

LIA: Libras com Inteligência Artificial – Aplicativo para ensino de libras com reconhecimento de gestos baseado em inteligência artificial

Elaine Gomes Moreira, 081240037@faculdade.cefsa.edu.br, Faculdade Engenheiro Salvador Arena
Joyce da Costa Peres, 081240045@faculdade.cefsa.edu.br, Faculdade Engenheiro Salvador Arena
Profa. Dra. Kátia Milani Lara Bossi, pro21002390@cefsa.edu.br, Faculdade Engenheiro Salvador Arena

Resumo

Este artigo apresenta o desenvolvimento do LIA (Libras com Inteligência Artificial), um aplicativo educacional interativo voltado ao ensino da Língua Brasileira de Sinais (Libras), com foco no aprendizado de vocabulário temático e do alfabeto manual. A solução utiliza técnicas avançadas de visão computacional, por meio da biblioteca MediaPipe, combinadas a modelos de inteligência artificial do tipo Long Short-Term Memory (LSTM) para reconhecer gestos capturados pela câmera em tempo real. A escolha do LSTM deve-se à sua capacidade intrínseca de processar sequências temporais, característica fundamental dos gestos em Libras que envolvem movimentos dinâmicos com dependência temporal. A metodologia contemplou a coleta e tratamento de dados gestuais, treinamento do modelo preditivo e o desenvolvimento de uma interface gamificada em Python com Tkinter. Os resultados indicam que o sistema alcançou acurácia de 93,94% de acurácia no reconhecimento do alfabeto manual, demonstrando viabilidade técnica para a expansão para outras categorias de gestos. Conclui-se que o LIA tem potencial para complementar métodos tradicionais de ensino, ampliando o acesso à Libras de forma inclusiva e tecnológica.

Palavras-chave: Educação inclusiva. Libras. Reconhecimento de gestos. Inteligência artificial. Aprendizado interativo.

Introdução

A Língua Brasileira de Sinais (Libras) é um sistema linguístico complexo, composto por elementos visuais-gestuais que constituem a principal forma de comunicação da comunidade surda no Brasil. Reconhecida oficialmente pela Lei nº 10.436, de 24 de abril de 2002 (BRASIL, 2002), a Libras é fundamental para a inclusão social e o exercício pleno dos direitos humanos das pessoas surdas. Apesar da importância social e legal, o ensino e a disseminação da Libras ainda enfrentam desafios estruturais que dificultam o acesso amplo a essa língua. Segundo o Censo Demográfico do IBGE (2022), apenas 3% da população ouvinte brasileira apresenta proficiência básica em Libras, e a escassez de profissionais qualificados, destacada pela Associação Brasileira de Cultura Surda (ABRAÇA, 2023), limita a oferta de cursos e materiais didáticos de qualidade.

O cenário atual evidencia a necessidade de novas abordagens pedagógicas que possam ampliar o acesso e a qualidade do ensino de Libras, conforme apontado por Silveira e Barros (2021) em estudo sobre estratégias de aprendizagem ativa. A evolução das tecnologias digitais e da inteligência artificial oferece ferramentas inovadoras para apoiar esse processo, por meio da criação de sistemas que possibilitam o reconhecimento automático de gestos e a interação em tempo real. Entre as tecnologias promissoras destaca-se a visão computacional, que, aliada a modelos de aprendizado profundo, tem demonstrado alta

eficiência em tarefas complexas de reconhecimento de padrões visuais, incluindo a interpretação de sinais manuais (BASTOS et al., 2021).

Bibliotecas modernas, como a MediaPipe (ZHANG et al., 2020), framework de percepção multimodal desenvolvido pelo Google, possibilitam a captura de 21 pontos articulatorios de cada mão em tempo real, gerando dados tridimensionais precisos que alimentam modelos de inteligência artificial. Redes neurais recorrentes, especialmente LSTM (Long Short-Term Memory), são particularmente indicadas para o tratamento de sequências temporais, características fundamentais no reconhecimento de gestos dinâmicos, conforme demonstrado por Hassanat et al. (2019).

