

# Inovações Tecnológicas na Comunicação Inclusiva: Detecção de LIBRAS por meio de Inteligência Artificial no Edge Impulse

Rafaela de Moraes Papale  
*Engenheira de Computação*  
INATEL  
Santa Rita do Sapucaí, Brasil  
rafaelamp@mtel.inatel.br

Vinicius Bottini Jardim  
*Engenheiro de Controle e Automação*  
INATEL  
Santa Rita do Sapucaí, Brasil  
vinicius.jardim@mtel.inatel.br

**Abstract**—This article investigates the implementation of artificial intelligence and neural networks, particularly within the Edge Impulse platform, for the detection of Brazilian Sign Language (LIBRAS) through mobile cameras. Using a Kaggle dataset, the study demonstrates the model's efficiency in recognizing a diverse array of LIBRAS signs. The project emphasizes real-world applicability, with promising results in terms of speed and adaptability. While challenges were acknowledged, the overall impact underscores the potential for technology to bridge linguistic gaps, contributing to a more inclusive and connected society.

**Index Terms**—Brazilian Sign Language; Neural Networks; Sign Detection;

**Resumo**—Este artigo investiga a implementação de inteligência artificial e redes neurais, especialmente na plataforma Edge Impulse, para a detecção da Língua Brasileira de Sinais (LIBRAS) por meio das câmeras de dispositivos móveis. Utilizando dois conjuntos diferentes de dados do Kaggle, o estudo demonstra a eficiência do modelo em reconhecer uma variedade de sinais de LIBRAS. O projeto destaca a aplicabilidade prática, com resultados promissores em termos de velocidade e adaptabilidade. Embora desafios tenham sido reconhecidos, o impacto global sublinha o potencial da tecnologia para superar barreiras linguísticas, contribuindo para uma sociedade mais inclusiva e conectada.

**Index Terms**—LIBRAS; Redes Neurais; Detecção de Sinais;

## I. INTRODUÇÃO

A comunicação é uma parte intrínseca da experiência humana, uma ponte que conecta indivíduos e sociedades. Contudo, para aqueles que se comunicam por meio da Língua Brasileira de Sinais (LIBRAS), os obstáculos linguísticos frequentemente representam barreiras significativas em sua interação com o mundo. A LIBRAS é uma forma rica e complexa de expressão, composta por gestos, expressões faciais e movimentos corporais. No entanto, a falta de compreensão generalizada da LIBRAS cria uma barreira comunicativa, limitando as interações e o acesso a informações cruciais [1].

Neste cenário desafiador, a inteligência artificial emerge como uma ferramenta promissora para superar as barreiras comunicativas. Algoritmos avançados de aprendizado de máquina, quando treinados para reconhecer padrões complexos presentes na LIBRAS, têm o potencial de traduzir eficientemente gestos e expressões em linguagem compreendida por

todos. Essa aplicação da IA não apenas facilita a comunicação entre pessoas surdas e ouvintes, mas também amplia o acesso a informações essenciais em ambientes que tradicionalmente excluem a expressão visual da linguagem [2].

Dentro deste contexto, destaca-se o *Edge Impulse*, uma plataforma que possibilita a construção e implementação direta de modelos de IA em dispositivos de borda, representando um avanço notável. Ao descentralizar o processamento de dados para o local de interação, o *Edge Impulse* diminui a dependência de conexões de internet robustas, resultando em uma detecção de LIBRAS mais ágil e eficaz. A utilização de redes neurais no *Edge Impulse* oferece a capacidade de desenvolver sistemas compactos e eficientes, particularmente adequados para dispositivos como *smartphones*. Assim, o propósito deste artigo é a implementação de um projeto que empregue um conjunto de dados de sinais de LIBRAS em uma rede neural, de modo que, por meio da câmera do celular, seja possível identificar o sinal correspondente ao gesto realizado.

O presente artigo é subdividido da seguinte forma: na seção II deste artigo, serão abordados os trabalhos relacionados que servem como base e contexto para a pesquisa atual. Na seção III, será detalhada a metodologia adotada para o desenvolvimento do projeto, descrevendo as ferramentas, técnicas e abordagens utilizadas para alcançar os objetivos propostos. Em seguida, na seção IV, serão apresentadas as análises e resultados obtidos ao longo do processo. Este segmento destacará as descobertas significativas e as contribuições específicas do projeto. Por fim, na seção V, a conclusão consolidará os principais pontos discutidos, resumindo as descobertas, discutindo as implicações práticas e sugerindo possíveis direções para pesquisas futuras nesta área.

