

Detecção de Piscadas Baseada em Aprendizado de Máquina para Controle de Dispositivos por Pessoas com Deficiência Motora Severa

Felipe Barradas Sebastião*, Henrique Coelho dos Santos[†],
Odair Moreira de Souza[‡], Thiago Berticelli Ló[§]
Instituto Federal de Educação, Ciência e Tecnologia do Paraná (IFPR)
Cascavel, Paraná, Brasil

Email: *febarradas13@gmail.com, [†]henriquecoelhodossantos2007@gmail.com,
[‡]odair.desouza@ifpr.edu.br, [§]thiago.lo@ifpr.edu.br

Abstract—In Brazil, approximately 14.4 million people report some type of disability, many of whom have severe motor impairments that limit autonomy in daily activities. This study proposes a system based on computer vision and machine learning to detect and classify blinks, using this signal as a control interface in smart home environments. Facial tracking libraries such as Dlib and MediaPipe were employed to extract facial landmarks and compute the Eye Aspect Ratio (EAR), enabling real-time identification of eye states. The development followed the OSEMN data science cycle, including data collection and cleaning, predictive modeling, and result interpretation. Four classification models were evaluated: Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest, all implemented using Scikit-Learn. Preliminary results indicate that the proposed approach outperforms traditional fixed-threshold methods, accounting for individual physiological variations and environmental conditions. The system features a modular architecture, facilitating expansion and integration with Internet of Things (IoT) devices, providing a low-cost solution with potential positive impact on autonomy and quality of life for individuals with severe motor disabilities.

Keywords—Assistive technologies; Computer vision; Predictive models; Human-computer interaction

Resumo—No Brasil, cerca de 14,4 milhões de pessoas apresentam algum tipo de deficiência, muitas com limitações motoras severas que comprometem a autonomia em atividades cotidianas. Este estudo propõe um sistema baseado em visão computacional e aprendizado de máquina para detectar e classificar piscadas, utilizando esse sinal como interface de controle em ambientes residenciais inteligentes. Foram empregadas bibliotecas de rastreamento facial, como Dlib e MediaPipe, para extração de pontos faciais e cálculo da razão de aspecto ocular (Eye Aspect Ratio), permitindo identificar estados oculares em tempo real. O desenvolvimento seguiu o ciclo OSEMN da Ciência de Dados, incluindo coleta e limpeza de dados, modelagem preditiva e interpretação dos resultados. Quatro modelos de classificação foram avaliados: Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression e Random Forest, todos implementados com Scikit-Learn. Os

resultados preliminares indicam que a abordagem proposta supera métodos tradicionais baseados em limiares fixos, considerando variações fisiológicas individuais e condições ambientais. O sistema apresenta arquitetura modular, favorecendo expansão e integração com dispositivos de Internet das Coisas (IoT), oferecendo uma solução de baixo custo com potencial impacto positivo na autonomia e na qualidade de vida de pessoas com deficiência motora severa.

Palavras-chave—Tecnologias assistivas; Visão computacional; Modelos preditivos; Interação homem-máquina.

I. INTRODUÇÃO

De acordo com o Censo de 2022, aproximadamente 14,4 milhões de pessoas no Brasil declararam possuir algum tipo de deficiência, o que corresponde a 7,3% da população. Dentre esses indivíduos, cerca de 5,2 milhões apresentam comprometimento motor ao subir degraus, enquanto 2,6 milhões relatam dificuldades ao manusear objetos simples, como abrir tampas ou segurar utensílios [1]. Além desses obstáculos, mais de um quinto da população com deficiência encontra-se em condição de analfabetismo, o que reforça a sobreposição de vulnerabilidades enfrentadas por essa parcela da sociedade. Nesse cenário, deficiências motoras graves, como a paraplegia e a tetraplegia, representam um fator de grande impacto, pois restringem a autonomia em tarefas básicas do cotidiano e limitam as oportunidades de inclusão educacional, social e profissional [2].

A legislação brasileira reconhece a importância de enfrentar as desigualdades vivenciadas pelas pessoas com deficiência. A Lei Brasileira de Inclusão, Lei nº 13.147/2015 [3], estabelece explicitamente o direito ao acesso a recursos tecnológicos para promover a acessibilidade e participação social. Em paralelo, no âmbito internacional, a Agenda 2030 das Nações Unidas [4] destaca a inclusão tecnológica como instrumento de garantia

de bem-estar e redução de desigualdades, especialmente nos Objetivos de Desenvolvimento Sustentável de número 3 e 10. Portanto, mais do que uma demanda técnica, o desenvolvimento de tecnologias assistivas se constitui como um compromisso social e político visando assegurar a participação plena e igualitária de pessoas com deficiência em uma sociedade cada vez mais mediada por tecnologias digitais.

Entre as diversas estratégias e tecnologias para promover acessibilidade, o controle ocular assistivo se apresenta como uma abordagem inovadora. Esse tipo de solução permite que movimentos voluntários dos olhos sejam interpretados como comandos para sistemas computacionais ou dispositivos eletrônicos. A detecção de piscadas é uma das técnicas investigadas, pois a contração das pálpebras pode ser realizada de maneira controlada por grande parte dos indivíduos com limitações motoras severas, mesmo em casos de tetraplegia.

