costa *et al.*, 2017), enquanto em disfonias orgânicas, estratégias educativas, demonstraram potencial para aumentar a prontidão e favorecer mudanças comportamentais relacionadas à voz (Kopf, Graetzer & Hun, 2015). Ademais, quando o impacto emocional da disfonia é elevado, maior é a adesão do paciente a terapia de voz, evidenciando a importância de predizer o estágio de prontidão para orientar intervenções motivacionais (Góes *et al.*, 2016).

Assim, o estado psicológico do paciente e o ambiente de tratamento exerce um papel fundamental na prontidão para a mudança (Krebs *et al.*, 2018), destaca-se a necessidade de estratégias terapêuticas que integrem tanto a motivação individual quanto o suporte externo. Esses fatores psicossociais podem influenciar o estágio de prontidão para a terapia de voz. Isso sugere que, embora as variáveis clínicas sejam importantes para o diagnóstico e o manejo da disfonia, é necessário reconhecer a influência de fatores psicológicos no estágio de prontidão.

5.5 Modelos de AM – Dados das Medidas Acústicas

Foram criados modelos também com as medidas acústica. A partir dos gráficos de efeitos marginais (Figura 27) gerados com base no modelo de regressão logística, foi possível observar a relação individual entre variáveis acústicas da voz e a variável resposta. Esses gráficos por permitirem visualizar o efeito isolado de cada preditor e manterem os demais constantes no modelo ajustado, tornam-se uma ferramenta útil para interpretação clínica e estatística dos modelos preditivos.

Figura 27 – Efeito Marginal do Modelo de Regressão Logística: Medidas Acústicas.



Fonte: Dados da Pesquisa, 2025.

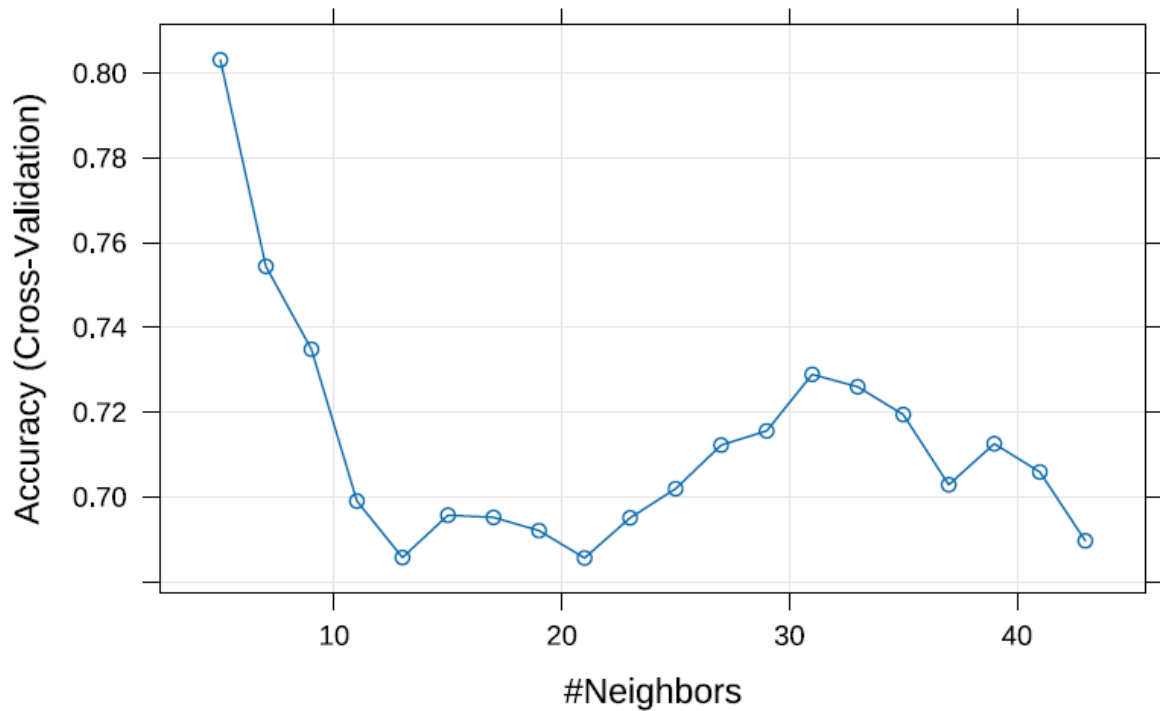
Na Figura 27 mostra que variáveis acústicas apresentam relações importantes com a prontidão para mudança vocal no modelo de Regressão logística. A Intensidade e o Período apresentaram relação inversa com a variável resposta, sugerindo que vozes mais intensas e com frequência fundamental mais alta estariam associadas ao estágio de contemplação. Isso pode indicar que indivíduos com vozes mais estáveis ou menos comprometidas percebem menor necessidade de modificar seu comportamento vocal, permanecendo em estágios iniciais, como a contemplação.

Entretanto, variáveis como Shimmer, Tilt e CPPS desvio padrão mostraram associação positiva com a prontidão, sugerindo que maior instabilidade ou distorção vocal está relacionada a maior prontidão para a mudança (manutenção), possivelmente devido à percepção mais clara do problema vocal. A AUC obtida foi de 0,6308, indicando que o modelo tem uma capacidade moderada de discriminar. O valor de corte utilizado foi 0,949 para realizar as predições.

No modelo KNN, a escolha do valor de k foi realizada por meio de validação cruzada, utilizando a acurácia como critério de desempenho. A Figura 28 apresenta a

variação da acurácia em diferentes valores de k, assim permite identificar o ponto ideal.

Figura 28 - Número de Vizinhos do KNN com Base na Acurácia: Medidas Acústicas



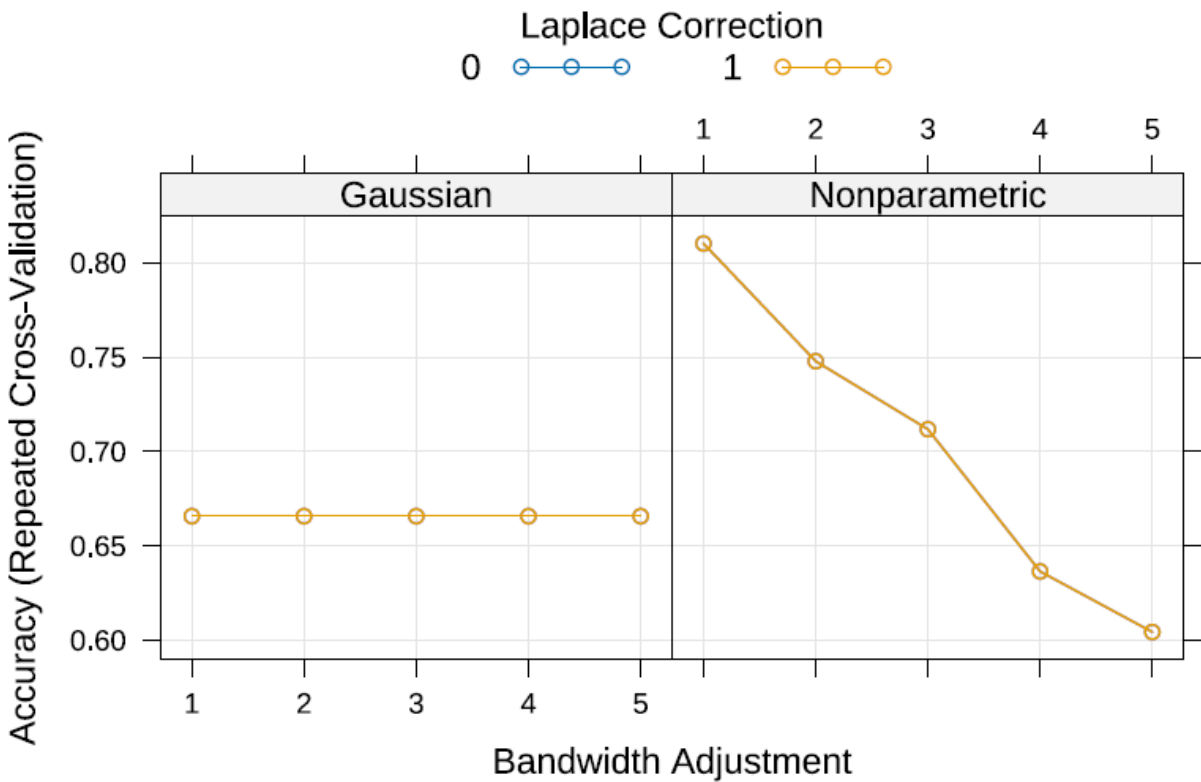
Fonte: Dados da Pesquisa, 2025.

Na Figura 28 demonstra que a maior acurácia (aproximadamente 0.80) é obtida com um baixo número de vizinhos, de $k=5$, assim sugere que modelos baseados nos 5 vizinhos mais próximos oferecem o melhor desempenho preditivo. A partir desse ponto, a acurácia cai significativamente, indicando que valores maiores de k podem introduzir ruído e reduzir a capacidade do modelo de predizer de forma eficaz. No KNN, a AUC foi de 0,5662, sugerindo uma capacidade preditiva moderada, levemente acima do acaso. Utilizou-se o ponto de corte de 0,069, sendo adotado como limiar para as previsões.

O modelo final de Naive Bayes foi selecionado com base na maior acurácia obtida durante o processo de validação. Os hiperparâmetros escolhidos foram: $fL = 0$, $usekernel = TRUE$ e $adjust = 1$ (Figura 29). Isso significa que o modelo utilizou estimativas de densidade kernel (em vez de assumir distribuição normal para as variáveis) com ajuste padrão, e sem correção de suavização ($fL = 0$). Esses parâmetros indicam que o modelo buscou capturar melhor a distribuição real dos

dados sem adicionar suavização artificial, o que pode melhorar o desempenho em conjuntos de dados com características não paramétricas.

Figura 29 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Medidas Acústicas.

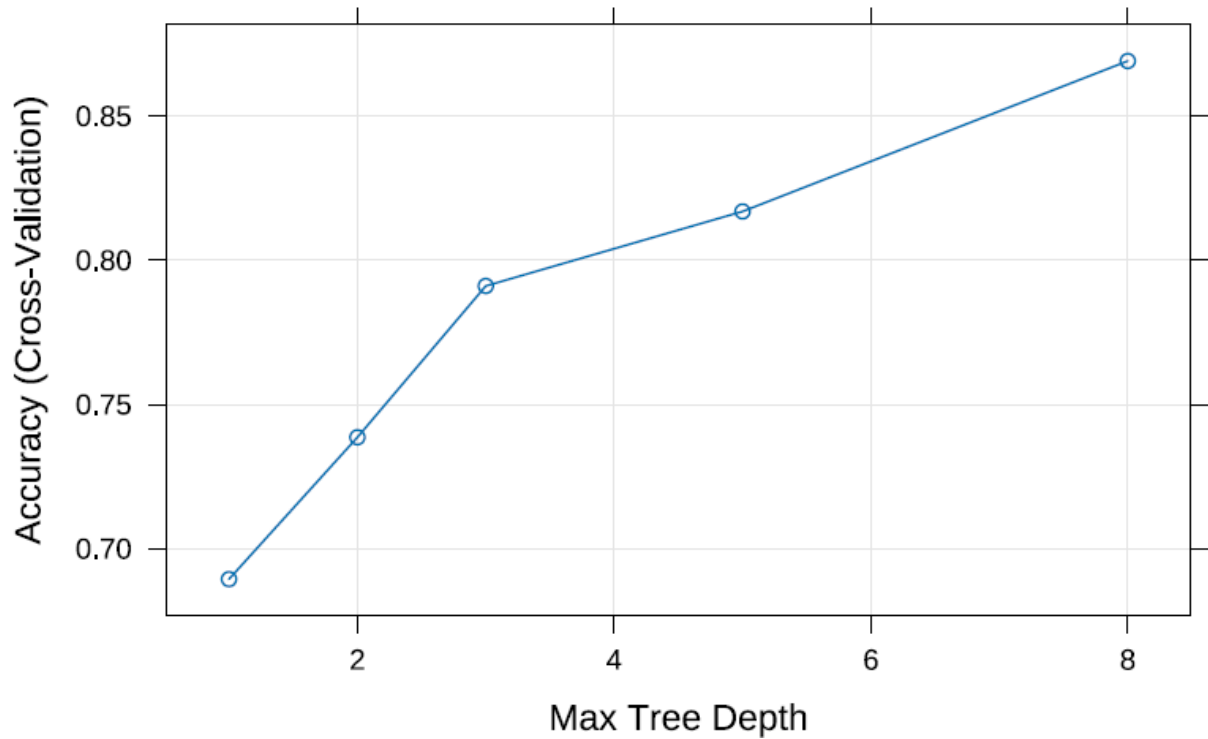


Fonte: Dados da Pesquisa, 2025.

Quando a AUC do modelo KNN, foi de 0,5262, esse valor indica um desempenho próximo ao aleatório na tarefa de discriminar entre os estágios da variável resposta. A partir da análise da curva, foi identificado o ponto de corte ótimo em 0,319.

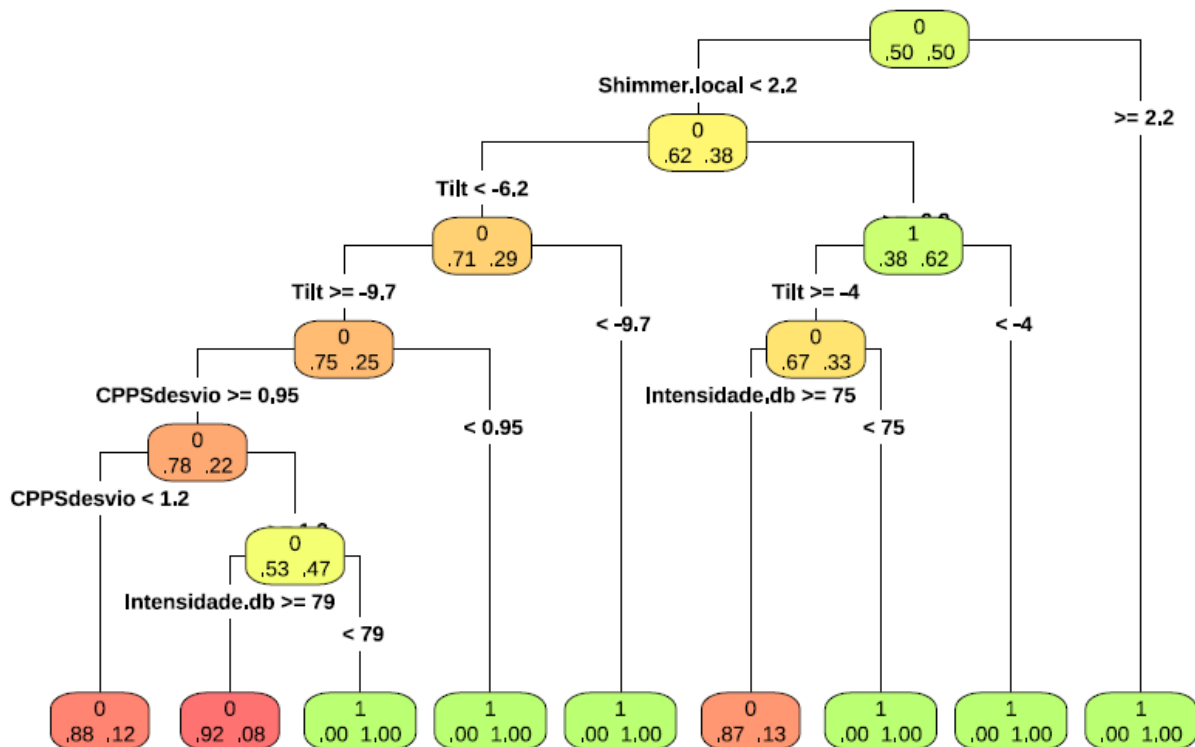
Em relação ao ajuste da Árvore de Decisão, o modelo final adotado foi com profundidade máxima igual a 8 (maxdepth = 8), pois apresentava a melhor acurácia, como apresenta a Figura 30, onde valores menores indicavam também acurácia menor na validação cruzada.

