ELOS: Sistema Embarcado de Tradução Bidirecional LIBRAS-Português para os Setores Administrativos de Instituições de Ensino

Carlos Eduardo Silva Correia[1], Guilherme Lago dos Santos[1], Tiago da Silva Gonçalves[1], Yanes Costa Nascimento[1], Ivanildo Gomes da Silva[1]

[1]Curso Técnico em Desenvolvimento de Sistemas, Serviço Nacional de Aprendizagem Industrial (SENAI), Unidade Feira de Santana, Feira de Santana, BA, Brasil

{carloscorreia.tech, guilherme.lago.santos, silvagoncalves0503, yanes.cnascimento, invanidlo99gomes}@gmail.com

Abstract. *This paper presents ELOS, an embedded system with artificial intelligence for real-time bidirectional translation between Brazilian Sign Language (LIBRAS) and Portuguese, designed for the administrative areas of educational institutions, such as registrar, reception, coordination, and library, where deaf students currently lack communication mediation. ELOS does not replace the human interpreter; it provides support where the interpreter cannot be. Gesture recognition relies on computer vision with MediaPipe and a rule-based classifier, with voice synthesis as output, targeting the Raspberry Pi 4 as the deployment platform. The functional MVP, at Sprint 5, recognizes 58 gestures from an administrative vocabulary in real time. Preliminary results indicate technical and economic viability for large-scale adoption in educational institutions.*

Resumo. *Este artigo apresenta o ELOS, um sistema embarcado com inteligência artificial para tradução bidirecional em tempo real entre a Língua Brasileira de Sinais (LIBRAS) e o português, voltado aos setores administrativos de instituições de ensino, como secretaria, recepção, coordenação e biblioteca, onde, atualmente, o aluno surdo permanece sem mediação comunicacional. A proposta não substitui o intérprete; é o suporte onde ele não pode estar. O reconhecimento gestual emprega visão computacional com MediaPipe e um classificador baseado em regras, com saída por síntese de voz, tendo o Raspberry Pi 4 como plataforma-alvo de implantação. O MVP funcional, na Sprint 5, reconhece 58 gestos de um vocabulário administrativo em tempo real. Os resultados preliminares indicam viabilidade técnica e econômica para adoção em escala em instituições de ensino.*

1. Introdução

A comunicação é essencial para a participação social, mas permanece um desafio para a comunidade surda, que depende da língua de sinais para interagir com o mundo ouvinte [Rastgoo et al. 2021]. A Língua Brasileira de Sinais (LIBRAS) é a língua da comunidade surda no Brasil, porém as soluções tecnológicas capazes de reduzir a barreira comunicacional entre surdos e ouvintes ainda são escassas [Silva et al. 2022]. A inclusão digital das pessoas surdas enfrenta grandes barreiras justamente pela falta de tecnologias acessíveis de tradução automática de LIBRAS [Alves et al. 2025].

Em instituições de ensino, os serviços de interpretação humana são limitados em disponibilidade e custo, além de dependerem da presença física do intérprete [Alsharif et al. 2025]. Quando presente, o intérprete atua principalmente na sala de aula, deixando descobertos os setores administrativos, como secretaria, recepção, coordenação e biblioteca, onde o aluno surdo precisa resolver matrículas, consultar documentos e obter orientações. Nesses espaços, o aluno frequentemente recorre à mímica ou a bilhetes escritos, o que compromete sua autonomia e sua permanência na instituição.

Este trabalho apresenta o ELOS, uma solução embarcada com inteligência artificial instalada de forma permanente nos setores administrativos de instituições de ensino, capaz de realizar tradução bidirecional em tempo real entre LIBRAS e português, sem necessidade de aplicativo, cadastro ou treinamento prévio do usuário. O escopo do sistema é delimitado: o ELOS não substitui o intérprete, e sim atua como suporte onde ele não pode estar, não se destinando à sala de aula. O artigo descreve a arquitetura da solução, a metodologia de desenvolvimento e os resultados parciais obtidos no MVP funcional demonstrável.

2. Trabalhos Relacionados

O reconhecimento automático de língua de sinais evoluiu de forma expressiva com o avanço da aprendizagem profunda, hoje predominante na área [Rastgoo et al. 2021, Al-Qurishi et al. 2021]. Boa parte dos trabalhos recentes adota a extração de pontos-chave da mão com a biblioteca MediaPipe [Zhang et al. 2020], seguida por classificadores baseados em redes neurais. Nessa linha, sistemas de reconhecimento em tempo real foram propostos para diferentes línguas de sinais, como a americana [Alsharif et al. 2025], a indiana [Khartheesvar et al. 2023] e a assamesa [Bora et al. 2023], combinando MediaPipe com arquiteturas de aprendizado profundo. Redes convolucionais tridimensionais [Al-Hammadi et al. 2020] e conjuntos de dados multimodais de larga escala [Sincan and Keles 2020, Sridhar et al. 2020] também foram explorados para ampliar a robustez do reconhecimento.