Um dos métodos amplamente utilizados para quantificar esse movimento é a Razão do Aspecto Ocular (*Eye Aspect Ratio* - *EAR*), que mede proporções geométricas entre os marcos faciais dos olhos. A simplicidade da *EAR* favorece a sua ampla utilização, mas ao mesmo tempo introduziu uma limitação: a maioria das soluções baseadas nesse índice estabelece um limiar fixo para distinguir olhos abertos e fechados, aplicando-o indistintamente a todos os usuários.

A utilização de limiares fixos, embora prática, desconsidera a grande variabilidade fisiológica existente entre indivíduos. O formato dos olhos, a amplitude da abertura palpebral e até características da iluminação ambiente influenciam os valores da *EAR*. Como consequência, sistemas que não se adaptam a essas variações tendem a apresentar falhas de detecção, dificultando o controle voluntário e comprometendo a experiência de uso [5]. Além disso, a falta de adaptação aos perfis individuais é um dos fatores que contribuem para o abandono de dispositivos, mesmo quando esses poderiam representar um ganho significativo de autonomia [6].

A superação dessas limitações exige a exploração de abordagens que permitam a adaptação personalizada de cada sistema ao usuário final. Nesse sentido, a incorporação de técnicas da Ciência de Dados (*Data Science* - *DS*), em especial o aprendizado de máquina, abre novas possibilidades como um eixo central para a inclusão tecnológica, uma vez que possibilita transformar dados biométricos em modelos de predição capazes de se adaptar às particularidades de cada usuário. Ao mesmo tempo, o paradigma da Inteligência Artificial das Coisas (*AIoT* - *Artificial Intelligence of Things*) potencializa esse processo, integrando sensores e dispositivos conectados que fornecem informações em tempo real. Essa combinação não apenas pode superar as limitações de abordagens baseadas em limiares fixos, como também estabelecer ciclo de adaptação contínua, em que o sistema aprende com a interação cotidiana e oferece uma

acessibilidade mais adaptativa.

Logo, em vez de depender de um limiar global pré-definido, modelos preditivos podem aprender a distinguir padrões específicos de piscadas com base nos dados coletados de cada pessoa [7]. Isso possibilita a construção de soluções mais robustas, capazes de acomodar diferenças fisiológicas e variações ambientais, além de oferecer maior precisão na diferenciação entre piscadas voluntárias e involuntárias.

Outro aspecto a ser considerado é a relação entre precisão e desempenho computacional. A implementação prática de sistemas de controle ocular exige que a detecção de piscadas ocorra em tempo real, sem atrasos perceptíveis para o usuário. Para isso, é relevante avaliar a acurácia dos modelos e o impacto da escolha das bibliotecas de rastreamento facial utilizadas. Assim, a análise comparativa dessas ferramentas constitui em uma etapa fundamental para orientar o desenvolvimento de soluções assistivas eficientes e acessíveis.

O objetivo principal desse trabalho é desenvolver um modelo preditivo de piscadas voltado ao controle de dispositivos em ambientes de automação residencial, destinado a pessoas com tetraplegia ou outras deficiências motoras severas. A proposta busca superar as limitações associadas a sistemas baseados em limiares fixos, por meio da adoção de um modelo adaptativo capaz de considerar as características específicas de cada usuário e diferentes ambientes. A integração da detecção ocular com métodos de aprendizado de máquina, busca-se fornecer uma solução que amplie a autonomia funcional, contribuindo para a redução das desigualdades e promover a inclusão tecnológica.

Este artigo está estruturado da seguinte forma: a Seção II apresenta a fundamentação teórica, que aborda tecnologias assistivas, trabalhos relacionados sobre controle ocular, o cálculo da *EAR* e a aplicação da *DS*. A Seção III detalha a metodologia utilizada, incluindo a coleta, processamento dos dados, treinamento dos modelos e métricas de avaliação. A Seção IV apresenta e discute os resultados dos modelos preditivos. Por fim, a Seção V conclui o estudo, destacando suas contribuições, limitações e trabalhos futuros.

II. FUNDAMENTAÇÃO TEÓRICA

A. Tecnologias Assistivas

As Tecnologias Assistivas (TA) descrevem um amplo conjunto de recursos, serviços e estratégias que visam ampliar a autonomia de pessoas com deficiência em atividades do cotidiano, na educação, no trabalho e no lazer [8]. Entre esses recursos, destacam-se soluções digitais mediadas por visão computacional, as quais permitem que movimentos oculares sejam interpretados como comandos de controle. Esse tipo de tecnologia é particularmente relevante para pessoas com deficiência motora severa, nos quais a preservação parcial

da motricidade ocular pode ser utilizada como canal de comunicação e interação com o ambiente [9].

Contudo, a falta de adaptação aos perfis individuais é um dos fatores mais recorrentes de abandono de dispositivos assistivos [10]. Diante dessa limitação, torna-se fundamental o desenvolvimento de sistemas capazes de incorporar personalização e calibragem dinâmica, adequando-se às necessidades fisiológicas e contextuais de cada usuário.

