



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS PICOS
TECNÓLOGO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

LETÍCIA BEATRIZ DE HOLANDA VIEIRA

**GUARDIAN A.N.Y: DETECÇÃO ALGORÍTMICA E CLASSIFICAÇÃO DE
HIPERESTÍMULOS VISUAIS EM VÍDEOS DO YOUTUBE PARA SUPORTE À
ACESSIBILIDADE SENSORIAL NO TRANSTORNO DO ESPECTRO AUTISTA (TEA)**

PICOS, PIAUÍ
2026

LETÍCIA BEATRIZ DE HOLANDA VIEIRA

GUARDIAN A.N.Y: DETECÇÃO ALGORÍTMICA E CLASSIFICAÇÃO DE
HIPERESTÍMULOS VISUAIS EM VÍDEOS DO YOUTUBE PARA SUPORTE À
ACESSIBILIDADE SENSORIAL NO TRANSTORNO DO ESPECTRO AUTISTA (TEA)

Trabalho de Conclusão de Curso (Artigo Científico) apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Picos.

Orientador: Prof. Dr. Rogério Figueredo de Sousa

PICOS, PIAUÍ
2026

GUARDIAN A.N.Y: DETECÇÃO ALGORÍTMICA E CLASSIFICAÇÃO DE HIPERESTÍMULOS VISUAIS EM VÍDEOS DO YOUTUBE PARA SUPORTE À ACESSIBILIDADE SENSORIAL NO TRANSTORNO DO ESPECTRO AUTISTA (TEA)

Letícia Beatriz de Holanda Vieira¹

Rogério Figueredo de Sousa²

RESUMO

O consumo de mídias virtuais por crianças tem crescido significativamente, destacando-se o YouTube como uma das principais plataformas de acesso. No entanto, essa exposição traz desafios de acessibilidade para crianças com Transtorno do Espectro Autista (TEA), uma vez que hiperestímulos visuais, como cortes rápidos e saturação, podem desencadear episódios de sobrecarga sensorial. Embora a plataforma forneça recursos como legendas e leitores de tela, há uma lacuna no controle de estímulos sensoriais. Este trabalho apresenta o desenvolvimento e a avaliação técnica do Guardian A.N.Y (Autista No YouTube), um sistema que atua como filtro automatizado categorizando conteúdos infantis segundo seu nível de estímulo visual. A partir de um conjunto de dados categorizado manualmente pela autora para estabelecer a referência de análise, o sistema realiza o processamento de vídeo, extrai métricas dos quadros e aplica um algoritmo de *Random Forest Classifier* com estratégia de validação cruzada estratificada. Os testes técnicos demonstraram a capacidade do modelo em diferenciar vídeos de baixo, moderado e alto estímulo com acurácia média de 78,46%, sugerindo a viabilidade de ferramentas assistivas focadas no alerta preventivo de conteúdo.

Palavras-chave: Tecnologia Assistiva; Processamento de Vídeo; Transtorno do Espectro Autista; Sobrecarga Sensorial; *Random Forest*.

ABSTRACT

The consumption of digital media by children has grown significantly, with YouTube standing out as one of the main platforms for access. However, this exposure presents accessibility challenges for children with Autism Spectrum Disorder (ASD), as visual overstimulation, such as rapid cuts and high saturation, can trigger episodes of sensory overload. Although the platform provides features such as captions and screen readers, there remains a gap in the control of sensory stimuli. This study presents the development and technical evaluation of Guardian A.N.Y (Autista No YouTube), a system that operates as an automated filter, categorizing children's content according to its level of visual stimulation. Based on a dataset manually categorized by the author as a reference for analysis, the system processes videos, extracts frame-level metrics, and applies a Random Forest Classifier algorithm with a stratified cross-validation strategy. The technical tests demonstrated the model's ability to distinguish between videos with low, moderate, and high levels of visual stimulation, reaching a mean accuracy of 78.46%, suggesting the feasibility of assistive tools focused on the preventive alerting of potentially overstimulating content.

¹ <http://lattes.cnpq.br/3360523954086833>. IFPI. capic.2024118tads0035@aluno.ifpi.edu.br.

² <http://lattes.cnpq.br/9346694618502913>. IFPI. rogerio.sousa@ifpi.edu.br.

Keywords: Assistive Technology; Video Processing; Autism Spectrum Disorder; Sensory Overload; Random Forest.

1 INTRODUÇÃO

Observa-se que, nas últimas décadas, as tecnologias se tornaram parte de diversas esferas da humanidade. O acesso à internet tem crescido consideravelmente: cerca de 95% dos domicílios particulares permanentes do país possuem esse recurso (IBGE, 2026). Juntamente com esse maior alcance ao espaço virtual, a mídia tem sido mais utilizada. Os produtos midiáticos são parte intrínseca da vida diária de uma pessoa e desempenham papel relevante na sociedade contemporânea (VIANA; SILVA; OLIVEIRA, 2023). Logo, o público infantil não fica de fora dessa ocorrência: cada vez mais as crianças e adolescentes têm acessado o mundo digital, e plataformas como o YouTube, TikTok e Instagram têm tido maior destaque entre elas (NIC.BR, 2024).

Para crianças de até 12 anos, o YouTube tem se destacado como principal plataforma de acesso, a qual impõe desafios relevantes a diferentes setores da sociedade, como o aumento do uso de telas (NIC.BR, 2026). Outro problema relevante é a acessibilidade, como a falta de adaptações adequadas para o público com o Transtorno do Espectro Autista (TEA), condição do neurodesenvolvimento caracterizada por particularidades na comunicação social, no comportamento e no processamento sensorial.

Elementos como cores intensas, transições rápidas, variações bruscas de luminosidade e excesso de estímulos visuais podem provocar sobrecarga sensorial, o que pode dificultar a utilização da plataforma por usuários que possuem a condição (SENA; BARROS, 2023). De maneira análoga, o uso de tecnologia assistiva é necessária para a autonomia, independência, qualidade de vida e inclusão social de indivíduos com deficiência (FERNÁNDEZ-BATANERO *et al.*, 2022). Embora o YouTube já possua algumas ferramentas de acessibilidade, como legendas automáticas, suporte a leitores de tela e player de acessibilidade com controles simplificados, não há mecanismos específicos para adaptação sensorial (YOUTUBE, 2023).