Neste contexto, o presente trabalho apresenta o LIA (Libras com Inteligência Artificial), um aplicativo educacional gamificado desenvolvido para apoiar o aprendizado de Libras por iniciantes. O aplicativo integra captura em tempo real, reconhecimento automático de gestos e uma interface orientada por níveis temáticos, combinando metodologias de ensino tradicionais com recursos tecnológicos inovadores para promover uma aprendizagem acessível e motivadora, seguindo princípios de gamificação descritos por Kapp (2012).

Metodologia

Esta pesquisa enquadra-se na natureza aplicada, pois busca gerar conhecimento com aplicação prática imediata, voltada à solução de um problema real: o reconhecimento automatizado de movimentos físicos por meio de visão computacional. De acordo com Gil (2010, p. 27), "a pesquisa aplicada é motivada pela necessidade de resolver problemas concretos, mais imediatos ou não". Quanto à abordagem, adota-se o método quantitativo, uma vez que os dados coletados e processados serão analisados por meio de métricas objetivas, como acurácia, precisão, recall e F1-score, derivadas de testes experimentais. Para Prodanov e Freitas (2013, p. 70), "a pesquisa quantitativa considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las". Os objetivos desta investigação são exploratórios e descritivos. É exploratória na fase inicial, pois visa investigar ferramentas, bibliotecas e técnicas adequadas para o reconhecimento de movimentos, sem buscar conclusões definitivas de imediato. Conforme Gil (2010, p. 41), "a pesquisa exploratória tem como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos". É também descritiva, pois visa caracterizar o funcionamento do sistema e descrever seu desempenho. Segundo Cervo, Bervian e Silva (2007, p. 61), "a pesquisa descritiva observa, registra, analisa e correlaciona fatos ou fenômenos sem manipulá-los". A metodologia adotada é experimental, pois envolve a implementação e teste de um protótipo para verificar sua eficácia em condições controladas. De acordo com Gil (2010, p. 47), "a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto".

Nesse sentido, em termos de coleta e pré-processamento, para validar a arquitetura do sistema, optou-se por iniciar com o reconhecimento do alfabeto manual em Libras, composto por 26 letras (A-Z). Dessa forma, foram coletadas 30 amostras para cada letra, totalizando 780 sequências temporais. Faz-se relevante salientar que as gravações de coletas foram realizadas com um único participante em ambiente controlado, com iluminação superior a 300 lux e fundo neutro, utilizando webcam Logitech C920 com resolução de 1280×720 pixels.

Ademais, a biblioteca MediaPipe Hands foi utilizada para extrair coordenadas tridimensionais dos 21 pontos articulatorios de cada mão. Nesse contexto, é importante ressaltar que cada amostra consiste em

uma sequência de 30 frames com taxa de captura de 30 FPS, resultando em vetores de 126 características por frame (21 pontos $\times$ 3 coordenadas $\times$ 2 mãos). Para garantir a uniformidade dos dados, todas as sequências foram padronizadas para 30 frames através de preenchimento com zeros quando necessário. Assim, visando aumentar a robustez do modelo, aplicou-se técnicas de data augmentation gerando 5 variações para cada amostra original através de: rotações aleatórias de $\pm 15^\circ$, variações de escala entre $\pm 10\%$ e deslocamentos espaciais de até 5% da dimensão da imagem. Esta abordagem ampliou o dataset efetivo para 3.900 amostras, simulando variações naturais de execução dos sinais.