B. Controle Ocular e Limiar Fixo

Grande parte das pesquisas sobre controle ocular para acessibilidade utiliza a EAR como métrica para diferenciar olhos abertos e fechados [11]. O método baseia-se em marcos faciais e cálculos de proporções geométricas, sendo computacionalmente simples e eficaz em condições controladas. Entretanto, a maioria das implementações adota limiares fixos, definidos empiricamente a partir de testes preliminares e aplicados indistintamente a todos os usuários [12].

Esse tipo de abordagem de limiar fixo é recorrente em diferentes domínios de aplicação. Por exemplo, sistemas de monitoramento de fadiga em motoristas também utilizam a EAR para detectar sinais de sonolência, aplicando limiares previamente definidos para identificar piscadas ou fechamentos prolongados dos olhos [13].

Apesar de sua simplicidade, a abordagem de limiares fixos apresenta sérias limitações: diferenças fisiológicas (formato dos olhos, amplitude de abertura palpebral), variações ambientais (iluminação, reflexos, uso de óculos) e características de aquisição (posição da câmera, taxa de quadros) afetam diretamente a acurácia da detecção. Consequentemente, o sistema pode apresentar falsos positivos (detecção de piscadas involuntárias como voluntárias) e falsos negativos (não detecção de piscadas intencionais), e isso compromete a confiabilidade do controle ocular.

C. Detecção Facial e Cálculo da EAR

A biblioteca MediaPipe, desenvolvida pelo Google, é uma das ferramentas mais utilizadas atualmente para rastreamento de marcos faciais, fornecendo até 478 pontos de referência distribuídos na face conforme apresentado na Figura 1a. Entre esses, um subconjunto é usado para modelagem dos olhos e cálculo da EAR. A Figura 1b ilustra a localização dos pontos oculares utilizados, destacando as regiões empregadas no cálculo da razão de aspecto ocular.

A EAR é definida pela razão entre a distância vertical média das pálpebras e a distância horizontal do olho, garantindo invariância à distância entre o usuário e a câmera. Na implementação mais comum, utiliza-se o ponto superior (386) e o inferior (374) para medir a abertura vertical, enquanto o ponto

lateral direito (466) e o lateral esquerdo (362) determinam a largura horizontal. A Equação 1 formaliza essa relação geométrica e serve como base matemática para os algoritmos de detecção de piscadas.

$$EAR = \frac{\sqrt{(x_{lat2} - x_{lat1})^2 + (y_{lat2} - y_{lat1})^2}}{\sqrt{(x_{sup} - x_{inf})^2 + (y_{sup} - y_{inf})^2} + 0.001} \quad (1)$$

Ao integrar a EAR em um pipeline de observadores, é possível notificar módulos consumidores com informações adicionais, como: valores da EAR por frame, tempo de processamento e índices de quadros, permitindo análises em tempo real e posterior mineração de dados.

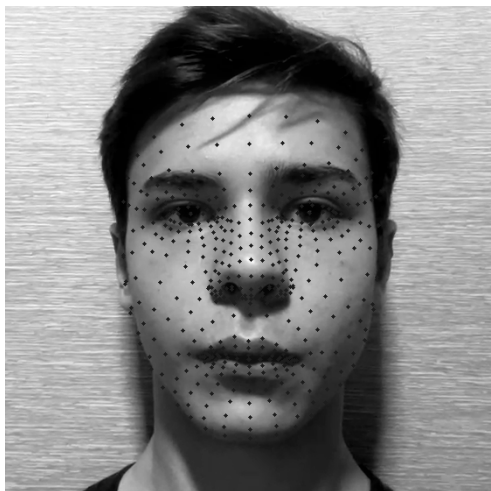
D. Ciência de Dados Aplicada a Tecnologias Assistivas

A aplicação de técnicas da DS em sistemas assistivos permite a transição de abordagens baseadas em regras fixas para modelos preditivos adaptativos. Entre os ciclos de organização do processo analítico, destaca-se o OSEMN (*Obtain, Scrub, Explore, Model, iNterpret* – Obtenção, Limpeza, Exploração, Modelagem e Interpretação), que orienta desde a obtenção dos dados até a interpretação dos resultados [14].

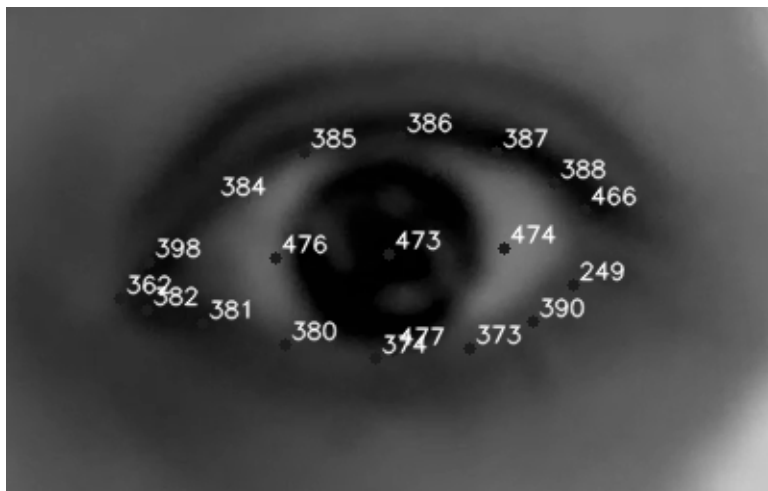
O ciclo de vida OSEMN fornece uma estrutura sistemática para a condução de projetos em DS, sendo amplamente utilizado em pesquisas aplicadas. Na primeira etapa, denominada obtenção (*Obtain*), o foco está na coleta dos dados necessários para a análise. No contexto deste trabalho, isso envolve a captura de informações relacionadas ao movimento ocular, como índices de frames, coordenadas dos pontos faciais, medidas de EAR, tempos de processamento e rótulos de piscadas. Essa fase garante que o conjunto de dados represente adequadamente o fenômeno observado e seja suficiente para sustentar a modelagem posterior.

