UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PRÓ-REITORIA DE GRADUAÇÃO
DEPARTAMENTO DE ENGENHARIAS E TECNOLOGIA
CURSO INTERDISCIPLINAR EM TECNOLOGIA DA INFORMAÇÃO

MARCOS MATEUS VIEIRA BARROS

RECONHECIMENTO AUTOMÁTICO DE SINAIS EM LIBRAS NA ÁREA DA TECNOLOGIA DA INFORMAÇÃO: ESTRUTURAÇÃO DE UMA BASE DE DADOS E AVALIAÇÃO DE UM MODELO LSTM

PAU DOS FERROS - RN
2026

MARCOS MATEUS VIEIRA BARROS

RECONHECIMENTO AUTOMÁTICO DE SINAIS EM LIBRAS NA ÁREA DA TECNOLOGIA DA INFORMAÇÃO: ESTRUTURAÇÃO DE UMA BASE DE DADOS E AVALIAÇÃO DE UM MODELO LSTM

Monografia apresentada à Universidade Federal Rural do Semi-Árido como requisito para obtenção do título de Bacharel em Interdisciplinar em Tecnologia da Informação.

Orientadora: Rosana Cibely Batista Rego, Profª. Dra.

PAU DOS FERROS - RN
2026

B277r Barros, Marcos Mateus Vieira.
Reconhecimento automático de sinais em libras na área da tecnologia da informação: estruturação de uma base de dados e avaliação de um modelo LSTM / Marcos Mateus Vieira Barros. - 2026.
38 f. : il.

Orientadora: Rosana Cibely Batista Rego.
Monografia (graduação) - Universidade Federal Rural do Semi-árido, Curso de Tecnologia da Informação, 2026.

1. Língua brasileira de sinais. 2. Reconhecimento de sinais. 3. Redes neurais recorrentes. 4. LSTM. 5. Tecnologia assistiva. I. Rego, Rosana Cibely Batista, orient. II. Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade com AACR2 e os dados fornecidos pelo) autor(a).
Biblioteca Campus Pau dos Ferros
Bibliotecária: Erlanda Maria Lopes da Silva
CRB: 15/966

MARCOS MATEUS VIEIRA BARROS

RECONHECIMENTO AUTOMÁTICO DE SINAIS EM LIBRAS NA ÁREA DA TECNOLOGIA DA INFORMAÇÃO: ESTRUTURAÇÃO DE UMA BASE DE DADOS E AVALIAÇÃO DE UM MODELO LSTM

Monografia apresentada à Universidade Federal Rural do Semi-Árido como requisito para obtenção do título de Bacharel em Interdisciplinar em Tecnologia da Informação.

Aprovado em: 22/06/2026.

Rosana C. B. Rego, Profª. Dra. - UFERSA
Presidente

Antonio Flávio Ferreira de Oliveira, Prof. Dr.
Membro Examinador

Náthalee Cavalcanti de Almeida Lima, Profª. Dra.
Membro Examinador

PAU DOS FERROS - RN
2026

AGRADECIMENTOS

Agradeço primeiramente aos meus familiares próximos: o meu pai, Marcos, e a minha mãe, Girlene, por sempre me incentivarem a buscar uma formação acadêmica e por me darem todo o apoio emocional e financeiro. Agradeço também ao meu irmão, Eduardo, por me motivar e me ajudar com o transporte para a universidade.

Tenho imensa gratidão à minha orientadora, a Professora Doutora Rosana Cibely, por todos os esclarecimentos, pela disponibilidade e pelas orientações ao longo da produção deste trabalho, e por ser extremamente paciente comigo. Agradeço também aos colegas de pesquisa: Letícia Gonçalves e Pedro Lucas, por me ajudarem com dúvidas e esclarecimentos.

À minha namorada, Clarisse, agradeço por ser meu ombro amigo, minha companhia e por me apoiar nos meus projetos.

Agradeço aos meus amigos veteranos: Pollyana, Lucas Anderson, Pedro Hércules e Pedro Vinícius, por me acolherem quando ingressei na universidade, pela ajuda e pela amizade. Sou muito grato à Ângela, Michely, Vladimyr, que ingressaram no curso junto comigo, pelo companheirismo e pela amizade. Agradeço a todos os meus amigos que fiz na universidade, por compartilharem os desafios e experiências da vida acadêmica.

RESUMO

A Língua Brasileira de Sinais (Libras) constitui a principal forma de comunicação da comunidade surda no Brasil, reconhecida oficialmente pela Lei nº 10.436/2002(Brasil, 2002). Apesar dos avanços legislativos, persiste uma lacuna significativa de ferramentas tecnológicas que facilitem a aprendizagem e o reconhecimento automatizado de sinais, especialmente em domínios especializados como a Tecnologia da Informação (TI). Este trabalho propõe o desenvolvimento de um modelo de aprendizado de máquina para reconhecimento e predição de sinais isolados em Libras voltados ao vocabulário técnico da área de TI. A metodologia empregada baseou-se na coleta de vídeos de sinais disponibilizados na plataforma do Youtube, no canal IFSC Chapecó. Implementou-se uma arquitetura LSTM (*Long Short-Term Memory*). O modelo foi treinado e validado com um conjunto de dados contendo sinais específicos da área tecnológica, alcançando uma precisão de 99,36% na fase de teste. Os resultados demonstram o potencial impacto social ao promover maior acessibilidade e inclusão digital para pessoas surdas em ambientes educacionais de TI.

Palavras-chave: língua brasileira de sinais; reconhecimento de sinais; redes neurais recorrentes; LSTM; tecnologia assistiva.

ABSTRACT

The *Língua Brasileira de Sinais* (Libras) is the primary form of communication for the deaf community in Brazil, officially recognized by Law No. 10.436/2002(Brasil, 2002). Despite legislative advances, there remains a significant gap in technological tools that facilitate learning and automated recognition of signs, especially in specialized domains such as Information Technology. This work proposes the development of a machine learning model for recognition and prediction of isolated signs in Libras focused on the technical vocabulary of the IT area. The methodology employed was based on the collection of videos of signs available on the IFSC Chapecó YouTube channel. An LSTM (Long Short-Term Memory) architecture was implemented. The model was trained and validated with a dataset containing specific signs from the technological area, achieving an accuracy of 99.36% in the testing phase. The results demonstrate the potential social impact by promoting greater accessibility and digital inclusion for deaf people in educational environments of IT.

Keywords: libras, sign language recognition, LSTM, information technology, accessibility.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

A comunicação humana constitui um direito fundamental e um instrumento indispensável à participação social, ao acesso à educação e ao pleno exercício da cidadania. Para a comunidade surda brasileira, esse direito é exercido primariamente por meio da Língua Brasileira de Sinais (Libras), sistema linguístico de modalidade visual-espacial dotado de gramática própria, independente da língua portuguesa e reconhecido oficialmente pela Lei n.º 10.436, de 24 de abril de 2002 (Brasil, 2002). O referido dispositivo legal estabelece o reconhecimento formal da língua e a obrigatoriedade de sua difusão e do apoio institucional ao seu uso, determinando a incorporação do ensino de Libras na formação de professores e fonoaudiólogos, o que evidencia sua relevância tanto social quanto pedagógica.

Apesar do avanço normativo, a efetivação desse direito ainda enfrenta obstáculos significativos. Segundo a Pesquisa Nacional de Saúde (PNS) de 2019, divulgada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2021, cerca de 1,7 milhão de brasileiros entre 5 e 40 anos declararam possuir algum grau de dificuldade auditiva (Ibge, 2021). Entre aqueles com surdez severa, apenas 22,4% afirmaram dominar a Libras, proporção que sobe para 61,3% entre os que relataram não conseguir ouvir som algum (Ibge, 2021). Esses indicadores demográficos revelam, simultaneamente, a dimensão da demanda por suporte comunicacional e a parcela expressiva da população que ainda não tem acesso pleno à própria língua.

Essa realidade é agravada pela escassez de intérpretes de Libras qualificados, especialmente em regiões afastadas dos grandes centros urbanos, e pela predominância de conteúdos educacionais e profissionais produzidos exclusivamente na modalidade oral ou escrita do português (Klein, 2001; Lemos et al., 2019). Em contextos técnico-científicos, como os cursos de Tecnologia da Informação (TI), essa lacuna é ainda mais pronunciada: o vocabulário especializado da área, composto por termos como *hardware*, *software*, algoritmo, banco de dados, rede de computadores e entre outros, não dispõe de representação sistematizada em bases de dados públicas de Libras. Consequentemente, estudantes surdos que ingressam nesses cursos enfrentam uma barreira linguística que não decorre de limitação cognitiva, mas da ausência de ferramentas de mediação terminológica adequadas ao seu contexto.

Nesse cenário, as tecnologias de aprendizado de máquina e visão computacional emergem como instrumentos promissores para o desenvolvimento de sistemas de reconhecimento automático de sinais. A literatura demonstra que algoritmos clássicos como *Extra Trees* e *Random Forest*, combinados com técnicas de extração de pontos-chave das mãos por meio do framework *MediaPipe*, são capazes de atingir acurácias superiores a 98% para vocabulários amplos de Libras (Morais; Almeida; Rego, 2025; Morais et al., 2025). Para sinais com dinâmicas temporais mais complexas, arquiteturas de redes neurais recorrentes, em particular as redes *Long Short-Term Memory* (LSTM) e suas variantes bidirecionais (BiLSTM), demonstraram superioridade na modelagem de dependências temporais de longo alcance, alcançando até 92% de acurácia em

conjuntos de sinais dinâmicos (Rego; Morais; Almeida, 2025). Contudo, o desenvolvimento de tais sistemas pressupõe a existência de bases de dados rotuladas e estruturadas que, no domínio técnico de TI, simplesmente não existem na literatura.