Diante dessa lacuna tecnológica, este trabalho investiga o nível de estímulos visuais em desenhos infantis por meio da ferramenta Guardian A.N.Y (Autista No YouTube), sistema que atua como um filtro categorizador de mídias. A proposta baseia-se na extração de métricas visuais a partir de um conjunto de dados categorizado manualmente para estabelecer o padrão de referência, aplicando um algoritmo de *Random Forest Classifier* com estratégia de validação cruzada estratificada para a classificação automática do nível de estímulo.

Nesse sentido, o objetivo geral é avaliar a viabilidade técnica de um sistema baseado em Visão Computacional e aprendizado de máquina para classificar vídeos infantis em níveis baixo, moderado e alto de estímulo visual. Busca-se responder à seguinte questão: em que medida características computacionais extraídas dos quadros de

vídeo permitem reproduzir uma categorização manual do nível de estímulo visual?

2 TRABALHOS RELACIONADOS

Esta seção apresenta uma revisão da literatura dividida em quatro eixos centrais. Inicialmente, discutem-se as particularidades do Transtorno do Espectro Autista (TEA) e suas implicações no processamento sensorial. Em seguida, aborda-se a relação entre a sobrecarga sensorial em ambientes digitais e as limitações de acessibilidade em plataformas de vídeo. Por fim, analisam-se os aspectos visuais presentes nas mídias infantis e as técnicas de processamento de vídeo e aprendizado de máquina aplicadas à categorização automatizada desses conteúdos.

2.1 TRANSTORNO DO ESPECTRO AUTISTA E PROCESSAMENTO SENSORIAL

O autismo caracteriza-se como uma condição do neurodesenvolvimento que afeta a interação social, a comunicação e o comportamento. Essas manifestações ocorrem em diferentes graus de severidade e dificilmente da mesma forma em cada sujeito (GUEDES; TADA, 2015). Além dos desafios na comunicação, o Transtorno do Espectro Autista está diretamente associado a alterações no processamento sensorial.

Essas alterações no processamento geram padrões de respostas atípicas que impactam negativamente a interação social e a adaptação em múltiplos contextos do cotidiano (MATTOS, 2019). Diante disso, indivíduos com TEA podem apresentar hipersensibilidade sensorial, resultando em reações comportamentais e emocionais intensas diante da sobrecarga visual do ambiente. Sob esses estímulos excessivos, é comum a ocorrência de reações de esquiva, retraimento, isolamento, explosões emocionais e respostas físicas de proteção (DUNN, 1997).

Nesse contexto, a compreensão dessas particularidades sensoriais torna-se fundamental, especialmente diante de ambientes digitais saturados por elementos gráficos, trocas bruscas de iluminação e cores intensas, os quais frequentemente atuam como gatilhos para essa sobrecarga.

2.2 SOBRECARGA SENSORIAL EM AMBIENTES DIGITAIS E ACESSIBILIDADE EM PLATAFORMAS DE VÍDEO

Em conteúdos audiovisuais, elementos como cores vibrantes, alternâncias de luminosidade e mudanças rápidas de cena são amplamente utilizados para prender a atenção do espectador; no entanto, esses recursos visuais podem ser extremamente prejudiciais para pessoas com TEA. Devido ao excesso de estímulos intensos, a capacidade de processar informações de maneira organizada torna-se um desafio, o que pode comprometer a autorregulação emocional de crianças autistas. Essa

sobrecarga frequentemente desencadeia estados de irritabilidade e agitação, além de estar associada a dificuldades de concentração, redução do tempo de atenção e impactos negativos no desenvolvimento cognitivo e emocional (LLAMAS; VEGA; RAMOS, 2026).

Curiosamente, o público infantil no espectro costuma ter um contato prolongado com mídias digitais. Crianças com TEA passam cerca de 62% mais tempo assistindo à televisão e jogando videogames do que o tempo que passam envolvidas em todas as atividades fora das telas somadas. Essa preferência pelo ambiente virtual costuma ser motivada pela busca por maior previsibilidade e controle em relação ao mundo físico (MAZUREK; WENSTRUP, 2013). Contudo, quando esse consumo é direcionado a plataformas audiovisuais populares, a exposição a vídeos que não foram projetados com foco na regulação sensorial atua como um gatilho direto para episódios de desregulação.

Apesar dessa alta adesão ao espaço virtual, a acessibilidade digital em plataformas de vídeo, como o YouTube, permanece majoritariamente restrita a deficiências motoras, visuais e auditivas clássicas, oferecendo recursos como legendas automáticas, leitores de tela e navegação simplificada (YOUTUBE, 2023). Não há, até o momento, mecanismos nativos específicos voltados à adaptação ou modulação de estímulos visuais para o público neurodivergente.

Diante da ausência de suporte tecnológico voltado à regulação sensorial nas plataformas de streaming, evidencia-se a urgência de abordagens que ampliem a inclusão real no consumo de conteúdos digitais. Nesse cenário, a tecnologia assistiva surge como uma estratégia fundamental para proporcionar maior segurança, autonomia e qualidade de experiência para indivíduos com TEA no ambiente virtual (FERNÁNDEZ-BATANERO *et al.*, 2022).

2.3 ANÁLISE DE CONTEÚDO E ESTÍMULOS SENSORIAIS EM MÍDIAS INFANTIS

Os produtos midiáticos podem ser categorizados pelo impacto de seus elementos constituintes em animações de alto e baixo estímulo. Desenhos de alto estímulo (ou *fast-paced*) caracterizam-se por cortes e cenas rápidas, movimentação intensa de objetos, trocas de plano aceleradas e paletas de cores saturadas e vibrantes (LLAMAS; VEGA; RAMOS, 2026). Essa sobrecarga na codificação dos eventos reduz a manutenção da atenção voluntária e pode desencadear estados de hiperexcitabilidade, fadiga e prejuízos imediatos nas funções executivas (SANTOS; VALE, 2025). Em contrapartida, produções de baixo estímulo utilizam ritmo cadenciado, pausas adequadas e cores suaves, favorecendo a autorregulação e a atenção sustentada.