## II. TRABALHOS RELACIONADOS

Nesta seção, são examinados estudos anteriores pertinentes. A revisão da literatura tem como objetivo contextualizar e posicionar o estudo dentro do panorama existente, identificando lacunas no conhecimento, avaliando abordagens metodológicas eficazes e destacando contribuições relevantes ao campo.

[L. C. Gonçalves, E. F. Saad, R. B. Andrade, B. A. Romero, and R. D. Campos] propõem o desenvolvimento de um programa que utiliza o sensor *Kinect*, redes neurais e processamento de imagens para traduzir eletronicamente a Linguagem Brasileira de Sinais (LIBRAS). O objetivo é capturar gestos realizados livremente com a mão do usuário, permitindo treinamento e reconhecimento de diversos tipos de gestos. O sensor *Kinect*, originalmente projetado para o videogame *Xbox 360* da *Microsoft*, é empregado para capturar imagens junto com mapas de profundidade. A partir do rastreamento e captura de imagens, é possível distinguir os gestos obtidos, que correspondem a letras específicas em LIBRAS. O trabalho inclui o acompanhamento de elementos de cena, aprendizado utilizando matrizes binárias geradas por processamento de imagens e reconhecimento em uma Rede Neural Artificial (RNA). Os resultados alcançados foram de 88%, superando significativamente o mínimo aceitável para uma rede neural, demonstrando a eficácia do sistema proposto [2].

[M. F. O. de M. Santos and C. C. Quarto] abordam a respeito do desafio da alfabetização de pessoas surdas no contexto educacional, destacando a escassez de capacitação para atividades de ensino nessa área, apesar da existência da LIBRAS. O crescimento da internet e das tecnologias computacionais motivou a criação de aplicações avançadas de inteligência artificial para aprimorar esse processo. O artigo foca em pesquisas relacionadas à aplicação de redes neurais convolucionais, especialmente no contexto da visão computacional. Experimentos foram conduzidos usando uma base de dados de LIBRAS para treinar imagens por meio do algoritmo YOLOv5, uma rede neural profunda, para realizar a classificação de grupos de 6 e 7 imagens. Os resultados dos testes foram comparados, destacando as épocas que obtiveram melhor desempenho na classificação das imagens de LIBRAS. Essa abordagem indica uma tentativa de integrar tecnologias avançadas de inteligência artificial, especificamente redes neurais convolucionais, para melhorar o processo de alfabetização de pessoas surdas, oferecendo uma contribuição significativa para a educação inclusiva [3].

[I. G. O. Bastos] explora o reconhecimento de sinais da Língua Brasileira de Sinais (LIBRAS) por meio de uma perspectiva que emprega descritores de forma e redes neurais artificiais. A pesquisa concentra-se nos gestos, considerados como ações corporais não-verbais, com o propósito de facilitar interações humanas com sistemas computacionais. A metodologia adotada compreende uma revisão da literatura sobre técnicas de reconhecimento de gestos, enfatizando a escolha de descritores e classificadores para extrair características relevantes das imagens. O estudo utiliza um vetor de características derivado da aplicação dos descritores Histograma de Gradientes Orientados (HOG) e Momentos Invariantes de Zernike (MIZ) para a extração de informações. O processo de reconhecimento é conduzido pelo classificador Perceptron Multicamada, implementado em uma arquitetura de dois estágios. Diante da ausência de *datasets* públicos da LIBRAS, os pesquisadores criaram um conjunto de dados composto por

9600 imagens representando 40 sinais da LIBRAS, envolvendo a colaboração de especialistas e alunos surdos. Os testes realizados resultaram em uma taxa de acerto de 96,77%, evidenciando a eficácia da metodologia. A validação abrange considerações sobre potenciais ameaças, como a realização de testes com indivíduos não incluídos no conjunto de treinamento, além da aplicação da técnica em um *dataset* público de gestos [4].

