



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CARLOS FREDERICO PEREIRA JUNIOR

**CLASSIFICAÇÃO DE ESTADOS AFETIVOS EM
PESSOAS IDOSAS: UMA ABORDAGEM DE
REPRESENTAÇÃO DE DADOS MULTIMODAIS DE EEG E VOZ**

Recife

2025

CARLOS FREDERICO PEREIRA JUNIOR

**CLASSIFICAÇÃO DE ESTADOS AFETIVOS EM
PESSOAS IDOSAS: UMA ABORDAGEM DE
REPRESENTAÇÃO DE DADOS MULTIMODAIS DE EEG E VOZ**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de bacharel em Ciência da Computação.

Orientador (a): Profa. Dra. Máira Araújo de Santana.

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Pereira Júnior, Carlos Frederico.

Classificação de estados afetivos em pessoas idosas: uma abordagem de representação de dados multimodais de EEG e voz / Carlos Frederico Pereira Júnior. - Recife, 2025.

24 p. : il., tab.

Orientador(a): Maíra Araújo de Santana

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Ciências da Computação - Bacharelado, 2025.

Inclui referências.

1. Computação Afetiva. 2. Inteligência Artificial. 3. Aprendizado de Máquina. 4. Reconhecimento das emoções. 5. Sinais EEG. I. Santana, Maíra Araújo de. (Orientação). II. Título.

000 CDD (22.ed.)

CARLOS FREDERICO PEREIRA JUNIOR

**CLASSIFICAÇÃO DE ESTADOS AFETIVOS EM
PESSOAS IDOSAS: UMA ABORDAGEM DE
REPRESENTAÇÃO DE DADOS MULTIMODAIS DE EEG E VOZ**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Ciência da Computação da
Universidade Federal de Pernambuco,
como requisito parcial para obtenção do
título de bacharel em Ciência da
Computação.

Aprovado em: 04/08/2025

BANCA EXAMINADORA

Profª. Dra. Maíra Araújo de Santana (Orientadora)

Universidade Federal de Pernambuco

Prof. Dr. Filipe Carlos de Albuquerque Calegario (Examinador Interno)

Universidade Federal de Pernambuco

CLASSIFICAÇÃO DE ESTADOS AFETIVOS EM PESSOAS IDOSAS: UMA ABORDAGEM DE REPRESENTAÇÃO DE DADOS MULTIMODAIS DE EEG E VOZ¹

Classification of Affective States in Elderly People: An Approach to Representing Multimodal EEG and Voice Data

Carlos Frederico Pereira Junior²

Orientação: Profa. Dra. Máira Araújo de Santana³

RESUMO

Impulsionada pelo avanço da inteligência artificial nos últimos anos, a computação afetiva vem conquistando cada vez mais destaque dentro da comunidade científica. Dentro desse contexto, grande parte das pesquisas dessa área utilizam dados de fontes públicas para investigar, através dos seus estudos, o reconhecimento das emoções. Entretanto, os dados disponíveis publicamente ainda não são capazes de contemplar as particularidades inerentes a diferentes grupos sociais, como pessoas idosas com quadro de demência, dada a limitação de casos existentes nesses corpora. Sendo assim, o estudo proposto utilizou dados de pessoas idosas, com e sem demência, dispostos em 3 modalidades: sinais EEG, sinais de voz e uma abordagem multimodal combinando ambos os tipos. Como forma de investigar a capacidade computacional no reconhecimento dos estados afetivos no grupo exposto, foi utilizada uma extração de atributos de cunho matemático e algoritmos rasos de aprendizado de máquina sobre as três abordagens. Os resultados destacam o desempenho da abordagem multimodal, cuja acurácia obtida alcançou 90,90% na etapa de teste, sinalizando positivamente para cenários de uso em aplicações reais, dentre elas, o tratamento do quadro de demência com uso da musicoterapia. Em seguida, é destacada a abordagem com uso de sinais EEG com 99,48% de acurácia e de voz com 90,41%.

Palavras-chave: computação afetiva; multimodal; EEG; voz; idosos; reconhecimento de emoções

ABSTRACT

Driven by advances in artificial intelligence in recent years, affective computing has been gaining increasing prominence in the scientific community. Within this context, much of the research in this area uses data from public sources to investigate, through their studies, the recognition of emotions. However, publicly available data are not yet capable of encompassing the particularities inherent to different social groups, such as elderly people with dementia, given the limited number of cases in these corpora. Therefore, the proposed study used data from elderly people, with and without dementia, arranged in 3 modalities: EEG signals, voice signals and a multimodal approach combining both types. As a way of investigating the computational capacity in recognizing affective states in the exposed group,

¹ Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal de Pernambuco (UFPE), cuja banca de defesa foi composta pelos seguintes membros: Profa. Dra. Máira Araújo de Santana; Prof. Dr. Filipe Carlos de Albuquerque Calegario, na seguinte data: 04 de agosto de 2025.

² Graduando em Ciência da Computação na UFPE.

³ Professora do Centro de Informática da UFPE

mathematical attribute extraction and shallow machine learning algorithms were used on the three approaches. The results highlight the performance of the multimodal approach, whose accuracy reached 90.90% in the testing stage, signaling positively for use scenarios in real applications, among them, the treatment of dementia using music therapy. Next, the approach using EEG signals with 99.48% accuracy and voice with 90.41% accuracy is highlighted.

Keywords: affective computing; multimodal; EEG; voice; elderly people; recognition of emotions

1 INTRODUÇÃO

A computação afetiva é um campo de pesquisa que busca desenvolver sistemas computacionais capazes de processar e reconhecer emoções humanas dentro do domínio da interação humano-computador (IHC), como apresentado por Picard (1997). Esse segmento vem ganhando cada vez mais visibilidade dentro da comunidade científica, impulsionado pelo avanço da inteligência artificial nos últimos anos e pela diversidade de possibilidades de uso do reconhecimento de emoções para a concepção de aplicações inovadoras para a sociedade. Uma dessas oportunidades está na musicoterapia, uma abordagem terapêutica que utiliza elementos musicais e sonoros para promoção da saúde (Benenson, 1988). Sob essa ótica, o estudo desenvolvido por Santana (2023a), tem como objetivo o desenvolvimento de uma aplicação que permita auxiliar a condução da musicoterapia em pacientes com demência, utilizando o reconhecimento das emoções manifestadas para customização de um tratamento mais adequado para um determinado indivíduo.

Existem estudos na literatura que buscam representar as diferentes formas de expressões emocionais. Um dos mais relevantes nesse campo de pesquisa é o modelo bidimensional proposto por Russel (1980), que define os estados afetivos com base na polaridade e na excitabilidade. A polaridade mensura o quão agradável ou desagradável é a emoção, sendo geralmente representada pelos rótulos positivo, negativo e neutro. Por sua vez, a excitabilidade define a intensidade da emoção manifestada, variando entre alta, média e baixa. Devido a complexidade inerente ao processo de reconhecimento das emoções, a maioria dos estudos se concentra em analisar os estados afetivos sob a perspectiva de apenas uma das representações, como forma de simplificar a diversidade das expressões emocionais existentes.

A natureza subjetiva e a diversidade de formas de expressão dos estados emocionais dificultam o processo de reconhecimento das emoções por especialistas, especialmente quando há alguma patologia que afeta a capacidade do indivíduo de expressar as emoções (Santana, 2023a). Nessa perspectiva, estudos voltados para essa temática buscam investigar alternativas mais objetivas para classificação dos estados afetivos. Dentre elas, estão os dados eletroencefalográficos (EEG), sinais elétricos medidos através de eletrodos não invasivos dispostos em torno do couro cabeludo, capazes de registrar as atividades geradas pelo sistema nervoso central, parte do sistema nervoso responsável pelo processamento e geração das emoções. Sendo assim, os sinais EEG podem ser usados para analisar as atividades elétricas relacionadas diretamente à manifestação das emoções, como afirmam Yu e Wang (2022). Entretanto, mesmo sendo uma das fontes de dados mais objetivas e confiáveis para classificação dos estados afetivos, os sinais EEG são complexos para análise devido sua natureza não linear, não estacionária e sua sensibilidade a ruídos fisiológicos e técnicos, causados principalmente por movimentos musculares e a interferência do ambiente externo no momento da captura desses sinais. Outras fontes de dados de cunho comportamental, como a

fala, também podem ser usadas para detecção das emoções expressadas em um discurso, como defende Alluhaidan et al. (2023) em seu estudo, podendo ser combinadas com os sinais EEG para uma análise mais robusta.

