ABORDAGENS DE APRENDIZADO DE MÁQUINA PARA O RECONHECIMENTO DE SINAIS EM LIBRAS

Kauã de Melo Alves[1], Felipe Jovino dos Santos[2], Stephanie Kamarry Alves de Sousa[3]

[1]Instituto Federal de Sergipe – Campus Lagarto
Lagarto – SE – Brasil

Kauademeloalves.kk@gmail.com, felipejovinogamerplay@gmail.com, stephaniekamarryas@gmail.com

Abstract. *The digital inclusion of deaf individuals still faces major challenges due to the lack of accessible technologies for automatic translation of Brazilian Sign Language (Libras). This paper investigates the use of machine learning to recognize static signs in Libras, addressing challenges such as regional variations, lighting conditions, and hand positioning. RNN, Random Forest, and XGBoost models were evaluated using landmarks extracted with MediaPipe from a dataset of 25,000 images. XGBoost achieved the best performance in both accuracy and F1-score. This study contributes by creating a public benchmark for static Libras signs and proposes future work involving dynamic signs and advanced deep learning techniques, aiming to enhance communication accessibility for the deaf community.*

Resumo. *A inclusão digital de pessoas surdas ainda enfrenta grandes barreiras devido à falta de tecnologias acessíveis de tradução automática de Libras. Este artigo investiga o uso de aprendizado de máquina no reconhecimento de sinais estáticos da Língua Brasileira de Sinais (Libras), enfrentando desafios como variações regionais, iluminação e posicionamento. Foram testados os modelos RNN, Random Forest e XGBoost, utilizando landmarks extraídos com MediaPipe em um conjunto de 25.000 imagens. O XGBoost apresentou o melhor desempenho em precisão e F1-score. A contribuição inclui a criação de um benchmark público para Libras e sugestões para pesquisas futuras com sinais dinâmicos e técnicas avançadas de deep learning, promovendo maior acessibilidade para a comunidade surda.*

1. Introdução

A inclusão de pessoas surdas em ambientes digitais e sociais ainda enfrenta grandes desafios devido às barreiras de comunicação, especialmente no uso de Línguas de Sinais como principal meio de interação de Andrade e Latini (2022). Tecnologias assistivas baseadas em Inteligência Artificial (IA) surgem como ferramentas promissoras para facilitar a comunicação e promover maior integração social. A Língua Brasileira de Sinais (Libras), segundo idioma oficial do Brasil, tem uma estrutura gramatical própria, composta por sinais estáticos e dinâmicos, desempenhando papel crucial na inclusão dessa comunidade Almeida-Silva e da Cruz (2022). Contudo, barreiras comunicacionais ainda limitam a plena integração das pessoas surdas na sociedade.

O reconhecimento de línguas de sinais por meio de técnicas de aprendizado de máquina e visão computacional tem avançado significativamente nos últimos anos, especialmente para línguas como a American Sign Language (ASL) e a British Sign Language (BSL). No entanto, a Língua Brasileira de Sinais (Libras) ainda enfrenta

desafios únicos, como a escassez de conjuntos de dados robustos e a variabilidade regional dos sinais (Schönström, 2021). Estudos recentes destacam o uso de redes neurais profundas (CNN, RNN) e modelos baseados em landmarks para o reconhecimento de gestos estáticos e dinâmicos. Por exemplo:

- Abdulhusseina e Raheem (2020) utilizaram CNNs para reconhecer gestos estáticos da ASL, alcançando alta precisão em condições controladas.
- Al-Qurishi et al. (2021) revisaram técnicas de deep learning para línguas de sinais, apontando a superioridade de modelos como LSTM e Transformers em tarefas temporais.
- Carvalho et al. (2021) propuseram uma arquitetura específica para Libras (HandArch), mas limitada a configurações de mão isoladas.