[J. P. C. Sobrinho, L. O. P. H. da Silva, G. C. B. C. Dalpra, and S. C. A. Basilio] discutem sobre o processo para o reconhecimento e tradução de sinais em LIBRA) utilizando redes neurais artificiais. Destacando a relevância da LIBRAS como segunda língua oficial do Brasil e a considerável comunidade de aproximadamente nove milhões de pessoas com deficiência auditiva, os autores propõem uma solução alternativa para a necessidade de intérpretes de LIBRAS. Em vez de depender exclusivamente de intérpretes, a pesquisa propõe o uso de métodos de redes neurais artificiais para o reconhecimento e tradução de LIBRAS. O processo apresentado utiliza vídeos como entrada e uma Rede Neural Convolucional-Recorrente, conhecida como ConvLSTM, para a análise sequencial de *frames* dos vídeos. A análise ocorre em duas etapas, primeiro na camada convolucional e, em seguida, na camada recorrente da rede. A versão atual da rede implementada já passou por coleta de dados e treinamento, permitindo o reconhecimento de vídeos que representam ou não uma sentença específica em LIBRAS. A proposta destaca-se como uma alternativa tecnológica para promover a inclusão de pessoas com deficiência auditiva na comunidade em geral [5].

### III. METODOLOGIA

A base fundamental para qualquer projeto de visão computacional repousa sobre um conjunto de dados representativo. Neste projeto, a escolha foi utilizar dois conjuntos de dados proveniente da plataforma *Kaggle*, reconhecida por sua especialização em fornecer dados para projetos de ciência de dados e aprendizado de máquina. Optou-se por empregar ambos os conjuntos de dados mesclados, somando no total 11548 imagens em 21 letras diferentes do alfabeto, para treinamento e teste, capturados em diversos ambientes e condições de iluminação. Ambos os conjuntos podem ser acessados por meio dos links: <https://www.kaggle.com/datasets/allanpardinho/libras-cnn/data> e <https://www.kaggle.com/datasets/williansoliveira/libras>.

Antes de alimentar o *dataset* ao modelo, realizou-se uma fase de pré-processamento para garantir a consistência e qualidade dos dados. Isso envolveu a normalização das imagens, redimensionamento para um formato padrão de 64x64 pixels, e a separação apropriada entre conjuntos de treinamento e teste.

A plataforma *Edge Impulse* foi selecionada como o ambiente de desenvolvimento para o modelo de detecção de LIBRAS. Por meio da interface do *Edge Impulse*, o modelo foi treinado utilizando a técnica de redes neurais convolucionais (CNNs). O conjunto de treinamento foi introduzido no modelo, e o processo iterativo de ajuste de hiperparâmetros foi conduzido para otimizar seu desempenho.

Na arquitetura desta rede neural de classificação de imagens, a primeira camada 2D Conv/Pool consiste em 32 filtros com um tamanho de *kernel* de 3x3, operando em uma única camada. A função de ativação utilizada é a ReLU (*Rectified Linear Unit*), seguida por uma camada de *pooling* que tem o objetivo de reduzir a dimensionalidade espacial das características extraídas. A segunda camada 2D Conv/Pool apresenta 16 filtros e um tamanho de *kernel* de 3x3, também operando em uma única camada, com a mesma função de ativação ReLU. Assim como na primeira camada, uma camada de *pooling* subsequente é empregada para continuar a redução da dimensionalidade.

Após estas camadas convolucionais, a arquitetura incorpora uma camada *Flatten*, cuja função é transformar os mapas de características tridimensionais em um vetor unidimensional. Essa transformação é essencial para a transição para camadas densas (*fully connected*) subsequentes. Para mitigar o risco de *overfitting*, uma camada *Dropout* é introduzida com uma taxa de 0.25, desativando aleatoriamente 25% dos neurônios durante o treinamento. Isso promove a robustez do modelo ao evitar dependências excessivas de neurônios específicos.

No que diz respeito à divisão dos dados, 80% são alocados para o conjunto de treino, enquanto os restantes 20% são destinados ao conjunto de teste. Essa prática comum permite que o modelo ajuste seus pesos durante o treinamento utilizando dados conhecidos, enquanto a eficácia do modelo é avaliada em dados não vistos, proporcionando uma estimativa realista de seu desempenho em situações do mundo real. Essa arquitetura representa, assim, uma abordagem equilibrada e eficaz para a tarefa de classificação de imagens.

