

Sistema Orientado por Inteligência Artificial para Auxílio ao Pré-diagnóstico de Crianças com Manifestações Comportamentais do TEA

Artificial Intelligence Oriented System for Assisting in the Pre-diagnosis of Children with Behavioral Manifestations of ASD

Aline Coelho Lauriano { aline.luriano@fatec.sp.gov.br }
Amanda Nogueira de Castro e Silva { amanda.silva226@fatec.sp.gov.br }
Adeldivo Alves de Sousa Junior { adeldivo.sousa@fatec.sp.gov.br }
Frederico Barbosa Muniz { frederico.muniz@fatec.sp.gov.br }
Thissiany Beatriz Almeida { thissiany.almeida@fatec.sp.gov.br }

RESUMO

O Transtorno do Espectro Autista (TEA) é um distúrbio de neurodesenvolvimento cujos sinais podem ser identificados a partir dos 18 meses. No entanto, o diagnóstico precoce é desafiado pela falta de profissionais especializados, especialmente em áreas rurais. Este estudo tem como objetivo desenvolver uma aplicação móvel, baseada em Inteligência Artificial (IA), para auxiliar profissionais de saúde no pré-diagnóstico de TEA em crianças de 0 a 2 anos na região do Vale do Ribeira. Foi implementado um modelo de rede neural do tipo *MultiLayer Perceptron* (MLP) para analisar dados de triagem coletados por meio do protocolo Q-CHAT-10. A aplicação foi desenvolvida com *Kotlin* para a interface, *textit.java* com *Spring Boot* para a API e *Python (FastAPI)* para a interação com o modelo de IA utilizando *Keras* e *TensorFlow*. A base de dados contou com 1.054 instâncias, e o modelo foi treinado com validação cruzada *k-fold* de 10 dobras, obtendo acurácia superior a 90%. Testes adicionais, com 54 novas instâncias, resultaram em uma precisão de 90,7% e sensibilidade de 92,6%. Conclui-se que a aplicação desenvolvida tem potencial para melhorar o pré-diagnóstico de TEA, especialmente em regiões de difícil acesso, ampliando assim a acessibilidade no processo de identificação precoce.

PALAVRAS-CHAVE: Transtorno do Espectro Autista; Inteligência Artificial; Pré-diagnóstico; Neurodesenvolvimento; Redes Neurais.

ABSTRACT

The Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder whose signs can be identified from 18 months of age. However, early diagnosis is challenged by the lack of specialized professionals, especially in rural areas. This study aims to develop a mobile application, based on Artificial Intelligence (AI), to assist healthcare professionals in the pre-diagnosis of ASD in children aged 0 to 2 years in the Vale do Ribeira region. A *MultiLayer Perceptron* (MLP) neural network model was implemented to analyze screening data collected through the Q-CHAT-10 protocol. The application was developed with *Kotlin* for the interface, *Java* with *Spring Boot* for the API, and *Python (FastAPI)* for interaction with the AI model using *Keras* and *TensorFlow*. The database comprised 1,054 instances, and the model was trained using 10-fold cross-validation, achieving an accuracy of over 90%. Additional tests, with 54 new instances, resulted in 90.7% accuracy and 92.6% sensitivity. It is concluded that the developed application has the potential to improve ASD pre-diagnosis, especially in hard-to-reach regions, thus enhancing accessibility in the early identification process.

KEYWORDS: Autism Spectrum Disorder; Artificial Intelligence; Early Diagnosis; Neurodevelopment; Neural Networks.

INTRODUÇÃO

O Transtorno do Espectro Autista (TEA) é caracterizado pelo Manual Diagnóstico e Estatístico de Transtornos Mentais - Quinta Edição (APA, 2014) como um Transtorno do Neurodesenvolvimento, cujos sinais se apresentam por meio de “prejuízos persistentes na interação social, comunicação e pela presença de padrões restritos e repetitivos de comportamentos, interesses ou atividades”. Embora o

diagnóstico possa ser realizado a partir dos 18 meses de idade (CDC, 2024), a média de idade do diagnóstico ainda é superior a 4 anos (Robins et al., 2019), com atrasos maiores para crianças pretas, de áreas rurais e baixa renda (Guthrie, 2019). Este cenário está fortemente vinculado à Agenda 2030 da Organização das Nações Unidas (ONU), que promove um acordo global para o desenvolvimento sustentável. O terceiro Objetivo de Desenvolvimento Sustentável (ODS) dessa agenda é "saúde e bem-estar", e busca garantir o acesso universal a serviços de saúde, contemplando assim ações e iniciativas voltadas a crianças autistas.

Nos Estados Unidos, a prevalência de TEA foi estimada em 2,8% em 2020, com um caso identificado a cada 36 crianças de 8 anos (CDC, 2023). Embora o Brasil não disponha de dados oficiais sobre a prevalência de TEA, ao aplicar essa taxa à população brasileira de 203 milhões (IBGE, 2022), estima-se que cerca de 5,6 milhões de brasileiros possam apresentar o transtorno. Este estudo concentra-se na região do Vale do Ribeira, no estado de São Paulo, composta por 22 municípios. Especificamente, em três cidades da região — Miracatu, Registro e Sete Barras —, aplicando a prevalência em dados do programa "Vale do Futuro", pode-se supor que aproximadamente 175 meninos e 167 meninas entre 0 e 9 anos podem ter TEA. Estudos apontam que meninos têm quatro vezes mais chances de serem diagnosticados com TEA em comparação com meninas (ICD-11, 2024; CDC, 2023), que tendem a exibir menos comportamentos restritivos e repetitivos (ICD-11, 2024).

O processo de diagnóstico do TEA, geralmente conduzido por equipes multidisciplinares, pode durar até 18 meses, devido ao volume elevado de solicitações, à escassez de profissionais qualificados, à baixa percepção da acurácia dos diagnósticos e à sobreposição de transtornos relacionados. Além disso, a falta de tempo para revisão dos resultados e a discussão sobre tratamentos com os responsáveis também contribuem para o atraso (Megerian et al., 2022). Ferramentas tradicionais de rastreamento muitas vezes não conseguem identificar todos os casos de TEA. Em um estudo com mais de 20 mil crianças, 61% das diagnosticadas com TEA apresentaram resultados negativos em rastreadores comuns (Guthrie, 2019). Muitos pediatras relatam não utilizar ferramentas formais de triagem, citando falta de familiaridade, necessidade de encaminhamento a especialistas e falta de tempo (Robins et al., 2019).

Além disso, os obstáculos logísticos e geográficos para acessar profissionais qualificados afetam significativamente o diagnóstico em populações vulneráveis, levando a atrasos que resultam na perda de oportunidades para intervenções precoces (Megerian et al., 2022). No entanto, o uso de sistemas eletrônicos de rastreamento tem se mostrado uma solução viável para melhorar o diagnóstico precoce do TEA nos cuidados primários (Guthrie, 2019).