Figura 30 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia: Medidas Acústicas.



Fonte: Dados da Pesquisa, 2025.

A Figura 31 apresenta a árvore de decisão construída para classificar o estágio de prontidão para a terapia de voz. A variável Shimmer local ocupa a divisão inicial da árvore, o que demonstra sua maior relevância para separar os estágios contemplação e manutenção. Em seguida, outras variáveis vocais como Tilt, Intensidade e CPPS desvio padrão aparecem em níveis subsequentes, com diferentes pontos de corte, para complementam a decisão ao longo da árvore. A variável Período não consta em nenhuma divisão, isso sugere que neste modelo específico, ela não apresentou capacidade discriminativa suficiente para diferenciar quanto à prontidão para a mudança.

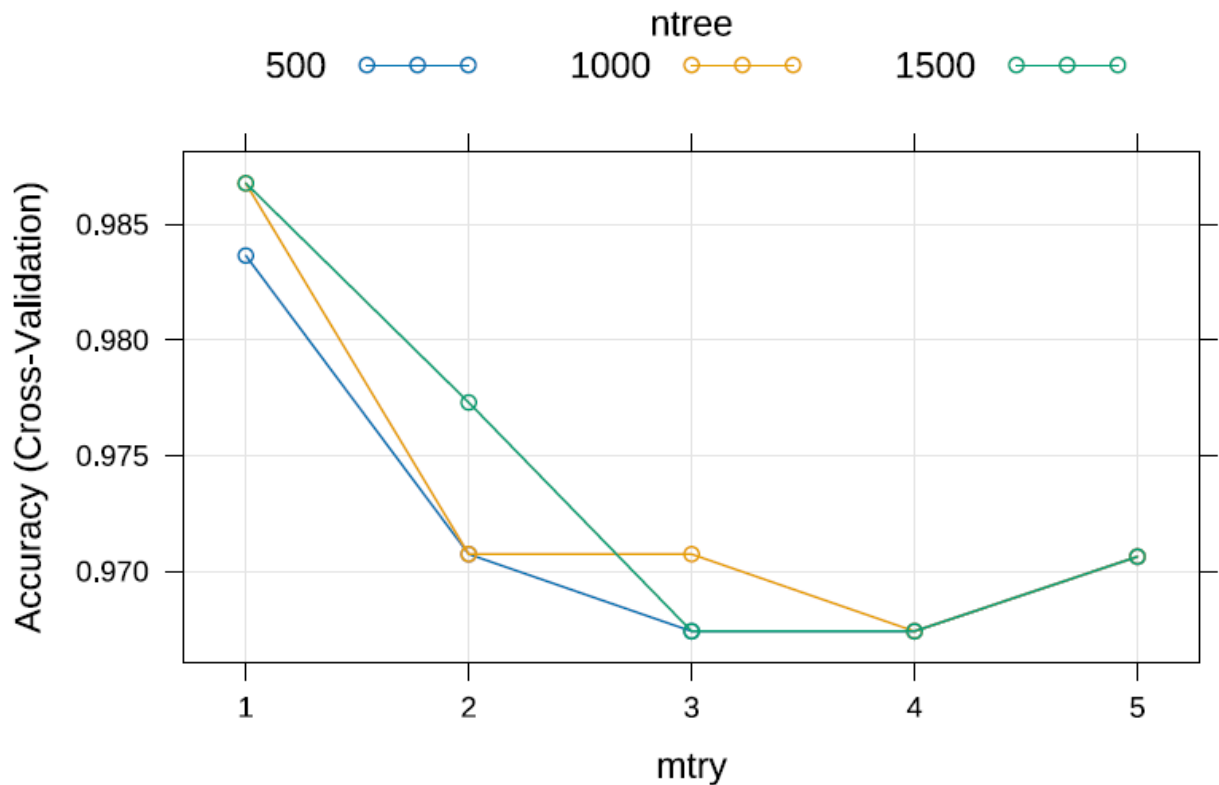
Figura 31 – Árvore de Decisão das Medidas Acústicas.

Fonte: Dados da Pesquisa, 2025.

O modelo de Árvore de Decisão obteve uma AUC de 0,6262, indicando capacidade moderada para classificar o estágio de prontidão para mudança. O ponto de corte ótimo foi identificado em 0,567, equilibrando sensibilidade e especificidade. Com esse limiar ajustado, os dados do teste foram classificados para avaliar o desempenho do modelo.

Além disso, o modelo Floresta Aleatória teve sua performance otimizada com base na métrica de acurácia, selecionando-se o valor mais alto obtido durante o processo de validação. Os melhores hiperparâmetros encontrados foram $mtry = 1$, indicando que uma variável preditora foi considerada a cada divisão das árvores, e $ntree = 1000$, o que significa que o modelo foi construído com 1000 árvores (Figura 32). Esses parâmetros refletem a configuração que proporcionou o melhor desempenho do modelo na tarefa de classificação.

Figura 32 - Hiperparâmetros do Floresta Aleatória com Base na Acurácia: Medidas Acústicas.



Fonte: Dados da pesquisa, 2025.

Constata-se um desempenho excelente do modelo Floresta Aleatória, com erro OOB) de 0,98%, indica alta capacidade de generalização (Figura 33). O modelo apresentou alta precisão e equilíbrio, o estágio de prontidão contemplação é classificada com erro nulo, enquanto o de manutenção tem erro de 1,96%. Como pode-se observar na matriz a seguir:

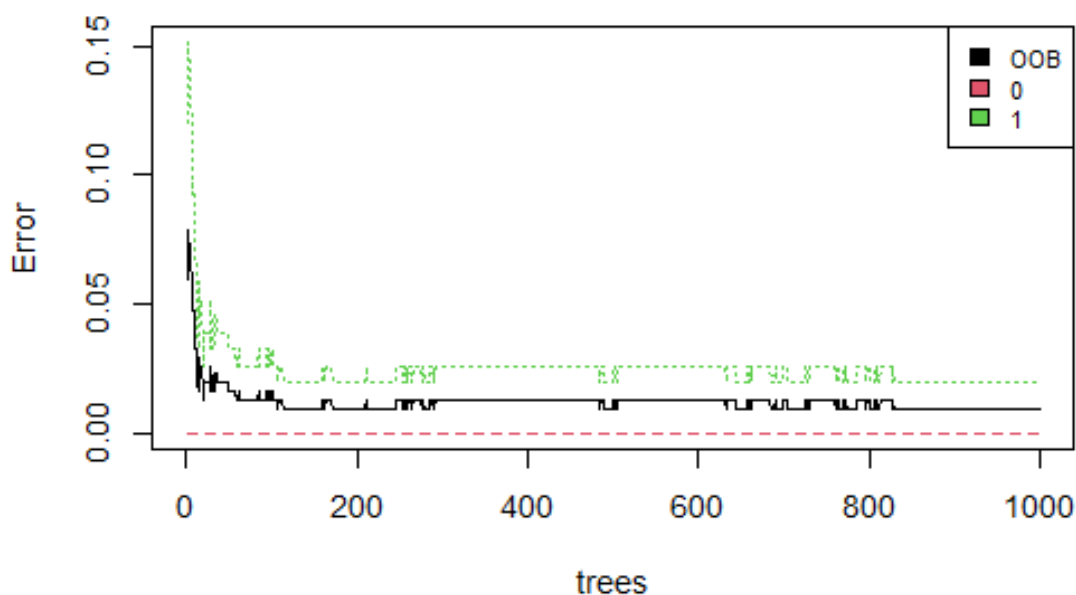
Tabela 13 – Matriz de Confusão OOB: Medidas Acústicas

	Contemplação	Manutenção	Erro da classe
Contemplação	153	0	0.000 (0%)
Manutenção	3	150	0.0196 (1,96%)

Fonte: Dados da pesquisa, 2025.

A matriz de confusão OOB indicou alta acurácia, com todos os acertos no estágio contemplação e apenas três erros no estágio manutenção, evidenciando boa capacidade de classificação. Contudo, a validação em dados de teste é necessária para confirmar a generalização do modelo.

Figura 33 – Erro OOB do Modelo de Floresta Aleatória: Medidas acústicas.



Fonte: Dados da pesquisa, 2025.

O valor da AUC obtido foi de 0,4738, indicando um desempenho inferior ao esperado, abaixo de 0,5, o que sugere que o modelo está classificando pior do que o acaso. O ponto de corte foi de 0,952, foi feita uma nova classificação e gerada a matriz de confusão a seguir:

Tabela 14 – Matriz de Confusão da Floresta Aleatória: Medidas Acústicas.

	Contemplação	Manutenção
Contemplação	2	36
Manutenção	3	29

Fonte: Dados da pesquisa, 2025.

Na matriz de confusão da validação externa, o modelo apresentou dificuldades para classificar corretamente o estágio contemplação, acertando apenas 2 casos e classificando erroneamente 36 indivíduos. Para o estágio manutenção, classificou

corretamente 29 casos, com 3 erros de classificação. Esses resultados diferem dos obtidos na validação interna, evidenciando limitação do modelo para generalizar a classificação dos estágios de prontidão.

Tabela 15 – Desempenho dos Modelos Para Classificação do Estágio de Prontidão com Dados das Medidas Acústicas.

	Regressão Logística	KNN	Naive Bayes	Árvore de Decisão	Floresta Aleatória
Acurácia	0,3571	0,8429	0,7857	0,8286	0,4429
Acurácia Balanceada	0,6538	0,6385	0,6077	0,6308	0,4230
Kappa	0,0597	0,1895	0,1176	0,1683	-0,038
Sensibilidade	0,3077	0,8769	0,8154	0,8615	0,4461
Especificidade	1,0000	0,4000	0,4000	0,4000	0,4000
Valor Preditivo Positivo	1,0000	0,9500	0,9464	0,9492	0,9062
Valor Preditivo Negativo	0,1000	0,2000	0,1429	0,1818	0,0526
Prevalência	0,9286	0,9286	0,9286	0,9286	0,9286
Taxa de Detecção	0,2857	0,8143	0,7571	0,8000	0,4143
Prevalência Estimada	0,2857	0,8571	0,8000	0,8429	0,4571
Taxa de Informação Nula	0,9286	0,9286	0,9286	0,9286	0,9286
P-Valor [Acurácia > Taxa Nula]	1	0,9962	0,99997	0,9987	1
P-Valor Teste de McNemar	5.41e-11	0,2278	0,03887	0,1489	2.99e-07

Fonte: Dados da Pesquisa, 2025.

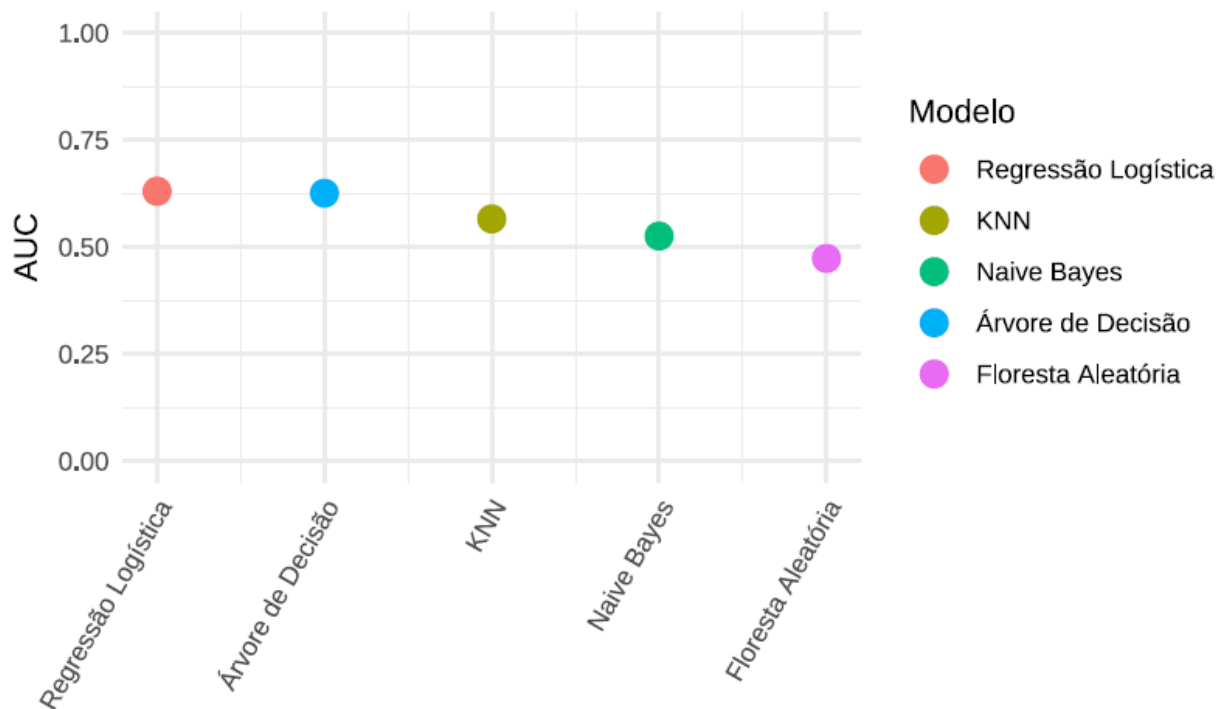
Conforme a Tabela 15, os modelos KNN e Árvore de Decisão apresentaram as melhores acurácias, com 0,8429 e 0,8286, respectivamente, e altas sensibilidades (acima de 0,86), indicando boa capacidade para identificar corretamente indivíduos no estágio de prontidão manutenção. No entanto, a especificidade desses modelos foi

baixa (0,4000), mostrando dificuldade na correta identificação dos indivíduos no estágio contemplação.

O modelo Naive Bayes apresentou acurácia de 0,7857 e sensibilidade de 0,8154, também com boa capacidade de identificação dos casos em manutenção, mas, assim como os demais, teve baixa especificidade (0,40), refletindo limitações na detecção dos indivíduos em contemplação. A acurácia balanceada, que pondera igualmente o desempenho em ambas as classes, foi semelhante entre Regressão Logística (0,6538), KNN (0,6385), Árvore de Decisão (0,6308) e Naive Bayes (0,6077), sugerindo que, apesar da alta acurácia geral do KNN, da Árvore e do Naive Bayes, o desempenho nesses dados desequilibrados não é tão robusto para os dois estágios.

A Floresta Aleatória apresentou a pior acurácia balanceada (0,4230), refletindo o baixo desempenho global, que difere dos modelos a partir de dados da Anamnese, onde apresentou um dos melhores desempenhos. A Regressão Logística, apesar da baixa acurácia geral (0,3571), alcançou a maior especificidade (1,0) e a melhor acurácia balanceada (0,6538), o que indica bom desempenho na identificação do estágio contemplação, mas falha na detecção do estágio manutenção, evidenciado pela baixa sensibilidade (0,3077).

Os valores de Kappa foram baixos para todos os modelos, indicando que a concordância entre as previsões dos modelos e as classificações reais é apenas ligeiramente superior ao que poderia ocorrer por acaso. Os testes de McNemar indicam diferenças estatisticamente significativas para Regressão Logística e Floresta Aleatória, mas não para KNN, Árvore de Decisão e Naive Bayes.

Figura 34 - Valores de AUC dos Modelos: Medidas Acústicas.

Fonte: Dados da Pesquisa, 2025.

A Regressão Logística obteve o melhor desempenho geral na métrica AUC (0,6308), seguida muito de perto pela Árvore de Decisão (0,6262). Ambos ficaram acima do valor de 0,60, o que indica uma capacidade moderada de discriminar entre as classes. O KNN (0,5662) e o Naive Bayes (0,5262) apresentaram desempenhos mais baixos, próximos ao limite de 0,50, o que representa desempenho pouco melhor que o acaso. Enquanto, a Floresta Aleatória, que normalmente tende a apresentar bom desempenho, obteve a menor AUC (0,4738), sugerindo que, no contexto e nos dados utilizados, o modelo não conseguiu aprender padrões eficazes para distinguir corretamente os estágios.