A ausência de um dataset específico para sinais de TI em Libras representa, portanto, um gargalo que impossibilita tanto o treinamento de modelos de reconhecimento nesse domínio quanto a avaliação comparativa de diferentes abordagens algorítmicas. Nenhum dos conjuntos de dados públicos existentes, tais como, *Brazilian Sign Language Words Recognition* (Carneiro; Silva; Salvadeo, 2021), Libras-10 (Almeida et al., 2019) ou MINDS-Libras (Guimarães et al., 2019), contempla terminologia técnica de informática, tendo sido construídos para vocabulários do cotidiano ou sinais emocionais e dinâmicos genéricos. Suprir essa lacuna é uma contribuição deste trabalho, pois viabiliza pesquisas futuras e abre caminho para o desenvolvimento de ferramentas assistivas no contexto educacional tecnológico.

Diante do exposto, este trabalho propõe a estruturação de um *dataset* de sinais de Libras relacionados ao vocabulário do curso de Bacharelado em Tecnologia da Informação, coletados a partir de vídeos do canal Libras IFSC Chapecó na plataforma YouTube, e o desenvolvimento de um modelo de reconhecimento automático de sinais isolados, baseado em redes neurais recorrentes LSTM, alimentadas por características espaciais extraídas dos pontos-chave anatômicos das mãos por meio do *MediaPipe Hands*. A representação por pontos-chave tridimensionais confere ao sistema robustez frente a variações de iluminação, cor de pele e cenário de fundo, além de reduzir a dimensionalidade do problema. Técnicas de aumento de dados foram empregadas para ampliar a variabilidade das amostras e mitigar o sobreajuste (*overfitting*) decorrente do volume limitado de dados disponíveis.

A relevância científica deste trabalho reside, primeiramente, na estruturação e disponibilização de uma base de dados[1] específico para sinais técnicos de TI em Libras, com 11 interpretes diferentes, foram selecionados 30 sinais distintos, a escolha desse conjunto buscou contemplar um vocabulário representativo do contexto acadêmico priorizando sinais efetivamente utilizados na comunicação de conteúdos técnicos.

O número de sinais não representa uma limitação metodológica da arquitetura LSTM, mas sim um recorte do domínio investigado. Observa-se que os conjuntos de dados disponíveis na literatura tem o diferentes comprimento de vocabulário, variando entre 10 e 20 sinais (Almeida et al., 2019; Rezende; Almeida; Guimarães, 2021) abordando termos do cotidiano. Embora este trabalho ainda represente um vocabulário inicial, esse conjunto amplia o número de classes em relação a parte dos datasets existentes e introduz um domínio temático ainda não explorado na literatura, constituindo um primeiro passo para a construção de bases de dados especializadas em Libras.

A segunda contribuição consiste na demonstração da viabilidade da utilização de redes neurais recorrentes do tipo LSTM para o reconhecimento automático de sinais isolados em Libras para um domínio técnico. A arquitetura proposta é capaz de capturar padrões temporais

[1] <https://github.com/mmvbs/Libras-Dataset>

nos sinais permitindo a classificação com alta precisão?

Por fim, a contribuição social do trabalho está relacionada à promoção da acessibilidade e da inclusão de estudantes e profissionais surdos na área de Tecnologia da Informação. A disponibilização de uma base de dados especializada pode subsidiar pesquisas futuras e o desenvolvimento de tecnologias assistivas, tradutores automáticos e ferramentas educacionais capazes de reduzir barreiras linguísticas no ensino e na aprendizagem de conteúdos técnicos em Libras.

1.1 Objetivos

1.1.1 Objetivo Geral

Desenvolver um modelo de aprendizado de máquina baseado em rede neural recorrente LSTM capaz de reconhecer e classificar sinais da Língua Brasileira de Sinais relacionados ao vocabulário técnico do curso de Bacharelado em Tecnologia da Informação, a partir da estruturação de um *dataset* construído especificamente para esse domínio.

1.1.2 Objetivos Específicos

- Estruturar um conjunto de dados (*dataset*) composto por 30 sinais de Libras associados a termos técnicos da área de Tecnologia da Informação, a partir de vídeos coletados em fonte audiovisual aberta, suprindo a lacuna de bases de dados especializadas nesse domínio.
- Aplicar técnicas de aumento de dados para ampliar a variabilidade amostral e reduzir o risco de *overfitting* no treinamento do modelo.
- Extrair representações espaciais tridimensionais das configurações manuais por meio dos pontos-chave anatômicos fornecidos pelo *framework MediaPipe Hands*, obtendo vetores de características compactos e invariantes a condições de captura.
- Implementar e treinar uma arquitetura de rede neural recorrente com camadas LSTM para a classificação dos sinais do vocabulário técnico de TI.
- Avaliar o desempenho do modelo proposto por meio das métricas de acurácia, precisão, revocação e F1-score, identificando padrões de confusão entre sinais visualmente similares.

1.2 Organização do Trabalho

Este trabalho está estruturado em cinco capítulos. A organização dos capítulos é descrita da seguinte forma:

- No Capítulo 1, apresenta-se a introdução, contextualizando a importância da Libras e os desafios enfrentados pela comunidade surda, especialmente no contexto técnico de TI, além de delinear os objetivos gerais e específicos do estudo.

- O Capítulo 2 revisa a literatura relacionada ao reconhecimento automático de sinais em Libras, destacando as metodologias empregadas, os datasets disponíveis e as lacunas existentes.
- O Capítulo 3 detalha a metodologia adotada para a construção do *dataset*, a extração de características e o desenvolvimento do modelo LSTM.
- O Capítulo 4 apresenta os resultados obtidos, incluindo análises quantitativas e qualitativas do desempenho do modelo.
- Finalmente, o Capítulo 5 discute as implicações dos resultados, as limitações do estudo e as direções para pesquisas futuras.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais que embasam o desenvolvimento deste trabalho, abrangendo desde a natureza linguística da Língua Brasileira de Sinais até as abordagens computacionais empregadas para o reconhecimento automático de sinais.

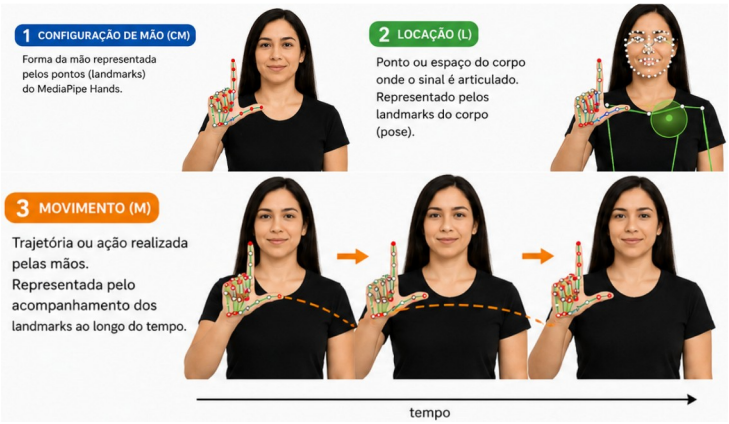
2.1 Língua Brasileira de Sinais

A Libras é a língua primária da comunidade surda brasileira, constituindo um sistema linguístico completo dotado de sintaxe, morfologia e semântica próprias (Governo Federal do Brasil, 2025). Vale destacar que a Libras não é uma representação gestual do português, mas uma língua natural de modalidade visual-espacial, reconhecida oficialmente pela Lei n.º 10.436, de 24 de abril de 2002, como meio legal de comunicação e expressão (Brasil, 2002). Sua adoção é essencial para a inclusão social de pessoas surdas, especialmente considerando a escassez de intérpretes qualificados e a dificuldade de acesso a recursos de comunicação em diversas regiões do país (Ibge, 2021; Morais; Almeida; Rego, 2025).

Segundo estimativas da Federação Mundial de Surdos, mais de 70 milhões de pessoas surdas vivem no mundo, muitas delas enfrentando barreiras de comunicação que restringem sua participação plena na sociedade (Wfd, 2003). No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) aponta que aproximadamente dez milhões de brasileiros possuem algum grau de deficiência auditiva, dos quais cerca de três milhões apresentam surdez completa (Ibge, 2021). Para essa parcela da população, a Libras representa um meio de comunicação e identidade cultural e linguística coletiva.

Cada sinal da Libras é composto por parâmetros fonológicos fundamentais conforme mostra a Figura 1. A configuração de mão (CM) refere-se à forma específica que a mão assume durante a execução do sinal, incluindo a posição dos dedos e a orientação da palma. A locação (L), ou ponto de articulação, indica o espaço de enunciação onde o sinal é produzido, como próximo ao corpo, na altura do rosto ou em um plano mais distante. O movimento (M) está relacionado à trajetória ou à ação realizada pelas mãos para representar determinado conceito, podendo envolver deslocamentos lineares, circulares ou movimentos mais complexos. Esses parâmetros funcionam como os fonemas das línguas orais, constituindo unidades mínimas de distinção de significado. No entanto, ao contrário das línguas orais, onde os fonemas são produzidos sequencialmente ao longo do tempo, os parâmetros dos sinais em Libras manifestam-se de forma simultânea (Quadros; Karnopp, 2007).

Figura 1 – Exemplo de sinal da Libras com seus parâmetros fonológicos: Configuração de Mão (CM), Locação (L) e Movimento (M).



Fonte: Gerado com o auxílio do ChatGPT Go, versão GPT-5.5.

Essa característica de simultaneidade introduz um nível adicional de complexidade para o processamento e o reconhecimento automático da língua usando algoritmos de Inteligência Artificial (IA). Enquanto nas línguas orais os elementos fonológicos podem ser analisados em sequência temporal, nas línguas de sinais é necessário considerar a ocorrência síncrona de múltiplos parâmetros, exigindo abordagens computacionais capazes de representar e interpretar adequadamente essa estrutura multidimensional. Além dos parâmetros manuais, muitos sinais incorporam expressões faciais e movimentos corporais como componentes gramaticais, o que amplia ainda mais a complexidade do reconhecimento automático (Morais; Almeida; Rego, 2025; Rego; Morais; Almeida, 2025).

