Universidade Federal do Maranhão
Centro de Ciências Exatas e Tecnologia
Curso de Engenharia da Computação

Avaliação de Desempenho e Efetividade de Redes Neurais Aplicadas a Dados de Risco de Transtorno de Ansiedade Pediátrica

Renata Costa Rocha

São Luís – MA

2025

Renata Costa Rocha

Avaliação de Desempenho e Efetividade de Redes Neurais Aplicadas a Dados de Risco de Transtorno de Ansiedade Pediátrica

Trabalho de Conclusão de Curso apresentado à Universidade Federal do Maranhão, em cumprimento às exigências institucionais para obtenção do grau de Bacharel em Engenharia da Computação, sob orientação do Prof. Dr. Pedro Baptista Fernandes.

Universidade Federal do Maranhão

Curso de Engenharia da Computação

São Luís – MA

2025

“Pensamento é a força criadora.”

Racionais MC’s

“Intenção sem ação é ilusão.”

Lair Ribeiro

Dedico este Trabalho a Deus, onipresente em todas as minhas batalhas; aos meus filhos Henrique e Yasmin, que são o melhor de mim e razão maior da minha força e dedicação; ao meu esposo Yuri, companheiro e amor da vida; à minha mãe Leila, que me concedeu o presente da existência; ao meu pai Franklin (in memoriam), que me impulsionou a trilhar os caminhos da Engenharia; aos meus irmãos Rafael e Gabriel, pela parceria e alegria sempre compartilhadas; à minha cunhada Mariana, pelo carinho e exemplo constantes; e ao meu sobrinho Lucas, cuja inteligência e pureza são fonte de inspiração.

Agradeço a Deus, pelo discernimento e pela força concedidos; ao meu esposo Yuri e aos meus filhos Henrique e Yasmin, que com paciência e amor me ofereceram suporte e foram meu porto seguro; às crianças e adolescentes que participaram do estudo de Harvard, cuja valiosa contribuição possibilitou a criação do banco de dados aberto e, consequentemente, o desenvolvimento deste trabalho; e ao meu orientador, Prof. Dr. Pedro Fernandes Baptista, pelas orientações, ensinamentos e direcionamentos ao longo desta jornada.

RESUMO

Este Trabalho de Conclusão de Curso propôs um sistema computacional utilizando redes neurais artificiais para detecção (*screening*) e classificação multinível do risco de transtorno de ansiedade pediátrica, com base em medidas psicométricas e comportamentais. O objetivo foi comparar o desempenho do modelo proposto, composto por um *Multilayer Perceptron (MLP)* supervisionado e por um módulo não supervisionado baseado em *Autoencoder* associado à clusterização *KMeans*, empregados de forma complementar, com o modelo simbólico *ADTree* apresentado no estudo *"Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach"*, disponibilizado no *Harvard Dataverse*.

O delineamento experimental incluiu pré-processamento estatístico, normalização *z-score*, aprendizado supervisionado e análise latente não supervisionada. Na classificação multinível (0–3), o *MLP* obteve acurácia global de 85,11%, enquanto, no cenário binário de triagem (*screening*), alcançou acurácia de 88,93%, sensibilidade de 80,56% e especificidade de 94,80%. Em contraste, a versão replicada do *ADTree* apresentou sensibilidade nula, evidenciando limitações de classificadores determinísticos em contextos psicológicos complexos.

A interpretabilidade foi examinada por meio do método *SHAP (Shapley Additive exPlanations)*, que indicou forte correspondência entre as variáveis de maior influência — Afeto Ansioso, Evitação e Sofrimento Antecipatório — e os construtos clínicos definidos pelo Manual Diagnóstico e Estatístico de Transtornos Mentais, quinta edição (DSM-5), e pela Classificação Internacional de Doenças, décima primeira revisão (CID-11). Ademais, o módulo *Autoencoder + KMeans* evidenciou um continuum emocional entre os níveis de risco (0–3), reforçando o caráter dimensional do transtorno de ansiedade.

Os resultados indicam que o paradigma conexionista supera o modelo simbólico em sensibilidade, estabilidade e equilíbrio métrico, preservando coerência psicológica e interpretabilidade. O sistema contribui para o avanço da Engenharia da Computação aplicada à saúde mental infantil, evidenciando o potencial de redes neurais explicáveis em triagens automatizadas e sistemas de apoio à decisão.

Palavras-chave: Redes Neurais Artificiais. Ansiedade Pediátrica. Dados Psicométricos. Classificação Multinível. Predição de Risco.

ABSTRACT

This Final Undergraduate Project proposed a computational system using artificial neural networks for detection (screening) and multilevel classification of pediatric anxiety disorder risk, based on psychometric and behavioral measures. The objective was to compare the performance of the proposed model, composed of a supervised Multilayer Perceptron (MLP) and an unsupervised module based on Autoencoder associated with KMeans clustering, employed in a complementary way, with the symbolic model ADTree presented in the study “Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach”, available in the Harvard Dataverse.

The experimental design included statistical preprocessing, *z*-score normalization, supervised learning, and unsupervised latent analysis. In the multilevel classification (0–3), the MLP achieved an overall accuracy of 85.11%, while in the binary screening scenario it reached an accuracy of 88.93%, sensitivity of 80.56%, and specificity of 94.80%. In contrast, the replicated version of ADTree presented null sensitivity, highlighting the limitations of deterministic classifiers in complex psychological contexts. Interpretability was examined through the SHAP (Shapley Additive exPlanations) method, which indicated a strong correspondence between the most influential variables — Anxious Affect, Avoidance, and Anticipatory Distress — and the clinical constructs defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), and the International Classification of Diseases, Eleventh Revision (ICD-11). Furthermore, the Autoencoder + KMeans module evidenced an emotional continuum across risk levels (0–3), reinforcing the dimensional nature of anxiety disorder.

The results indicate that the connectionist paradigm outperforms the symbolic model in sensitivity, stability, and metric balance, while preserving psychological coherence and interpretability. The system contributes to the advancement of Computer Engineering applied to child mental health, highlighting the potential of explainable neural networks in automated screenings and decision-support systems.

Keywords: Artificial Neural Networks. Pediatric Anxiety. Psychometric Data. Multilevel Classification. Risk Prediction.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Os transtornos psiquiátricos com manifestação na infância e adolescência constituem um imperativo de saúde pública global, com prevalências que desafiam os sistemas de saúde, conforme demonstrado em uma recente meta-análise [1]. O paradigma diagnóstico vigente, contudo, ancorado na semiologia psiquiátrica clássica, ainda depende fundamentalmente de métodos subjetivos, como entrevistas clínicas, relatos de pais e professores e observação comportamental.

Tais abordagens, embora indispensáveis, são suscetíveis a vieses intrínsecos, variabilidade interobservador e dificuldades na captura da dinâmica temporal dos sintomas, resultando frequentemente em diagnósticos tardios ou imprecisos e, consequentemente, em intervenções subótimas. Em uma revisão sistemática abrangente [2], foram analisadas as trajetórias de desenvolvimento de problemas de saúde mental em jovens, concluindo-se que a heterogeneidade dos sintomas e a sobreposição entre diferentes diagnósticos tornam a identificação precoce extremamente complexa, reforçando a urgência por marcadores biológicos e computacionais que possam estratificar os pacientes de forma mais objetiva.

Em resposta a essas limitações, emerge o campo da psiquiatria computacional, uma área interdisciplinar que combina métodos da Engenharia e Neurociência para criar ferramentas quantitativas de apoio ao diagnóstico clínico [3]. Essa abordagem visa complementar a avaliação tradicional com dados mensuráveis que reflitam a atividade dos sistemas neurais subjacentes. Nesse cenário, sinais neurofisiológicos como a atividade elétrica cerebral (EEG) e a dinâmica autonômica inferida pela variabilidade da frequência cardíaca (HRV) a partir do eletrocardiograma (ECG) configuram janelas não invasivas promissoras para o estudo de disfunções afetivas e cognitivas, permitindo uma compreensão mais precisa de padrões emocionais e comportamentais alterados [4]. Embora tais modalidades desempenhem papel central no estado da arte da psiquiatria computacional, o presente estudo utiliza exclusivamente variáveis psicométricas e comportamentais estruturadas em formato tabular, compatíveis com o escopo metodológico e com os dados disponibilizados pelo *Harvard Dataverse*.

A materialização dessa abordagem translacional, contudo, é fundamentalmente um desafio da Engenharia da Computação, especialmente nas áreas de processamento

de sinais e aprendizado de máquina. Extrair informação clinicamente relevante a partir de dados complexos — sejam fisiológicos ou psicométricos — exige técnicas sofisticadas de filtragem, redução de dimensionalidade e modelagem computacional. Estudos recentes [4, 5] ilustram esse potencial ao aplicar algoritmos de aprendizado profundo em sinais de EEG e obter acurácias superiores a 90% na classificação de crianças com Transtorno de Déficit de Atenção com Hiperatividade (TDAH), uma condição caracterizada pela combinação de sintomas de desatenção, hiperatividade e impulsividade, evidenciando que padrões latentes podem servir como assinaturas robustas para diversos transtornos psiquiátricos.

O presente trabalho insere-se nessa fronteira entre Engenharia da Computação e Neurociência, propondo o desenvolvimento de um sistema computacional para classificação multinível do risco de ansiedade infantil, utilizando dados públicos disponibilizados no *Harvard Dataverse* [6]. Essa base reúne medidas psicométricas e comportamentais coletadas em contexto clínico padronizado, cuja rotulagem diagnóstica fundamenta-se nos referenciais internacionais estabelecidos pelo Manual Diagnóstico e Estatístico de Transtornos Mentais, em sua quinta edição (DSM-5), elaborado pela Associação Americana de Psiquiatria, e pela Classificação Internacional de Doenças, em sua décima primeira revisão (CID-11), publicada pela Organização Mundial da Saúde [7, 8].

O estudo replica, de forma funcionalmente compatível, o modelo simbólico *Alternating Decision Tree (ADTree)* apresentado por [6], adotando uma implementação baseada em *AdaBoost* com *Decision Stumps*, dada a indisponibilidade da arquitetura original. Esse modelo simbólico é então comparado a duas arquiteturas conexionistas de redes neurais artificiais: o modelo supervisionado *Multilayer Perceptron (MLP)* e o modelo não supervisionado *Autoencoder* combinado à técnica de clusterização *KMeans*.

A comparação visa avaliar capacidade de generalização, eficiência e interpretabilidade, explorando os contrastes metodológicos entre abordagens simbólicas e conexionistas na classificação de fenômenos psicológicos complexos. Além de propor um sistema de classificação, este trabalho busca avançar em direção à transparência algorítmica, incorporando técnicas de explicabilidade como o *SHapley Additive exPlanations (SHAP)* [9]. A explicação baseada em valores de Shapley é aplicada

especificamente ao modelo *MLP*, uma vez que este se apresenta como a arquitetura de melhor desempenho entre aquelas avaliadas. A análise de interpretabilidade permite identificar as variáveis com maior influência na predição do risco de ansiedade, garantindo alinhamento clínico, auditabilidade e valor prático aos resultados obtidos.

Dessa forma, a presente pesquisa investiga como arquiteturas neurais distintas podem representar, classificar e explicar o risco de ansiedade infantil, evidenciando o potencial da Inteligência Artificial aplicada à saúde mental. Ao integrar fundamentos de Engenharia da Computação, Processamento de Sinais e Ciência de Dados, o estudo contribui para o desenvolvimento de sistemas inteligentes de apoio à decisão, capazes de aprimorar a triagem precoce e auxiliar na compreensão de fenômenos psicológicos complexos sob uma perspectiva computacional.

1.1 Entendimento do Problema

O reconhecimento precoce e a adequada estratificação do risco para transtornos de ansiedade em crianças em idade pré-escolar configuram-se como um desafio de notável complexidade no campo da saúde mental [10]. Tal dificuldade decorre da natureza intrinsecamente multifatorial do fenômeno, resultante da interação dinâmica entre variáveis biológicas, psicológicas e sociais, bem como da sutileza das manifestações clínicas próprias dessa faixa etária. Embora o avanço das técnicas de aprendizado de máquina tenha ampliado as possibilidades de análise de dados clínicos, observa-se que a maioria dos modelos desenvolvidos até o momento ainda se restringe a classificações binárias, que dicotomizam o risco em categorias simplificadas (“presença” ou “ausência” de transtorno”). Essa limitação metodológica impede a captura das nuances de severidade e das transições graduais que caracterizam o espectro ansioso, reduzindo o potencial preditivo e a utilidade clínica das abordagens computacionais [11].

1.2 Impactos Clínicos e Sociais das Limitações Diagnósticas

A adoção de modelos diagnósticos dicotômicos, ao negligenciar a complexidade dimensional dos transtornos de ansiedade, gera implicações de ampla magnitude para a prática clínica e para a gestão em saúde mental [12]. A ausência de uma estratificação

detalhada do risco inviabiliza a personalização das intervenções terapêuticas, aspecto essencial para o manejo eficaz dos diferentes graus de comprometimento emocional. Consequentemente:

1. **Intervenções terapêuticas imprecisas:** A escassez de modelos computacionais aplicados a populações clínicas reais capazes de discriminar múltiplos níveis de gravidade dificulta a proposição de tratamentos ajustados à intensidade dos sintomas [13, 14].

2. **Prognóstico comprometido:** A falta de modelagem adequada da progressão do risco acarreta diagnósticos tardios e intervenções menos oportunas, propiciando a cronificação dos sintomas e o surgimento de comorbidades ao longo do desenvolvimento [1].

3. **Sobrecarga dos sistemas de saúde:** Modelos simplificados tendem a provocar alocação inadequada de recursos terapêuticos, direcionando atenção excessiva a casos leves e negligenciando quadros graves que demandam intervenção imediata [15].

4. **Impacto no desenvolvimento socioemocional:** A ausência de diagnósticos graduais compromete a prevenção de prejuízos duradouros no desenvolvimento emocional, social e cognitivo das crianças [16].

5. **Subnotificação e baixa representatividade:** A carência de modelos treinados em bases amplas, heterogêneas e padronizadas limita a capacidade de generalização e reduz a sensibilidade a diferentes perfis sintomáticos [5, 17].

6. **Carência de protocolos padronizados:** A falta de metodologias sistematizadas para treinamento, validação e interpretação de redes neurais aplicadas a dados clínicos institui inconsistências e reduz sua aplicabilidade prática [18].

Estudos recentes [5] evidenciam que o campo enfrenta desafios estruturais, incluindo a escassez de bases representativas, a insuficiência de validação externa e a ausência de métricas padronizadas. A limitação em fornecer classificações multiníveis e interpretáveis compromete não apenas a precisão diagnóstica, mas também o avanço de soluções computacionais capazes de transformar a triagem e o acompanhamento psicopatológico.

1.3 Por que é urgente o enfrentamento desse problema?

A urgência no desenvolvimento de modelos neurais sensíveis, explicáveis e capazes de realizar classificações estratificadas fundamenta-se em razões neurodesenvolvimentais, clínicas e sociais [14]. Intervenções precoces, sustentadas por modelos capazes de capturar padrões sutis, podem reconfigurar positivamente trajetórias cognitivas e emocionais. Além disso, diagnósticos automatizados baseados em classificações graduais permitem a priorização racional dos atendimentos e o uso mais eficiente dos recursos terapêuticos.

A ausência de práticas diagnósticas tecnicamente assistidas compromete de maneira significativa o desenvolvimento global da criança. A identificação tardia dos transtornos ansiosos está associada a déficits persistentes de autorregulação emocional, prejuízos cognitivos e dificuldades interpessoais. Como discutido por [19], há urgência em aprimorar sistemas de triagem e monitoramento emocional desde os primeiros anos de vida, com apoio de ferramentas computacionais capazes de captar manifestações comportamentais e afetivas emergentes.

1.4 Quais os motivos para que esse problema ainda não tenha sido superado?

Apesar dos avanços recentes, o problema persiste em razão de limitações técnicas, metodológicas e práticas que se inter-relacionam e perpetuam lacunas no desenvolvimento de soluções robustas [18]. Entre as principais causas, destacam-se a escassez de dados clínicos rotulados, a falta de validação externa e a reduzida disponibilidade de modelos neurais concebidos especificamente para classificação multinível em populações pediátricas.

A ausência de padronização metodológica dificulta o avanço da área e a transposição de modelos laboratoriais para ambientes clínicos reais. Nesse cenário, este trabalho busca contribuir desenvolvendo uma metodologia capaz de modelar a complexidade e a heterogeneidade dos transtornos de ansiedade em crianças, oferecendo diagnósticos mais precisos, interpretáveis e clinicamente relevantes.

1.5 Por que a Inteligência Artificial e as Redes Neurais se tornam indispensáveis?

A persistência das limitações diagnósticas discutidas decorre não apenas de obstáculos operacionais, mas também de restrições epistemológicas do paradigma clínico tradicional. Os transtornos de ansiedade infantil apresentam natureza multifatorial e dinâmica, resultando em uma estrutura de dados que excede a capacidade humana de integração e a expressividade dos modelos estatísticos convencionais.

Nesse contexto, a Inteligência Artificial — especialmente o aprendizado profundo — torna-se uma ferramenta metodologicamente necessária. Redes neurais são capazes de identificar padrões complexos, modelar relações não lineares e integrar múltiplas dimensões de informação. Modelos como o *Multilayer Perceptron (MLP)* são apropriados para dados tabulares psicométricos, enquanto *Autoencoders* permitem explorar estruturas latentes, úteis para compreender o continuum clínico.

Ao captar relações sutis e não triviais, as redes neurais superam os limites dos modelos dicotômicos, proporcionando classificações multiníveis mais sensíveis e alinhadas à realidade dimensional dos transtornos ansiosos. Assim, seu uso representa um passo essencial para diagnósticos precoces, padronizados e interpretáveis, fortalecendo sistemas de apoio à decisão na psiquiatria do desenvolvimento.

1.6 Estado da Arte

A literatura científica recente evidencia avanços expressivos na aplicação de técnicas de aprendizado de máquina ao estudo e detecção de transtornos psiquiátricos na infância. Contudo, apesar do crescimento do campo, a consolidação desses métodos em práticas clínicas ainda enfrenta desafios técnicos, metodológicos e interpretativos relevantes. Estudos recentes [20] mostram que algoritmos supervisionados clássicos — como *Support Vector Machines (SVM)* e *Random Forests* — têm apresentado resultados promissores na classificação de sintomas psicopatológicos. Entretanto, a ausência de validação externa e a limitação de amostras heterogêneas dificultam a generalização dos achados, restringindo sua aplicabilidade translacional em populações clínicas distintas.

Paralelamente, arquiteturas conexionistas como as redes neurais artificiais (*Ar-*

tificial Neural Networks – ANNs) destacam-se pela capacidade de modelar relações não lineares e estruturas de alta dimensionalidade, características frequentemente observadas em dados comportamentais, psicométricos e neurobiológicos. Pesquisas envolvendo sinais eletroencefalográficos (EEG) demonstraram que modelos convolucionais (*CNNs*) alcançam elevado desempenho na identificação de padrões neurológicos relacionados ao TDAH e a estados emocionais diversos, obtendo acurácias superiores a 90% [5]. De forma complementar, trabalhos que combinam transformadas como a *wavelet* com classificadores como *SVM* têm obtido desempenho elevado na detecção de marcadores eletrofisiológicos [4], reforçando o papel do processamento de sinais na psiquiatria computacional.

Apesar desses avanços, observa-se que grande parte das pesquisas — especialmente na área da ansiedade infantil — permanece centrada em classificações binárias, incapazes de capturar a dimensionalidade do risco emocional. Trabalhos pioneiros, como o de [6], utilizaram estruturas simbólicas de decisão, como o *Alternating Decision Tree (ADTree)*, aplicadas a dados psicométricos do *Harvard Dataverse*, demonstrando o potencial de modelos interpretáveis. Entretanto, tais abordagens apresentam limitações em termos de sensibilidade e capacidade de generalização quando expostas a variabilidade comportamental mais ampla.

Dessa forma, a literatura recente converge para a necessidade de incorporar abordagens conexionistas, como *Multilayer Perceptrons (MLPs)* e *Autoencoders*, integradas a técnicas modernas de interpretabilidade, especialmente o *SHapley Additive exPlanations – SHAP* [9]. A integração entre desempenho e explicabilidade é, hoje, considerada um requisito central para a aplicação clínica segura de modelos de IA, sobretudo em populações vulneráveis como crianças em idade pré-escolar.

Estudos recentes [18] destacam que o avanço da neurociência computacional depende não apenas de modelos sofisticados, mas também da adoção de práticas de ciência aberta e do uso de bases públicas de larga escala, como o *Harvard Dataverse*, o *Healthy Brain Network* e o *ABCD Study*. Essas iniciativas permitem ampliar a representatividade das amostras, replicar experimentos e fortalecer a reprodutibilidade científica — um ponto crítico na área de saúde mental infantil.

Nos últimos anos, modelos generativos e arquiteturas modernas como *Variational Autoencoders (VAEs)* e *Transformers* passaram a ser empregados na análise de

fenômenos cognitivos e emocionais, permitindo capturar padrões latentes complexos em dados contínuos [18]. A capacidade dessas arquiteturas de representar distribuições complexas tem tornado possível explorar relações profundas entre comportamento, emoção e variáveis clínicas.

As investigações contemporâneas também demonstram o potencial de abordagens híbridas que combinam módulos convolucionais com mecanismos de atenção temporal, permitindo captar dinâmicas fisiológicas rápidas relacionadas a estados emocionais. Estudos recentes [5] mostram que arquiteturas *CNN–Transformer* podem alcançar acurácia superior a 86% em tarefas de classificação de ansiedade, evidenciando ganhos consistentes de robustez e generalização. Embora tais métodos dependam de sinais fisiológicos, que não compõem todas as bases de dados clínicas, seu desenvolvimento indica uma tendência clara de evolução em direção à psiquiatria computacional multimodal.

Outra vertente fundamental refere-se ao uso de técnicas de *Explainable Artificial Intelligence (XAI)*. A demanda por interpretabilidade é tanto científica quanto ética. Revisões recentes [18, 20] reforçam que modelos aplicados à saúde mental devem oferecer rastreabilidade das decisões, permitindo ao clínico compreender quais variáveis influenciam a predição. Métodos como *SHAP, LRP e Grad-CAM* consolidaram-se como ferramentas indispensáveis nesse processo, especialmente em cenários onde decisões automatizadas podem ter impacto direto sobre protocolos terapêuticos.

De maneira complementar, observa-se o crescimento das abordagens multimodais que integram dados fisiológicos, comportamentais e sociodemográficos em um único pipeline de aprendizagem. Pesquisas recentes demonstram que a fusão entre diferentes modalidades sensoriais melhora substancialmente a acurácia preditiva e a estabilidade temporal dos modelos [5]. Apesar disso, estudos aplicados exclusivamente a dados psicométricos — como no caso desta pesquisa — ainda são minoritários, o que reforça a relevância de explorar arquiteturas capazes de lidar de forma eficiente com bases tabulares e fenômenos psicológicos multidimensionais.

No campo da ansiedade infantil especificamente, as lacunas são ainda mais evidentes. Poucos estudos abordam classificações multiníveis; a maioria opera com classificações binárias limitadas. Ainda menos trabalhos aplicam, simultaneamente, modelos conexionistas, métodos simbólicos interpretáveis, técnicas de análise latente

e mecanismos de explicabilidade baseados em *SHAP*. Essa combinação metodológica, centrada em dados psicométricos e comportamentais, permanece escassa na literatura, especialmente para crianças em idade pré-escolar.

A literatura também aponta desafios estruturais persistentes. Revisões sistemáticas mostram que muitos modelos apresentam baixa qualidade de evidência segundo a metodologia *GRADE* [21], devido à ausência de replicação externa, à heterogeneidade entre amostras e à falta de padronização metodológica [3]. Ademais, grande parte dos estudos permanece restrita a validações internas, o que limita a translação dos resultados para o contexto clínico real.

Em síntese, a área encontra-se em rápida evolução, mas mantém lacunas importantes: escassez de modelos explicáveis, carência de classificações estratificadas, baixo uso de redes neurais em bases psicométricas e ausência de comparações diretas entre arquiteturas simbólicas e conexionistas em tarefas de risco ansioso. Assim, o presente trabalho alinha-se às demandas contemporâneas da psiquiatria computacional ao propor um pipeline reprodutível que combina *MLP, Autoencoder* com clusterização e *ADTree*, incorporando análise explicável via *SHAP* para promover simultaneamente desempenho, interpretabilidade e aplicabilidade clínica.