O desempenho geral foi limitado em todos os modelos. A Regressão Logística, embora tenha apresentado a maior AUC e especificidade perfeita, revelou baixa sensibilidade, o que significa que o modelo é eficaz para identificar corretamente os indivíduos no estágio de contemplação, porém falha em detectar adequadamente aqueles no estágio de manutenção. Esse comportamento sugere que o modelo pode estar subestimando os casos do estágio de manutenção, o que limita seu uso prático para intervenções direcionadas a esses pacientes.

Por outro lado, os modelos KNN, Árvore de Decisão e Naive Bayes demonstraram boa capacidade em reconhecer os indivíduos no estágio de manutenção, com sensibilidades elevadas, mas apresentaram baixa especificidade, o que implica dificuldade em distinguir os indivíduos no estágio de contemplação. Essa limitação pode decorrer do desequilíbrio entre as classes, em que o estágio de manutenção possui maior representatividade, influenciando os modelos a favorecerem a classe majoritária. Além disso, a baixa especificidade reflete a complexidade intrínseca da prontidão para mudança, um fenômeno que envolve aspectos cognitivos, emocionais e comportamentais que podem não estar totalmente capturados pelas variáveis utilizadas.

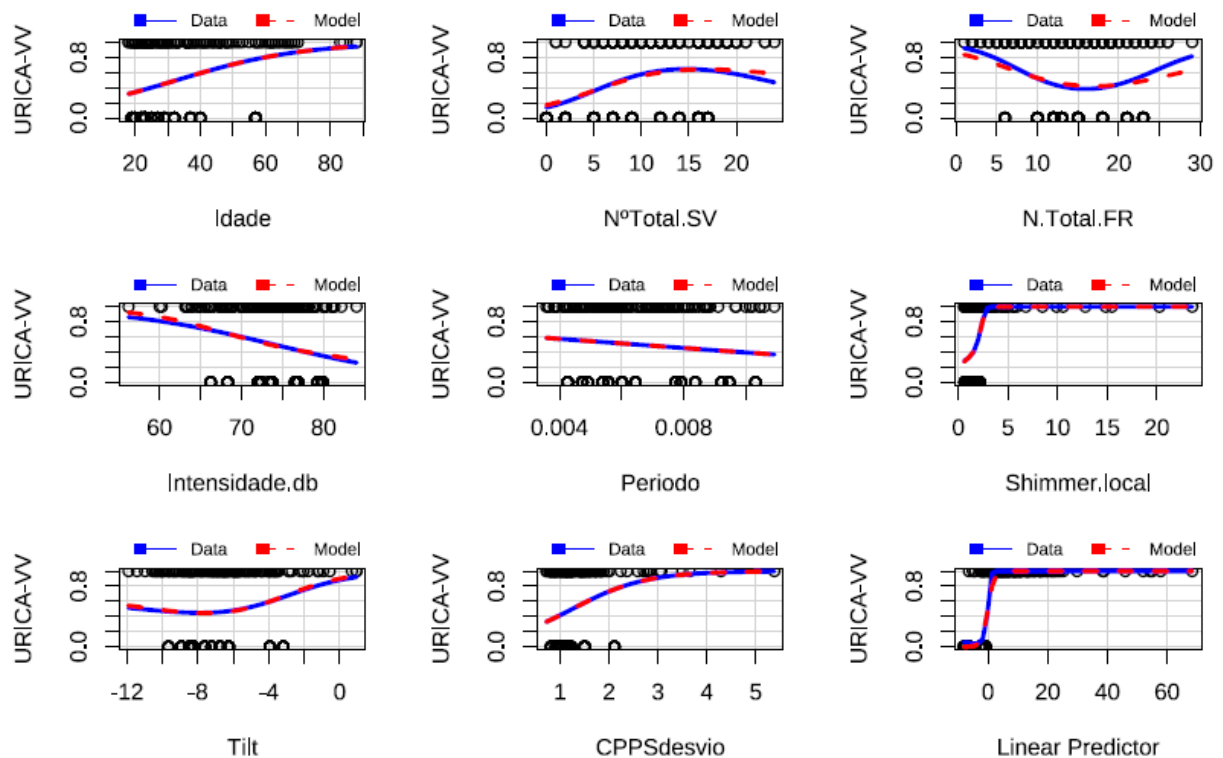
A Floresta Aleatória, apesar de ser um modelo robusto para muitas aplicações, apresentou baixo desempenho geral neste contexto, isso reforça a importância de ajustes específicos e testes adicionais para cada algoritmo, bem como a potencial necessidade de inclusão de variáveis complementares que melhorem a discriminação entre os estágios. Apesar de fornecerem informações importantes, os modelos atuais ainda carecem de precisão suficiente para uso clínico isolado, indicando a necessidade de estratégias adicionais para melhorar sua capacidade preditiva e apoiar intervenções mais eficazes.

5.6 Modelos de AM – Dados Multidimensionais da Voz

Por fim, foram propostos modelos com todas as variáveis anteriores que foram analisadas nos modelos isolados, usando as variáveis de cada uma das dimensões da Avaliação Multidimensional da Voz, compreendendo os dados da anamnese, dos instrumentos de autoavaliação, clínicos e das medidas acústicas.

Para a regressão logística, foi elaborada a Figura 35 que apresenta a relação entre cada variável preditora contínua e a variável resposta, que é o estágio de prontidão para mudança. O gráfico do Linear Predictor mostra a relação esperada entre o preditor linear do modelo e a variável resposta. Observa-se um crescimento monotônico, indicando que, à medida que o valor do preditor linear aumenta, também aumenta a probabilidade de o indivíduo estar no estágio de prontidão para mudança. Esse comportamento confirma que o modelo está ajustado conforme o esperado, reforçando a coerência entre os efeitos combinados das variáveis e a saída predita pela regressão logística.

Figura 35 - Efeitos Marginais do Modelo de Regressão Logística: Dados Multidimensionais da Voz.

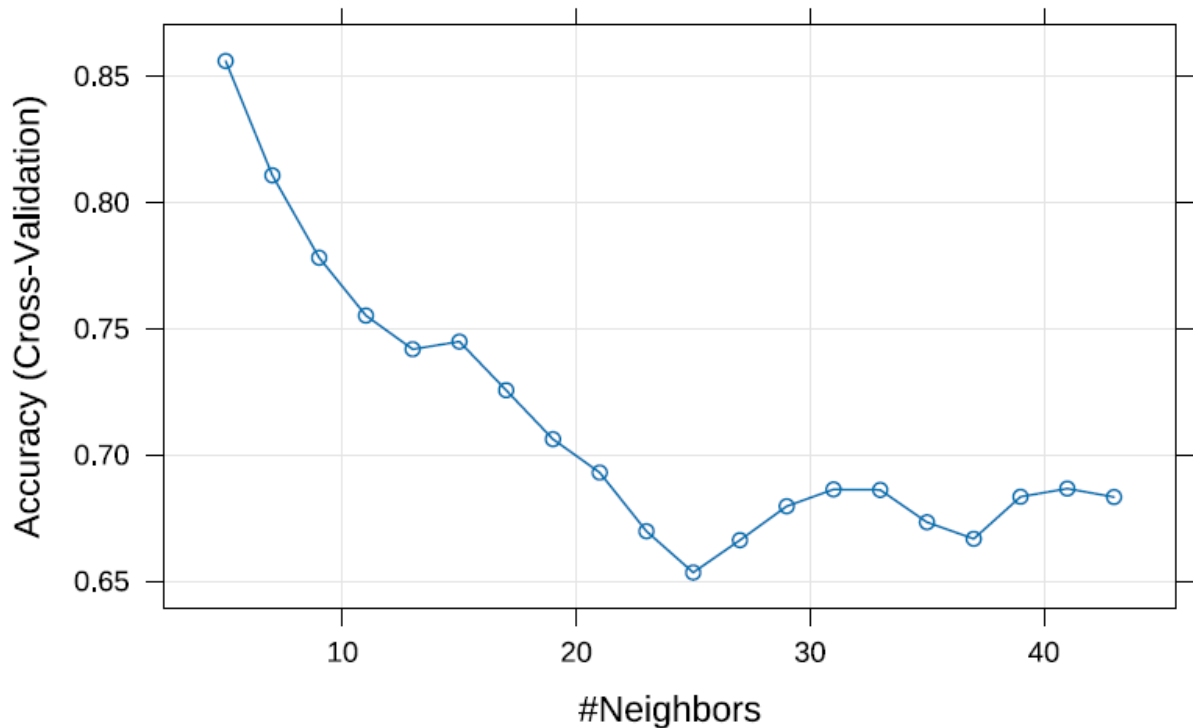


Fonte: Dados da Pesquisa, 2025.

O desempenho da Regressão Logística também foi avaliado pelo valor da AUC, foi obtido 0,5662, indicando um desempenho próximo ao aleatório, com baixa capacidade em distinguir entre os estágios de prontidão. O ponto de corte utilizado foi de 0,9 em busca do melhor equilíbrio entre sensibilidade e especificidade.

Quanto ao modelo KNN, em relação ao ajuste do k , através da Figura 36, a escolha de $K=5$ vizinhos é o ideal porque ele apresenta a maior acurácia de validação cruzada, aproximadamente 0,85. Isso sugere que, para este conjunto de dados, considerar os 5 vizinhos mais próximos é a estratégia que oferece o melhor equilíbrio entre viés e variância, e resulta na melhor capacidade de generalização do modelo.

Figura 36 - Número de Vizinhos do KNN com Base na Acurácia: Dados Multidimensionais da Voz.

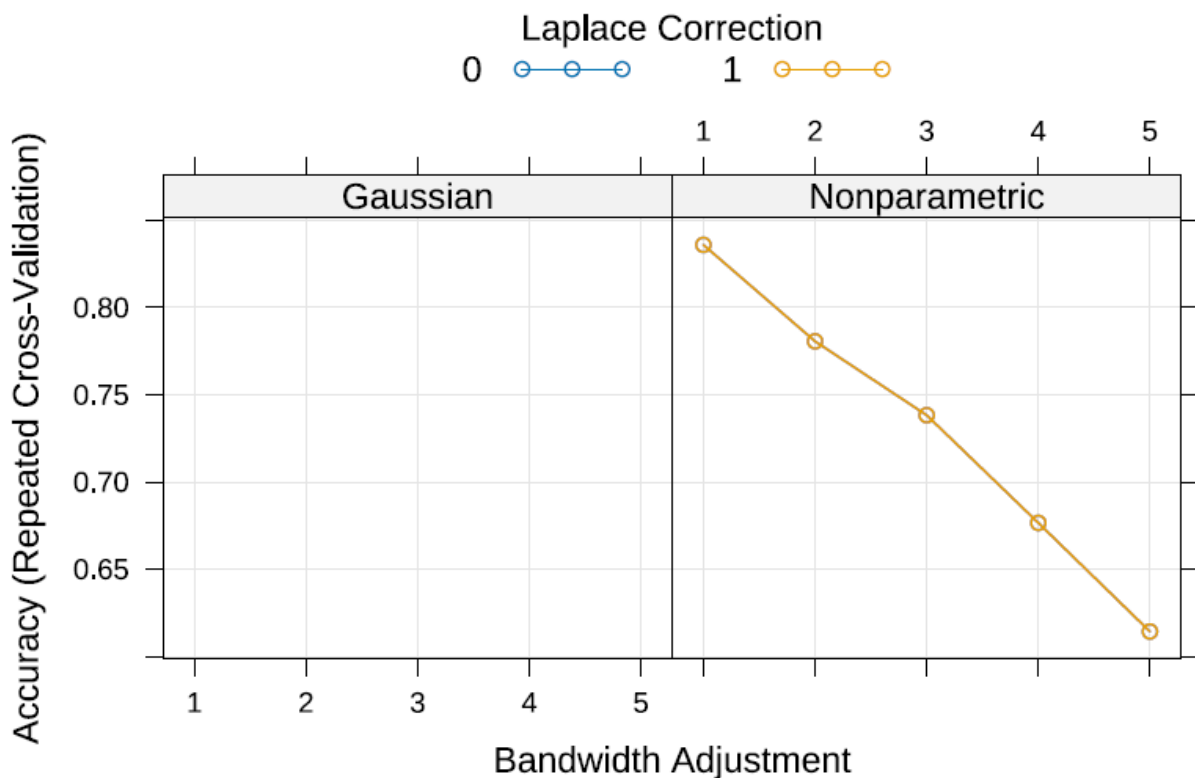


Fonte: Dados da Pesquisa.

Para o KNN, o valor da AUC foi de 0,6185, o que indica um desempenho moderado na discriminação entre os estágios de prontidão. Foi identificado um ponto de corte ótimo de 0,086. Com a adoção desse limiar, foram realizadas as predições para verificar as métricas de desempenho.

No modelo de Naive Bayes é possível na Figura 37 verificar como a acurácia foi utilizada como critério para a seleção dos hiperparâmetros, sendo escolhido aquele que apresentou o maior valor. Os hiperparâmetros finais adotados para o modelo Naive Bayes foram: $fL = 0$, $usekernel = TRUE$ e $adjust = 1$. Esses parâmetros indicam que não foi aplicado um fator de suavização (fL), foi utilizada uma estimativa por kernel para modelar a distribuição dos preditores contínuos, e o parâmetro $adjust$ foi mantido em seu valor padrão, controlando a largura da banda do kernel.

Figura 37 - Hiperparâmetros do Naive Bayes com Base na Acurácia: Dados Multidimensionais da Voz.

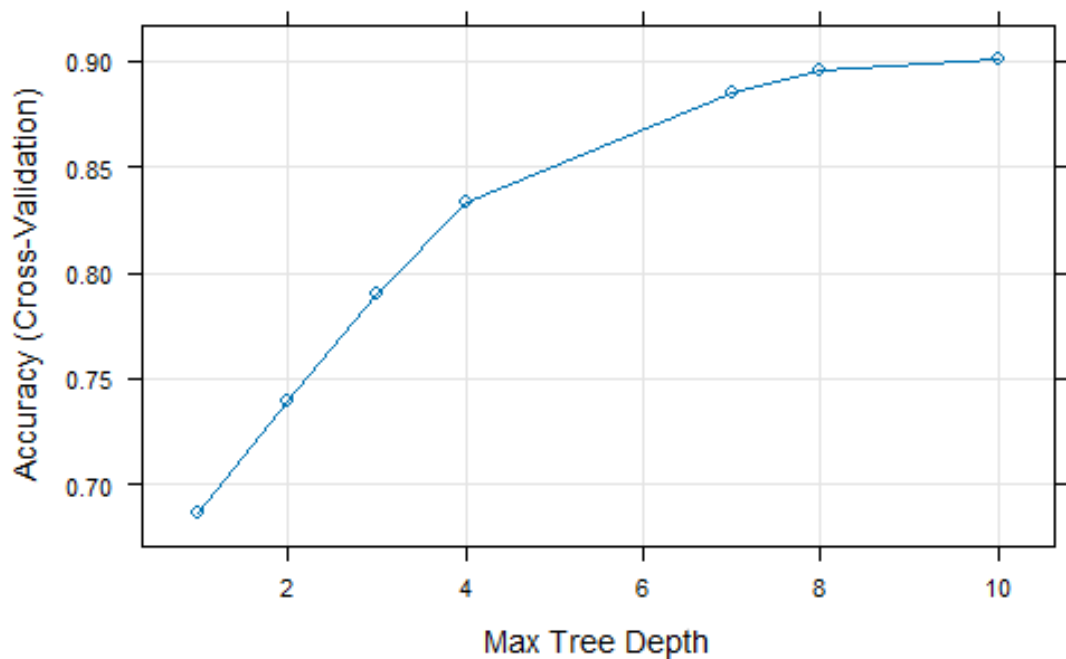


Fonte: Dados da Pesquisa, 2025.

Quanto o valor da AUC para o modelo Naive Bayes foi de 0,59, indica um desempenho apenas ligeiramente superior ao acaso. Quanto ao ponto de corte, foi definido 0,5, que visa uma boa capacidade discriminativa.