Com o objetivo de desenvolver um treinamento e avaliação do modelo preditivo, implementou-se uma arquitetura LSTM com 2 camadas (128 e 64 unidades), camada densa de 64 neurônios com ativação ReLU, dropout de 0.3 entre camadas, e camada de saída com ativação Softmax. Além disso, é importante compreender que o modelo foi compilado com otimizador Adam (learning rate de 0.001) e função de perda `sparse_categorical_crossentropy`. Ressalta-se que, os dados foram divididos estratificadamente em: 70% treino (546 amostras), 15% validação (117 amostras) e 15% teste (117 amostras). O treinamento foi conduzido por 30 épocas com batch size de 32, utilizando early stopping baseado na acurácia de validação com paciência de 10 épocas. Dessa forma, foram empregadas as seguintes métricas para avaliação do desempenho: acurácia, precisão, recall, F1-score e matriz de confusão. Faz-se necessário identificar que a latência média de processamento identificada foi medida para garantir viabilidade de aplicação em tempo real.

Torna-se importante desenvolver que, a interface do LIA foi desenvolvida em Python utilizando Tkinter para os componentes gráficos e OpenCV para captura e exibição de vídeo. A arquitetura modular inclui: módulo de captura (MediaPipe), módulo de reconhecimento (LSTM), módulo de interface (Tkinter) e módulo de gamificação. Ela incorpora: captura e exibição de vídeo em tempo real, processamento simultâneo dos dados, módulos temáticos organizados em níveis de dificuldade progressiva, sistema de feedback instantâneo visual e textual, e elementos de gamificação como pontuação progressiva, temporizadores, desbloqueio gradual de conteúdos e recompensas virtuais.

Nesse contexto, a integração e testes do sistema foi realizada em ciclos iterativos, adotando metodologia incremental para garantir estabilidade. Foram conduzidos testes técnicos incluindo: avaliação de performance computacional (tempo de resposta, consumo de memória), testes de robustez em diferentes condições de iluminação e verificação da taxa de frames por segundo (FPS) em tempo real.

Assim como pode ser observado na Tabela 1, o modelo LSTM alcançou um desempenho notável, com 93,94% de acurácia nos conjuntos de validação e teste, indicando uma robustez satisfatória e uma generalização adequada para novos dados. A acurácia ligeiramente superior no conjunto de treino (95,45%) é esperada, sendo a pequena diferença entre elas (aproximadamente 1,5 ponto percentual) um forte indicativo de que o modelo não sofreu de overfitting significativo. O tempo médio de inferência de 59 milissegundos por amostra atesta a viabilidade da aplicação para o funcionamento em tempo real, um requisito crítico para a experiência interativa do usuário.

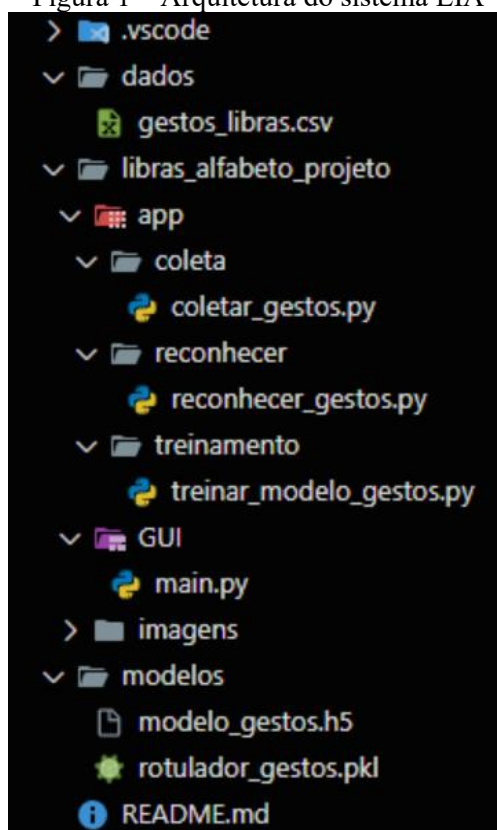
Tabela 1 - Métricas de desempenho do modelo LSTM no reconhecimento do alfabeto manual

Conjunto de Dados	Acurácia	Perda (Loss)	Tempo de Inferência
Treino	95,45%	0,1648	-
Validação	93,94%	0,3159	-
Teste	93,94%	0,3159	59ms

Fonte: Autoria própria (2025)

A Figura 1 ilustra a arquitetura modular do sistema LIA, a qual fora desenvolvida no Visual Studio Code:

Figura 1 – Arquitetura do sistema LIA



Fonte: Autoria própria (2025)

A integração e testes do sistema foram realizados em ciclos iterativos, adotando metodologia incremental para garantir a estabilidade do aplicativo. Foram conduzidos testes técnicos, que incluíram avaliação de performance computacional (tempo de resposta, consumo de memória), testes de robustez em diferentes condições de iluminação e verificação da taxa de frames por segundo (FPS) em tempo real.