A pesquisa de campo realizada nas três cidades citadas confirmou esse cenário, com questionários aplicados a 8 pais e responsáveis por crianças diagnosticadas com TEA, além de 11 profissionais de saúde envolvidos na triagem e/ou intervenção. A coleta de dados revelou que seis das crianças diagnosticadas eram meninos e duas eram meninas; duas delas tinham idade igual ou inferior a 4 anos, enquanto seis tinham mais de 4 anos. Observou-se que mais de 60% das crianças foram diagnosticadas após os 4 anos de idade, sendo que 37% consultaram mais de cinco médicos durante o processo. Mesmo assim, 88% dos pais relataram não haver conflitos de diagnóstico, com exceção de um caso (12%), em que o TEA foi confundido com Transtorno de Déficit de Atenção e Hiperatividade (TDAH). Mais de 50% das crianças receberam o diagnóstico conclusivo após 6 meses ou mais. As principais razões para essa demora foram a falta de especialistas na região (38%), a necessidade de consultar múltiplos especialistas (25%) e a ausência de sinais físicos perceptíveis do TEA, sendo os sinais comportamentais predominantes (25%).

A análise dos dados coletados junto aos pais indicou que estereótipias foram identificadas por 75% deles como o principal sintoma de alerta, enquanto distúrbios de comunicação e fala foram

mencionados por 63%, sendo uma preocupação importante. Além disso, 50% dos pais notaram distúrbios de atenção, como a falta de concentração, como um fator relevante. A persistência dos pais em buscar um diagnóstico é evidenciada pelo fato de muitos terem consultado mais de cinco profissionais, o que revela a complexidade do processo diagnóstico do TEA. Este cenário corrobora a necessidade de otimizar o processo de diagnóstico, já que os atrasos reduzem a janela de oportunidade para intervenções precoces, que são essenciais para o desenvolvimento da criança.

Por sua vez, a entrevista com profissionais de saúde revelou que 55% dos entrevistados não possuem especialização em TEA, mas todos já participaram de diagnósticos relacionados. A maioria relatou que o neuropediatra é o profissional mais qualificado para fornecer o diagnóstico definitivo (55%). O tempo médio para obter o diagnóstico variou de menos de 1 mês a 1-2 meses, embora possa se estender por mais de 6 meses devido à falta de especialistas (82%). Nenhum dos profissionais utilizava *softwares* auxiliares para o diagnóstico, utilizando apenas gravações de celular, e 7 deles relataram o uso de protocolos de triagem (como M-CHAT) e/ou intervenções (ABA). Entre os sinais comportamentais mais evidentes identificados pelos profissionais estavam estereotípias e distúrbios de comunicação/fala. Disfunções sensoriais, como seletividade alimentar, também foram mencionadas por 64% dos entrevistados, reforçando a consistência entre as percepções dos pais e dos profissionais de saúde. Embora não existam exames específicos para o TEA, exames complementares são usados para descartar outros transtornos, com 36% dos profissionais destacando a importância de exames fonoaudiológicos, como a Audiometria e o Processamento Auditivo Central (PAC).

Dada a complexidade do diagnóstico do TEA e a importância de uma abordagem multidisciplinar, o presente estudo propõe o desenvolvimento de uma aplicação móvel voltada para auxiliar profissionais de saúde no pré-diagnóstico de TEA em crianças na região do Vale do Ribeira. As Redes Neurais apresentam-se como ferramentas computacionais eficazes para prever resultados com base em grandes volumes de dados (Haykin, 2009). No projeto proposto, foi utilizado um modelo de rede neural do tipo *MultiLayer Perceptron (MLP)*, onde as conexões entre os neurônios seguem uma única direção, da camada de entrada até a camada de saída. Este modelo é composto por múltiplas camadas intermediárias que aplicam funções de ativação para extrair características relevantes dos dados, enquanto a camada de saída realiza a previsão final, gerando o resultado do pré-diagnóstico de TEA.

OBJETIVO

Desenvolver uma aplicação móvel baseada em inteligência artificial para auxiliar profissionais de saúde na identificação de sinais precoces do TEA em crianças de 0 a 2 anos, visando aumentar a eficiência e rapidez no pré-diagnóstico na região do Vale do Ribeira.

OBJETIVOS ESPECÍFICOS

1. Implementar um modelo de rede neural para análise de dados de triagem de TEA em crianças de 0 a 2 anos;
2. Integrar o modelo de IA em uma aplicação móvel acessível para os profissionais da saúde;
3. Realizar uma pesquisa de campo local para o levantamento de dados.

ESTADO DA ARTE

Em estudos atuais sobre o TEA, o uso de tecnologias para auxiliar no diagnóstico tornou-se essencial.

O primeiro estudo, conduzido por Megerian et al. (2022), avaliou um *Software* como Dispositivo Médico (*SaMD*) que usa IA para diagnóstico de TEA. A análise foi feita a partir de três fontes: questionário para cuidadores, vídeos curtos das crianças e um questionário para profissionais de saúde. O algoritmo utilizado foi uma árvore de decisão com gradiente reforçado, treinado com dados de crianças com TEA e outras condições neurodesenvolvimentais, além de crianças sem distúrbio. O estudo comparou os resultados do dispositivo com a concordância diagnóstica de dois ou mais especialistas. Os resultados possíveis eram “positivo para TEA”, “negativo para TEA” e “indeterminado”. Entre os participantes, 31,8% foram classificados como positivo ou negativo para TEA, enquanto 91% dos casos indeterminados apresentaram ao menos outro distúrbio neurodesenvolvimental, confirmado pelos especialistas. O tempo médio de preenchimento foi de 5 minutos para o questionário do cuidador, 11 minutos para revisão dos vídeos e 10 minutos para o questionário do profissional de saúde.

O estudo de Robins et al. (2019), autora do M-CHAT, propõe o uso de uma rede neural artificial *feed-forward* (*fANN*) para automatizar a pontuação do M-CHAT e aprimorar a triagem do TEA. Utilizando técnicas de aprendizado de máquina, como treinamento, validação cruzada (CV) e testes, o sistema atenua falhas humanas na pontuação e no acompanhamento das respostas. Os dados analisados envolveram mais de 14 mil crianças de 16 a 30 meses, divididas em subgrupos: 46,51% meninos, 44,8% meninas, com 50,68% brancos e 20,30% negros. A escolaridade materna média foi de 15 anos, destacando a correlação entre baixa escolaridade e dificuldades no diagnóstico. Os melhores resultados alcançaram 99% de precisão em todos os subgrupos. O estudo conclui que a *fANN* pode não apenas aprimorar a triagem do M-CHAT, mas também adaptar questões para subgrupos específicos. Limitações incluem possíveis correlações entre fatores sociodemográficos e o desempenho da ferramenta.