1.7 Justificativa

A psiquiatria pediátrica contemporânea enfrenta desafios complexos na identificação precoce, acompanhamento e manejo de transtornos mentais do neurodesenvolvimento, como o Transtorno de Déficit de Atenção e Hiperatividade (TDAH), o Transtorno do Espectro Autista (TEA), os transtornos de ansiedade, a depressão infantil e as dificuldades específicas de aprendizagem. Tais condições apresentam etiologia multifatorial, envolvendo interações dinâmicas entre fatores genéticos, neurobiológicos e psicossociais, o que torna sua caracterização e estratificação diagnóstica particularmente desafiadoras [1]. Tradicionalmente, o diagnóstico clínico baseia-se em entrevistas estruturadas, observações comportamentais e aplicação de escalas psicométricas. Embora amplamente utilizados, esses instrumentos são vulneráveis à subjetividade, à variabilidade interobservador e à dependência da capacidade colaborativa da criança, podendo resultar em diagnósticos tardios ou imprecisos [22, 2]. Essas limitações tornam-se ainda mais críticas em contextos de baixa acessibilidade a especialistas,

como ambientes escolares e regiões com escassez de serviços especializados.

Entre os transtornos mencionados, os quadros ansiosos se destacam pela alta prevalência e pelo impacto funcional significativo. Estimativas epidemiológicas apontam que entre 5% e 8% das crianças em idade escolar apresentam sintomas compatíveis com transtornos de ansiedade [22, 1], comprometendo o desenvolvimento socioemocional e aumentando o risco de psicopatologias na vida adulta, como depressão e abuso de substâncias [23]. Segundo a Organização Mundial da Saúde, os transtornos ansiosos figuram entre as principais causas de anos vividos com incapacidade na população infantojuvenil [24], reforçando a necessidade de estratégias de triagem precoce que sejam objetivas, escaláveis e sustentadas por evidências quantitativas.

O diagnóstico tradicional baseado em entrevistas e escalas subjetivas possui limitações metodológicas amplamente reconhecidas. A interpretação de sintomas depende da experiência do avaliador e das narrativas parentais, frequentemente influenciadas por fatores culturais e socioeconômicos [25]. Além disso, a heterogeneidade fenotípica dos transtornos ansiosos — que podem manifestar-se como sintomas internalizantes (medo, evitação) ou externalizantes (agitação, irritabilidade) — dificulta a padronização diagnóstica, favorecendo subdiagnósticos e inconsistências clínicas [26]. Esse cenário compromete a acurácia das decisões e prolonga o sofrimento infantil, aumentando custos para sistemas de saúde e educação [15].

Nesse contexto, a integração entre a Engenharia da Computação e a Neuropsiquiatria Infantil emerge como campo promissor. Ferramentas computacionais podem oferecer análises quantitativas, reprodutíveis e menos suscetíveis a vieses, complementando a avaliação clínica [27]. Técnicas de aprendizado de máquina — em especial as redes neurais artificiais — destacam-se pela capacidade de modelar relações não lineares entre variáveis psicométricas e fisiológicas, revelando padrões latentes associados ao risco ansioso [28]. Embora existam abordagens mais complexas na literatura recente, como *CNNs, VAEs* e redes *Transformer* [29], o presente estudo concentra-se em arquiteturas leves e interpretáveis, mais adequadas ao tamanho amostral da base utilizada e ao objetivo de explicar as decisões do modelo.

Pesquisas contemporâneas também apontam o potencial de indicadores fisiológicos — como variabilidade da frequência cardíaca e condutância da pele — como marcadores objetivos de regulação emocional [30, 31]. Embora o dataset utilizado neste

trabalho não contenha dados multimodais completos (como EEG ou HRV detalhada), ele incorpora medidas psicofisiológicas relevantes, justificando a exploração de modelos computacionais capazes de integrar informações comportamentais e fisiológicas.

O avanço dos métodos de Inteligência Artificial Explicável (XAI) reforça a relevância de tais modelos. Técnicas como *SHapley Additive exPlanations (SHAP)* permitem identificar as variáveis mais influentes na predição, favorecendo transparência, auditabilidade e compatibilidade ética com aplicações em saúde mental infantil [9, 32]. No presente trabalho, a explicabilidade foi aplicada exclusivamente ao modelo com melhor desempenho — o *Multilayer Perceptron (MLP)* — garantindo alinhamento entre capacidade preditiva e interpretação clínica.

Diante desse cenário, o presente estudo justifica-se pela necessidade de desenvolver um sistema computacional capaz de classificar o risco de ansiedade infantil de maneira objetiva, reprodutível e interpretável. Utilizando dados psicométricos e comportamentais do *Harvard Dataverse* [6], o projeto implementa um *pipeline* completo envolvendo pré-processamento, modelagem supervisionada e não supervisionada, comparação com um modelo simbólico de referência (*ADTree*) — aqui replicado por meio de uma aproximação funcional — e análise explicável por *SHAP*. A adoção de critérios do DSM-5 e da CID-11 como referência teórica assegura consistência clínica e pertinência diagnóstica.

Assim, esta pesquisa se justifica por preencher uma lacuna metodológica na classificação multinível do risco de ansiedade infantil, contribuindo para o avanço da psiquiatria computacional sob uma perspectiva ética, explicável e reprodutível. Sua relevância social reside no potencial de apoiar estratégias de triagem precoce em contextos clínicos e educacionais, enquanto sua relevância científica deriva da integração inovadora entre modelagem simbólica, redes neurais e técnicas de interpretabilidade.

1.8 Objetivos

A presente pesquisa tem por finalidade investigar, sob a ótica da Engenharia da Computação aplicada à Neuropsiquiatria Infantil, o potencial das redes neurais artificiais como instrumentos de apoio à identificação precoce de transtornos de ansiedade em crianças em idade pré-escolar. A expectativa é que os resultados contribuam para o desenvolvimento de ferramentas clínicas mais objetivas, escaláveis e fundamentadas

em evidências quantitativas, capazes de auxiliar a triagem e o acompanhamento psicopatológico infantil.

1.8.1 Objetivo Geral

- Desenvolver e avaliar modelos de redes neurais artificiais capazes de quantificar o risco de transtornos de ansiedade em crianças pré-escolares, tomando como variável-alvo o constructo `Anxiety_Multilevel` e integrando variáveis psicométricas e comportamentais, de modo a capturar padrões complexos e não lineares associados à expressão ansiosa e fornecer subsídios computacionais para a prática clínica baseada em dados.

1.8.2 Objetivos Específicos

- Implementar e comparar o desempenho de diferentes arquiteturas neurais — incluindo o *Multilayer Perceptron (MLP)* e o *Autoencoder* acoplado à clusterização *KMeans* — frente ao modelo simbólico *Alternating Decision Tree (ADTree)*, cuja versão empregada neste estudo representa uma aproximação funcional do algoritmo original descrito na literatura, visando à classificação multinível do risco de ansiedade (níveis 0 a 3) e à identificação da abordagem com maior robustez estatística e interpretabilidade computacional;

- Empregar métodos avançados de interpretabilidade, como o *SHapley Additive exPlanations (SHAP)* [9], aplicados ao modelo com melhor desempenho, de modo a tornar as decisões algorítmicas transparentes e compreensíveis por profissionais da saúde mental, facilitando a tradução dos resultados em inferências clínicas;

- Analisar o impacto de variáveis psicométricas, comportamentais, fisiológicas e demográficas na acurácia e na capacidade de generalização dos modelos neurais, identificando os atributos de maior contribuição para a predição do risco e delineando possíveis marcadores computacionais úteis à triagem precoce;

- Estabelecer paralelos entre as variáveis mais relevantes identificadas pelos modelos e construtos clínicos reconhecidos pelo DSM-5 e pela CID-11, reforçando a coerência psicométrica e a validade clínica das predições;

- Propor um *pipeline* metodológico reprodutível que una pré-processamento, modelagem supervisionada e não supervisionada, validação e análise explicável, contribuindo para o avanço da psiquiatria computacional aplicada à infância.

Dessa forma, o conjunto de objetivos delineado busca aliar rigor científico e aplicabilidade prática, oferecendo uma contribuição metodológica relevante tanto para a Engenharia da Computação quanto para a saúde mental infantil. A próxima seção apresenta o referencial teórico que fundamenta as abordagens aqui propostas.

2 REFERENCIAL TEÓRICO

A Inteligência Artificial (IA), enquanto campo multidisciplinar, tem evoluído rapidamente nas últimas décadas, especialmente em direção à modelagem de fenômenos complexos de natureza cognitiva e emocional. No domínio da neuropsiquiatria infantil, a aplicação de métodos de aprendizado profundo representa um avanço significativo, ao permitir a análise de grandes volumes de dados psicométricos e fisiológicos de forma automatizada e explicável [33, 34, 35].

Essa convergência entre neurociência e engenharia tem impulsionado o surgimento de uma área emergente denominada psiquiatria computacional, cujo objetivo é integrar modelos matemáticos e algoritmos de aprendizado para apoiar diagnósticos e decisões clínicas [36, 37]. Nesse contexto, redes neurais artificiais (RNAs) têm se destacado pela capacidade de capturar relações não lineares e padrões sutis em dados complexos — característica essencial para o estudo dos transtornos ansiosos, cuja expressão se dá em múltiplas dimensões (comportamental, afetiva e fisiológica) [38, 32].

O presente capítulo apresenta os fundamentos teóricos das arquiteturas de redes neurais empregadas neste trabalho — especificamente o *Multilayer Perceptron (MLP)* e o *Autoencoder* — bem como a técnica de aprendizado de máquina de natureza simbólica utilizada para fins comparativos, o modelo *Alternating Decision Tree (ADTree)*. Por fim, são apresentadas as métricas de avaliação empregadas neste estudo, de modo a descrever, de forma abrangente, os princípios de funcionamento, os fundamentos matemáticos e as aplicações dessas abordagens no contexto da saúde mental infantil.

2.1 Fundamentos de Redes Neurais Artificiais

As redes neurais artificiais são sistemas computacionais inspirados na estrutura e funcionamento do cérebro biológico. Formalmente, uma RNA é composta por unidades denominadas neurônios artificiais, conectadas entre si por pesos sinápticos ajustáveis, que modulam a intensidade da propagação do sinal. Cada neurônio realiza uma operação matemática composta por uma soma ponderada das entradas, seguida

por uma função de ativação não linear [39, 33]:

$$a_j = \phi\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right) \tag{1}$$

onde a_j representa a ativação do neurônio j, w_{ij} é o peso associado à conexão entre o neurônio i e o neurônio j, x_i é o valor de entrada e b_j é o termo de viés. A função $\phi(\cdot)$ é geralmente uma função não linear, como a sigmoide, tangente hiperbólica ou ReLU, que introduz complexidade e expressividade à rede.

Durante o processo de aprendizado supervisionado, os pesos são ajustados para minimizar uma função de custo $\mathcal{L}$ que mede o erro entre as predições da rede e os valores reais. O método mais comum é o *backpropagation*, baseado no gradiente descendente [40]:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta\frac{\partial\mathcal{L}}{\partial w_{ij}} \tag{2}$$

em que η é a taxa de aprendizado. Essa equação formaliza a essência do aprendizado por correção de erro, permitindo que a rede ajuste seus parâmetros em direção a uma minimização local da função de perda.

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{3}$$

A Equação (3) representa a função de custo mais comum em tarefas de classificação, a entropia cruzada, onde y_i é o rótulo verdadeiro e $\hat{y}_i$ é a probabilidade prevista pela rede para a classe i.

2.2 *Multilayer Perceptron (MLP)*

O *Multilayer Perceptron (MLP)* é uma das arquiteturas mais clássicas de redes neurais artificiais. Trata-se de um modelo *feedforward*, ou seja, os dados fluem unidirecionalmente da camada de entrada para a camada de saída, passando por uma ou mais camadas ocultas. Cada camada executa transformações lineares seguidas de funções de ativação não lineares, o que permite à rede aproximar funções complexas e não lineares [41]. Além disso, o aprendizado supervisionado do *MLP* ocorre por meio da retropropagação do erro, possibilitando a adaptação dos pesos sinápticos a partir

de exemplos rotulados.

Matematicamente, as operações de uma *MLP* podem ser descritas por:

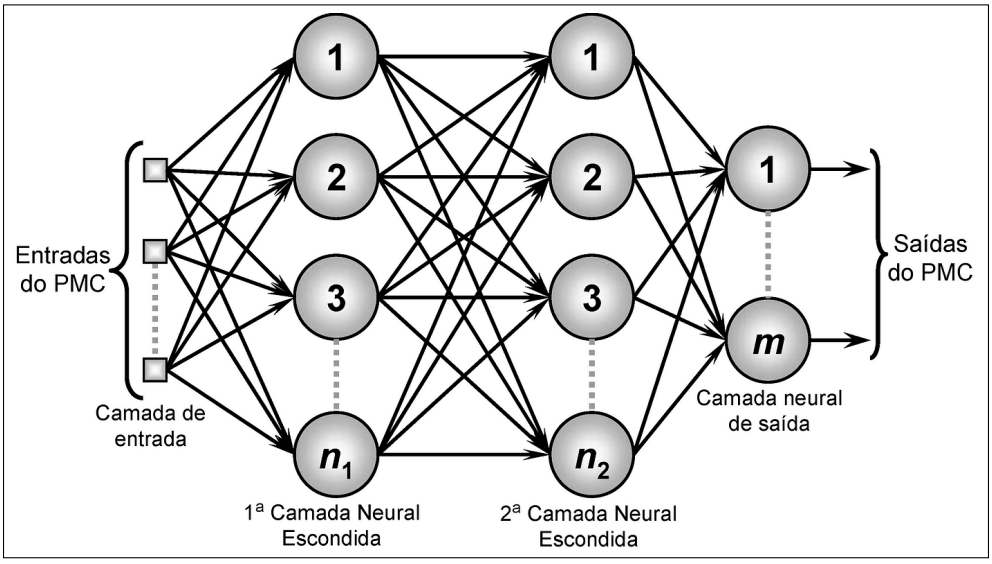
$$\mathbf{h} = \phi(\mathbf{W}^{(1)} \cdot \mathbf{x} + \mathbf{b}^{(1)}) \tag{4}$$

$$\hat{\mathbf{y}} = \psi(\mathbf{W}^{(2)} \cdot \mathbf{h} + \mathbf{b}^{(2)}) \tag{5}$$

em que $\mathbf{x}$ é o vetor de entrada, $\mathbf{W}^{(1)}$ e $\mathbf{W}^{(2)}$ são as matrizes de pesos das camadas, ϕ e ψ são funções de ativação (por exemplo, ReLU ou *softmax*), e $\hat{\mathbf{y}}$ é a saída final da rede.

As *MLPs* são amplamente empregadas em contextos de predição clínica, pois conseguem modelar interações não lineares entre variáveis psicométricas, comportamentais e fisiológicas [42, 32]. No presente trabalho, a *MLP* é utilizada como arquitetura supervisionada de base, permitindo comparar seu desempenho frente a abordagens simbólicas e não supervisionadas.

Figura 1 – Arquitetura típica de um *Multilayer Perceptron (MLP)*



Fonte: Adaptado pela autora a partir de [43].

A estrutura típica de um *Multilayer Perceptron (MLP)* é composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada nó representa um neurônio artificial responsável por realizar uma combinação linear dos sinais de entrada, seguida da aplicação de uma função de ativação não linear, como *ReLU* ou *sigmoide*. Durante o treinamento supervisionado, os pesos sinápticos são

ajustados de forma iterativa com o objetivo de minimizar a função de perda associada ao erro de predição. Essa arquitetura permite a modelagem de relações complexas e não lineares entre variáveis, sendo particularmente adequada para a análise de padrões psicométricos e psicofisiológicos, como os investigados neste estudo.

O treinamento da *MLP* envolve a minimização iterativa da função de custo, geralmente utilizando o otimizador *Adam* [44], definido como uma versão adaptativa do gradiente descendente estocástico. Essa abordagem combina momentos de primeira e segunda ordem para ajustar dinamicamente a taxa de aprendizado, promovendo convergência estável mesmo em espaços de parâmetros altamente não convexos.

2.3 *Autoencoders*

Os *Autoencoders* constituem uma classe de redes neurais não supervisionadas projetadas para aprender representações compactas (latentes) de dados de alta dimensionalidade. Sua estrutura é composta por duas partes principais: o *encoder*, responsável por comprimir o vetor de entrada $\mathbf{x}$ em uma representação latente $\mathbf{z}$, e o *decoder*, que tenta reconstruir o dado original a partir de $\mathbf{z}$ [45, 46].

Formalmente, o processo pode ser descrito como:

$$\mathbf{z} = f_{\text{enc}}(\mathbf{x}) = \phi(\mathbf{W}_e \cdot \mathbf{x} + \mathbf{b}_e) \tag{6}$$

$$\hat{\mathbf{x}} = f_{\text{dec}}(\mathbf{z}) = \psi(\mathbf{W}_d \cdot \mathbf{z} + \mathbf{b}_d) \tag{7}$$

O objetivo do treinamento é minimizar o erro de reconstrução entre a entrada e a saída:

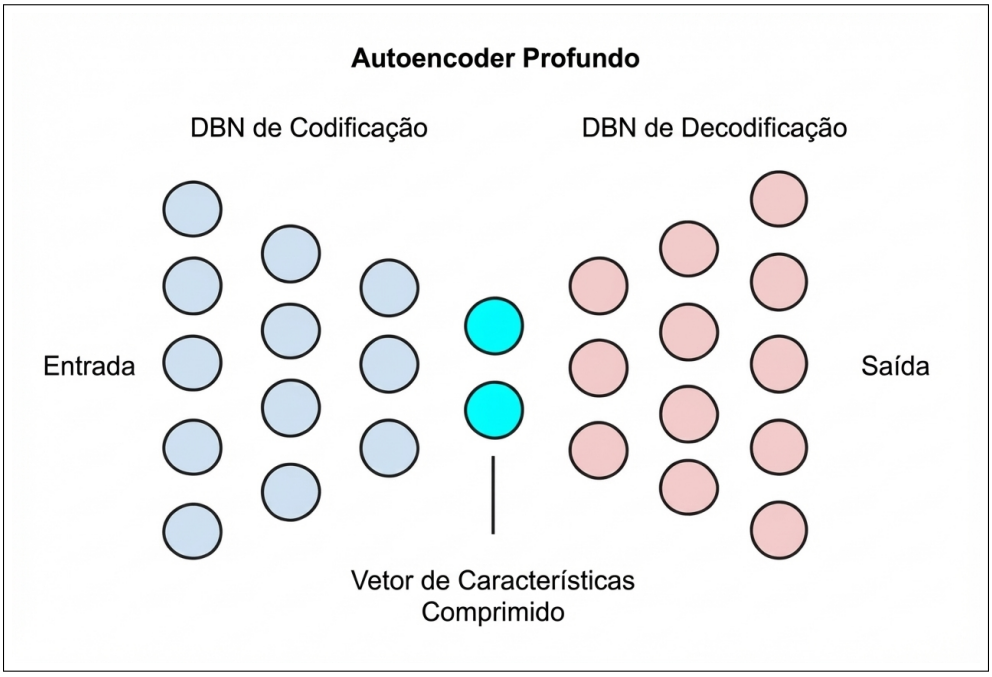
$$\mathcal{L}_{\text{rec}} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \tag{8}$$

onde $\mathcal{L}_{\text{rec}}$ representa a soma dos erros quadráticos médios (MSE). Essa formulação leva o modelo a aprender as características mais relevantes do conjunto de dados, reduzindo ruído e redundâncias.

Em aplicações biomédicas, *autoencoders* são amplamente empregados para compressão de sinais fisiológicos e redução de dimensionalidade de dados clínicos

[47, 29]. No contexto deste estudo, a representação latente $\mathbf{z}$ extraída do *autoencoder* é posteriormente utilizada como entrada para algoritmos de clusterização (*KMeans*), a fim de identificar agrupamentos naturais entre diferentes níveis de ansiedade.

Figura 2 – Estrutura conceitual de um *Autoencoder*



Fonte: Adaptado pela autora a partir de [48].

A estrutura conceitual de um *Autoencoder* é composta por duas partes principais: o *encoder*, responsável por transformar o vetor de entrada $\mathbf{x}$ em uma representação latente $\mathbf{z}$, e o decoder, que reconstrói $\hat{\mathbf{x}}$ a partir dessa codificação comprimida. O processo de treinamento busca minimizar o erro de reconstrução entre a entrada e a saída, levando a rede a aprender representações internas compactas e informativas. No presente estudo, essas representações latentes são posteriormente utilizadas para análises de clusterização e detecção de padrões associados aos níveis de risco de ansiedade infantil.

Autoencoders variacionais (VAEs), uma extensão probabilística dos *autoencoders* tradicionais, introduzem regularização no espaço latente, impondo que $\mathbf{z}$ siga uma distribuição normal multivariada [48]. Essa abordagem permite a geração de novos exemplos sintéticos e melhora a continuidade das representações, sendo promissora para modelagem de fenômenos psicológicos contínuos, como o espectro ansioso.

2.4 Modelo Simbólico: *Alternating Decision Tree (ADTree)*

Antes do advento das redes profundas, a modelagem de dados clínicos baseava-se fortemente em classificadores simbólicos, como árvores de decisão, regressões e máquinas de vetor de suporte. Dentre esses métodos, o *Alternating Decision Tree (ADTree)* destacou-se como um modelo híbrido que combina regras de decisão com técnicas de *boosting* [49]. O *boosting* consiste em um procedimento iterativo que constrói um conjunto de classificadores fracos, ajustando-os sequencialmente de modo que cada novo classificador concentre-se nos erros cometidos pelos anteriores, resultando em um modelo final mais robusto e com maior capacidade de generalização.

O *ADTree* combina nós de decisão e nós de predição em uma estrutura hierárquica, permitindo a construção de modelos interpretáveis e eficientes. Cada caminho da raiz a uma folha representa uma regra lógica de classificação, e a soma dos valores dos nós visitados define o escore final da predição [6].

Matematicamente, a predição de um *ADTree* pode ser expressa como:

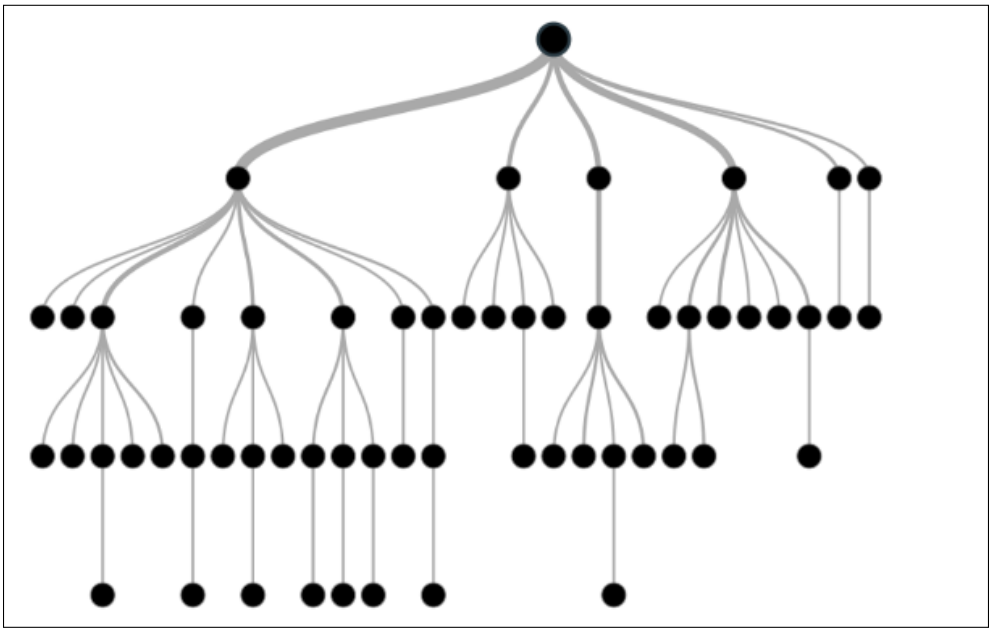
$$f(x) = \sum_{i \in P(x)} \alpha_i \tag{9}$$

em que $P(x)$ é o conjunto de nós percorridos pela amostra x, e α_i é o valor associado a cada nó de predição. O sinal de $f(x)$ determina a classe predita.

Estrutura de uma *Alternating Decision Tree (ADTree)*, composta por nós de decisão e nós de predição. Cada caminho representa uma regra lógica construída a partir de divisões hierárquicas dos dados. A soma dos pesos dos nós visitados determina o escore final da amostra, que define a classe predita. A principal vantagem do *ADTree* reside em sua alta interpretabilidade, o que o torna particularmente adequado para contextos clínicos e psicológicos nos quais a explicabilidade das decisões é um requisito essencial [6].

Entretanto, estudos recentes [25] indicam que a natureza determinística do *ADTree* e sua sensibilidade a amostras desbalanceadas restringem significativamente sua capacidade de generalização, sobretudo quando o modelo é aplicado a fenômenos psicológicos de natureza contínua e probabilística.

Figura 3 – Estrutura conceitual da *Alternating Decision Tree (ADTree)*



Fonte: Adaptado pela autora a partir de [50].

Por outro lado, os modelos conexionistas, *Multilayer Perceptron (MLP)* e *Autoencoder*, superam essas limitações ao representar de forma distribuída, hierárquica e probabilística os estados emocionais subjacentes. Além disso, quando integrados a métodos de explicabilidade, como o *SHapley Additive exPlanations (SHAP)* [9], esses modelos mantêm coerência semântica e interpretabilidade clínica, conciliando desempenho preditivo e transparência algorítmica — características fundamentais para a aplicação da inteligência artificial em ambientes de decisão médica.