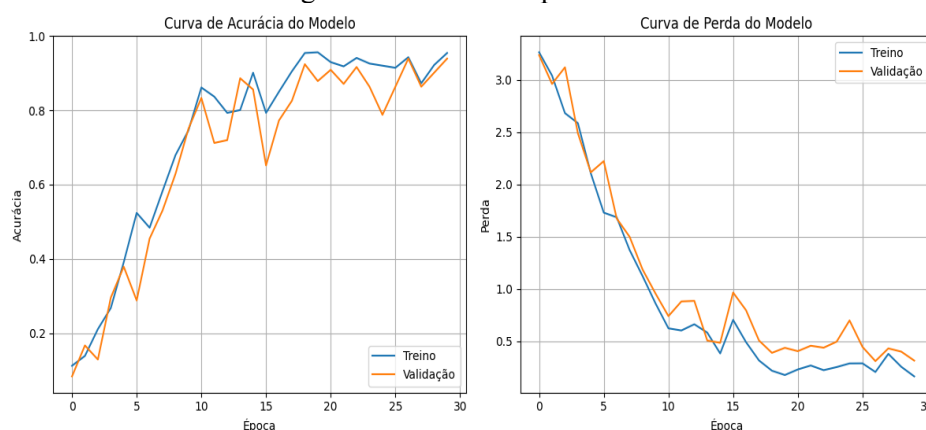
Resultados e discussão

Os testes realizados evidenciaram o funcionamento do modelo preditivo aplicado ao reconhecimento de gestos em Libras. Durante a fase de experimentação, foram adotados procedimentos padronizados para coleta de dados, incluindo o posicionamento das câmeras a uma distância variável e o enquadramento centralizado das mãos, de forma a reduzir ruídos visuais e manter a uniformidade entre as amostras. O classificador LSTM apresentou desempenho robusto no reconhecimento do alfabeto manual, atingindo 93,94% de acurácia no conjunto de teste.

A curva de aprendizado (Figura 2) mostra convergência estável, com acurácia de treino atingindo 95,45% e acurácia de validação estabilizando em 93,94% na época final. A pequena diferença entre as

curvas ($\approx 1,5\%$) indica boa generalização do modelo sem overfitting significativo. O tempo médio de inferência de 59ms por batch demonstra a viabilidade para aplicações em tempo real.

Figura 2 – Curva de Aprendizado



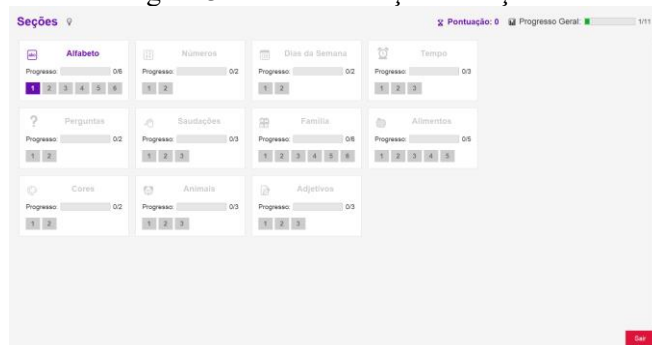
Fonte: Autoria própria (2025)

Observou-se que, para sinais com formas bem definidas e consistentes entre usuários, como as vogais A, E, I, O, U, o índice de acertos aproximou-se de 97%, enquanto sinais com maior variação posicional ou que exigiam movimentos complexos obtiveram desempenho menos consistente. A análise qualitativa dos resultados indica que os erros se concentram em pares de letras que compartilham configurações de mão semelhantes, exigindo do modelo a detecção de nuances sutis na posição dos dedos, um desafio consistentemente reportado na literatura da área (BASTOS et al., 2021).