Para o público com Transtorno do Espectro Autista, essa dinâmica apresenta desdobramentos ainda mais críticos. Crianças autistas apresentam maior predisposição ao uso prolongado e contínuo de mídias baseadas em tela em comparação aos seus pares neuro típicos (MAZUREK; WENSTRUP, 2013). Contudo, devido a atipicidades no

processamento sensorial, como a redução da habituação a estímulos repetitivos e dificuldades no filtro atencional, a exposição a animações de alto impacto gera uma carga cognitiva desproporcional. Elementos como transições visuais aceleradas e trocas de cena frequentes sobrecarregam o sistema nervoso de indivíduos hiper-reativos, aumentando a probabilidade de estresse e *meltdowns* (LEAL-SILVA; ALVES; ALVES, 2025).

Diante desse cenário, evidencia-se a relevância de mapear quantitativamente as métricas que diferenciam o nível de impacto das produções infantis. A identificação objetiva de parâmetros visuais em desenhos animados torna-se, portanto, um passo fundamental para viabilizar ferramentas de tecnologia assistiva capazes de filtrar e adequar o conteúdo audiovisual às especificidades sensoriais do autismo.

2.4 PROCESSAMENTO DE VÍDEO E APRENDIZADO DE MÁQUINA

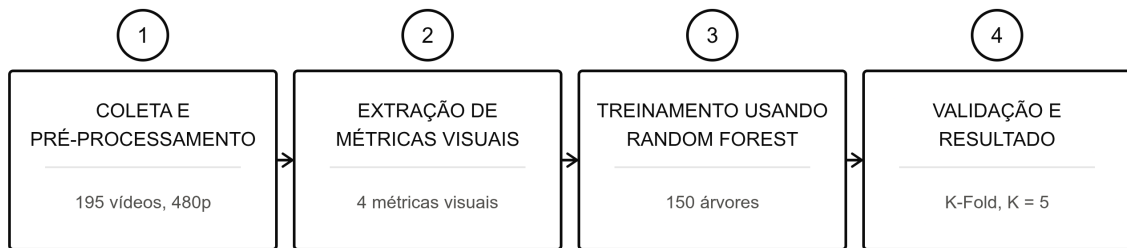
A análise automatizada de vídeos para a identificação de estímulos visuais fundamenta-se em técnicas de Visão Computacional e Aprendizado de Máquina (*Machine Learning*). O processo consiste em decompor o vídeo digital em quadros sequenciais (*frames*), viabilizando a mensuração quantitativa e temporal de propriedades visuais ao longo da exibição (SZELISKI, 2022).

Para quantificar o nível de estímulo de animações infantis, a literatura utiliza descritores voltados a aspectos como movimento, variação de cor e frequência de transições. Ferramentas de processamento de imagem permitem calcular a diferença visual entre quadros consecutivos para estimar a agitação física de uma cena (BRADSKI, 2000), bem como analisar o contraste cromático e a correlação de cores para identificar cortes rápidos de câmera (GARGI; KASTURI; ANTANI, 2000).

Na etapa de classificação, modelos supervisionados baseados em *ensembles* de árvores de decisão, como o *Random Forest*, destacam-se por sua capacidade de lidar com múltiplas variáveis preditoras e por reduzir o sobreajuste (*overfitting*) em relação a árvores isoladas (BREIMAN, 2001).

3 METODOLOGIA

Figura 1 – Fluxo de trabalho para construção do modelo de aprendizado de máquina.



Fonte: Elaborada pela autora (2026).

Conforme ilustrado na Figura 1, a metodologia deste trabalho é estruturada em um *pipeline* computacional composto por três etapas principais: amostragem e pré-processamento do sinal de vídeo (*streams*), extração de métricas visuais, e classificação supervisionada do nível de estímulos utilizando algoritmos de aprendizado de máquina com validação cruzada.

3.1 AMOSTRAGEM E PRÉ-PROCESSAMENTO DE VÍDEO

3.1.1 Coleta dos Vídeos e Construção do Dataset

A amostragem de dados consistiu na construção de um *dataset* próprio composto por 195 vídeos infantis, catalogados em um repositório estruturado em formato CSV contendo os identificadores, as URLs e os respectivos rótulos atribuídos. A coleta foi realizada entre os meses de maio e julho de 2026, diretamente na plataforma YouTube, por meio de termos de busca como "desenhos infantis", "desenhos animados", "desenhos antigos", "desenhos musicais" e variações semelhantes. Foram incluídos vídeos de classificação livre destinados a crianças de até 12 anos, contemplando animações em geral e conteúdos musicais infantis, com duração mínima de 2 minutos e 40 segundos, tempo necessário para a extração das métricas a partir do minuto inicial de corte, conforme detalhado na Subseção 3.1.2. O repositório inclui, propositalmente, diferentes episódios de uma mesma produção (como *Bolofofos*, *Galinha Pintadinha*, *Três Palavrinhas*, além de conteúdos musicais de artistas como Aline Barros e Xuxa), uma vez que episódios de um mesmo desenho podem apresentar níveis distintos de estímulo visual entre si.

A rotulagem manual foi realizada pela autora deste trabalho, que assistiu aos vídeos sem áudio no mesmo intervalo posteriormente processado pelo sistema, do minuto 01:00 até completar os 100 segundos de exibição (01:00 a 02:40), e classificou cada um em três níveis de estímulo: baixo, moderado ou hiperestimulante. O critério adotado considerou a paleta de cores, a quantidade de elementos em tela e a frequência de troca de cena: vídeos com cores suaves, ritmo lento e baixa variação de cena foram

classificados como baixo estímulo; vídeos com cores vivas, poucos elementos em tela e variação de cena moderada foram classificados como estímulo moderado; e vídeos com trocas de cena frequentes, alta densidade de informação visual e cores vivas foram classificados como hiperestimulantes. Os rótulos foram atribuídos exclusivamente por percepção visual da autora, de forma independente de qualquer valor computacional, garantindo que a rotulagem de referência não fosse enviesada pelas métricas posteriormente extraídas. O balanceamento exato entre as três classes (65 vídeos cada) não foi um resultado espontâneo da rotulagem, mas um critério de coleta: novos vídeos continuaram sendo buscados e classificados até que cada categoria atingisse 65 exemplares, momento em que a coleta daquela classe era encerrada.