O terceiro estudo, realizado por Küpper et al. (2020), aplicou técnicas de *ML* ao questionário *Autism Diagnostic Observation Schedule* (*ADOS*) para analisar características comportamentais sócio-comunicativas e estereotipadas (repetitivas) presentes no Módulo 4, voltado para adolescentes e adultos com comunicação verbal típica. A pesquisa envolveu uma amostra composta por 673 adolescentes, 385 adultos com TEA que realizam atividades de rotina e 288 indivíduos com suspeita de TEA, mas que possuem outros diagnósticos psiquiátricos bem definidos. O procedimento diagnóstico incluiu a observação dos comportamentos, entrevistas com familiares presentes e a utilização de ferramentas padronizadas, permitindo que clínicos experientes chegassem a um diagnóstico mais preciso. A abordagem de *ML*, validada com uma técnica de validação cruzada 5x10, alcançou 83% de especificidade e 71% de sensibilidade. Os testes do modelo de *ML* foram realizados com um conjunto de dados com recursos reduzidos que não haviam sido usados no treinamento inicial do modelo, garantindo maior robustez nos resultados.

O método do presente estudo apresenta similaridades com os estudos de Megerian, ao sugerir questionários com base na faixa etária da criança para maximizar a precisão diagnóstica. Assim como no estudo de Robins, foram utilizadas técnicas de *ML*, como treinamento, validação cruzada e testes, para otimizar a pontuação automatizada do Q-CHAT-10, que é uma versão simplificada do M-CHAT utilizado por Robins. Além disso, à semelhança do estudo de Robins, que considerou fatores socioeconômicos, este estudo também incorporou questões demográficas pertinentes ao diagnóstico de TEA.

METODOLOGIA

No contexto do diagnóstico de TEA, os protocolos de triagem são ferramentas aplicadas por médicos para identificar sinais do TEA nos pacientes. Estes geralmente possuem um formato de

questionário com perguntas e respostas estruturadas. Os protocolos de intervenção são utilizados no processo de acompanhamento do paciente após o diagnóstico.

Neste estudo, optou-se pelo uso do protocolo de triagem para atender o objetivo de atuar no auxílio do processo de diagnóstico. Especificamente, foi abordado o Q-CHAT-10, indicado para a faixa etária de 0 a 2 anos, pela disponibilidade de uma base pública com registros de respostas da triagem (Thabtah, F, 2018) e pela similaridade com o M-CHAT, frequentemente citado na pesquisa de campo realizada. O questionário Q-CHAT-10 é composto por 10 perguntas que buscam identificar traços comportamentais associados ao TEA (Quadro 1).

Quadro 1 – Questionário Q-CHAT-10

Variável no Conjunto de Dados	Perguntas Correspondentes do Q-CHAT-10
A1	Seu filho olha para você quando você chama o nome dele/dela?
A2	Quão fácil é para você obter contato visual com seu filho?
A3	Seu filho aponta para indicar que quer algo? (ex.: um brinquedo que está fora de alcance)
A4	Seu filho aponta para compartilhar interesse com você? (ex.: apontando para algo interessante)
A5	Seu filho faz de conta? (ex.: cuidar de bonecas, falar em um telefone de brinquedo)
A6	Seu filho segue onde você está olhando?
A7	Se você ou outra pessoa da família estiver visivelmente chateada, seu filho mostra sinais de querer confortá-los? (ex.: acariciando o cabelo, abraçando)
A8	Como você descreveria as primeiras palavras do seu filho?
A9	Seu filho usa gestos simples? (ex.: acenar adeus)
A10	Seu filho olha para o nada sem um propósito aparente?

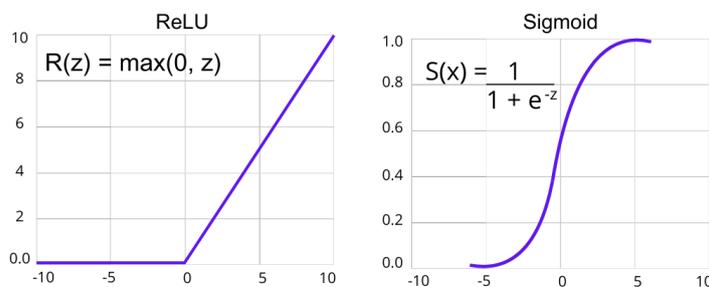
Fonte: (Thabtah, F, 2018); Adaptado pelos autores (2024)

As ferramentas escolhidas para o desenvolvimento deste projeto foram selecionadas para garantir eficiência e eficácia. A interface do aplicativo móvel foi desenvolvida utilizando a linguagem de programação *Kotlin*, especialmente adequada para criar aplicativos *Android*, proporcionando uma experiência fluida para o usuário. Já para a parte lógica do sistema, foi empregado *Java Spring Boot*, uma ferramenta robusta que permite a criação de *Application Programming Interfaces (APIs)* de maneira eficiente, facilitando a comunicação entre o aplicativo e o servidor. Além disso, o processamento da rede neural foi realizado utilizando o *FastAPI*, um *framework* leve em *Python* que facilita na execução de tarefas complexas, como o processamento de dados para análises e previsões. Por fim, a base de dados dos registros dos cadastros e protocolos de triagem foi armazenada no *MongoDB Atlas*, uma plataforma *NoSQL* que organiza os dados em documentos *JavaScript Object Notation (JSON)*, um formato leve e de fácil leitura. Essa estrutura oferece maior flexibilidade no armazenamento e gerenciamento de grandes volumes de informações, como respostas e registros da triagem, além de garantir alta escalabilidade.

Nas camadas intermediárias da rede, foi utilizada a função de ativação *Rectified Linear Unit (ReLU)*, amplamente adotada por sua simplicidade e eficácia em lidar com grandes volumes de dados. A *ReLU* ativa apenas valores positivos, o que ajuda a resolver o problema do gradiente desaparecido em redes profundas, acelerando o processo de aprendizado (Goodfellow; Bengio; Courville, 2016).

Para a camada de saída, foi escolhida a função *Sigmoid*, que comprime os valores entre 0 e 1, facilitando a interpretação probabilística dos resultados. Essa função é ideal para tarefas de classificação binária, como o pré-diagnóstico de TEA, onde as respostas do Q-CHAT-10 foram convertidas em variáveis binárias.

