

Plataforma Colaborativa para Validação de Dados em Reconhecimento de Sinais

George Henrique Santos, Carla Diacui Medeiros Berkenbrock,
Fabiola Sucupira Ferreira Sell

Universidade do Estado de Santa Catarina (UDESC)
Programa de Pós-Graduação em Ensino de Ciências, Matemática e
Tecnologias (PPGECMT)
Joinville - SC - Brasil

georgehenriquessantos@gmail.com,

{carla.berkenbrock, fabiola.sell}@udesc.br

Abstract. *The lack of accessibility for the deaf community is aggravated by the critical data bottleneck for training Libras recognition systems. The central problem lies in the fact that current datasets depend on specialists, which can generate overfitting in real-world scenarios and fail to represent the natural variations (in the wild) of everyday users. The objective of this research design is to validate the hypothesis that training with large-scale collaborative data, even if noisy, can achieve or even surpass the robustness of models trained purely on specialist datasets. As a solution, a collaborative ecosystem is proposed in which volunteers reproduce reference signs, providing the necessary real-world variability. The plan integrates the use of Reinforcement Learning with Human Feedback (RLHF) for continuous refinement of the models and tests in a real scenario.*

Resumo. *A falta de acessibilidade para a comunidade surda é agravada pelo gargalo crítico de dados para treinar sistemas de reconhecimento de Libras. O problema central reside no fato de que as bases atuais dependem de especialistas, o que pode gerar overfitting em cenários reais e não contemplar as variações naturais de usuários cotidianos. O objetivo deste desenho de pesquisa é validar a hipótese de que o treinamento com dados colaborativos em larga escala, ainda que ruidosos, pode atingir ou superar a robustez de modelos treinados puramente com bases especialistas. Como solução, propõe-se um ecossistema colaborativo onde voluntários reproduzem sinais de referência, fornecendo a variabilidade real necessária. O planejamento integra o uso de Aprendizado por Reforço com Feedback Humano (RLHF) para refinamento contínuo dos modelos a partir da validação humana e testes dos algoritmos em um cenário real.*

1. Introdução

A Língua Brasileira de Sinais (Libras) é o meio de comunicação da comunidade surda no Brasil, reconhecido pela Lei [BRASIL 2002]. Assim, o desenvolvimento de tecnologias que auxiliem na sua disseminação e compreensão é uma demanda social e

científica relevante [de Freitas 2025]. Contudo, depara-se com um desafio: a escassez e a artificialidade dos dados disponíveis.

Embora o idioma americano (ASL) tenha amenizado esse gargalo por meio de volumes massivos de dados [Tao et al. 2024], a realidade brasileira ainda se apoia em bases controladas [da Silva Almeida et al. 2022]. Nesse sentido, grandes Modelos de Linguagem (LLMs) já comprovaram que dados imperfeitos e ruidosos da internet, quando em larga escala, produzem sistemas altamente robustos. No entanto, a aplicação desse paradigma sociotécnico para Libras permanece inexplorado.

A construção de uma plataforma colaborativa justifica-se por ser um meio viável e de fácil montagem, utilizando-se tecnologias web como HTML5, CSS3 e JavaScript. Além de favorecerem a implementação, tais tecnologias fazem parte da expertise dos autores. Dessa forma, busca-se capturar a variabilidade *in the wild* (isto é, variações de iluminação, sotaques regionais, hesitações e imperfeições motoras), que muitas vezes está ausente nos *datasets* de estúdio. Para estruturar o desenho de pesquisa, estabelecem-se as seguintes Questões de Pesquisa (RQ):

RQ Central: Modelos treinados com dados ruidosos podem atingir performance similar a modelos treinados com bases especialistas?

RQ1 (Degradação): Algoritmos de reconhecimento de sinais sofrem degradação significativa de performance ao saírem de ambientes controlados para cenários reais com execuções erráticas?