3.1.2 Extração das Métricas e Preparação dos Dados de Entrada

Após a rotulagem manual, o link de cada vídeo e seu respectivo rótulo foram armazenados em formato CSV, servindo de entrada para o *pipeline* automatizado de extração de métricas. Para cada registro, utilizou-se a biblioteca *yt_dlp*³ para a extração direta da URL do *stream* em formato MP4, com limitação de resolução em 480p, visando otimizar o tempo de processamento e reduzir o consumo de recursos computacionais. Com o objetivo de descartar vinhetas de canais, anúncios ou introduções que não representam o conteúdo efetivo da animação, o processamento de cada vídeo salta automaticamente o primeiro minuto, iniciando a análise a partir de 01:00. A partir desse ponto, o algoritmo realiza a captura de amostras de quadros (*frames*) em intervalos fixos de 0,5 segundo até atingir a cota de 200 amostras por vídeo, equivalente a 100 segundos de mídia, redimensionados para a resolução padronizada de 320×240 pixels. Esse intervalo de amostragem foi definido em função das limitações de hardware disponíveis para o processamento, de modo a manter o *pipeline* computacionalmente leve e viável para a extração das métricas ao longo dos 195 vídeos do *dataset*.

Para cada quadro amostrado, foram calculadas quatro métricas visuais por meio da biblioteca *OpenCV*⁴ em *Python*⁵: (i) *similaridade visual*, obtida pela correlação entre histogramas RGB de quadros consecutivos; (ii) *frequência de cortes de cena*, calculada a partir da proporção de quadros cuja similaridade ficou abaixo do limiar de 0,70, normalizada por minuto de vídeo; (iii) *contraste cromático*, obtido a partir do desvio padrão dos canais *a* e *b* do espaço de cor CIELAB; e (iv) *movimento físico da cena*, calculado pela diferença absoluta entre quadros em escala de cinza consecutivos, quantificando o percentual de pixels alterados acima de um limiar de intensidade. Ao final do processamento de cada vídeo, os valores obtidos por quadro foram agregados por meio da média aritmética, gerando um vetor de quatro métricas resumidas por

³ Disponível em: <<https://github.com/yt-dlp/yt-dlp.git>>.

⁴ Versão 5.0.0.93 (pacote *opencv-python*).

⁵ Versão 3.12.13.

vídeo, associado ao respectivo rótulo manual.

Esse vetor de métricas médias constituiu a base de entrada para o modelo de classificação. Utilizou-se a biblioteca *scikit-learn*⁶ para a implementação de um classificador *Random Forest*, treinado e validado por meio de validação cruzada estratificada em 5 *folds* (*StratifiedKFold*), de forma a preservar a proporção original das três classes (baixo, moderado e hiperestimulante) em cada partição.

3.2 MÉTRICAS VISUAIS ANALISADAS

O sistema concentra-se na interpretação comportamental de quatro indicadores computacionais extraídos a partir da comparação de quadros consecutivos:

- **Similaridade de Cor:** Avalia o grau de continuidade cromática entre quadros vizinhos nos canais *Red*, *Green* e *Blue* (RGB). Quanto menor o valor de similaridade, maior é a ruptura visual entre as cenas.
- **Frequência de Cortes por Minuto:** Contabiliza as trocas bruscas de plano na animação. O sistema registra um corte de cena sempre que a similaridade de cor cai abaixo do limiar crítico de 0,70. Esse valor foi obtido empiricamente por meio de testes preliminares de calibração: diferentes limiares foram testados sobre uma amostra do *dataset* e seus resultados comparados em termos de taxa de acerto na identificação de cortes de cena. O valor de 0,70 apresentou o melhor desempenho entre os limiares avaliados, sendo então adotado como parâmetro final do sistema. Cabe registrar que essa calibração foi realizada sobre vídeos pertencentes ao próprio conjunto de 195 amostras utilizado posteriormente na avaliação do modelo, o que é discutido como limitação do trabalho na Seção 5.
- **Contraste Cromático:** Avalia a saturação e vivacidade das cores. Para evitar distorções do espaço RGB, a imagem é convertida para o espaço perceptual CIELAB, medindo-se a variação e intensidade das cores nos eixos cromáticos.
- **Agitação e Movimento Físico:** Mede a quantidade de movimento na tela por meio da diferença absoluta de intensidade dos pixels entre o quadro atual e o anterior, isolando alterações significativas em termos da porcentagem da área alterada.

A Tabela 1 ilustra, a partir de uma amostra de vídeos do *dataset*, os valores obtidos para cada uma dessas quatro métricas antes da etapa de treinamento.

⁶ Versão 1.6.1.

Tabela 1 – Exemplo dos dados de entrada (CSV) fornecidos ao classificador Random Forest.

Classe	Similaridade	Cortes	Contraste	Agitacao
Hiperestimulante	0,7824	25,2	39,59	44,40%
Estímulo Moderado	0,9053	15,0	25,99	25,77%
Baixo Estímulo	0,8955	5,4	17,17	15,24%

Fonte: Elaborada pela autora (2026).