Figura 1 – Funções



Fonte: Autoria Própria (2024)

O conjunto de dados utilizado para este estudo é denominado "*Autistic Spectrum Disorder Screening Data for Toddlers*", disponível na plataforma *Kaggle*, amplamente conhecida pelo compartilhamento de bases de dados para uso livre voltadas para pesquisas. Com 1054 registros (também chamados de instâncias) e 18 atributos (características do base de dados), esse conjunto fornece uma combinação de indicadores comportamentais do questionário Q-Chat-10, além de atributos demográficos e de saúde (Quadro 2). Para viabilizar a aplicação em redes neurais, foi necessário transformar os atributos categóricos em valores numéricos, uma vez que as redes neurais interpretam apenas dados numéricos. Para isso, utilizou-se o *LabelEncoder* e o *OneHotEncoder*. O *LabelEncoder* converte variáveis categóricas em valores inteiros únicos, atribuindo um número a cada categoria distinta. Enquanto o *OneHotEncoder* expande essa codificação ao representar cada categoria como uma coluna binária separada, permitindo que o modelo capture a independência entre as categorias sem estabelecer uma ordem entre elas. Este processo é especialmente útil para atributos como "Sexo" ou "Histórico familiar de TEA", que não possuem uma hierarquia numérica. Além disso, a idade em meses foi normalizada utilizando o *StandardScaler*, que ajusta os dados para que apresentem média zero e desvio padrão igual a um. Esse procedimento garante que as variáveis numéricas estejam na mesma escala, prevenindo que valores maiores dominem o aprendizado da rede neural e, assim, melhorando a eficiência do modelo. Alguns atributos também foram excluídos para evitar vieses ou porque não apresentavam valor preditivo relevante. A coluna "Etnia" foi omitida para evitar o risco de enviesamento demográfico, uma vez que as categorias de etnia utilizadas não correspondem aos padrões brasileiros. Já as colunas "Responsável pelo teste" e "Motivo da triagem" foram removidas por possuírem uma natureza administrativa e por não influenciarem diretamente a análise dos traços do TEA.

Quadro 2 – Características coletadas e suas descrições

Característica	Tipo	Descrição
Perguntas A1 a A10: respostas correspondentes a cada pergunta do Q-CHAT-10	Binário (0, 1)	Código das respostas para cada pergunta com base no método de triagem utilizado
Idade	Number	Crianças pequenas (meses)
Pontuação pelo Q-chat-10	Number	A pontuação varia de 0 a 10 (1-10)
Sexo	Character	Masculino ou Feminino
Etnia	String	Lista de etnias comuns em formato de texto
Nascido com Icterícia	Boolean (yes or no)	Se a criança nasceu com Icterícia (coloração amarelada da pele, mucosas e olhos)
Histórico familiar de TEA	Boolean (yes or no)	Se algum membro imediato da família tem TEA
Responsável pelo teste	String	Pai, mãe, cuidador, equipe médica, clínico, etc.
Motivo da triagem	String	Entrada de texto do usuário
Variável de classe	String	Traços de TEA ou Sem traços de TEA

Fonte: (Thabtah, F, 2018); Adaptado pelos autores (2024)

O treinamento da rede foi otimizado usando o algoritmo *Adam* (*Adaptive Moment Estimation*), que combina as vantagens dos métodos *AdaGrad* e *RMSProp* (Goodfellow; Bengio; Courville, 2016). O *Adam* adapta as taxas de aprendizado para cada parâmetro com base em momentos de primeira e segunda ordem (média e variância), permitindo uma convergência mais rápida. A atualização dos pesos na rede neural é realizada pela seguinte equação:

$$\begin{aligned}
 m &= \beta_1 \cdot m + (1 - \beta_1) \cdot \Delta W \\
 v &= \beta_2 \cdot v + (1 - \beta_2) \cdot (\Delta W)^2 \\
 W &= W - \frac{\alpha \cdot m}{\sqrt{v + \epsilon}}
 \end{aligned}
 \tag{1}$$

Para o processo de treinamento, utilizou-se a técnica de Validação Cruzada *K-Fold*, onde o conjunto de dados foi dividido em *k* subconjuntos, garantindo que o modelo fosse treinado e testado *k* vezes, utilizando diferentes subconjuntos a cada vez. Isso aumenta a robustez do modelo, evitando o *overfitting* — situação em que o modelo memoriza padrões específicos do conjunto de treino, comprometendo sua capacidade de generalização para novos dados — e, assim, gerando resultados mais confiáveis (Goodfellow; Bengio; Courville, 2016).

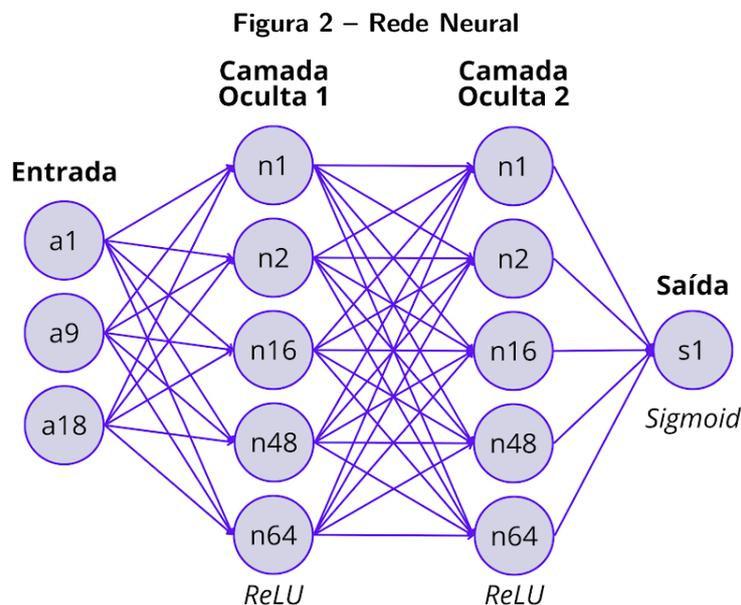
Após o treinamento completo da rede neural, além do arquivo *.pkl*, que armazena os parâmetros do processo de aprendizado, conexões entre os neurônios e as taxas de aprendizado ajustadas durante o treinamento, também é gerado um arquivo com a extensão *.keras*, contendo a arquitetura da rede, os pesos ajustados e as configurações do modelo, permitindo que o modelo seja reutilizado para fazer previsões em dados novos sem a necessidade de um novo treinamento.

A extensão *.pkl* é criada utilizando o módulo *pickle* do *Python*, que serializa os objetos, ou seja, converte-os em uma sequência de *bytes* para serem armazenados e recuperados posteriormente (PYTHON, 2024). Já o arquivo *.keras* é gerado com o auxílio da biblioteca *Keras*, que é uma interface de alto nível para a construção e treinamento de redes neurais, integrada ao *TensorFlow*,

uma plataforma de aprendizado de máquina desenvolvida pelo *Google* (KERAS, 2024). Com isso, é possível carregar e replicar o comportamento do modelo em futuras execuções de maneira eficiente.