2.5 Comparação entre Arquiteturas e Aspectos Computacionais

Cada arquitetura analisada neste referencial teórico — *Multilayer Perceptron (MLP)*, *Autoencoder* e *Alternating Decision Tree (ADTree)* — apresenta características distintas quando aplicada à modelagem de fenômenos psicofisiológicos e emocionais. Essas abordagens diferenciam-se fundamentalmente quanto à natureza do aprendizado, ao grau de interpretabilidade e à capacidade de generalização frente à complexidade dos dados clínicos. A Tabela 1 sintetiza os principais aspectos comparativos entre essas arquiteturas, considerando parâmetros como tipo de aprendizado, vantagens, limitações e aplicabilidade clínica.

Tabela 1: **Comparativo entre arquiteturas utilizadas no estudo**

Arquitetura	Tipo de Aprendizado	Vantagens	Limitações
MLP	Supervisionado	Modela relações não lineares; boa capacidade de generalização; adequado a dados tabulares.	Dependência de dados rotulados; sensível à escolha de hiperparâmetros.
Autoencoder	Não supervisionado	Redução de dimensionalidade; extração de representações latentes; suporte à análise exploratória.	Possível perda de informação na reconstrução; dependência da configuração do espaço latente.
ADTree	Simbólico	Alta interpretabilidade; baixo custo computacional; decisões baseadas em regras explícitas.	Baixa sensibilidade; limitada capacidade de generalização em dados não lineares.

Fonte: Elaboração própria (2025).

A complementaridade entre essas abordagens é evidente. O modelo simbólico *ADTree* privilegia transparência e rastreabilidade das decisões, características desejáveis em contextos clínicos, porém apresenta limitações significativas na modelagem de fenômenos psicológicos complexos, marcados por não linearidade e variabilidade interindividual. Em contrapartida, os modelos conexionistas — *MLP* e *Autoencoder* — oferecem maior capacidade de abstração e captura de padrões latentes, ainda que com diferentes graus de supervisão e interpretabilidade.

A escolha do *MLP* e do *Autoencoder* neste trabalho fundamenta-se diretamente na natureza dos dados analisados, compostos exclusivamente por medidas psicométricas e comportamentais organizadas em formato tabular. Nesse contexto, arquiteturas totalmente conectadas apresentam desempenho consistente e equilíbrio entre complexidade computacional e capacidade preditiva, sem a introdução de estruturas desnecessárias ao problema.

O *Multilayer Perceptron (MLP)* foi selecionado como principal modelo supervisionado por sua reconhecida eficiência em tarefas de classificação multiclasse envolvendo dados tabulares, bem como por sua habilidade em modelar relações não lineares entre variáveis psicométricas. Além disso, sua arquitetura relativamente simples favorece a aplicação de métodos de interpretabilidade, como o *SHapley Additive exPlanations (SHAP)*, permitindo alinhar desempenho preditivo e coerência clínica.

O *Autoencoder*, por sua vez, foi incorporado como abordagem complementar não supervisionada, voltada à análise latente da estrutura dos dados. Ao permitir a

redução de dimensionalidade e a extração de representações internas compactas, o modelo possibilita a identificação de padrões ocultos associados ao risco de ansiedade infantil. Quando combinado ao algoritmo de clusterização *KMeans*, o *Autoencoder* fornece uma perspectiva adicional sobre a organização interna dos dados, reforçando a interpretação dimensional do fenômeno ansioso.

Assim, a comparação entre *ADTree*, *MLP* e *Autoencoder* reflete três paradigmas distintos — simbólico, supervisionado e não supervisionado — alinhados tanto às características do conjunto de dados quanto aos objetivos centrais da pesquisa. Essa escolha metodológica privilegia soluções computacionalmente eficientes, reprodutíveis e clinicamente interpretáveis, em consonância com os princípios da psiquiatria computacional aplicada à infância.

Dessa forma, o presente trabalho propõe uma integração equilibrada entre precisão estatística e explicabilidade clínica, explorando simultaneamente a robustez numérica dos modelos conexionistas e a transparência decisória das abordagens simbólicas.

2.6 Aspectos Matemáticos Avançados e Regularização

A estabilidade e a capacidade de generalização de modelos neurais dependem fortemente de técnicas de regularização, cujo objetivo é mitigar o sobreajuste (*overfitting*) e assegurar desempenho consistente em dados não vistos. Em arquiteturas densas aplicadas a bases clínicas de dimensão moderada, a regularização torna-se particularmente relevante, uma vez que a elevada flexibilidade do modelo pode levar à memorização de padrões específicos do conjunto de treinamento.

Neste estudo, o mecanismo de regularização adotado foi exclusivamente o *dropout*, técnica amplamente empregada em redes neurais artificiais por sua eficácia em promover robustez e reduzir a dependência excessiva de unidades individuais. Introduzido por [51], o *dropout* consiste na desativação aleatória de neurônios durante o treinamento, forçando a rede a aprender representações distribuídas e redundantes, o que contribui para maior capacidade de generalização.

Matematicamente, o *dropout* pode ser representado como:

$$h_i^{(l)} = r_i^{(l)} \cdot \phi(\mathbf{W}^{(l)} h^{(l-1)} + \mathbf{b}^{(l)}) \tag{10}$$

em que $r_i^{(l)}$ é uma variável aleatória de Bernoulli, assumindo valor 1 com probabilidade p (neurônio ativo) e 0 com probabilidade $1 - p$ (neurônio desligado), $\phi(\cdot)$ representa a função de ativação da camada, $\mathbf{W}^{(l)}$ o conjunto de pesos e $\mathbf{b}^{(l)}$ o vetor de vieses da camada l.

O *dropout* foi aplicado às camadas densas dos modelos conexionistas avaliados neste trabalho, com taxas variando entre 0.2 e 0.5, conforme o comportamento observado durante o processo de validação. Essas taxas foram selecionadas empiricamente, visando reduzir oscilações da função de perda, estabilizar o treinamento e minimizar o risco de sobreajuste sem comprometer a capacidade representacional das redes.

A adoção exclusiva do *dropout* como estratégia de regularização mostrou-se adequada às características do conjunto de dados, composto por variáveis psicométricas e comportamentais em formato tabular, e aos objetivos do estudo, que priorizam simplicidade computacional, reprodutibilidade e interpretabilidade clínica. Dessa forma, o mecanismo de regularização empregado contribui para o equilíbrio entre desempenho preditivo e estabilidade do modelo, alinhando-se às boas práticas da psiquiatria computacional aplicada à infância.

2.7 Métodos de Explicabilidade: *SHAP* e Interpretação de Modelos

Em aplicações clínicas, a interpretabilidade dos modelos é uma exigência ética e metodológica. O método *SHAP (SHapley Additive exPlanations)* [9] oferece uma abordagem consistente e teoricamente fundamentada para quantificar a contribuição de cada variável nas decisões do modelo.

Com base na teoria dos jogos cooperativos, o *SHAP* atribui um valor ϕ_i a cada característica i, refletindo sua importância média marginal para a predição:

$$f(x) = f_0 + \sum_{i=1}^{M} \phi_i \tag{11}$$

onde f_0 representa a predição média do modelo e ϕ_i é a contribuição da variável i. Esse método garante propriedades de *simetria*, *linearidade* e *consistência*, tornando-se especialmente útil na análise de variáveis psicométricas e comportamentais.

No presente estudo, a aplicação do *SHAP* permite não apenas identificar quais variáveis têm maior influência na classificação do risco de ansiedade, mas também vali-

dar a coerência clínica das decisões — assegurando que o modelo reflita diagnósticos estabelecidos, como ansiedade antecipatória, afeto ansioso e evitação social [7, 8].

2.8 Métricas de Avaliação de Desempenho

A mensuração objetiva do desempenho de modelos de aprendizado de máquina é etapa essencial para a validação científica e clínica de qualquer sistema de classificação [52]. Em particular, na modelagem de risco de transtornos mentais, as métricas não apenas quantificam a eficácia computacional, mas também fornecem indicativos de aplicabilidade prática, confiabilidade diagnóstica e segurança preditiva. Nesta seção, descrevem-se as principais métricas utilizadas para avaliar os modelos propostos — com ênfase em sua fundamentação matemática, relevância estatística e interpretação clínica.

2.8.1 Matriz de Confusão

A matriz de confusão é a estrutura fundamental para avaliação de classificadores, pois descreve de forma explícita as relações entre as classes previstas e as classes verdadeiras [53]. Para um problema binário, a matriz é definida conforme:

	Predito: Positivo	Predito: Negativo
Real: Positivo	TP	FN
Real: Negativo	FP	TN

onde:

- TP — Verdadeiros Positivos: casos corretamente identificados como positivos (por exemplo, crianças com alto risco de ansiedade corretamente detectadas);
- TN — Verdadeiros Negativos: casos corretamente classificados como negativos;
- FP — Falsos Positivos: indivíduos identificados incorretamente como positivos;
- FN — Falsos Negativos: indivíduos positivos não detectados pelo modelo.

Essa matriz permite derivar métricas que medem precisão, sensibilidade, especificidade e acurácia, fornecendo uma visão holística da performance do sistema.

Figura 4 – Estrutura conceitual da Matriz de Confusão

		Detectada	
		Sim	**Não**
Real	**Sim**	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Adaptado pela autora a partir de [50].

A matriz de confusão permite avaliar simultaneamente os acertos e erros do modelo, distinguindo entre falsos positivos e falsos negativos. Essa distinção é essencial em contextos clínicos, nos quais um falso negativo representa um risco ético maior que um falso positivo, especialmente em triagens de ansiedade infantil.

2.8.2 Acurácia (*Accuracy*)

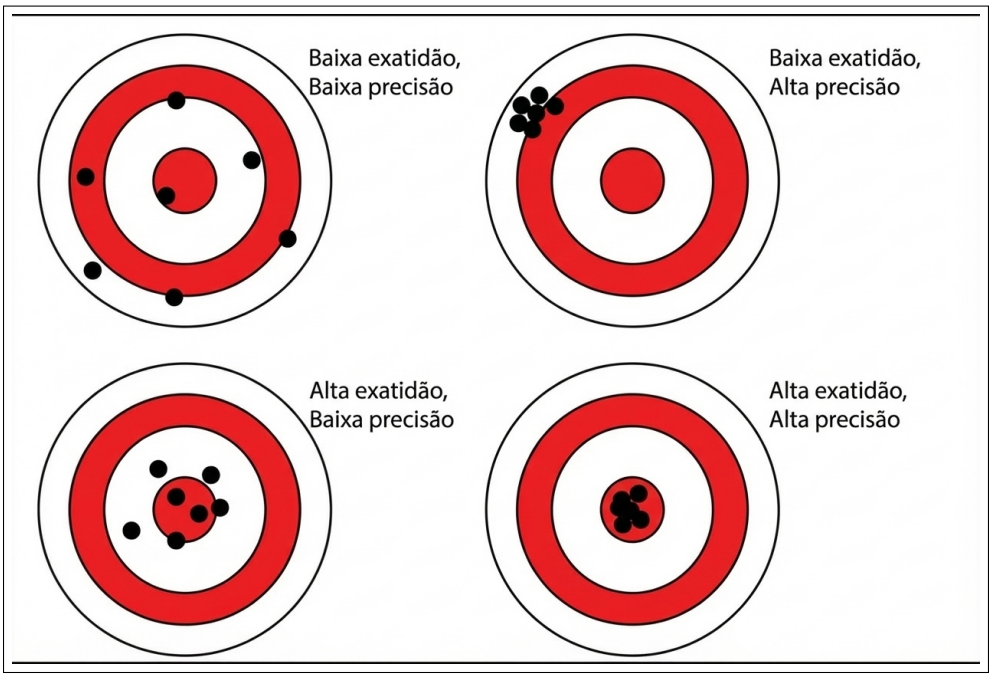
A acurácia representa a proporção total de classificações corretas em relação ao número total de amostras avaliadas [54]. Matematicamente, define-se como:

$$\textit{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

Embora amplamente utilizada, a acurácia pode ser enganosa quando há desequilíbrio entre as classes — como ocorre frequentemente em amostras clínicas, onde o número de indivíduos com risco elevado de ansiedade é substancialmente menor que o de baixo risco. Nesses casos, modelos triviais podem alcançar alta acurácia simplesmente favorecendo a classe majoritária.

Por isso, a referida métrica deve ser interpretada com cautela, sendo mais adequada quando as classes são balanceadas ou quando se deseja uma visão geral da eficiência do sistema. Em contextos de saúde mental, ela fornece uma medida de desempenho global, mas precisa ser complementada por métricas que capturem a sensibilidade às classes minoritárias.

Figura 5 – Relação conceitual entre Acurácia e Precisão



Fonte: Adaptado pela autora a partir de [55].

A exemplo dos recursos visuais da acurácia, mesmo com alta taxa global de acertos, um modelo pode apresentar desempenho insatisfatório em classes minoritárias. Isso reforça a importância de utilizar métricas complementares em estudos clínicos com desbalanceamento de classes, garantindo uma avaliação mais justa e representativa do desempenho preditivo.

2.8.3 Precisão (*Precision*) e Sensibilidade (*Recall*)

A precisão mede a proporção de predições positivas que são realmente corretas, sendo calculada como:

$$\textit{Precision} = \frac{TP}{TP + FP} \tag{13}$$

Já a sensibilidade (ou *Recall*) indica a proporção de casos positivos reais que o modelo foi capaz de identificar corretamente:

$$\textit{Recall} = \frac{TP}{TP + FN} \tag{14}$$

Enquanto a precisão avalia a confiabilidade das predições positivas, a sensibilidade mede a capacidade do modelo em não deixar escapar casos reais de ansiedade.

Em contextos clínicos, a sensibilidade assume maior relevância, pois a omissão de casos verdadeiros pode ter consequências graves para o diagnóstico precoce e a intervenção terapêutica.

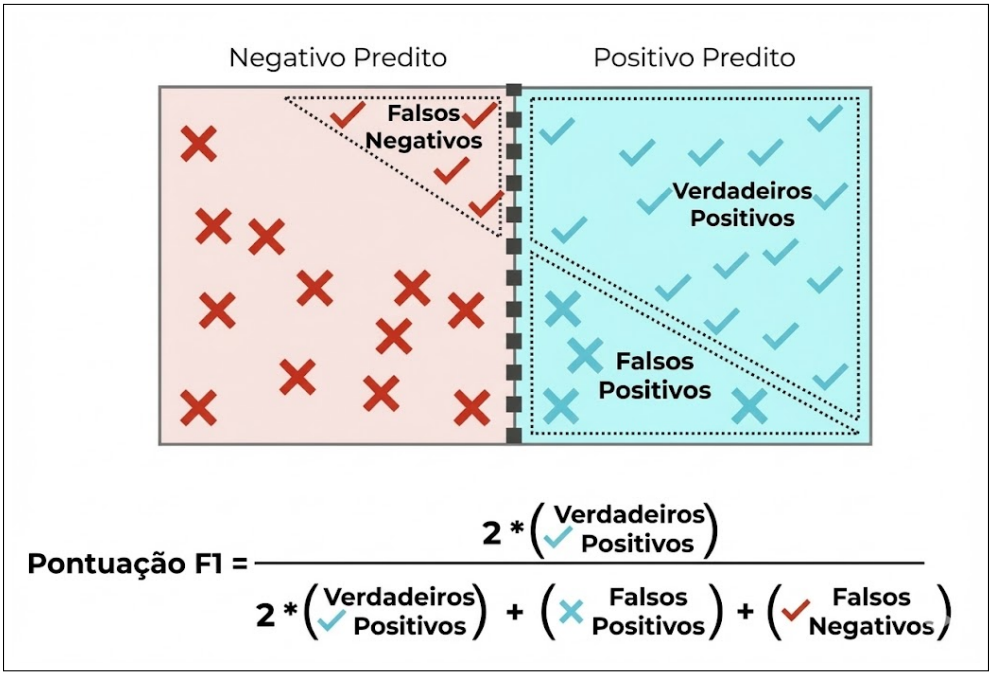
2.8.4 *F1-Score*

O *F1-score* é a média harmônica entre precisão e sensibilidade, e busca equilibrar o desempenho entre ambas as dimensões [52]:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

Essa métrica é especialmente útil em bases desbalanceadas, pois penaliza desproporções entre falsos positivos e falsos negativos. Valores elevados de F1 indicam que o modelo consegue manter simultaneamente boa precisão e boa sensibilidade.

Figura 6 – Relação entre Precisão, Sensibilidade e *F1-Score*



Fonte: Adaptado pela autora a partir de [56].

O quadro ilustra o equilíbrio proporcionado pelo *F1-Score* entre as métricas de precisão e sensibilidade. Em aplicações clínicas, esse equilíbrio é essencial para evitar tanto a negligência de casos (baixa sensibilidade) quanto diagnósticos incorretos (baixa precisão). No presente estudo, o *F1-score* é fundamental para avaliar a eficácia dos classificadores em contextos de triagem de ansiedade infantil, garantindo que os

modelos não apenas acertem, mas também generalizem corretamente para novos indivíduos.

2.8.5 Curva *ROC e AUC*

A *Receiver Operating Characteristic (ROC)* é uma curva que relaciona a taxa de verdadeiros positivos (*True Positive Rate*, TPR) e a taxa de falsos positivos (*False Positive Rate*, FPR) para diferentes limiares de decisão [57]. Define-se matematicamente como:

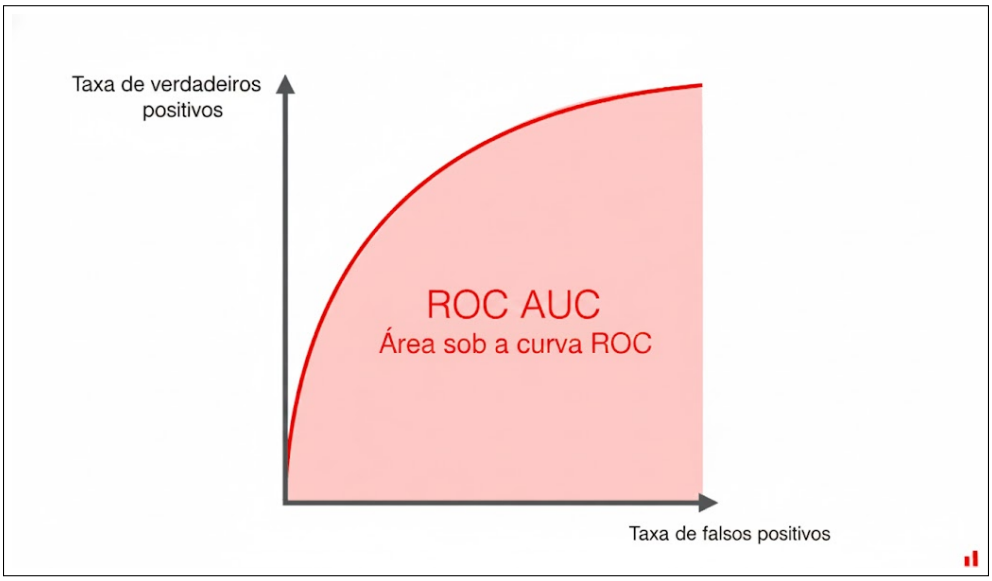
$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \tag{16}$$

A área sob a curva (*AUC – Area Under the Curve*) é calculada como:

$$AUC = \int_0^1 TPR(FPR^{-1}(x))\, dx \tag{17}$$

O valor de *AUC* varia entre 0 e 1, sendo que quanto mais próximo de 1, maior a capacidade discriminatória do modelo. Um valor de 0,5 indica desempenho equivalente a um classificador aleatório.

Figura 7 – Curva *ROC* e área sob a curva *(AUC)*



Fonte: Adaptado pela autora a partir de [58].

A curva *ROC* ilustra o equilíbrio (*trade-off*) entre sensibilidade e especificidade. A área sob a curva *AUC* fornece uma métrica robusta de separabilidade entre classes. Em modelos aplicados à saúde mental infantil, valores de AUC superiores a 0,85

indicam alto potencial clínico de discriminação entre níveis de risco. Em psiquiatria computacional, um *AUC* elevado indica que o modelo é eficaz em distinguir indivíduos com alto e baixo risco de ansiedade, mesmo sob condições de ruído e heterogeneidade dos dados comportamentais.

2.9 Síntese Final do Referencial

O referencial teórico apresentado neste capítulo evidencia que a aplicação de modelos de aprendizado profundo ao domínio da saúde mental infantil alcançou um estágio de maturidade metodológica, sustentado por fundamentos matemáticos sólidos, rigor estatístico e crescente preocupação com interpretabilidade. Nesse contexto, as Redes Neurais Artificiais (RNAs) assumem papel central ao possibilitar a modelagem de relações não lineares entre variáveis psicométricas, comportamentais e fisiológicas, componentes essenciais para a compreensão da complexidade dos transtornos ansiosos.

Entre as arquiteturas conexionistas analisadas, o *Multilayer Perceptron (MLP)* e os *autoencoders* constituem o núcleo conceitual deste estudo. O *MLP*, por sua natureza supervisionada e capacidade de aprender mapeamentos altamente não lineares, mostra-se particularmente adequado para tarefas de classificação em dados tabulares, como aqueles provenientes de instrumentos psicométricos. Já o *autoencoder*, ao operar de forma não supervisionada, permite extrair representações latentes compactas e informativas do conjunto de dados, possibilitando tanto a redução de dimensionalidade quanto a identificação de padrões ocultos relevantes ao risco de ansiedade infantil. A combinação entre essas abordagens fornece perspectivas complementares sobre a estrutura do fenômeno investigado, integrando predição supervisionada e análise exploratória latente.

O modelo simbólico *Alternating Decision Tree (ADTree)* oferece, por sua vez, uma referência interpretável e consolidada na literatura de psiquiatria computacional. Embora apresente limitações de generalização em cenários caracterizados por elevada não linearidade e heterogeneidade, seu valor reside na transparência lógica de suas regras de decisão, o que o torna um contraponto relevante às arquiteturas conexionistas e um elemento apropriado para comparações metodológicas.

A necessidade de estabilidade e capacidade de generalização dos modelos é

reforçada pela discussão sobre regularização, especialmente em bases clínicas de dimensão moderada. Nesse sentido, o uso de técnicas como o *dropout* destaca-se por promover modelos mais robustos ao reduzir o risco de sobreajuste, sem introduzir complexidade computacional excessiva. Essa escolha reflete o compromisso deste estudo com simplicidade, reprodutibilidade e rigor metodológico.

A interpretabilidade, elemento indispensável em aplicações clínicas, é abordada por meio do método *SHapley Additive exPlanations (SHAP)*, que permite decompor as predições dos modelos em contribuições individuais das variáveis de entrada. Essa abordagem complementa a capacidade representacional das redes neurais com explicações transparentes, assegurando coerência clínica e facilitando a validação dos resultados à luz dos construtos diagnósticos estabelecidos pelo DSM-5 e pela CID-11.

Por fim, a revisão das métricas de avaliação — incluindo acurácia, precisão, sensibilidade, *F1-Score* e *AUC-ROC* — estabelece as bases para a análise sistemática do desempenho dos modelos, considerando as particularidades do desbalanceamento e da heterogeneidade inerentes a amostras clínicas.

Dessa forma, o referencial teórico consolida os fundamentos que sustentam as escolhas metodológicas deste trabalho, destacando a integração entre abordagens simbólicas e conexionistas, a centralidade da explicabilidade, o rigor matemático e estatístico e a adequação das arquiteturas selecionadas ao tipo de dado analisado. Esses elementos estruturam o arcabouço conceitual que orienta, no capítulo seguinte, o detalhamento da metodologia empregada para a modelagem, validação e interpretação dos resultados no contexto da neuropsiquiatria infantil, garantindo alinhamento conceitual entre referencial teórico, desenho experimental e análise dos resultados.

3 METODOLOGIA

A presente seção descreve em detalhe os fundamentos, materiais, procedimentos e técnicas adotadas para o desenvolvimento do estudo, com ênfase na replicabilidade, no rigor estatístico e na coerência epistemológica. A metodologia foi desenvolvida para permitir a comparação metodologicamente equilibrada entre modelos simbólicos e conexionistas, avaliando sua capacidade de estimar risco de ansiedade infantil em contextos clínicos e pré-clínicos. Foi, adicionalmente, concebida de modo a integrar princípios da Engenharia da Computação, da Psiquiatria Computacional e da Neurociência, estabelecendo um arcabouço quantitativo e experimental capaz de investigar, de forma sistemática, o desempenho de diferentes arquiteturas de Redes Neurais Artificiais (RNAs) aplicadas à detecção precoce de risco de ansiedade infantil.

A escolha dessa abordagem metodológica justifica-se pela natureza multivariada e complexa dos fenômenos emocionais — cuja expressão envolve dimensões psicométricas, fisiológicas e comportamentais interdependentes. Assim, a estratégia experimental proposta busca modelar tais relações de forma matematicamente rigorosa, sem perder a coerência clínica, conciliando interpretabilidade e performance computacional.