O processo de reconhecimento das emoções pode ser dividido em indução, aquisição dos sinais, pré-processamento, processamento e classificação, conforme exposto por Yu e Wang (2022). Todas essas etapas são fundamentais para a construção de um sistema de reconhecimento de emoções eficiente. A fase de processamento tem como objetivo principal a extração das características mais relevantes dos sinais, que posteriormente serão dadas como entrada nos classificadores para reconhecimento das emoções na etapa de classificação. Mesmo diante do avanço dos algoritmos de aprendizado profundo, capazes de realizar a extração de características de forma automática, a engenharia manual nessa fase ainda é uma etapa preferível para a retenção das informações mais humanamente significativas dos dados brutos. Nesse contexto, a fase de processamento se destaca pela diversidade de estudos que buscam investigar diferentes métodos de extração, a fim de analisar a confiabilidade dos atributos derivados com base no desempenho obtido durante os testes dos modelos. Dentre os testes realizados, a avaliação da capacidade de generalização dos modelos é uma das principais abordagens discutidas, visto que se trata de uma etapa imprescindível para investigar a eficiência da técnica de extração, quando são utilizadas amostras com maior variabilidade. Entretanto, a maioria das pesquisas nessa área esbarram na limitação do escopo dos dados disponíveis publicamente, que ainda não são capazes de abranger a diversidade das expressões emocionais, comprometendo assim, a avaliação da robustez das técnicas propostas.

Nesse contexto, este trabalho tem como objetivo aprimorar o reconhecimento de estados afetivos, categorizados em positivo, negativo e neutro, a partir de dados provenientes de sinais EEG e de voz, de pessoas idosas com e sem demência, coletados durante a elaboração do trabalho proposto por Santana (2023a). A interpretação das emoções em pessoas idosas comumente é desafiadora devido a sua alta capacidade de regulação emocional, o que dificulta o processo de detecção desses estados afetivos. Além disso, a presença da demência neste estudo adiciona uma complexidade extra, uma vez que essa doença pode impactar diretamente a forma como as emoções são manifestadas nesse público. Essas particularidades, elencadas pela base de dados utilizada, se tornam propícias para avaliação da capacidade de generalização dos modelos de classificação. Portanto, será investigada uma forma de extração de características matemáticas, do domínio de tempo e frequência, sobre os sinais (Santana, 2023b), como forma de avaliar os possíveis impactos dessa representação no desempenho de algoritmos de classificação utilizando a base de dados supracitada.

O conteúdo do trabalho foi estruturado em 4 seções: Trabalhos Relacionados, Metodologia, Resultados e Discussões e Conclusão. Primeiramente, a seção de trabalhos relacionados reúne estudos relevantes publicados nos últimos anos sobre a temática discutida. Em seguida, a seção de metodologia descreve as principais etapas e processos necessários para condução do trabalho apresentado. Por sua vez, a seção de resultados destaca e analisa os desempenhos obtidos nos experimentos, fazendo comparações com outros estudos desenvolvidos na área. Por fim, a etapa de conclusão retoma os principais pontos discutidos ao longo do trabalho, aponta para os desafios e limitações enfrentados, bem como, os trabalhos futuros que podem surgir a partir desse.

2 TRABALHOS RELACIONADOS

No geral, o tema fornece uma literatura robusta, composta por uma diversidade de pesquisas publicadas na área. Para seleção desses trabalhos, foi utilizado o motor de busca, Google Acadêmico, filtrando por trabalhos relevantes e recentes, publicados nos últimos 5 anos, sobre a temática. Como resultado das buscas, foi observada uma grande concentração no uso de arquiteturas profundas para desenvolvimento das soluções. Mesmo diante dessa vertente, nota-se, também, trabalhos relevantes com resultados expressivos utilizando arquiteturas rasas, como SVM. Para os tipos de dados abordados, é evidente a utilização de sinais EEG na maioria das pesquisas conduzidas, com poucos trabalhos voltados para dados superficiais, como os sinais de voz, e menos ainda combinando ambos os tipos.

Yu e Wang (2022) elaboraram uma revisão da literatura sobre as técnicas de pré-processamento, métodos de extração de características e os classificadores mais utilizados em reconhecimento de emoções a partir de sinais EEG. O foco principal das técnicas de pré-processamento apresentadas está na detecção e descarte dos ruídos de natureza fisiológica e técnicas, geralmente sendo necessário a aplicação de mais de um método para eliminação de cada tipo de ruído observado nos sinais. Já para o processamento dos sinais, são apresentadas técnicas de extração de características que sejam capazes de capturar as informações mais relevantes presentes nos sinais EEG para reconhecimento das emoções. Foram apresentadas técnicas de diferentes tipos de domínio que conseguiram alcançar bons resultados em outras pesquisas relacionadas com a temática, são elas: Modelo Auto-Regressivo Adaptativo (AAR), Transformada Rápida de Fourier (FFT), Transformada de Wavelet e o Padrão Espacial Comum (CSP). Além disso, em relação aos classificadores usados, o estudo evidenciou os modelos SVM, Rede Neural Convolutiva (CNN), Rede Neural Recorrente (RRN), Memória de Curto e Longo Prazo (LSTM) e Análise Discriminante Linear (LDA).

Uma dessas abordagens é apresentada no estudo de Ahmadzadeh et al. (2024), que propõe uma Rede Neural Convolutiva Fuzzy (CFNN), uma arquitetura híbrida de aprendizado profundo que combina CNN (Rede Neural Convolutiva) com FNN (Rede Neural Fuzzy) para reconhecimento de parâmetros emocionais (valência e excitabilidade) a partir de sinais EEG. A técnica de extração de características usada foi a FFT, aplicado sobre cada canal de eletrodo para análise das frequências relacionadas às manifestações emocionais. Além do método proposto, o estudo avalia o desempenho obtido nos modelos SVM, KNN, Bi-LSTM e CNN para classificação das emoções quanto à sua valência e excitabilidade utilizando a base DEAP (Database for Emotion Analysis using Physiological Signals). Os resultados obtidos evidenciaram a dificuldade dos modelos SVM e KNN, que obtiveram acurácias abaixo dos 62% nas duas abordagens de classificação. Os modelos Bi-LSTM e CNN alcançaram um bom resultado, mas ainda distante quando comparado ao modelo proposto, que obteve uma acurácia de aproximadamente 98%, tanto na valência quanto na excitabilidade. O estudo reconhece o escopo limitado dos dados utilizados, que representa um universo de pouca diversidade, portanto não sendo possível avaliar o método proposto quanto à sua capacidade de generalização.

Em outro estudo, conduzido por Chowdary et al. (2022), os autores avaliaram o desempenho obtido por algoritmos de aprendizado profundo projetados especificamente para análise de dados sequenciais e temporais, como os sinais EEG. Com foco no reconhecimento de emoções a partir de sinais EEG, em seu estado positivo, negativo e neutro, foram avaliados os desempenhos entre três variações de redes neurais recorrentes: Rede Neural Recorrente

(RNN), Memória de Longo e Curto Prazo (LSTM) e Unidade Recorrente com Portas (GRU). Para extração das características dos sinais, foi utilizada uma abordagem estatística para registrar as variações ao longo do tempo. Os três modelos obtiveram ótimos resultados, alcançando acurácias de 95%, 97% e 96%, respectivamente, para os modelos RNN, LSTM e GRU. Entretanto, uma limitação do estudo foi a homogeneidade da base de dados utilizada, que possui um pouco mais de 2 mil amostras de apenas duas pessoas avaliadas.

Diante da falta de consenso na comunidade científica acerca de uma abordagem geral de representação dos dados provenientes de sinais EEG para reconhecimento de emoções, o estudo proposto por Yuvara et al. (2023) foi desenvolvido para investigar diferentes métodos de extração das características provenientes dos sinais EEG, com intuito de recomendar a aquele que obtém a maior confiabilidade e generalização para o reconhecimento das emoções. Nesse sentido, foram avaliados os seguintes métodos: Estatística, Transformada de Wavelet Contínua (CWT), Dimensão Fractal, Parâmetros de Hjorth e Espectros de Ordem Superior. Como forma de avaliar a capacidade de generalização, o estudo avaliou os métodos em cinco bases de dados separadamente: MAHNOB-HCI, DEAP, SEED, AMIGOS, DREAMER. Apesar de utilizar várias fontes de dados EEG para avaliar a capacidade de generalização dos métodos, o estudo optou por uma abordagem de treino e teste dependentes do sujeito, limitando a avaliação proposta. A pesquisa mostra os resultados obtidos por meio de 2 modelos, Árvore de Classificação e Regressão (CART) e SVM com kernel Gaussiano (GSVM), que classificaram as emoções quanto a valência e a excitabilidade. A combinação da Dimensão Fractal com o modelo CART teve o melhor desempenho médio geral, tanto na valência quanto na excitabilidade, alcançando uma acurácia média de 85,06% e 84,55%, respectivamente.