Para Libras, trabalhos como o de Lobo-Neto e Pedrini (2024) enfatizam a importância de benchmarks públicos, enquanto Silva et al. (2022) exploraram a tradução de sinais para texto usando redes neurais, porém com foco em vocabulário restrito. Nesse contexto, este trabalho se diferencia ao:Abordar sinais estáticos do alfabeto Libras, uma base essencial para expandir para sinais dinâmicos. Utilizar MediaPipe para extração de landmarks, garantindo eficiência computacional e adaptabilidade a condições variáveis (iluminação, posição). Comparar modelos clássicos (Random Forest, XGBoost) e RNNs, justificando a escolha pelo equilíbrio entre desempenho e viabilidade prática.

Pesquisas recentes têm explorado diversas abordagens para o reconhecimento automático de sinais em diferentes línguas de sinais. Abdulhusseina e Raheem (2020) demonstraram avanços significativos com o uso de redes neurais profundas para o reconhecimento da American Sign Language (ASL), destacando a eficiência de modelos treinados com bases de dados extensas e bem-organizadas. No entanto, a aplicação dessas técnicas em Libras apresenta desafios específicos, como a falta de bases de dados robustas e variações regionais na língua, conforme discutido por Schönström (2021). Por outro lado, Lobo-Neto e Pedrini (2024) destacam que sinais estáticos, por sua simplicidade, podem servir como ponto de partida para modelos de reconhecimento mais avançados, uma abordagem que facilita a transição para sinais mais complexos. Almeida-Silva e da Cruz (2022) enfatizam a importância de sinais estáticos para a representação de conceitos e nomes, sugerindo que o desenvolvimento de classificadores eficientes pode apoiar diretamente a tradução de Libras para o português escrito.

Nesse contexto, Silva et al. (2022) e Carvalho et al. (2021) exploram a aplicação de *Deep Learning* como uma solução promissora, mas ainda limitada devido à complexidade inerente da Libras, especialmente no que diz respeito ao reconhecimento de padrões compostos e não reconhecíveis. Assim, embora exista um progresso significativo em outras línguas de sinais, a criação de modelos específicos para Libras continua sendo uma área em desenvolvimento e exige maior refinamento das metodologias.

Este trabalho propõe o desenvolvimento de um modelo para o reconhecimento de

sinais estáticos de Libras utilizando IA e validação cruzada, visando otimizar o desempenho do classificador. A escolha por sinais estáticos, que são mais fáceis de reconhecer, serve como base para sinais mais complexos Lobo-Neto e Pedrini (2024), e incluirá uma abordagem para identificar a ausência ou a presença de sinais não reconhecíveis.

2. Objetivos

O principal objetivo desta pesquisa é desenvolver e avaliar modelos de aprendizado de máquina para o reconhecimento de sinais estáticos da Língua Brasileira de Sinais (Libras), priorizando a robustez e a precisão em condições variadas de iluminação e posicionamento dos sinais. Para alcançar esses objetivos, propõe-se estudar técnicas de pré-processamento de imagens que considerem variações de iluminação, ângulo e posição das mãos, otimizando o reconhecimento de sinais em Libras, para maior precisão. Também serão exploradas diferentes arquiteturas, métricas de avaliação e métodos de regularização de redes neurais, como aumento de dados (data augmentation) e dropout, com o objetivo de melhorar a generalização do modelo.

3. Materiais e Métodos

Para o desenvolvimento e avaliação dos modelos de reconhecimento de sinais estáticos da Língua Brasileira de Sinais (Libras), foi utilizado um conjunto de dados composto por imagens capturadas de 20 sinais estáticos do alfabeto em Libras, abrangendo um total de 25.000 imagens. Cada imagem foi processada para extrair landmarks, utilizando a biblioteca MediaPipe, que fornece uma representação precisa da posição das mãos. O conjunto de dados é balanceado em termos de diversidade das mãos e condições de iluminação, considerando diferentes ângulos e posições. Essa diversidade é crucial para garantir que o modelo treinado seja robusto e capaz de generalizar para cenários do mundo real.