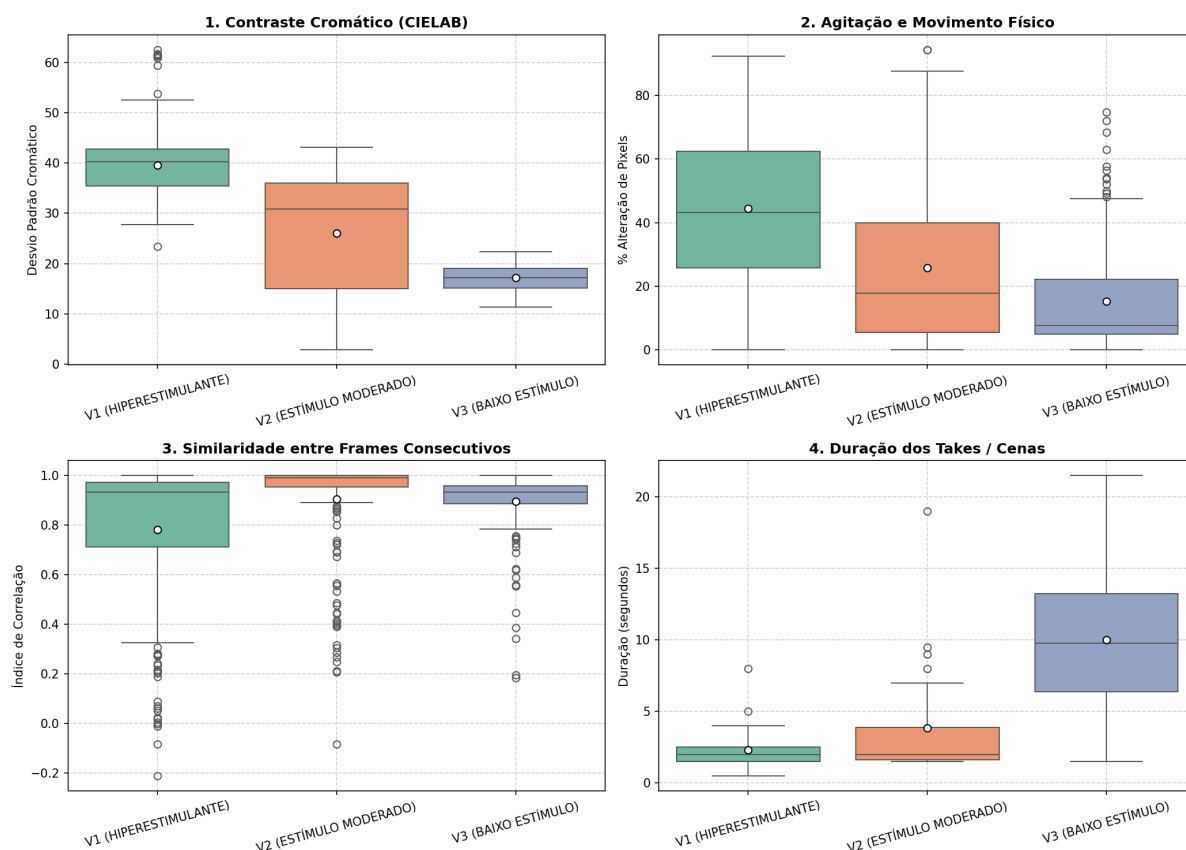
3.3 DEFINIÇÃO DAS CLASSES DE ESTÍMULO SENSORIAL

Com base nos padrões identificados pelas métricas visuais e na fundamentação sobre hipersensibilidade no público com TEA, os conteúdos do *dataset* foram categorizados em três classes de estímulo:

- **Baixo Estímulo:** Animações com ritmos contemplativos, paletas de cores suaves, pouca movimentação física e baixa frequência de cortes por minuto.
- **Estímulo Moderado:** Animações com ritmo narrativo intermediário, apresentando variação equilibrada de cores e edições com contraste cromático moderado.
- **Hiperestimulante:** Desenhos com edições aceleradas, movimentação constante na tela e cores primárias altamente saturadas.

Como observado no comparativo geral (Figura 2), as métricas apresentam divergências claras entre cada categoria.

Figura 2 – Comparativo das métricas visuais extraídas de três vídeos de exemplo, um representante de cada classe de estímulo: **V1 (Hiperestimulante)** ("*Casinha de Jesus*", Aline Barros, <https://youtu.be/L_Jmdp69T8Q>), **V2 (Estímulo Moderado)** ("*Pedro, Tiago, João no barquinho*", 3Palavrinhas, <<https://youtu.be/Tdwy3BZe61s>>) e **V3 (Baixo Estímulo)** ("*Os Três Porquinhos*", Disney Desenhos, <https://youtu.be/M_QpsigrVaM>).



Fonte: Elaborada pela autora (2026).

A Figura 2 apresenta, de forma exemplificativa, a distribuição em *boxplot* de quatro das métricas visuais extraídas quadro a quadro ao longo dos 100 segundos analisados desses três vídeos, um de cada classe de estímulo. Em cada gráfico, a linha central da caixa indica a mediana, os limites superior e inferior correspondem ao terceiro e ao primeiro quartil, o círculo branco assinala a média, as hastes representam a amplitude dos valores considerados normais e os pontos isolados são *outliers*. Os painéis 1 e 2 mostram que o contraste cromático e a agitação/movimento físico são consistentemente mais altos e dispersos no vídeo hiperestimulante (V1), diminuindo progressivamente em V2 e V3, o que reflete visualmente cenas mais coloridas e mais movimentadas. O painel 3 evidencia o padrão inverso na similaridade entre quadros consecutivos: em V1 os valores são mais baixos e espalhados, já que trocas de cena e movimentação constante reduzem a semelhança entre *frames* vizinhos, enquanto em V2 e, principalmente em V3, a similaridade se mantém alta e concentrada, indicando cenas mais estáveis. Por fim, o painel 4 mostra que a duração dos cenas é bem maior

e mais variável em V3, ao passo que V1 e V2 concentram cortes de cena mais curtos e frequentes. Vale ressaltar que esses três vídeos servem apenas como exemplo ilustrativo do comportamento das métricas em cada classe, já que o mesmo padrão é observado, de forma estatisticamente mais robusta, no conjunto completo dos 195 vídeos que compõem o *dataset*.

3.4 CLASSIFICAÇÃO E VALIDAÇÃO DO MODELO

As métricas calculadas (Similaridade, Cortes, Contraste e Agitação) compõem o conjunto de atributos de entrada do arquivo compilado. A etapa de inteligência do sistema utiliza o algoritmo *Random Forest Classifier*, escolhido por sua robustez e por combinar múltiplas árvores de decisão para chegar a um resultado mais confiável. O modelo foi configurado de forma a equilibrar a complexidade das árvores, evitando tanto a simplificação excessiva quanto o ajuste exagerado aos dados de treino, garantindo que nenhuma classe fosse favorecida indevidamente durante o aprendizado.