Uma vez treinado e validado, o modelo foi integrado ao *Edge Impulse* para implementação em dispositivos de borda, como *smartphones*. A plataforma facilitou esse processo, permitindo a conversão do modelo em um formato otimizado para o consumo eficiente de recursos do dispositivo. Ademais, a configuração de uma interface de câmera no *Edge Impulse* possibilitou a captura de imagens em tempo real, tornando a aplicação apta a reconhecer sinais de LIBRAS através da câmera do celular.

Para avaliar a eficácia do modelo implementado, uma série de testes foi conduzida utilizando o conjunto de dados de teste separado anteriormente. A precisão, sensibilidade e especificidade do modelo foram avaliadas, proporcionando uma visão abrangente de seu desempenho em situações do mundo real. O projeto encontra-se disponível no GitHub e pode ser acessado pelo link [https://github.com/bottinivj/TP557\\_PROJETO](https://github.com/bottinivj/TP557_PROJETO).

#### IV. ANÁLISE E RESULTADOS

Após a implementação do projeto de detecção de sinais de LIBRAS utilizando o *Edge Impulse* e um *dataset* adquirido da plataforma *Kaggle*, conduziu-se uma análise abrangente para avaliar a eficácia e a viabilidade prática do sistema proposto.

O modelo de detecção de LIBRAS, treinado e implementado no *Edge Impulse*, demonstrou um desempenho notável durante os testes. A taxa de acerto na identificação de sinais de LIBRAS atingiu uma acurácia de 98%, isso representa

a proporção de predições corretas em relação ao total de exemplos. Neste caso, sugere que o modelo está acertando a classificação da grande maioria das imagens no conjunto de treinamento. Isso é um indicativo de que a rede neural está aprendendo bem as características das imagens e é capaz de generalizar eficientemente para novos exemplos.

Além disso, encontrou-se um valor de perda de 0.12, que indica quão bem o modelo está performando durante o treinamento. O objetivo é minimizar o valor de perda. Um valor de perda de 0.12 é relativamente baixo e sugere que o modelo está fazendo boas previsões com uma quantidade mínima de erro. Isso significa que o modelo está ajustando seus pesos de maneira eficiente para se adaptar aos dados de treinamento.

A seguir apresenta-se a matriz de confusão indicando os valores de precisão (*precision*), *recall* e *F1-Score*. A precisão mede a proporção de imagens positivas classificadas corretamente entre todas as imagens classificadas como positivas. O *recall* mede a proporção de imagens positivas corretamente identificadas em relação a todas as imagens que são verdadeiramente positivas. O *F1-Score* é uma média harmônica entre precisão e *recall*, proporcionando uma métrica que considera ambos os tipos de erros.

	F1-SCORE	PRECISION	RECALL
A	0.94	0.95	0.92
B	0.96	0.99	0.94
C	1.00	1.00	1.00
D	0.97	0.96	0.99
E	0.97	0.98	0.96
F	0.98	0.99	0.97
G	0.98	1.00	0.97
I	0.98	0.98	0.99
L	0.99	1.00	0.97
M	0.99	1.00	0.98
N	0.99	0.99	0.99
O	0.97	0.99	0.96
P	1.00	1.00	1.00
Q	0.99	0.98	1.00
R	1.00	1.00	0.99
S	0.95	0.91	0.99
T	0.97	0.94	1.00
U	0.97	0.94	1.00
V	1.00	0.99	1.00
W	0.98	1.00	0.96
Y	1.00	1.00	1.00

Figura 1. Matriz de Confusão

Ao analisar a matriz de confusão, destaca-se que várias classes, nomeadamente C, P, Q, R e Y, demonstram F1-Scores perfeitos de 1.00, evidenciando um equilíbrio ótimo entre precisão e *recall*. Além disso, classes como C, P, Q, R, M e L exibem valores de *precision* perfeitos, sinalizando que o modelo tem uma baixa propensão para classificar erroneamente instâncias negativas como positivas. Da mesma forma, classes como C, P, Q, R, N, U, V e Y apresentam *recalls* perfeitos de 1.00, indicando que o modelo consegue identificar todas as instâncias positivas dessas classes.

Notavelmente, as classes C, P, Q, R e Y se destacam com resultados excepcionais, exibindo F1-Scores, *precision* e *recall* perfeitos, sugerindo um desempenho notável nessas categorias.

Classes com F1-Scores superiores a 0.95, como A, B, D, E, F, G, I, L, M, N, O, T, U, V e W, denotam um equilíbrio robusto entre precisão e *recall*.