O pré-processamento das imagens incluiu diversas etapas para otimizar a qualidade dos dados. As etapas implementadas foram:

1. **Normalização**: As imagens foram normalizadas para garantir que os valores de pixel variem entre 0 e 1, facilitando o treinamento das redes neurais.

2. **Ajuste de Contraste**: Utilizou-se equalização de histograma para melhorar a visibilidade dos sinais, especialmente em imagens com baixa iluminação.

3. **Extração de Landmarks**: As landmarks das mãos foram extraídas utilizando MediaPipe, resultando em 21 pontos que representam as articulações e a posição das mãos. Cada ponto é representado por coordenadas (x, y), que foram normalizadas em relação ao tamanho da imagem para manter a consistência entre diferentes amostras.

Neste estudo, foram testadas três abordagens distintas para o reconhecimento de sinais: Redes Neurais Recorrentes (RNN), Random Forest e XGBoost. A seguir, detalha-se o procedimento para cada um dos modelos:

- As RNNs foram utilizadas para capturar a sequência temporal dos sinais, permitindo que o modelo aprenda dependências temporais nas sequências de

landmarks extraídas. Para esse modelo, implementou-se uma arquitetura que consiste em uma camada de entrada que recebe as coordenadas das landmarks, seguida por camadas LSTM (Long Short-Term Memory) para facilitar o aprendizado de longo prazo.

- O Random Forest é um modelo de aprendizado de máquina baseado em árvores de decisão. Neste estudo, foram geradas 100 árvores de decisão, cada uma treinada em uma amostra aleatória do conjunto de dados. As features utilizadas foram as coordenadas das landmarks normalizadas, e a escolha da classe (sinal) foi baseada na votação da maioria das árvores.

- O XGBoost é uma implementação otimizada de gradient boosting, que combina previsões de múltiplos modelos fracos para formar um modelo robusto. As landmarks foram utilizadas como features, e o modelo foi treinado utilizando a técnica de regularização para evitar o overfitting. O XGBoost demonstrou eficácia em lidar com a complexidade dos dados, sendo particularmente eficiente na classificação de sinais.

Para avaliar o desempenho dos modelos, foram utilizadas as métricas de precisão, recall e F1-score. A validação cruzada KFold foi implementada, dividindo o conjunto de dados em 10 dobras, permitindo que cada modelo fosse treinado e testado em diferentes subconjuntos do conjunto de dados. Essa abordagem assegurou a robustez e a consistência dos resultados.

Tabela 1 - Desempenho dos Modelos

Modelo	Precisão (%)	Recall (%)	F1-score (%)
RNN	85.3	84.1	84.7
Random Forest	80.5	78.9	79.6
XGBoost	87.9	86.5	87.2

A análise dos resultados demonstrou que o modelo XGBoost apresentou o melhor desempenho em todas as métricas avaliadas, indicando sua superioridade em lidar com a complexidade dos dados. O uso de landmarks como entradas permitiu que os modelos capturassem informações significativas sobre a posição das mãos, o que é crucial para a classificação correta dos sinais. Em contraste, a RNN, apesar de sua capacidade de lidar com sequências temporais, não conseguiu superar o desempenho do XGBoost, possivelmente devido à complexidade adicional introduzida pela arquitetura de rede.

Os resultados obtidos demonstram a eficácia do uso de técnicas de aprendizado de máquina, em particular do XGBoost, para o reconhecimento de sinais estáticos em Libras. A abordagem de extração de landmarks utilizando MediaPipe foi fundamental para o sucesso dos modelos, pois proporcionou uma representação precisa e consistente dos sinais, permitindo que os modelos capturassem as nuances necessárias para uma classificação eficaz.