Na etapa de limpeza (*Scrub*) é realizado o tratamento de inconsistências e valores anômalos que podem comprometer a qualidade dos modelos. Para isso, é aplicada a remoção de *outliers* e valores fora do padrão por meio de técnicas estatísticas, como o intervalo interquartil. Esse processo garante que os dados resultantes sejam mais confiáveis e homogêneos, reduzindo o risco de enviesamentos durante a fase de treinamento dos algoritmos de aprendizado de máquina.

Na fase de exploração (*Explore*), realizam-se análises preliminares para compreender as características do conjunto de dados. Isso inclui a verificação da distribuição das variáveis, a identificação de padrões visuais por meio de gráficos e a análise de medidas estatísticas descritivas. Essa exploração fornece uma visão inicial sobre as correlações entre atributos, ajudando a identificar quais fatores mais contribuem para diferenciar piscadas voluntárias de estados de olhos abertos.



(a) 478 Pontos MediaPipe. Fonte: Autoria Própria, 2025.



(b) 19 Pontos Oculares. Fonte: Autoria Própria, 2025.

Fig. 1: Exemplo de detecção de pontos faciais e oculares utilizando o MediaPipe.

Seguindo para a etapa de modelagem (*Model*), os dados processados são utilizados para treinar algoritmos para realizar a predições sobre o comportamento ocular. Nessa fase, aplicam-se técnicas de aprendizado de máquina supervisionado, que recebem como entrada os atributos extraídos (como a EAR) e produzem classificações sobre os estados do olho. O objetivo principal é definir modelos que superem as limitações de métodos tradicionais baseados em limiares fixos, proporcionando maior adaptabilidade e precisão frente à diversidade fisiológica dos usuários.

Por fim, a etapa de interpretação (*Interpret*) tem como objetivo transformar os resultados obtidos pela modelagem em informações úteis para avaliar a viabilidade do sistema no contexto da tecnologia assistiva. Nessa fase, não apenas métricas quantitativas, como acurácia, *precision*, *recall* e *f1-score*, foram consideradas, mas também a análise qualitativa de usabilidade, buscando compreender em que medida o modelo preditivo de piscadas seria adequado para um usuário real.

Esse processo de interpretação permite identificar limitações práticas, como a influência de condições de iluminação, variações fisiológicas entre indivíduos e a latência do processamento, fatores que podem comprometer a adoção contínua da solução. Essa sistematização garante rastreabilidade e reprodutibilidade, além de facilitar a integração com sistemas de Internet das Coisas (IoT) e AIoT, que fornecem dados contínuos para a adaptação dos modelos em tempo real.

III. METODOLOGIA

O desenvolvimento do modelo de predição foi estruturado de acordo com ciclo de vida de projetos da DS OSEMN, esse modelo garante a organização e rastreabilidade das etapas. A implementação desenvolvida nesse trabalho está licenciada sob a licença MIT e disponibilizada em repositório público no *GitHub*¹. Para a documentação e execução das análises da DS, foram utilizados *notebooks Jupyter*² e a implementação foi realizada em Python.

No escopo deste projeto, foram consideradas exclusivamente piscadas voluntárias, isto é, aquelas realizadas de maneira consciente e controlada pelo usuário. Para fins de classificação, os eventos oculares foram inicialmente divididos em duas categorias principais: piscou e não piscou.

Para realizar o controle, é analisada uma sequência de imagens, a partir da qual são extraídos os pontos faciais Figura 1a, resultando em uma janela de amostras representadas como uma série temporal de pontos oculares. Esses pontos são então interpretados e classificados pelo modelo de aprendizado de máquina. A extração dos pontos faciais foi realizada utilizando um software desenvolvido com as bibliotecas *MediaPipe*³ e *OpenCV*⁴, permitindo o processamento de visão computacional e a captura dos dados do rosto do usuário. Os pontos utilizados na análise correspondem à região dos

¹GitHub - <https://github.com/Barradas13/AssistiveDomotical>. Acessado em: 06 out. 2025

²Notebooks Jupyter - <https://jupyter.org/>. Acessado em: 06 out. 2025

³MediaPipe - <https://pypi.org/project/mediapipe/>. Acessado em: 06 out. 2025

⁴OpenCV - <https://opencv.org/>. Acessado em: 06 out. 2025

olhos, conforme ilustrado na Figura 1b. Embora existam outras bibliotecas para detecção facial, como a *Dlib*, a *MediaPipe* foi escolhida por oferecer instalação intuitiva, maior velocidade de processamento e um número maior de pontos faciais. Dessa forma, desenvolveu-se um software capaz de fornecer as posições faciais do usuário em cada imagem, utilizando uma sequência temporal de imagens para acompanhar e processar o movimento do rosto ao longo do tempo.