Devido à complexidade envolvida na análise de sinais EEG, em razão da sua natureza não estacionária e não linear, o estudo proposto por Alhalaseh and Alasasfeh (2020) explora técnicas de extração de características baseadas na não linearidade e não estacionaridade de sinais temporais. Diante disso, foi utilizada uma combinação híbrida que utiliza Entropia e Dimensão Fractal de Higuchi para derivar as características relevantes. Como forma de investigar a confiabilidade do método proposto, a pesquisa procurou avaliar seus impactos nos classificadores Naive Bayes (NB), KNN, Árvore de Decisão (DT) e CNN. Para isso, foi utilizado a base de dados DEAP com intuito de classificar as emoções quanto sua valência e excitabilidade. O modelo CNN se destacou com uma acurácia geral de 95,26%. Os demais modelos também obtiveram ótimos resultados, chegando a alcançar uma acurácia acima dos 90%.

O estudo por Karthiga et al. (2024) propõe uma arquitetura de aprendizado profundo para classificação das emoções a partir de sinais EEG provenientes das bases de dados DEAP e SEED, com base na valência (positivo, negativo, neutro). Para isso, o método proposto utiliza um algoritmo híbrido de meta heurística, que combina Colônia Artificial de Abelhas com Otimização por Lobos Cinzentos (ABC-GWO) para realização da seleção das características mais relevantes, extraídas através da Transformada Wavelet Complexa de Árvore Dupla (DT-CWT), e da otimização dos hiperparâmetros do classificador CNN. Além do método proposto, o estudo testa as características selecionadas pelo algoritmo híbrido de meta heurística nos modelos CNN, LSTM, CNN-LSTM, RNN. Sendo assim, o resultado obtido nos modelos destaca a qualidade das características selecionadas, com todos os modelos alcançando uma acurácia de mais de 90% em ambas as bases, com maior destaque para o modelo proposto que obteve 99% e 100% de acurácia. Vale destacar também que o estudo pontua a necessidade de investigar o modelo quando exposto a mais dados

heterogêneos para verificação da capacidade de generalização e, conseqüentemente, da eficiência do método proposto.

Em relação ao reconhecimento de emoções a partir de sinais de voz, Jain et al. (2020) classifica as emoções em quatro tipos (felicidade, tristeza, raiva e medo), utilizando as bases de dados LDC e UGA com registros de falas e SVM como algoritmo de classificação. São utilizadas features de caráter acústico para treinamento dos modelos, como tom, energia, taxa de fala, MFCC e LPCC. O estudo comparou o desempenho obtido utilizando as técnicas de extração de atributos MFCC e LPCC. Foi observado um melhor desempenho usando MFCC, que obteve um uma acurácia geral de aproximadamente 85%. Por sua vez, a técnica de representação LPCC obteve uma acurácia geral de aproximadamente 73%.

Por sua vez, Alluhaidan et al. (2023) destaca a importância do processo de extração de características dos sinais de voz para detecção de emoções. Dessa forma, foram analisados três tipos de extração de atributos: uma técnica baseada em atributos do domínio do tempo, outra abordagem no domínio da frequência (MFCC) e uma técnica híbrida que combina ambos os métodos (MFCCT). Além disso, foram analisados diferentes algoritmos de classificação de aprendizado de máquina e aprendizado profundo. Entre os modelos testados, a arquitetura baseada em CNN associada a abordagem híbrida MFCCT apresentou os melhores resultados, alcançando uma acurácia de 96,6%, 92,6% e 91,4%, respectivamente, nas bases EMO-DB, SAVEE e RAVDESS. O ótimo resultado apresentado evidencia a robustez do método proposto quanto a sua capacidade de generalização.

Tabela 1 – Síntese dos trabalhos relacionados

Autores	Processamento	Classificador	Dados	Resultados (acurácia)
Ahmadzadeh et al. (2024)	FFT	CFNN	Sinais de EEG da base DEAP	98% em valência e excitabilidade
Chowdary et al. (2022)	Estatística	RNN, LSTM e GRU	Sinais de EEG da base EEG Brain Wave Dataset: Feeling Emotions	95%, 97% e 96%, respectivamente, para RNN, LSTM e GRU em valência
Yuvara et al. (2023)	Estatística, CWT, Dimensão Fractal, Parâmetros de Hjorth e Espectros de Ordem Superior	CART e GSVM	Sinais de EEG da combinação das bases de dados MAHNOB-HCI, DEAP, SEED, AMIGOS e DREAMER	Respectivamente para valência e excitabilidade. Estatística: 84,03% e 83,45%; CWT: 84,76% e 83,94%; Dimensão Fractal: 85,06% e 84,55%; Parâmetros de Hjorth: 80,43% e 79,66%; Espectros de Ordem Superior: 81,11% e 80,52%
Alhalaseh and Alasasfeh (2020)	Entropia e Dimensão Fractal de Higuchi	NB, KNN, DT e CNN	Sinais de EEG da base DEAP	Média obtida para valência e excitabilidade. CNN: 95,26% KNN: 94,03% DT: 91,03% NB: 92,27%

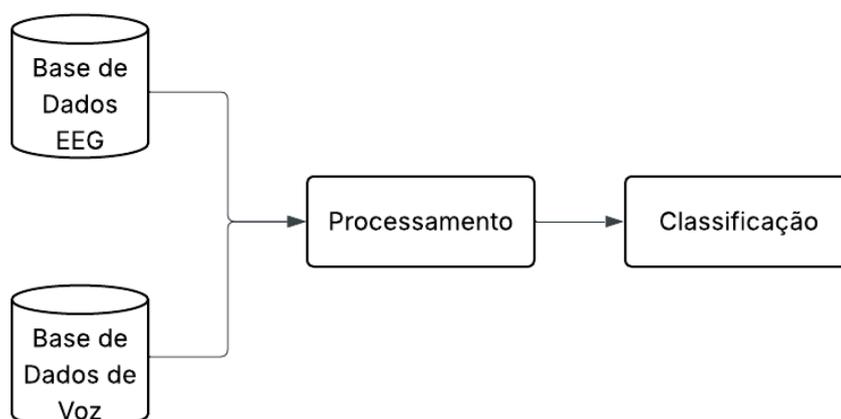
Karthiga et al. (2024)	DT-CWT e ABC-GWO	CNN	Sinais de EEG das bases DEAP e SEED	Classificação quanto à valência, com 99% e 100%, respectivamente, para as bases SEED e DEAP.
Alluhaidan et al. (2023)	MFCC, Domínio de Tempo e MFCCT	CNN	Sinais de voz das bases de dados EMO-DB, SAVEE e RAVDESS	Classificação das emoções em raiva, tédio, calma, nojo, medo, felicidade, neuro, tristeza, e surpresa, respectivamente para as bases EMO-DB, SAVEE e RAVDESS MFCC: 53,1%, 47,3% e 32,8%; Domínio de Tempo: 49,2%, 42,1% e 24,7%; MFCCT: 96,6%, 92,2% e 91,4%
Jain et al. (2020)	MFCC e LPCC	SVM	Sinais de voz das bases LDC e UGA	Classificação das emoções em felicidade, tristeza, raiva e medo. MFCC: 85,08%; LPCC: 73,12%

Fonte: O Autor (2025)

3 METODOLOGIA

A metodologia seguida aborda três principais etapas, conforme ilustradas na Figura 1. A primeira etapa corresponde ao estudo das bases de dados utilizadas, seguida do processamento dos dados brutos, com objetivo de preparação dos dados que serão utilizados na fase de classificação. Por fim, a etapa de classificação, que fica responsável pela condução dos experimentos de reconhecimento das emoções.

Figura 1– Metodologia adotada.