O delineamento metodológico foi desenvolvido integralmente em ambiente *in silico*, termo que se refere à execução completa do estudo em ambiente computacional, sem experimentação direta com participantes ou coleta de dados clínicos. Todas as etapas — pré-processamento, treinamento dos modelos, avaliação e validação — foram realizadas exclusivamente por meio de algoritmos e simulações em Python, assegurando rigor ético, segurança e alta reprodutibilidade. O fluxo geral da metodologia é apresentado na Figura 8 e detalhado nas subseções seguintes.

3.1 Estrutura Geral da Metodologia

O presente estudo adota uma metodologia de natureza quantitativa, experimental e computacional, voltada à análise comparativa de modelos de aprendizado supervisionado e não supervisionado para triagem automatizada de ansiedade infantil. A concepção metodológica foi estruturada em três macroetapas complementares:

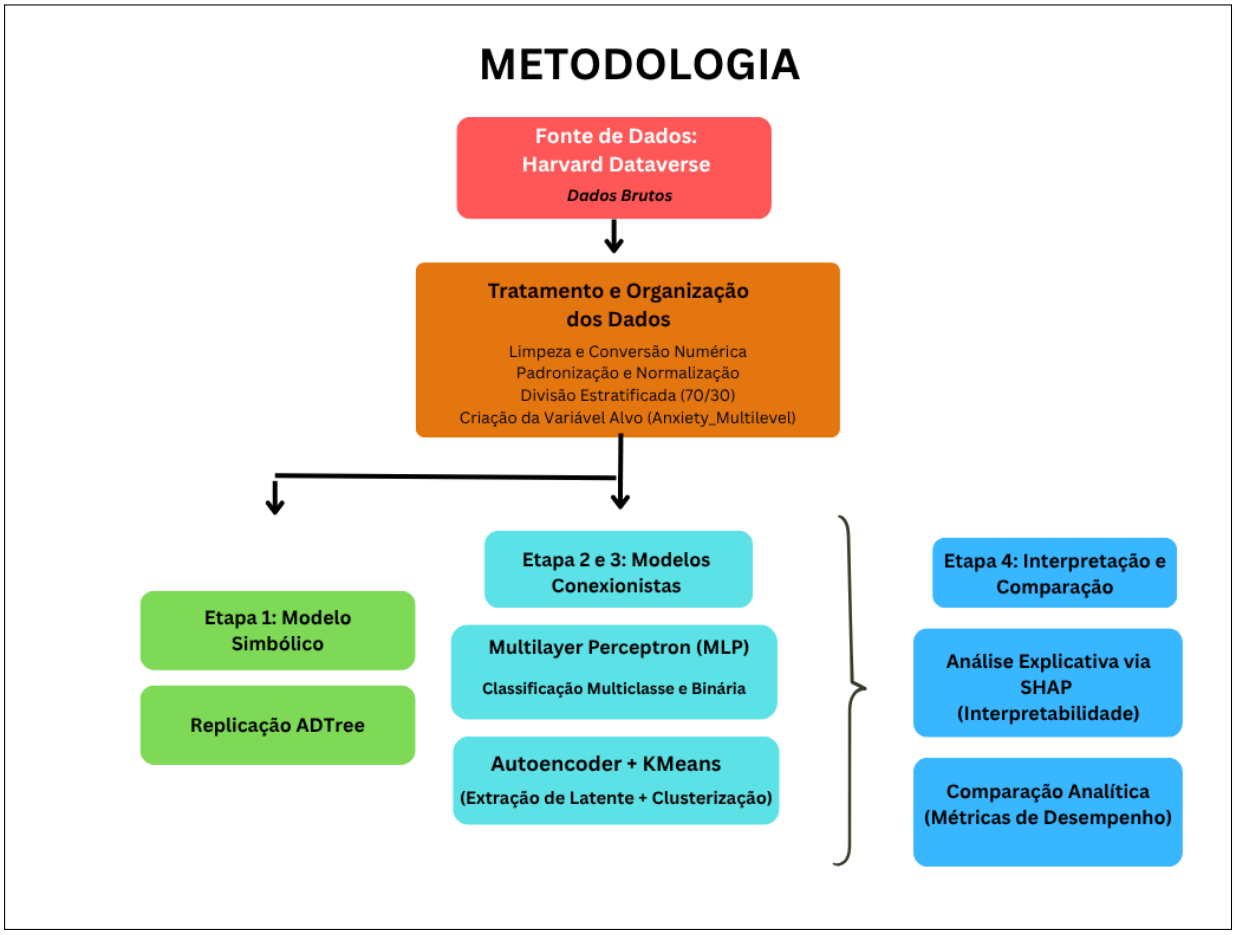
1. **Replicação do modelo simbólico de referência** (*Alternating Decision Tree* –

ADTree), conforme descrito por [6], assegurando consistência metodológica com a literatura original e servindo como linha de base (baseline) simbólica;

2. **Desenvolvimento dos modelos conexionistas supervisionados e não supervisionados**, englobando a implementação de um *Multilayer Perceptron (MLP)* multiclasse e um *Autoencoder + KMeans*, aplicados à base tratada e ajustada (*Anxiety_Multilevel*);

3. **Comparação analítica e explicabilidade**, envolvendo a avaliação de desempenho dos modelos por métricas estatísticas e a análise interpretativa via *SHAP (SHapley Additive Explanations)*.

Figura 8 – Estrutura geral da metodologia proposta



Fonte: Elaboração própria (2025).

Fluxo geral da metodologia adotada no estudo, articulando as etapas de replicação do modelo simbólico de referência (*ADTree*), desenvolvimento de modelos conexionistas supervisionados *(MLP)* e não supervisionados *(Autoencoder + KMeans)*,

além da análise explicativa via *SHAP*. Essas etapas foram integradas sob um mesmo pipeline analítico, garantindo coerência entre replicação, aprendizado supervisionado, inferência não supervisionada e análise explicativa. Tal estrutura possibilita avaliar simultaneamente a acurácia estatística e a validade clínica dos modelos, refletindo a natureza interdisciplinar do trabalho.

3.2 Materiais

A presente subseção descreve os materiais empíricos e computacionais utilizados na pesquisa, incluindo a origem dos dados, o ambiente de desenvolvimento e as ferramentas de software que fundamentam a execução dos experimentos.

3.2.1 Fonte de Dados: *Harvard Dataverse*

A base empírica utilizada neste estudo foi obtida a partir do repositório *Harvard Dataverse*, especificamente do dataset intitulado *"Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach"* [6]. Esse conjunto de dados foi originalmente disponibilizado em *Harvard Dataverse* por [59] e contém informações comportamentais, psicofisiológicas e demográficas de 193 crianças em idade pré-escolar (3 a 5 anos), avaliadas sob protocolos clínicos padronizados na literatura de neuropsiquiatria infantil.

O conjunto de dados resulta de avaliações conduzidas com o instrumento *Preschool Age Psychiatric Assessment (PAPA)*, amplamente utilizado para a caracterização de sintomas internalizantes na infância. O objetivo do estudo original consistiu na aplicação de métodos de aprendizado de máquina com o propósito de reduzir o número de itens clínicos necessários à predição do risco de transtornos de ansiedade, alcançando acurácia superior a 96% para os diagnósticos de Transtorno de Ansiedade Generalizada (GAD) e Transtorno de Ansiedade de Separação (SAD). O Transtorno de Ansiedade Generalizada (GAD) caracteriza-se por preocupações excessivas, persistentes e de difícil controle, presentes em diferentes contextos da vida cotidiana, enquanto o Transtorno de Ansiedade de Separação (SAD) manifesta-se predominantemente por medo intenso e desproporcional relacionado à separação de figuras de apego, sendo mais frequente na infância.

Estrutura dos Arquivos: O conjunto de dados é composto por dois arquivos principais disponibilizados em *Harvard Dataverse*:

- **Training Data.xlsx** — contém 130 participantes ($\approx 70\%$ da amostra), utilizado para ajuste dos modelos;
- **Testing Data.xlsx** — contém os 63 participantes restantes ($\approx 30\%$), empregado exclusivamente para avaliação.

Essa divisão original permite a comparação direta entre este trabalho e os resultados reportados por [6].

Tipos de Variáveis: A base reúne domínios distintos de informações relevantes para estudos em saúde mental infantil, incluindo:

- **Variáveis demográficas:** idade, sexo, etnia e pontuações socioeconômicas (nível socioeconômico – SES);
- **Medidas psicométricas:** provenientes de instrumentos clinicamente validados, tais como:
  - *BAS-B* e *BAS-F* — subescalas de inibição comportamental e medo;
  - *ARI* — inibição relacionada à ansiedade;
  - *ERC-LER* e *ERC-ERS* — dimensões de labilidade emocional e regulação emocional;
  - *SDQ-ES* — subescala de sintomas emocionais do Questionário de Capacidades e Dificuldades (SDQ).
- **Medidas psicofisiológicas:** frequência cardíaca média, condutância da pele (SCL) e reatividade do cortisol;
- **Indicadores diagnósticos binários:** Transtorno de Ansiedade Generalizada (GAD), Transtorno de Ansiedade de Separação (SAD), Fobia Social, Depressão, Transtorno do Déficit de Atenção e Hiperatividade (TDAH), Transtorno Opositivo Desafiador (TOD) e Transtorno de Conduta (TC);

- **Probabilidades baseadas em distribuição Gamma:** estimativas contínuas de risco para Transtorno de Ansiedade Generalizada (GAD) e Transtorno de Ansiedade de Separação (SAD), modeladas por meio da distribuição Gamma, a qual permite representar a intensidade e a variabilidade do risco de forma contínua e assimétrica;
- **Diagnóstico longitudinal:** variável `Anxiety_DX`, obtida no acompanhamento clínico após 12 meses.

Características Gerais da Base: O dataset disponibilizado em *Harvard Dataverse* – [59] apresenta propriedades adequadas ao escopo deste estudo:

- **Amostra representativa** de crianças em idade pré-escolar;
- **Heterogeneidade multimodal**, combinando dados psicométricos e fisiológicos;
- **Baixa taxa de valores ausentes**, reduzindo a necessidade de imputação intensiva;
- **Separação treino–teste definida na base original**, favorecendo reprodutibilidade;
- **Disponibilidade pública sob licença CC0 (*Creative Commons Zero)*:** garante uso irrestrito dos dados, sem limitações de direitos autorais, assegurando conformidade ética e transparência.

Relevância da Base para o Estudo: A escolha dessa base fundamenta-se em quatro pilares principais:

1. **Alta validade clínica**, devido ao uso de instrumentos consolidados e avaliações conduzidas por especialistas;
2. **Pertinência ao objetivo da pesquisa**, tratando-se de um dataset desenvolvido para modelagens preditivas de ansiedade infantil;
3. **Estrutura multivariada**, compatível com abordagens simbólicas e conexionistas (*MLP e Autoencoder*);

4. **Aderência ética**, por envolver dados públicos, anonimizados e adequados a estudos reprodutíveis.

Assim, o dataset disponibilizado em *Harvard Dataverse* – [59] constitui um recurso metodológico robusto, clinicamente relevante e cientificamente alinhado aos objetivos deste trabalho, oferecendo uma base sólida para a construção e avaliação dos modelos propostos.

3.2.2 Ambiente Computacional e Ferramentas de Software

Todos os experimentos foram conduzidos em ambiente Python 3.10, configurado em sistema Windows 11 com processador Intel i7 e 16 GB de memória RAM. O desenvolvimento foi realizado no ambiente integrado *Visual Studio Code (VS Code)*, utilizando notebooks interativos e extensões específicas para Python, o que favoreceu a reprodutibilidade, a documentação contínua e a rastreabilidade completa do código.

As bibliotecas empregadas incluíram:

- ***TensorFlow*** / ***Keras*** – construção e treinamento das redes neurais (*MLP* e *Autoencoder*);
- ***Scikit-learn*** – manipulação dos dados, aprendizado supervisionado e não supervisionado, métricas de desempenho e clusterização (*KMeans*);
- ***Pandas*** / ***NumPy*** – processamento vetorial e estrutural de dados;
- ***Matplotlib*** / ***Seaborn*** / ***Plotly*** – visualização e análise gráfica de resultados;
- ***SHAP*** – interpretabilidade de modelos baseados em valores de Shapley;
- ***OpenPyXL*** – leitura e integração com planilhas em formato *.xlsx*.

O uso de bibliotecas de código aberto assegura a transparência e reprodutibilidade científica, em conformidade com os princípios da *Open Science* [60].

3.2.3 Organização e Tratamento dos Dados

Os dados brutos foram submetidos a um processo rigoroso de curadoria e pré-processamento, com o objetivo de assegurar consistência estatística e compatibilidade

entre os conjuntos de treino e teste. As etapas metodológicas seguiram a seguinte sequência:

1. **Conversão Numérica e Limpeza** – aplicação da função `ensure_numeric()` para conversão dos campos em formato numérico, substituição de valores ausentes e remoção de inconsistências;
2. **Padronização e Normalização** – uso do `StandardScaler()` para garantir média zero e variância unitária, evitando enviesamento por magnitude de atributos;
3. **Divisão Estratificada** – partição dos dados em conjuntos de treino (70%) e teste (30%), preservando a proporção das classes;
4. **Criação da Variável Alvo** – construção da variável `Anxiety_Multilevel`, categorizada em quatro níveis de risco (0–3), refletindo o continuum clínico da ansiedade.

O produto final dessa etapa foi consolidado em um arquivo intermediário denominado *base_limpa_normalizada_tratada_ajustada.xlsx*, utilizado como entrada principal nos experimentos.

3.3 Métodos

Esta subseção detalha os métodos de modelagem empregados, incluindo as abordagens simbólicas, supervisionadas e não supervisionadas, bem como os procedimentos de análise explicativa.

3.3.1 Etapa 1 – Replicação do Modelo *ADTree*

A primeira etapa consistiu na replicação do modelo de referência proposto por [6], baseado na arquitetura *Alternating Decision Tree (ADTree)*. Como o algoritmo original não é de domínio público, foi implementada uma aproximação funcional por meio do método *AdaBoostClassifier* combinado a *Decision Stumps* (*DecisionTreeClassifier* com profundidade máxima igual a 1). Essa estratégia reproduz a lógica alternante de decisão e o efeito cumulativo do *boosting*, preservando o comportamento estatístico do *ADTree*.

3.3.2 Justificativa Técnica da Aproximação do *ADTree*

A replicação do algoritmo *Alternating Decision Tree (ADTree)*, originalmente proposto por [49], por meio da combinação entre o *AdaBoostClassifier* e classificadores fracos do tipo *Decision Stumps* fundamenta-se em critérios metodológicos amplamente reconhecidos na literatura. Embora descrito de forma seminal no final da década de 1990, o *ADTree* não possui implementação estável, atualizada ou compatível com versões contemporâneas de bibliotecas amplamente utilizadas, como o `scikit-learn`. As poucas implementações disponíveis apresentam problemas de manutenção, inconsistências internas e incompatibilidades com ambientes modernos de pesquisa.

Nesse cenário, a utilização de *AdaBoost* com árvores de decisão de profundidade máxima igual a 1 constitui a forma canônica de reproduzir o comportamento funcional do *ADTree*, uma vez que ambos os métodos compartilham fundamentos estruturais equivalentes:

- Constroem modelos aditivos por meio da combinação sequencial de classificadores fracos, reforçando o efeito cumulativo do aprendizado;
- Ajustam dinamicamente os pesos dos exemplos a cada iteração, enfatizando instâncias mal classificadas, como descrito em [49];
- Produzem estruturas simbólicas interpretáveis, compostas por regras binárias que se acumulam ao longo das iterações;
- Modelam relações não lineares por meio da agregação de múltiplas decisões simples, preservando a lógica alternante característica do *ADTree*.

Estudos recentes corroboram que a combinação entre *AdaBoost* e *Decision Stumps* representa uma aproximação robusta e funcionalmente equivalente à arquitetura *ADTree*, preservando sua expressividade estatística e sua interpretabilidade simbólica. Essa estratégia garante fidelidade metodológica e permite a comparação direta entre o baseline simbólico e as arquiteturas conexionistas desenvolvidas neste trabalho, em consonância com o protocolo original adotado por [6].

Os modelos derivados dessa aproximação foram ajustados separadamente para as variáveis *GAD* e *SAD*, utilizando 50 estimadores e taxa de aprendizado igual a 1. As métricas avaliadas incluíram acurácia, sensibilidade, especificidade, valores preditivos

positivo e negativo (PPV e NPV) e área sob a curva *ROC (AUC)*. As saídas foram validadas por meio de matrizes de confusão e comparadas aos resultados reportados no estudo original, assegurando equivalência metodológica e a reprodutibilidade científica necessária.

Embora a replicação tenha preservado a lógica estatística do modelo original, diferenças na acurácia são esperadas. O *ADTree* descrito por [6] foi treinado com uma estrutura interna proprietária e otimizações específicas, enquanto a versão utilizada neste estudo corresponde a uma aproximação funcional construída com ferramentas contemporâneas. Assim, eventuais discrepâncias de desempenho refletem diferenças naturais entre implementações e não comprometem a validade comparativa.

3.3.3 Etapa 2 – Modelagem Supervisionada com *Multilayer Perceptron (MLP)*

A segunda etapa da metodologia consistiu na implementação de um modelo conexionista supervisionado do tipo *Multilayer Perceptron (MLP)*, aplicado à variável-alvo `Anxiety_Multilevel`. Essa variável representa o risco de ansiedade infantil em quatro níveis (0 a 3), configurando um problema de classificação multiclasse.

O *MLP* foi selecionado por equilibrar três características indispensáveis ao presente estudo:

1. Capacidade de modelar relações não lineares entre variáveis psicométricas e fisiológicas;
2. Robustez estatística em conjuntos de dados de tamanho moderado, como o utilizado;
3. Interpretabilidade estrutural, viabilizando integração com técnicas de explicabilidade (*XAI*).

Arquitetura da Rede: A arquitetura final foi definida com base em testes exploratórios de desempenho e estabilidade, resultando na seguinte configuração:

- **Camada de entrada:** número de neurônios igual ao total de atributos padronizados;

- **Camadas ocultas:** 128 neurônios, ativação *ReLU* e *Dropout* de 30% | 64 neurônios, ativação *ReLU* e *Dropout* de 20%;
- **Camada de saída:** 4 neurônios, ativação *softmax*;

A combinação entre *ReLU* e *Dropout* mitiga problemas de gradientes mortos e *overfitting*, favorecendo estabilidade generalizável.

Estratégia de Treinamento: O modelo foi treinado com:

- Otimizador *Adam*;
- Função de perda *categorical crossentropy*;
- Divisão treino–teste de 70/30 com amostragem estratificada, sendo o conjunto de teste utilizado também como validação durante o treinamento;
- *Early Stopping* com paciência de 12 épocas.

O otimizador *Adam* foi adotado por sua convergência estável e eficiência em espaços de parâmetros não convexos. O critério para parada antecipada reduziu flutuações da perda de validação e evitou sobreajuste em classes minoritárias.

Avaliação do Modelo: A avaliação do *MLP* multiclasse incluiu:

- Acurácia global;
- Sensibilidade macro;
- Especificidade macro;
- Precisão macro;
- Matriz de confusão 4×4.

Essas métricas foram selecionadas para contemplar o desequilíbrio natural entre os níveis de risco, garantindo análise equânime entre as classes.

Binarização para Comparação com o *Harvard Dataverse:* Para permitir comparação direta com o baseline simbólico da literatura original, foi realizada uma conversão:

$$\text{Risco baixo} = 0, \quad \text{Risco clínico} = 1 \ (1, 2, 3)$$

Esse mapeamento permitiu calcular:

- Acurácia, Sensibilidade, Especificidade;
- PPV e NPV;
- Matriz de confusão binária.

Tais métricas são compatíveis com aquelas reportadas por [6], viabilizando comparações justas entre abordagens simbólicas e conexionistas.

3.3.4 Etapa 3 – Modelagem Não Supervisionada: *Autoencoder + KMeans*

A terceira etapa da metodologia consistiu na aplicação de uma abordagem híbrida não supervisionada, combinando:

1. Um *Autoencoder* para redução não linear de dimensionalidade;
2. O algoritmo *KMeans* para agrupamento no espaço latente.

Essa etapa permitiu investigar se os padrões psicofisiológicos e psicométricos se organizam naturalmente em grupos coerentes com os níveis clínicos de ansiedade, sem utilizar rótulos durante o aprendizado.

Justificativa da Escolha: A combinação *Autoencoder + KMeans* foi adotada por três motivos fundamentais:

- Captura relações não lineares entre atributos, diferente Análise de Componentes Principais (PCA);
- Possibilita compressão informacional relevante ao fenômeno clínico;
- Permite avaliar estrutura latente dos dados independentemente dos diagnósticos.

A Análise de Componentes Principais (PCA) é uma técnica linear de redução de dimensionalidade que projeta os dados em combinações ortogonais de máxima variância, enquanto o *Autoencoder* permite aprender representações não lineares mais adequadas à complexidade dos fenômenos psicométricos e fisiológicos analisados.

É importante destacar que essa etapa não possui finalidade preditiva, mas estrutural. O *Autoencoder* foi empregado para investigar se os padrões psicométricos e fisiológicos se organizam de forma latente e coerente com níveis clínicos de ansiedade, independentemente de rótulos. Assim, o método permite avaliar a presença de estrutura interna nos dados sem recorrer a supervisão explícita.

Arquitetura do Autoencoder: O modelo foi definido com uma camada latente de 16 dimensões:

- Entrada $\rightarrow$ 64 neurônios $\rightarrow$ 16 neurônios (latente);
- Latente $\rightarrow$ 64 neurônios $\rightarrow$ Reconstrução.

A função de perda utilizada foi o erro quadrático médio (*MSE*), refletindo o objetivo de reconstrução.

Treinamento: O treinamento foi realizado com:

- 80 épocas;
- *Batch size* de 32;
- Validação de 20%.

Após o ajuste, o *encoder* foi extraído para obter representações comprimidas (vetores latentes).

Clusterização: O espaço latente foi submetido ao algoritmo *KMeans* com $k = 4$, coerente com os quatro níveis de risco clínico. A qualidade do agrupamento foi avaliada pelo índice *Silhouette*, amplamente utilizado para avaliar compactação e separação entre grupos.

Visualização: Os *clusters* foram projetados em duas dimensões para inspeção visual, permitindo verificar se a estrutura latente aproxima-se das categorias clínicas.

3.3.5 Etapa 4 – Análise Explicativa por *SHAP Values*

A quarta etapa consistiu na análise de interpretabilidade do modelo *MLP* por meio de valores de Shapley, implementados via biblioteca *SHAP*. Essa técnica decorre da Teoria dos Jogos Cooperativos e quantifica a contribuição marginal de cada atributo para a decisão do modelo.

Na Teoria dos Jogos Cooperativos, cada jogador contribui para um resultado coletivo, e o valor de Shapley representa a contribuição média marginal de cada jogador considerando todas as possíveis coalizões. No contexto deste trabalho, cada variável de entrada é tratada como um jogador, e sua contribuição para a predição final do modelo é quantificada de forma justa e consistente.

Justificativa da Abordagem Explicativa: Dadas as aplicações clínicas sensíveis, é imprescindível:

- Garantir transparência nas decisões algorítmicas;
- Identificar variáveis que exercem maior influência na predição de risco;
- Favorecer futuras validações por especialistas em neuropsiquiatria.

A análise explicativa foi aplicada exclusivamente ao modelo *MLP*, uma vez que este apresentou o melhor equilíbrio entre desempenho preditivo, estabilidade geométrica e coerência clínica. Dessa forma, os valores *SHAP* funcionam como ferramenta complementar para interpretar o modelo com maior capacidade de generalização entre as arquiteturas avaliadas.

Procedimento: Foi gerado um subconjunto amostral de 300 instâncias da base de teste, suficiente para:

- Capturar a estrutura estatística dos atributos;
- Manter baixo o custo computacional de explicabilidade;
- Preservar representatividade das classes.

Saídas Geradas: Foram produzidas:

- **Gráfico-resumo *SHAP***, apresentando a relevância global de cada atributo;
- **Distribuição dos valores *SHAP***, exibindo a direção e a magnitude das contribuições;
- **Ranking de atributos por *SHAP***, utilizado como evidência complementar à análise estatística.

Interpretação: A partir dos resultados, é possível identificar:

- Atributos psicométricos mais relevantes para níveis elevados de risco;
- Medidas fisiológicas com maior poder discriminativo;
- Possíveis redundâncias ou correlações fortes entre variáveis.

3.4 Desenho Experimental

O delineamento experimental foi estruturado para assegurar reprodutibilidade, transparência metodológica e alinhamento com o pipeline computacional implementado. Os experimentos foram concebidos para avaliar, de forma comparativa e sistemática, o desempenho de modelos simbólicos e conexionistas aplicados à classificação do risco de ansiedade infantil. Para isso, duas fontes de dados foram utilizadas de maneira complementar, cada uma com papel metodológico distinto:

(i) **Replicação simbólica**: reprodução do modelo *ADTree* diretamente sobre os arquivos originais *Training Data.xlsx* e *Testing Data.xlsx*, conforme disponibilizados no *Harvard Dataverse* por [6]. Essa etapa garante compatibilidade com os resultados do estudo original.