Por outro lado, a classe S, com um F1-Score de 0.95, demonstra um equilíbrio razoável entre *precision* e *recall*, embora com um *recall* ligeiramente inferior. Isso pode indicar uma capacidade um pouco menor do modelo de identificar todas as instâncias positivas dessa classe. Ainda mais, a classe S exibe uma *precision* mais baixa de 0.91, sugerindo uma propensão maior a classificar instâncias negativas como positivas.

Por fim, as classes A, B, D, E, F, G, I, L, M, N, O, T, U, V e W revelam *precision* e *recall* superiores a 0.94, indicando um desempenho sólido nessas métricas e reforçando a eficácia do modelo para essas categorias específicas.

A figura a seguir representa a *feature explorer* (explorador de características) e exibe um subconjunto do conjunto de treinamento completo, composto por 2000 amostras de um total de 9183, classificadas pela rede neural. Nesta visualização, as amostras corretamente classificadas são representadas por itens em verde, enquanto as amostras incorretamente classificadas são destacadas em vermelho. O destaque visual das amostras classificadas corretamente em verde proporciona uma rápida percepção da eficácia do modelo em relação a esse subconjunto específico. Por outro lado, as amostras em vermelho indicam as instâncias em que o modelo falhou na classificação, possibilitando uma análise mais detalhada dos casos de erro.

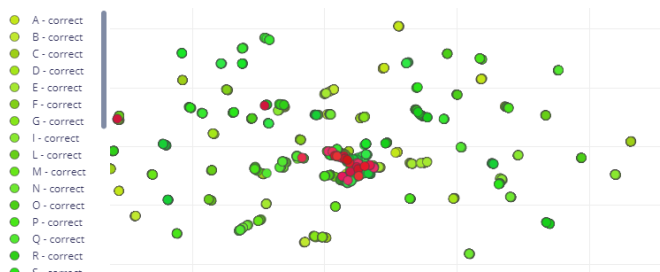


Figura 2. Feature explorer

A avaliação da eficácia do modelo foi conduzida em diversos cenários do mundo real, revelando resultados que destacam a sua elevada confiabilidade na detecção de sinais da Língua Brasileira de Sinais (LIBRAS). A capacidade de capturar e interpretar esses sinais de maneira consistente, mesmo em ambientes dinâmicos, sugere uma aplicabilidade prática em situações cotidianas. A seguir, apresentam-se algumas imagens que ilustram o funcionamento do projeto, validando de forma concreta a precisão na interpretação dos sinais.

O modelo revelou uma notável capacidade de adaptação a uma ampla gama de sinais de LIBRAS. Isso destaca a versatilidade do sistema em reconhecer a riqueza da linguagem gestual da LIBRAS, contribuindo para uma experiência de comunicação mais abrangente. O tempo de resposta para o reconhecimento de sinais de LIBRAS através da câmera do celular foi avaliado e demonstrou um desempenho rápido o suficiente para suportar a comunicação em tempo real.