A literatura aponta que soluções tecnológicas para reconhecimento de sinais podem reduzir as barreiras de comunicação, viabilizando o desenvolvimento de sistemas assistivos e ferramentas educativas que favoreçam a interação entre surdos e ouvintes (Rego; Morais; Almeida, 2025; Rezende; Almeida; Guimarães, 2021). Diferentemente de abordagens antigas que dependiam de luvas sensorizadas ou equipamentos de alto custo, as tecnologias atuais de visão computacional e aprendizado de máquina permitem trabalhar com câmeras convencionais, tornando as soluções significativamente mais acessíveis e práticas (Morais; Almeida; Rego, 2025; Arcanjo et al., 2024).

2.2 Base de Dados para Reconhecimento de Libras

Um dos principais desafios no reconhecimento automático de Libras é a disponibilidade de bases de dados diversificadas. Diferentemente de línguas de sinais mais estudadas, como a Língua de Sinais Americana (ASL - *American Sign Language*), que dispõe de *datasets* bem estabelecidos, os conjuntos de dados para Libras são limitados tanto em quantidade quanto em diversidade de sinais representados (Rego; Morais; Almeida, 2025).

Existem, no entanto, alguns conjuntos de dados públicos que têm sido utilizados como referência na literatura. Por exemplo, o *dataset* denominado *Brazilian Sign Language – Words*

Recognition, disponível na plataforma Kaggle, contém 15 sinais do cotidiano, entre os quais figuram vocabulários como “Eu”, “Você”, “Carro”, “Moto”, “Hospital” e “Banco”. Os vídeos foram capturados com variações de iluminação e diferentes cenários de fundo, o que favorece a capacidade de generalização dos modelos treinados a partir desse conjunto (Carneiro; Silva; Salvadeo, 2021). Trabalhos como o de Morais, Almeida e Rego (2025) demonstraram que, mesmo com um número relativamente pequeno de sinais, é possível alcançar bons níveis de precisão utilizando técnicas de aprendizado de máquina adequadas.

Outro exemplo é o *dataset Libras-10*, que reúne 10 sinais que incorporam expressões faciais e emocionais, tais como “Acalmar”, “Felicidade”, “Zangado” e “Sortudo”. Esse conjunto apresenta um desafio adicional, pois muitos de seus sinais dependem não apenas da configuração das mãos, mas também de componentes faciais para a transmissão completa do significado (Almeida et al., 2019).

Por fim, um dos conjuntos de dados mais robustos e amplamente utilizados é o MINDS-Libras, disponibilizado no *Zenodo*. Ele compreende 1.200 sequências de vídeo cobrindo 20 sinais dinâmicos, executados por 12 pessoas distintas com cinco repetições cada. A diversidade de sinalizadores é um aspecto crucial, pois permite avaliar se os modelos aprenderam efetivamente a representação dos sinais ou apenas memorizaram peculiaridades de um sinalizador específico (Guimarães et al., 2019).

Além desses, o presente trabalho apresenta a estruturação de um novo conjunto de dados, extraído de vídeos gravados e disponibilizados no canal do Libras IFSC Chapecó no YouTube[2], que foi construído especificamente para sinais relacionados ao vocabulário técnico de tecnologia da informação, cujo detalhamento é apresentado no capítulo de metodologia. A construção desse dataset se justifica pela escassez de bases públicas voltadas a terminologias da área da informática em Libras.

2.3 Visão Computacional e Extração de Pontos-Chave

A extração de características visuais das mãos constitui o primeiro estágio do *pipeline* de reconhecimento de sinais automático com algoritmos de IA. No contexto deste trabalho, é empregado o *framework MediaPipe Hands*, desenvolvido pelo *Google Research*, para detectar e extrair os pontos-chave das mãos a partir de imagens RGB comuns (Edge, 2025).

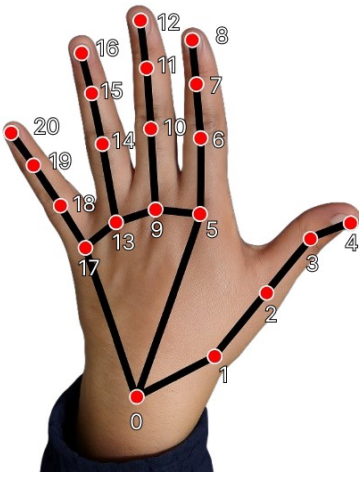
O funcionamento interno do *MediaPipe Hands* baseia-se em uma arquitetura de dois estágios. No primeiro, uma rede neural convolucional fundamentada no detector *Single Shot Detector* (SSD) varre a imagem em busca de regiões que possam conter mãos. Identificada uma região candidata, uma segunda rede neural localiza com precisão 21 pontos-chave, denominados *landmarks*, que representam o pulso, as articulações e as pontas dos dedos de cada mão detectada. As coordenadas desses pontos são normalizadas para o intervalo $[0, 1]$ nos eixos X e Y, enquanto a coordenada Z representa a profundidade relativa de cada ponto tomando o pulso como referência

[2] <https://www.youtube.com/@librasifscchapeco666/videos>

(Edge, 2025).

A Figura 2 apresenta um exemplo de detecção de pontos-chave utilizando o *MediaPipe Hands*. Conforme a documentação do *MediaPipe Hands*(Edge, 2025), o sistema é capaz de identificar até 21 pontos-chave por mão, incluindo o pulso, as articulações e as pontas dos dedos. As coordenadas desses pontos são normalizadas para o intervalo $[0, 1]$ nos eixos X e Y, enquanto a coordenada Z representa a profundidade relativa de cada ponto em relação ao pulso.

Figura 2 – Exemplo de detecção de pontos-chave utilizando o *MediaPipe Hands.*



Fonte: Rego, Morais e Almeida (2025)

Para cada vídeo processado, o sistema extrai até 63 coordenadas por mão (21 pontos multiplicados por três eixos), totalizando 126 atributos por quadro quando ambas as mãos são detectadas simultaneamente. Quando apenas uma mão é visível, os atributos da mão ausente são preenchidos com zeros. Quadros em que nenhuma mão é detectada ou em que duas mãos são classificadas com a mesma lateralidade são descartados do conjunto de dados (Rego; Morais; Almeida, 2025).

Para além das coordenadas brutas dos pontos-chave, essas informações estão diretamente relacionadas aos parâmetros linguísticos discutidos por Quadros e Karnopp (2007). A configuração de mão (CM) pode ser inferida a partir da disposição relativa dos pontos-chave identificados através do módulo Hands, enquanto a locação (L) se dá a partir pontos-chave estipulados pelo módulo *Holistics*. O movimento (M) é capturado pela variação temporal das posições dos pontos-chave ao longo dos quadros do vídeo. Assim, a extração de pontos-chave não apenas fornece uma representação compacta e eficiente das mãos, mas também preserva informações linguísticas essenciais para o reconhecimento automático de sinais em Libras.

2.4 Algoritmos de Aprendizado de Máquina Clássico

Quando a discriminação entre sinais pode ser realizada a partir de padrões espaciais e de configuração da mão, algoritmos tradicionais de aprendizado de máquina demonstram eficácia notável. A literatura e os resultados experimentais convergem para o uso de métodos *ensemble*

baseados em árvores de decisão e técnicas de vizinhança (Morais; Almeida; Rego, 2025; Morais et al., 2025).

O *Random Forest*, introduzido por Breiman (2001), combina múltiplas árvores de decisão independentes, cada uma treinada em uma amostra aleatória com reposição do conjunto de dados, e produz a predição final por votação majoritária. Morais et al. (2025) demonstraram que o *Random Forest* pode alcançar alta precisão no reconhecimento de sinais da Libras. No entanto, para sinais não estáticos, cuja distinção depende de padrões temporais, os modelos baseados em árvores de decisão podem apresentar limitações, uma vez que tratam cada quadro de forma independente, sem capturar as dependências temporais intrínsecas aos sinais dinâmicos (Morais; Almeida; Rego, 2025).

O algoritmo *Extra Trees* (*Extremely Randomized Trees*) foi proposto por Geurts, Ernst e Wehenkel (2006) como uma variante do *Random Forest* que introduz aleatoriedade adicional na seleção dos pontos de corte em cada nó das árvores. Morais, Almeida e Rego (2025) demonstraram que o *Extra Trees* pode superar o *Random Forest* em tarefas de reconhecimento de sinais, especialmente quando o conjunto de dados é relativamente pequeno e apresenta alta dimensionalidade, como é o caso dos pontos-chave extraídos dos vídeos de Libras. Morais, Almeida e Rego (2025) realizaram uma comparação entre diferentes algoritmos de aprendizado de máquina, incluindo o *Random Forest*, o *Extra Trees*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM) e o LightGBM, concluindo que o *Extra Trees* apresentou melhor desempenho em termos de acurácia e tempo de treinamento para o reconhecimento de sinais estáticos da Libras.

2.4.1 Redes Neurais Artificiais e Trabalhos Relacionados em Libras

Dentro do campo do aprendizado de máquina, as redes neurais artificiais representam uma classe de modelos computacionais inspirados na estrutura e no funcionamento do sistema nervoso biológico. Elas são amplamente utilizadas em tarefas de reconhecimento de padrões, classificação e regressão, devido à sua capacidade de aprender representações complexas a partir dos dados (Goodfellow; Bengio; Courville, 2016). Especialmente no contexto do reconhecimento de sinais da Libras, as redes neurais têm se mostrado eficazes para capturar as nuances visuais e temporais dos sinais, superando frequentemente os métodos tradicionais de aprendizado de máquina em termos de precisão e capacidade de generalização (Rezende; Almeida; Guimarães, 2021; Castro; Guerra; Guimarães, 2023; Rezende; Almeida; Guimarães, 2021; Arcanjo et al., 2024; Rego; Morais; Almeida, 2025).