Em seguida, foi realizado o ajuste da Árvore de Decisão, a Figura 38 exibe a relação entre a acurácia medida por validação cruzada e a profundidade máxima da árvore. Fica evidente que, ao aumentar a profundidade, a acurácia do modelo cresce significativamente até um ponto de estabilização. A escolha foi de uma profundidade máxima de 10, pois é nesse patamar que o modelo atinge sua acurácia máxima, em torno de 0.90.

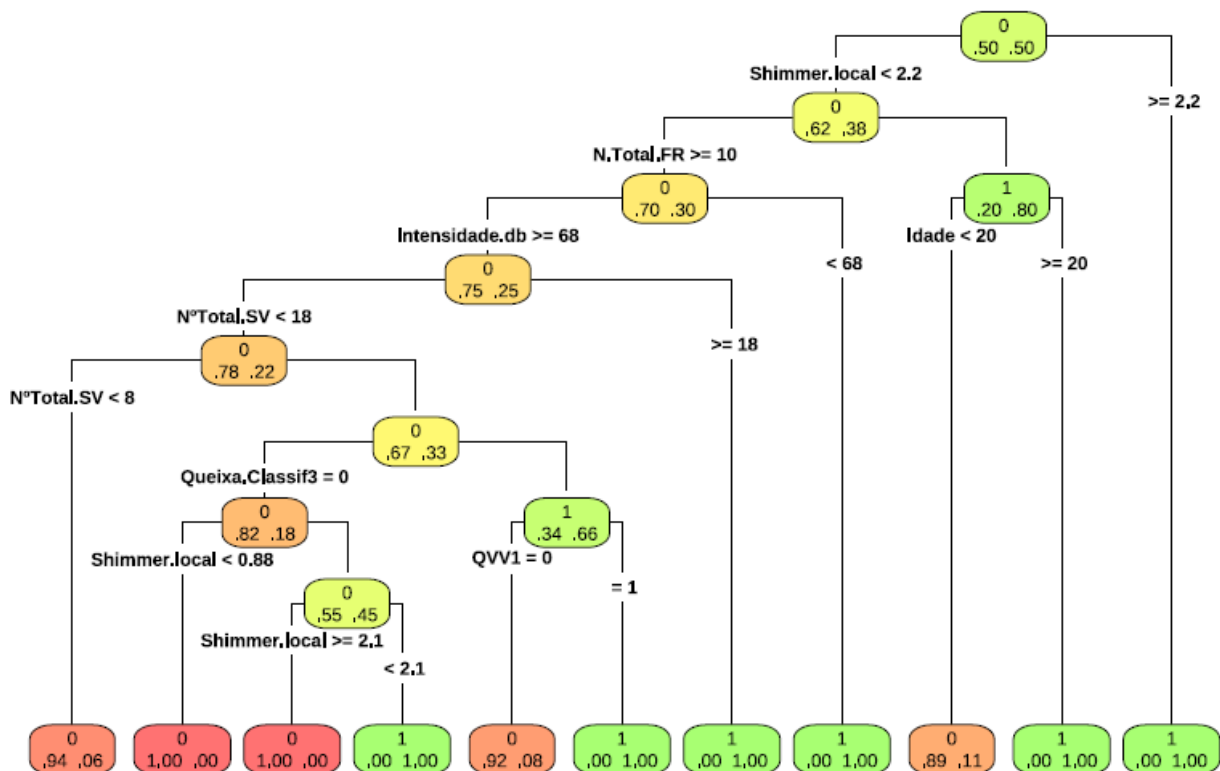
Figura 38 - Profundidade Máxima da Árvore de Decisão com Base na Acurácia:
Dados Multidimensionais da Voz.



Fonte: Dados da Pesquisa, 2025.

Na Figura 39, apresenta-se a árvore de decisão, a qual ilustra visualmente o processo de classificação com base nas variáveis preditoras selecionadas. A variável Shimmer local demonstra ser um preditor de grande relevância, uma vez que aparece como o primeiro nó de decisão da árvore, indicando sua importância na classificação. Valores elevados dessa variável (acima ou iguais a 2.2) conduzem diretamente à predição da classe 1, o que reforça seu peso no modelo. Além disso, outras variáveis como Nº Total de FR, Intensidade, Nº Total de SV, Classificação da Queixa, QVV e Idade também se mostram relevantes na distinção entre as classes nos casos em que o valor de Shimmer local é baixo. Observa-se que alguns caminhos da árvore levam a predições bastante claras, com 100%, enquanto outros apresentam uma mistura entre as classes, o que sugere que, para determinadas combinações de características, a separação entre os grupos não é tão evidente.

Figura 39 – Árvore de Decisão dos Dados Multidimensionais da Voz.

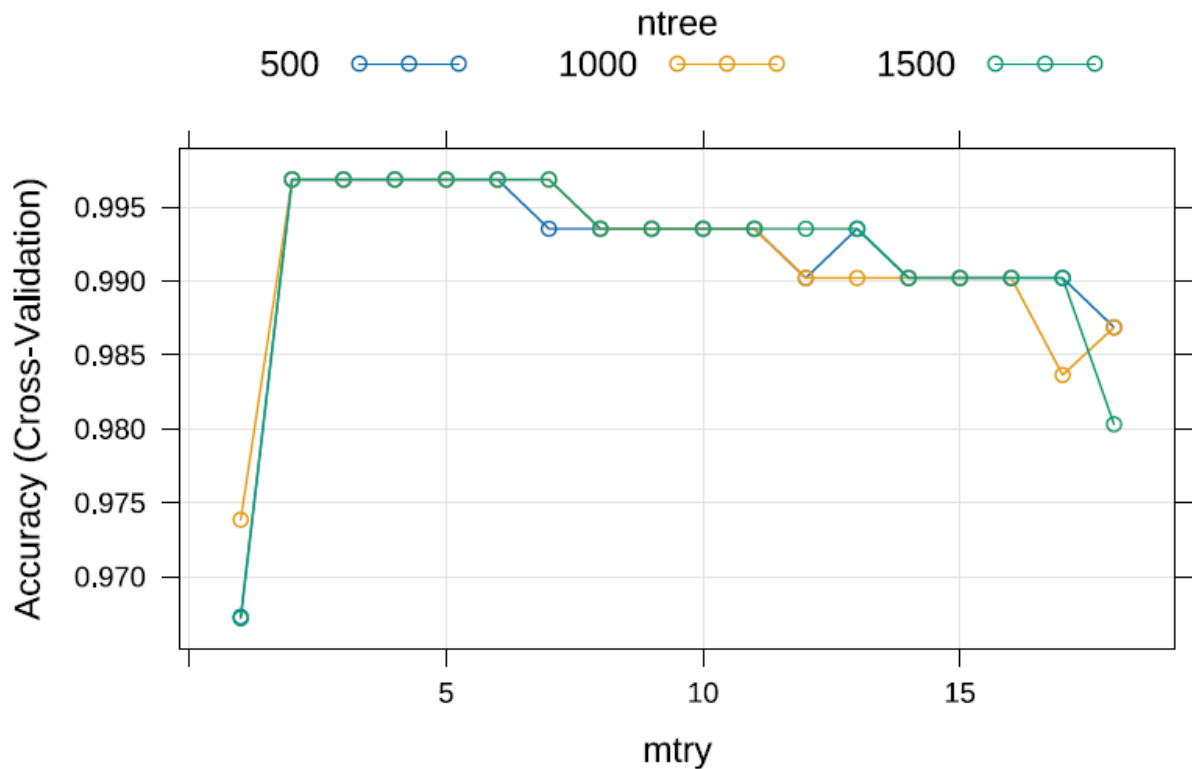


Fonte: Dados da pesquisa, 2025.

Em relação ao AUC do modelo de Árvore de Decisão foi de 0,45, indicando que o desempenho do modelo está abaixo do valor esperado por acaso. Isso sugere uma baixa capacidade discriminativa para diferenciar corretamente entre os estágios de prontidão. Foi identificado do ponto de corte de 0,06.

No modelo Floresta Aleatória também foi realizado ajuste de hiperparâmetros, é possível visualizar na Figura 40. Como a acurácia foi o critério de seleção, o modelo ideal de Floresta Aleatória foi definido com $mtry = 2$ e $ntree = 500$. Essa combinação oferece a maior acurácia (próxima a 0.995), com $mtry = 2$ garantindo um bom poder preditivo e $ntree = 500$ sendo a escolha mais eficiente computacionalmente, já que mais árvores não trariam ganhos significativos de desempenho nesse ponto. Sendo o ponto no gráfico de maior acurácia.

Figura 40 - Hiperparâmetros da Floresta Aleatória com Base na Acurácia: Dados Multidimensionais da Voz.



Fonte: Dados da pesquisa, 2025.

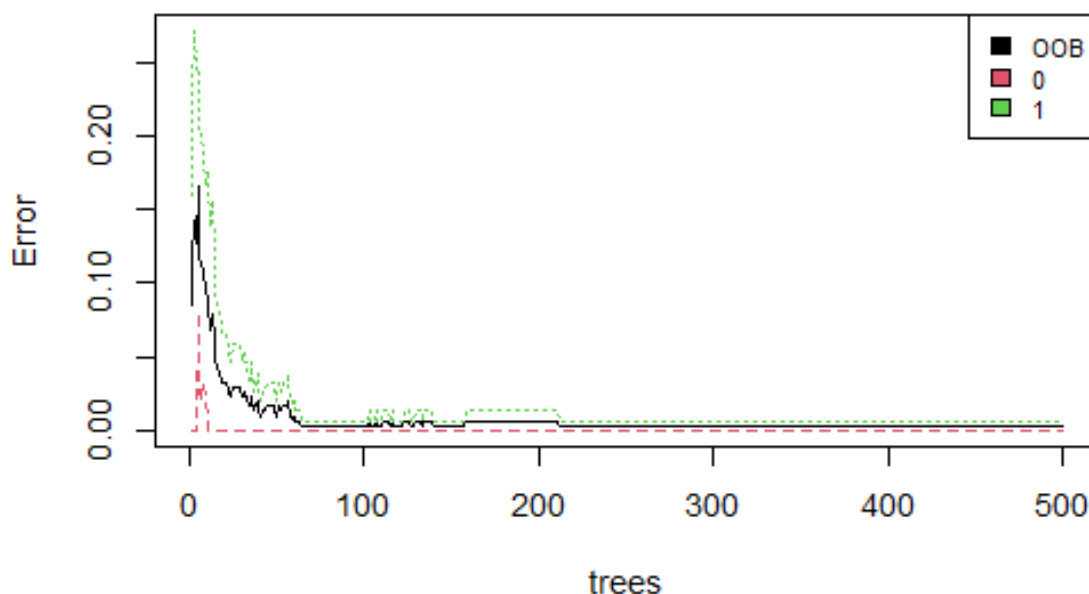
Na validação interna, obteve-se um desempenho excelente do modelo Floresta Aleatória, com erro OOB de 0,33%, com alta capacidade de generalização (Figura 41). Como pode-se observar na matriz de confusão a seguir:

Tabela 16 – Matriz de Confusão OOB: Dados Multidimensionais da Voz.

	Contemplação	Manutenção	Erro da classe
Contemplação	153	0	0.000 (0%)
Manutenção	1	152	0.0065 (0,65%)

Fonte: Dados da pesquisa, 2025.

Figura 41 - Erro OOB do Modelo Floresta Aleatória: Dados Multidimensionais da Voz.



Fonte: Dados da pesquisa, 2025.

A matriz de confusão OOB indicou alta acurácia, com todos os acertos no estágio contemplação e apenas um erro no estágio manutenção, com boa capacidade de classificação. Contudo, a validação externa em dados de teste é necessária para confirmar a generalização do modelo. O valor da AUC foi de 0,5585, indica que o modelo possui uma performance limitada para distinguir os estágios de prontidão. O ponto de corte foi de 0,849 no modelo de Floresta Aleatória, com esse limiar utilizando os dados do conjunto teste foi gerada a matriz de confusão a seguir:

Tabela 17 – Matriz de Confusão da Floresta Aleatória: Dados Multidimensionais da Voz.

	Contemplação	Manutenção
Contemplação	3	23
Manutenção	2	42

Fonte: Dados da pesquisa, 2025.

O modelo apresentou dificuldades para classificar corretamente. No estágio contemplação, acertou 3 casos e classificou erroneamente 23 indivíduos. Para o estágio manutenção, classificou corretamente 29 casos, com 3 erros de classificação.

Por serem resultados que diferem da validação interna, evidencia a limitação do modelo para generalizar a classificação do estágio de prontidão.

Tabela 18 - Desempenho dos Modelos para Classificação do Estágio de Prontidão com os Dados Multidimensionais da Voz.

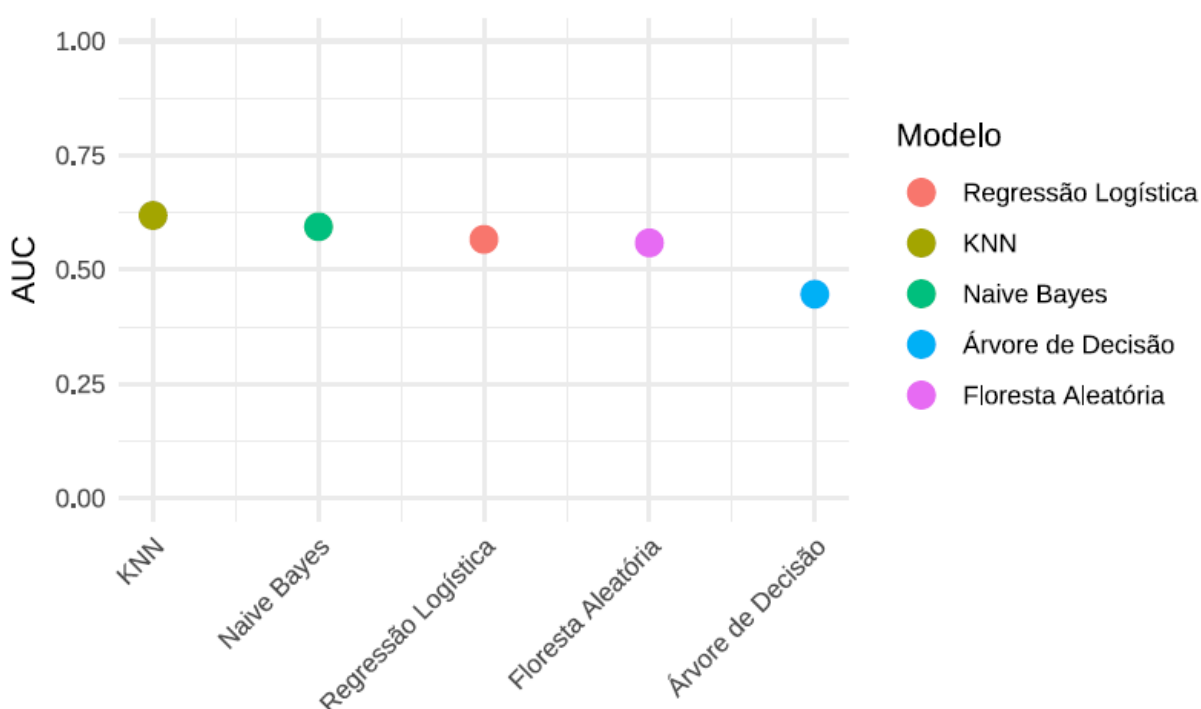
	Regressão Logística	KNN	Naive Bayes	Árvore de Decisão	Floresta Aleatória
Acurácia	0.7000	0.8857	0.8143	0.8857	0.6429
Acurácia	0.4692	0.6615	0.6231	0.4769	0.6231
Balanceada					
Kappa	-0.028	0.2727	0.1495	-0.0566	0.0838
Sensibilidade	0.73846	0.9231	0.8462	0.9538	0.6462
Especificidade	0.2000	0.4000	0.4000	0.0000	0.6000
Valor Preditivo	0.92308	0.9524	0.9483	0.9254	0.9545
Positivo					
Valor Preditivo	0.05556	0.2857	0.1667	0.0000	0.1154
Negativo					
Prevalência	0.92857	0.9286	0.9286	0.9286	0.9286
Taxa de	0.68571	0.8571	0.7857	0.8857	0.6000
Deteção					
Prevalência	0.74286	0.9000	0.8286	0.9571	0.6286
Estimada					
Taxa de	0.9286	0.9286	0.9286	0.9286	0.9286
Informação					
Nula					
P-Valor	1.000000	0.9391	0.99962	0.9391	1
[Acurácia >					
Inf. Nula]					
P-Valor do	0.008829	0.7237	0.09609	0.7237	6.334e-05
Teste de					
McNemar					

Fonte: Dados da pesquisa, 2025.