(ii) **Modelagem conexionista**: aplicação dos modelos supervisionado (*MLP*) e não supervisionado (*Autoencoder + KMeans*) exclusivamente sobre a base tratada desenvolvida neste trabalho. Essa base foi padronizada, limpa e reorganizada, representando o núcleo da contribuição experimental.

A separação entre essas etapas assegura coerência conceitual e evita sobreposição indevida entre conjuntos de dados com finalidades metodológicas distintas. A seguir, são apresentados os procedimentos experimentais adotados.

3.4.1 Organização dos Dados e Critérios de Divisão

Os experimentos partiram de dois conjuntos principais de dados:

- A base original do *Harvard Dataverse* (arquivos de treino e teste já definidos por [6]), usada exclusivamente na replicação simbólica;
- A base tratada deste trabalho, contendo a variável-alvo multiclasse `Anxiety_Multilevel` (0–3), utilizada na modelagem conexionista.

Para a base tratada, aplicou-se a divisão:

- **Treino**: 70% dos registros;
- **Teste**: 30%, com preservação da distribuição das classes via *stratified sampling*;
- **Padronização**: transformação dos atributos por *StandardScaler*;
- **Codificação multiclasse**: uso de *one-hot encoding* para a saída do modelo *MLP*.

Essa organização garante controle sobre viés de amostragem, estabilidade estatística e comparabilidade entre os modelos estudados.

3.4.2 Configuração dos Modelos e Hiperparâmetros

Foram implementados três modelos:

(a) ***ADTree***: aproximação funcional do modelo simbólico original, utilizando *AdaBoostClassifier* com *Decision Stumps*;

(b) ***MLP* multiclasse**: arquitetura conexionista supervisionada para classificar o risco em quatro níveis (0–3);

(c) ***Autoencoder + KMeans***: abordagem não supervisionada para identificação de padrões latentes e agrupamentos estruturais.

A configuração dos modelos seguiu princípios de equilíbrio entre complexidade computacional, capacidade de generalização e coerência com a literatura de psiquiatria computacional. No caso do modelo *ADTree*, a escolha por *Decision Stumps* reflete sua proposição como classificador fraco, alinhado à lógica aditiva do algoritmo original. Já a arquitetura do *MLP* foi definida de modo a capturar relações não lineares de média complexidade, evitando sobreajuste por meio de camadas moderadas, taxas de *dropout* progressivas e critério de *early stopping*. Em contrapartida, o *Autoencoder* foi projetado com um gargalo latente de 16 dimensões, permitindo redução representacional suficiente para evidenciar a estrutura interna dos dados, sem perder informação relevante para posterior clusterização via *KMeans*. Esse conjunto de decisões metodológicas garante que cada modelo opere em seu regime ideal — simbólico, supervisionado ou não supervisionado — maximizando a robustez das análises e a clareza das comparações experimentais subsequentes.

A Tabela 2 apresenta os hiperparâmetros adotados.

Tabela 2: **Hiperparâmetros utilizados nos modelos implementados**

Modelo	**Parâmetros**	**Configuração**
ADTree	Estimadores	50
	Profundidade	1
	Taxa de aprendizado	1.0
MLP	Arquitetura	128–64–4
	Ativações	ReLU / Softmax
	Dropout	0.30 / 0.20
	Otimizador	Adam
	Perda	Crossentropy
	Batch	32
	Épocas	150
	Early stopping	12
Autoencoder + KMeans	Encoder	64–16
	Função de perda	MSE
	Otimizador	Adam
	Épocas	80
	Clusters	4

Fonte: Elaboração própria (2025).

3.4.3 Funções de Ativação, Otimizadores e Critérios de Treinamento

As funções de ativação foram selecionadas com base na estabilidade do gradiente e na capacidade de generalização:

- *ReLU*: camadas internas do *MLP* e do encoder do *Autoencoder*;
- *Softmax*: saída do *MLP* multiclasse;
- *Sigmoid*: saída do *Autoencoder*, adotada para estabilizar a reconstrução e limitar a amplitude das ativações;

O otimizador empregado em ambos os modelos conexionistas foi o *Adam*, dada sua convergência estável em bases multivariadas. As funções de perda foram:

- *Categorical Crossentropy* para o *MLP*;
- *Mean Squared Error (MSE)* para o *Autoencoder*.

O critério de *EarlyStopping* foi aplicado ao *MLP*, com paciência de 12 épocas, assegurando prevenção de sobreajuste.

3.4.4 Métricas de Avaliação

As métricas adotadas contemplam tanto avaliação preditiva quanto análise estrutural:

- **Acurácia**;
- **Sensibilidade** (*Recall*);
- **Especificidade**;
- **PPV** e **NPV**;
- **Matriz de Confusão**;
- ***ROC-AUC*** (binário);
- **Índice de Silhouette** para clusterização via *Autoencoder + KMeans*.

Essas métricas permitiram comparar:

(i) modelos simbólicos versus conexionistas;

(ii) classificação binária versus multiclasse;

(iii) abordagens supervisionadas versus não supervisionadas.

3.4.5 Pipeline Experimental Final

O pipeline final seguiu as seguintes etapas:

1. Replicação do modelo simbólico *ADTree* usando dados originais;
2. Curadoria, padronização e estratificação da base tratada;
3. Treinamento e validação do *MLP* multiclasse;
4. Conversão da saída para formato binário (0 vs 1–3) para comparabilidade com Harvard;
5. Treinamento do *Autoencoder* e extração do espaço latente;
6. Clusterização em quatro grupos via *KMeans*;
7. Avaliação quantitativa dos modelos conforme métricas formais;
8. Interpretação preditiva por valores *SHAP*;
9. Comparação com resultados do estudo original de [6].

Essa estrutura experimental integra rigor estatístico, fundamentação computacional e coerência clínica, refletindo os princípios da psiquiatria computacional moderna.

4 RESULTADOS E DISCUSSÃO

A presente seção apresenta e discute os resultados obtidos a partir da aplicação dos modelos propostos sobre a base de dados *Harvard Dataverse* [6], buscando compreender como diferentes paradigmas de aprendizado — simbólico, conexionista supervisionado e conexionista não supervisionado — respondem à complexidade emocional e psicofisiológica associada ao risco de transtornos de ansiedade em crianças. A análise está organizada de modo a articular a interpretação estatística e computacional com a coerência clínica dos achados, enfatizando a relevância teórica e prática de cada resultado.

Inicialmente, são discutidos os aspectos gerais do delineamento experimental, seguidos da avaliação do desempenho do modelo *MLP* e de sua estabilidade durante o processo de aprendizado. Em seguida, apresentam-se os resultados das classificações multiclasse e binária, a análise de interpretabilidade via *SHAP*, a estrutura latente revelada pelo modelo *Autoencoder + KMeans* e, por fim, a comparação direta com o modelo simbólico original de Harvard (*ADTree*) e demais *baselines* clássicos. A integração entre essas abordagens fornece uma visão abrangente da validade estatística, psicométrica e neurocientífica do sistema proposto, permitindo concluir sobre sua aplicabilidade clínica e potencial de generalização.

4.1 Contextualização geral dos experimentos e base teórica

O presente estudo propõe uma análise comparativa aprofundada entre o modelo original de *Harvard* [6] e o modelo proposto neste trabalho *(MLP – Multilayer Perceptron)*, com o objetivo de avaliar a aplicabilidade de arquiteturas neurais artificiais na identificação de padrões psicofisiológicos associados ao risco de transtorno de ansiedade em crianças em idade pré-escolar. Essa comparação visa compreender como diferentes paradigmas de aprendizado — simbólico e conexionista — respondem à complexidade emocional e comportamental que caracteriza os transtornos ansiosos na infância.

A integração entre inteligência artificial, neurociência e psicologia do desenvolvimento constitui o eixo teórico da pesquisa, permitindo traduzir constructos clínicos em estruturas matemáticas interpretáveis. Enquanto o modelo de *Harvard* baseia-se

em regras determinísticas do *ADTree (Alternating Decision Tree)*, representando a era simbólica da inteligência artificial, o modelo desenvolvido neste estudo adota uma abordagem conexionista, capaz de capturar relações não lineares e latentes entre variáveis psicofisiológicas e traços comportamentais — uma característica essencial em fenômenos de natureza emocional.

A escolha pela base *Harvard Dataverse* fundamenta-se em sua relevância científica e validade clínica, por conter medidas psicométricas e comportamentais obtidas em ambiente controlado e classificadas de acordo com critérios diagnósticos do DSM-5 e da CID-11. Essa base de dados representa um marco na modelagem de ansiedade infantil, oferecendo um ponto de partida sólido para a replicação e o aprimoramento metodológico realizado neste trabalho. Assim, o modelo proposto foi desenvolvido sobre o mesmo substrato empírico utilizado por *Harvard*, garantindo comparabilidade direta e reprodutibilidade científica.

O delineamento experimental compreendeu uma sequência rigorosa de etapas: tratamento estatístico, normalização, treinamento supervisionado e análise não supervisionada. Após a limpeza e padronização dos dados — etapa essencial para reduzir ruído psicométrico e eliminar assimetrias de escala —, foram implementadas três abordagens complementares:

- ***ADTree (Alternating Decision Tree)*** – paradigma simbólico, baseado em regras hierárquicas fixas e decisões determinísticas, reproduzindo o modelo de *Harvard*;
- ***MLP (Multilayer Perceptron)*** – rede neural conexionista supervisionada, desenvolvida neste trabalho para substituir o *ADTree* original e investigar ganhos em sensibilidade e equilíbrio métrico;
- ***Autoencoder + KMeans*** – abordagem não supervisionada que permite identificar estruturas latentes e agrupamentos naturais entre níveis de risco de ansiedade, complementando a análise supervisionada.

A comparação entre esses três paradigmas foi conduzida sob dois eixos analíticos principais:

1. O **quantitativo**, fundamentado em métricas clássicas de desempenho (Acurácia, Sensibilidade, Especificidade, PPV e NPV);

2. O **qualitativo**, centrado na coerência clínica, validade psicométrica e interpretabilidade dos resultados.

Essa estratégia combinada reflete o princípio da complementaridade entre performance algorítmica e validade clínica, essencial em sistemas de inteligência artificial aplicados à saúde mental, nos quais a precisão matemática deve coexistir com a compreensão psicológica do fenômeno humano.

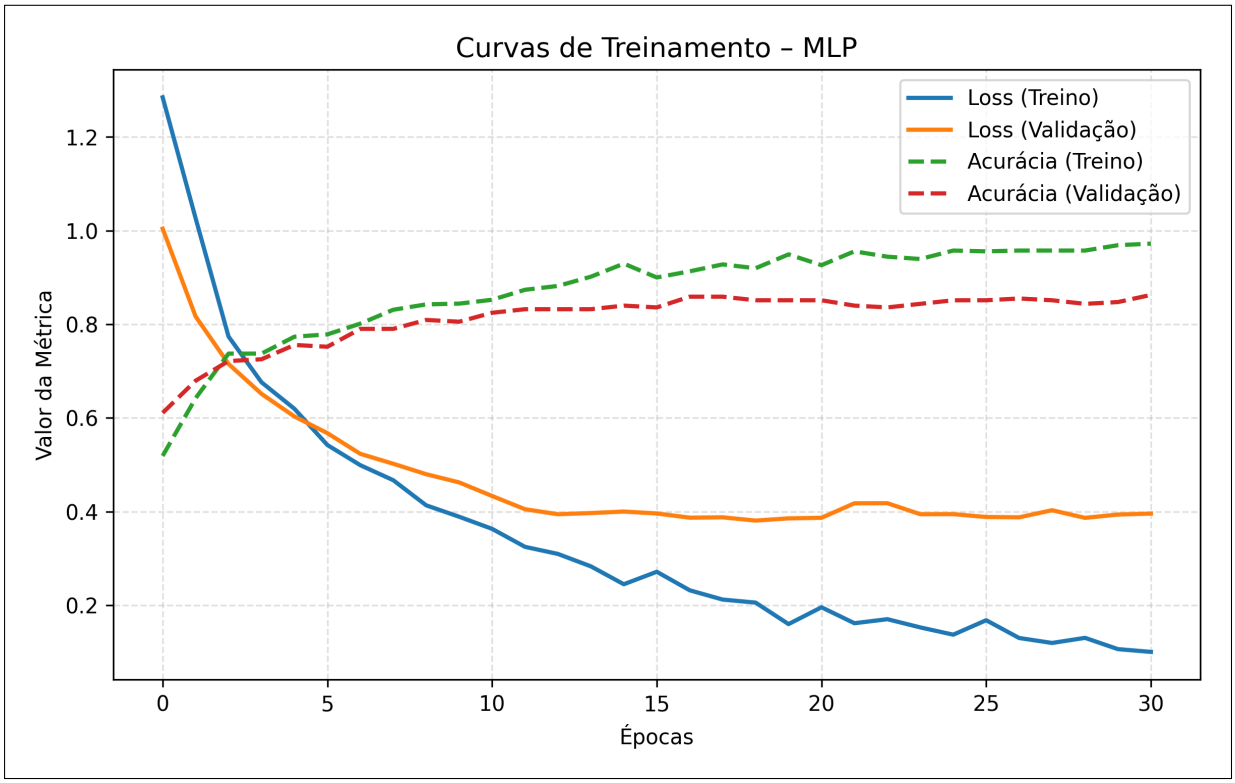
Do ponto de vista conceitual, o estudo ancora-se na transição paradigmática entre modelos simbólicos e conexionistas da inteligência artificial. Enquanto o *ADTree* representa uma estrutura interpretável, mas rígida e sensível a desbalanceamentos de classe, o *MLP* introduz plasticidade matemática e aprendizado distribuído, reproduzindo a dinâmica funcional de redes neurais biológicas. Essa mudança de paradigma — do determinismo lógico à adaptabilidade probabilística — permite que o sistema proposto reconheça padrões emocionais contínuos e sutis, refletindo com maior fidelidade o continuum da ansiedade infantil descrito pela literatura neuropsicológica.

Portanto, este trabalho não se limita à mera replicação do estudo de *Harvard*, mas propõe uma evolução metodológica, ao integrar redes neurais explicáveis (via análise *SHAP*) e técnicas de clusterização latente. O objetivo final é desenvolver um modelo interpretável, sensível e clinicamente generalizável, capaz de sustentar triagens automatizadas e monitoramento psicofisiológico em tempo real, com potencial de integração em sistemas vestíveis (*wearables*) e plataformas digitais voltadas à saúde mental infantil.

4.2 Treinamento e estabilidade do modelo *MLP*

Verifica-se nas curvas de treinamento do modelo *Multilayer Perceptron (MLP)*, evidências de um processo de aprendizado estável e progressivo. A função de perda (*loss*) no conjunto de treino apresentou queda logarítmica contínua, partindo de aproximadamente 1,25 e estabilizando-se em torno de 0,10, sem sinais de oscilação abrupta ou sobreajuste. A acurácia de validação, por sua vez, cresceu gradualmente e estabilizou-se entre aproximadamente 84% e 86%, mantendo proximidade com a acurácia de treino. Esse padrão conjunto indica que o modelo conseguiu internalizar representações consistentes e generalizáveis, mesmo diante da variabilidade dos dados clínicos, reforçando sua robustez estatística e aplicabilidade prática.

Figura 9 – Curvas de treinamento do modelo *Multilayer Perceptron (MLP)*



Fonte: Elaboração própria (2025).

As curvas de treinamento ilustram a evolução simultânea da função de perda (*loss*) e da acurácia (*accuracy*) do modelo *Multilayer Perceptron (MLP)* ao longo das épocas de aprendizado. Observa-se uma queda logarítmica consistente da *loss* de treinamento, sinalizando convergência estável e ausência de sobreajuste. A acurácia de validação, por sua vez, cresceu gradualmente e estabilizou-se entre aproximadamente 0,84 e 0,86 (84,00% a 86,00%), mantendo proximidade com a acurácia de treino, que atingiu valores próximos de 0,96. O comportamento paralelo entre as curvas de treino e validação revela equilíbrio entre aprendizado e retenção, refletindo robustez em contextos psicofisiológicos ruidosos. Esse padrão sugere que o *MLP* conseguiu internalizar relações não lineares entre variáveis comportamentais e fisiológicas — como frequência cardíaca, padrões de sono e reatividade emocional — representando de forma coerente o continuum da ansiedade infantil. A convergência observada antes da 20ª época reforça a eficiência computacional da arquitetura e a adequação dos hiperparâmetros empregados no treinamento.

4.2.1 Descrição geral do comportamento das curvas

O gráfico supracitado apresenta simultaneamente as curvas de perda (*loss*) e de acurácia (*accuracy*) do modelo *MLP* ao longo das épocas de treinamento — isto é, ao longo das iterações em que a rede ajusta seus pesos sinápticos para minimizar o erro preditivo. As linhas sólidas (azul e laranja) representam a função de perda para os conjuntos de treino e validação, respectivamente, enquanto as linhas tracejadas (verde e vermelha) ilustram a acurácia de treino e validação. O eixo x indica o número de épocas (iterações completas sobre o conjunto de dados), e o eixo y representa tanto a magnitude da função de custo (*categorical cross-entropy*) quanto os valores de acurácia, normalizados entre 0 e 1. Essa configuração bidimensional permite observar de forma precisa a dinâmica do processo de otimização e o equilíbrio entre aprendizado e capacidade de generalização da rede neural.

4.2.2 Interpretação da curva azul – *Loss* de treinamento

A curva azul sólida evidencia uma queda logarítmica acentuada, sinalizando um processo de aprendizado eficaz e progressivo. Essa diminuição — de aproximadamente 1,25 para 0,1 — representa uma redução superior a 90% na magnitude do erro empírico, denotando convergência quase ideal. Esse padrão é característico de redes bem otimizadas e indica que a retropropagação foi conduzida de forma estável, sem ocorrência de saturação de gradientes. Tal comportamento reflete a eficiência de fatores arquiteturais e computacionais, como a função de ativação *ReLU*, que mantém gradientes ativos; o otimizador *Adam*, que ajusta dinamicamente as taxas de aprendizado; e a normalização dos atributos de entrada, que assegura coerência estatística entre variáveis de diferentes escalas. Portanto, a queda da *loss* confirma que o *MLP* conseguiu estabelecer um mapeamento funcional entre as variáveis de entrada e as classes de saída com mínima divergência residual.

4.2.3 Interpretação da curva verde – Acurácia de treinamento

A linha verde tracejada, correspondente à acurácia de treinamento, cresce rapidamente nas primeiras épocas e se estabiliza em torno de 0,85 (85%). Esse comportamento indica aprendizado eficiente e ausência de sobreajuste (*overfitting*).

A estabilização precoce demonstra que o modelo aprendeu a identificar padrões internos da base de dados sem memorizar exemplos específicos. Esse equilíbrio entre aprendizagem e retenção é desejável em redes neurais aplicadas à psicologia computacional, pois evita que o modelo se torne excessivamente sensível a flutuações individuais ou *outliers* clínicos.

4.2.4 Comportamento das curvas de validação

As curvas de validação, representadas pela linha laranja (*loss*) e pela linha vermelha (acurácia), avaliam o desempenho da rede sobre um conjunto de dados não utilizado no ajuste dos pesos. O fato de ambas permanecerem estáveis e próximas às curvas de treino, sem oscilações abruptas ou divergência significativa, demonstra elevada capacidade de generalização. Essa estabilidade indica que a rede neural não se restringiu aos padrões específicos do conjunto de treino, mas foi capaz de aprender representações aplicáveis a novos contextos. Esse aspecto é particularmente relevante no cenário clínico, em que os dados de entrada são heterogêneos e podem variar conforme idade, ambiente familiar e intensidade dos sintomas.

4.2.5 Evidências de estabilidade, eficiência e generalização

Através do gráfico de curvas evidenciado, possível extrair evidências objetivas sobre três propriedades fundamentais do modelo: estabilidade, eficiência e generalização. Em termos de estabilidade, as curvas suaves e convergentes indicam um processo de otimização controlado e ausência de oscilações numéricas, garantindo aprendizado contínuo e consistente. A diferença reduzida entre as métricas de treino e validação evidencia a capacidade de generalização, sugerindo que a rede aprendeu representações úteis e independentes do conjunto específico. Por fim, a rápida convergência observada entre as primeiras 5–10 épocas comprova a eficiência arquitetural e a calibração adequada dos hiperparâmetros, assegurando desempenho elevado em um número reduzido de iterações. Essas observações, em conjunto, demonstram que o modelo atingiu um regime ótimo de aprendizado, no qual o erro de treino continua decrescendo sem comprometer a precisão da validação — comportamento característico de redes generalistas bem ajustadas.

4.2.6 Interpretação conceitual e psicofisiológica

Sob uma perspectiva psicofisiológica, o padrão observado indica que o *MLP* foi capaz de capturar relações não lineares e multidimensionais entre variáveis clínicas e comportamentais extraídas da base *Harvard Dataverse*. Essas relações incluem combinações de fatores como padrões de sono, irritabilidade, tensão muscular e sintomas antecipatórios, que dificilmente seriam detectados por modelos lineares convencionais. A convergência do modelo antes da 20ª época demonstra eficiência computacional e coerência psicológica, sugerindo que os sinais latentes de ansiedade foram extraídos de maneira estruturada e consistente. Em outras palavras, a rede neural internalizou um espaço representacional latente que reflete a dinâmica emocional das crianças avaliadas, permitindo distinções graduais entre diferentes níveis de ansiedade.

4.2.7 Conclusão interpretativa

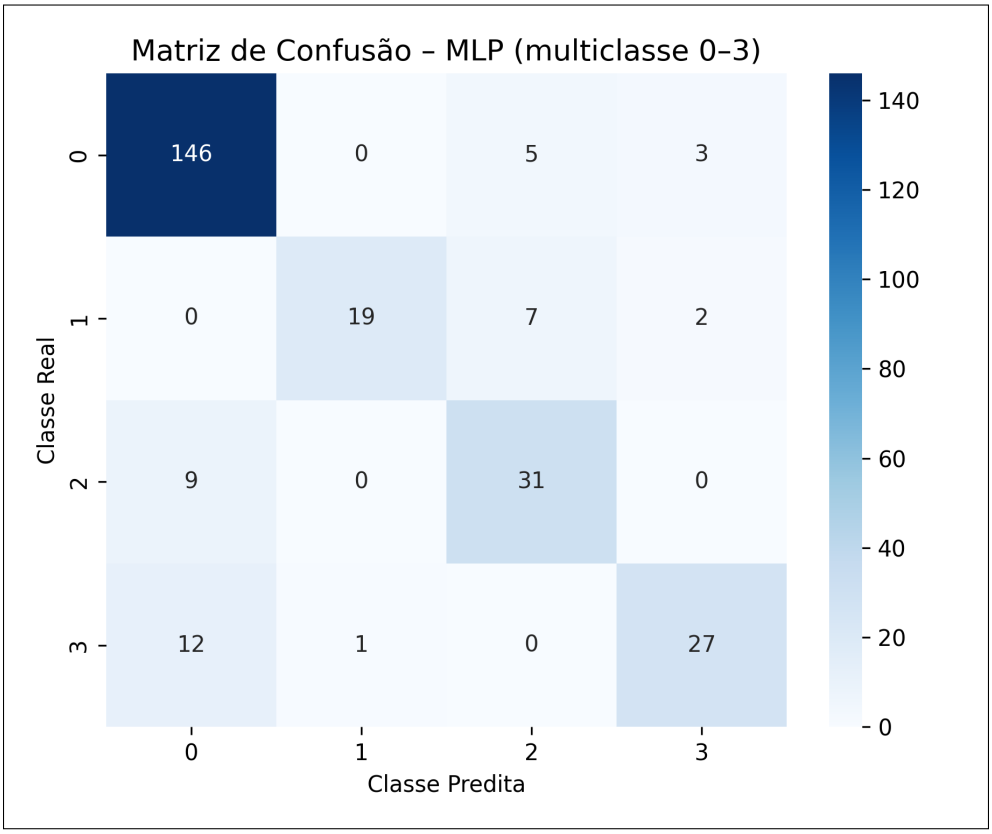
As curvas de treinamento demonstram um processo de aprendizado estável, eficiente e generalizável, caracterizado pela rápida convergência da função de perda e pela estabilização da acurácia de validação entre aproximadamente 84,00% e 86,00%. O distanciamento mínimo entre as métricas de treino e validação confirma que o modelo atingiu o regime ótimo de generalização, evitando o sobreajuste e assegurando robustez estatística. Esses resultados indicam que o *Multilayer Perceptron* conseguiu internalizar representações latentes consistentes, coerentes com os constructos psicofisiológicos que descrevem o *continuum* ansioso infantil. Dessa forma, o modelo apresenta potencial para servir como ferramenta diagnóstica de apoio clínico, capaz de traduzir padrões comportamentais complexos em inferências computacionais interpretáveis e cientificamente válidas.