A avaliação de desempenho da rede neural foi feita utilizando a função de perda *Cross Entropy*, que mede a discrepância entre as probabilidades preditas pelo modelo e os rótulos verdadeiros, ajustando os pesos do modelo para melhorar a precisão da classificação. Além disso, a performance da rede neural foi validada por meio de uma matriz de confusão, que detalha a distribuição dos erros e acertos em diferentes categorias de predição. Esse método é amplamente utilizado em aprendizado de máquina para avaliar o desempenho de classificadores, fornecendo uma visão clara sobre os tipos de erros cometidos e a eficácia do modelo em diferentes classes (Bishop, 2006).

A rede neural *MLP* (Figura 2) foi configurada com 18 neurônios na camada de entrada, duas camadas ocultas com 64 neurônios cada e ativação *ReLU*, e uma camada de saída com 1 neurônio e ativação *Sigmoid* para a classificação binária. O treinamento foi realizado em diferentes épocas, com tamanho de lote (*batch size*) de 32, utilizando a função de perda binária *Cross Entropy* e o otimizador *Adam*. A divisão dos dados foi feita por validação cruzada *k-fold* com 10 dobras, onde 900 instâncias (90% da base) foram usadas para o treinamento e 100 instâncias para os testes.



- Função de perda: Entropia Cruzada Binária
- Otimizador: Adam
- Tamanho do lote (*batch size*): 32
- Validação cruzada *k-fold* (10 dobras)
- 900 instâncias para treinamento (90%) e 100 para teste (10%)

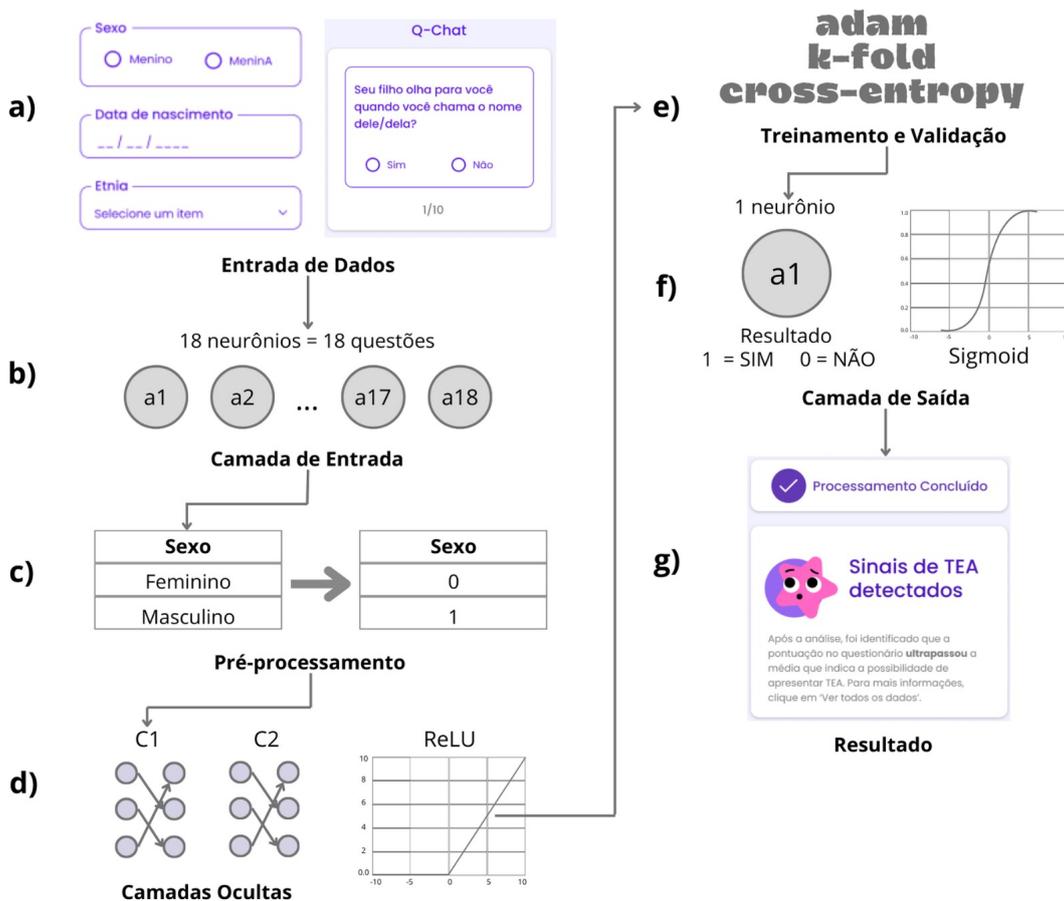
Fonte: Autoria Própria (2024)

Na prática, o funcionamento da IA no aplicativo SOFIA segue o fluxo descrito a seguir, conforme ilustrado na Figura 3:

- a) **Entrada de Dados:** O processo inicia com a coleta de informações demográficas do paciente, como sexo, data de nascimento e etnia. Essas informações são essenciais para estabelecer o perfil inicial e calcular a idade exata em meses da criança.
- b) **Camada de Entrada:** A camada de entrada da rede neural contém 18 neurônios, cada um correspondente a uma das questões do questionário Q-CHAT-10, refletindo as respostas dadas pelo usuário.

- c) Pré-processamento: Antes de enviar os dados para a rede neural, as informações demográficas, como o sexo, são convertidas pelo *LabelEncoder* e pelo *OneHotEncoder* para um formato numérico (0 para Feminino e 1 para Masculino).
- d) Camadas Ocultas: O modelo conta com camadas ocultas intermediárias que utilizam a função de ativação *ReLU*. Essas camadas ocultas processam as informações, aplicando transformações que tornam os dados mais adequados para a classificação final.
- e) Treinamento e Validação: A rede neural foi treinada e validada utilizando algoritmos *Adam* e validação cruzada *k-fold*, além de empregar a *cross-entropy* como função de perda, o que contribui para a precisão e a confiabilidade dos resultados.
- f) Camada de Saída: A última camada ou camada de saída, gera a probabilidade de a criança apresentar sinais de TEA. Ela utiliza a função *Sigmoid*, que converte os resultados em uma escala binária (Sim/Não), conforme a probabilidade obtida.
- g) Resultado: Após o processamento completo, o aplicativo exibe o resultado ao usuário, indicando a presença ou ausência de sinais de TEA com base nas respostas fornecidas e na análise da rede neural.