Para o estudo foram utilizadas duas abordagens, uma com todos os 19 pontos extraídos para o treinamento e uma segunda abordagem utilizando o índice EAR, definido na Equação 1. Além disso, a modelagem foi feita a partir de um aprendizado supervisionado e sua tarefa especificamente será a classificação de uma sequência de dados do olho do usuário em duas possíveis categorias: não piscou ou piscou.

Para a construção das séries temporais dos pontos extraídos e utilizados no treinamento dos modelos, foi desenvolvido um software que permite a anotação dos dados por meio do pressionamento de um botão. Esse software foi implementado utilizando a biblioteca *OpenCV* para o processamento de imagens e a alocação dos metadados, e a *Pynput* para a detecção do pressionamento de teclas. A captura dos dados foi realizada com uma câmera de resolução 720p (16:9), taxa de 30 quadros por segundo (fps), e um computador equipado com processador Intel i5-8265U 1,60 GHz e 16 GB de RAM.

Para a gravação, sete voluntários foram instruídos a pressionar uma tecla sempre que fechassem os olhos propositalmente. Dessa forma, foi gerada uma sequência de metadados com 15 fps, facilitando o processamento, totalizando aproximadamente 100 mil imagens organizadas temporalmente, cada uma contendo a informação sobre o estado do olho (aberto ou fechado de forma intencional).

Em seguida, o software desenvolvido com *MediaPipe* forneceu as posições do olho, permitindo o cálculo da EAR na primeira abordagem e, na segunda abordagem, fornecendo diretamente as posições oculares. Na segunda abordagem, foi necessária a normalização das posições em relação à pupila, isto é, ao ponto central do olho, uma vez que movimentos da cabeça poderiam deslocar os pontos oculares sem que o usuário estivesse realmente piscando, interferindo na classificação do modelo. A normalização foi realizada tomando como referência o ponto 473 da pupila Figura 1b, que passou a ser o ponto de origem para os demais pontos oculares. Posteriormente, o estado do olho foi classificado a cada cinco imagens. Se pelo menos quatro dessas imagens correspondessem ao olho aberto ou fechado, de acordo com as anotações dos voluntários, essa sequência era registrada para o treinamento do modelo. Com esse procedimento, foram obtidas aproximadamente 11.000 ações de não piscada e 7.100 ações de piscadas.

Para a limpeza dos dados, foi utilizada uma estratégia

de preenchimento e substituição para tratar valores nulos e *outliers*. A limpeza foi dividida em duas etapas: a primeira focada no tratamento de *outliers* nas posições do olho e a segunda no tratamento de *outliers* do valor da EAR.

No caso de posições oculares nulas ou fora do padrão esperado, os valores foram substituídos pelo valor do frame anterior ou posterior. Como a EAR é calculado a partir dessas posições, valores nulos nas posições implicam também em EAR nulo, sendo este substituído pelo valor do frame anterior ou posterior.

Para *outliers* da EAR, foram definidos limites mínimo e máximo. Valores abaixo do limite mínimo foram substituídos pelo limite mínimo (0), enquanto valores acima do limite máximo foram substituídos pelo limite máximo (35). Testes com diferentes indivíduos indicaram que, na prática, a EAR dificilmente ultrapassa esses limites.

A Figura 2 representa dois tipos de piscada: a primeira curva corresponde a uma piscada curta e a segunda a uma piscada longa. Observa-se a diferença nos valores de EAR entre as duas piscadas, sendo que a primeira apresenta uma EAR média de 10, enquanto a segunda apresenta EAR média de 9. Embora essa diferença possa parecer sutil, em abordagens que utilizam limites fixos, valores que não ultrapassam o limiar podem não ser considerados piscadas. Dessa forma, uma diferença pequena como essa pode determinar a classificação de uma ação como piscada ou não. Além disso, essa variação foi observada nos olhos de uma mesma pessoa em um curto intervalo de tempo, indicando que a EAR apresenta flutuações naturais mesmo sem alteração do usuário, o que reforça a necessidade de considerar tais variações na definição de limiares ou no desenvolvimento de modelos de classificação.

A Figura 3 apresenta dois estados do olho, a imagem do lado esquerdo é de um olho aberto e do lado direito mostra um olho fechado. É possível notar a diferença entre elas e como isso afeta a função aproximada formada por ambas.

Para a modelagem, foram avaliados quatro algoritmos de aprendizado de máquina: Árvores de Decisão (*Decision Tree Classifier* -DT), K-Vizinhos Mais Próximos (*K-Nearest Neighbors* -KNN), Regressão Logística (*Logistic Regression* - LR) e Florestas Aleatórias (*Random Forest* - RF). Todos implementados por meio da biblioteca *Scikit-Learn*⁵.