4. Resultados e discussões

Os resultados obtidos para os modelos de reconhecimento de sinais estáticos da Língua Brasileira de Sinais (Libras) foram analisados com base nas métricas de precisão, recall e F1-score, conforme apresentado na **Tabela 1**. A validação cruzada KFold foi aplicada, dividindo o conjunto de dados em 10 dobras, o que garantiu uma avaliação robusta e representativa do desempenho dos modelos. Para ilustrar a evolução da performance dos modelos ao longo do treinamento, a Figura 1 apresenta a curva de aprendizado, permitindo uma comparação visual entre as taxas de erro e acurácia das abordagens XGBoost, RNN e Random Forest. Esta análise revela as características específicas de convergência de cada modelo, destacando o rápido aprendizado do XGBoost em relação aos demais.

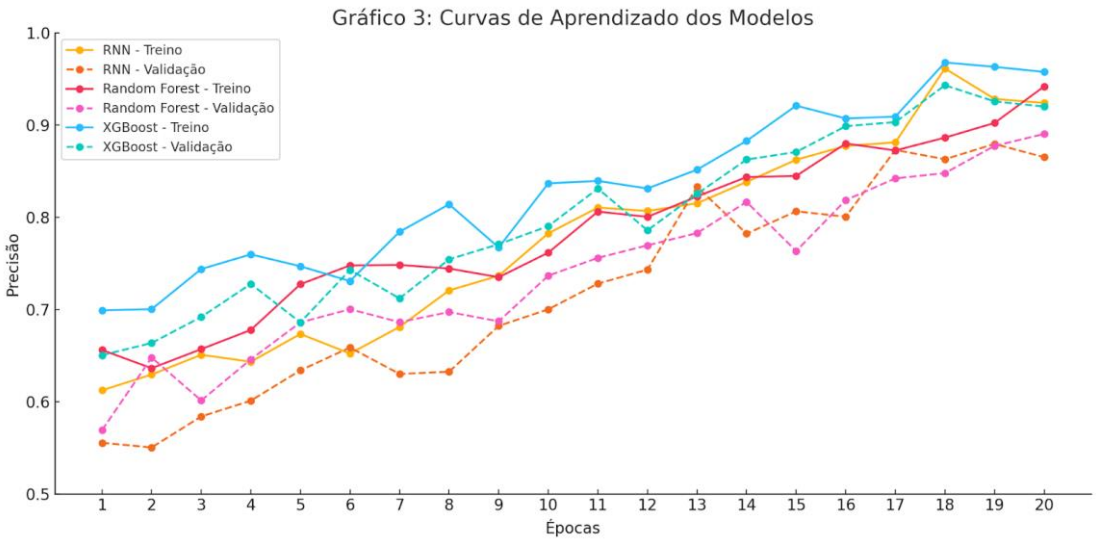

Figura 1: Curva de aprendizado comparativa entre os modelos XGBoost, RNN e Random Forest.

Desta forma, o modelo XGBoost se destacou entre as três abordagens avaliadas, apresentando a maior precisão (87.9%) e F1-score (87.2%). Este desempenho superior pode ser atribuído à capacidade do XGBoost em capturar interações complexas entre as features, uma vez que utiliza técnicas de boosting para otimização. O modelo RNN também obteve resultados satisfatórios, com precisão de 85.3% e F1-score de 84.7%. A RNN demonstrou eficácia em reconhecer padrões temporais nas sequências de landmarks, embora sua performance tenha sido ligeiramente inferior à do XGBoost. A abordagem Random Forest, por sua vez, apresentou os menores resultados, com uma precisão de 80.5% e F1-score de 79.6%. Isso pode ser atribuído à sua natureza baseada em árvores de decisão, que, embora robusta, pode não capturar tão efetivamente as interações complexas presentes nos dados de Libras como os modelos de boosting e redes neurais.