Rezende, Almeida e Guimarães (2021) introduziram o dataset MINDS-Libras, contendo mais de 1.000 vídeos RGB e de profundidade de 20 sinais dinâmicos da Libras executados por 12 participantes. Os autores conduziram experimentos de referência com uma arquitetura de Rede Neural Convolucional (CNN - *Convolutional Neural Networks*), alcançando acurácia de 93,3%. O trabalho estabeleceu uma base de dados fundamental para pesquisas posteriores em reconhecimento de Libras, destacando a importância da diversidade de sinalizadores para a

generalização dos modelos.

Castro, Guerra e Guimarães (2023) propuseram um método baseado em CNN tridimensional (3D CNN) que processa exclusivamente entrada RGB, estimando indiretamente informações de pose e profundidade sem necessidade de sensores especializados. O modelo alcançou 96% de acurácia em validação cruzada de dez partições sobre o MINDS-Libras; contudo, ao empregar o protocolo de deixar-um-sinalizador-de-fora (*leave-one-person-out*), a acurácia caiu para aproximadamente 57%, evidenciando overfitting às características particulares dos sinalizadores presentes no treino.

Arcanjo et al. (2024) exploraram uma abordagem não baseada em redes neurais profundas, utilizando o algoritmo *Fast Dynamic Time Warping* (FastDTW) sobre características angulares extraídas dos pontos-chave das mãos. O método alcançou 96% de acurácia no MINDS-Libras e 90% no conjunto INCLUDE-50, porém apresentou dificuldades em discriminar sinais com configurações manuais similares, como “Banheiro” e “Sapo”.

Rego, Morais e Almeida (2025) propuseram uma arquitetura BiLSTM (*Bidirectional Long Short-Term Memory*) combinada com características multimodais como entradas espaciais, cinemáticas (velocidade e aceleração) e espectrais obtidas via Transformada Rápida de Fourier (FFT) para o reconhecimento de 20 sinais dinâmicos do MINDS-Libras. O modelo alcançou 92% de acurácia e F1-Score, superando *baselines* como LSTM (*Long Short-Term Memory*) simples (84%), CNN (89%) e BiLSTM sem características multimodais (89%). Os resultados validaram empiricamente que, para sinais dinâmicos com dependências temporais complexas, arquiteturas recorrentes bidirecionais se enquadram significativamente melhor do que algoritmos tradicionais de aprendizado de máquina.

o Quadro 1 sintetiza os principais trabalhos relacionados ao reconhecimento automático de Libras identificados na literatura, evidenciando as abordagens utilizadas, os conjuntos de dados, as acurácias reportadas e as limitações apontadas por cada estudo.

Quadro 1 – Resumo dos principais trabalhos de reconhecimento automático de Libras.

Referência	**Abordagem**	**Dataset**	**Acurácia**	**Limitação principal**
Rezende, Almeida e Guimarães (2021)	CNN	MINDS-Libras	93,3%	Base de dados limitada e sem avaliação entre diferentes sinalizadores.
Castro, Guerra e Guimarães (2023)	CNN	MINDS-Libras, LIBRAS-UFOP e INCLUDE-50	96% (10-fold) / 57% (LOSO)	Forte dependência dos sinalizadores utilizados no treinamento.
Arcanjo et al. (2024)	FastDTW com características angulares	MINDS-Libras e INCLUDE-50	96% / 90%	Dificuldade na distinção de sinais com configurações de mão semelhantes.
Morais, Almeida e Rego (2025)	*Extra Trees* e *Random Forest*	MINDS-Libras, Libras-10, *Brazilian Sign Language Words Recognition*	98% e 97%	Não considera expressões faciais e restringe-se a sinais isolados.
Morais et al. (2025)	*Random Forest* com expressões faciais	MINDS-Libras, Libras-10, *Brazilian Sign Language Words Recognition*	99%	Variabilidade entre sinalizadores e sinais estáticos.
Rego, Morais e Almeida (2025)	BiLSTM e características cinemáticas	MINDS-Libras	92%	Avaliação limitada a sinais isolados e vocabulário reduzido.

Fonte: Autoria própria (2026)

A análise conjunta desses trabalhos permite identificar uma progressão metodológica dos modelos convolucionais para extração de características espaciais em quadros isolados até arquiteturas recorrentes bidirecionais capazes de modelar dependências temporais de longo alcance. Os modelos convencionais, embora eficazes para sinais estáticos, apresentam limitações para sinais dinâmicos, cuja distinção depende de padrões temporais complexos. Por outro lado, as redes neurais recorrentes, especialmente as BiLSTM, demonstraram superioridade ao capturar essas dependências temporais, mas ainda enfrentam desafios relacionados à generalização entre diferentes sinalizadores e à necessidade de conjuntos de dados mais diversificados. Por esse motivo, a construção de novos *datasets* e a exploração de arquiteturas híbridas que integrem características espaciais, cinemáticas e espectrais representam direções promissoras para futuras pesquisas no reconhecimento automático de Libras. Dessa forma, a principal contribuição deste trabalho reside na construção de um novo conjunto de dados focado em sinais técnicos de informática e na avaliação do conjunto de dados em um modelo LSTM, visando ampliar a diversidade de sinais disponíveis para pesquisa e explorar o potencial de reconhecimento de terminologias específicas da área de tecnologia da informação em Libras.

3 METODOLOGIA

Este trabalho adota uma abordagem experimental baseada na construção de um *dataset* personalizado, na extração de características espaciais por meio de pontos-chave anatômicos (*landmarks*) e no treinamento de um modelo de aprendizado de máquina para o reconhecimento de sinais da Libras voltados à área de TI. As etapas metodológicas estão descritas nas subseções a seguir.

3.1 Coleta de Dados

Os sinais utilizados neste trabalho foram coletados a partir do canal *Libras IFSC Chapecó*[3], disponível na plataforma YouTube, cujo conteúdo é voltado à difusão da Língua Brasileira de Sinais aplicada ao contexto da informática (Libras. . . , 2025). A escolha dessa fonte se justifica pelo fato de ser um canal de uma instituição federal de ensino. Todas as gravações do sinais são padronizadas, primeiro é apresentado o sinal em linguagem escrita, depois é apresentado o sinal em Libras. Além disso a qualidade do vídeos variava entre 480p e 720p, com iluminação adequada, fundo neutro, e o sinalizador enquadrado de forma centralizada.

A partir dos vídeos disponibilizados, foram selecionados 30 sinais distintos. A seleção priorizou termos técnicos recorrentes em disciplinas do curso de Bacharelado em Tecnologia da Informação, abrangendo áreas como programação, banco de dados, redes de computadores, sistemas operacionais, arquitetura de computadores e desenvolvimento de software. Essa seleção específica justifica-se pela escassez de bases de dados públicas voltadas a terminologias técnicas, uma vez que datasets populares como o MINDS-Libras (Guimarães et al., 2019) focam em vocabulário cotidiano. A Figura 3 apresenta a lista dos sinais coletados, juntamente com um quadro representativo de cada um.

[3] <https://www.youtube.com/@librasifscchapeco666/videos>

Figura 3 – Sinais de Libras presentes no dataset *(continua)*

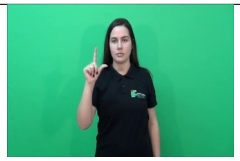
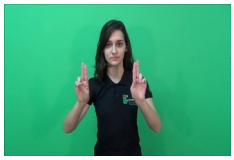
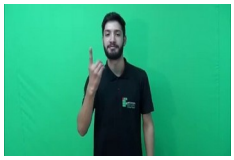
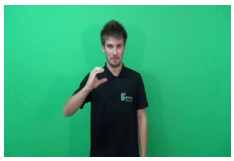
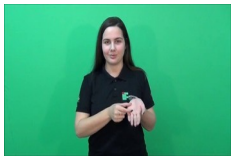
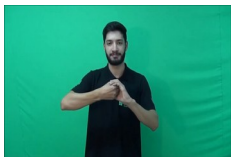
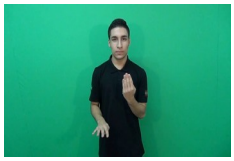
Sinal	**Frame Representativo**
Algoritmo	
Backup	
Binário	
Bloqueado	
CPU	
Cartão de Memória	
Compactar	
Compilar	
Computador	

Figura 3 – Sinais de Libras presentes no dataset *(continuação)*

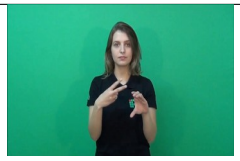
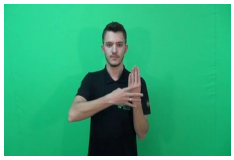
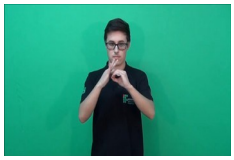
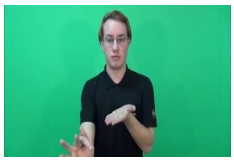
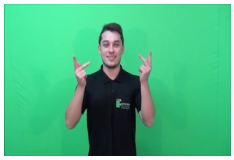
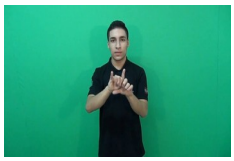
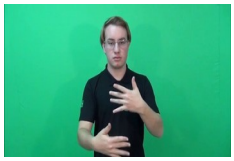
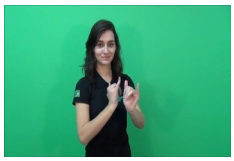
Sinal	**Frame Representativo**
Cooler	
Código	
Descompactar	
Fibra Óptica	
Formatar	
Hardware	
Importar	
Interface	
Internet	