Figura 3 – Funcionamento da IA

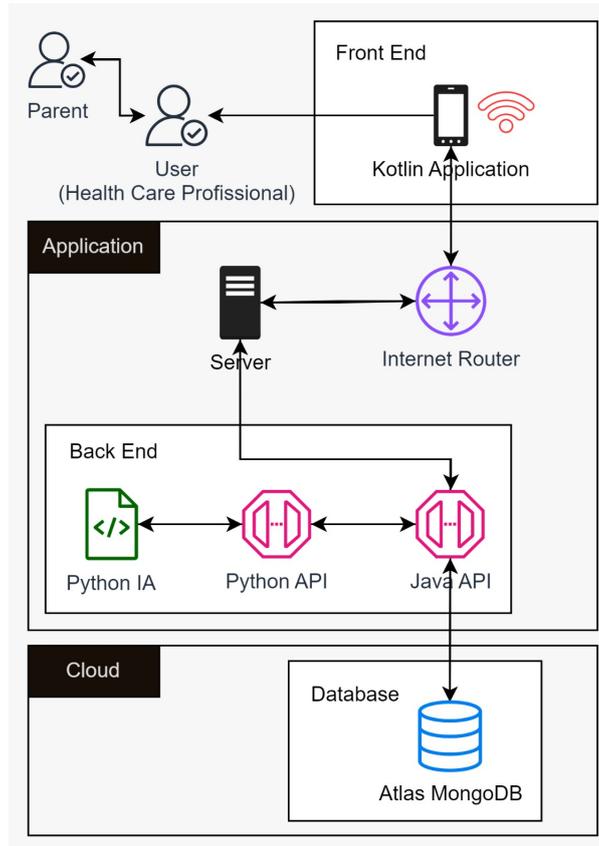


Fonte: Autoria Própria (2024)

O sistema de estrutura de rede do aplicativo opera utilizando uma arquitetura distribuída composta por *front-end*, *back-end* e armazenamento *NoSQL* em nuvem. O *front-end* é representado pelo aplicativo móvel desenvolvido em *Kotlin*, onde o profissional de saúde insere os dados demográficos do paciente. Essas informações são enviadas para o *back-end*, hospedado em uma instância *EC2* da

AWS, que gerencia as *APIs* responsáveis pelo processamento das informações. A *API* em *Java Spring Boot* realiza o armazenamento inicial dos dados no banco de dados *MongoDB*, que está hospedado na plataforma em nuvem *MongoDB Atlas*. Após o preenchimento do protocolo Q-CHAT-10, as respostas são encaminhadas da *API* em *Java* para a *API* em *Python*, desenvolvida com *FastAPI*, onde a rede neural pré-treinada calcula a probabilidade de sinais de TEA. O resultado é então enviado de volta para a *API* em *Java*, que atualiza o banco de dados no *MongoDB Atlas* e exibe o pré-diagnóstico na interface do aplicativo.

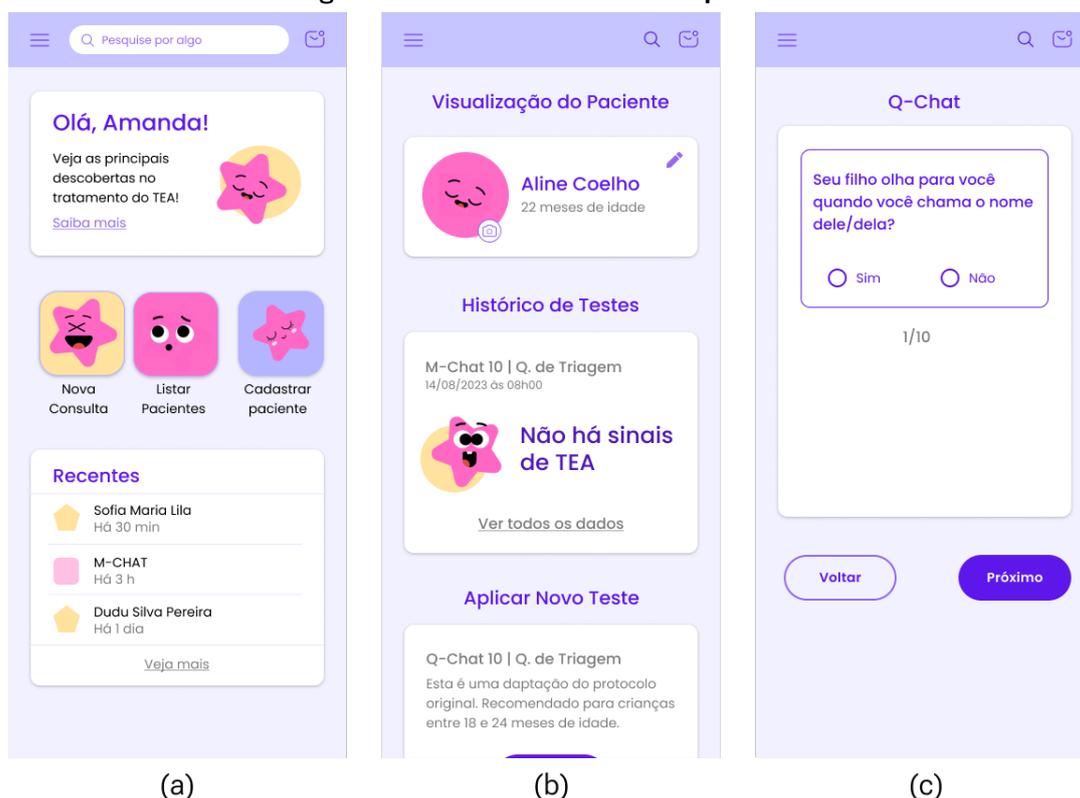
Figura 4 – Diagrama de metodologia



Fonte: Autoria Própria (2024)

O protótipo foi desenvolvido com foco na simplicidade e intuitividade para o usuário. A interface do aplicativo apresenta três telas principais: Na Tela Inicial (Figura 5a), o profissional de saúde é recebido com um cartão de boas-vindas, um registro das atividades recentes e três funções principais: agendar consulta, cadastrar pacientes e listar pacientes. Na Tela de Visualização Individual (Figura 5b) é possível acessar informações detalhadas sobre o paciente, incluindo histórico e resultados de testes anteriores e acesso para aplicação de novos protocolos de triagem. Cada questão do protocolo de triagem será mostrada de forma individual na Tela da Aplicação (Figura 5c). Todas as respostas são armazenadas no sistema e podem ser revisadas pelo profissional de saúde a qualquer momento para complementar o diagnóstico final.

Figura 5 – Interfaces iniciais do aplicativo



Fonte: Autoria Própria (2024)

RESULTADOS PRELIMINARES

Os resultados deste estudo foram obtidos a partir de testes com a aplicação e a rede neural, realizados em ambiente controlado. A análise inclui métricas de desempenho da aplicação, abordando variações no número de épocas e instâncias para avaliar o comportamento da solução em diferentes cenários. Além disso, foi realizada uma avaliação do modelo de rede neural quanto à sua capacidade de predição e de aprendizado contínuo.