No contexto brasileiro, diversos trabalhos investigam o reconhecimento de LIBRAS. Silva et al. [2022] compararam arquiteturas LSTM e BiLSTM para reconhecimento contínuo de sinais em tempo real, obtendo melhor desempenho com a LSTM. Carvalho et al. [2021] propuseram uma arquitetura de aprendizado profundo para o reconhecimento de configurações de mão em LIBRAS, enquanto Alves et al. [2025] avaliaram modelos de aprendizado de máquina sobre pontos-chave extraídos com MediaPipe, com destaque para o XGBoost. A construção de bases de dados próprias de LIBRAS também tem sido apontada como fator crítico para o desempenho dos modelos [Lobo-Neto and Pedrini 2024].

A tradução automática de língua de sinais, por sua vez, ainda enfrenta desafios de naturalidade e generalização [Núñez-Marcos et al. 2023, Liang et al. 2023]. Aplicações voltadas ao atendimento, como sistemas de apoio à consulta médica de pessoas surdas [Xia et al. 2022] e ao atendimento em situações críticas [Areeb et al. 2022], reforçam o

potencial dessas tecnologias em contextos de serviço. O ELOS se diferencia por combinar tradução bidirecional, operação autônoma sem dispositivo pessoal e instalação fixa especificamente nos setores administrativos de instituições de ensino, um nicho ainda pouco explorado pela literatura.

3. Arquitetura e Metodologia

3.1. Componentes e Fluxo de Comunicação

O ELOS é estruturado como um sistema embarcado composto por componentes comerciais de fácil aquisição: uma unidade de processamento Raspberry Pi 4 (plataforma-alvo de implantação), câmera HD USB, microfone direcional e uma tela sensível ao toque de dimensões reduzidas. O fluxo de comunicação opera em dois sentidos. No sentido de LIBRAS para voz, a câmera capta os gestos do aluno surdo, a inteligência artificial reconhece o sinal e o sistema sintetiza a mensagem em áudio para o atendente. No sentido de voz para LIBRAS, o atendente responde por voz e o sistema apresenta a resposta em texto na tela, com exibição de sinais como evolução planejada. A Figura 1 sintetiza esses dois fluxos e os módulos envolvidos.

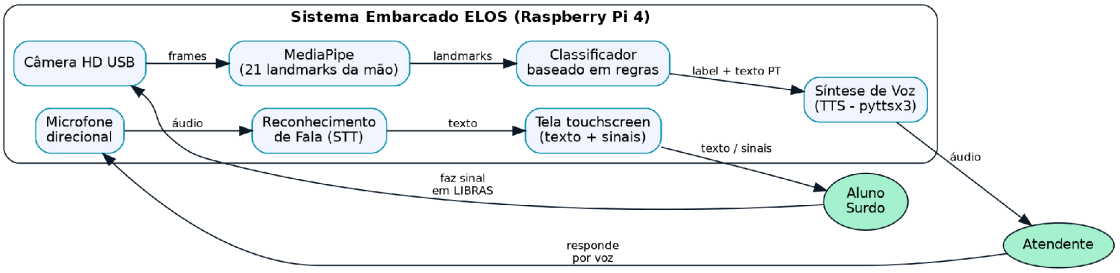


Figura 1. Arquitetura do sistema embarcado ELOS e os dois fluxos de tradução: de LIBRAS para voz e de voz para LIBRAS.

3.2. Arquitetura de Software

A detecção gestual emprega o MediaPipe [Zhang et al. 2020], que identifica 21 pontos-chave da mão em tempo real. Essa representação compacta e robusta do gesto é também explorada em abordagens baseadas em esqueleto da mão [Rastgoo et al. 2020]. No MVP atual, a classificação dos sinais é realizada por um classificador baseado em regras, que mapeia padrões de dedos abertos e fechados para rótulos e textos em português. A síntese de voz utiliza uma biblioteca de código aberto com processamento local, sem dependência de internet, enquanto a captação da fala do atendente usa reconhecimento automático de voz. Como evolução, prevê-se a substituição do classificador por regras por um modelo de rede neural recorrente treinado com auxílio de intérpretes reais, conforme detalhado na Seção 5.