A análise comparativa entre os modelos de classificação indica que o KNN obteve a maior acurácia (0.8857) e sensibilidade (0.9231), além da melhor acurácia balanceada (0.6615), evidenciando sua capacidade de identificar corretamente os indivíduos em manutenção sem negligenciar completamente os em contemplação, mesmo diante do forte desequilíbrio entre as classes. Seu valor de Kappa (0.2727) foi o mais alto, indicando uma concordância razoável com os dados observados. A especificidade (0.4000) foi superior à da Regressão Logística e da Árvore de Decisão, que não identificou nenhum caso em contemplação. Apesar da sensibilidade elevada da Árvore de Decisão (0.9538), sua especificidade nula comprometeu sua acurácia balanceada e resultou em Kappa negativo.

O Naive Bayes e a Floresta Aleatória apresentaram desempenhos intermediários, sendo a floresta o modelo com maior especificidade (0.6000), embora com menor sensibilidade e acurácia. A Regressão Logística foi o modelo que não apresentou bom desempenho, com baixa especificidade (0.2000), acurácia balanceada reduzida (0.4692) e Kappa negativo, demonstrando fragilidade na identificação de indivíduos em contemplação. Assim, o KNN se sobressai como a alternativa mais equilibrada para a classificação dos estágios de prontidão, contribuindo de forma mais confiável para a triagem e planejamento da terapia de voz.

Figura 42 – Valores de AUC dos Modelos: Dados Multidimensionais da Voz.



Fonte: Dados da Pesquisa, 2025.

Os resultados mostram que os modelos KNN e Naive Bayes apresentaram os melhores desempenhos, com AUC de 0,6185 e 0,5938, respectivamente, indicando uma capacidade moderada para distinguir os estágios de prontidão para a terapia de voz. A Regressão Logística e a Floresta Aleatória tiveram desempenhos semelhantes, com AUCs próximas de 0,56, sugerindo uma discriminação um pouco abaixo dos modelos mencionados. E o modelo de Árvore de Decisão apresentou o pior desempenho, com AUC de 0,4462, indicando uma capacidade inferior à aleatoriedade na classificação dos estágios. Esses resultados indicam que, entre os modelos avaliados, KNN e Naive Bayes mostram maior potencial para aplicação. Contudo, o modelo KNN apresentou métricas um pouco superiores em comparação, sendo assim o mais recomendável entre os avaliados, embora o desempenho geral do modelo ainda seja considerado moderado.

Em suma, o modelo KNN se destaca como o mais equilibrado e eficaz entre os avaliados com variáveis que contemplam todas dimensões da avaliação vocal, especialmente diante do desbalanceamento. Ele mantém boas taxas de acerto, o que o torna mais confiável para aplicações práticas nesse contexto. Além disso, o modelo KNN tem se destacado na área da voz. Sua eficácia foi comprovada na diferenciação entre vozes patológicas e saudáveis por meio de gravações feitas com smartphones, demonstrando viabilidade em contextos de triagem acessível (Cesare *et al.*, 2024).

De forma semelhante, o KNN também foi utilizado com sucesso na identificação da doença de Parkinson a partir de alterações acústicas da voz, evidenciando sua capacidade de captar padrões sutis associados a distúrbios neurológicos (Ouhmida *et al.*, 2021). Esses achados corroboram os resultados da presente pesquisa, na qual o KNN apresentou um bom desempenho, evidenciando seu potencial como suporte à tomada de decisões na fonoaudiologia.

No entanto, é importante reconhecer algumas limitações inerentes ao KNN, como sua sensibilidade a ruídos nos dados e à escolha do parâmetro k , que pode impactar o desempenho do modelo. Além disso, o KNN exige um pré-processamento cuidadoso dos dados e pode se beneficiar de técnicas de seleção e engenharia de variáveis para otimizar a predição (Shetty *et al.*, 2022). A utilização de variáveis provenientes de múltiplas dimensões da avaliação vocal, incluindo aspectos clínicos, acústicos, laringológicos e de autoavaliação, foi determinante para o desempenho dos modelos. Essa abordagem multidimensional reflete a complexidade do distúrbio vocal

e oferece uma visão mais abrangente do paciente, contribuindo para a identificação mais precisa dos estágios de prontidão.

De modo geral, os resultados dos modelos em diferentes dimensões evidenciam um desafio recorrente em bases de dados desbalanceadas, que é o viés dos algoritmos em privilegiar a classe mais representada, mesmo após tentativas de correção por meio de técnicas de reamostragem. Esse desequilíbrio pode ser parcialmente explicado pelo tipo de amostragem utilizada na pesquisa. Por se tratar de uma amostra por conveniência, composta por indivíduos que já estavam em busca de atendimento fonoaudiológico, é esperado que a maioria dos participantes apresentasse um maior nível de consciência sobre o problema vocal e, conseqüentemente, estivesse em estágios mais avançados de prontidão para a mudança, como o estágio de manutenção. Essa característica da amostra contribuiu para a predominância dessa classe na variável resposta, dificultando o desempenho dos modelos de classificação ao tentar identificar adequadamente os casos pertencentes aos demais estágios de prontidão.

Entretanto, ainda foi possível identificar modelos com desempenho razoável, capazes de lidar, mesmo que parcialmente, com o desbalanceamento da amostra. Isso sugere que, mesmo diante de limitações amostrais, é viável empregar algoritmos de aprendizado de máquina como ferramenta auxiliar no reconhecimento de diferentes estágios de prontidão para a mudança. Tal resultado reforça o potencial da aplicação de modelos em contextos clínicos, desde que acompanhados de estratégias que considerem as características da população-alvo e busquem mitigar o impacto do desbalanceamento nas análises.

Finalmente, cabe destacar que o estágio de prontidão para a terapia de voz, que o modelo busca prever, é um fator essencial para o sucesso do tratamento. Compreender em que fase o paciente se encontra, desde a contemplação até a manutenção, permite ao fonoaudiólogo adaptar as intervenções, aumentar a adesão ao tratamento e otimizar os resultados terapêuticos. Assim, modelos preditivos representam uma ferramenta valiosa para a prática clínica, oferecendo suporte na tomada de decisão e na personalização da terapia fonoaudiológica.

6. CONCLUSÃO

É possível identificar modelos com desempenho moderado, como o KNN e a Floresta Aleatória nos dados da anamnese, a Regressão Logística e o Naive Bayes nos dados clínicos (JPA, diagnóstico laríngeo e classificação da disfonia), e o KNN nos dados combinados das diferentes dimensões da avaliação vocal. Nos dados dos instrumentos de autoavaliação e das medidas acústicas, os modelos não apresentaram o desempenho esperado, demonstrando limitações para a predição do estágio de prontidão. Embora ainda apresentem limitações em seu poder preditivo, esses achados sugerem que modelos de AM podem contribuir para a identificação do estágio de prontidão para mudança, para possibilitar estratégias de intervenção mais direcionadas desde o início da terapia de voz.

REFERÊNCIAS

- AGUIAR, A. C.; ALMEIDA, L. N. A.; PERNAMBUCO, L.; RAMOS, N.; ANDRADE, J. M.; BEHLAU, M.; ALMEIDA, A. A. Urica-VV Scale: A New Research Perspective of The Stage of Readiness for Treatment in Patients with Dysphonia. **Journal of Voice**, v. 37, n. 6, p. 807-821, 2023.
- AGUIAR, A. C.; ALMEIDA, L. N. A.; PERNAMBUCO, L.; RAMOS, N.; ANDRADE, J. M.; BEHLAU, M.; ALMEIDA, A. A. Urica-VV Scale: A New Research Perspective of The Stage of Readiness for Treatment in Patients with Dysphonia. **Journal of voice**, 2021.
- ARAIN, M.; HAQUE, M.; JOHAL, L.; MATHUR, P.; NEL, W.; RAIS, A.; KHAN, S.; CRAIG, M. Maturation of the adolescent brain. **Neuropsychiatric Disease and Treatment**, v. 9, p. 449–461, 2013.
- ALLOGHANI, Mohamed; AL-JUMEILY, Dhiya; MUSTAFINA, Jamila; HUSSAIN, Abir; ALJAAF, Ahmed J. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: BERRY, M.; MOHAMED, A.; YAP, B. (Org.). Supervised and unsupervised learning for data science. **Springer**, p. 3-21, 2020.
- ALMEIDA, A. A.; LOPES, L. W.; AGUIAR, A. C.; OLIVEIRA, P.; BEHLAU, M. Avaliação e diagnóstico do comportamento vocal. In: PERNAMBUCO, Leandro; ASSENÇO, Ana Manhani Assenço (Org.). **Fonoaudiologia: avaliação e diagnóstico**. 1. ed. Rio de Janeiro: Thieme Revinter, 2021. v. 1, p. 1-430.
- ALMEIDA, L. N. A. ; NASCIMENTO, J. A. ; BEHLAU, M. ; VITOR, A. ; AGUIAR, A. C.; ALMEIDA, AA . Processo de validação de instrumentos de autoavaliação da voz no Brasil. **AUDIOLOGY - COMMUNICATION RESEARCH (ACR)**, v. 26, p. e2364, 2021.
- ALMEIDA, AA; LIMA, L.; ZAMBON, F. ; ALMEIDA, L. N. . **Instrumentos de autoavaliação da voz: fundamentos e atualidades**. In: Leonardo Lopes; Ana Paula Lefèvre Machado; Cíntia Alves Salgado Azoni; Juliana Ferreira Benatti; Rosane Sampaio Santos; Vanessa Veis Ribeiro; Giorvan Ânderson dos Santos Alves; Leandro de Araújo Pernambuco. (Org.). Tratado de Fonoaudiologia. 3ªed.Barueri - SP: Manole Saúde, 2024, v. 0, p. 763-770.
- ALTALHAN, Manahel; ALGARNI, Abdulmohsen; MONIA, Turki Hadj Alouane. Imbalanced data problem in machine learning: a review. **IEEE Access**, p. 1–1, jan. 2025.
- ALVIM, João Paulo Nóbrega. **Aplicação Web para geração de modelos automatizados de Aprendizado de Máquina**. 2019. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação pela Faculdade de Engenharia Elétrica) Universidade Federal de Uberlândia, 2019.

BARBOSA, R. L.; GIVISIEZ, B. S.; FERNANDES, C. L. C.; LICHTENFELS, P. Estão os participantes de grupos de cessação do tabagismo prontos a parar de fumar? Estudo sobre avaliação do nível motivacional em serviço de saúde comunitária / Are the participants in smoking cessation groups ready to quit smoking? Study on assessment of motivational level in a community health service. **Brazilian Journal of Development**, [S. l.], v. 6, n. 10, p. 83604–83618, 2020.

BEHLAU, M. The 2016 G. Paul Moore Lecture: Lessons in Voice Rehabilitation: Journal of Voice and Clinical Practice. **Journal of Voice**. V. 33, n. 5, p. 669-681, 2018.

BEHLAU, M.; ALMEIDA, A. A.; AMORIM, G. O. ; BALATA, P. ; BASTOS, S. ; CASSOL, M. ; CONSTANTINI, A. C. ; ECKLEY, C. ; ENGLERT, M. ; GAMA, A. C. C. ; GIELOW, I. ; GUIMARAES, B. ; LIMA, L. R. ; LOPES, L. W. ; MADAZIO, G. ; MORETI, F. T. ; MOUFFRON, V. ; NEMR, K. ; OLIVEIRA, P. ; PADOVANI, M. . Reduzindo o GAP entre a ciência e a clínica: lições da academia e da prática profissional - parte A: julgamento perceptivo-auditivo da qualidade vocal, análise acústica do sinal vocal e autoavaliação em voz. **CODAS**, v. 34, p. 1-12, 2022.

BEHLAU, M.; MADAZIO, G.; MORETI, F.; OLIVEIRA, G.; SANTOS, L. M. A.; PAULINELLI, B. R.; COUTO JUNIOR, E. de B. Efficiency and Cutoff Values of Self-Assessment Instruments on the Impact of a Voice Problem. **Journal of Voice**, v. 30, n. 4, p. 506.e9-506.e18, 2016.

BEHLAU, M.; MADAZIO, G.; YAMASAKI, R. Campo dinâmico vocal: avaliação da funcionalidade da voz. **CoDAS**, v. 35, n. 5, e20210083, 2023.

BEHLAU, M.; ZAMBON, F.; MORETI, F.; OLIVEIRA, G.; COUTO, E. B. Voice Self-assessment Protocols: Different Trends Among Organic and Behavioral Dysphonias. **Journal of Voice**, v. 31, n. 1, p. 112.e13-112.e27, 2017.

BEHLAU, Mara. **Voz: o livro do especialista**. Volume I. Rio de Janeiro: Revinter, 2001.

BODT, M.; WUYTS, F. L.; VAN DE HEYNING, P. H. Test-retest study of the GRBAS scale: influence of experience and professional back ground on perceptual rating of voice quality. **Journal of Voice**. v. 11, p. 74-80, 1997.

BREIMAN, Leo; FRIEDMAN, Jerome H.; OLSHEN, Richard A.; STONE, Charles J. **Classification and Regression Trees**. New York: Chapman & Hall, 1984.

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

BRUCE, P.; BRUCE, A. **Estatística Prática para Cientistas de Dados: 50 conceitos essenciais**. Alta Books, Rio de Janeiro-RJ, 2019.

BULC, Mateja; SVAB, Igor; GODYCKI-CWIRKO, Maciek. Factors that affect readiness to change lifestyle: A 22-country survey from primary care. **The European Journal of General Practice**, v. 21, p. 1-6, 2014.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 10. ed. SaraivaUni, 2023.

BYEON, H. Comparing ensemble-based machine learning classifiers developed for distinguishing hypokinetic dysarthria from presbyphonia. **Applied Sciences**, v. 11, n. 5, p. 2235, 2021.

CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Tradução. [s.l.] Artmed Editora, 2009.

CAPUCHO, Maria Clara Pinheiro. **Avaliação Multidimensional na voz Profissional**. 2017. Tese (Doutorado em Medicina), NOVA Medical School: Ciências Médicas e Ciências da Nutrição. Universidade NOVA de Lisboa, Portugal. 2017.

CARDING, P. N.; WILSON, J. A.; MACKENZIE, K.; DEARY, I. J. Measuring voice outcomes: state of the science review. **Journal of Laryngology & Otology**, v. 123, n. 8, p. 823-829, 2009.

CESARE, Michele Giuseppe; PERPETUINI, David; CARDONE, Daniela; MERLA, Arcangelo. Assessment of voice disorders using machine learning and vocal analysis of voice samples recorded through smartphones. **BioMedInformatics**, v. 4, n. 1, p. 549-565, 2024.

CONSTANTINI, A. C.; RIBEIRO, V. V.; BEHLAU, M. Voice Disorder Classifications: A Scoping Review - Part A. **Journal of Voice**: Official Journal of the Voice Foundation, v. 36, n. 5, p. S0892-1997, 2022.

CONWAY, D.; WHITE, J. M. **Machine learning for hackers**: case studies and algorithms to get you started. Sebastopol: O'Reilly Media, 2012.

COSTA, B. O. I. da; SILVA, P. O. C.; PINHEIRO, R. S. de A.; SILVA, H. F. da; ALMEIDA, A. A. F. de. Estágio de prontidão de pacientes com disfonia comportamental pré e pós-terapia de voz de grupo. **CoDAS**, v. 29, n. 4, e20160198, 2017.