Figura 3 – Sinais de Libras presentes no dataset *(continuação)*

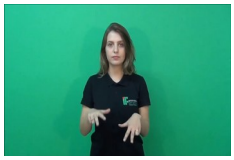
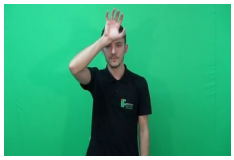
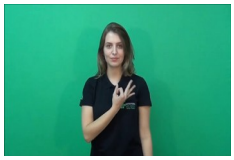
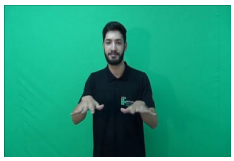
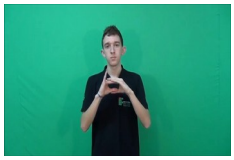
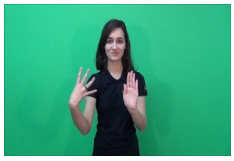
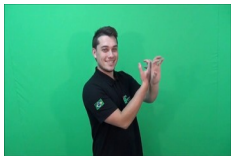
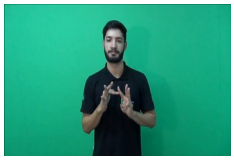
Sinal	**Frame Representativo**
Link	
Linux	
Memória RAM	
Número de IP	
Redes	
Roteador	
Servidor	
Site	
Software	

Figura 3 – Sinais de Libras presentes no dataset *(Conclusão)*

Sinal	**Frame Representativo**
Tecnologia	
WebCam	
Wi-Fi	

Fonte: Adaptado de *Libras IFSC Chapecó* (2016)

3.2 Organização da Base de Dados

A extração dos quadros (*frames*) dos vídeos coletados foi realizada com o auxílio da biblioteca OpenCV. Cada vídeo do dataset foi lido quadro a quadro, e cada imagem extraída foi processada por algoritmos de detecção de mãos da biblioteca MediaPipe, com o objetivo de verificar se havia uma mão visível e corretamente reconhecida naquele instante. Somente os quadros em que essa detecção foi bem-sucedida foram considerados válidos e armazenados para uso posterior; quadros em que a mão estava ausente ou não reconhecida foram descartados.Com o objetivo de reduzir a redundância entre imagens consecutivas, adotou-se uma taxa de amostragem de um quadro a cada cinco exibidos. Essa escolha se justifica pelo fato de que os vídeos foram capturados a uma taxa de quadros relativamente alta, de modo que quadros muito próximos no tempo tendem a registrar praticamente a mesma pose da mão, com variações mínimas de posição e configuração dos dedos. Por carregarem pouca informação adicional em relação ao quadro anterior, esses quadros intermediários representam redundância visual: mantê-los aumentaria o volume de dados sem agregar variabilidade relevante ao conjunto. Ao descartar parte desses quadros, reduziu-se esse excesso sem comprometer a representatividade do gesto ao longo do vídeo.

Cada imagem extraída e validada foi armazenada em um diretório correspondente à sua respectiva classe (ou seja, ao sinal representado), o que resultou em uma estrutura de pastas organizada por classe, contendo os quadros estáticos extraídos de todos os vídeos daquele sinal. Essa organização foi posteriormente utilizada como base para a etapa de aumento de dados, na qual a sequência de quadros de cada classe deu origem à geração de amostras sintéticas adicionais, conforme detalhado na seção seguinte.

3.3 Geração de Amostras Sintéticas

A fim de ampliar o volume de dados disponíveis e aumentar a variabilidade das amostras, foi gerado um *dataset* sintético a partir do *dataset* original, com 100 amostras por classe. A geração das novas imagens foi realizada por meio da classe `ImageDataGenerator` da biblioteca TensorFlow/Keras, que aplica transformações geométricas e fotométricas controladas sobre as imagens originais.

O processo foi conduzido da seguinte forma: para cada classe, a sequência de quadros estáticos extraída na etapa anterior foi carregada, e então para cada amostra sintética a ser gerada, foi sorteado um conjunto único de parâmetros de transformação (por exemplo, um ângulo de rotação específico, um fator de zoom, um grau de cisalhamento). Esse mesmo conjunto de parâmetros foi então aplicado de forma consistente a todos os quadros pertencentes àquela sequência — ou seja, todos os quadros de uma mesma amostra sofreram exatamente a mesma transformação, evitando distorções incoerentes entre quadros consecutivos de um mesmo sinal. Esse procedimento foi repetido sucessivamente até que o número desejado de amostras sintéticas por classe fosse atingido, e cada amostra gerada foi armazenada em um subdiretório próprio.

As transformações aplicadas incluíram: *zoom* aleatório, rotação da imagem, cisalhamento (*shear*) e espelhamento horizontal (*flip*). Tais transformações visam simular variações naturais na execução dos sinais, como diferenças de posicionamento, ângulo e lateralidade, contribuindo para a robustez do modelo treinado. A Figura 4 ilustra a comparação entre um quadro original e uma amostra gerada sinteticamente com transformações aleatórias.

Figura 4 – Comparação entre frame original e frame gerado sinteticamente para o sinal de Código.

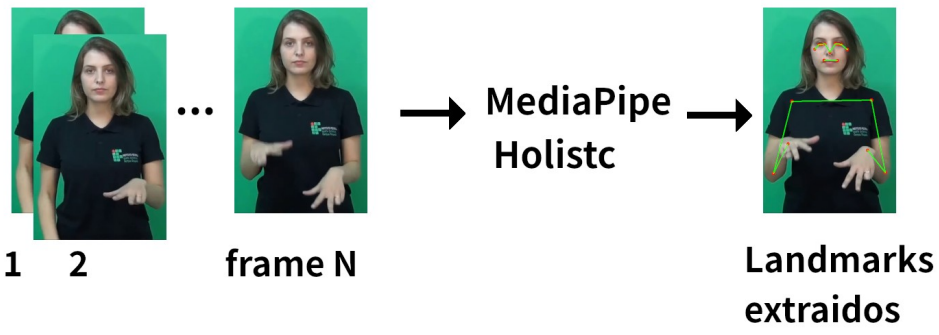
(a) Frame original (b) Frame gerado

Fonte: Adaptado de *Libras IFSC Chapecó* (2016)

3.4 Extração de *Landmarks*

A representação dos sinais foi realizada por meio da extração de pontos-chave anatômicos (*landmarks*), os quais correspondem a coordenadas espaciais tridimensionais de regiões de referência do corpo humano. O procedimento, ilustrado na Figura 5, compreende desde o recebimento do quadro em vídeo até a geração do vetor de características. Essa abordagem permite representar a postura e o movimento das mãos e do corpo de forma compacta e independente de fatores como iluminação e cor de pele.

Figura 5 – visualização simplificada da extração de *Landmarks*



Fonte: Adaptado de *Libras IFSC Chapecó* (2016)

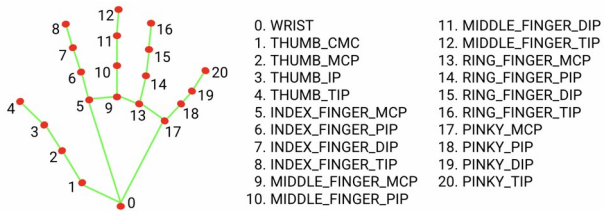
Nesse fluxo, a imagem passa por um pré-processamento de cor antes de ser submetida

aos algoritmos de detecção, todas as imagens são verificadas e convertidas para o espaço de cores RGB (*Red, Green, Blue*) quando necessário, uma vez que a biblioteca utilizada exige esse padrão de entrada.

Para a extração dos *landmarks*, foram empregados dois módulos da biblioteca Google *MediaPipe*: *Hands* e *Holistic*. O módulo *Hands* é responsável pela detecção e extração dos pontos-chave das mãos. Esse modelo identifica até 21 *landmarks* por mão, representando as articulações e as extremidades dos dedos. Esses pontos capturam a posição aproximada das mãos no espaço tridimensional, a extração desses *landmarks* são fundamentais para a identificação do parâmetro de Configuração de Mão, uma vez que a disposição relativa e a orientação desses pontos definem a forma específica que a mão assume para realizar um sinal de Libras. Adicionalmente, a variação temporal dessas coordenadas ao longo dos quadros permite ao sistema capturar o parâmetro de Movimento.

Ademais para garantir a qualidade das amostras, foram descartados os quadros que apresentassem mais de duas mãos detectadas, duas mãos pertencentes ao mesmo lado do corpo (ambas direitas ou ambas esquerdas) ou ausência total de mãos na imagem. A Figura 6 ilustra os *landmarks* extraídos das mãos.

Figura 6 – Representação dos 21 *landmarks* da mão detectados pelo módulo *Hands* do *MediaPipe*.

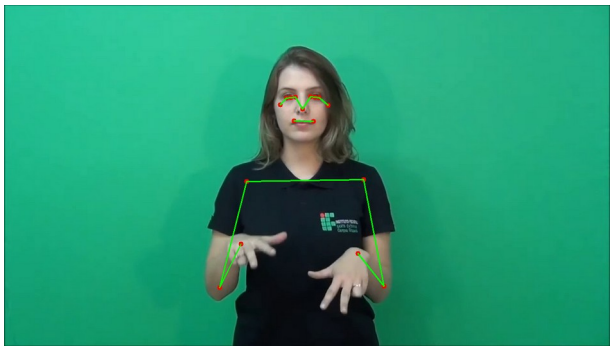


Fonte: MediaPipe Hands (2026)

O módulo *Holistic* foi utilizado para a detecção dos pontos-chave da parte superior do corpo, incluindo ombros, cotovelos, punhos e regiões do rosto relevantes para a interpretação dos sinais. Do ponto de vista fonológico, a extração desses *landmarks* é fundamental para determinar o parâmetro de Locação, pois fornece as referências espaciais (como o rosto ou o tronco) onde o sinal é articulado. Foram considerados 17 pontos da região superior do corpo. Quadros nos quais não houve detecção desses pontos foram descartados do *dataset*. A Figura 7 apresenta um exemplo dos *landmarks* corporais extraídos para o sinal de Linux.