A interface desenvolvida para este projeto se mostrou uma ferramenta com potencial pedagógico. A progressão por níveis, aliada ao sistema de feedback instantâneo, favoreceu o aprendizado incremental e contribuiu para a fixação dos sinais.

A Figura 3 apresenta a tela inicial do aplicativo, onde o usuário seleciona a seção de aprendizado desejada assim como seu respectivo nível de aprendizado e a visualização de seu progresso entre as seções e sua pontuação geral na aplicação:

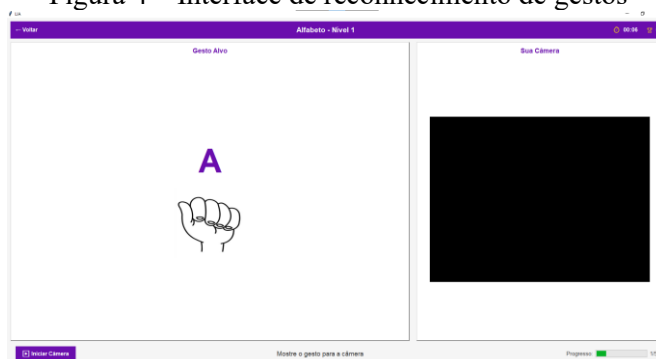
Figura 3 – Tela de seleção de seções



Fonte: Autoria própria (2025)

A Figura 4 ilustra a interface de aprendizado ativo, composta por um comando de qual sinal ele deve realizar assim como sua representação gramatical na língua portuguesa e sua representação gesto-visual em libras:

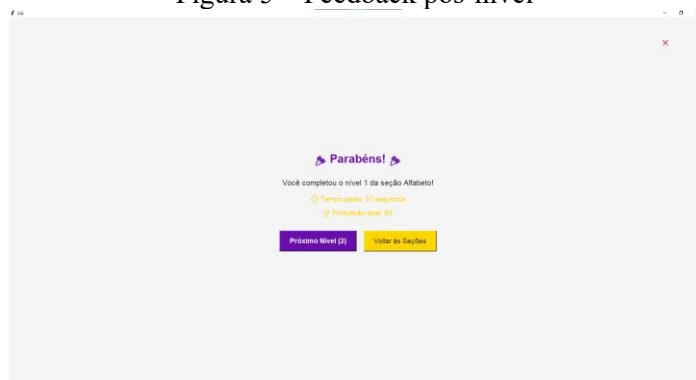
Figura 4 – Interface de reconhecimento de gestos



Fonte: Autoria própria (2025)

A Figura 5 mostra a tela de parabéns exibida ao completar um nível, representando uma recompensa ao usuário do LIA:

Figura 5 – Feedback pós-nível



Fonte: Autoria própria (2025)

Entretanto, algumas limitações foram observadas, como a dependência de condições ambientais favoráveis (luminosidade adequada), queda de desempenho em casos de oclusão parcial das mãos e sensibilidade a variações na velocidade de execução dos gestos.

Considerações finais

O LIA demonstrou ser uma ferramenta tecnológica promissora para o ensino da Língua Brasileira de Sinais, unindo aprendizado interativo, acessibilidade e inteligência artificial. Nesse sentido, os resultados obtidos, embora validados inicialmente apenas por métricas técnicas de desempenho, demonstram a viabilidade da abordagem proposta. Ademais, o desenvolvimento do projeto revelou que, ao integrar técnicas de visão computacional com modelos de aprendizado profundo, é possível criar um ambiente educacional mais inclusivo e dinâmico. Assim, para a comunidade surda, o LIA representa um recurso

tecnológico de apoio que respeita a estrutura linguística e cultural da Libras, enquanto para ouvintes interessados no aprendizado, oferece um método acessível, interativo e autônomo.