COSTA, D. B. da; LOPES, L. W.; SILVA, E. G.; CUNHA, G. M. S. da; ALMEIDA, L. N. A.; ALMEIDA, A. A. F. de. Fatores de risco e emocionais na voz de professores com e sem queixas vocais. **Revista CEFAC**, v. 15, n. 4, p. 1001-1010, jul. 2013.

COX, D. R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society**: Series B (Methodological), London, v. 20, n. 2, p. 215–242, 1958.

DEJONCKERE, P. H.; BRADLEY, P.; CLEMENTE, P.; CORNUT, G.; CREVIER-BUCHMAN, L.; FRIEDRICH, G.; VAN DE HEYNING, P.; REMACLE, M.; WOISARD, V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of phonosurgical treatments and evaluating new assessment techniques: guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). **European Archives of Oto-Rhino-Laryngology**, v. 258, p. 77-82, 2001.

DENG, Wei; GUO, Yixiu; LIU, Jie; LI, Yong; LIU, Dingguo; ZHU, Liang. A missing power data filling method based on improved random forest algorithm. **Chinese Journal of Electrical Engineering**, Changsha, v. 5, n. 4, p. 37–43, dez. 2019.

DEARY, I. J.; WILSON, J. A.; CARDING, P. N.; MACKENZIE, K. VoISS: a patient-derived Voice Symptom Scale. **Journal of Psychosomatic Research**, v. 54, n. 5, p. 483-489, 2003

DOZOIS, David J. A.; WESTRA, Henny A.; COLLINS, Kerry A.; FUNG, Tak S.; GARRY, Jennifer K. F. Stages of change in anxiety: psychometric properties of the University of Rhode Island Change Assessment (URICA) scale. **Behaviour Research and Therapy**, v. 42, n. 6, p. 711-729, jun. 2004.

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science**: Algoritmos de Machine Learning e métodos de análise. São Paulo: Bookwire, 2020.

FIX, E.; HODGES, J. L. **Discriminatory analysis**. Nonparametric discrimination: consistency properties. Technical Report, University of California, Los Angeles, 1951.

FLORENCIO, Vanessa de Oliveira; ALMEIDA, Anna Alice; BALATA, Patrícia; NASCIMENTO, Silvana; BROCKMANN-BAUSER, Meike; LOPES, Leonardo Wanderley. Differences and reliability of linear and nonlinear acoustic measures as a function of vocal intensity in individuals with voice disorders. **Journal of Voice**, v. 37, n. 5, p. 663-681, set. 2023.

GHIRARDI, A. C. A. M.; CONSTANTINI, A. C. Distúrbio de voz relacionado ao trabalho – DVRT: considerações sobre o tratamento e retorno ao trabalho. In: FERREIRA, L. P.; ANDRADA E SILVA, M. A. (Orgs.). **Distúrbio de voz relacionado ao trabalho: conquistas e desafios na América Latina**. São Paulo: Sintropia, 2022. p. 178-184.

GÓES, T. R. V. de; FERRACCIU, C. C. S.; SILVA, D. R. O. da. Associação entre a adesão da terapia vocal e perfil de atividades vocais em pacientes disfônicos comportamentais. **CoDAS**, v. 28, n. 5, p. 595–601, 2016.

HABEHH, H.; GOHEL, S. Machine Learning in Healthcare. **Curr Genomics**, v. 22, n. 4, p. 291-300, 2021.

OUHMIDA, Asmae; RAIHANI, Abdelhadi; CHERRADI, Bouchaib; TERRADA, Oumaima. A novel approach for Parkinson's disease detection based on voice classification and features selection techniques. **International Journal of Online Engineering**, v. 17, n.

JOHN, George H.; LANGLEY, Pat. Estimating continuous distributions in Bayesian classifiers. **Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence**, p. 338–345, 1995.

KAUR, P. **A comparison of machine learning classifiers for use on historical record linkage**. 2020. Tese (Master of Science in Computer Science) – University of Guelph, Ontario, Canadá, 2020.

KEMPSTER, G. B.; GERRATT, B. R.; VERDOLINI ABBOTT, K.; BARKMEIERKRAEMER, J.; HILLMAN, R. E. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. **Am J Speech Lang Pathol**, v. 18, p.124-132, 2009.

KOPF, L. M.; GRAETZER, S.; HUH, J. Videos Influence Behavior Change Measures for Voice and Speech in Individuals with Parkinson's Disease. **Proc Wirel Health**, v. 18, 2015.

KREBS, P.; NORCROSS, J. C.; NICHOLSON, J. M.; PROCHASKA, J. O. Stages of change and psychotherapy outcomes: A review and meta-analysis. **Journal of Clinical Psychology**, v. 74, n. 11, p. 1964–1979, 2018.

LANTZ, Brett. **Machine learning with R: expert techniques for predictive modeling**. 4th ed. Birmingham: Packt Publishing, 2023.

LAPLANTE-LÉVESQUE, A.; BRÄNNSTRÖM, K. J.; INGO, E.; ANDERSSON, G.; LUNNER, T. Stages of change in adults who have failed an online hearing screening. **Ear and Hearing**, v. 36, n. 1, p. 92-101, 2015.

LEE, Y., KIM, G., SOHN, K. T., LEE, B. J., LEE, J. C., & KWON, S. B. The Usefulness of Auditory Perceptual Assessment and Acoustic Analysis as a Screening Test for Voice Problems. **Folia Phoniatr Logop**, v. 73, n.1, p. 34-41, 2021.

LEITE, D. R. A.; MORAES, R. M. de; LOPES, L. W. Different performances of machine learning models to classify dysphonic and non-dysphonic voices. **Journal of Voice**, v. 39, n. 3, p. 577–590, 2025.

LIMA, N. T. X.; LOPES, M. R.; SOUZA, M. J. S.; REIS, I. A.; REINALDO, A. M. dos S.; RABELO, J. L., MOTIVAÇÃO PARA MUDANÇA NO CONSUMO DE BEBIDAS ALCOÓLICAS: INTERVENÇÃO BREVE COMO ESTRATÉGIA MOTIVACIONAL. **Cogitare Enferm.**, v. 27, e84261, 2022.

LOPES, Diana Filipa Rodrigues. **KNN com relevância**: uma extensão do classificador de Machine Learning KNN tradicional. 2022. Dissertação (Mestrado em Estatística para Ciência de Dados) – Escola de Ciências, Universidade do Minho, Braga, 2022.

LOPES, L. W.; CAVALCANTE, D. P.; COSTA, P. O. da. Severity of voice disorders: integration of perceptual and acoustic data in dysphonic patients. **CoDAS**, São Paulo, v. 26, n. 5, p. 382-388, set. 2014.

LOPES, L. W.; GODOY, J. F.; FRANÇA, F. P.; PAZ, K. E. S.; ALMEIDA, A. A. Avaliação instrumental da voz. In: PERNAMBUCO, Leandro; ASSENÇO, Ana Manhãni (Org.). **Fonoaudiologia**: avaliação e diagnóstico. 1. ed. Rio de Janeiro: Thieme Revinter, 2021. v. 1, p. 159-182.

LOPES, L.; VIEIRA, V.; BEHLAU, M. Performance of different acoustic measures to discriminate individuals with and without voice disorders. **Journal of Voice**, v. 36, n. 4, p. 487-98, 2022.

LOPES, Leonardo Wanderley; SOUSA, Eliezer Silva da Silva; SILVA, Andreia Cristina Fernandes da; SILVA, Isabela Martins da; PAIVA, Maurício André Araújo de; VIEIRA, Valéria Janayna Dantas; et al. Medidas cepstrais na avaliação da intensidade do desvio vocal. **CoDAS**, v. 31, n. 4, e20180175, 2019.

LUDERMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, São Paulo, Brasil, v. 35, n. 101, p. 85–94, 2021.

LUZ, C. R.; FARIAS, F. B. Q.; NEVES, J. G.; MACEDO, A. B.; CABRINI, D. Avaliação dos estágios de comportamento alimentar de indivíduos com obesidade em ambulatório da atenção especializada. **Comunicação em Ciências da Saúde**, [S. l.], v. 32, n. Suppl1, 2021.

MARTINS, L. K. G.; MENDES, A. L. F.; OLIVEIRA, P.; ALMEIDA, A. A. Distúrbio de voz e fatores de risco em profissionais da voz falada: uma revisão integrativa. **Audiology - Communication Research**, v. 29, e2809, 2024.

MCCONNAUGHY, E. A.; PROCHASKA, J. O.; VELICER, W. E. Stages of change in psychotherapy: measurement and sample profiles. **Psychotherapy: Theory, Research and Practice**, v. 20, p. 368-375, 1983.

MCNEMAR, Quinn. Note on the sampling error of the difference between correlated proportions or percentages. **Psychometrika**, v. 12, n. 2, p. 153–157, 1947.

MIOTTO, R.; WANG, F.; WANG, S.; JIANG, X.; DUDLEY, J. T. Deep learning for healthcare: review, opportunities and challenges. **Brief Bioinform**, v. 19, n. 6, p. 1236-1246, 2018.

MITCHELL, T. M. **Machine learning**. New York: McGraw-Hill, 1997.

MOLNAR, C. **Interpretable machine learning**: a guide for making black box models explainable. [S.l.]: self-published, 2022. Disponível em: <https://christophm.github.io/interpretable-ml-book/>. Acesso em: 7 jul. 2025.

PAIVA, G. M.; SILVA, P. O. C.; SILVA, L. J. A. da; NASCIMENTO, K. A.; SILVA, A. B. da V. e; ABREU, S. R. de, et al. Medidas espectrais e cepstrais em mulheres com disfonia comportamental. **CoDAS**, v. 36, n. 1, e20220327, 2024.

PATEL, R. R.; AWAN, S. N.; BARKMEIER-KRAEMER, J.; COUREY, M.; DELISKY, D.; EADIE, T.; PAUL, D.; SVEC, J. G.; HILLMAN, R. Recommended protocols for instrumental assessment of voice: American Speech Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. **American Journal of Speech-Language Pathology**, v. 25, p. 1-19, 2018.

PAYTEN, C. L.; CHIAPELLO, G.; WEIR, K. A.; MADILL, C. J. Frameworks, Terminology and Definitions Used for the Classification of Voice Disorders: A Scoping Review. **Journal of Voice: Official Journal of the Voice Foundation**, v. 36, n. 5, p. S0892-1997, 2022.

PAZ, K. E. da S.; PAIVA, M. A. A. de; LIMA, D. O.; RIBEIRO, V. V.; MORAES, R. M. de; LOPES, L. W. Treinamento para análise perceptivo-auditiva da voz: revisão de escopo. **Audiol, Commun Res**, v. 28, e2768, 2023.

PEDERSEN, P. V.; KJØLLER, M.; EKHOLM, O.; GRØNBAEK, M.; CURTIS, T. Readiness to change level of physical activity in leisure time among physically inactive Danish adults. **Scand J Public Health**, v. 37, n. 8, p. 785-792, nov. 2009.

PROCHASKA, J. O.; DICLEMENTE, C. C. Transtheoretical therapy: Toward a more integrative model of change. **Psychotherapy: Theory, Research & Practice**, v. 19, p. 276-288, 1982.

PROCHASKA, J. O.; DICLEMENTE, C. C.; VELICER, W. F.; ROSSI, J. S. Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. **Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association**, v. 12, n. 5, p. 399-405, 1993.

PROCHASKA, J. O.; NORCROSS, J. C.; DICLEMENTE, C. C. Applying the stages of change. **Psychotherapy in Australia**, v. 19, n. 2, p. 10-15, 2013.

PROCHASKA, J. O.; NORCROSS, J. C.; FOWLER, J. L.; FOLLICK, M. J.; ABRAMS, D. B. Attendance and outcome in a worksite weight control program: Processes and stages of change as process and predictor variables. **Addictive Behaviors**, v. 17, n. 1, p. 35-45, 1992.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2025.

REZENDE, S. O. **Sistemas inteligentes**. 1. ed. São Paulo: Manole, 2003.

ROCHA, Bianca Silva da. **O engajamento no tratamento de poliusuários de substâncias por meio de um serviço de telessaúde**. Dissertação (Mestrado em Ciências da Saúde) – Programa de Pós-Graduação em Ciências da Saúde, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, 2020.

SANTIAGO, J. C. dos S.; MOREIRA, T. M. M.; FLORÊNCIO, R. S.; BORGES, J. W. P.; PESSOA, V. L. M. de P.; SOUZA, A. C. C. de. Efetividade tecnológica na prontidão para mudança comportamental em hipertensão e excesso ponderal. **Acta Paulista de Enfermagem**, v. 34, eAPE00395, 2021.

SILVA, Bruna Karolina Pereira da. **Corpo e voz: relação entre postura corporal inadequada e suas implicações na voz**. 2021. Trabalho de Conclusão de Curso (Bacharel em Fonoaudiologia) - Escola de Ciências Sociais e da Saúde, Pontifícia Universidade Católica de Goiás, Goiânia, 2021.

SILVA, F. M.; LUZ, C. R. A. N. Avaliação de indivíduos diabéticos quanto ao autocuidado e ao estágio de mudança de comportamento segundo o Modelo Transteórico na Atenção Secundária do Distrito Federal. **Brasília Med**, v. 59, p. 1-11, 2022.

SILVA, Hêmmilly Farias da. **Etapas finais do processo de validação da escala URICA-VV: responsabilidade e relação com outras variáveis da avaliação multidimensional da voz**. 2022. Tese (Doutorado em Modelos de Decisão e Saúde) - Programa de Pós-Graduação em Modelos de Decisão e Saúde, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa - PB, Brasil. 2022.

SILVA, R. Vieira da; OLIVEIRA, I. R. de; VELASQUEZ, M. Lopes. Estágios de prontidão para mudança e escala de ansiedade para tratamento em sobrepeso e obesidade: propriedades psicométricas (SOCRATES-OO). **Journal of Clinical Psychology in Medical Settings**, v. 27, p. 805-817, 2020.

SOARES, J.; VARGAS, D. de. Group Brief Intervention: effectiveness in motivation to change alcohol intake. **Rev Bras Enferm**, v. 73, n. 1, e20180138, 2020.

SOARES, Waldirene Rodrigues de Souza. **Ansiedade e prontidão para mudança de comportamento em indivíduos com obesidade, usuários da Atenção Primária no interior de Minas Gerais**. 2022. 122 p. Dissertação (Mestrado em Ciências da Nutrição) – Programa de Pós-Graduação em Ciências da Nutrição, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2022.

SHETTY, Roopashri; GEETHA, M.; ACHARYA, Dinesh U.; GURUVARE, Shyamala. Data Preprocessing and Finding Optimal Value of K for KNN Model. **Advances in Intelligent Systems and Computing**, 2022.

SHIN, D.; CHO, W. K.; PARK, C. H. K.; RHEE, S. Y.; KIM, M. J.; LEE, H.; KIM, N.-Y.; AHN, Y. M. Detection of minor and major depression through voice as a biomarker using machine learning. **Journal of Clinical Medicine**, v. 10, n. 14, p. 3046, 2021.

SULICA, Lucian. Laryngoscopy, stroboscopy and other tools for the evaluation of voice disorders. **Otolaryngologic Clinics of North America**, v. 46, n. 1, p. 21-30, 2013.

SUPPA, A.; COSTANTINI, G.; ASCI, F.; DI LEO, P.; AL-WARDAT, M. S.; DI LAZZARO, G.; SCALISE, S.; PISANI, A.; SAGGIO, G. Voice in Parkinson's Disease: A Machine Learning Study. **Frontiers in Neurology**, v. 13, p. 831428, 2022.

SUTTON, K.; LOGUE, E.; JARJOURA, D.; BAUGHMAN, K.; SMUCKER, W.; CAPERS, C. Assessing dietary and exercise stage of change to optimize weight loss interventions. **Obesity Research**, v. 11, n. 5, p. 641-652, maio 2003. DOI: 10.1038/oby.2003.92. PMID: 12740454.

TEIXEIRA, L. C.; RODRIGUES, A. L. V.; SILVA, Á. F. G. da; AZEVEDO, R.; GAMA, A. C. C.; BEHLAU, M. Escala URICA-VOZ para identificação de estágios de adesão ao tratamento de voz. **CoDAS**, v. 25, n. 1, p. 8–15, 2013.