A entrada dos dados consiste em 5 frames sequenciais, sendo que para a primeira abordagem são extraídos os valores dos 19 pontos do olho, enquanto a segunda abordagem é utilizada o valor da EAR obtida de cada frame.

A coleta de dados para este estudo foi realizada com sete voluntários, caracterizando-se como uma etapa inicial de viabilidade do projeto. Para a validação dos modelos a partir

⁵Scikit-Learn - <https://scikit-learn.org/>. Acessado em: 06 out. 2025

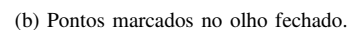
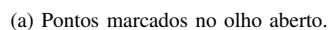


Fig. 3: Gráfico de diferença entre os olhos aberto e fechado. Fonte: Autoria Própria, 2025.

6 conjuntos (representados por diferentes participantes) para treinamento e 1 conjunto para teste em cada iteração, assim cobrindo todas as combinações possíveis.

Essa abordagem permite avaliar a capacidade de generalização dos modelos, considerando variações entre diferentes ambientes e indivíduos, o que é essencial para aplicações futuras, pois possibilita o uso dos modelos por novos usuários sem necessidade de treinamento.

IV. RESULTADOS E DISCUSSÕES

A avaliação dos modelos foi realizada utilizando o conjunto de dados dos sete voluntários. A Tabela I resume o desempenho de quatro algoritmos de classificação distintos, contrastando as duas abordagens de extração de características: EAR e Pontos Faciais. Destaca-se o modelo *Random Forest Classifier*, que obteve as melhores métricas em ambas as abordagens. Também, embora ambas as entradas de dados sejam eficazes, a abordagem com EAR apresentou uma ligeira vantagem em termos de precisão para a maioria dos modelos.

TABELA I: Métricas de avaliação dos modelos de acordo com o tipo de entrada (EAR ou Pontos Faciais)

Modelo	Métrica	Com EAR	Com Pontos
Decision Tree Classifier	Acurácia	0,998	0,992
	Precisão	0,996	0,982
	Revocação	0,998	0,998
	F1-Score	0,997	0,990
Random Forest Classifier	Acurácia	0,998	0,997
	Precisão	0,997	0,993
	Revocação	0,999	0,999
	F1-Score	0,998	0,996
Logistic Regression	Acurácia	0,942	0,966
	Precisão	0,944	0,941
	Revocação	0,902	0,972
	F1-Score	0,922	0,956
K-Nearest Neighbors (KNN)	Acurácia	0,986	0,985
	Precisão	0,976	0,967
	Revocação	0,989	0,993
	F1-Score	0,982	0,980

Para uma análise mais aprofundada do comportamento dos classificadores, a Figura 4 apresenta as matrizes de confusão acumuladas para cada modelo nas abordagens EAR e Pontos, respectivamente (A coluna da esquerda representa a abordagem EAR e a da direita, a abordagem Pontos).

Para comparação com um método de detecção de piscadas de limiar fixo, utilizou-se o algoritmo *Grid Search* para encontrar o limiar de EAR que maximizasse a acurácia, tanto de forma global (para todos os voluntários) quanto individual.

Os limiares ótimos resultantes para cada participante são apresentados na Figura 5. Como pode ser observado, os valores variam consideravelmente entre os usuários; a grande diferença entre o limiar do Voluntário 1 e os demais, por exemplo,