Fonte: O Autor (2025)

3.1 BASE DE DADOS

Foram utilizados os dados de sinais de voz e EEG, de 38 pessoas idosas com e sem demência, coletados durante os estudos de Santana (2023a), cujos experimentos conduzidos combinaram, como método de indução emocional, estímulos musicais e relatos autobiográficos feitos pelos próprios participantes da pesquisa. As emoções foram rotuladas por meio da autodeclaração, onde os participantes foram capazes de classificar o estado emocional com base na polaridade da música interagida. Ao final desse exercício, as bases ficaram compostas com 33 casos de estados afetivos positivos, 3 negativos e 2 neutros. Nitidamente desbalanceada, foi necessário a aplicação de uma técnica de balanceamento sobre as bases na etapa de classificação, como forma de atenuar o viés imposto pela classe predominante.

Cada modalidade de dados estão distribuídas em 38 arquivos, correspondentes a cada um dos participantes. Em referência a composição desses dados, as bases EEG são constituídas por 21 atributos que representam os sinais captados, de forma não invasiva e indolor, por 21 canais de eletrodos de prata com gel condutor, dispostos em torno do couro cabeludo dos participantes. O equipamento utilizado para registro desses sinais foi o polissonógrafo Neurovirtual, modelo Brain Wave III. Enquanto os dados de voz são oriundos dos relatos autobiográficos e estão distribuídos em faixas de áudios no formato wav. As bases utilizadas estão pré-processadas com a aplicação de filtros para remoção de ruídos originados pela interferência elétrica do ambiente.

Como forma de investigar a capacidade computacional no reconhecimento das emoções, quando submetidos a diferentes formas de dados, foram realizados experimentos utilizando três tipos de bases: uma apenas com sinais EEG, outra composta apenas com sinais de voz e uma base multimodal que combina os sinais EEG e de voz.

3.2 PROCESSAMENTO

O processamento dos dados é uma etapa primordial nos estudos sobre o reconhecimento das emoções, com impacto direto sobre o desempenho computacional durante a classificação dos estados afetivos, conforme é observado nos estudos de Saucedo et al. (2024). Nesse sentido, durante essa etapa, são realizadas ações sobre os dados brutos como forma de sintetizá-los em uma representação mais significativa, visto que esses dados, por natureza, não representam relações claras em sua composição, o que dificulta a sua análise por parte dos algoritmos de classificação.

Nessa perspectiva, dentre as ações aplicadas durante a fase de processamento dos dados, destacam-se a segmentação e a extração de atributos. Para as bases de sinais EEG e voz, foram realizadas segmentações em janelas de 5 segundos, seguida da extração dos atributos para cada uma das janelas. Já para etapa de extração, foi utilizada a proposta apresentada no estudo de Santana (2023b), cujos atributos definidos capturam relações estatísticas no domínio de tempo e frequência, fornecendo assim uma representação mais relevante desses sinais. Ao final do processo de extração, alguns atributos apresentaram algumas indefinições (NaN), fruto das operações matemáticas realizadas, e portanto, para esses casos, foi aplicada uma substituição com base na média observada.

Após os procedimentos de segmentação e extração dos atributos, foi realizada a fusão dos dados com objetivo de gerar um único vetor de representação multimodal, por meio da

junção dos vetores das modalidades de EEG e voz. Entretanto, devido à disparidade existente na quantidade de instâncias entre as bases de voz e EEG, para cada participante, foi necessária a replicação das instâncias de cada indivíduo na base EEG, por meio de uma seleção com base na aleatoriedade sobre os registros do sujeito observado, a fim de igualar o número de amostras por participante nas duas modalidades. Embora essa técnica introduza redundância nos dados, ela não compromete a consistência das informações, visto que os dados de EEG e voz não estão completamente sincronizados de forma temporal.

Tabela 2 – Lista de atributos extraídos para as representações de EEG, voz e multimodal

Atributo	Expressão	Atributo	Expressão
Valor Médio (μ)	$\frac{1}{N} \sum_{n=1}^N x_n$	Comprimento da Forma de Onda	$\sum_{n=1}^{N-1} x_{n+1} - x_n $
Variância	$\frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2$	Cruzamento por Zero	$\sum_{n=1}^{N-1} \text{sgn}(x_n \cdot x_{n+1}) \wedge x_n - x_{n+1} \geq \text{threshold} $ onde $\text{sgn}(x) = \begin{cases} 1, & \text{Se } x \geq \text{threshold} \\ 0, & \text{caso contrário} \end{cases}$
Desvio Padrão (σ)	$\sqrt{\frac{1}{N-1} \sum_{n=1}^N x_n - \mu ^2}$	Mudanças de Sinal da Inclinação	$\sum_{n=1}^{N-1} f(x_n - x_{n-1})(x_n - x_{n+1}) $ onde $f(x) = \begin{cases} 1, & \text{se } x \geq \text{threshold} \\ 0, & \text{caso contrário} \end{cases}$
Raiz Quadrada da Média dos Quadrados	$\sqrt{\frac{1}{N} \sum_{n=1}^N (x_n)^2}$	Frequência Média	$\frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j}$ onde, f_j é a frequência P_j é o poder espectral M é o comprimento da frequência
Varição Média da Amplitude	$\frac{1}{N} \left(\sum_{n=1}^N \left \frac{dx(t)}{dt} \right \right)$	Frequência Mediana	$\frac{1}{2} \sum_{j=1}^M P_j$
Desvio Absoluto da Diferença	$\sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{dx(t)}{dt} \right)^2}$	Potência Média	$\frac{\sum_{j=1}^M P_j}{M}$
Valor Absoluto Integrado	$\sum_{n=1}^N x_n$	Frequência de Pico (PKF)	$\max(P_j)$
Detector Logarítmico	$e^{\left(\frac{1}{N} \sum_{n=1}^N \log(x_n)\right)}$	Razão do Espectro de Potência	$\frac{PKF}{\sum_{j=1}^M P_j}$
Integral Simples dos Quadrados	$\sum_{n=1}^N x_n^2$	Primeiro Momento Espectral (SM1)	$\sum_{j=1}^M f_j P_j$
Valor Absoluto Médio	$\frac{1}{N} \sum_{n=1}^N x_n $	Segundo Momento Espectral (SM2)	$\sum_{j=1}^M f_j^2 P_j$
Média do Núcleo Logarítmico	$\frac{1}{N} \left \sum_{n=1}^N x_n \right $	Terceiro Momento Espectral	$\sum_{j=1}^M f_j^3 P_j$
Potência Total (TP)	$\sum_{j=1}^M P_j$	Variância da Frequência Central	$\frac{SM2}{TP} - \left(\frac{SM1}{TP}\right)^2$
Kurtosis	$\frac{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^4}{\sigma^4}$	Quinto Momento	$\left \frac{1}{N} \sum_{n=1}^N (x_n)^5 \right $
Amplitude Máxima	$\max(x_n)$	Quarto Momento	$\left \frac{1}{N} \sum_{n=1}^N (x_n)^4 \right $
Terceiro Momento	$\left \frac{1}{N} \sum_{n=1}^N (x_n)^3 \right $		

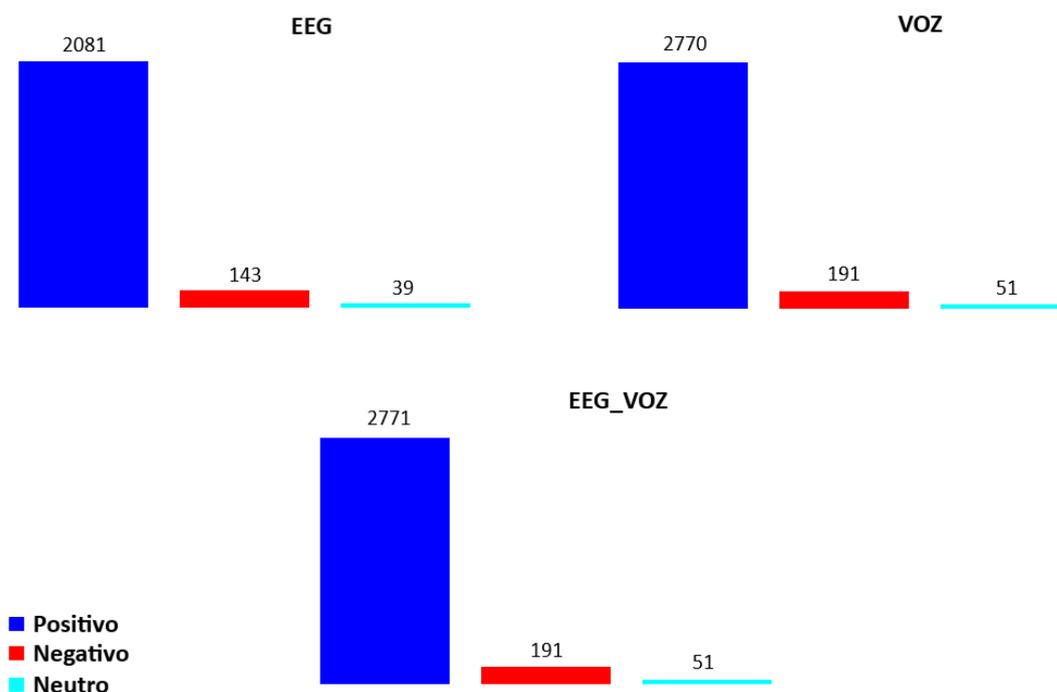
Fonte: Santana (2023b)

3.3 CLASSIFICAÇÃO

É durante a etapa de classificação que acontece a condução dos processos necessários para realização dos experimentos de reconhecimento das emoções através dos algoritmos de classificação. Todas as etapas dessa fase foram realizadas através da ferramenta Weka, um software de código aberto voltado para execução de experimentos de aprendizado de máquina e análise de dados (Frank, Eibe, et al., 2016).