É importante ressaltar que, em nenhuma etapa do processo, houve participação de usuários com Transtorno do Espectro Autista ou coleta de dados pessoais de qualquer natureza. Os rótulos utilizados como referência de treinamento e validação do modelo foram atribuídos manualmente pela autora deste trabalho, a partir da análise visual dos vídeos, dispensando qualquer envolvimento direto de participantes humanos como sujeitos de pesquisa.

A validação da capacidade de generalização do modelo foi executada de forma inteiramente computacional, utilizando o método de Validação Cruzada Estratificada (*Stratified K-Fold*, com $K = 5$). Esse procedimento garante a preservação da proporção exata dos rótulos em cada partição de treino e teste, permitindo consolidar os resultados de acurácia, precisão, *recall* e *F1-score* de forma estatisticamente confiável.

4 RESULTADOS

Esta seção apresenta a avaliação técnica do Guardian A.N.Y, com foco na validação do algoritmo de classificação quanto à capacidade de reproduzir a categorização manual do nível de estímulos visuais em vídeos infantis. Para tanto, submeteu-se o modelo *Random Forest Classifier* a um procedimento de validação cruzada estratificada utilizando um conjunto de dados composto por 195 amostras balanceadas, distribuídas igualmente com 65 vídeos para cada uma das classes estabelecidas: Baixo Estímulo, Estímulo Moderado e Hiperestimulante.

4.1 VALIDAÇÃO CRUZADA E ESTABILIDADE DO MODELO

A aplicação do método de validação cruzada estratificada em cinco rodadas permitiu aferir a capacidade de generalização do classificador frente a diferentes partições dos dados de treinamento e teste, preservando a proporção original das classes em cada *fold*. Os índices de acurácia obtidos em cada iteração encontram-se dispostos na Tabela 2.

Tabela 2 – Resultados de acurácia por rodada de validação cruzada.

Rodada de Validação (<i>Fold</i>)	Acurácia (%)
Rodada 1	69,23%
Rodada 2	79,49%
Rodada 3	84,62%
Rodada 4	76,92%
Rodada 5	82,05%
Acurácia Média	78,46%
Desvio Padrão	5,28%

Fonte: Elaborada pela autora (2026).

Ao término das cinco iterações, o modelo atingiu uma acurácia média de 78,46%. Observou-se um desvio padrão de 5,28%, o que indica uma variação moderada entre as partições avaliadas. Esse resultado, isoladamente, não permite descartar sobreajuste, mas sugere que o desempenho do modelo não dependeu de uma única divisão dos dados, característica que é favorecida pela natureza de *ensemble* do *Random Forest*, ao combinar múltiplas árvores de decisão treinadas sobre subconjuntos distintos dos dados.

4.2 DESEMPENHO POR CATEGORIA SENSORIAL

Para além da acurácia global, analisou-se o desempenho distinto do algoritmo em relação a cada nível de estímulo visual. A Tabela 3 detalha os resultados acumulados considerando as métricas de precisão, revocação (*recall*) e *F1-Score*.

Tabela 3 – Relatório de classificação acumulado por classe sensorial.

Categoria Sensorial	Precisão	Revocação (<i>Recall</i>)	F1-Score	Amostras (<i>Support</i>)
Baixo Estímulo	0,88	0,88	0,88	65
Estímulo Moderado	0,70	0,66	0,68	65
Hiperestimulante	0,77	0,82	0,79	65
Acurácia		0,78		195
Média Ponderada (<i>Weighted Avg</i>)	0,78	0,78	0,78	195

Fonte: Elaborada pela autora (2026).

Para além dos índices agregados de precisão, revocação e *F1-Score*, a matriz de confusão acumulada das cinco rodadas de validação cruzada (Figura 3) permite

identificar para quais classes os erros de classificação são direcionados.

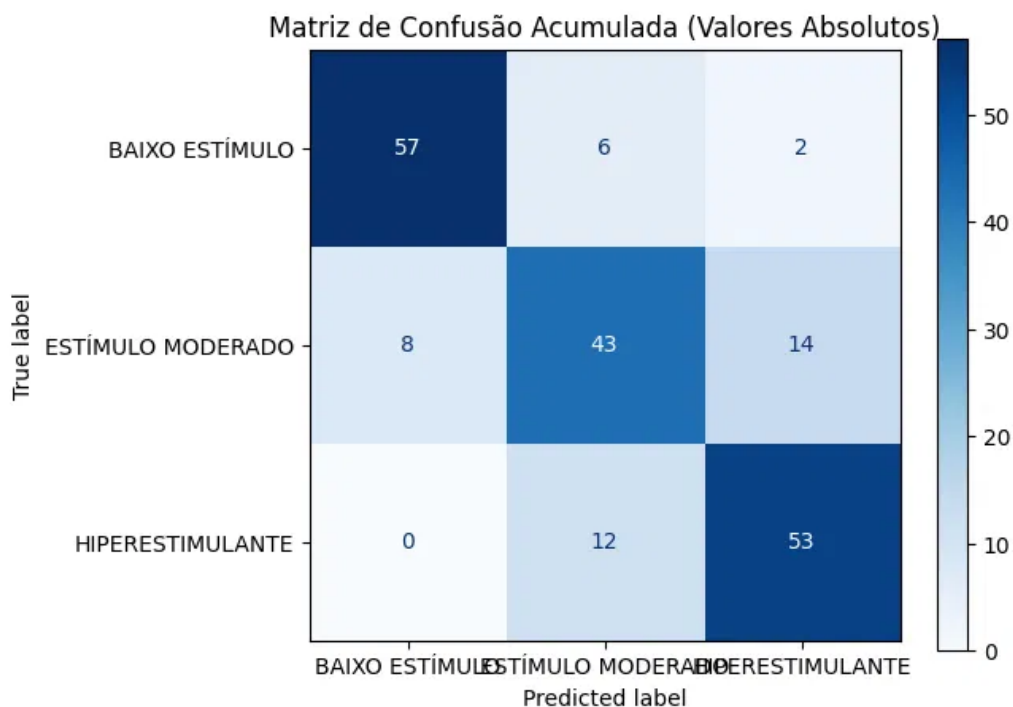


Figura 3 – Matriz de confusão acumulada das cinco rodadas de validação cruzada (valores absolutos).