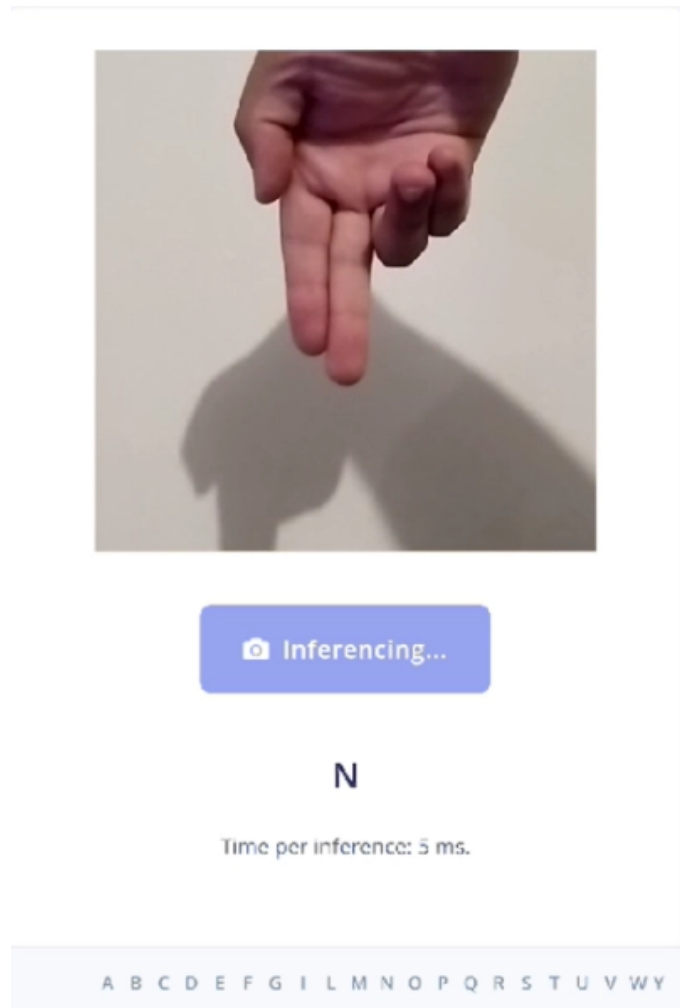


Figura 3. Letra N

Embora os resultados tenham sido promissores, identificaram-se algumas limitações durante a análise. Em certas situações de iluminação extrema ou quando os sinais eram realizados em velocidades muito altas, houve uma diminuição na precisão do modelo. Essas observações indicam áreas potenciais para melhorias e otimizações futuras.

Além da análise técnica, o projeto foi avaliado quanto ao seu impacto social. A implementação bem-sucedida desse sistema de detecção de LIBRAS promove a inclusão e a acessibilidade, possibilitando uma comunicação efetiva para indivíduos surdos ou com deficiência auditiva. Essa dimensão social ressalta a importância e o potencial transformador dessa tecnologia na promoção da igualdade de oportunidades.

## V. CONCLUSÃO

Este estudo explorou as aplicações inovadoras da inteligência artificial na detecção de sinais de LIBRAS, focalizando na implementação eficiente dessa tecnologia por meio do *Edge Impulse*. Ao longo deste projeto, foi evidente o potencial transformador dessa abordagem para superar as

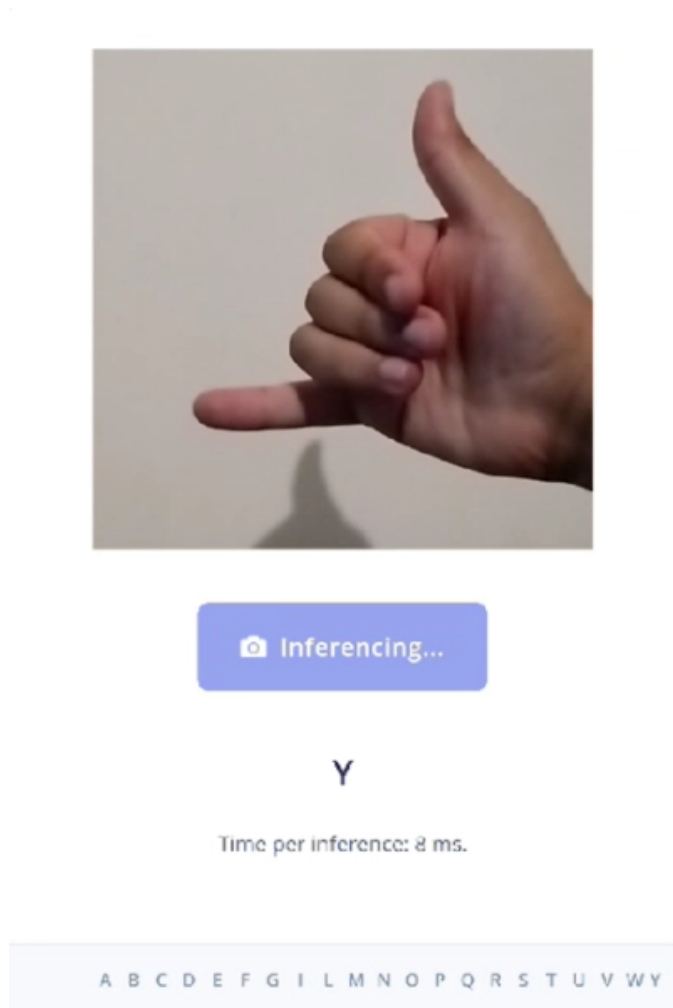


Figura 4. Letra Y

barreiras comunicativas enfrentadas pela comunidade surda, promovendo inclusão e acessibilidade.