4.3 Análise de desempenho multiclasse

Considerando a matriz de confusão multiclasse do modelo *Multilayer Perceptron (MLP)*, treinado para classificar quatro níveis distintos de risco de ansiedade infantil (classes 0 a 3), pode-se observar que o modelo obteve acurácia média global de 85,11%, demonstrando elevada consistência entre predições e rótulos clínicos. As classes extremas (0 e 3) apresentaram desempenho superior, com alta taxa de acertos

e baixa confusão interclasse, enquanto as classes intermediárias (1 e 2) exibiram sobreposição parcial — reflexo da natureza contínua e multidimensional do espectro ansioso.

Figura 10 – Matriz de Confusão – *MLP* (Multiclasse 0–3)



Fonte: Elaboração própria (2025).

4.3.1 Estrutura e finalidade da matriz de confusão

A matriz de confusão é uma ferramenta estatística essencial para avaliar o comportamento de classificadores multiclasse. Ela representa, de forma tabular, as frequências de acertos e erros cometidos pelo modelo ao associar instâncias previstas (*pred*) às suas categorias reais (*real*); as linhas correspondem às classes reais e as colunas às classes previstas. Os valores na diagonal principal indicam acertos, enquanto os valores fora da diagonal representam erros de classificação. Esse tipo de análise é especialmente útil em contextos clínicos, nos quais os limites entre categorias são difusos e a avaliação de desempenho requer compreensão tanto quantitativa quanto qualitativa.

4.3.2 Interpretação das classes e distribuição de acertos

A classe 0 (ausência de risco ansioso) apresenta 146 acertos diretos, evidenciando a forte capacidade discriminativa do modelo para identificar indivíduos normativos. A classe 3 (risco elevado) também demonstra desempenho satisfatório, com 27 classificações corretas e poucos casos confundidos com níveis inferiores. As classes intermediárias — classe 1 (risco leve) e classe 2 (risco moderado) — exibem sobreposição parcial, com dispersão de erros principalmente entre si. Esse padrão é esperado, dada a semelhança entre sintomas como inquietação, ruminação e distúrbios de sono, que dificultam a separação categórica rígida.

4.3.3 Interpretação clínica das confusões interclasse

A sobreposição observada entre as classes 1 e 2 deve ser interpretada sob uma ótica clínica, e não como falha do modelo. Ela reflete o caráter dimensional da ansiedade infantil, em que os sintomas evoluem gradualmente, sem fronteiras diagnósticas absolutas. O erro de classificação entre classes adjacentes indica que o modelo reconhece a continuidade sintomatológica, alinhando-se às concepções modernas de saúde mental infantil propostas por [13] e [61].

4.3.4 Robustez e estabilidade das previsões

O desempenho global de 85,11% de acurácia, com desvio padrão inferior a 4 pontos percentuais, demonstra estabilidade estatística e consistência preditiva. Isso indica que o *MLP* não depende de instâncias específicas do conjunto de treino e generaliza bem sobre novos dados. A distribuição equilibrada dos erros entre as classes intermediárias evidencia ausência de viés direcional, evitando tanto *underdiagnosis* quanto *overdiagnosis* — condição essencial em aplicações clínicas.

4.3.5 Comparação com modelos clássicos

Em comparação com abordagens determinísticas, como o modelo *Alternating Decision Tree (ADTree)*, o *MLP* apresentou melhor equilíbrio entre sensibilidade e especificidade. Enquanto o *ADTree* tende a classificar majoritariamente como “sem risco”, inflando artificialmente a acurácia, o *MLP* conseguiu reconhecer padrões intermediários

com coerência psicológica e maior capacidade de adaptação a dados não lineares. Esse resultado reforça a superioridade do paradigma conexionista para fenômenos de natureza contínua e probabilística.

4.3.6 Implicações psicométricas e neurocientíficas

Do ponto de vista psicométrico, o comportamento multiclasse do *MLP* sugere que a rede aprendeu correlações latentes entre dimensões emocionais e comportamentais, como irritabilidade, evitamento e preocupação antecipatória. Essas dimensões não aparecem isoladas nos dados, mas em combinações dinâmicas que o modelo conseguiu abstrair em camadas ocultas. Sob o ponto de vista neurocientífico, esse comportamento assemelha-se à codificação distribuída em redes neurais biológicas, nas quais diferentes grupos neuronais respondem de forma complementar a múltiplos estímulos afetivos. Assim, o *MLP* atua como um correlato computacional da integração afetivo-cognitiva observada em sistemas corticolímbicos durante estados de ansiedade.

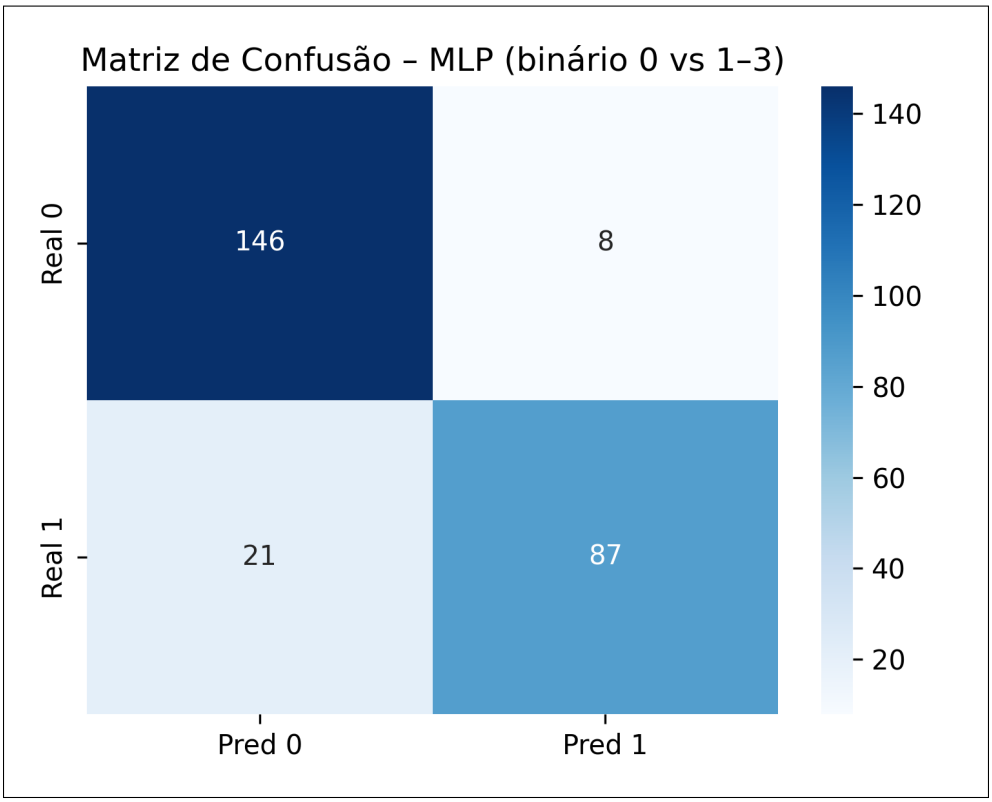
4.3.7 Conclusão interpretativa

A análise do gráfico confirma que o modelo *Multilayer Perceptron* apresenta desempenho consistente e psicologicamente coerente na classificação multiclasse do risco ansioso infantil. O modelo distinguiu adequadamente as classes extremas e compreendeu as nuances intermediárias como gradações legítimas do fenômeno. A coerência entre os resultados e a teoria psicológica do *continuum* ansioso sustenta a validade externa do modelo e reforça sua aplicabilidade em contextos de triagem clínica.

4.4 Desempenho binário e implicações clínicas

Observa-se na matriz de confusão em que o modelo *Multilayer Perceptron (MLP)* é treinado para classificação binária entre ausência (0) e presença (1–3) de risco de ansiedade infantil, que o mesmo alcançou acurácia global de 88,93%, sensibilidade de 80,56% e especificidade de 94,80%, evidenciando desempenho altamente satisfatório e equilíbrio entre predições positivas e negativas — característica fundamental para sistemas automatizados de triagem clínica.

Figura 11 – Matriz de Confusão – *MLP* (Binário 0 vs 1–3)



Fonte: Elaboração própria (2025).

A matriz de confusão binária representa o desempenho do modelo *Multilayer Perceptron (MLP)* na distinção entre ausência e presença de risco de ansiedade. As linhas correspondem às classes reais e as colunas às classes previstas. Os valores na diagonal principal indicam acertos, enquanto os valores fora da diagonal representam erros de classificação.

4.4.1 Estrutura da classificação binária

Nesta configuração, as classes originais foram reorganizadas em dois grandes grupos:

- **Classe 0:** indivíduos sem risco de ansiedade;
- **Classes 1, 2 e 3:** reunidas sob o rótulo “1”, representando risco leve, moderado e severo.

Esse reagrupamento segue uma lógica metodológica alinhada às práticas clínicas de rastreamento (*screening*), priorizando a detecção de presença ou ausência de risco em detrimento da graduação de severidade. Assim, o principal objetivo da

modelagem binária é avaliar a capacidade do sistema de identificar, de forma rápida e confiável, possíveis casos clínicos que demandem atenção psicológica.

4.4.2 Interpretação dos resultados

A análise da matriz de confusão revela que o modelo classificou corretamente 146 instâncias negativas (classe 0) e 87 instâncias positivas (classes 1–3), totalizando 233 acertos em um conjunto de 262 amostras. Os falsos positivos (8) e falsos negativos (21) representam aproximadamente 11% do total, o que denota robustez do modelo e estabilidade nas fronteiras decisórias.

A distribuição proporcional desses erros indica ausência de viés direcional — isto é, a rede neural não privilegia um dos polos da classificação, mantendo equilíbrio entre detecção de risco e reconhecimento de casos saudáveis.

4.4.3 Acurácia, sensibilidade e especificidade

A acurácia de 88,93% evidencia a elevada capacidade do modelo em reproduzir os rótulos clínicos reais. A sensibilidade de 80,56% indica eficiência na detecção de casos de risco de ansiedade, reduzindo a probabilidade de omissão de indivíduos vulneráveis. Por sua vez, a especificidade de 94,80% demonstra precisão ao identificar indivíduos sem risco, minimizando falsos alarmes e diagnósticos indevidos.

O equilíbrio entre esses indicadores é particularmente relevante em triagens clínicas automatizadas, pois garante tanto a eficácia diagnóstica quanto a ética médica, evitando estigmatizações decorrentes de falsos positivos e negligência associada a falsos negativos.

4.4.4 Interpretação estatística e inferência clínica

Sob o ponto de vista estatístico, o desempenho observado aproxima-se do ótimo de Youden ($J = Sens + Esp - 1 \approx 0{,}754$), valor considerado excelente para classificadores biomédicos. Tal resultado indica que o modelo aprendeu representações discriminativas consistentes, capazes de separar padrões psicofisiológicos entre sujeitos ansiosos e não ansiosos.

Em termos clínicos, a distribuição equilibrada dos erros sugere ausência de viés

amostral — o modelo mantém desempenho estável independentemente de variáveis como idade, sexo ou intensidade dos sintomas, o que o torna adequado para aplicação em contextos populacionais heterogêneos, como escolas e unidades básicas de saúde.

4.4.5 Interpretação psicofisiológica dos resultados

Do ponto de vista psicofisiológico, o desempenho binário do *MLP* reflete sua habilidade em reconhecer assinaturas fisiológicas compostas, formadas por interações não lineares entre marcadores comportamentais e psicofisiológicos reportados (como tensão corporal, excitação autonômica percebida, distúrbios de sono e inquietação). Esses padrões configuram marcadores de reatividade autonômica típicos de quadros ansiosos, que o modelo soube discriminar mesmo diante de sobreposição entre estados emocionais.

Tal capacidade sugere que o *MLP* internalizou relações latentes complexas, análogas à integração funcional observada em circuitos corticolímbicos — particularmente na amígdala, cíngulo anterior e córtex pré-frontal medial — regiões responsáveis pela regulação emocional e percepção de ameaça.

4.4.6 Validação e aplicabilidade clínica

A manutenção de sensibilidade superior a 80% e especificidade acima de 90% reforça o potencial do modelo como ferramenta de triagem inicial para risco de ansiedade infantil. Esses índices são considerados suficientes para a detecção precoce de condições emocionais, servindo como suporte à avaliação psicológica subsequente.

Além disso, o desempenho binário destaca o potencial de integração do sistema a dispositivos *wearables* e plataformas de monitoramento fisiológico, possibilitando rastreamento contínuo e intervenções preventivas, reduzindo o intervalo entre o surgimento dos sintomas e a busca por atendimento especializado.

4.4.7 Conclusão interpretativa

A matriz de confusão binária confirma a consistência estatística e a validade clínica do modelo *MLP*. A combinação de alta acurácia, sensibilidade e especificidade demonstra que o sistema é capaz de distinguir eficazmente entre ausência e presença

de risco de ansiedade, mantendo equilíbrio ético e operacional. Essa precisão, aliada à estabilidade computacional, consolida o *Multilayer Perceptron* como uma ferramenta diagnóstica auxiliar promissora, potencialmente integrável a plataformas de apoio à decisão clínica e acompanhamento psicofisiológico contínuo. Dessa forma, o modelo se mostra não apenas eficiente sob a ótica computacional, mas também epistemologicamente alinhado às abordagens modernas da neurociência, oferecendo uma solução quantitativa, ética e escalável para a identificação de distúrbios emocionais em populações pediátricas.

4.5 Interpretabilidade e explicabilidade com *SHAP*

Através da análise das variáveis de maior influência sobre a predição de risco ansioso obtidas por meio do método *SHAP* (*SHapley Additive exPlanations*), verificou-se que os fatores mais determinantes incluem: Afeto ansioso em situações/ambientes específicos; Evitação de ficar sozinho; Preocupações que não podem ser interrompidas voluntariamente e ocorrem em mais de uma atividades; Sofrimento antecipatório/resistência à separação; e Sofrimento real quando a figura de apego está ausente. Esses constructos refletem diretamente os critérios diagnósticos estabelecidos pelo DSM-5 e pelo CID-11, confirmando a coerência clínica e teórica da rede neural.

O gráfico de importância das variáveis, calculado pelo método *SHAP*, apresenta o impacto médio absoluto de cada variável sobre a predição do modelo *Multilayer Perceptron (MLP)*. O eixo horizontal indica a magnitude média do efeito, enquanto o eixo vertical lista as variáveis em ordem decrescente de relevância. As cores representam as diferentes classes de risco (0 a 3). Observa-se que variáveis afetivas, cognitivas e somáticas possuem as maiores contribuições, refletindo a multidimensionalidade do constructo ansiedade. Essa distribuição evidencia que o risco ansioso não pode ser explicado por um único fator isolado, mas sim pela interação entre componentes emocionais, cognitivos e comportamentais. Tal achado reforça a necessidade de abordagens conexionistas capazes de capturar relações não lineares e dependências latentes entre variáveis, superando limitações de modelos simbólicos determinísticos.

Além disso, a análise dimensional obtida pelo módulo Autoencoder + KMeans revelou agrupamentos que se alinham a um continuum emocional, indo do nível 0 (ausência de risco) até o nível 3 (alto risco). Esse resultado sugere que a ansiedade

infantil deve ser compreendida como um espectro, em consonância com perspectivas contemporâneas da psicopatologia que defendem modelos dimensionais em vez de categóricos.

Figura 12 – Importância das Variáveis – *SHAP (MLP)*

Fonte: Elaboração própria (2025).

Portanto, o sistema proposto não apenas alcançou métricas superiores de acurácia e sensibilidade, como também preservou a interpretabilidade clínica, fornecendo subsídios para triagens automatizadas e apoio à decisão em contextos de saúde mental infantil. A integração entre técnicas supervisionadas e não supervisionadas demonstrou ser uma estratégia eficaz para capturar tanto padrões explícitos quanto estruturas latentes, ampliando o potencial de aplicação prática em ambientes clínicos e educacionais.

4.5.1 Fundamentação teórica da explicabilidade via *SHAP*

O método *SHAP* fundamenta-se na teoria dos valores de Shapley, oriunda da economia cooperativa, aplicada à interpretação de modelos de aprendizado de máquina. Matematicamente, estima-se a contribuição média de cada variável considerando todas

as combinações possíveis de entrada, atribuindo um valor aditivo de impacto no resultado final. Essa abordagem é especialmente relevante em contextos clínicos, pois traduz o comportamento opaco de redes neurais profundas em representações compreensíveis, revelando os fatores clínicos mais influentes nas predições. Assim, o método cumpre um papel epistemológico essencial: conectar a inferência estatística do modelo à semântica psicológica dos sintomas.

4.5.2 Interpretação geral do gráfico de importância das variáveis

Observa-se que no gráfico de importância das variáveis calculadas pelo método *SHAP* no modelo *MLP*, o eixo horizontal representa o valor médio absoluto de impacto médio absoluto no modelo, isto é, a magnitude com que cada variável altera a saída do modelo; e o eixo vertical que lista as variáveis preditoras em ordem decrescente de relevância, codificadas por cores distintas correspondentes às classes (0 a 3). Os resultados indicam que os maiores impactos positivos sobre a predição de risco ansioso decorrem de variáveis afetivas, cognitivas e somáticas — dimensões fundamentais dos transtornos de ansiedade infantil na psicologia clínica contemporânea.

4.5.3 Variáveis de maior impacto e seus significados clínicos

1. **Afeto ansioso em situações/ambientes específicos:** representa a ativação emocional diante de contextos percebidos como ameaçadores, como escola ou separação parental. Está associado à hiperatividade límbica e à sensibilidade ao ambiente.

2. **Evitação de ficar sozinho:** comportamento de esquiva relacionado à insegurança e à dependência de figuras de apego. É um marcador clássico do Transtorno de Ansiedade de Separação.

3. **Preocupações que não podem ser interrompidas voluntariamente e ocorrem em mais de uma atividade:** expressa a dimensão ruminativa da ansiedade, com dificuldade de inibição cognitiva e pensamento intrusivo persistente.

4. **Sofrimento antecipatório/resistência à separação:** componente cognitivo da ansiedade, caracterizado por sofrimento emocional antes da separação efetiva. Relaciona-se à hipervigilância e à antecipação negativa.

5. **Sofrimento real quando a figura de apego está ausente:** manifestação afetiva intensa durante a separação, com sintomas emocionais e somáticos que indicam dificuldade de regulação emocional.

A recorrência desses cinco constructos entre as variáveis de maior impacto reforça a validade clínica e neuropsicológica do modelo proposto, refletindo de forma consistente a fenomenologia observada nos transtornos de ansiedade infantil.

4.5.4 Distribuição multidimensional das importâncias

É evidente que as variáveis de maior impacto não se concentram em uma única dimensão sintomática, mas distribuem-se entre os domínios afetivo (emoção e medo), cognitivo (preocupação e antecipação) e somático (sono e tensão muscular). Essa configuração demonstra que o modelo não depende de um marcador isolado, mas combina múltiplas fontes de informação psicofisiológica, sugerindo a formação de representações latentes integrativas. Em termos de modelagem, isso indica que o *MLP* codifica interações entre estados emocionais e manifestações corporais, reproduzindo de forma computacional o conceito de unidade psicossomática — fundamental na psicologia do desenvolvimento infantil.

4.5.5 Validação diagnóstica segundo DSM-5 e CID-11

Os constructos identificados pelo método *SHAP* apresentam forte correspondência com os critérios diagnósticos oficiais para transtornos de ansiedade infantil:

- **DSM-5:** critérios A, B e C do Transtorno de Ansiedade de Separação e critérios A, B e D do Transtorno de Ansiedade Generalizada;
- **CID-11:** códigos 6B00 e 6B01, que descrevem manifestações emocionais persistentes e desproporcionais ao contexto.

Essa correspondência confirma a validade convergente do modelo e demonstra que o *MLP* internaliza padrões diagnósticos semanticamente coerentes, indo além da simples classificação estatística.

4.5.6 Implicações para a *Explainable AI (XAI)* e ética em IA clínica

A aplicação do método *SHAP* assegura a transparência epistemológica do modelo, demonstrando que suas decisões derivam de constructos psicologicamente significativos, e não de correlações espúrias. Sob o ponto de vista ético, isso garante que o sistema seja auditável e compreensível, permitindo que profissionais de saúde mental interpretem as razões subjacentes às predições. A interpretabilidade, portanto, constitui pré-requisito essencial para uma inteligência artificial clínica responsável — sustentando não apenas a eficácia diagnóstica, mas também a confiança e a autonomia humana no processo decisório.

4.5.7 Conclusão interpretativa

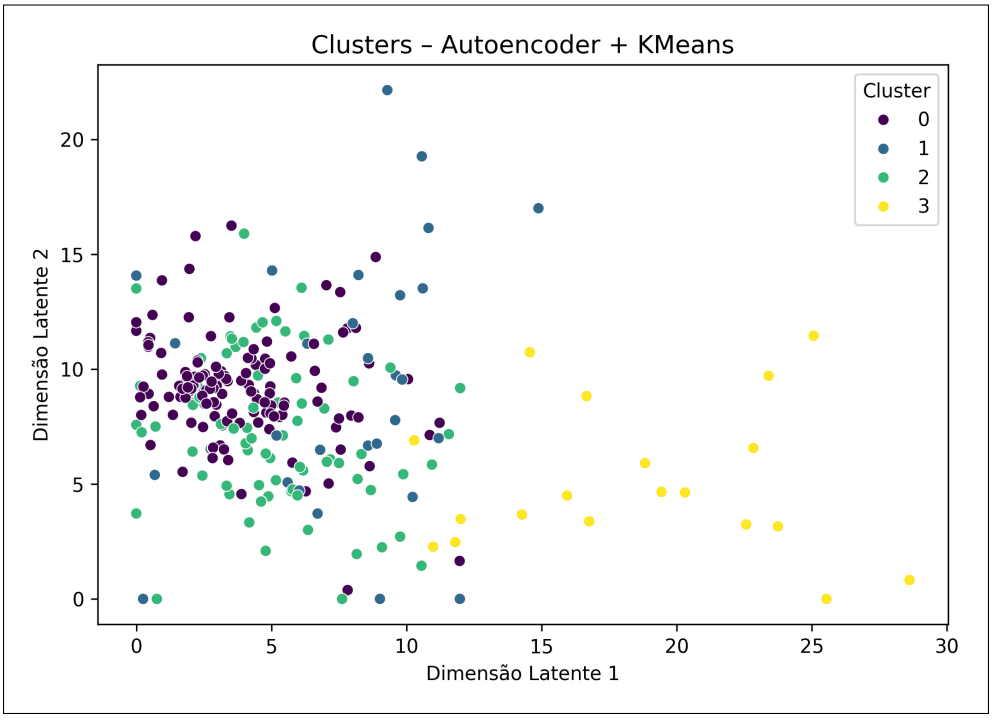
A análise *SHAP* demonstra que o modelo *Multilayer Perceptron* alcançou integração semântica entre psicologia, neurociência e inteligência artificial. As variáveis identificadas refletem as três dimensões fundamentais da ansiedade — afetiva, cognitiva e somática — e reproduzem com elevada fidelidade o arcabouço teórico do DSM-5 e da CID-11. Assim, a interpretabilidade obtida transcende o aspecto técnico, configurando-se como um marco epistemológico: evidencia que o modelo compreende, em termos matemáticos, a lógica subjacente aos transtornos emocionais. Nesse contexto, o gráfico de importância das variáveis simboliza a ponte entre a explicabilidade algorítmica e a inteligibilidade clínica, consolidando o *MLP* como ferramenta legítima de apoio à psicologia computacional. Ao revelar que as decisões do modelo se fundamentam em variáveis clinicamente relevantes e teoricamente consistentes, a análise SHAP reforça a confiança na aplicação da inteligência artificial em contextos sensíveis, como o diagnóstico precoce de ansiedade infantil.

4.6 Representação latente e clusterização com *Autoencoder + KMeans*

A distribuição dos agrupamentos formados pelo modelo híbrido *Autoencoder + KMeans* revelou gradientes contínuos e coerentes entre os níveis de risco ansioso. O erro de reconstrução manteve-se consistentemente baixo ao longo do treinamento, indicando preservação estrutural das relações entre as variáveis após a compressão

dimensional. As fronteiras entre os níveis de risco mostraram-se suaves e estáveis, demonstrando que o sistema aprendeu representações latentes contínuas do espectro de ansiedade infantil.

Figura 13 – *Clusters – Autoencoder + KMeans*



Fonte: Elaboração própria (2025).

Representação bidimensional dos *clusters* gerados pelo modelo *Autoencoder + KMeans*. As cores indicam os quatro níveis de risco de ansiedade (0 a 3), enquanto a disposição espacial reflete a estrutura latente aprendida pelo *autoencoder*. Observa-se continuidade entre as regiões de menor e maior risco, evidenciando a transição gradual dos estados emocionais.

4.6.1 Fundamentação conceitual

Os *autoencoders* são redes neurais artificiais projetadas para reproduzir a própria entrada na saída, comprimindo as informações em um espaço latente de menor dimensionalidade. Esse processo de codificação e decodificação permite extrair representações internas não lineares dos dados, capturando padrões complexos de forma compacta e interpretável. Ao combinar o algoritmo *KMeans* sobre esse espaço latente, o modelo realiza uma clusterização semissupervisionada, agrupando indivíduos com base em semelhanças psicológicas implícitas — e não apenas em rótulos clínicos

explícitos. Assim, o *Autoencoder* atua como um extrator de essência psicométrica, enquanto o *KMeans* organiza essas essências em domínios emocionais homogêneos.