Para garantir representatividade e eficácia em condições reais, como iluminação variável, fundos não controlados e diferentes ângulos, um conjunto de dados proprietário de LIBRAS está sendo construído com a participação direta de usuários surdos. Esse conjunto, coletado no próprio ambiente de operação, constitui um ativo estratégico e uma barreira de replicação frente a bases públicas obtidas em laboratório.

3.3. Especificação e Desenvolvimento

O desenvolvimento segue um processo iterativo organizado em Sprints, com especificação formal de 8 Requisitos Funcionais e 9 Requisitos Não Funcionais. Entre os requisitos não funcionais prioritários estão a precisão mínima de reconhecimento gestual de 85%, a disponibilidade do sistema de 95% e o tempo de resposta de até 2 segundos por interação. A validação da qualidade tem como critério central a precisão do reconhecimento aferida em condições reais de uso com usuários surdos.

4. Resultados Parciais e Viabilidade

O estágio atual do projeto corresponde a um MVP funcional demonstrável, concluído na Sprint 5, capaz de reconhecer 58 gestos de um vocabulário orientado ao atendimento administrativo, em tempo real. A Figura 2 mostra o sistema em operação, reconhecendo o sinal correspondente a *BOM* a partir dos pontos-chave da mão detectados pelo MediaPipe, com a interface exibindo o estado dos dedos, a transcrição e a saída de voz. A Tabela 1 apresenta a distribuição desse vocabulário por categoria.

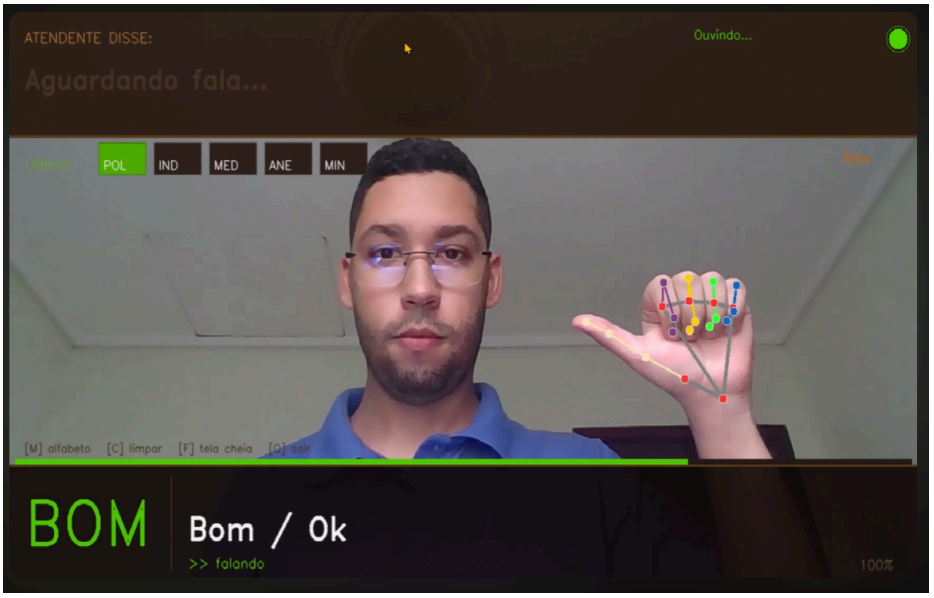


Figura 2. MVP do ELOS em operação: detecção de pontos-chave da mão e reconhecimento de um sinal em tempo real.

Tabela 1. Vocabulário de 58 gestos reconhecidos pelo MVP, por categoria.

Categoria	**Exemplos de sinais**
Cumprimentos	OLÁ, TCHAU, OBRIGADO, POR_FAVOR, DESCULPA
Respostas básicas	SIM, NÃO, ENTENDI, REPETIR, DEVAGAR, TALVEZ
Comunicação	AJUDA, ESPERA, JÁ_VOLTO, AGORA, DEPOIS
Apresentação pessoal	NOME, ALUNO, PROFESSOR, SENAI, TRABALHO, ESTUDO
Localização e tempo	ONDE, AQUI, PERTO, HOJE, AMANHÃ, ONTEM
Necessidades	BANHEIRO, COMER, BEBER, ÁGUA

Setores administrativos	SECRETARIA, MATRÍCULA, CADASTRO, DOCUMENTO, PROTOCOLO, PRAZO, PAGAMENTO, BOLETO
Acadêmico	AULA, PROVA, NOTA, BIBLIOTECA, LABORATÓRIO, REFEITÓRIO, COORDENAÇÃO
Números	UM, DOIS, TRÊS, CINCO

As métricas de precisão de 85%, disponibilidade de 95% e tempo de resposta de até 2 segundos constituem metas declaradas do sistema, a serem aferidas na fase de validação com usuários reais, e funcionam como critério de aceitação: modelos que não as atinjam retornam para retreinamento antes de qualquer implantação.