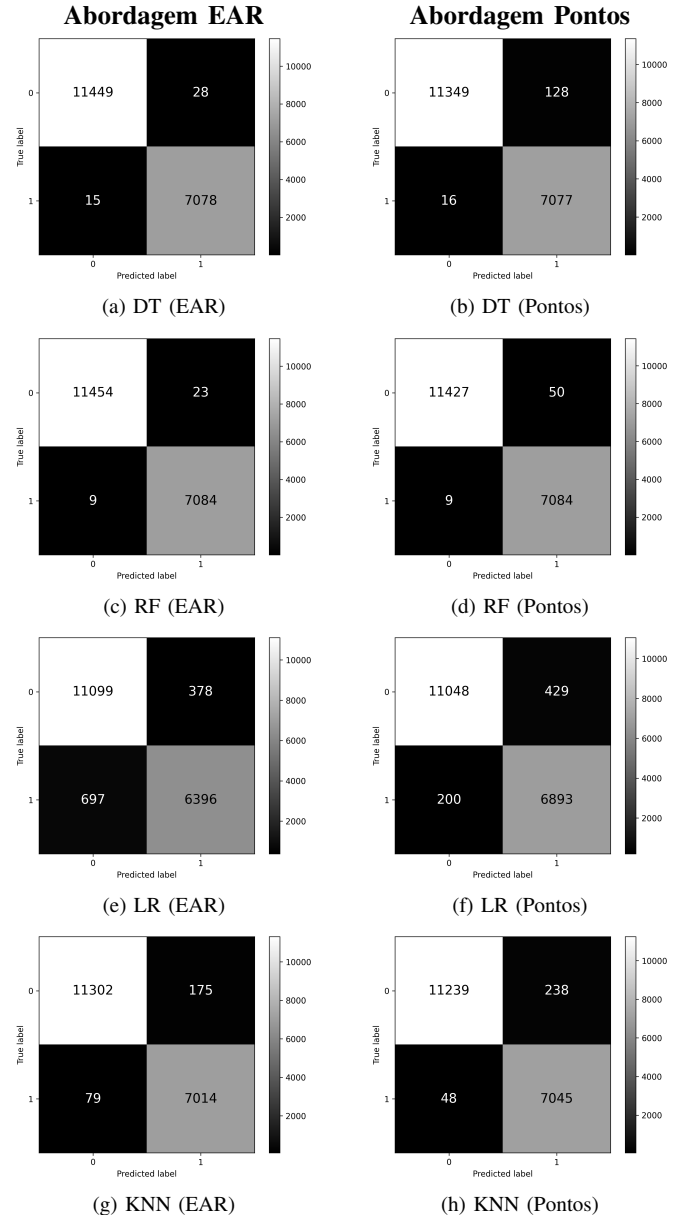


Fig. 4: Comparativo das matrizes de confusão (média por modelo).

evidencia que um valor fixo único não é uma solução eficaz para todos os casos.

Além disso, a Tabela II mostra os valores das métricas de avaliação com o método de limiar fixo geral que foi aferido como $EAR = 7,116$ a partir do *Grid Search*. Pode-se observar uma diferença considerável ao comparar os resultados descritos na Tabela I. Essa diferença reforça que os modelos preditivos foram capazes de aprender padrões mais complexos e robustos

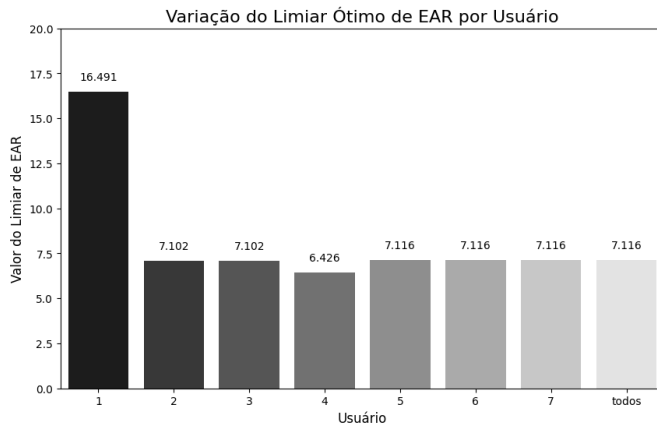


Fig. 5: Variação do Limiar Ótimo de EAR por Usuário.

do que uma simples heurística de limiar.

TABELA II: Métricas de avaliação para o método de limiar fixo.

Métrica	Valor
Acurácia	0,937
Precisão	0,908
Revocação	0,930
F1-Score	0,919

Para uma validação qualitativa, foi desenvolvida uma implementação visual que sobrepõe as previsões do modelo aos vídeos dos voluntários, disponível publicamente no repositório do projeto. Durante essa análise, observou-se que as anotações manuais continham pequenas imprecisões de sincronia, por exemplo, um leve atraso entre o fechamento do olho e o pressionamento da tecla de anotação. Notavelmente, o modelo treinado demonstrou robustez a esse ruído, aprendendo o padrão visual da piscada em si, em vez de simplesmente memorizar os rótulos, o que evidencia sua capacidade de generalização.

O processamento do modelo, realizado em um computador com processador Intel i5-8265U 1,60 GHz e 16 GB de RAM, levou em média 17 milissegundos por avaliação do conjunto de entradas, permanecendo dentro da taxa de amostragem de 1/15 frames (66 milissegundos por frame).

Para a implantação desses modelos em um aplicativo, basta substituir a função que determina se o olho está aberto ou fechado pelo modelo treinado, no lugar da avaliação por limiar fixo. Essa abordagem facilita a construção de módulos adicionais, como interfaces gráficas para a comunicação de pessoas com deficiência utilizando movimentos das pálpebras. Como o modelo foi treinado com vídeos de 15 fps, recomenda-se manter aproximadamente a mesma taxa de quadros,

considerando que a janela de classificação utiliza 5 frames.

Durante os testes, foram observados poucos erros de classificação. Em particular, diferenças sutis entre piscadas curtas e longas podiam gerar classificações incorretas. Além disso, o modelo não conseguiu identificar exclusivamente piscadas propositalis: piscadas não propositalis com duração superior a 5 frames eram reconhecidas como piscadas. Outro problema identificado foi a dificuldade em diferenciar entre piscadas e olhares para baixo; nesses casos, o modelo classificou erroneamente o olhar para baixo como uma piscada.

V. CONCLUSÃO

O presente trabalho apresentou o desenvolvimento de um modelo de predição de piscadas com aplicação em tecnologias assistivas, explorando o ciclo OSEMN como guia metodológico e comparando diferentes abordagens de extração de características a partir do rastreamento ocular. A principal contribuição reside na superação das limitações impostas por métodos baseados em limiares fixos de EAR, uma vez que o modelo treinado demonstrou maior adaptabilidade às variações fisiológicas e ambientais dos usuários.