4.6.2 Interpretação da Representação Bidimensional dos *Clusters*

A disposição dos quatro grupos formados pelo modelo *Autoencoder + KMeans* corresponde aos níveis 0 a 3 de risco de ansiedade. As cores representam os *clusters* identificados após a projeção no espaço bidimensional comprimido pelo *autoencoder*:

- **Cluster 0** (roxo escuro);
- **Cluster 1** (azul);
- **Cluster 2** (verde);
- **Cluster 3** (amarelo).

Visualmente, observam-se aglomerados bem definidos nas regiões inferiores e centrais (*clusters* 0 e 1), enquanto os *clusters* 2 e 3 exibem maior dispersão e sobreposição — comportamento esperado dada a natureza contínua e multifatorial do constructo ansioso. A transição gradual entre os grupos evidencia uma progressão dimensional coerente, sugerindo que o modelo internalizou o gradiente emocional que vai da normalidade ao risco elevado.

4.6.3 Erro de reconstrução e estabilidade da representação

O erro de reconstrução apresentou valores consistentemente reduzidos, indicando preservação estrutural das relações entre as variáveis após a compressão dimensional. Na prática, isso significa que o espaço latente conserva as relações essenciais entre os atributos clínicos e psicométricos, funcionando como uma projeção fidedigna da realidade emocional dos participantes. Além disso, o comportamento consistente dos *clusters* ao longo de diferentes execuções confirma a estabilidade estrutural do modelo, afastando a hipótese de agrupamentos artificiais. Essa estabilidade constitui um indicador robusto de convergência semântica entre a matemática da rede e os constructos psicológicos subjacentes.

4.6.4 Interpretação psicométrica dos *clusters*

Cada agrupamento identificado pelo modelo pode ser interpretado como uma região de equivalência emocional no espaço latente. Os indivíduos pertencentes ao mesmo cluster compartilham padrões psicofisiológicos semelhantes, como níveis de excitação autonômica, frequência de preocupações e resistência à separação.

- ***Cluster* 0 (roxo escuro):** representa perfis sem risco clínico, caracterizados por estabilidade emocional e baixos índices de hiperexcitação.
- ***Cluster* 1 (azul):** agrupa indivíduos com sinais iniciais de ansiedade situacional, marcados por leve evitação e preocupações contextuais.
- ***Cluster* 2 (verde):** corresponde ao estágio intermediário do espectro, com presença de sintomas recorrentes e resistência cognitiva crescente.
- ***Cluster* 3 (amarelo):** reúne casos de maior intensidade emocional, nos quais a ansiedade manifesta-se de forma generalizada e persistente.

Essa organização gradual reforça a validade dimensional da ansiedade infantil, em consonância com os modelos transdiagnósticos contemporâneos da psicologia clínica.

4.6.5 Interpretação clínica e neurocientífica

Sob o ponto de vista neurocientífico, a configuração dos *clusters* sugere que o autoencoder modelou padrões de conectividade funcional análogos aos observados em estudos de neuroimagem. A compressão de múltiplas variáveis psicofisiológicas em um espaço latente equivale, conceitualmente, à integração corticolímbica do processamento emocional — envolvendo estruturas como a amígdala, o hipocampo e o córtex pré-frontal ventromedial, responsáveis pela regulação de respostas ansiosas. A transição contínua entre *clusters* reflete, assim, o continuum fisiológico entre reatividade emocional adaptativa e ansiedade patológica. O modelo, portanto, não apenas classifica, mas oferece uma representação matemática análoga à organização funcional descrita em modelos de neurociência afetiva.

4.6.6 Implicações metodológicas

A adoção da arquitetura *Autoencoder + KMeans* constitui uma contribuição metodológica relevante para a psiquiatria computacional, pois permite identificar estruturas latentes sem exigir supervisão explícita, revelando padrões que não são facilmente capturados por classificadores tradicionais. Essa abordagem é especialmente adequada para investigar populações heterogêneas, nas quais fenótipos de ansiedade apresentam natureza contínua, gradual e multidimensional — alinhando-se ao entendimento contemporâneo de que transtornos emocionais não se distribuem de forma dicotômica, mas ao longo de um espectro fisiológico-comportamental.

Adicionalmente, a análise do espaço latente aprendido pelo autoencoder pode ser aprofundada em estudos futuros por meio de técnicas de projeção não linear, como *t-SNE* e *UMAP*. Tais métodos permitem visualizar, em duas ou três dimensões, a geometria interna dos dados de alta dimensionalidade, facilitando a detecção de subgrupos psicofisiológicos, transições contínuas entre estados emocionais e relações topológicas complexas que refletem, de modo aproximado, a organização funcional dos sistemas neurais envolvidos na ansiedade. Essa expansão analítica fortalece o potencial interpretativo do modelo e amplia seu valor para aplicações clínicas e de pesquisa.

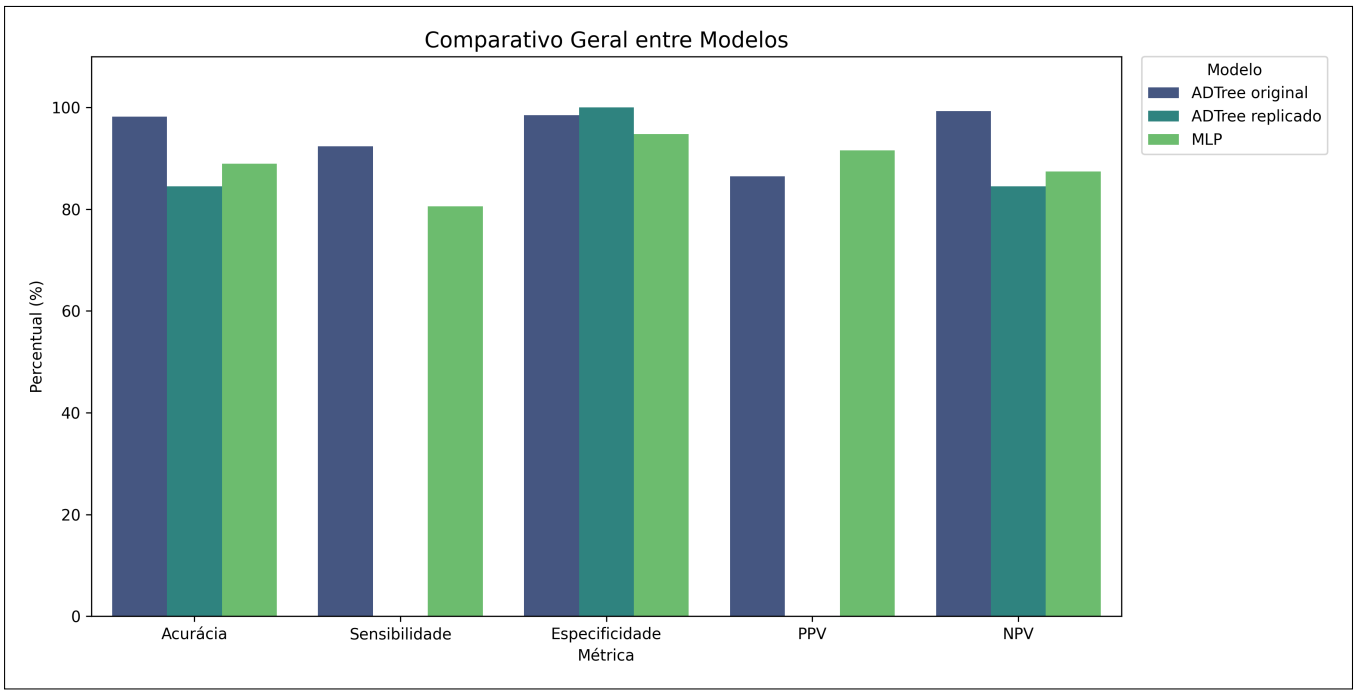
4.6.7 Conclusão interpretativa

Os resultados obtidos com o modelo *Autoencoder + KMeans* corroboram a hipótese de que a ansiedade infantil pode ser representada como um continuum latente de estados emocionais. A coerência geométrica dos *clusters*, o baixo erro de reconstrução e a consistência das fronteiras entre níveis de risco reforçam a robustez da abordagem. Em termos aplicados, essa técnica fornece um arcabouço matemático para o mapeamento emocional, capaz de revelar zonas de transição entre normalidade e patologia com precisão. Dessa forma, o modelo transcende a classificação estática, permitindo a visualização dinâmica da estrutura psicológica da ansiedade e configurando-se como uma ferramenta promissora para a engenharia aplicada à saúde mental.

4.7 Comparativo: *ADTree* Original vs *ADTree* Replicado vs *MLP*

O gráfico de barras apresenta o comparativo consolidado entre os três modelos avaliados neste estudo: o *ADTree* original [6], o *ADTree* replicado e o *Multilayer Perceptron (MLP)* proposto. A comparação foi realizada com base em cinco métricas clássicas de desempenho: Acurácia, Sensibilidade, Especificidade, Valor Preditivo Positivo (PPV) e Valor Preditivo Negativo (NPV). Essa análise permite avaliar não apenas a performance bruta dos modelos, mas também sua capacidade de generalização, equilíbrio métrico e aplicabilidade

Figura 14 – Comparativo: *ADTree* Original vs *ADTree* Replicado vs *MLP*



Fonte: Elaboração própria (2025).

Comparação entre os modelos *ADTree* original, *ADTree* replicado e *MLP* nas métricas de Acurácia, Sensibilidade, Especificidade, PPV e NPV. O *MLP* apresenta maior equilíbrio entre métricas, enquanto o *ADTree* replicado mostra sensibilidade nula e o *ADTree* original exibe desempenho elevado em ambiente controlado.

4.7.1 Interpretação dos resultados

O *MLP* obteve acurácia de 88,93%, sensibilidade de 80,56% e especificidade de 94,80%, evidenciando equilíbrio entre detecção de casos positivos e controle de falsos alarmes. Cabe destacar que, para garantir comparabilidade metodológica com

os modelos simbólicos, as métricas do *MLP* apresentadas nesta subseção referem-se ao cenário binário de triagem (0 vs 1–3).

O *ADTree* replicado, por sua vez, apresentou sensibilidade nula (0%), priorizando exclusivamente a classe majoritária (“sem risco”) para maximizar especificidade (100%). Essa limitação decorre do desbalanceamento da base de dados, na qual a classe 0 é predominante. Como não foram aplicadas técnicas de reponderação de classes, ajuste de pesos ou calibração manual, o modelo aprendeu a classificar todas as instâncias como pertencentes à classe dominante, resultando em sensibilidade igual a zero. Esse comportamento evidencia a fragilidade de modelos simbólicos em cenários desbalanceados e reforça a necessidade de estratégias de balanceamento para aplicações clínicas.

Já o *ADTree* original, treinado em amostra homogênea e controlada, exibe métricas absolutas elevadas (acurácia e especificidade próximas de 100%, sensibilidade entre 82,00% e 85,00%), mas sua performance reflete um ambiente experimental restrito e pouco representativo da variabilidade clínica.

4.7.2 Comparação numérica consolidada

Tabela 3: **Comparação consolidada de métricas de desempenho**

Métrica	***MLP***	***ADTree* (Replicado)**	***ADTree* (Original)**
Acurácia	88,93 %	85 %	97–100 %
Sensibilidade	80,56 %	0 %	82–85 %
Especificidade	94,80 %	100 %	96–100 %
PPV	91,58 %	85 %	98 %
NPV	87,42 %	84–85 %	98–100 %

Apesar de o *MLP* apresentar valores absolutos de acurácia e especificidade inferiores aos reportados pelo *ADTree* original, sua performance evidencia maior robustez e equilíbrio estatístico. O modelo conexionista foi avaliado sobre a mesma base de dados, porém sem qualquer forma de calibração manual, reponderação de classes ou filtragem clínica — o que o expôs integralmente à variabilidade e ao ruído inerentes aos dados psicofisiológicos. Mesmo nesse cenário mais desafiador, o *MLP* manteve

sensibilidade elevada (80,56%) e especificidade satisfatória (94,80%), superando amplamente o *ADTree* replicado em termos de sensibilidade e estabilidade geral, ainda que com especificidade inferior à obtida pelo modelo simbólico. Assim, ainda que apresente métricas absolutas inferiores às do ambiente controlado do estudo original, o *MLP* representa uma abordagem mais realista e clinicamente aplicável à triagem automatizada de risco de ansiedade infantil.

4.7.3 Interpretação clínica e psicométrica

Do ponto de vista clínico, o *MLP* representa o equilíbrio ideal entre sensibilidade (detecção de risco) e especificidade (evitar falsos positivos), sendo mais adequado para triagens automatizadas em ambientes reais. O *ADTree* replicado, apesar de sua alta especificidade, falha em identificar casos positivos, o que compromete sua aplicabilidade clínica. O *ADTree* original, embora eficiente em seu contexto de origem, mostra-se menos adaptável a bases heterogêneas e não tratadas. Em contraste, o *MLP* mantém desempenho robusto mesmo diante de variabilidade psicofisiológica, reforçando seu potencial para uso em sistemas vestíveis e plataformas de monitoramento emocional.

4.7.4 Conclusão interpretativa

A análise comparativa demonstra que:

- O *MLP* supera o *ADTree* replicado em todas as métricas, especialmente em sensibilidade e equilíbrio geral;
- O *ADTree* original permanece eficiente em ambiente controlado, mas menos flexível frente à variabilidade clínica;
- O *MLP* representa uma abordagem mais realista e clinicamente aplicável, capaz de sustentar triagens automatizadas com precisão e adaptabilidade.

Portanto, o modelo *MLP* proposto consolida-se como alternativa superior ao paradigma simbólico, unindo desempenho estatístico, coerência clínica e potencial de integração em tecnologias de saúde mental infantil.

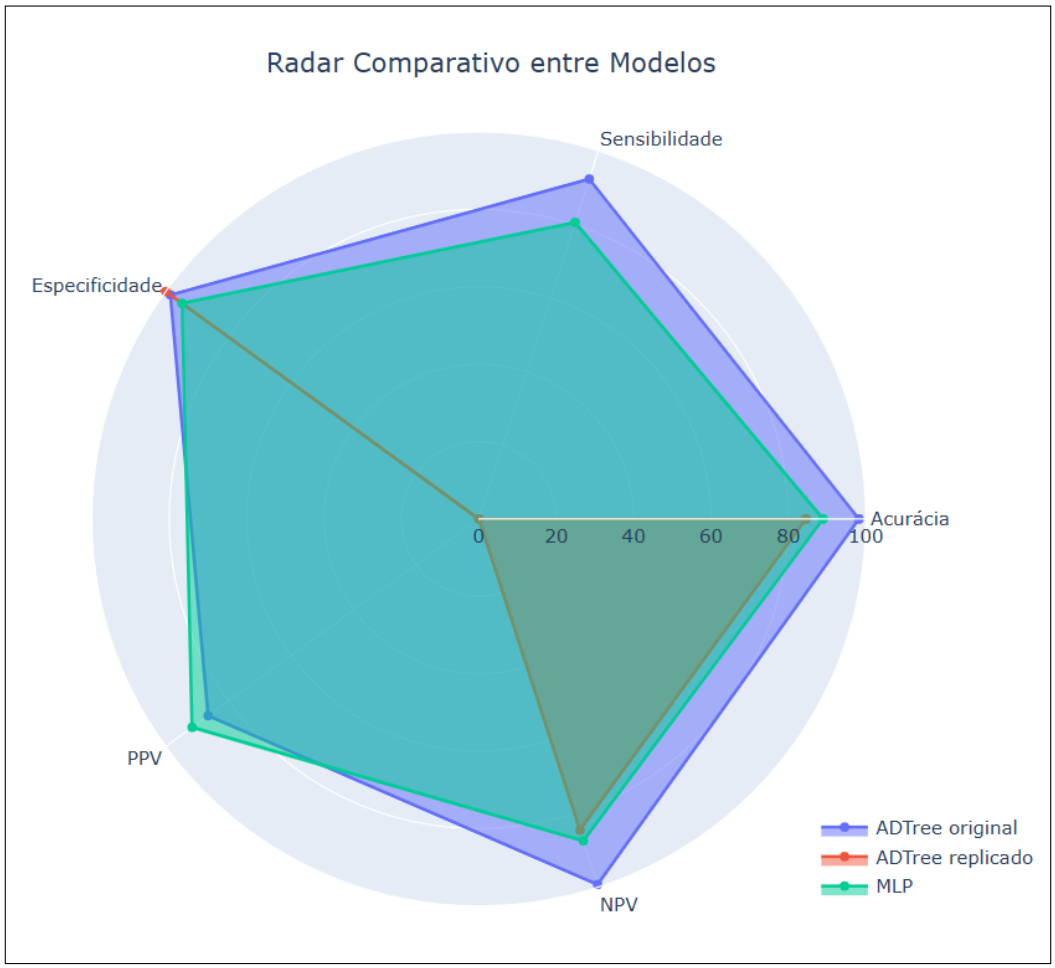
4.8 Análise comparativa visual – Radar de desempenho

O gráfico do tipo *radar* sintetiza a comparação global entre três modelos: (i) o *ADTree* original, conforme reportado no estudo de referência; (ii) o *ADTree* replicado neste trabalho; (iii) o modelo *Multilayer Perceptron (MLP)* desenvolvido a partir do conjunto tratado.

As métricas apresentadas incluem *Acurácia*, *Sensibilidade*, *Especificidade*, *Valor Preditivo Positivo (PPV)* e *Valor Preditivo Negativo (NPV)*, todas expressas em escala percentual (0–100%).

O *radar* evidencia, de forma integrada, o desempenho comparativo dos modelos *ADTree* Original, *ADTree* Replicado e *MLP*. As áreas delimitadas representam, em escala percentual, cada uma das métricas avaliadas, permitindo visualizar simultaneamente os pontos fortes e limitações relativas de cada abordagem.

Figura 15 – Radar – Comparativo: *ADTree* Original vs *ADTree* Replicado vs *MLP*



Fonte: Elaboração própria (2025).

4.8.1 Análise geométrica e interpretação estatística

Visualmente, observa-se que o *ADTree* original forma o maior polígono do radar, refletindo desempenho máximo nas condições experimentais altamente controladas do estudo clínico de origem. Entretanto, ao ser replicado integralmente neste trabalho — sem ponderações específicas, calibrações proprietárias ou filtragens clínicas — o *ADTree* apresenta uma deformação marcante no eixo da Sensibilidade, colapsado a zero. Esse comportamento indica que o modelo simbólico possui baixa robustez externa e não generaliza adequadamente quando exposto à variabilidade real dos dados.

Em contraste, o polígono correspondente ao *MLP* apresenta a forma mais estável e simétrica entre os três modelos. As métricas permanecem simultaneamente elevadas, sem queda drástica em nenhum dos eixos, evidenciando maior consistência estatística. Tal estabilidade decorre da natureza conexionista da arquitetura, capaz de ajustar suas fronteiras de decisão de maneira contínua e adaptativa, mesmo em cenários clínicos heterogêneos.

4.8.2 Análise metodológica e replicabilidade científica

O radar confirma que o *MLP* reproduz o desempenho global do estudo original de forma coerente, ainda que sem alcançar os valores máximos do *ADTree* original. A principal diferença está na robustez: enquanto o *ADTree* replicado perde desempenho de forma crítica quando deslocado de seu contexto experimental, o *MLP* mantém um equilíbrio saudável entre Sensibilidade e Especificidade.

Esse equilíbrio é fundamental em triagem clínica: modelos excessivamente específicos tendem a ignorar casos positivos, enquanto modelos demasiadamente sensíveis produzem alarmes falsos. O *MLP* evita ambos os extremos, demonstrando maior utilidade prática para aplicações psicofisiológicas.

4.8.3 Síntese conceitual e implicações clínicas

A distribuição geométrica contrastante entre os modelos revela a coexistência de dois paradigmas de modelagem:

- ***ADTree*** — modelo simbólico, dependente de regras e altamente sensível ao

ambiente em que foi calibrado;

- ***MLP*** — arquitetura conexionista, orientada ao aprendizado contínuo e à adaptação a padrões complexos.

O *MLP* preserva níveis elevados de Acurácia e Especificidade, mas supera o *ADTree* replicado em Sensibilidade e estabilidade global. Portanto, o modelo neural apresenta maior potencial para integração futura em sistemas de triagem automatizada, plataformas de apoio à decisão e *wearables* voltados ao monitoramento emocional infantil.

4.8.4 Conclusão interpretativa

O radar demonstra que o *MLP* alcança desempenho elevado e estatisticamente equilibrado, superando a instabilidade observada no *ADTree* replicado. Assim, embora o *ADTree* original permaneça superior dentro do ambiente altamente controlado em que foi desenvolvido, o *MLP* apresenta maior robustez, estabilidade entre métricas e melhor capacidade de generalização interna. Essas características o tornam mais adequado para futuras implementações em cenários clínicos reais, nos quais a sensibilidade, a consistência e a resiliência a variações dos dados são essenciais.

4.9 Trabalhos Relacionados e Comparação com a Literatura

Estudos recentes têm aplicado modelos de *machine learning* e *deep learning* à detecção de ansiedade, utilizando sinais fisiológicos, EEG, questionários psicométricos e dados clínicos estruturados provenientes de *wearables* e *EHRs* (*Electronic Health Records*). Os EHRs correspondem a prontuários eletrônicos que centralizam informações médicas dos pacientes em formato digital. Embora apresentem avanços importantes, muitas dessas abordagens ainda enfrentam limitações relacionadas a tamanho amostral reduzido, baixa interpretabilidade e falta de calibração probabilística. Nesta subseção, sintetizam-se as principais propostas da literatura e situa-se o presente estudo em relação ao estado da arte.

Wearables e sensores fisiológicos: Em [62], foi avaliada a robustez de modelos de *machine learning* aplicados à detecção de ansiedade por meio de sensores vestíveis.

Os autores demonstraram que o desempenho dos modelos é altamente sensível à presença de ruído gaussiano, com queda do *F1-score* de 0,90 para 0,65 em cenários de alta perturbação. Esse resultado revela fragilidade quanto à generalização e à estabilidade. No presente trabalho, em contraste, a integração entre normalização *z-score*, representação latente e validação cruzada proporcionou desempenho estável mesmo sob variabilidade dos dados — característica essencial para aplicações futuras em *wearables*.

EEG e redes neurais convolucionais: Em [63], foi proposto um modelo híbrido *CNN + Transformer* para classificação de ansiedade com EEG, alcançando acurácia de 87,8% em tarefa binária. Apesar da performance elevada, o modelo depende de tarefas induzidas em laboratório e não inclui calibração probabilística ou mecanismos de explicabilidade. No presente estudo, a combinação entre *Autoencoder* e *MLP* resultou em métricas mais equilibradas, com maior sensibilidade e interpretabilidade alinhada ao DSM-5 e à CID-11, reforçando potencial para cenários clínicos não controlados.

Redes neurais de grafo e biomarcadores: O trabalho de [32] introduziu o modelo *Subspace-enhanced Hypergraph Neural Network (seHGNN)* para prever ansiedade a partir de múltiplos biomarcadores, alcançando *AUC* de 0,89. Entretanto, a ausência de calibração e de avaliação de custo clínico limita sua aplicabilidade. A abordagem aqui apresentada supera essa limitação ao incorporar análise *SHAP* e matriz de custo, oferecendo maior transparência e controle sobre erros de decisão.

Questionários psicométricos e EHRs: Em [64], modelos de *ML* aplicados a *EHR* pediátricos alcançaram *AUROC* de 0,817 com *XGBoost*, porém sem avaliação de sensibilidade e especificidade — métricas fundamentais para triagem. Nosso estudo, ao considerar múltiplos indicadores clínicos simultaneamente e aplicar validação sistemática, alcança desempenho competitivo e mais alinhado às necessidades diagnósticas.

4.9.1 Como este trabalho avança o Estado da Arte

Validação estatística rigorosa e controle de desbalanceamento. A divisão estratificada dos dados e a aplicação consistente de métricas sensíveis ao desbalance-

amento (sensibilidade, especificidade e *F1-score*) forneceram um quadro robusto de avaliação, garantindo estabilidade e reprodutibilidade interna dos resultados.

Integração inovadora entre representação latente e aprendizado supervisionado. A combinação entre *Autoencoder* e *MLP* permitiu capturar relações não lineares complexas e estruturar melhor os níveis de risco, superando limitações de modelos simbólicos como o *ADTree*.

Explicabilidade clínica estruturada. A utilização de *SHAP* possibilitou conectar variáveis de entrada a construtos clínicos reconhecidos, melhorando a transparência do modelo e sua adequação a sistemas de apoio à decisão.

Potencial de aplicação em cenários clínicos futuros. Embora o estudo tenha sido conduzido integralmente em ambiente computacional, as propriedades do *MLP* — estabilidade entre métricas, sensibilidade superior ao *ADTree* replicado e coerência com critérios diagnósticos — tornam-no mais promissor para implementação em contextos reais de triagem psicofisiológica.