A implementação de redes neurais no *Edge Impulse* revelou-se uma estratégia eficaz para a identificação precisa de sinais de LIBRAS, constituindo uma ferramenta poderosa para fortalecer a comunicação inclusiva. A capacidade do modelo em reconhecer uma diversidade de gestos contribui diretamente para aprimorar a interação entre indivíduos surdos e ouvintes, rompendo com barreiras linguísticas persistentes.

A escolha estratégica do *Edge Impulse* como plataforma de implementação destacou-se pela eficiência do projeto. Ao permitir a execução do modelo diretamente em *smartphones*, o sistema ofereceu uma experiência ágil para o usuário e evidenciou a praticidade e a aplicabilidade do modelo em ambientes do cotidiano. A implementação em tempo real, através da câmera do celular, sublinha o potencial deste projeto para aplicação prática.

Embora os resultados tenham sido encorajadores, foram identificados desafios, como variações de iluminação, que afetaram a precisão do modelo em contextos específicos. Essas observações abrem caminho para possíveis aprimoramentos

futuros, seja por meio de ajustes nos algoritmos, ampliação do conjunto de dados ou otimizações direcionadas a cenários mais desafiadores.

A linguagem de sinais vai além do alfabeto, incorporando uma variedade de gestos que representam palavras e conceitos. Diante dessa complexidade, uma potencial área de aprimoramento para trabalhos futuros consiste na inclusão de vídeos provenientes de *datasets* pré-existentes. Essa abordagem permitiria que o modelo fosse treinado para classificar não apenas letras, mas também a vasta gama de gestos que compõem a linguagem de sinais. Isso enriqueceria a capacidade do modelo em interpretar e compreender a linguagem de sinais em sua totalidade, expandindo suas capacidades além da simples identificação do alfabeto manual.

Além da dimensão técnica, a análise do impacto social destacou a relevância significativa deste projeto. A capacidade de facilitar a comunicação efetiva para a comunidade surda não é apenas uma inovação tecnológica; é um instrumento para a promoção da igualdade de oportunidades. A implementação bem-sucedida do sistema reforça a ideia de que a tecnologia pode desempenhar um papel crucial na construção de uma sociedade mais inclusiva.

Este projeto representa um avanço significativo em direção a um futuro no qual as barreiras linguísticas deixarão de ser obstáculos à comunicação. A convergência entre inteligência artificial, redes neurais e a implementação no *Edge Impulse* oferece uma base sólida para a contínua evolução dessa tecnologia. À medida que a pesquisa e o desenvolvimento avançam, torna-se crucial manter uma colaboração contínua com a comunidade surda, garantindo que as soluções sejam verdadeiramente inclusivas e abordem as diversas nuances da linguagem gestual.

## REFERÊNCIAS

- [1] M. G. D. A. M. e Batista, "A importância da comunicação em libras para o surdo brasileiro," 2018, Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/14177>
- [2] L. C. Gonçalves, E. F. Saad, R. B. Andrade, B. A. Romero, and R. D. Campos, "Redes Neurais Artificiais e Processamento de Imagem no Reconhecimento de LIBRAS, Usando o Kinect," *Jornal de Engenharia, Tecnologia e Meio Ambiente* XI Computer on the Beach. (JETMA), vol. 1, no. 1, 2016.
- [3] M. F. O. de M. Santos and C. C. Quarto, "Classificação de libras em imagens através de redes neurais convolucionais: Libra classification in images through convolutional neural networks", *Braz. J. Develop.*, vol. 8, no. 12, pp. 80897–80915, Dec. 2022.
- [4] I. G. O. Bastos, "Reconhecimento de sinais da Libras utilizando descritores de forma e redes neurais artificiais," *Dissertação de Mestrado, Universidade Federal da Bahia, Mestrado Multiinstitucional de Pós-Graduação em Ciência da Computação*, jun. 2016. [Online]. Disponível: <http://repositorio.ufba.br/ri/handle/ri/19374>
- [5] J. P. C. Sobrinho, L. O. P. H. da Silva, G. C. B. C. Dalpra, and S. C. A. Basilio, "Processo para Reconhecimento e Tradução de Sinais em LIBRAS Utilizando Redes Neurais Artificiais," *XI Computer on the Beach*, pp. 084-086, 2020. DOI: 10.14210/cotb.v11n1.p084-086.