Apesar dos avanços obtidos, ainda existem desafios e oportunidades de evolução. Entre eles, destaca-se a implementação de um módulo de tradução automática de vídeos em Libras, capaz de interpretar conteúdos audiovisuais de maneira prática e eficiente. Essa funcionalidade ampliaria significativamente o escopo da aplicação, permitindo que usuários acessem e compreendam vídeos sem a necessidade de um intérprete humano. Paralelamente, a ampliação do banco de dados para incluir regionalismos, variações dialetais e sinais específicos de diferentes contextos culturais é fundamental para aumentar a representatividade linguística e garantir precisão em um país com diversidade linguística expressiva dentro da própria Libras.

A adaptação do sistema para dispositivos móveis representa outro passo estratégico para sua popularização. Ao permitir acesso a qualquer hora e lugar, o LIA poderia integrar-se ao cotidiano dos usuários, favorecendo o aprendizado informal e contínuo. Além disso, a incorporação de recursos multimodais – como suporte a entrada por voz e acompanhamento textual – tem potencial para tornar a ferramenta útil a diferentes perfis: pessoas com deficiência auditiva parcial, ouvintes iniciantes, intérpretes em formação e educadores que buscam recursos complementares em sala de aula.

Sob uma perspectiva social mais ampla, o LIA não apenas facilita o aprendizado individual, mas também atua como um agente de inclusão, fortalecendo a comunicação entre surdos e ouvintes e incentivando a empatia linguística. Ao se consolidar como uma plataforma acessível, adaptável e inovadora, o sistema pode se tornar uma referência no ensino de Libras mediado por tecnologia, contribuindo para a quebra de barreiras linguísticas e culturais, a valorização da diversidade comunicacional e a promoção de uma sociedade mais inclusiva. Em síntese, o LIA simboliza a convergência entre tecnologia, educação e inclusão, destacando-se pelo seu potencial na promoção de ambientes de aprendizagem mais acessíveis, dinâmicos e equitativos.

O projeto está disponível no repositório: [https://github.com/ElaineFesa/Projeto_Simac]

Referências

ASSOCIAÇÃO BRASILEIRA DE CULTURA SURDA. *Panorama da formação de intérpretes no Brasil*. Brasília, 2023.

BASTOS, G. L. et al. Reconhecimento de gestos para Libras com visão computacional e redes neurais. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE, 2021, Recife. *Anais [...]*. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 100–110.

BRASIL. Lei nº 10.436, de 24 de abril de 2002. Dispõe sobre a Língua Brasileira de Sinais – Libras. *Diário Oficial da União*: seção 1, Brasília, DF, 25 abr. 2002.

CERVO, A. L.; BERVIAN, P. A.; SILVA, R. *Metodologia científica*. 6. ed. São Paulo: Pearson Prentice Hall, 2007.

GIL, A. C. *Como elaborar projetos de pesquisa*. 5. ed. São Paulo: Atlas, 2010.

HASSANAT, A. B. et al. LSTM and GRU neural networks for gesture recognition: a comparative study. *International Journal of Advanced Computer Science and Applications*, [S.l.], v. 10, n. 11, p. 1–8, 2019. DOI: <https://doi.org/10.14569/IJACSA.2019.0101110>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico 2022: características da população surda*. Rio de Janeiro, 2022.

KAPP, K. M. *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco: Pfeiffer, 2012.

PRODANOV, C. C.; FREITAS, E. C. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. 2. ed. Novo Hamburgo: Feevale, 2013.

SILVEIRA, R. M.; BARROS, F. A. Gamificação como estratégia de aprendizagem ativa no ensino de Libras. *Revista Brasileira de Informática na Educação*, Porto Alegre, v. 29, p. 1–15, 2021. DOI: <https://doi.org/10.5753/rbie.2021.xxxxx>.

ZHANG, F. et al. MediaPipe Hands: on-device real-time hand tracking. *arXiv*, 2020. Disponível em: <https://arxiv.org/abs/2006.10214>. Acesso em: 21 ago. 2025.