VASQUEZ, C. R.; GARRIDO, M. E. L. Transitando desde las fases de cambio de comportamiento hacia el automanejo de personas con condiciones crónicas. **Enfermería Actual de Costa Rica, San José**, n. 43, 51438, 2022.

VILELA JUNIOR, Guanís de Barros; LIMA, Bráulio Nascimento; PEREIRA, Adriano Almeida; RODRIGUES, Marcelo Francisco; OLIVEIRA, José Ricardo Lourenço de; SILIO, Luís Felipe; CARVALHO, Anderson dos Santos; FERREIRA, Heros Ribeiro; PASSOS, Ricardo Pablo. Métricas utilizadas para avaliar a eficiência de classificadores em algoritmos inteligentes. **Revista CPAQV - Centro de Pesquisas Avançadas em Qualidade de Vida**, [S. l.], v. 14, n. 2, 2022.

VITAL H.R.M.C., LIMA-SILVA M.F.B., ALMEIDA L.N.A., ALMEIDA A.A.F. Sintomas vocais auditivos e proprioceptivos pré e pós-terapia de grupo de pacientes com disfonia. **Revista CEFAC**, v.58051, p.900. 2016.

XING, W.; DU, D. Dropout prediction in MOOCs: using deep learning for personalized intervention. **Journal of Educational Computing Research**, v. 57, n. 4, p. 974–996, mar. 2018.

YAGNAVAJJULA, M. K.; MITTAPALLE, K. R.; ALKU, P.; RAO, S. K.; MITRA, P. Automatic classification of neurological voice disorders using wavelet scattering features. **Speech Communication**, [S.l.], v. 157, p. 103040, 2024.

YILMAZ, Yetkin Zeki; UĞURLAR, Müge; YILMAZ, Begüm Bahar; GÜLMEZ, Züleyha Dilek; ÖZDOĞAN, Hasan Ahmet; ATAŞ, Ahmet; BATIOĞLU-KARAALTIN, Ayşegül. The Comparison of Narrow Band Imaging, White Light Laryngoscopy and Videolaryngostroboscopy in the Evaluation of Benign Vocal Fold Lesions. **Journal of Voice**, v. 37, n. 2, p. 275-281, mar. 2023.

ANEXO 1

PROTOCOLO DE ANAMNESE E AVALIAÇÃO VOCAL

(Anna Alice Almeida; Leonardo Lopes; Alexandra Christine de Aguiar; Priscila Oliveira; Mara Behlau, 2022)

DATA: ____/____/____ Supervisor(a): _____
Responsável pela coleta(a): _____

() Disfonia funcional | () Disfonia organofuncional | () Disfonia orgânica

I) Identificação Pessoal

Nome: _____ Idade: _____
D.N.: ____/____/____ Local de nascimento: _____ UF: _____
Sexo: F () M () Estado conjugal: _____ Grau de instrução: _____
Profissão: _____ Período de trabalho: _____ Carga horária: _____
Endereço: _____

Contato (telefone / e-mail): _____
Encaminhado por: _____ Tel.: _____
Informante: _____ Parentesco: _____

II) Queixa e duração

1) Motivo da consulta/duração: _____

2) História pregressa da disfonia
a) Como ocorreu o início do problema da voz (brusco, gradual)? _____

b) De quem foi a sugestão de procurar o nosso serviço? _____

c) Qual o impacto da disfonia na comunicação/sociabilização (pessoal e profissional)? _____

d) O que conseguia fazer antes com a voz que não consegue agora?

3) Sintomas vocais (Nº = _____)

Auditivos (Nº = _____)

- | | | |
|--|---|--|
| <input type="checkbox"/> rouquidão | <input type="checkbox"/> voz monótona | <input type="checkbox"/> instabilidade na voz |
| <input type="checkbox"/> voz muda depois do uso | <input type="checkbox"/> dificuldade para agudos | <input type="checkbox"/> dificuldade para graves |
| <input type="checkbox"/> dificuldade em projetar a voz | <input type="checkbox"/> dificuldade de falar baixo | <input type="checkbox"/> falhas na voz |
| <input type="checkbox"/> mudança vocal no mesmo dia | <input type="checkbox"/> presença de ar na voz | <input type="checkbox"/> perda da voz constante |

Sensoriais / Cinestésicos (Nº = _____)

- | | | |
|---|---|---|
| <input type="checkbox"/> fadiga ao falar | <input type="checkbox"/> desconforto ao falar | <input type="checkbox"/> esforço para falar |
| <input type="checkbox"/> "bolo" na garganta | <input type="checkbox"/> garganta seca | <input type="checkbox"/> dor na garganta |
| <input type="checkbox"/> tensão no pescoço | <input type="checkbox"/> pigarro | <input type="checkbox"/> tosse improdutiva |
| <input type="checkbox"/> formação de muco | <input type="checkbox"/> gosto ácido na boca | <input type="checkbox"/> dor para engolir |

Os sintomas relatados acima pioram no final do dia/semana? () sim () não

4) Fatores de risco (Nº = _____)

a) Organizacionais (Nº = _____)

- | | | |
|--|--|--|
| <input type="checkbox"/> Jornada de trabalho longa | <input type="checkbox"/> Acúmulo de atividades | <input type="checkbox"/> Demanda vocal excessiva |
| <input type="checkbox"/> Alto número de ouvintes | <input type="checkbox"/> Tempo de serviço | |

b) Ambientais (Nº = _____)

- | | | |
|--|---|--|
| <input type="checkbox"/> Ruído de fundo | <input type="checkbox"/> Acústica pobre | <input type="checkbox"/> Distância interfalantes |
| <input type="checkbox"/> Baixa umidade do ar | <input type="checkbox"/> Poluição | <input type="checkbox"/> Poeira e mofo |
| <input type="checkbox"/> Fatores ergonômicos | <input type="checkbox"/> Ambiente estressante | <input type="checkbox"/> Equipamento inadequado |

c) Pessoais (Nº = _____)

- | | | |
|---|---|---|
| <input type="checkbox"/> Fuma | <input type="checkbox"/> Bebe | <input type="checkbox"/> Usa drogas |
| <input type="checkbox"/> Fala muito | <input type="checkbox"/> Fala alto | <input type="checkbox"/> Fala rápido |
| <input type="checkbox"/> Fala muito ao telefone | <input type="checkbox"/> Fala com esforço | <input type="checkbox"/> Fala agudo/grave demais |
| <input type="checkbox"/> Fala acima do ruído | <input type="checkbox"/> Fala em público | <input type="checkbox"/> Imita (atores, cantores) |
| <input type="checkbox"/> Grita com frequência | <input type="checkbox"/> Torce com frequência | <input type="checkbox"/> Canta fora do tom |
| <input type="checkbox"/> Vida social intensa | <input type="checkbox"/> Tosse constante | <input type="checkbox"/> Hidratação insuficiente |
| <input type="checkbox"/> Automedicação | <input type="checkbox"/> Repouso inadequado | <input type="checkbox"/> Alimentação inadequada |

III) Tratamentos anteriores para disfonia (medicamentoso, cirúrgico, fonoterápico, outros):

IV) Investigação complementar (distúrbios alérgicos, faríngeos, bucais, nasais, otológicos, pulmonares, digestivos, hormonais e neurovegetativos): _____

1) Submeteu-se a alguma cirurgia? Qual?(ias)? Data(s)? Houve complicações decorrentes do ato cirúrgico? _____

2) Apresenta história de patologia laríngea (tumor, obstrução, inflamação, trauma, etc..)? Especificar abordando a data e cuidados dispensados. _____

3) Sofreu algum ferimento, especialmente na área do nariz e da garganta? Especifique a natureza, data e extensão? _____

4) Houve algo incomum no período da mudança de voz na puberdade? (referir sobre o acontecimento e indicar a idade na ocasião)? _____

5) Há alguma menção considerada importante sobre o seu desenvolvimento? (desde o período da gestação até o presente momento) _____

V) Antecedentes familiares

VI) Motivação para a terapia e execução das condutas em casa

VII) Outras observações

ANEXO 2

ESCALA URICA-Vv

(Aguilar, Almeida, 2020)

Queremos conhecer sua opinião sobre como você lida com sua voz. Abaixo estão escritas 25 frases que as pessoas usam sobre isso. Por favor, leia com atenção e marque o quanto você discorda ou concorda com as afirmações. Não existem respostas certas ou erradas; as afirmações abaixo apenas refletem modos diferentes de agir.

Use: DT=Discordo Totalmente; D=Discordo; NS=Não sei; C=Concordo; CT=Concordo totalmente

1.	Eu acho que estou pronto para melhorar minha voz.	DT	D	NS	C	CT
2.	O meu problema de voz me incomoda e eu estou tentando resolvê-lo.	DT	D	NS	C	CT
3.	Acho que vale a pena cuidar da minha voz.	DT	D	NS	C	CT
4.	Fico preocupado em ter um novo problema de voz, por isto estou procurando ajuda.	DT	D	NS	C	CT
5.	Já estou tratando o meu problema de voz.	DT	D	NS	C	CT
6.	Eu acho que quero mudar a minha voz.	DT	D	NS	C	CT
7.	Tenho tido sucesso no meu tratamento de voz, mas não tenho certeza se consigo mantê-la boa sem ajuda.	DT	D	NS	C	CT
8.	Às vezes é difícil tratar a minha voz, mas estou me dedicando para isto.	DT	D	NS	C	CT
9.	Eu espero compreender melhor o meu problema de voz.	DT	D	NS	C	CT
10.	Estou me dedicando bastante para melhorar o meu problema de voz.	DT	D	NS	C	CT
11.	Eu tenho um problema de voz e estou certo de que vou resolvê-lo.	DT	D	NS	C	CT
12.	Não estou conseguindo manter minha voz "boa" e quero evitar um novo problema.	DT	D	NS	C	CT
13.	Ainda que minha voz não esteja boa o tempo todo, estou me dedicando para melhorá-la.	DT	D	NS	C	CT
14.	Achei que depois de tratar a voz eu me livraria deste problema, mas algumas vezes ele ainda me incomoda.	DT	D	NS	C	CT
15.	Eu gostaria de saber mais como melhorar minha voz.	DT	D	NS	C	CT
16.	Eu comecei a tratar a minha voz, mas preciso de mais ajuda.	DT	D	NS	C	CT
17.	Talvez um fonoaudiólogo ou algum tratamento possa ajudar a resolver meu problema de voz.	DT	D	NS	C	CT
18.	Eu preciso de um incentivo para manter o que consegui mudar na minha voz.	DT	D	NS	C	CT
19.	Talvez eu seja responsável por parte do meu problema de voz, mas não sou o único responsável.	DT	D	NS	C	CT
20.	Eu tenho esperança de que alguém me ajude a melhorar a minha voz.	DT	D	NS	C	CT
21.	Eu já estou fazendo a minha parte para melhorar a minha voz.	DT	D	NS	C	CT
22.	Estou me esforçando muito para não ter uma recaída no meu problema de voz.	DT	D	NS	C	CT
23.	É frustrante, mas eu sinto que minha voz está piorando de novo.	DT	D	NS	C	CT
24.	Eu estou trabalhando ativamente para resolver meu problema de voz.	DT	D	NS	C	CT
25.	Depois de tudo o que fiz para melhorar a minha voz, ela às vezes, ainda me preocupa.	DT	D	NS	C	CT
Fator 1 – Contemplação: itens 1, 2, 3, 6, 9, 10, 11, 13, 15, 17, 19 e 20 sendo o item 15 o mais representativo deste fator						
Fator 2 – Manutenção: itens 4, 5, 7, 8, 12, 14, 16, 18, 21, 22, 23, 24 e 25, teve o item 14 como mais representativo						

ANEXO 3

CENTRO DE CIÊNCIAS DA
SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA -
CCS/UFPB

**PARECER CONSUBSTANCIADO DO CEP****DADOS DO PROJETO DE PESQUISA**

Título da Pesquisa: Responsividade de instrumentos de autoavaliação de voz validados para o português brasileiro com base na Teoria de Resposta ao Item (TRI)

Pesquisador: Anna Alice Figueiredo de Almeida

Área Temática:

Versão: 3

CAAE: 46601421.2.0000.5188

Instituição Proponente: Universidade Federal da Paraíba

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.866.871

Apresentação do Projeto:

Responsável pelo projeto Anna Alice Figueiredo de Almeida. A pesquisa será desenvolvida inicialmente na modalidade remota utilizando um formulário do Google Forms e reunião via Zoom Video Communications. Em um segundo momento, será realizada a coleta de dados na modalidade presencial realizada no Laboratório Integrado de Estudos da Voz (LIEV) do Departamento de Fonoaudiologia da Universidade Federal da Paraíba (UFPB), Campus I, município de João Pessoa-PB, Brasil. Pretende-se desenvolvê-la no período de junho de 2021 a Fevereiro de 2025. Estudo 1: a pesquisa possui delineamento observacional transversal Estudo 2: Consiste em um estudo de intervenção, pois haverá interferência do pesquisador na amostra estudada. Estudo 1: Participarão destes estudos pessoas com e sem disfonia, adultos, de ambos os sexos de todo o Brasil. Os participantes não poderão ter realizado nenhuma intervenção que influenciará na voz. Este estudo ocorrerá em ambiente virtual. Todos os participantes terão sua voz gravada a partir de duas tarefas de fala (vogal sustentada // e fala encadeada a partir da contagem de números de 1 a 10) e responderão os instrumentos de autoavaliação da voz validados para o português brasileiro a partir da TRI, a saber: Índice de Desvantagem Vocal (IDV), Questionário de Qualidade de Vida em Voz (QVV), Escala de Sintomas Vocais (ESV), Escala de Desconforto do Trato Vocal (EDTV), Protocolo de Estratégias de

Endereço: Prédio da Reitoria da UFPB 2º 1º Andar

Bairro: Cidade Universitária

CEP: 58.051-900

UF: PB

Município: JOAO PESSOA

Telefone: (83)3216-7791

Fax: (83)3216-7791

E-mail: comitedeetica@ccs.ufpb.br

**CENTRO DE CIÊNCIAS DA
SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA -
CCS/UFPB**



Continuação do Parecer: 4.866.871

Enfrentamento na Disfonia (PEED), Escala URICA-Voz. validada (URICA-VV) e a Escala de Controle Percebido no Presente sobre a Voz (CPPV). Os dados coletados serão migrados para um banco de dados digital que será composto por dados pessoais sociodemográficos, bem como todos os itens de cada um dos protocolos que serão utilizados. Posteriormente, serão realizadas as devidas análises dos dados para verificar a acurácia dos instrumentos para discriminar indivíduos com e sem disfonia.

O Estudo 2 será composto por pacientes adultos, de ambos os sexos, com diagnóstico de disfonia que procurarão voluntariamente a intervenção fonoaudiológica na Clínica Escola de Fonoaudiologia/UFPB. Esses serão alocados aleatoriamente em dois grupos: o grupo com disfonia que recebeu fonoterapia (GDF) e grupo com disfonia que estará na fila de espera (GDE). Todos passarão pelas mesmas fases de coleta de tarefas de fala e instrumentos apontados no Estudo 1. Neste estudo, passarão pela coleta pré e pós fonoterapia de voz que ocorrerá em seis sessões. A intervenção abordará princípios da terapia direta e indireta com foco na melhora da qualidade vocal. Os dados coletados serão inseridos em um banco de dados digital que será composto por dados pessoais sociodemográficos, bem como todos os itens de cada um dos instrumentos que serão utilizados pré e pós fonoterapia. Posteriormente, serão realizadas as devidas análises dos dados para verificar a acurácia dos instrumentos para monitorar o paciente com disfonia que foi submetido à fonoterapia na área de voz. Inicialmente as amostras vocais serão analisadas de forma perceptivo-auditiva por fonoaudiólogos

especialistas em voz, para esta avaliação será utilizada a Escala Analógica Visual (EAV). Em seguida todos os dados serão inseridos em uma planilha de dados digital para realização da análise estatística descritiva e inferencial. Na análise estatística descritiva, será obtido informações sintéticas sobre os dados como média, desvio padrão, valores máximo e mínimo. Na análise estatística inferencial, inicialmente será investigada a normalidade dos dados através do teste de Shapiro-wilk. Caso esse pressuposto seja atendido, será realizado testes comparação a partir do teste T de Student, investigação do tamanho do efeito e a análise de Regressão Logística.