Apesar dos resultados positivos, algumas limitações foram observadas. Primeiramente, o modelo ainda não diferencia piscadas voluntárias e involuntárias de forma plena, o que pode comprometer a confiabilidade em situações de uso contínuo. Também foram detectadas dificuldades na distinção entre piscadas e movimentos oculares específicos, como o ato de olhar para baixo, que em determinados contextos foi classificado erroneamente como piscada. Outro aspecto a ser considerado é a dependência de condições controladas de captura, como taxa de quadros e iluminação mínima aceitável, fatores que ainda restringem a aplicabilidade em ambientes reais mais adversos. Por fim, embora os algoritmos apresentem bom desempenho em termos de acurácia, a avaliação prática de usabilidade com um grupo maior e mais diverso de usuários permanece necessária para consolidar a robustez da solução.

Para trabalhos futuros, a pesquisa se concentrará na validação da robustez do sistema com um número maior de voluntários e na expansão de sua capacidade interpretativa, investigando o reconhecimento de microexpressões faciais e padrões oculares complexos para enriquecer a interação do usuário. Paralelamente, a integração com dispositivos AIoT será explorada para permitir a adaptação contínua do modelo em tempo real. Por fim, avaliar a eficácia do sistema em cenários de uso real

AGRADECIMENTOS

Ao Instituto Federal de Educação, Ciência e Tecnologia do Paraná (IFPR) pelo auxílio financeiro por meio de bolsa estudantil via Edital nº 07/2024 - Unificado de Inovação 2024

(Bolsa estudante) - Edital de Pesquisa e oportunidade de pesquisa.

REFERÊNCIAS

- [1] Instituto Brasileiro de Geografia e Estatística (IBGE), “Censo demográfico 2022: Pessoas com deficiência e pessoas diagnosticadas com transtorno do espectro autista – resultados preliminares da amostra,” <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/43463-censo-2022-brasil-tem-14-4-milhoes-de-pessoas-com-deficiencia,2022, acessado em: 9 out. 2025>.
- [2] V. Baldassin, C. Lorenzo, and H. E. Shimizu, “Tecnologia assistiva e qualidade de vida na tetraplegia: abordagem bioética,” *Revista Bioética*, vol. 26, no. 4, pp. 574–586, 2018. [Online]. Available: <https://doi.org/10.1590/1983-80422018264276>
- [3] Brasil, “Lei no 13.146, de 6 de julho de 2015: Institui a lei brasileira de inclusão da pessoa com deficiência,” http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/Lei/L13146.htm, 2015, acessado em: 21 de agosto de 2025.
- [4] Nações Unidas, “Transformando nosso mundo: A agenda 2030 para o desenvolvimento sustentável,” https://www.mds.gov.br/webarquivos/publicacao/brasil_amigo_pesso_idosa/agenda2030.pdf, 2015, acesso em: 26 set. 2025.
- [5] N. Modi and J. Singh, “A survey of research trends in assistive technologies using information modelling techniques,” *Disability and Rehabilitation: Assistive Technology*, vol. 17, no. 6, pp. 605–623, 2022. [Online]. Available: <https://doi.org/10.1080/17483107.2020.1817992>
- [6] A. T. Sugawara, V. D. Ramos, F. M. Alfieri, and L. R. Battistella, “Abandonment of assistive products: assessing abandonment levels and factors that impact on it,” *Disability and Rehabilitation: Assistive Technology*, vol. 13, no. 7, pp. 716–723, 2018. [Online]. Available: <https://doi.org/10.1080/17483107.2018.1425748>
- [7] B. R. Ibrahim, F. M. Khalifa, S. R. M. Zeebaree, N. A. Othman, A. Alkhayyat, R. R. Zebari, and M. A. M. Sadeeq, “Embedded System for Eye Blink Detection Using Machine Learning Technique,” in *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*. Babil, Iraq: IEEE, Apr. 2021, pp. 58–62. [Online]. Available: <https://doi.org/10.1109/BICITS51482.2021.9509908>
- [8] R. Bersch, “Introdução à tecnologia assistiva,” *Porto Alegre: CEDI*, vol. 21, pp. 1–20, 2008.
- [9] World Health Organization (WHO), “Assistive technology: Fact sheet,” <https://www.who.int/news-room/fact-sheets/detail/assistive-technology>, 2023, acesso em: 21 ago. 2025.
- [10] S. Federici, F. Meloni, and S. Borsci, “The abandonment of assistive technology in Italy: a survey of national health service users,” *European journal of physical and rehabilitation medicine*, vol. 52, no. 4, pp. 516–526, 2016.
- [11] J. Cech and T. Soukupova, “Real-time eye blink detection using facial landmarks,” *Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague*, pp. 1–8, 2016.
- [12] A. Królak and P. Strumillo, “Eye-blink detection system for human–computer interaction,” *Universal Access in the Information Society*, vol. 11, no. 4, pp. 409–419, 2012. [Online]. Available: <http://link.springer.com/10.1007/s10209-011-0256-6>
- [13] L. C. COSTA *et al.*, “Monitoramento inteligente de fadiga: detecção de sonolência em motoristas com ia e visão computacional para aumentar a segurança no trânsito,” 2024.
- [14] H. Mason and C. Wiggins, “A taxonomy of data science,” 2010, acesso em: 6 out. 2025. [Online]. Available: <https://web.archive.org/web/20160220042455/dataists.com/2010/09/a-taxonomy-of-data-science>