Fonte: Elaborada pela autora (2026).

A matriz reforça a adequação do modelo ao propósito assistivo da aplicação: dos 65 vídeos de *Baixo Estímulo*, apenas 2 foram classificados como *Hiperestimulante* (falso positivo, ou seja, um conteúdo calmo sinalizado como mais estimulante do que de fato é), e nenhum dos 65 vídeos *Hiperestimulantes* foi confundido com *Baixo Estímulo*. Os erros do modelo concentram-se, portanto, na fronteira entre categorias adjacentes: 8 vídeos de *Estímulo Moderado* foram classificados como *Baixo Estímulo* e 14 como *Hiperestimulante*, enquanto 12 vídeos *Hiperestimulantes* foram classificados como *Estímulo Moderado*. Essa distribuição é relevante para uma aplicação assistiva, pois indica que, quando o modelo erra, ele tende a fazê-lo entre classes vizinhas, e não entre os dois extremos do espectro sensorial.

A partir do detalhamento por categoria, nota-se uma assertividade expressiva nas classes situadas nos extremos do espectro sensorial:

- **Baixo Estímulo:** Apresentou o maior índice de precisão do modelo (0,88), o que indica baixa incidência de falsos positivos para conteúdos calmos. Como detalhado na matriz de confusão (Figura 3), esses erros correspondem a apenas 2 dos 65 vídeos de Baixo Estímulo classificados como Hiperestimulante, reforçando a relevância dessa métrica para a segurança do filtro.

- **Hiperestimulante:** Obteve precisão de 0,77, revocação de 0,82 e *F1-Score* de 0,79. O recall de 0,82 indica que o modelo identificou a maior parte dos vídeos hiperestimulantes, embora aproximadamente 18% dos exemplos dessa classe ainda tenham sido confundidos com outras categorias. Esse resultado demonstra potencial, mas evidencia a necessidade de reduzir falsos negativos antes de um uso assistivo.
- **Estímulo Moderado:** Registrou índices intermediários, com *F1-Score* de 0,68. Esse comportamento decorre do fato de que produções situadas na zona de transição compartilham limiares métricos próximos tanto às produções de baixo impacto quanto às animações mais aceleradas, gerando sobreposições pontuais nas regras de decisão aprendidas pelo conjunto de árvores do modelo.

4.3 ANÁLISE DA VIABILIDADE TÉCNICA

A divergência observada entre os extremos e a classe intermediária não compromete o objetivo central da ferramenta. Em uma aplicação voltada à tecnologia assistiva para o público com TEA, a prioridade do filtro reside justamente na identificação precisa de conteúdos hiperestimulantes e na garantia de segurança das produções de baixo estímulo.

Diante disso, a capacidade do Guardian A.N.Y em isolar vídeos de alto impacto visual com 82% de revocação sustenta a viabilidade técnica da proposta, demonstrando o potencial de utilizar métricas extraídas automaticamente de arquivos de vídeo para a categorização objetiva de mídias infantis, ainda que a taxa de falsos negativos remanescente reforce a necessidade de refinamentos antes de um uso assistivo direto.

5 CONSIDERAÇÕES FINAIS

Este trabalho investigou a viabilidade técnica do Guardian A.N.Y, um sistema automatizado para identificação e classificação de hiperestímulos visuais em vídeos infantis do YouTube. A proposta surge para preencher uma lacuna das plataformas de vídeo, cujos recursos de acessibilidade focam em deficiências motoras, visuais e auditivas, mas carecem de mecanismos específicos para modulação sensorial do público com TEA. Para isso, foi construído um *pipeline* computacional para extrair métricas de similaridade de cor, contraste cromático, frequência de cortes e agitação física a partir dos frames, submetendo esses dados a um classificador *Random Forest* com validação cruzada estratificada (*Stratified K-Fold*).

Os resultados obtidos sugerem a viabilidade técnica da ferramenta. O modelo alcançou uma acurácia média de 78,46% nas cinco rodadas de validação cruzada,

com desvio padrão de 5,28 pontos percentuais. Essa variação moderada indica que o desempenho do classificador não dependeu de uma única partição dos dados, o que aponta para um comportamento preditivo relativamente consistente entre as diferentes divisões, embora esse indicador, isoladamente, não permita descartar sobreajuste (*overfitting*).

Além da acurácia global, o desempenho se destacou nos extremos do espectro sensorial, que são os pontos críticos para a aplicação assistiva, ao passo que a classe de *Estímulo Moderado*, situada na zona de transição entre as demais categorias, apresentou o desempenho mais modesto, conforme detalhado na Seção 4.2.

Uma vez que a prioridade de um filtro assistivo para o público com TEA é identificar conteúdos hiperestimulantes e buscar assegurar que produções indicadas como calmas sejam de fato adequadas, os dados sugerem a viabilidade do Guardian A.N.Y como ferramenta preventiva, ainda que a faixa intermediária e a taxa de falsos negativos na classe Hiperestimulante indiquem a necessidade de refinamentos antes de uma aplicação assistiva direta.