Inicialmente os dados foram divididos em 70% treino e 30% teste, sobre cada base utilizada. Ao final desse processo, foi constatado um desbalanceamento nas amostras de treinamento para os três tipos de bases, com uma acentuada representação da classe positiva, conforme ilustrado na Figura 2. Diante dessa conjuntura, as amostras foram balanceadas utilizando a técnica de balanceamento por SMOTE (Synthetic Minority Over-sampling Technique), como forma de atenuar possíveis casos de enviesamento nos resultados. Essa abordagem é capaz de criar instâncias sintéticas das classes minoritárias considerando os vizinhos mais próximos da instância selecionada. Com base nisso, o estudo proposto parametrizou a quantidade de vizinhos para 3.

Figura 2 – Distribuição das classes nas amostras de treinamento EGG, VOZ e EEG_VOZ.



Fonte: Adaptado do Weka (2025)

Para a realização dos experimentos de aprendizagem de máquina aplicados no campo do reconhecimento das emoções, foram empregados algoritmos clássicos de classificação supervisionada, tais como Random Forest, SVM, J48, Bayes Network e Naive Bayes. Esses modelos foram selecionados em razão das vantagens que apresentam em relação às abordagens de aprendizado profundo, como alta capacidade de interpretabilidade, onde é possível compreender com mais clareza as decisões tomadas a partir de uma instância submetida, e a menor complexidade computacional, cujo tempo de execução se torna evidente em grandes volumes de dados.

Popularmente conhecido como *Árvore de Decisão*, o algoritmo J48 segue uma estrutura em forma de árvore, cujo os nós representam abstrações de regras condicionais aprendidas durante o treinamento do modelo. Cada uma dessas regras é composta por um determinado atributo, onde seus possíveis valores expressam as condições que orientam o fluxo para tomada da decisão final. Esses atributos são dispostos na árvore de maneira hierárquica utilizando-se a medida de pureza como critério de seleção, de modo a identificar o atributo capaz de promover a segmentação mais homogênea possível entre as classes.

De maneira similar a *Árvore de Decisão*, a *Random Forest* utiliza os mesmos princípios de aprendizagem, mas adaptado para um comitê de árvores de decisão. A ideia central por trás desse algoritmo trata do compartilhamento da responsabilidade de decisão para cada uma dessas árvores, assim como em um sistema eleitoral, cujo objetivo é eleger um representante, no caso deste algoritmo, a decisão final a ser tomada. Como consequência da corresponsabilidade imposta, a *Random Forest* apresenta uma maior robustez em comparação ao algoritmo de *Árvore de Decisão*, visto que cada árvore é treinada sob diferentes perspectivas de dados, tornando-se menos propício ao risco de *overfitting*.

Outro algoritmo tradicionalmente reconhecido no campo da inteligência artificial, é o *Naive Bayes*. Esse algoritmo utiliza como técnica de aprendizagem a probabilidade condicional expressa através do teorema de Bayes, tomando como premissa a independência das variáveis usadas para predição. Entretanto, o pressuposto é um forte atenuante desse modelo, uma vez que a maioria dos problemas de classificação usam atributos que possuem algum grau de correlação. Mesmo violando-se a restrição, o modelo ainda é capaz de obter bons resultados no geral, o que faz com que ele seja amplamente adotado na literatura.

Por outro lado, o algoritmo *Bayes Network* também faz uso da estatística probabilística com base no teorema de Bayes, mas adaptado para assumir dependências existentes entre atributos, dada a classe observada. Isso faz com que esse modelo tenha mais flexibilidade para lidar com diversos problemas de classificação quando comparados ao algoritmo de *Naive Bayes*.

Ainda nesse contexto, vale ressaltar, também, o modelo de classificação *SVM* (*Support Vector Machine*). Caracterizado como um algoritmo de otimização, o *SVM* possui como princípio de busca, encontrar a melhor forma de separar linearmente os dados em grupos que representam as classes. Para isso, o modelo visa maximizar as margens entre o separador utilizado e os pontos mais próximos que representam classes distintas, mais conhecidos como vetores de suporte. Nesse sentido, quanto maior a maximização obtida, maior será a capacidade de generalização do modelo. Entretanto, podem existir casos em que não seja possível separar linearmente conjunto de dados dentro do espaço dimensional considerado. Diante desses cenários, pode ser necessária a parametrização do modelo com base em funções kernel, responsáveis por expandir a representação dos dados para um espaço de maior dimensão, no qual a separação linear pode tornar-se possível.

Todos os modelos foram avaliados na etapa de treinamento, utilizando validação cruzada como técnica de aprendizagem. Esse método consiste em particionar o conjunto de dados em k amostras de tamanho iguais (*folds*), dos quais $k-1$ são usados para treinamento do modelo e 1 amostra destinada para teste. O experimento é executado k vezes, de modo a utilizar uma amostra de teste distinta para cada repetição, com intuito de avaliar o modelo de forma mais generalizável. No estudo proposto, foi usada uma validação cruzada 10-fold

configurada para ser reproduzida 30 vezes, visando avaliar também a dispersão dos resultados.

Em relação às métricas utilizadas para avaliação dos resultados dos modelos, tanto na etapa de treinamento quanto na etapa de teste, foram adotadas a acurácia (ACC), sensibilidade (SEN), especificidade (ESP), índice kappa (κ) e a área sob a curva ROC (AUC). Sob uma perspectiva mais geral, a acurácia é utilizada para analisar a precisão de acerto das classes, diferentemente das métricas de especificidade e sensibilidade que calculam, respectivamente, a taxa de verdadeiros positivos e negativos, consideradas abordagens mais específicas voltadas para análise da precisão obtida em cada classe. O índice kappa, por sua vez, é utilizado para calcular estatisticamente a concordância entre a predição obtida e o fato observado, com objetivo de mensurar o grau de incerteza do modelo. Já a área sob a curva ROC, trata da relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos, de maneira que quanto maior for a área observada, maior será a capacidade de distinção do modelo entre as classes.

4 RESULTADOS E DISCUSSÕES

Para análise dos resultados, foram obtidos nas etapas de treinamento e teste dos modelos, a média e o desvio padrão das métricas avaliadas. Na etapa de teste, foi selecionado o modelo que obteve o melhor desempenho geral na etapa de treinamento, para cada tipo de base de dados. Diante dessa perspectiva, é válido destacar que para base de dados de sinais EEG, dois modelos apresentaram um desempenho geral semelhante na etapa de treinamento: Random Forest com 150 e 250 árvores. Sendo assim, a acurácia média foi usada como critério de desempate nesse cenário. Os resultados podem ser visualizados com mais detalhes, em suas respectivas etapas, nas tabelas presentes nas seções abaixo.

De início, é fundamental pontuar, a diferença no desempenho geral observado nas três bases de dados estudadas. Os experimentos conduzidos em ambas as etapas, mostram que a base de dados de sinais EEG combinados com sinais de voz, obtiveram o melhor resultado geral, seguida da base de dados EEG e, por fim, pela base de dados de sinais de voz. Isso pode ser observado através da acurácia, que apresentou uma média de 99,94% e 99,90%, respectivamente, na etapa de treinamento e teste, através da Random Forest com 150 árvores, representando o modelo de melhor configuração para a base de dados de sinais EEG combinados com sinais de voz.