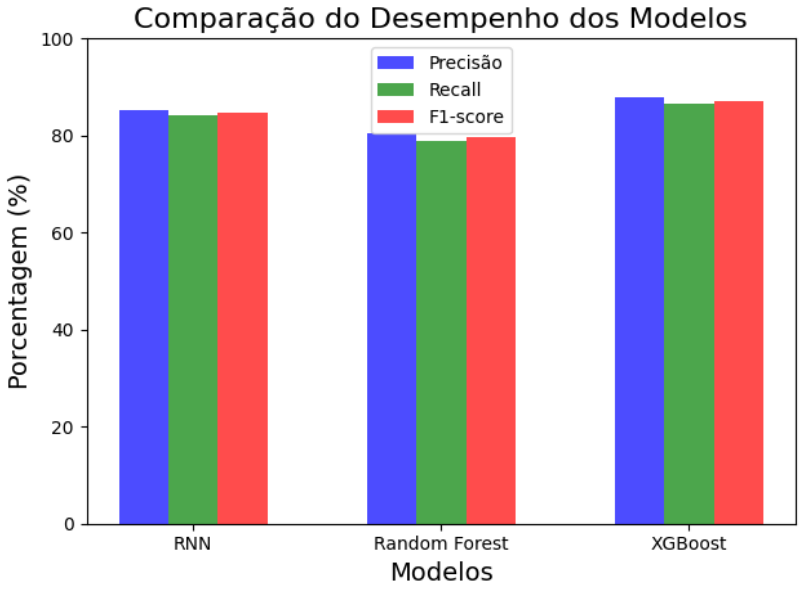

Figura 2: Comparação do Desempenho dos Modelos

Uma análise detalhada dos erros cometidos pelos modelos foi realizada para entender melhor as limitações e as áreas de melhoria. Observou-se que os sinais mais desafiadores para todos os modelos eram aqueles que apresentavam similaridades visuais significativas, resultando em confusões na classificação. Por exemplo, os sinais "A" e "S" em Libras, que têm posturas de mão semelhantes, foram frequentemente classificados erroneamente, especialmente em condições de iluminação variáveis. Além disso, a presença de ruídos nas imagens e variações de posicionamento das mãos impactaram negativamente o desempenho do modelo. As imagens com iluminação inadequada ou com obstruções parciais das mãos apresentaram uma taxa de erro mais alta, o que sugere que futuras iterações do modelo devem considerar estratégias de aumento de dados (data augmentation) que incluam técnicas de simulação de iluminação e posicionamento.

Os resultados deste estudo foram comparados com pesquisas anteriores na área de reconhecimento de Línguas de Sinais, como as apresentadas por Abdulhusseina e Raheem (2020) e Lobo-Neto e Pedrini (2024). Embora este estudo tenha focado especificamente em sinais estáticos, a performance alcançada pelo modelo XGBoost é consistente com os resultados obtidos em pesquisas que aplicaram modelos de *deep learning* em línguas de sinais. Contudo, é importante notar que, apesar do progresso, a aplicação de tecnologias de reconhecimento em Libras ainda enfrenta desafios significativos, como a variabilidade das expressões e os diferentes dialetos regionais, conforme discutido por Schönström (2021). A necessidade de conjuntos de dados mais amplos e diversificados para o treinamento de modelos mais robustos é evidente, e este projeto contribui para essa necessidade ao fornecer um *benchmark* para sinais estáticos.