Como limitações, destacam-se o escopo do conjunto de dados, composto por 195 vídeos categorizados manualmente para a extração e validação das métricas, a condução de testes estritamente técnicos, sem etapas operacionais com participantes humanos, e o fato de a rotulagem de referência ter sido realizada por uma única avaliadora (a autora deste trabalho). Por não haver mais de um avaliador envolvido na rotulagem, não foi utilizado o coeficiente Kappa nem qualquer outra métrica de concordância entre avaliadores, o que introduz um componente de subjetividade individual e impede atestar a confiabilidade dos rótulos por essa via. Destaca-se ainda que a calibração do limiar de 0,70 utilizado na detecção de cortes de cena foi realizada sobre vídeos pertencentes ao próprio conjunto de 195 amostras posteriormente avaliado, o que pode ter favorecido de forma otimista os resultados de acurácia obtidos. Por fim, o intervalo de amostragem de 0,5 segundo entre quadros foi definido em função de limitações de hardware disponíveis para o processamento, o que impõe um teto de detecção de cortes de cena e pode subestimar a intensidade de transições em vídeos com edições muito aceleradas. Para trabalhos futuros, recomenda-se a expansão do *dataset*, a rotulagem por múltiplos avaliadores independentes com cálculo de concordância entre avaliadores (Kappa) para validar a confiabilidade dos rótulos, a inclusão de métricas sonoras, como variações de volume e frequência, a exploração de modelos de aprendizado profundo para aperfeiçoar a classificação da categoria moderada e a redução de falsos negativos na classe Hiperestimulante, além da validação do uso do sistema por famílias e profissionais da área.

Conclui-se que o Guardian A.N.Y constitui uma contribuição com potencial para a acessibilidade digital, indicando que a extração automática de métricas de vídeo oferece uma base capaz de reproduzir, em boa medida, os rótulos manuais atribuídos

às produções infantis segundo seu nível de estímulo sensorial, com espaço para aprimoramentos antes de uma aplicação assistiva plena.

REFERÊNCIAS

BRADSKI, G. The opencv library. **Dr. Dobb's Journal of Software Tools**, v. 25, n. 11, p. 120–125, 2000. Citado na página 6.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. Citado na página 6.

DUNN, W. The impact of sensory processing abilities on the daily lives of young children and their families: A conceptual model. **Infants & Young Children**, v. 9, n. 4, p. 23–35, 1997. Citado na página 4.

FERNÁNDEZ-BATANERO, J. M. *et al.* Digital inclusion and connectivity of people with disabilities: a systematic review. **Educational Technology Research and Development**, v. 70, n. 4, p. 1421–1444, 2022. Citado 2 vezes nas páginas 3 e 5.

GARGI, U.; KASTURI, R.; ANTANI, S. Performance characterization of video-shot-change detection algorithms. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 10, n. 1, p. 1–13, 2000. Citado na página 6.

GUEDES, N. P. d. S.; TADA, I. N. C. A produção científica brasileira sobre autismo na psicologia e na educação. **Psicologia: Teoria e Pesquisa**, Brasília, v. 31, n. 3, p. 303–309, 2015. Citado na página 4.

IBGE. **Internet chega a 95% de domicílios do país em 2025**. Agência de Notícias IBGE, 2026. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/47410-internet-chega-a-95-de-domicilios-do-pais-em-2025>>. Acesso em: 21 jul. 2026. Citado na página 3.

LEAL-SILVA, L.; ALVES, J. L. O.; ALVES, J. P. Alterações visuoauditivas em pacientes pediátricos com TEA. **RECIMA21 – Revista Científica Multidisciplinar**, v. 6, n. 5, p. e656419, 2025. Citado na página 6.

LLAMAS, C. R.; VEGA, D. O. C.; RAMOS, M. G. D. Light and sound hypersensitivity in autism spectrum disorder: a systematic review focusing on age and gender bias. **Frontiers in Psychiatry**, v. 17, p. 1771956, 2026. Disponível em: <<https://doi.org/10.3389/fpsy.2026.1771956>>. Citado na página 5.

MATTOS, J. C. Alterações sensoriais no transtorno do espectro autista (tea): implicações no desenvolvimento e na aprendizagem. **Revista Psicopedagogia**, São Paulo, v. 36, n. 109, p. 87–95, 2019. Citado na página 4.

MAZUREK, M. O.; WENSTRUP, C. Television, video game and social media use among children with asd and typically developing siblings. **Journal of Autism and Developmental Disorders**, v. 43, n. 6, p. 1258–1271, 2013. Disponível em: <<https://doi.org/10.1007/s10803-012-1659-9>>. Citado na página 5.

NIC.BR. **Pesquisa sobre o Uso da Internet por Crianças e Adolescentes no Brasil: TIC Kids Online Brasil 2023**. São Paulo: Comitê Gestor da Internet no Brasil (CGI.br), 2024. Livro eletrônico. ISBN 978-65-85417-29-7. Disponível em: <<https://cetic.br/pt/pesquisa/kids-online/>>. Citado na página 3.

NIC.BR. **Pesquisa sobre o uso da Internet por crianças e adolescentes no Brasil: TIC Kids Online Brasil 2025**. 1. ed. São Paulo: Comitê Gestor da Internet no Brasil, 2026. ISBN 978-65-6165-007-6. Disponível em: <<https://cetic.br/pt/pesquisa/kids-online/>>. Acesso em: 28 jul. 2026. Citado na página 3.

SANTOS, A. L. d. B.; VALE, R. M. d. S. **Da Tela ao Cérebro: Efeitos Neurocomportamentais dos Estímulos Audiovisuais na Primeira Infância**. Monografia (Trabalho de Conclusão de Curso (Graduação em Pedagogia)) — UNINASSAU Salvador, Salvador, 2025. Citado na página 5.

SENA, B. U.; BARROS, T. d. S. Hipersensibilidade em crianças com transtorno do espectro autista (TEA). **Revista Foco**, v. 16, n. 11, p. e3502, 2023. Citado na página 3.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. 2nd. ed. [S.l.]: Springer Nature, 2022. Citado na página 6.

VIANA, E. E.; SILVA, L. W.; OLIVEIRA, L. A. d. Hábitos de consumo de mídia sob o olhar das teorias da comunicação. In: **Anais do 46º Congresso Brasileiro de Ciências da Comunicação (Intercom)**. [S.l.: s.n.], 2023. Citado na página 3.

YOUTUBE. **Acessibilidade no YouTube**. 2023. Disponível em: <<https://support.google.com/youtube/answer/189278>>. Acesso em: 5 fev. 2026. Citado 2 vezes nas páginas 3 e 5.