Cada ponto extraído é representado por três coordenadas normalizadas no espaço tridimensional (x, y, z). Considerando os 21 pontos de cada mão (42 no total, para ambas as mãos) e os 17 pontos corporais, obtêm-se $59 \times 3 = 177$ características por quadro. Juntamente com as colunas de identificação da classe (nome e índice), o *dataframe* resultante, construído com a biblioteca Pandas, totaliza 179 colunas e 28.250 linhas, compostas pelas amostras de todos os sinais coletados e sintetizados.

Figura 7 – *Landmarks* corporais extraídos pelo módulo *Holistic* do MediaPipe para o sinal de Linux.



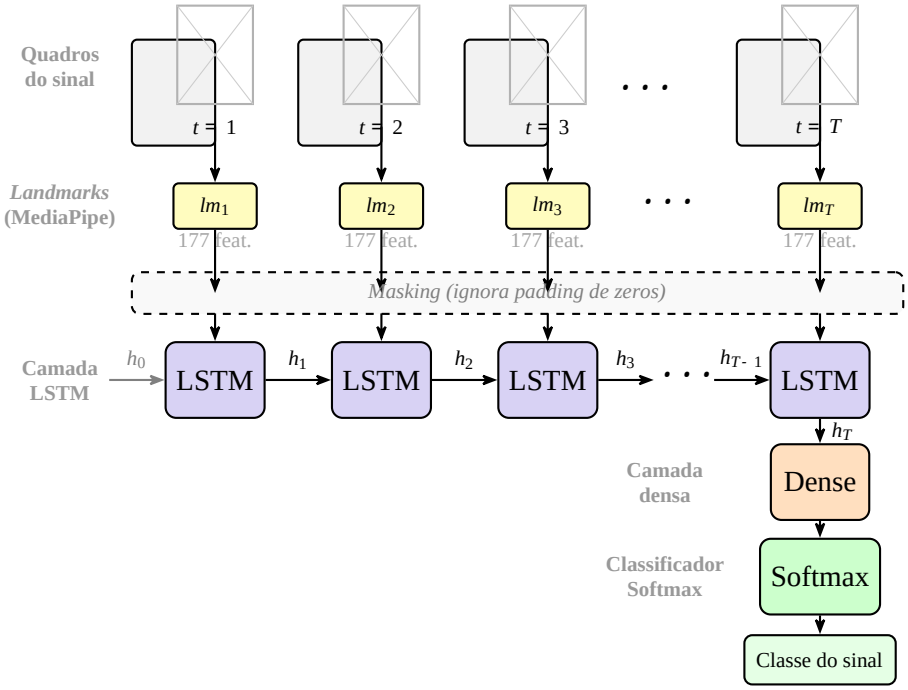
Fonte: Adaptado de *Libras IFSC Chapecó* (2016)

3.5 Treinamento do Modelo

O modelo selecionado para a tarefa de reconhecimento dos sinais foi uma rede neural recorrente do tipo LSTM, arquitetura especialmente adequada para o processamento de sequências temporais, o que é inerente à natureza dinâmica dos sinais em Libras.

A Figura 8 ilustra o fluxo completo de processamento da rede LSTM ao longo de uma sequência de quadros extraídos de um sinal, desde a entrada dos vetores de *landmarks* até a classificação final.

Figura 8 – Arquitetura da rede LSTM.



Fonte: Autoria própria (2026)

Cada quadro do sinal gera um vetor de 177 *landmarks*, que é processado sequencialmente pelas células LSTM; o estado oculto final h_T alimenta uma camada densa seguida de *Softmax* para classificação do sinal.

O conjunto de dados foi particionado em 80% para treinamento e 20% para teste, respeitando a distribuição proporcional entre as classes. Devido à variação no comprimento das sequências entre as diferentes amostras, foi aplicada a técnica de *padding* com zeros, de modo a padronizar todas as entradas com o comprimento máximo observado no *dataset* (Géron, 2022). Para que os valores inseridos artificialmente não influenciassem o aprendizado, foi incorporada uma camada de *Masking*, responsável por ignorar as posições preenchidas com zero durante o processamento da rede.

O critério de parada do treinamento foi definido por meio de *Early Stopping*: caso a acurácia de validação não apresentasse melhora por 15 épocas consecutivas, o treinamento era interrompido automaticamente, prevenindo o sobreajuste (*overfitting*) ao conjunto de treinamento (Géron, 2022). Adicionalmente, foi aplicada a técnica de *Dropout* com taxa de 10%, que desativa aleatoriamente uma fração dos neurônios em cada iteração, forçando a rede a aprender representações mais generalizáveis e menos dependentes de neurônios individuais.

3.6 Avaliação de Desempenho

O desempenho do modelo treinado foi avaliado por meio de métricas consolidadas para tarefas de classificação multiclasse: acurácia, precisão (*precision*), revocação (*recall*) e F1-score (Géron, 2022). Essas métricas são derivadas dos valores da matriz de confusão, onde VP denota os verdadeiros positivos, VN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos.

A acurácia expressa a proporção global de predições corretas em relação ao total de amostras avaliadas:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}. \tag{1}$$

A precisão mensura, dentre todas as amostras classificadas como positivas pelo modelo, qual fração é de fato positiva:

$$\text{Precisão} = \frac{VP}{VP + FP}. \tag{2}$$

A revocação (*recall*) indica, dentre todas as amostras verdadeiramente positivas, qual fração foi corretamente identificada pelo modelo:

$$\text{Revocação} = \frac{VP}{VP + FN}. \tag{3}$$

O *F1-score* representa a média harmônica entre precisão e revocação, combinando ambas as perspectivas em um único indicador e sendo particularmente informativo em cenários com possível desbalanceamento entre classes:

$$\text{F1-score} = 2 \cdot \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}. \tag{4}$$

No contexto multiclasse, as métricas de precisão, revocação e F1-score foram calculadas por meio da estratégia *macro average*, na qual o valor de cada classe é computado individualmente e então se obtém a média aritmética entre todas as classes, conferindo peso igual a cada uma delas independentemente do número de amostras.

4 SIMULAÇÃO E RESULTADOS

Este capítulo apresenta os resultados obtidos a partir da implementação do modelo de aprendizado de máquina proposto para a predição de sinais em Libras relacionados à área de Tecnologia da Informação. A análise dos resultados é realizada por meio de métricas quantitativas, bem como pela interpretação qualitativa dos padrões de classificação observados na matriz de confusão. A discussão dos resultados inclui a identificação de casos específicos de confusão entre sinais visualmente semelhantes, proporcionando *insights* sobre as limitações e potencialidades do modelo desenvolvido.

4.1 Resultados

Após o treinamento do modelo LSTM, avaliou-se a eficiência do sistema de classificação por meio de métricas quantitativas. Os resultados obtidos são apresentados no Quadro 2, os quais permitem uma análise abrangente do desempenho do modelo proposto.

Quadro 2 – Métricas do modelo LSTM.

Métrica	**Valor**
Acurácia	99,34%
Precisão	99,36%
Revocação	99,34%
F1-score	99,34%

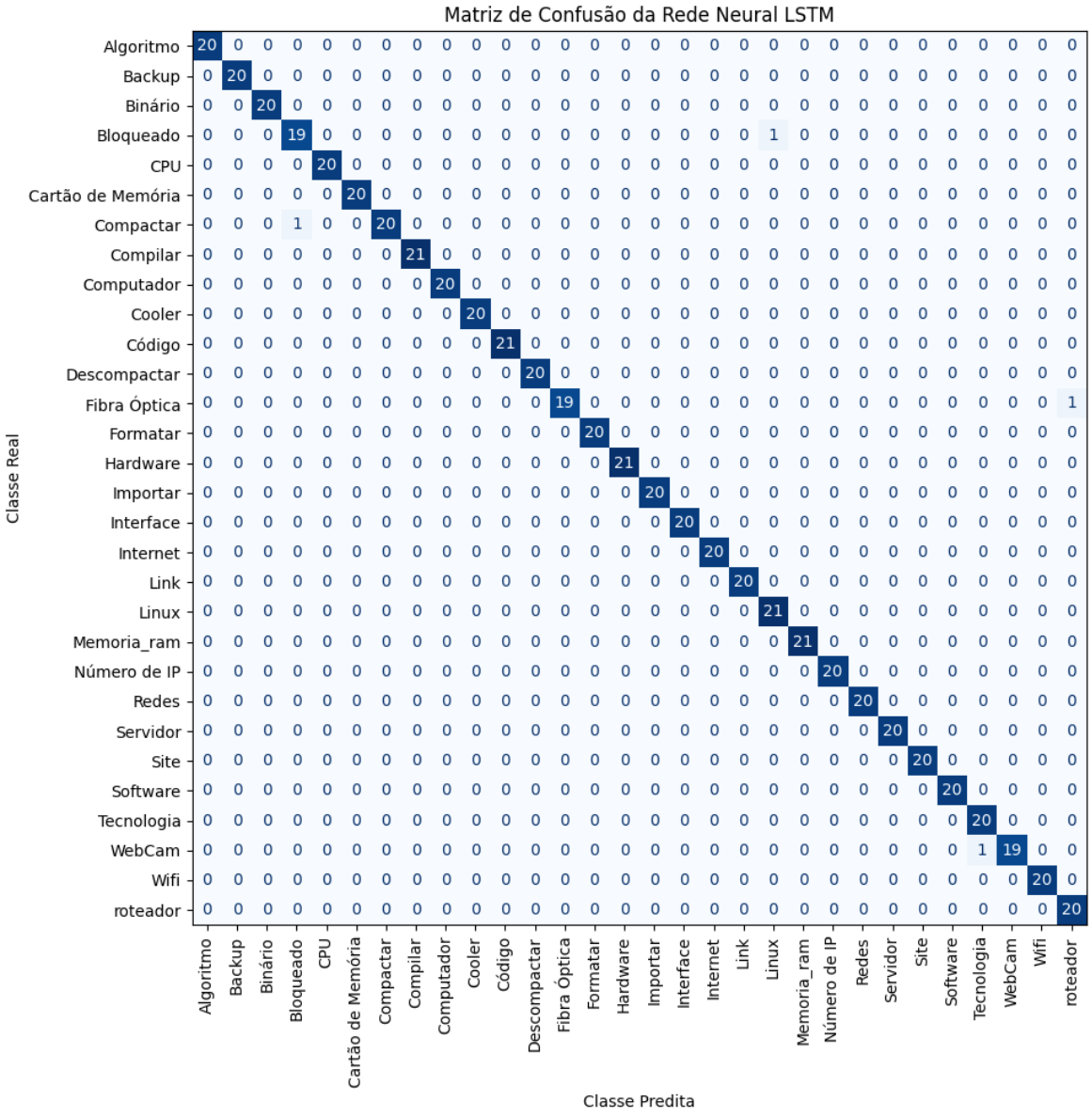
Fonte: Autoria própria (2026)