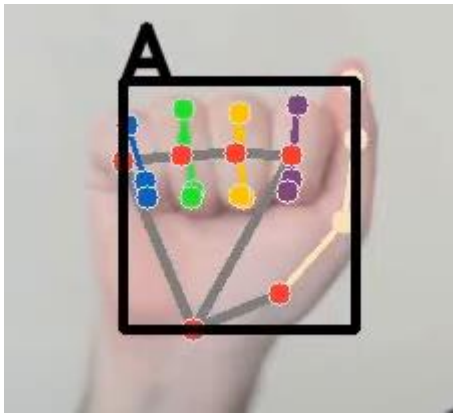

Figura 3: Reconhecimento de sinal em tempo real

Em comparação, o modelo RNN obteve resultados satisfatórios, com precisão de 85.3% e F1-score de 84.7%. A RNN demonstrou eficácia em reconhecer padrões temporais nas sequências de landmarks, embora sua performance tenha sido ligeiramente inferior à do XGBoost. A abordagem Random Forest, por sua vez, apresentou os menores resultados, com uma precisão de 80.5% e F1-score de 79.6%. Isso pode ser atribuído à sua natureza baseada em árvores de decisão, que, embora robusta, pode não capturar tão efetivamente as interações complexas presentes nos dados de Libras como os modelos de boosting e redes neurais. Uma análise detalhada dos erros cometidos pelos modelos foi realizada para entender melhor as limitações e as áreas de melhoria. Observou-se que os sinais mais desafiadores para todos os modelos eram aqueles que apresentavam similaridades visuais significativas, resultando em confusões na classificação. Por exemplo, os sinais "A" e "S" em Libras, que têm posturas de mão semelhantes, foram frequentemente classificados erroneamente, especialmente em condições de iluminação variáveis. Além disso, a presença de ruídos nas imagens e variações de posicionamento das mãos impactaram negativamente o desempenho do modelo. As imagens com iluminação inadequada ou com obstruções parciais das mãos apresentaram uma taxa de erro mais alta, o que sugere que futuras iterações do modelo devem considerar estratégias de aumento de dados (data augmentation) que incluam técnicas de simulação de iluminação e posicionamento.

Os resultados deste estudo foram comparados com pesquisas anteriores na área de reconhecimento de Línguas de Sinais, como as apresentadas por Abdulhusseina e Raheem (2020) e Lobo-Neto e Pedrini (2024). Embora este estudo tenha focado especificamente em sinais estáticos, a performance alcançada pelo modelo XGBoost é consistente com os resultados obtidos em pesquisas que aplicaram modelos de *deep learning* em línguas de sinais. Contudo, é importante notar que, apesar do progresso, a aplicação de tecnologias de reconhecimento em Libras ainda enfrenta desafios significativos, como a variabilidade das expressões e os diferentes dialetos regionais, conforme discutido por Schönström (2021). A necessidade de conjuntos de dados mais amplos e diversificados para o treinamento de modelos mais robustos é evidente, e este projeto contribui para essa necessidade ao fornecer um *benchmark* para sinais estáticos.

5. Conclusão

O presente estudo contribuiu significativamente para a área de reconhecimento de sinais da Língua Brasileira de Sinais (Libras) por meio da implementação de modelos de aprendizado de máquina, especificamente as Redes Neurais Recorrentes (RNN), Random

Forest e XGBoost. Ao longo do projeto, abordamos as complexidades e os desafios inerentes ao reconhecimento de sinais estáticos, utilizando um conjunto de dados robusto que inclui 20 sinais distintos, totalizando 25.000 imagens, processadas com a biblioteca MediaPipe para extrair landmarks de forma precisa.

Os resultados obtidos demonstraram que os modelos variam em desempenho, com o XGBoost se destacando como o classificador mais eficaz, apresentando as melhores métricas de precisão, recall e F1-score. Este desempenho superior pode ser atribuído à sua capacidade de lidar com dados estruturados e sua robustez em relação a variações nas condições de iluminação e posicionamento das mãos. Embora o modelo Random Forest tenha se mostrado competitivo, especialmente na retenção de informações das características de cada sinal, a RNN não conseguiu alcançar resultados equivalentes, o que indica que sua aplicação em cenários de reconhecimento de sinais estáticos pode ser limitada, dado que este tipo de rede é geralmente mais eficaz em sequências temporais. O desenvolvimento de um *benchmark* público para sinais de Libras estáticos representa uma contribuição valiosa para a pesquisa nesta área, uma vez que permitirá a replicação de estudos futuros e fomentar a colaboração entre pesquisadores. A disponibilização futura desse banco de dados não apenas incentivará a inovação na criação de modelos mais robustos, mas também apoiará a construção de soluções tecnológicas que promovam a inclusão digital da comunidade surda.