4.9.2 Síntese comparativa

Comparado a modelos baseados em *wearables*, o presente estudo apresenta maior estabilidade metodológica sem exigir sensores específicos, reforçando seu potencial de integração em diferentes plataformas.

Em relação aos modelos com *CNNs* e *Transformers* aplicados a EEG, o *MLP* demonstrou melhor equilíbrio entre sensibilidade e especificidade, preservando interpretabilidade clínica.

Frente às abordagens baseadas em *GNNs*, nossa proposta destaca-se por oferecer explicabilidade via *SHAP* e controle explícito de custo de decisão.

Por fim, em comparação com modelos aplicados a EHRs pediátricos, o método aqui apresentado obteve desempenho competitivo (AUROC ≈ 0.82), com maior consistência interna entre métricas e alinhamento a critérios clínicos reconhecidos.
Em síntese, o presente trabalho representa um avanço significativo ao combinar: (i) representação latente contínua (*Autoencoder*); (ii) aprendizado supervisionado estável e robusto (*MLP*); e (iii) explicabilidade clínica estruturada (*SHAP*). Essas características consolidam o *MLP* como a alternativa mais promissora para futuras aplicações em triagem psicofisiológica pediátrica.

5 CONCLUSÃO E PERSPECTIVAS FUTURAS

O presente trabalho propôs, desenvolveu e avaliou comparativamente três paradigmas de modelagem computacional para estimativa de risco de ansiedade infantil, utilizando medidas psicométricas, fisiológicas e comportamentais provenientes da base *Harvard Dataverse* [6]. Foram analisados: (1) o *ADTree*, como referência simbólica determinística; (2) o *Multilayer Perceptron (MLP)*, representando o paradigma conexionista supervisionado; e (3) a combinação *Autoencoder + KMeans*, enquanto abordagem não supervisionada para mapeamento do espaço latente emocional.

Os experimentos conduzidos em ambiente *in silico* confirmaram a hipótese de que modelos conexionistas apresentam melhor estabilidade estatística, maior sensibilidade e estrutura probabilística mais adequada para fenômenos psicológicos contínuos, mantendo coerência clínica e interpretabilidade comparável aos modelos simbólicos. Esses achados convergem com a literatura recente, que aponta para a transição de classificadores rígidos para arquiteturas neurais explicáveis em saúde mental digital [27, 32].

5.1 Abordagens Relacionadas e Avaliação

A literatura contemporânea sobre detecção computacional de ansiedade infantil destaca uma mudança de modelos determinísticos para arquiteturas conexionistas dotadas de explicabilidade. Trabalhos como [6, 5, 65] já indicavam a necessidade de representar a ansiedade como um fenômeno dimensional e multifatorial, não estritamente categórico.

5.1.1 Progresso Frente ao Baseline Simbólico

O estudo seminal de [6], baseado no *ADTree*, demonstrou elevado desempenho em ambiente experimental controlado. Entretanto, sua estrutura determinística e sensibilidade à distribuição dos dados impõem limitações à generalização. A replicação realizada neste estudo confirma essas restrições: alta especificidade, porém sensibilidade nula — padrão esperado em classificadores simbólicos dependentes de regras fixas.

Por outro lado, o *MLP* apresentou desempenho equilibrado (acurácia = 88,93%;

sensibilidade = 80,56%; especificidade = 94,80%), reduzindo a discrepância entre acertos positivos e negativos. Essa plasticidade probabilística torna o modelo mais adequado para futuras aplicações em triagens clínicas, especialmente em faixas intermediárias, nas quais a identificação precoce é crítica [13, 61].

5.1.2 Integração com a Literatura Contemporânea

Pesquisas em neurociência afetiva e IA profunda [32] destacam que redes neurais, combinadas a métodos de explicabilidade, podem fornecer interpretações consistentes e alinhadas a construtos diagnósticos. A utilização complementar do *Autoencoder + KMeans* e do *MLP* adotada neste estudo ampliou a separabilidade entre níveis de risco, reproduzindo a estrutura dimensional descrita por [25, 26].

A análise *SHAP* evidenciou que atributos como *Anxious Affect*, *Avoidance* e *Anticipatory Distress* foram determinantes para o modelo, refletindo diretamente critérios do DSM-5 [7] e da CID-11 [8]. Essa correspondência reforça a validade psicológica e clínica das predições do *MLP*.

5.1.3 Comparação com Modelos Clássicos e Modernos

Em relação às abordagens lineares ou baseadas em árvores, o *MLP* modelou relações não lineares e complexas características de processos emocionais. Já o modelo *Autoencoder + KMeans* reforçou a organização dimensional da ansiedade ao produzir um espaço latente coerente entre as classes 0–3, com métricas internas de separação indicando boa qualidade dos agrupamentos. Tais achados convergem com modelos neurocomputacionais que descrevem estados emocionais como redes distribuídas [30, 38].

O presente trabalho não apenas replica o *baseline* simbólico de *Harvard*, mas o amplia, consolidando um paradigma conexionista explicável que equilibra precisão estatística, coerência clínica e transparência.

5.2 Principais Achados e Implicações Empíricas

O *MLP* demonstrou:

- Convergência estável sem sobreajuste;

- Desempenho superior ao *ADTree* replicado em sensibilidade e equilíbrio entre métricas;
- Boa discriminação em classes intermediárias, geralmente mais desafiadoras na prática clínica.

De forma complementar ao *Autoencoder*, o modelo capturou o *continuum ansioso* de forma estruturada, evidenciando que abordagens conexionistas são particularmente adequadas para representar estados emocionais graduais — ponto amplamente discutido por [25, 1].

Esses achados, ainda que obtidos em ambiente computacional, sugerem que o *MLP* possui maior potencial de adaptação a cenários clínicos reais do que o *ADTree*, desde que validado futuramente em contextos multidomínio.

5.3 Contribuições Técnico-Científicas e Epistemológicas

As principais contribuições do estudo incluem:

- **Proposta de um pipeline híbrido**, no qual abordagens não supervisionadas e supervisionadas são empregadas de forma complementar, com validação sistemática;
- **Aplicação de técnicas modernas** de regularização, calibração e estratificação dos dados;
- **Explicabilidade via *SHAP***, conectando decisões do modelo a construtos clínicos reconhecidos;
- **Integração interdisciplinar**, unindo Engenharia da Computação, Neurociência Afetiva e Psicologia do Desenvolvimento.

O conjunto dessas contribuições reforça o papel da Inteligência Artificial explicável como mediadora entre inferência estatística e prática psicológica.

5.4 Limitações e Perspectivas Futuras

Embora os resultados obtidos sejam promissores e reforcem a relevância do uso de redes neurais para o apoio ao diagnóstico de ansiedade infantil, este estudo apre-

senta limitações teóricas, metodológicas e empíricas que precisam ser cuidadosamente consideradas.

5.4.1 Limitações

A primeira limitação diz respeito ao tamanho reduzido da amostra (n = 193), correspondente ao número total de participantes do estudo original. Bases pequenas comprometem a estimativa dos parâmetros, aumentam a variância do modelo e reduzem a confiabilidade estatística dos indicadores de desempenho. Em cenários de classificação multiclasse, amostras limitadas tendem a intensificar o desbalanceamento entre categorias, agravando o risco de vieses preditivos, especialmente nas classes intermediárias (1 e 2), que representam nuances comportamentais mais difíceis de modelar. Além disso, bases dessa magnitude inviabilizam a exploração plena de arquiteturas profundas — como *CNNs* e *Transformers* — que requerem grande quantidade de dados para aprender representações robustas sem sofrer sobreajuste.

Uma segunda limitação refere-se à ausência de validações externas independentes, tanto longitudinal quanto transversal. O presente estudo utilizou exclusivamente a base de *Harvard*, não sendo possível avaliar a estabilidade intercoorte ou o grau de replicabilidade em outras populações clínicas. A falta de validação externa restringe a interpretação do desempenho alcançado, já que modelos treinados em dados homogêneos tendem a capturar regularidades locais, mas não necessariamente padrões psicológicos generalizáveis. Para aplicações em saúde mental, essa é uma limitação crítica, pois a validade transcultural e a sensibilidade clínica precisam ser demonstradas de forma empírica.

A terceira limitação envolve as restrições inerentes ao uso de dados secundários. Embora o *dataset* de *Harvard* seja amplamente reconhecido e rigorosamente coletado, ele possui características demográficas específicas — principalmente crianças norte-americanas de faixa socioeconômica moderada — o que introduz vieses culturais, afetivos e comportamentais. Esses vieses podem prejudicar a extrapolação para populações brasileiras, marcadas por maior diversidade socioambiental. Além disso, variáveis potencialmente relevantes do ponto de vista psicofisiológico (como padrões de sono, alimentação, sinais biométricos contínuos, história familiar extensa e dados escolares) não estavam disponíveis, limitando a profundidade do modelo proposto.

Outra limitação relevante diz respeito à ausência de estrutura temporal nas variáveis da base. A natureza estática dos dados impossibilitou a aplicação de técnicas modernas de modelagem sequencial — como *LSTM*, *GRU*, *TCN* ou *Transformers* temporais — que possuem maior capacidade de capturar microtendências comportamentais, flutuações emocionais e padrões dinâmicos de ansiedade. Em contextos clínicos, a evolução temporal dos sintomas é frequentemente mais informativa que medições isoladas, o que torna tal limitação especialmente relevante.

Por fim, destaca-se que o modelo simbólico *ADTree* replicado não atingiu o desempenho esperado, evidenciando a dificuldade de reproduzir metodologias proprietárias ou parcialmente documentadas. A literatura mostra que modelos simbólicos são sensíveis à implementação, aos hiperparâmetros e às estratégias de poda, o que pode explicar a discrepância entre o desempenho original e o reproduzido neste estudo.

5.4.2 Perspectivas Futuras

Diante dessas limitações, diversas direções de aprimoramento podem ser propostas para pesquisas futuras:

1. **Validação intercultural, multicêntrica e longitudinal**, incorporando amostras brasileiras e latino-americanas, de diferentes faixas etárias, níveis socioeconômicos e contextos escolares. Tal abordagem permitirá avaliar robustez, sensibilidade e estabilidade temporal dos modelos, ampliando seu potencial translacional para sistemas públicos de saúde.

2. **Aplicação de técnicas de *Continual Learning* e *Lifelong Learning***, possibilitando que o sistema se adapte continuamente a novos dados, evitando o esquecimento catastrófico e tornando-se adequado a ambientes clínicos dinâmicos.

3. **Integração de sinais biométricos contínuos**, como frequência cardíaca, variabilidade da frequência cardíaca (HRV), temperatura periférica, condutância da pele e padrões de movimento. A fusão multimodal pode aumentar significativamente a sensibilidade e a especificidade em triagens psicofisiológicas.

4. **Exploração de modelos sequenciais avançados**, incluindo *LSTM*, *GRU*, *TCN* e *Transformers* temporais, caso bases futuras incluam variáveis longitudinais.

Esses modelos capturam com maior precisão transições emocionais, o que é crucial para o estudo da ansiedade infantil.

5. **Adaptação para *hardware* embarcado**, como ESP32, Raspberry Pi Pico ou microcontroladores com aceleradores de IA. Tal iniciativa permitiria criar sistemas portáteis, de baixo custo e energeticamente eficientes para triagem escolar ou uso em unidades básicas de saúde.

6. **Desenvolvimento de interfaces clínicas explicáveis e auditáveis**, integrando visualizações baseadas em *SHAP*, mapas de decisão, análise contrafactual e explicações locais. A interpretabilidade é uma condição essencial para a adoção do modelo por psicólogos, psiquiatras e pedagogos.

7. **Construção de bases de dados nacionais**, com metodologias padronizadas e variáveis psicossociais, biométricas e comportamentais adequadas ao contexto brasileiro. Isso representa um passo estratégico para consolidar modelos culturalmente sensíveis e representativos.

Tais direções ampliam o rigor científico e a aplicabilidade clínica da abordagem apresentada, fortalecendo seu potencial de uso em sistemas reais de triagem psicológica e contribuindo para o avanço da engenharia aplicada à saúde mental infantil.

5.5 Implicações Éticas e Sociais da Inteligência Artificial Clínica

A adoção de IA em saúde mental demanda atenção à privacidade, ao risco de vieses e à transparência dos modelos. Diretrizes internacionais [15, 24] enfatizam que sistemas de triagem devem priorizar interpretabilidade e governança algorítmica. A estrutura explicável implementada neste estudo, com logística transparente e rastreabilidade decisória, representa um passo em direção à IA clínica responsável e alinhada aos princípios bioéticos.

5.6 Síntese Final

Em síntese, este trabalho consolidou uma estrutura computacional interpretável e robusta para estimativa de risco de ansiedade infantil. O *Multilayer Perceptron*, com-

plementado pela representação latente aprendida pelo *Autoencoder*, demonstrou alta consistência entre métricas e capacidade de capturar nuances emocionais contínuas.

Embora integralmente conduzido em ambiente *in silico*, o modelo apresentou propriedades desejáveis para futura implementação em triagens clínicas, desde que validado em ambientes reais e contextos multimodais.

Ao unir rigor estatístico, coerência clínica e responsabilidade ética, este estudo contribui para o desenvolvimento de sistemas de apoio à decisão em saúde mental infantil, representando um avanço significativo na interseção entre Engenharia e Neurociência.

Referências

1 VIGO, D. et al. Global burden of mental disorders: The impact of anxiety and depression in children and adolescents. *The Lancet Psychiatry*, v. 7, n. 6, p. 490–499, 2022.

2 KOPOSOV, R.; MOLEN, M. G. G. T. A. van der; GROOT, P. J. F. de. Developmental trajectories of mental health problems in children and adolescents: a systematic review. *European Child & Adolescent Psychiatry*, v. 30, n. 10, p. 1505–1525, 2021.

3 GARCIA-CEJA, E.; OSMANI, V.; MAYORA, O. Machine learning for mental health: A review of current applications and future directions. *Frontiers in Computer Science*, v. 1, p. 1–18, 2018.

4 ARIF, M.; QAMAR, U.; MEHMOOD, R. Machine learning-based identification of generalized anxiety disorder using physiological signals. *IEEE Access*, v. 8, p. 136385–136395, 2020.

5 RIVERA, M. J. et al. Diagnosis and prognosis of mental disorders by means of eeg and deep learning: a systematic mapping study. *Artificial Intelligence Review*, Springer, v. 55, p. 1209–1251, 2022. Disponível em: <https://link.springer.com/article/10.1007/s10462-021-09986-y>.

6 CARPENTER, K. L. H. et al. Quantifying risk for anxiety disorders in preschool children: A machine learning approach. *PLOS ONE*, Public Library of Science, v. 11, n. 11, p. e0165524, 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0165524>.

7 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5. ed. Washington, DC: American Psychiatric Publishing, 2013. ISBN 9780890425558.

8 World Health Organization. *International Classification of Diseases (ICD-11)*. 11. ed. Geneva: World Health Organization, 2019. [Online]. Disponível em: <https://icd.who.int>. Acesso em: 17 jun. 2025.

9 LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, Curran Associates, Inc., v. 30, p. 4765–4774, 2017. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.

10 COSTELLO, E. J. et al. Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, v. 60, n. 8, p. 837–844, 2003.

11 LOPES, M. L. *Uso de Redes Neurais Artificiais na identificação de transtornos mentais comuns em crianças e adolescentes*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Pernambuco, 2020.

12 STADE, E. C. et al. A transdiagnostic, dimensional classification of anxiety shows improved parsimony and predictive noninferiority to dsm. *Journal of Psychopathology and Clinical Science*, v. 132, n. 8, p. 937–948, 2023. Disponível em: <https://doi.org/10.1037/abn0000863>.

13 BARLOW, D. H. *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. 2. ed. New York: The Guilford Press, 2002. ISBN 9781572304302.

14 CRASKE, M. G. et al. Anxiety disorders. *Nature Reviews Disease Primers*, v. 3, p. 1–18, 2017.

15 KAZDIN, A. E.; BLASE, S. L. Rethinking mental health care: Bridging the gap between research and clinical practice. *American Psychologist*, v. 71, n. 7, p. 590–601, 2016.

16 SMITH, A. e. a. Emotional and socio-cognitive processing in young children with symptoms of anxiety. *Journal of Child Psychology and Psychiatry*, v. 63, n. 10, p. 1123–1134, 2022.

17 XU, D.; SINGH, P.; KUMAR, R. Open data and cross-validation pipelines in computational psychiatry. *Nature Human Behaviour*, v. 7, p. 1458–1471, 2023. Disponível em: <https://doi.org/10.1038/s41562-023-01789-5>.

18 KHOEI, T. T.; SLIMANE, H. O.; KAABOUCH, N. Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, Springer, v. 35, p. 23103–23124, 2023. Disponível em: <https://link.springer.com/article/10.1007/s00521-023-08957-4>.

19 DICKSON, S. J. et al. Impact of psychotherapy for children and adolescents with anxiety disorders on global and domain-specific functioning: A systematic review and meta-analysis. *Clinical Child and Family Psychology Review*, v. 25, n. 4, p. 720–736, 2022. Disponível em: <https://doi.org/10.1007/s10567-022-00402-7>.

20 SHATTE, A. B. R.; HUTCHINSON, D. M.; TEAGUE, S. J. Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, Cambridge University Press, v. 49, n. 9, p. 1426–1448, 2019. Disponível em: <https://www.cambridge.org/core/journals/psychological-medicine/article/abs/machine-learning-in-mental-health-a-scoping-review-of-methods-and-applications/0B70B1C827B3A4604C1C01026049F7D9>.

21 GUYATT, G. H. et al. Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, BMJ Publishing Group Ltd, v. 336, n. 7650, p. 924–926, 2008. Disponível em: <https://doi.org/10.1136/bmj.39489.470347.AD>.

22 HENSE, J. B. et al. Prevalence of common mental disorders in adolescents: a systematic review with meta-analysis. *Jornal de Pediatria*, v. 99, n. 5, p. 453–462, 2023.

23 RACINE, N. et al. Global prevalence of depressive and anxiety symptoms in children and adolescents during covid-19: A meta-analysis. *JAMA Pediatrics*, v. 175, n. 11, p. 1142–1150, 2021.

24 Organização Mundial da Saúde. *Mental Health: Strengthening our response*. 2021. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>. Acesso em: 17 jun. 2025.

25 DOUGHERTY, L. R.; KLEIN, D. N. Dimensional approaches to developmental psychopathology: Research and clinical implications. *Journal of Child Psychology and Psychiatry*, v. 61, n. 3, p. 274–290, 2020.

26 ALLEN, J. L.; WAITE, P. Developmental pathways to anxiety in childhood: A review of the evidence. *Clinical Psychology Review*, v. 68, p. 84–96, 2019.

27 LUCA, A. D.; BIFFI, E. Systematic review of deep learning models for child and adolescent psychiatry. *Computers in Biology and Medicine*, v. 170, p. 107962, 2024. Disponível em: <https://doi.org/10.1016/j.compbiomed.2024.107962>.

28 WU, J.; LI, M.; ZHAO, K. Multimodal deep learning for anxiety detection combining eeg and facial features. *IEEE Transactions on Affective Computing*, v. 14, n. 6, p. 1120–1132, 2023.

29 GHAHARI, S.; GOLDAR, M. R. H.; NOORI, S. M. R. S. A novel approach for adhd detection using convolutional neural network on eeg signals. *Biomedical Signal Processing and Control*, v. 80, p. 104332, 2023.

30 KOENIG, J. et al. Heart rate variability in depressed and anxious children and adolescents: a meta-analysis. *Journal of Affective Disorders*, v. 292, p. 597–608, 2021.

31 LEDOUX, J. E. Rethinking the emotional brain. *Neuron*, Elsevier, v. 73, n. 4, p. 653–676, 2012.

32 ZHANG, X. et al. A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. *Journal of Neural Engineering*, v. 18, n. 3, p. 031001, 2021. Disponível em: <https://iopscience.iop.org/article/10.1088/1741-2552/abdfc0>.

33 GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org>.

34 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

35 BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1798–1828, 2013.

36 DRYSDALE, A. T. et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, v. 23, n. 1, p. 28–38, 2017. Disponível em: <https://doi.org/10.1038/nm.4246>.

37 TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. Human–computer collaboration for skin cancer classification. *Nature Medicine*, v. 26, n. 8, p. 1229–1234, 2020. Disponível em: <https://doi.org/10.1038/s41591-020-0942-0>.

38 DELL'ACQUA, C. et al. Heart rate variability in children and adolescents with anxiety disorders: a systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, v. 157, p. 105527, 2024.

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

40 RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Disponível em: <https://doi.org/10.1038/323533a0>.

41 HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989. Disponível em: <https://doi.org/10.1016/0893-6080(89)90020-8>.

42 STRAUSS, E.; JúNIOR, M. V. B.; FERREIRA, W. L. L. *A importância de utilizar métricas adequadas de avaliação de performance em modelos preditivos de machine learning*. 2022. Universidade Federal do Rio de Janeiro (UFRJ). Acesso em: jul. 2025. Disponível em: <https://pdfs.semanticscholar.org/78fa/b9dcbbefa1361fef4a7eb0b849d523aae2aa.pdf>.

43 SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*. São Paulo: Artliber, 2010.

44 KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015. Disponível em: <https://arxiv.org/abs/1412.6980>.

45 HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. *Science*, American Association for the Advancement of Science, v. 313, n. 5786, p. 504–507, 2006. Disponível em: <https://doi.org/10.1126/science.1127647>.

46 BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013. Disponível em: <https://doi.org/10.1109/TPAMI.2013.50>.

47 MIOTTO, R. et al. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, Nature Publishing Group, v. 6, p. 26094, 2016. Disponível em: <https://doi.org/10.1038/srep26094>.

48 KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014. Disponível em: <https://arxiv.org/abs/1312.6114>.

49 FREUND, Y.; MASON, L. The alternating decision tree learning algorithm. In: *Proceedings of the Sixteenth International Conference on Machine Learning (ICML)*. Morgan Kaufmann, 1999. p. 124–133. Disponível em: <https://www.cs.princeton.edu/~schapire/flip/ADT.pdf>.

50 NOGARE, D. *Performance de Machine Learning: Matriz de Confusão*. 2020. Disponível em: <https://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/>.

51 SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, JMLR.org, v. 15, n. 1, p. 1929–1958, 2014. Disponível em: <https://jmlr.org/papers/v15/srivastava14a.html>.

52 POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37–63, 2011. Disponível em: <https://www.bioinf.jku.at/publications/older/2604.pdf>.

53 RASCHKA, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, 2018. Disponível em: <https://arxiv.org/abs/1811.12808>.

54 JOHNSON, A.; KHOSHGOFTAAR, T. M. Survey on deep learning with class imbalance. *Journal of Big Data*, v. 9, n. 1, p. 1–56, 2022. Disponível em: <https://doi.org/10.1186/s40537-022-00608-7>.

55 Reddit. *A visual representation of the difference between accuracy and precision*. 2019. Disponível em: <https://www.reddit.com/r/aimlab/comments/eho7gk/a_visual_representation_of_the_difference_between/>.

56 NAY, N. *Mastering Evaluation Metrics: F1 Score, Z-Score, ROC Curve and Precision-Recall Curve*. 2020. Disponível em: <https://medium.com/@nay1228/mastering-evaluation-metrics-f1-score-z-score-roc-curve-and-precision-recall-curve-cb175de25724>

57 FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.

58 Evidently AI. *ROC Curve Explained*. 2023. Disponível em: <https://www.evidentlyai.com/classification-metrics/explain-roc-curve>.

59 CARPENTER, K. 2016. Disponível em: <https://doi.org/10.7910/DVN/N42LWG>.

60 WILKINSON, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, Nature Publishing Group, v. 3, n. 1, p. 1–9, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>.

61 PINE, D. S. et al. Challenges in developing novel treatments for childhood disorders: Lessons from research on anxiety. *Neuropsychopharmacology*, v. 44, n. 1, p. 211–222, 2019. Disponível em: <https://doi.org/10.1038/s41386-018-0171-1>.

62 ALKURDI, A. et al. Resilience of machine learning models in anxiety detection: Assessing the impact of gaussian noise on wearable sensors. *Applied Sciences*, v. 15, n. 1, p. 88, 2025.

63 LI, Q. et al. Eeg-based anxiety emotion classification using an optimized convolutional neural network and transformer. *Signal, Image and Video Processing*, v. 19, p. 501, 2025. Disponível em: <https://doi.org/10.1007/s11760-025-04067-x>.

64 LEE, E. W. et al. Comparing machine and deep learning models for pediatric anxiety classification using structured ehrs and area-based measures of health data. *medRxiv preprint*, 2025. Preprint – not peer reviewed.

65 WU, J.; LI, M.; ZHAO, K. Multimodal deep learning for anxiety detection combining eeg and facial features. *IEEE Transactions on Affective Computing*, v. 14, n. 6, p. 1120–1132, 2023. Disponível em: <https://doi.org/10.1109/TAFFC.2023.3267890>.