A análise dos dados apresentados no Quadro 2 revela que o modelo desenvolvido demonstrou capacidade preditiva, evidenciada pela acurácia de 99,34%, o que indica que o sistema classificou corretamente 99,34% das amostras do conjunto de teste. A precisão de 99,36% atesta que, dentre todas as predições positivas realizadas pelo modelo, a grande maioria corresponde a classificações verdadeiramente positivas, minimizando a ocorrência de falsos positivos. A revocação de 99,34% demonstra que o modelo foi capaz de identificar corretamente a quase totalidade dos casos positivos reais, apresentando apenas um número reduzido de falsos negativos. Por fim, o F1-score de 99,34% confirma o equilíbrio harmônico entre precisão e revocação, consolidando a robustez do modelo para a tarefa de classificação dos sinais de Libras relacionados à Tecnologia da Informação.

Para uma análise mais detalhada do comportamento do classificador, examinou-se a matriz de confusão apresentada na Figura 9. Nessa representação, a diagonal principal contém os valores de classificações corretas, enquanto os elementos fora da diagonal correspondem aos erros de classificação, permitindo identificar padrões específicos de confusão entre as classes.

A análise da matriz de confusão revela alguns casos pontuais de classificação incorreta que merecem atenção. O sinal de Compactar, por exemplo, foi confundido com o sinal de Bloqueado

Figura 9 – Resultado da classificação via modelo LSTM.



Fonte: Autoria própria (2026)

em uma instância. Conforme ilustrado na Figura 10, esses sinais apresentam similaridade estrutural considerável, compartilhando a mesma configuração manual básica: um punho fechado conectando-se à palma da outra mão, com a diferença que para o sinal Compactar os dedos cobrem a mão fechada. A Locação das mãos não é tão próxima, para o sinal de Compactar as mãos são posicionadas mais próximas ao rosto. O Movimento de ambos sinais envolve a aproximação das mãos, mas o punho do sinal Bloqueado gira ao encontrar a palma da outra mão. O punho fechado e maneira como as mão se aproximam, podem ter contribuído para a confusão do modelo, resultando em uma classificação incorreta. Adicionalmente, observou-se que o sinal de Bloqueado também foi erroneamente classificado como Linux em uma ocorrência. Ambos sinais apresentam momentos em que as mãos se localizam próximos a base do tronco, a

Configuração de Mão e Movimento são diferentes.

Figura 10 – Comparação entre os sinais Bloqueado e Compactar.

(a) Bloqueado (b) Compactar

Fonte: Canal *IFSC Chapecó* YouTube (2016)

Na análise da coluna correspondente ao sinal de Tecnologia na matriz de confusão, identificou-se um valor unitário na linha de WebCam. Analisando parâmetros fonológicos vemos que para os dois sinais temos um movimento ascendente que levas as duas mãos para o mesmo lado do corpo, a Configuração de Mão não é semelhante, ainda assim o modelo identificou similaridades entre essas classes e produziu um falso negativo.

Outro padrão de confusão observado envolve os sinais de Fibra Óptica e Roteador. Como demonstrado na Figura 11, esses sinais exibem padrões linguísticos com elevado grau de semelhança, especialmente no que concerne ao posicionamento das mãos à frente do tronco e à configuração manual adotada, além disso o Movimento de aproximação da mãos é semelhante, porém para o sinal de Fibra Óptica, após as mãos se juntarem ela se separam. A estrutura linguística semelhante explicam a dificuldade do modelo em estabelecer uma distinção clara entre essas classes, resultando em classificações imprecisas pontuais.

Figura 11 – Comparação entre os sinais Fibra Óptica e Roteador.

(a) Fibra Óptica (b) Roteador

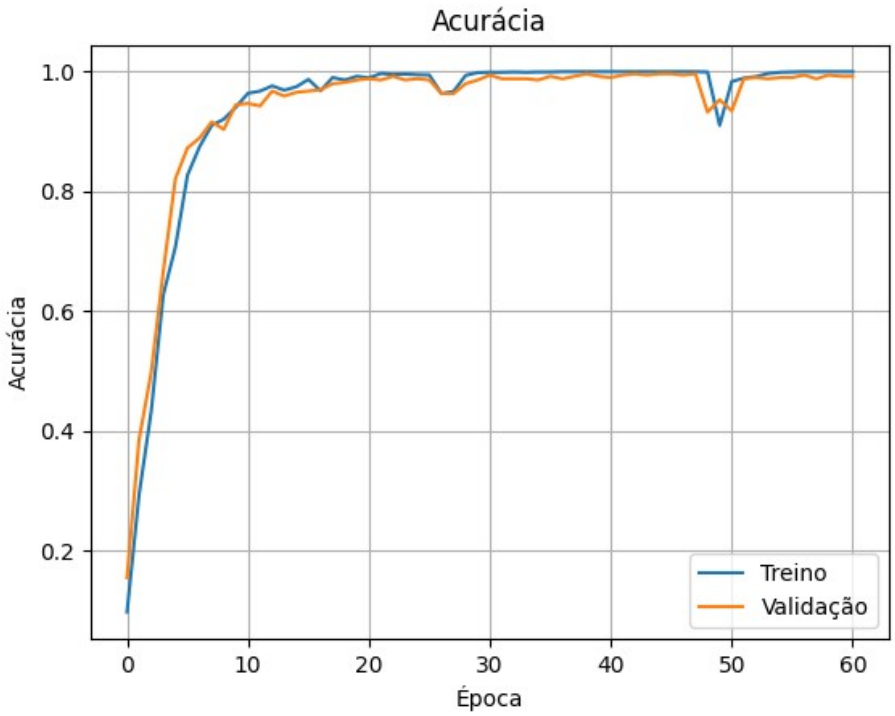
Fonte: Canal *IFSC Chapecó* YouTube (2016)

Apesar dos casos pontuais de confusão entre sinais visualmente semelhantes, os resul-

tados obtidos demonstram que a arquitetura LSTM apresentou bom desempenho na tarefa de reconhecimento de sinais de Libras relacionados à área de Tecnologia da Informação. A capacidade intrínseca das redes LSTM de manter e processar informações ao longo de sequências temporais (Hochreiter; Schmidhuber, 1997) mostrou-se particularmente adequada à natureza sequencial dos dados gestuais, resultando em métricas superiores a 99% em todas as medidas de desempenho avaliadas. Os erros observados, embora raros, são atribuíveis a similaridades estruturais legítimas entre determinados sinais, representando desafios naturais do domínio de reconhecimento gestual que poderiam ser mitigados com o enriquecimento do conjunto de dados de treinamento ou com o emprego de técnicas de aumento de dados focadas nas classes que apresentaram confusão.

A evolução do modelo durante o treino pode ser observada na Figura 12 que apresenta a acurácia do modelo ao longo das épocas. É possível observar que as curvas de aprendizagem de Treino e Validação convergem durante todo o período. Ligeiras sobreposições da curva de Validação em relação ao Treino podem ser indicativas da atuação do mecanismo de *Dropout*, que é empregado somente na fase de treino. Visto que o *Dropout* desativa aleatoriamente uma parcela dos neurônios a cada iteração para evitar sobreajuste, a robustez do modelo durante o teste de validação é nominalmente maior e, portanto, tende a ter melhores valores.

Figura 12 – Curvas de acurácia durante o treino e a validação do modelo LSTM.