RQ2 (RLHF): Constatada a degradação (RQ1), o Aprendizado por Reforço com Feedback Humano (RLHF) é capaz de recuperar e otimizar a performance desses modelos em ambientes reais? (*Nota: Esta questão pressupõe que a eficácia original dos modelos sofra queda in the wild; caso a eficácia permaneça muito alta, pode não ser possível verificar, estatisticamente, o impacto do RLHF.*)

1.1. Proposta e Abordagem

Para responder às questões de pesquisa, propõe-se:

1. **Seleção de algoritmos via Revisão Sistemática de Literatura (RSL):** Serão identificados, na literatura, algoritmos de reconhecimento de sinais isolados (*Isolated Sign Recognition*, ISR) que se destacam por dois critérios: (i) alta acurácia, seguindo protocolos rigorosos (*leave-one-person-out*, métricas como F1-score, WER, CER) [Tao et al. 2024]; ou (ii) alta eficiência computacional - menor custo de inferência, como redes leves do tipo MobileNet [Howard et al. 2017]. A razão para escolher também por eficiência é a aplicabilidade, pois modelos pesados podem não ser úteis em situações cotidianas devido ao alto custo de processamento, exigindo versões mais leves para aplicação/implementação [Howard et al. 2017].
2. **Desenvolvimento da Plataforma Colaborativa:** Será desenvolvida uma plataforma web (com elementos de jogos) que utiliza o Dicionário da Língua Brasileira de Sinais do INES [INES 2011] como fonte de vídeos de referência produzidos por especialistas. Usuários voluntários serão convidados a escolher categorias temáticas (Saúde, Educação, Tecnologia etc.) e a reproduzir os sinais assistindo aos vídeos.

2. Trabalhos Correlatos

Rezende, Almeida e Guimarães [Rezende et al. 2021] apresentaram a criação e validação de uma base pública de Libras utilizando sensores RGB-D, estabelecendo protocolos de gravação rigorosos com doze sinalizadores. Visando otimizar o processamento, Alves, Boldt e Paixão [Alves et al. 2024] propuseram a representação de sinais por meio de imagens de esqueletos (*skeletons*), alcançando alta eficiência e superando o estado da arte em bases isoladas.

Nesse sentido, esses trabalhos representam uma das lacunas que este projeto visa preencher: a dependência de ambientes artificiais e dados sintéticos/especialistas. Tao et al. [Tao et al. 2024] fornecem a prova de conceito de que, em ASL, a barreira da acurácia só foi rompida quando o gargalo de dados foi resolvido por meio de volume escalável. Esta pesquisa, em consonância com o supracitado, propõe um meio de capturar dados de forma massiva, colaborativa e com ruídos.

3. Metodologia e Desenho de Pesquisa

O presente estudo caracteriza-se como uma Iniciação Científica baseada em um desenho experimental em três fases focadas em coleta, estresse algorítmico e validação de hipóteses por meio de uma plataforma web gamificada (Figura 1). A plataforma exibirá sinais do INES [INES 2011] para que voluntários os reproduzam, gerando o *dataset* colaborativo, conforme mostra a Figura 1.

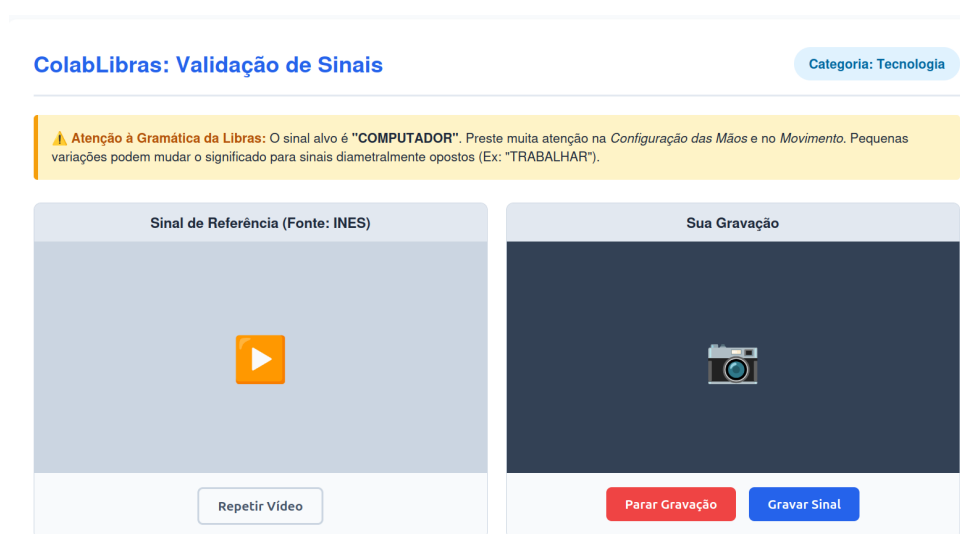


Figura 1. Protótipo da interface de coleta e validação, ilustrando a referência do sinal.