Investigar a evidência de validade baseada na consequência do teste de instrumentos de autoavaliação da voz validados para o português brasileiro a partir da TRI no processo de avaliação e monitoramento de pacientes com disfonia

Objetivo da Pesquisa:

Investigar a evidência de validade baseada na consequência do teste de instrumentos de

Endereço: Prédio da Reitoria da UFPB - 1º Andar

Bairro: Cidade Universitária

CEP: 58.051-900

UF: PB

Município: JOAO PESSOA

Telefone: (83)3216-7791

Fax: (83)3216-7791

E-mail: comitedeetica@ccs.ufpb.br

**CENTRO DE CIÊNCIAS DA
SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA -
CCS/UFPB**



Continuação do Parecer: 4.866.871

autoavaliação da voz validados para o português brasileiro a partir da TRI no processo de avaliação e monitoramento de pacientes com disfonia

analisar o poder discriminativo do instrumento para as disfonias comportamentais.

Investigar a eficácia dos instrumentos: Índice de Desvantagem Vocal (IDV), Questionário de Qualidade de Vida em Voz (QVV), Escala de Sintomas Vocais (ESV), Escala de Desconforto do Trato Vocal (EDTV), Protocolo de Estratégias de Enfrentamento na Disfonia (PEED), Escala URICA-Voz validada (URICA-VV) e a Escala de Controle Percebido no Presente sobre a Voz (CPPV) nas versões validadas com base na TRI; Analisar a habilidade desses instrumentos detectarem mudanças ao longo do tempo nos constructos a serem medidos. Avaliar a responsividade desses instrumentos para avaliação e monitoramento de indivíduos disfônicos;

investigar a relação desses instrumentos com a intensidade do desvio vocal.

Avaliação dos Riscos e Benefícios:

Pode ocorrer desconfortos ao responder os questionários, como arrependimento ou constrangimento. Será conferido ao participante o direito de expor toda e qualquer insatisfação caso sinta-se desconfortável. Neste caso, o pesquisador deverá adotar a conduta de interrupção total da pesquisa com o mesmo, a fim de evitar qualquer tipo de prejuízo ao participante.

Após toda a coleta dos dados, será fornecido por e-mail um breve relatório sobre as informações da sua avaliação vocal, bem como uma cartilha com orientações básicas de cuidados vocais tendo em vista que todos os participantes. Além disso, o LIEV coloca-se disponível para receber os voluntários que acreditem que precisem de atendimento fonoaudiológico na área de voz.

Comentários e Considerações sobre a Pesquisa:

Em relação à avaliação do binômio risco-benefício, o projeto considera o grau de vulnerabilidade dos participantes e apresenta as medidas protetoras cabíveis. TCLE foi re escrito e ficou mais fácil de ser compreendido.

A pesquisa Garante os direitos fundamentais do sujeito de pesquisa (informação, privacidade,

Endereço: Prédio da Reitoria da UFPB - 1º Andar

Bairro: Cidade Universitária

CEP: 58.051-900

UF: PB

Município: JOAO PESSOA

Telefone: (83)3216-7791

Fax: (83)3216-7791

E-mail: comitedeetica@ccs.ufpb.br

**CENTRO DE CIÊNCIAS DA
SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA -
CCS/UFPB**



Continuação do Parecer: 4.866.871

recusa inócua, desistência, indenização, ressarcimento, acesso ao pesquisador e ao CEP). O mesmo também permite aos participantes parcialmente conhecimento dos objetivos do estudo, dos procedimentos de coleta, dos riscos possíveis e dos benefícios que o estudo poderá trazer. A metodologia foi adequada aos objetivos pretendidos. Quanto aos instrumentos de coleta, os mesmos foram devidamente apresentados no projeto ajudando a compreender as variáveis que serão contempladas. Em relação à metodologia de análise dos dados a mesma se encontra bem definida quantos aos procedimentos cabíveis. Quanto aos instrumentos de coleta, os mesmos foram devidamente apresentados no projeto ajudando a compreender as variáveis que serão contempladas. Em relação à metodologia de análise dos dados a mesma se encontra bem definida quantos aos procedimentos cabíveis.

Considerações sobre os Termos de apresentação obrigatória:

Termos apresentados e pendências resolvidas

Conclusões ou Pendências e Lista de Inadequações:

Projeto de pesquisa em acordo com a resolução 466/12 e portanto aprovado

Considerações Finais a critério do CEP:

Certifico que o Comitê de Ética em Pesquisa do Centro de Ciências da Saúde da Universidade Federal da Paraíba – CEP/CCS aprovou a execução do referido projeto de pesquisa. Outrossim, informo que a autorização para posterior publicação fica condicionada à submissão do Relatório Final na Plataforma Brasil, via Notificação, para fins de apreciação e aprovação por este egrégio Comitê.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações	PB_INFORMAÇÕES_BÁSICAS_DO_P	06/07/2021		Aceito

Endereço: Prédio da Reitoria da UFPB - 1º Andar

Bairro: Cidade Universitária

CEP: 58.051-900

UF: PB

Município: JOAO PESSOA

Telefone: (83)3216-7791

Fax: (83)3216-7791

E-mail: comitedeetica@ccs.ufpb.br

**CENTRO DE CIÊNCIAS DA
SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA -
CCS/UFPB**



Continuação do Parecer: 4.866.871

Básicas do Projeto	ETO_1749758.pdf	11:25:30		Aceito
Declaração de Instituição e Infraestrutura	CERTIDAO_DepFono_V2.pdf	06/07/2021 11:24:59	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Anuencia_Clinica_Escola_V2.pdf	06/07/2021 11:24:47	Anna Alice Figueiredo de Almeida	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_V3.pdf	06/07/2021 11:23:36	Anna Alice Figueiredo de Almeida	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_CEP_V3.pdf	06/07/2021 11:23:16	Anna Alice Figueiredo de Almeida	Aceito
Outros	Carta_de_respostaCEP.pdf	04/06/2021 16:17:53	Anna Alice Figueiredo de Almeida	Aceito
Declaração de concordância	anuencia_LIEV.pdf	07/05/2021 15:46:42	Anna Alice Figueiredo de Almeida	Aceito
Folha de Rosto	folhaDeRosto_CEP_AAFAassinada.PDF	07/05/2021 15:44:53	Anna Alice Figueiredo de Almeida	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

JOAO PESSOA, 26 de Julho de 2021

Assinado por:
Eliane Marques Duarte de Sousa
(Coordenador(a))

Endereço: Prédio da Reitoria da UFPB - 1º Andar

Bairro: Cidade Universitária

CEP: 58.051-900

UF: PB

Município: JOAO PESSOA

Telefone: (83)3216-7791

Fax: (83)3216-7791

E-mail: comitedeetica@ccs.ufpb.br

APÊNDICE

```
##%#####%##
#
####          REGRESSÃO LOGISTICA          ####
#
##%#####%##
# Carregue as bibliotecas
library(DMwR2)
library(caTools)
library(dplyr)
library(caret)
library(readxl)

# Divisão treino e teste
set.seed(123)
training.samples <- createDataPartition(dados$Estagio.Prontidao, p = 0.7, list = FALSE)
train_a <- dados[training.samples, ]
test_a <- dados[-training.samples, ]
str(train_a)
str(test_a)

#### Classificação ####

### Pré-processamento dos dados ###

# Imputação KNN
set.seed(123)
train_imp <- knnImputation(train_a, k = 5)
test <- knnImputation(test_a, k = 5, distData = train_a)

colSums(is.na(train_imp))
colSums(is.na(test))

# Aplicar o oversampling no conjunto de treino
train <- upSample(x = train_imp[, -which(names(train_imp) == "Estagio.Prontidao")],
                 y = train_imp$Estagio.Prontidao)

#### Classificação ####

### Pré-processamento dos dados ###

# Verificando as variáveis
str(train)

## Verificar se há NA nos dados
colSums(is.na(train))
colSums(is.na(test))

# Modelo 1
mod1 <- glm(Class ~ ., data = train, family = binomial(link = "logit"))
summary(mod1)

# Razão de Chances
library(questionr)
odds.ratio(mod1)

### Diagnóstico do Modelo ###
marginalModelPlots(mod1)

### Predições ###
pred <- predict(mod1, test, type = "response")
result <- as.factor(ifelse(pred > 0.5, 1, 0))

### Desempenho do modelo ###

# Matriz de confusão e medidas
library(caret)
confusionMatrix(result, test$Estagio.Prontidao, positive = "1")

# Curva ROC e AUC
```

```

library(pROC)
auc <- roc(test$Estagio.Prontidao, pred)
plot.roc(auc, print.thres = T)

# Usando o novo ponto de corte
result2 <- as.factor(ifelse(pred > 0.5,1,0))
confusionMatrix(result2, test$Estagio.Prontidao, positive = "1")

#####
#                                     #
####                               ####
#                                     #
#####

# Definindo o diretório

#### Classificação ####

# Verificar a codificação das variáveis
str(train)
str(test)

### Ajuste dos Parâmetros ###

# Escolher o melhor k

# Carregando pacotes
# install.packages("caret")
library(caret)
library(Amelia)
library(pROC)

# Definindo a semente
set.seed(123)

# Vamos utilizar uma validação-cruzada 10-fold
ctrl <- trainControl(method = "cv",
                     number = 10)

knnFit <- train(Class ~ .,
               method = "knn",
               preProcess = c("center","scale"), # normalização dos dados
               tuneLength = 20, # testa 20 valores
               trControl = ctrl,
               metric = "Accuracy", # avalia pela acurácia
               data = train)

knnFit #k=5
plot(knnFit) # gráfico do desempenho segundo os valores de k
knnFit$finalModel # melhor modelo

### Predições ###
predknn <- predict(knnFit, test, type = "prob")
resultknn <- as.factor(ifelse(predknn[,2] > 0.5,1,0))

### Desempenho do modelo ###

# Matriz de confusão e medidas
library(caret)
confusionMatrix(resultknn, test$Estagio.Prontidao, positive = "1")

# Curva ROC e AUC
library(pROC)
aucknn <- roc(test$Estagio.Prontidao, predknn[,2])
aucknn
plot.roc(aucknn, print.thres = T)

#####

```



```

#####          NAIVE BAYES          #####
#                                     #
##%#####%##

#### Classificação ####

# Usandos os mesmos dados da Regressão Logística

# Verficar a codificação das variáveis
str(train)
str(test)

### Ajuste dos Parâmetros ###

# Definindo a semente
set.seed(123)

# Vamos utilizar uma validação-cruzada 10-fold
library(caret)
ctrl <- trainControl(method = "repeatedcv",
                      number = 10)

# Gerando as combinações de parâmetros
tuneGrid <- expand.grid(
  usekernel = c(TRUE, FALSE),
  fL = 0:1,
  adjust = 1:5
)

nbFit <- train(Class ~ .,
               method      = "nb",
               preProcess = c("center", "scale"),
               tuneLength = 20,
               trControl  = ctrl,
               tuneGrid   = tuneGrid,
               metric      = "Accuracy",
               data        = train)

```

```

nbFit
plot(nbFit)

### Predições ###

prednb <- predict(nbFit, test, type = "prob")
resultnb <- as.factor(ifelse(prednb[,2] > 0.5,1,0))

### Desempenho do modelo ###

# Matriz de confusão e medidas
library(caret)
confusionMatrix(resultnb, test$Estagio.Prontidao, positive = "1")

# Curva ROC e AUC
library(pROC)
aucnb <- roc(test$Estagio.Prontidao, prednb[,2])
aucnb
plot.roc(aucnb, print.thres = T)

#####%##
#                                     #
####          ÁRVORES DE DECISÃO          ####
#                                     #
#####%##

#### Classificação ####

# Usandos os mesmos dados da Regressão Logística

# Verificar a codificação das variáveis
str(train)
str(test)

### Ajuste dos Parâmetros ###

# Carregando pacotes
# install.packages("caret")
library(caret)
library(Amelia)
library(pROC)

# Definindo a semente
set.seed(123)

# Vamos utilizar uma validação-cruzada 10-fold
ctrl <- trainControl(method = "cv",
                     number = 10)

dtFit <- train(Class ~ .,
               method = "rpart2", # usa profundidade máxima
               tuneLength = 20,
               trControl = ctrl,
               metric = "Accuracy",
               data = train)

```

```

dtFit
plot(dtFit)

# Desenho da árvore
library(rpart.plot)
rpart.plot(dtFit$finalModel,
            cex = 0.5, #tamanho do texto
            extra = 4, #informações extras no nós
            type = 4, #tipo de gráfico
            box.palette = "RdYlGn") #cor

### Predições ###

predtdt <- predict(dtFit, test, type = "prob")
resultdt <- as.factor(ifelse(predtdt[,2] > 0.5,1,0))

### Desempenho do modelo ###

# Matriz de confusão e medidas
library(caret)
confusionMatrix(resultdt, test$Estagio.Prontidao, positive = "1")

# Curva ROC e AUC
library(pROC)
aucdt <- roc(test$Estagio.Prontidao, predtdt[,2])
aucdt
plot.roc(aucdt, print.thres = T)

#####%##
#                                     #
####          RANDOM FOREST          ####
#                                     #
#####%##

# Carregando os pacotes
library(randomForest)
library(caret)

##### Random Forest #####

#### Classificação ####

# Verificar a codificação das variáveis
str(train)

```

```

str(test)

### Ajuste dos Parâmetros ###

# Definindo a semente
set.seed(123)

# Ajustar manualmente mtry e ntree
customRF <- list(type = "Classification",
                 library = "randomForest",
                 loop = NULL)

customRF$parameters <- data.frame(parameter = c("mtry", "ntree"),
                                   class = rep("numeric", 2),
                                   label = c("mtry", "ntree"))

customRF$grid <- function(x, y, len = NULL, search = "grid") {}

customRF$fit <- function(x, y, wts, param, lev, last, weights, classProbs) {
  randomForest(x, y,
              mtry = param$mtry,
              ntree=param$ntree)
}

# Predizer a classe
customRF$predict <- function(modelFit, newdata, preProc = NULL, submodels = NULL)
  predict(modelFit, newdata)

# Predizer a probabilidade
customRF$prob <- function(modelFit, newdata, preProc = NULL, submodels = NULL)
  predict(modelFit, newdata, type = "prob")

customRF$sort <- function(x) x[order(x[,1]),]
customRF$levels <- function(x) x$classes

# Vamos utilizar uma validação-cruzada 10-fold
ctrl <- trainControl(method = "cv",
                    number = 10,
                    allowParallel = T)

grid <- expand.grid(.mtry = c(1:18), # número de variáveis
                  .ntree = c(500,1000,1500)) # número de árvores

set.seed(123)
rfFit <- train(Class ~ .,
              method = customRF,
              tuneGrid = grid,
              trControl = ctrl,
              metric = "Accuracy",
              data = train)

rfFit
plot(rfFit)
plot(rfFit$finalModel) # erro OOB
legend("topright", colnames(rfFit$finalModel$serr.rate),
      col = 1:3,
      cex = 0.8,
      fill = 1:3)
summary(rfFit)

rf <- randomForest(Class ~ ., data = train,
                  importance = T, mtry = 2, ntree = 500)
rf
rfFit$finalModel$xNames

### Predições ###

predrf <- predict(rfFit, test, type = "prob")
resultrf <- as.factor(ifelse(predrf[,2] > 0.5, 1, 0))

```

```
# Matriz de confusão e medidas
library(caret)
confusionMatrix(resultrf, test$Estagio.Prontidao, positive = "1")

# Curva ROC e AUC
library(pROC)
aucrf <- roc(test$Estagio.Prontidao, predrf[,2])
aucrf
plot.roc(aucrf, print.thres = T)
```