É importante destacar, também, o ótimo desempenho obtido nos demais modelos treinados com a base multimodal, que no geral também foram superiores aos modelos treinados nas demais bases, tanto por sua acurácia média quanto pelo seu baixo desvio padrão apresentado, o que evidencia uma alta confiabilidade e um alto grau de estabilidade e consistência nesses modelos. Na etapa de teste, o melhor modelo também atingiu ótimos resultados, mesmo diante de uma base originalmente desbalanceada, com destaque para a classe positiva que obteve uma sensibilidade de 100%.

De maneira semelhante, o modelo de melhor configuração treinado com a base de sinais EEG, a Random Forest com 250 árvores, alcançou uma acurácia média de 99,79% e

99,48%, respectivamente, nas etapas de treino e teste. Ainda que satisfatórios, esse resultado, atrelado ao desvio padrão moderado presentes em sua acurácia, indica um desempenho levemente inferior ao observado na base multimodal. Analisando através da perspectiva dos resultados na etapa de teste, é possível que o modelo tenha apresentado a limitação observada devido a escassez de representatividade das classes minoritárias. Ainda que a base tenha sido balanceada, a introdução de instâncias sintéticas pode não ter sido capaz de proporcionar a variabilidade necessária para o modelo. Isso pode ser observado pela dificuldade enfrentada em distinguir a classe negativa e ainda mais a classe neutro, quando comparado a abordagem multimodal, cujo os impactos do desbalanceamento da base, foi atenuado pelo processo de enriquecimento dos dados.

Já para a base de dados com sinais de voz, a Random Forest com 300 árvores se destacou como sendo o modelo de melhor configuração, chegando a obter uma acurácia média de 96,83% e 90,41%, respectivamente, para fase de treinamento e teste. Entretanto, para esse cenário em específico, somente analisar a acurácia não é suficiente para interpretar o desempenho obtido nessa configuração para etapa de teste, visto que o desbalanceamento existente na base utilizada contribuiu para uma acurácia mais elevada.

Diante dessa conjuntura, é necessário analisar esse caso sob a perspectiva dos resultados apresentados nas Tabelas 7 e 8, que servem de apoio para avaliação da capacidade de generalização do modelo. Com base nessas informações, torna-se evidente a limitação dessa configuração quando exposta a dados com maior variabilidade, uma vez que o modelo apresentou confusão acentuada das classes negativas e neutras com a classe majoritária positiva, o que sugere que essa abordagem foi a mais impactada pelo desbalanceamento da base dentre as outras, comprometendo assim o seu desempenho final. Além da dificuldade de discriminação das classes negativas e neutras, os modelos treinados com a base de dados de sinais de voz demonstraram um alto grau de dispersão da acurácia e resultados insatisfatórios para os classificadores SVM e Naive Bayes.

Dentre os algoritmos tradicionais utilizados nos experimentos, é válido evidenciar o desempenho obtido pelo Naive Bayes. Esse algoritmo acabou se destacando negativamente pelo baixo desempenho apresentado nos três tipos de bases de dados. Esse comportamento, que também foi observado nos estudos conduzidos por Alluhaidan et al. (2023), ao investigar o desempenho de diferentes classificadores em 3 bases de dados, pode estar associado à relação de correlação existentes entre os atributos, violando assim, a premissa da independência condicional exigida pelo algoritmo de Naive Bayes.

É imprescindível destacar o desempenho satisfatório obtido com algoritmos rasos de aprendizado de máquina, mesmo na ausência de técnicas mais sofisticadas de pré-processamento. Estudos recentes, no campo de pesquisa do reconhecimento das emoções, se concentram predominantemente a estudar técnicas mais inovadoras, como os algoritmos de aprendizado profundo, atrelados ao uso de pré-processamento sobre os dados brutos, para investigar os impactos desses modelos na capacidade de reconhecimento das emoções. Apesar dos bons resultados que podem ser conquistados com uso dessas abordagens, o custo

computacional acaba se tornando o seu principal gargalo. Nesse contexto, a pesquisa conduzida por Karthiga et al. (2024) chega a abordar a complexidade computacional como uma das limitações do estudo, que propõe uma arquitetura profunda com uso de técnicas de pré-processamentos mais refinadas. Portanto, os resultados obtidos no presente estudo, com uso de algoritmos tradicionais de aprendizado de máquina e sem a aplicação de técnicas refinadas de pré-processamento, demonstra que é possível atingir uma alta confiabilidade, assim como os algoritmos de aprendizado profundo, porém com um menor custo computacional, um requisito fundamental para conceber aplicações em tempo real nessa área.

É necessário ressaltar, ainda, a importância da representação dos dados proposta para alcançar os resultados supracitados. O desempenho positivo obtido por meio do processamento desses dados, torna-se evidente quando comparado com outros estudos que utilizaram abordagens metodológicas semelhantes. Em sua pesquisa, Santana (2023a) utilizou a rede LeNet para extração dos atributos, empregando a mesma base de dados utilizada no presente estudo. O modelo de melhor configuração, uma Random Forest configurada com 450 árvores, obteve uma acurácia de 60,52% e 63,97% nos dois tipos de experimentos conduzidos. Apesar dos resultados expressivos derivados através da representação matemática proposta, faz necessário o desenvolvimento de mais estudos com intuito de explorar os impactos da exposição dos modelos quando submetidos a uma maior variabilidade dos dados, principalmente de instâncias negativas e neutras, que foram sub-representadas nas bases utilizadas.

4.1 RESULTADOS NA ETAPA DE TREINAMENTO

Nessa seção são apresentados os resultados obtidos durante a etapa de treinamento. O desempenho nos três tipos de bases de dados pode ser visualizado através da Tabela 3, Tabela 4 e da Tabela 5, com base na média e o desvio padrão das métricas avaliadas.

Tabela 3 – Resultado com a base de dados de sinais EEG na fase de treinamento. Em negrito estão as maiores médias e os menores desvios padrão observado em cada métrica avaliada.

Algoritmo	Configuração	ACC̄	ACCσ	κ̄	κσ	SEN̄	SENσ	ESP̄	ESPσ	AUC̄	AUCσ
Random Forest	10 árvores	99,6934	0,2164	0,9954	0,0032	0,9989	0,0021	0,9959	0,0031	0,9997	0,0004
Random Forest	20 árvores	99,7709	0,1785	0,9965	0,0026	0,9996	0,0013	0,9967	0,0026	0,9998	0,0003
Random Forest	50 árvores	99,7922	0,1709	0,9968	0,0025	0,9998	0,0010	0,9969	0,0025	0,9999	0,0002
Random Forest	100 árvores	99,7928	0,1728	0,9968	0,0025	0,9999	0,0004	0,9969	0,0025	0,9999	0,0002
Random Forest	150 árvores	99,7944	0,1675	0,9969	0,0025	0,9999	0,0003	0,9969	0,0024	0,9999	0,0002
Random Forest	200 árvores	99,7944	0,1705	0,9969	0,0025	0,9999	0,0002	0,9969	0,0025	0,9999	0,0002
Random Forest	250 árvores	99,7960	0,1707	0,9969	0,0025	0,9999	0,0002	0,9969	0,0025	0,9999	0,0002
Random Forest	300 árvores	99,7949	0,1689	0,9969	0,0025	0,9999	0,0002	0,9969	0,0025	0,9999	0,0002

SVM	linear	99,1882	0,3533	0,9878	0,0053	0,9812	0,0090	0,9972	0,0025	0,9871	0,0053
SVM	polinomial 2	99,5412	0,2784	0,9931	0,0041	0,9916	0,0066	0,9973	0,0026	0,9945	0,0036
SVM	RBF 0.01	90,4955	1,0981	0,8574	0,0164	0,9534	0,0153	0,8807	0,0148	0,9203	0,0098
SVM	RBF 0.25	99,5551	0,2557	0,9933	0,0038	0,9931	0,0054	0,9967	0,0027	0,9949	0,0030
SVM	RBF 0.50	99,7164	0,2070	0,9957	0,0031	0,9966	0,0038	0,9974	0,0025	0,9970	0,0022
J48	-	98,4993	0,5052	0,9774	0,0075	0,9787	0,0102	0,9893	0,0052	0,9815	0,0081
Bayes Network	-	96,4765	0,7487	0,9471	0,0112	0,9990	0,0020	0,9476	0,0111	0,9810	0,0045
Naive Bayes	-	74,4125	1,0348	0,6162	0,0155	0,2682	0,0290	0,9823	0,0064	0,6638	0,0179

Nota: $ACC\bar{}$, $ACC\sigma$, $\kappa\bar{}$, $\kappa\sigma$, $SEN\bar{}$, $SEN\sigma$, $ESP\bar{}$, $ESP\sigma$, $AUC\bar{}$ e $AUC\sigma$, representam, respectivamente, a acurácia média, desvio padrão da acurácia, média do índice de kappa, desvio padrão do índice de kappa, média da sensibilidade, desvio padrão da sensibilidade, média da especificidade, desvio padrão da especificidade, média da área sob a curva ROC e o desvio padrão da área sob a curva ROC.