Fase 1: Benchmarking de Estresse Real - O primeiro passo metodológico é avaliar a queda de acurácia (*accuracy drop*) de algoritmos existentes quando expostos a usuários comuns. Serão selecionados via Revisão Sistemática (RSL) dois perfis de algoritmos: (i) Modelos SOTA (alta eficácia, alto custo computacional) e (ii) Modelos eficientes (menor processamento, como redes baseadas em MobileNets) [Howard et al. 2017]. Os sinais submetidos pelos usuários serão processados por esses algoritmos, e a discrepância entre os resultados laboratoriais dos autores originais e o cenário *in the wild* responderá à **RQ1**.

Fase 2: Ciclo de Otimização RLHF - Para responder à **RQ2**, será implementado um protocolo de *Reinforcement Learning With Human Feedback* [Sutton and Barto 2018]. Após o modelo tentar classificar o vídeo do voluntário, o próprio usuário validará se o sinal foi reconhecido corretamente. Esse *feedback* binário atuará como função de recompensa para o *fine-tuning* dos pesos da rede neural, ajustando os modelos à variabilidade real da sociedade.

Fase 3: Treinamento “Do Zero”(Validação da Hipótese Central) - A fase mais crítica do planejamento consiste em isolar o *dataset* colaborativo construído nas fases anteriores. As mesmas redes neurais testadas na Fase 1 serão treinadas desde o início (pesos aleatórios) exclusivamente com os dados ruidosos e em massa obtidos pela plataforma. O desempenho final desses modelos será estatisticamente comparado com os modelos treinados nas bases originais (especialistas), permitindo validar se a colaboração em escala supera a curadoria restrita.

4. Considerações Finais

Este trabalho propõe uma abordagem para o enfrentamento do gargalo de dados no reconhecimento de Libras, deslocando o foco de bases especialistas restritas para uma plataforma colaborativa de larga escala. A principal contribuição reside na proposição de um ecossistema que utiliza o *Reinforcement Learning with Human Feedback* (RLHF) para, inclusive, refinar modelos a partir da variabilidade real de usuários não especialistas.

Espera-se que este desenho metodológico resulte nas seguintes contribuições técnicas e acadêmicas após sua execução:

1. **Protocolo de validação via RLHF** aplicado especificamente à superação de sutilezas e imperfeições motoras na sinalização de Libras.
2. **Construção e disponibilização de um Dataset *In the Wild*** de Libras, refletindo condições de iluminação, enquadramento e execuções do mundo real.
3. **Estudo de eficiência computacional vs. colaboração humana.**
4. **Estresse real de algoritmos**, revelando a verdadeira acurácia deles quando submetidos a dados não controlados.

Agradecimentos

Ao apoio financeiro da Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina – FAPESC. Os autores também agradecem às ferramentas de inteligência artificial Gemini (Google) e Poe pelo suporte na revisão gramatical.

Referências

- Alves, C. E. G. R., de Assis Boldt, F., and Paixão, T. M. (2024). Enhancing Brazilian Sign Language recognition through skeleton image representation. *arXiv preprint arXiv:2404.19148*.
- BRASIL (2002). *Lei nº 10.436, de 24 de abril de 2002. Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências*. Brasil, Brasília, DF.
- da Silva Almeida, A. C., Rezende, T. M., and Almeida, S. G. M. (2022). Ampliação da base MINDS-Libras: um estudo de aplicação de técnicas de aumento sintético de dados e da inclusão de novos conjuntos de vídeos disponíveis na literatura. In *Anais do XXIV Congresso Brasileiro de Automática (CBA)*, Ouro Preto, MG. SBA.

- de Freitas, L. E. R. (2025). Aplicação de metodologias Ágeis no desenvolvimento de um produto mínimo viável acessível para conectar pessoas com deficiência a prestadores de serviços.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- INES (2011). Dicionário da Língua Brasileira de Sinais. Instituto Nacional de Educação de Surdos.
- Rezende, T. M., Almeida, S. G. M., and Guimarães, F. G. (2021). Development and validation of a Brazilian sign language database for human gesture recognition. *Neural Computing and Applications*, 33:10449–10467.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2 edition.
- Tao, T., Zhao, Y., Liu, T., and Zhu, J. (2024). Sign language recognition: A comprehensive review of traditional and deep learning approaches, datasets, and challenges. *IEEE Access*, 12:75034–75060.