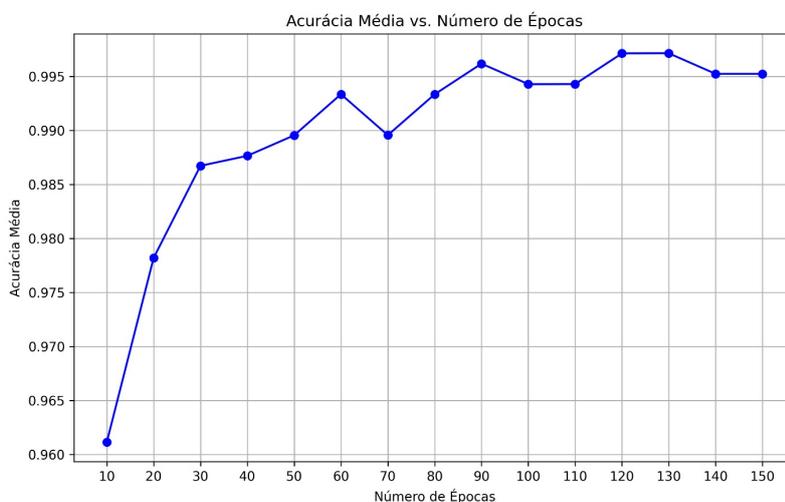
As principais conclusões obtidas desses testes serão discutidas a seguir, com destaque no impacto da aplicação e nas implicações do uso de redes neurais no processo de diagnóstico.

REDE NEURAL ARTIFICIAL

A análise dos testes com a rede neural demonstrou um padrão consistente no processo de aprendizado do modelo. Como mostrado no Gráfico 1, a acurácia cresce de forma progressiva à medida que o número de épocas aumenta, atingindo mais de 90% de precisão, com o maior desempenho observado na execução com 90 épocas. Esse aumento contínuo reflete o aprimoramento dos parâmetros do modelo a cada nova iteração, permitindo que ele extraia informações mais detalhadas dos dados.

Contudo, o Gráfico 1 também mostra um ponto de exaustão, onde o aumento no número de épocas deixa de proporcionar ganhos significativos na acurácia. Esse comportamento indica que o modelo alcançou seu limite de aprendizado com o conjunto de dados disponível. A partir desse estágio, prolongar o treinamento pode resultar em *overfitting*, comprometendo a qualidade das previsões.

Gráfico 1 – Acurácia média em função do número de épocas

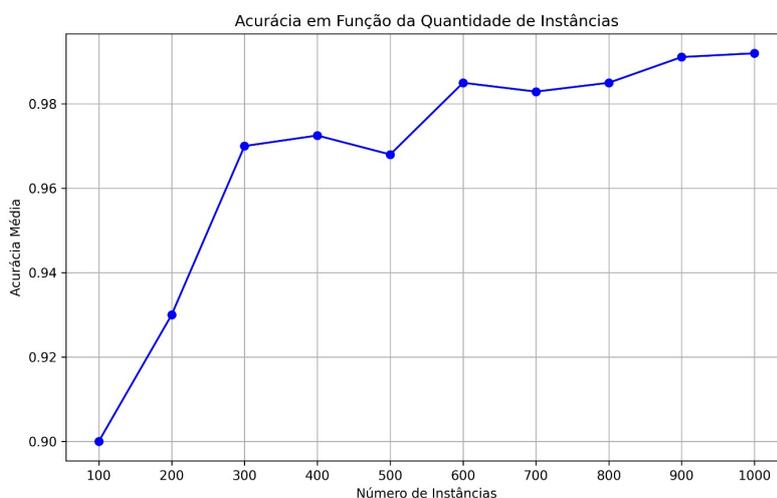


Fonte: Autoria Própria (2024)

Além disso, os testes realizados com menos instâncias revelaram que a acurácia média do modelo ficou abaixo de 90% quando havia menos de 300 exemplos disponíveis, como mostrado no Gráfico 2. Esse comportamento reforça a necessidade de um conjunto de dados robusto e diversificado para o treinamento eficaz da rede neural, já que um volume menor de informações não é suficiente para o modelo aprender de maneira eficiente.

À medida que o número de instâncias aumenta, o desempenho do modelo melhora consideravelmente, alcançando uma acurácia média de aproximadamente 98%. No entanto, após esse ponto, o gráfico indica uma estabilização, onde o aumento no número de instâncias não melhora significativamente a acurácia. Esse comportamento sugere que, após atingir uma quantidade mínima de dados, adicionar mais instâncias resulta em apenas pequenas melhorias no desempenho do modelo.

Gráfico 2 – Acurácia média em função da quantidade de Instâncias



Fonte: Autoria Própria (2024)

Esses resultados destacam a importância de um volume adequado de dados no treinamento,

garantindo que o modelo aprenda padrões significativos e gere previsões precisas. A eficácia do modelo *MLP* é evidente, especialmente ao combinar a função de ativação *ReLU* nas camadas intermediárias e *Sigmoid* na camada de saída. Isso permite que o modelo classifique os dados do protocolo de triagem de maneira eficiente, gerando um pré-diagnóstico baseado nos padrões comportamentais extraídos das instâncias analisadas.

AVALIAÇÃO DE USABILIDADE COM GRUPOS DE CONTROLE E TEA

Para avaliar a eficácia e a usabilidade da aplicação, foram realizados testes com 54 instâncias da base de dados pública, que não foram utilizadas no treinamento da rede neural. Destas, 27 instâncias correspondem a crianças sem o distúrbio (grupo de controle) e 27 a crianças diagnosticadas com TEA. O conjunto de teste foi dividido de forma equitativa para garantir que as variáveis estivessem equilibradas e que os resultados fossem representativos para ambos os grupos.

Os resultados do teste foram organizados em uma matriz de confusão Tabela 1 para avaliar o desempenho do modelo. As previsões foram classificadas nas seguintes categorias:

1. *True Positive* ou Verdadeiro Positivo (TP): Refere-se aos casos em que o modelo classificou corretamente crianças diagnosticadas com TEA. O teste obteve 25 verdadeiros positivos.
2. *False Negative* ou Falso Negativo (FN): São os casos em que o modelo não conseguiu identificar corretamente crianças com TEA, classificando-as equivocadamente como "sem TEA". Foram registrados 2 falsos negativos.
3. *False Positive* ou Falso Positivo (FP): Casos em que o modelo classificou incorretamente crianças sem o distúrbio como possuidoras de TEA. Houve 3 falsos positivos.
4. *True Negative* ou Verdadeiro Negativo (TN): Representa as previsões corretas do modelo para crianças sem o distúrbio, classificando-as como "sem TEA". O teste identificou 24 verdadeiros negativos.