Fonte: O Autor (2025)

Tabela 4 – Resultado com a base de dados de sinais de voz na fase de treinamento. Em negrito estão as maiores médias e os menores desvios padrão observado em cada métrica avaliada.

Algoritmo	Configuração	$ACC\bar{}$	$ACC\sigma$	$\kappa\bar{}$	$\kappa\sigma$	$SEN\bar{}$	$SEN\sigma$	$ESP\bar{}$	$ESP\sigma$	$AUC\bar{}$	$AUC\sigma$
Random Forest	10 árvores	95,6844	0,7142	0,9352	0,0107	0,9433	0,0142	0,9641	0,0079	0,9862	0,0035
Random Forest	20 árvores	96,3464	0,6626	0,9451	0,0099	0,9497	0,0133	0,9708	0,0071	0,9897	0,0030
Random Forest	50 árvores	96,6674	0,6289	0,9500	0,0094	0,9546	0,0130	0,9731	0,0068	0,9913	0,0027
Random Forest	100 árvores	96,7585	0,6096	0,9513	0,0091	0,9562	0,0124	0,9736	0,0068	0,9917	0,0025
Random Forest	150 árvores	96,8058	0,6250	0,9520	0,0093	0,9572	0,0125	0,9738	0,0068	0,9919	0,0025
Random Forest	200 árvores	96,8135	0,6091	0,9522	0,0091	0,9572	0,0123	0,9740	0,0068	0,9920	0,0025
Random Forest	250 árvores	96,8311	0,6222	0,9524	0,0093	0,9575	0,0126	0,9741	0,0068	0,9920	0,0025
Random Forest	300 árvores	96,8331	0,6016	0,9524	0,0090	0,9575	0,0122	0,9741	0,0068	0,9921	0,0024
SVM	linear	56,0865	1,5581	0,3413	0,0233	0,4224	0,0274	0,8074	0,0167	0,6208	0,0169
SVM	polinomial 2	66,6337	1,5101	0,4995	0,0226	0,5486	0,0289	0,8769	0,0140	0,7627	0,0159
SVM	RBF 0.01	49,6646	1,4914	0,2449	0,0223	0,4718	0,0280	0,7283	0,0191	0,5990	0,0171
SVM	RBF 0.25	61,4817	1,4709	0,4222	0,0220	0,4523	0,0276	0,8801	0,0137	0,7147	0,0159
SVM	RBF 0.50	66,6798	1,4549	0,5002	0,0218	0,5412	0,0275	0,8817	0,0131	0,7604	0,0155
J48	-	94,0378	0,9003	0,9105	0,0135	0,9152	0,0182	0,9568	0,0091	0,9439	0,0111
Bayes Network	-	92,5879	0,9372	0,8888	0,0140	0,9061	0,0158	0,9496	0,0099	0,9718	0,0065

Naive Bayes	-	42,4770	1,2523	0,1371	0,0187	0,1021	0,0170	0,9682	0,0074	0,6526	0,0197
-------------	---	---------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Nota: ACC , $ACC\sigma$, κ , $\kappa\sigma$, SEN , $SEN\sigma$, ESP , $ESP\sigma$, AUC e $AUC\sigma$, representam, respectivamente, a acurácia média, desvio padrão da acurácia, média do índice de kappa, desvio padrão do índice de kappa, média da sensibilidade, desvio padrão da sensibilidade, média da especificidade, desvio padrão da especificidade, média da área sob a curva ROC e o desvio padrão da área sob a curva ROC.

Fonte: O Autor (2025)

Tabela 5 – Resultado com a base de dados de sinais EEG combinados com sinais de voz na fase de treinamento. Em negrito estão as maiores médias e os menores desvios padrão observado em cada métrica avaliada.

Algoritmo	Configuração	ACC	$ACC\sigma$	κ	$\kappa\sigma$	SEN	$SEN\sigma$	ESP	$ESP\sigma$	AUC	$AUC\sigma$
Random Forest	10 árvores	99,8335	0,1424	0,9975	0,0021	0,9992	0,0016	0,9979	0,0019	0,9999	0,0001
Random Forest	20 árvores	99,9262	0,0904	0,9988	0,0013	0,9995	0,0012	0,9990	0,0012	0,9999	0,0001
Random Forest	50 árvores	99,9394	0,0802	0,9990	0,0012	0,9997	0,0008	0,9991	0,0011	0,9999	6,0401
Random Forest	100 árvores	99,9378	0,0814	0,9990	0,0012	0,9997	0,0009	0,9991	0,0011	0,9999	4,6358
Random Forest	150 árvores	99,9402	0,0796	0,9991	0,0011	0,9997	0,0009	0,9992	0,0011	0,9999	3,7171
Random Forest	200 árvores	99,9386	0,0796	0,9990	0,0011	0,9997	0,0008	0,9991	0,0011	0,9999	3,1497
Random Forest	250 árvores	99,9378	0,0808	0,9990	0,0012	0,9997	0,0008	0,9991	0,0011	0,9999	2,9015
Random Forest	300 árvores	99,9394	0,0814	0,9990	0,0012	0,9997	0,0008	0,9992	0,0011	0,9999	2,7265
SVM	linear	99,0234	0,3294	0,9853	0,0049	0,9850	0,0073	0,9928	0,0033	0,9888	0,0039
SVM	polinomial 2	99,6827	0,1881	0,9952	0,0028	0,9961	0,0035	0,9971	0,0021	0,9964	0,0021
SVM	RBF 0.01	95,5726	0,6537	0,9335	0,0098	0,9368	0,0136	0,9716	0,0067	0,9649	0,0058
SVM	RBF 0.25	99,6775	0,1993	0,9951	0,0029	0,9981	0,0026	0,9961	0,0025	0,9971	0,0019
SVM	RBF 0.50	99,8191	0,1483	0,9972	0,0022	0,9996	0,0010	0,9974	0,0021	0,9985	0,0012
J48	-	98,6846	0,4264	0,9802	0,0063	0,9847	0,0082	0,9892	0,0044	0,9851	0,0061
Bayes Network	-	96,0362	0,6222	0,9405	0,0093	0,9975	0,0030	0,9417	0,0092	0,9773	0,0039
Naive Bayes	-	74,7998	1,0455	0,6219	0,0156	0,3189	0,0270	0,9727	0,0066	0,6900	0,0160

Nota: ACC , $ACC\sigma$, κ , $\kappa\sigma$, SEN , $SEN\sigma$, ESP , $ESP\sigma$, AUC e $AUC\sigma$, representam, respectivamente, a acurácia média, desvio padrão da acurácia, média do índice de kappa, desvio padrão do índice de kappa, média da sensibilidade, desvio padrão da sensibilidade, média da especificidade, desvio padrão da especificidade, média da área sob a curva ROC e o desvio padrão da área sob a curva ROC.

Fonte: O Autor (2025)

4.2 RESULTADOS NA ETAPA DE TESTE

Nessa seção são apresentadas tabelas com os resultados obtidos durante a etapa de teste dos modelos com as melhores configurações encontradas. A Tabela 6, Tabela 8 e a Tabela 10, representam os resultados para os três tipos de abordagem, com base na média das métricas adotadas. Já a Tabela 7, Tabela 9 e Tabela 11, correspondem às matrizes de confusão de cada experimento realizado nessa etapa.

A representação EEG se destacou com um ótimo resultado no geral, apresentando uma precisão elevada e um alto grau de concordância, mesmo diante da queda de precisão da classe neutro, quando comparado com as demais. De maneira semelhante, a abordagem multimodal apresentou altos índices em todas as métricas observadas, destacando-se como modelo de maior capacidade de generalização observada. Em contraste com os resultados observados anteriormente, a configuração com base em sinais de voz apresentou um baixo grau de concordância e uma baixa precisão entre as classes, falhando no teste de robustez.

Tabela 6 – Resultado com a base de dados de sinais EEG na fase de teste.