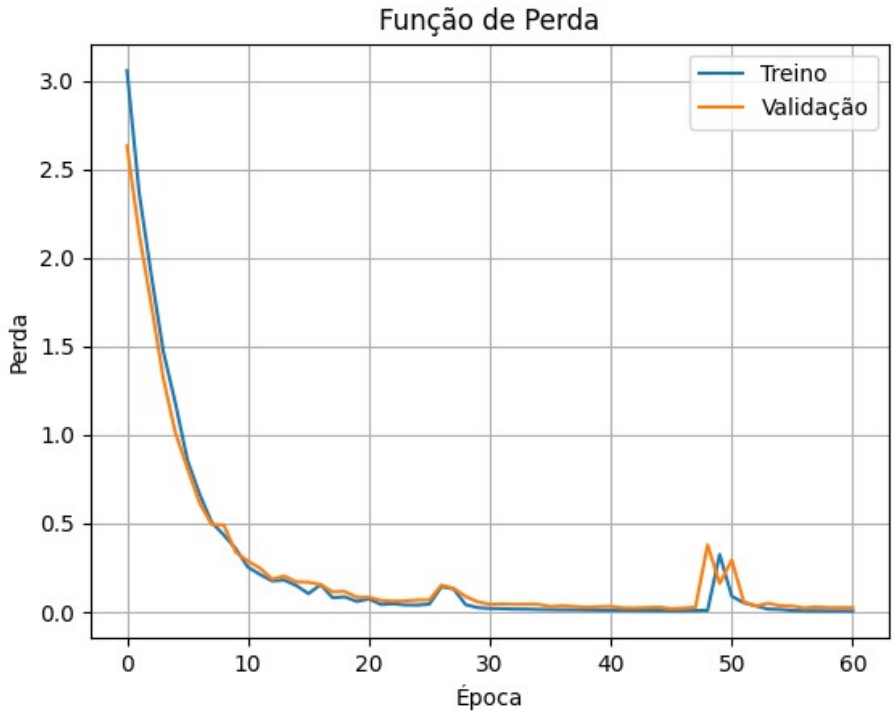


Fonte: Autoria própria (2026)

Em complemento, a Figura 13 detalha a função da Perda tanto para o conjunto de treino quanto para o de validação, mapeando o erro do classificador ao longo das épocas. Observa-se

uma queda constante nas dez primeiras épocas seguida de uma estabilização próxima a zero que só é interrompida em duas ocasiões. Conforme observado nas oscilações ocorridas entre as épocas 20 e 30, e de maneira mais proeminente próximo à época 49, esses picos pontuais de erro são prontamente mitigados devido à própria dinâmica da arquitetura do modelo LSTM. Através do mecanismo de *forget gate*(portão de esquecimento) introduzido por Gers, Schmidhuber e Cummins (1999), informações irrelevantes são descartadas. Esse isolamento impede que erros sejam passados adiante ou afetem o aprendizado das demais classes, garantindo o rápido reestabelecimento da estabilidade do modelo.

Figura 13 – Curvas de perda durante o treino e validação do modelo LSTM.



Fonte: Autoria própria (2026)

5 CONSIDERAÇÕES FINAIS

Este trabalho buscou responder se uma arquitetura de rede neural recorrente do tipo LSTM seria capaz de capturar padrões temporais para permitir a classificação de sinais de Libras, no domínio técnico de TI, com alta precisão. Os resultados alcançados confirmam positivamente essa hipótese, uma vez que o modelo desenvolvido obteve uma acurácia, Revocação e *F1-score* de 99,34% e precisão de 99,36% na identificação dos sinais propostos.

Dessa forma, os objetivos estabelecidos para esta pesquisa foram integralmente atingidos. A técnica de aumento de dados foi usada para diversificar as amostras do base de dados e diminuir o risco de *overfitting*. Foi possível estruturar um *dataset* inédito com 30 sinais específicos da área de TI, suprindo uma lacuna identificada na literatura. Os *Landmarks* foram extraídos através dos módulos da biblioteca *MediaPipe*, enquanto a arquitetura LSTM provou-se eficaz para processar a natureza simultânea dos fonemas da Libras.

Foi possível observar que para o caso de sinais com parâmetros fonológicos semelhante, como no caso dos pares, Fibra Óptica e Roteador, e também Bloqueado e Compactar, o modelo apresentou dificuldade na classificação. Esses erros, demonstram que ainda há potencial de evolução, ainda que estes não comprometam a robustez geral do modelo. Vale pontuar também que embora a geração de amostras sintéticas tenha sido empregada para aumentar a variabilidade da base de dados e mitigar o *overfitting*, essas técnicas não substituem a necessidade de coleta de dados reais. Portanto, o modelo não pode ser exposto à variabilidade natural de sinais apresentados por diferentes intérpretes, em diferentes cenários e diferentes velocidade de execução.

Para futuros trabalhos pode-se aumentar o vocabulário de sinais de TI. Fazer uso de outras arquiteturas baseadas em redes neurais recorrentes como BiLSTM. Outra possibilidade seria o desenvolvimento de uma aplicação que aplica a classificação de sinais em tempo real.

Por fim, espera-se que este trabalho contribua para a democratização do acesso ao conhecimento de TI em Libras, para reduzir a barreira linguística e contribuir para o desenvolvimento de ferramentas assistivas.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Em conformidade com as diretrizes de integridade científica e boas práticas acadêmicas recomendadas pela CAPES, os autores declaram que utilizaram ferramentas de inteligência artificial, incluindo ChatGPT exclusivamente para geração de imagem e auxiliar na melhoria da linguagem e legibilidade deste manuscrito. O uso dessas ferramentas foi realizado sob supervisão humana, e os autores revisaram e editaram cuidadosamente o conteúdo, assumindo total responsabilidade pelo texto final publicado.

REFERÊNCIAS

ALMEIDA, S. G. M. et al. **Libras-10 Dataset**. Zenodo, 2019. Disponível em: <https://doi.org/10.5281/zenodo.3229958>. Acesso em: 4 jun. 2026.

ARCANJO, L. d. S. et al. Automatic time-aware recognition of brazilian sign language based on dynamic time warping. In: SBC. **Brazilian Symposium on Multimedia and the Web (WebMedia)**. [S.l.], 2024. p. 72–79.

BRASIL, L. d. D. Lei nº 10.436, de 24 de abril de 2002. dispõe sobre a língua brasileira de sinais-libras e dá outras providências. **Diário Oficial da União**, Poder Legislativo Brasília, DF, v. 1, p. 23–23, 2002.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.

CARNEIRO, A. C.; SILVA, L. B.; SALVADEO, D. P. Efficient sign language recognition system and dataset creation method based on deep learning and image processing. In: SPIE. **Thirteenth International Conference on Digital Image Processing (ICDIP 2021)**. [S.l.], 2021. v. 11878, p. 11–19.

CASTRO, G. Z. D.; GUERRA, R. R.; GUIMARÃES, F. G. Automatic translation of sign language with multi-stream 3d cnn and generation of artificial depth maps. **Expert Systems with Applications**, Elsevier, v. 215, p. 119394, 2023.

EDGE, G. A. **Guia de detecção de pontos de referência manuais**. 2025. Disponível em: <https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker?hl=pt-br>. Acesso em: 4 jun. 2026.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S.l.]: "O'Reilly Media, Inc.", 2022.

GERS, F.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: continual prediction with lstm. In: **1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)**. [S.l.: s.n.], 1999. v. 2, p. 850–855 vol.2.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine learning**, Springer, v. 63, n. 1, p. 3–42, 2006.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Disponível em: <http://www.deeplearningbook.org>. Acesso em: 26 mai. 2026.

Governo Federal do Brasil. **VLibras – Plataforma de Acessibilidade em Língua Brasileira de Sinais**. 2025. Disponível em: <https://www.gov.br/governodigital/pt-br/acessibilidade-e-usuario/vlibras>. Acesso em: 13 mar. 2026.

GUIMARÃES, G. et al. **MINDS-Libras Dataset**. 2019. Version 1. Dataset (Lexical/Conceptual Resource). Disponível em: <https://live.european-language-grid.eu/catalogue/lcr/21705>. Acesso em: 4 jun. 2026.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa nacional de saúde**: 2019: ciclos de vida: Brasil. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2021.

KLEIN, M. Movimentos surdos e os discursos sobre surdez, educação e trabalho: a constituição do surdo trabalhador. In: **24ª Reunião Nacional da ANPEd**. Caxambu, Minas Gerais: [s.n.], 2001. p. 1–18.

LEMOS, S. K. d. S. de et al. Tecnologias digitais e acesso à informação: uma pesquisa com pessoas surdas. **Revista ACB: Biblioteconomia em Santa Catarina**, Associação Catarinense de Bibliotecários, v. 24, n. 1, p. 129–143, 2019.

LIBRAS IFSC Chapecó. 2025. Canal do YouTube. Disponível em: <https://www.youtube.com/@librasifscchapeco666/videos>. Acesso em: 03 fev. 2026.

MORAIS, L.; ALMEIDA, W.; REGO, R. Machine learning approaches for efficient recognition of brazilian sign language. In: **Anais do XXI Simpósio Brasileiro de Sistemas de Informação**. Porto Alegre, RS, Brasil: SBC, 2025. p. 49–56. ISSN 3086-4836. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/34320>.

MORAIS, L. M. G. de et al. Análise da aplicação de visão computacional e aprendizado de máquina na tradução da libras para texto. In: AES, F. G.; FERREIRA, D.; BARRETO, G. (Ed.). **Anais do XVII Congresso Brasileiro de Inteligência Computacional (CBIC 2025)**. Belo Horizonte, MG: SBIC, 2025. p. 1–8.

QUADROS, R. M. d.; KARNOPP, L. B. **Língua de sinais brasileira**: estudos linguísticos. São Paulo: Artmed, 2007. 221 p. (Biblioteca Artmed. Linguística).

REGO, R. C.; MORAIS, L. M. de; ALMEIDA, W. M. Brazilian sign language recognition using deep learning based on fast fourier transform and kinematic features. **IEEE Access**, IEEE, v. 13, p. 202875–202892, 2025.

REZENDE, T. M.; ALMEIDA, S. G. M.; GUIMARÃES, F. G. Development and validation of a brazilian sign language database for human gesture recognition. **Neural Computing and Applications**, Springer, v. 33, n. 16, p. 10449–10467, 2021.

WFD, W. F. of the D. **UN Enable - Promoting the rights of Persons with Disabilities - Contribution by WFD**. 2003. Disponível em: <https://www.un.org/esa/socdev/enable/rights/contrib-wfd.htm>. Acesso em: 4 jun. 2026.