No aspecto econômico, o uso de bibliotecas de código aberto e de hardware de baixo custo torna a solução financeiramente viável para adoção desde pequenas instituições até grandes redes de ensino, característica também apontada como vantagem em sistemas de reconhecimento de sinais voltados ao atendimento [Xia et al. 2022]. Essa relação favorável entre custo e benefício amplia o potencial de escala da solução em instituições de ensino.

5. Considerações Finais

O ELOS demonstra ser uma iniciativa de alta aplicabilidade e impacto social. Ao ocupar o vácuo comunicacional dos setores administrativos, a solução promove a autonomia do aluno surdo sem substituir o intérprete, reafirmando seu posicionamento como o suporte onde ele não pode estar.

Os próximos passos da pesquisa incluem a substituição do classificador por regras por um modelo de rede neural recorrente treinado com intérpretes reais, a adoção de reconhecimento de fala offline, o suporte simultâneo a duas mãos, a incorporação de expressões faciais para a gramática da LIBRAS e a finalização da integração entre hardware e software do sistema, seguida de testes práticos em cenários reais diversificados.

Referências

Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., & Mekhtiche, M. A. (2020). Hand gesture recognition for sign language using 3DCNN. IEEE Access, 8, 79491-79509.

Al-Qurishi, M., Khalid, T., & Souissi, R. (2021). Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. IEEE Access, 9, 126917-126951.

Alsharif, B., Alalwany, E., Ibrahim, A., Mahgoub, I., & Ilyas, M. (2025). Real-time American Sign Language interpretation using deep learning and keypoint tracking. Sensors, 25(7), 2138.

Alves, K. de M., Santos, F. J. dos, & Sousa, S. K. A. de. (2025). Abordagens de aprendizado de máquina para o reconhecimento de sinais em Libras. In Anais da XXV Escola Regional de Computação Bahia, Alagoas e Sergipe (pp. 327-335). SBC.

Areeb, Q. M., Nadeem, M., Alroobaea, R., & Anwer, F. (2022). Helping hearing-impaired in emergency situations: A deep learning-based approach. IEEE Access, 10, 8502-8517.

Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time Assamese sign language recognition using MediaPipe and deep learning. Procedia Computer Science, 218, 1384-1393.

Carvalho, G., Brandão, A., & Ferreira, F. (2021). HandArch: A deep learning architecture for LIBRAS hand configuration recognition. In Anais do XVII Workshop de Visão Computacional (pp. 19-24). SBC.

Khartheesvar, G., Kumar, M., Yadav, A. K., & Yadav, D. (2023). Automatic Indian sign language recognition using MediaPipe Holistic and LSTM network. Multimedia Tools and Applications, 83(20), 58329-58348.

Liang, Z., Li, H., & Chai, J. (2023). Sign language translation: A survey of approaches and techniques. Electronics, 12(12), 2678.

Lobo-Neto, V. C., & Pedrini, H. (2024). LSWH100: A handshape dataset for Brazilian Sign Language (Libras) using SignWriting. Data in Brief, 56, 110780.

Núñez-Marcos, A., Perez-de-Viñaspre, O., & Labaka, G. (2023). A survey on sign language machine translation. Expert Systems with Applications, 213, 118993.

Rastgoo, R., Kiani, K., & Escalera, S. (2020). Hand sign language recognition using multi-view hand skeleton. Expert Systems with Applications, 150, 113336.

Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 164, 113794.

Silva, J. L. S., Vieira, G. S., Fonseca, A. U., & Soares, F. (2022). Reconhecimento e tradução de sinais de Libras para língua portuguesa escrita usando redes neurais profundas. In Anais do Congresso Brasileiro de Automática (Vol. 3, No. 1). SBA.

Sincan, O. M., & Keles, H. Y. (2020). AUTSL: A large scale multi-modal Turkish sign language dataset and baseline methods. IEEE Access, 8, 181340-181355.

Sridhar, A., Ganesan, R. G., Kumar, P., & Khapra, M. (2020). INCLUDE: A large scale dataset for Indian sign language recognition. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 1366-1375).

Xia, K., Lu, W., Fan, H., & Zhao, Q. (2022). A sign language recognition system applied to deaf-mute medical consultation. Sensors, 22(23), 9107.

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). MediaPipe Hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214.