Tabela 1 – Matriz de Confusão

Matriz de Confusão		
	Predição: TEA	Predição: Sem TEA
Diagnosticado com TEA	25 (TP)	2 (FN)
Diagnosticado sem TEA	3 (FP)	24 (TN)

Fonte: Autoria Própria (2024)

Esses dados proporcionam uma visão detalhada do desempenho do modelo, destacando os acertos e erros cometidos pela rede neural. A precisão do modelo foi calculada com base na proporção de previsões corretas (TP e TN) em relação ao total de previsões, resultando em aproximadamente 90,7% — Equação (2). Isso indica que o modelo acertou em 90,7% dos casos ao classificar corretamente as crianças com ou sem TEA.

Além disso, a sensibilidade (ou taxa de TP) foi de 92,6% — Equação (3), evidenciando a capacidade do modelo de identificar corretamente a maioria das crianças com TEA. Esses números confirmam a eficácia do modelo, que se mostrou capaz de classificar adequadamente tanto os casos positivos quanto os negativos.

$$\text{Precisão} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{25 + 24}{25 + 24 + 3 + 2} = \frac{49}{54} \approx 0.907 \quad (2)$$

$$\text{Sensibilidade} = \frac{TP}{TP + FN} = \frac{25}{25 + 2} = \frac{25}{27} \approx 0.926 \quad (3)$$

Por fim, os resultados obtidos demonstram que o modelo de rede neural teve um bom desempenho na identificação de crianças com TEA. No entanto, há espaço para aprimoramentos, como a expansão do conjunto de dados, a redução dos erros de classificação e a adoção de classificações mais detalhadas. Essas melhorias podem aumentar a precisão e a confiabilidade da ferramenta, ampliando seu potencial de aplicação em ambientes clínicos.

CONCLUSÃO

Este projeto desenvolveu uma ferramenta móvel inovadora utilizando inteligência artificial para o pré-diagnóstico do TEA em crianças. A ferramenta se destaca por agilizar a triagem, auxiliar os profissionais de saúde na identificação precoce de sinais de TEA e tornar o diagnóstico precoce mais acessível e eficiente para populações que frequentemente enfrentam dificuldades com escassez de especialistas.

A integração de redes neurais, treinadas com dados de triagem validados, permitiu alcançar uma acurácia superior a 90%. Isso reforça a capacidade do sistema de fornecer pré-diagnósticos de alta precisão, auxiliando profissionais de saúde nas decisões clínicas. A pesquisa de campo realizada nas cidades do Vale do Ribeira forneceu uma visão clara dos desafios locais na detecção do TEA, especialmente de infraestrutura e falta de especialistas, contribuindo para o design da aplicação de modo a atender às necessidades específicas da região.

Focada na usabilidade, a ferramenta foi projetada para ser intuitiva, garantindo que mesmo profissionais sem treinamento técnico avançado possam utilizá-la facilmente. Contudo, há limitações, como a necessidade de validação em cenários clínicos reais e com um conjunto maior de amostras para assegurar sua aplicabilidade em larga escala. Além disso, a utilização de *hardware* mais avançado, como GPUs, poderia também ampliar o desempenho da aplicação, permitindo análises mais rápidas e eficientes.

Até o presente momento, os testes preliminares com as instâncias do Q-CHAT-10 do banco de dados público ofereceram uma base sólida para validações e implementações futuras. A próxima etapa envolve testes práticos em condições reais de uso sob supervisão de profissionais da saúde para uma avaliação completa em ambientes clínicos e a avaliação do impacto no processo de diagnóstico atual, verificando se a aplicação promove o diagnóstico mais rápido e, conseqüentemente, em idade mais precoce.

Em suma, este projeto não só facilita o pré-diagnóstico do TEA, como também serve como um modelo para incorporar tecnologias emergentes na saúde, representando um avanço importante para o diagnóstico precoce de transtornos neurodesenvolvimentais e potencializando melhorias no acesso a diagnósticos eficientes. A mobilidade da aplicação para celular permite que a solução seja expandida para outras regiões, e a virtualização do processo de preenchimento manual de protocolos promove a redução do tempo necessário para a triagem.

REFERÊNCIAS

APA. **Manual Diagnóstico e Estatístico de Transtornos Mentais DSM-5**. 5ª. Porto Alegre: Artmed, 2014. Acesso em: 12 de setembro de 2023.

BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. [S. l.]: Springer New York, 2006. <https://link.springer.com/book/9780387310732>. ISBN 978-0-387-31073-2.

CDC. **Key Findings from the ADDM Network: A Snapshot of Autism Spectrum Disorder in 2020**. [S. l.: s. n.], 2023. Acesso em: 11 de setembro de 2023.

Disponível em:

<https://www.cdc.gov/ncbddd/autism/addm-community-report/key-findings.html>

CDC. **Screening and Diagnosis of Autism Spectrum Disorder**. [S. l.: s. n.], 2024. Disponível em: <https://www.cdc.gov/autism/diagnosis/index.html>. Acesso em: 29 de setembro de 2024.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S. l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

GUTHRIE, Whitney; WALLIS, Kate; BENNETT, Amanda et al. Accuracy of autism screening in a large pediatric network. **Pediatrics**, v. 144, n. 4, 2019.

HAYKIN, S.S. **Neural Networks and Learning Machines**. [S. l.]: Pearson, 2009. (Pearson International Edition). ISBN 9780131293762.

Disponível em: <https://books.google.com.br/books?id=KCwW0AAACAAJ>

IBGE. **Panorama do Censo de 2022**. [S. l.: s. n.], 2022. Acesso em: 18 de setembro de 2023.

Disponível em: <https://censo2022.ibge.gov.br/panorama/>

ICD-11. **ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS)**. [S. l.: s. n.], 2024.

Acesso em: 21 de setembro de 2023.

Disponível em: <https://icd.who.int/browse11/1-m/en>

KERAS. **About Keras**. [S. l.: s. n.], 2024. Acesso em: 26 de setembro de 2024.

Disponível em: <https://keras.io/about/>

KÜPPER, Charlotte et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. **Scientific reports**, Nature Publishing Group UK London, v. 10, n. 1, p. 4805, 2020.

MEGERIAN, Jonathan T et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. **NPJ digital medicine**, Nature Publishing Group UK London, v. 5, n. 1, p. 57, 2022.

PYTHON. **Pickle** — **Python object serialization**. [S. l.: s. n.], 2024. Acesso em: 19 de setembro de 2024.

Disponível em: <https://docs.python.org/3/library/pickle.html>

ROBINS, Diana L. et al. A machine learning strategy for autism screening in toddlers. **Journal of Developmental and Behavioral Pediatrics: JDBP**, NIH Public Access, v. 40, n. 5, p. 369, 2019.

THABTAH, F. **Autism Screening for Toddlers**. [S. l.: s. n.], 2018. Acesso em: 12 de junho de 2024.

Disponível em: <https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers?resource=download>