Classe	ACC	κ	SEN	ESP	AUC
Positivo	99,4861	0,9659	1,000	0,938	1,000
Negativo	99,4861	0,9659	0,952	1,000	1,000
Neutro	99,4861	0,9659	0,889	1,000	0,999
AVG	99,4861	0,9659	0,995	0,943	1,000

Nota: ACC, κ , SEN, ESP e AUC, representam, respectivamente, a acurácia média, média do índice de kappa, média da sensibilidade, média da especificidade e a média da área sob a curva ROC.

Fonte: O Autor (2025)

Tabela 7 – Matriz de confusão com a base de dados de sinais EEG na fase de teste.

		Previsto		
		Positivo	Negativo	Neutro
Verdadeiro	Positivo	892	0	0
	Negativo	3	60	0
	Neutro	2	0	16

Fonte: O Autor (2025)

Tabela 8 – Resultado com a base de dados de sinais de voz na fase de teste.

Classe	ACC	κ	SEN	ESP	AUC
Positivo	90,4187	0,1092	0,976	0,094	0,665
Negativo	90,4187	0,1092	0,117	0,979	0,714

Neutro	90,4187	0,1092	0,009	0,996	0,658
AVG	90,4187	0,1092	0,904	0,175	0,668

Nota: ACC , κ , SEN , ESP e AUC , representam, respectivamente, a acurácia média, média do índice de kappa, média da sensibilidade, média da especificidade e a média da área sob a curva ROC.

Fonte: O Autor (2025)

Tabela 9 – Matriz de confusão com a base de dados de sinais voz na fase de teste.

		Previsto		
		Positivo	Negativo	Neutro
Verdadeiro	Positivo	11.585	241	49
	Negativo	724	96	3
	Neutro	213	8	2

Fonte: O Autor (2025)

Tabela 10 – Resultado com a base de dados de sinais de EEG combinado com sinais de voz na fase de teste.

Classe	ACC	κ	SEN	ESP	AUC
Positivo	99,9072	0,9938	1,000	0,989	1,000
Negativo	99,9072	0,9938	0,998	1,000	1,000
Neutro	99,9072	0,9938	0,955	1,000	1,000
AVG	99,9072	0,9938	0,999	0,989	1,000

Nota: ACC , κ , SEN , ESP e AUC , representam, respectivamente, a acurácia média, média do índice de kappa, média da sensibilidade, média da especificidade e a média da área sob a curva ROC.

Fonte: O Autor (2025)

Tabela 11 – Matriz de confusão com a base de dados de sinais EEG combinado com sinais de voz na fase de teste.

		Previsto		
		Positivo	Negativo	Neutro
Verdadeiro	Positivo	11.880	0	0
	Negativo	2	821	0
	Neutro	10	0	213

Fonte: O Autor (2025)

5 CONCLUSÃO

O reconhecimento dos estados afetivos a partir de técnicas de aprendizagem de máquina já é uma realidade dentro da comunidade científica, mas ainda carece de estudos que procuram explorar as manifestações emocionais a partir de particularidades inerentes aos seres humanos. Sob essa ótica, e motivado pela possibilidade de contribuição no tratamento

da demência com auxílio da musicoterapia, o estudo proposto buscou investigar os impactos no campo do reconhecimento das emoções, utilizando dados de sinais EEG, voz e uma abordagem multimodal combinando ambos os tipos, oriundos de pessoas idosas com e sem demência.

O estudo foi segmentado em macro etapas responsáveis pelo estudo da base de dados coletada, pela aplicação de técnicas de processamento dos dados observados e pela condução dos processos necessários para execução dos experimentos de classificação das emoções através dos algoritmos de aprendizado de máquina. Dentre as técnicas de processamento, vale destacar a extração de atributos aplicada, onde foi utilizado uma representação matemática no domínio de tempo e frequência. Os resultados derivados superam o desempenho obtido em outras pesquisas e destacam o uso da representação multimodal, que obteve uma acurácia de 99,90%. Tais evidências apontam para a importância do enriquecimento dos dados com múltiplas modalidades, atrelada à extração de atributos adotada, para aprimorar a eficiência na identificação dos estados afetivos. Além disso, o estudo aponta para a possibilidade de obter resultados satisfatórios com uso de algoritmos rasos de aprendizagem de máquina e sem a utilização de técnicas de pré-processamento com alto custo computacional, mesmo diante da complexidade existente na base de dados empregada, imposta principalmente pela sua composição etária, a condição patológica observada e o desbalanceamento das classes.

Finalmente, é importante ressaltar que, embora tenham sido obtidos resultados valiosos, não é possível assegurar a escalabilidade do modelo proposto, uma vez que a amostra utilizada é relativamente pequena e carece de mais variabilidade em relação às formas de expressões das emoções, sobretudo em casos de estados afetivos negativo e neutro. Portanto, para conseguir avaliar a capacidade de robustez do modelo em questão, faz-se necessário o desenvolvimento de mais estudos que explorem outras amostras sobre o público supracitado, podendo ser combinado com outras técnicas de pré-processamento, dependendo dos resultados obtidos.

REFERÊNCIAS

Yu, Chaofei, and Mei Wang. "Survey of emotion recognition methods using EEG information." *Cognitive Robotics*, vol. 2, 2022, p. 15. Accessed 04 June 2025.

Ahmadzadeh, Nasim, et al. "Detecting emotions through EEG signals based on modified convolutional fuzzy neural network." *Scientific Reports*, vol. 14, 2024, p. 14. Accessed 05 June 2025.

Chowdary, M. Kalpana, et al. "Emotion Recognition from EEG Signals Using Recurrent Neural Networks." *Electronics*, vol. 11, 2022, p. 20. Accessed 05 June 2025.

Yuvara, Rajamanickam, et al. "Comprehensive Analysis of Feature Extraction Methods for Emotion Recognition from Multichannel EEG Recordings." *Sensors*, vol. 23, 2023, p. 19. Accessed 06 June 2025.

Alhalaseh, Rania, and Suzan Alasasfeh. "Machine-Learning-Based Emotion Recognition System Using EEG Signals." *Computers*, vol. 9, 2020, p. 15. Accessed 07 June 2025.

Karthiga, M., et al. “Eeg based smart emotion recognition using meta heuristic optimization and hybrid deep learning techniques.” *Scientific Reports*, vol. 14, 2024, p. 24. Accessed 08 June 2025.

Jain, Manas, et al. “Speech Emotion Recognition using Support Vector Machine.” 2020, p. 6. Accessed 09 June 2025.

Alluhaidan, Ala Saleh, et al. “Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network.” *Applied Sciences*, vol. 13, 2023, p. 15. Accessed 10 June 2025.

Santana, Maíra Araújo. *Arquiteturas de redes profundas para interfaces musicais humano-máquina para reconhecimento de emoções em sinais eletroencefalográficos e de voz como apoio à musicoterapia*. 19 Dezembro 2023a, p. 176.

Santana, Maíra Araújo. “Emotion Recognition from Multimodal Data: a machine learning approach combining classical and hybrid deep architectures.” *Research on Biomedical Engineering*, vol. 39, 2023b, p. 26. Accessed 21 June 2025.

Russel, James A. “A Circumplex Model of Affect.” *Journal of Personality and Social Psychology*, vol. 39, no. 6, 1980, p. 19. Accessed 16 July 2025.

Sauceda, Jesus Arturo, et al. “Emotion Classification from Electroencephalographic Signals Using Machine Learning.” *Brain Sciences*, vol. 14, 2024, p. 25. Accessed 16 July 2025.

Picard, Rosalind Wright. *Affective Computing*. MIT Press, 1997. Accessed 22 July 2025.

Benenson, Rolando. *Teoria da musicoterapia: contribuição ao contexto do conhecimento não-verbal*. 3 ed., Summus Editorial, 1988. Accessed 22 July 2025.

Frank, Eibe, et al. *Data Mining: Practical Machine Learning Tools and Techniques*. 4 ed., Morgan Kaufmann Publishers In, 2016. Accessed 07 August 2025.

AGRADECIMENTOS

Aos meus pais, Carlos Frederico e Lúcia Maria, por me mostrarem desde cedo a importância dos estudos e do conhecimento, e que, com muito esforço demandado pela vida, permitiram-me chegar até aqui. Às minhas irmãs, Thaís e Tarsila, que sempre me inspiraram e incentivaram durante essa trajetória. À minha orientadora, Maíra, pelos seus ensinamentos, direcionamentos e incentivos durante a realização deste trabalho. À Universidade Federal de Pernambuco e ao Centro de Informática, por me proporcionar toda a infraestrutura necessária para construção de uma sólida formação acadêmica. E por fim, aos amigos que fiz durante essa caminhada, laços que levarei comigo para o resto da vida.