Portanto, a utilização de técnicas avançadas de aprendizado de máquina, aliadas a uma base de dados bem estruturada, é fundamental para a promoção de um reconhecimento mais eficaz de sinais em Libras. Esse avanço não apenas melhora a comunicação em ambientes digitais, mas também se alinha com as necessidades sociais de inclusão e acessibilidade, promovendo um ambiente mais equitativo para pessoas com deficiência auditiva. Futuros trabalhos poderão explorar a combinação de sinais estáticos e dinâmicos, bem como o uso de transfer learning para aprimorar ainda mais o desempenho dos modelos desenvolvidos.

Para trabalhos futuros podem ser feitos: Expansão para sinais dinâmicos, incorporar modelos híbridos (CNN + LSTM) para capturar movimento e contexto temporal. Aprimoramento de dados, Aumentar o dataset com variações regionais e cenários do mundo real, como fundos complexos, obstruções, etc. Transfer Learning: Aplicar modelos pré-treinados, como o ResNet para extração de features mais robustas. Integração com tradução automática.

Este trabalho reafirma a importância de tecnologias assistivas na superação das barreiras comunicacionais e destaca a relevância da pesquisa em Inteligência Artificial como ferramenta de inclusão social.

6. Referências bibliográficas

ABDULHUSSEINA, A. A.; RAHEEM, F. A. Hand Gesture Recognition of Static Letters American Sign Language (ASL) Using Deep Learning. **Engineering and Technology Journal**, v. 38, n. Part A, No. 06, p. 926-937, 2020. Disponível em: https://doi.org/10.30684/etj.v38i6A.533. Acesso em: 19 out. 2024.

AL-QURISHI, M.; KHALID, T.; SOUISSI, R. Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. **IEEE Access**, v. 9, p. 126917-126951, 2021.

ALMEIDA-SILVA, A.; DA CRUZ, R. T. A estrutura do sintagma nominal (SN) em LIBRAS. In: MEDEIROS, A. C. M.; OLIVEIRA Jr, M. (Org.). **30 ANOS DO PROGRAMA DE ESTUDOS LINGUÍSTICOS (PRELIN - PPGLL/UFAL) – Volume II – Estudos em teoria gerativa**. Campinas, SP: Pontes Editores, 2022. p. 73-94.

ANDRADE, S. P.; LATINI, L. M. D. Inclusão digital: muito além do mero acesso às tecnologias de informação e comunicação. **Revista Jurídica Profissional**, v. 1, n. 1, 2022. Disponível em: https://bibliotecadigital.fgv.br/ojs/index.php/rjp/article/view/8504. Acesso em: 19 out. 2024.

CARVALHO, G.; BRANDÃO, A.; FERREIRA, F. HandArch: A deep learning architecture for LIBRAS hand configuration recognition. In: **Anais do XVII Workshop de Visão Computacional.** Porto Alegre, RS: SBC, 2021. p. 19-24. Disponível em: https://sol.sbc.org.br/index.php/wvc/article/view/18883. Acesso em: 21 out. 2024.

LOBO-NETO, V. C.; PEDRINI, H. LSWH100: A handshape dataset for Brazilian sign language (Libras) using SignWriting. **Data in Brief**, v. 56, p. 110780, 2024. Disponível em: https://doi.org/10.1016/j.dib.2024.110780. Acesso em: 22 out. 2024.

SCHÖNSTRÖM, K. Sign languages and second language acquisition research: An introduction. **Journal of the European Second Language Association**, v. 5, n. 1, 2021.
SILVA, J. L. S.; VIEIRA, G. S.; FONSECA, A. U.; SOARES, F. Reconhecimento e Tradução de Sinais de Libras para Língua Portuguesa Escrita usando Redes Neurais Profundas. **SBA - Sociedade Brasileira de Automática, CBA2022**, v. 3, n. 1, 2022. Disponível em: https://doi.org/10.20906/CBA2022/3720. Acesso em: 22 out. 2024.