AI-Based Approaches for Brazilian Sign Language Recognition: A Systematic Literature Review

Insights on Methods, Metrics, and Resources for LIBRAS Recognition

Victor Costa
vcrcosta@inf.ufpel.edu.br
PPGC, ViTech - UFPel
Pelotas/RS, Brazil

João Pedro Tomaszewski
jptomaszewski@inf.ufpel.edu.br
PPGC - UFPel
Pelotas/RS, Brazil

Lucas Pereira
lspereira@inf.ufpel.edu.br
PPGC - UFPel
Pelotas/RS, Brazil

Brenda Salenave Santana
bssalenave@inf.ufpel.edu.br
PPGC - UFPel
Pelotas/RS, Brazil

Guilherme Corrêa
gcorrea@inf.ufpel.edu.br
PPGC, ViTech - UFPel
Pelotas/RS, Brazil

ABSTRACT

This Systematic Literature Review (SLR) investigates the application of Artificial Intelligence (AI), specifically with Computer Vision (CV) and Natural Language Processing (NLP) techniques for Brazilian Sign Language (LIBRAS) recognition, translation, and interpretation. Following the PRISMA protocol and using the Parsifal tool to support protocol definition and study management, 77 studies were initially identified. After removing duplicates and applying exclusion criteria, 23 studies were selected for in-depth analysis. The review addresses three research questions: (i) AI-based solutions currently employed for LIBRAS recognition, (ii) reported performance metrics and evaluation results, and (iii) datasets used for training and testing, along with their characteristics and limitations. The findings highlight advances in deep learning architectures and computer vision pipelines, while revealing persistent challenges such as limited integration of non-manual features, scarcity of large-scale and multimodal datasets, difficulties in continuous sign recognition, heterogeneous annotation practices, and inadequacy of common bilingual evaluation metrics for capturing syntactic and semantic differences between LIBRAS and Portuguese. Moreover, deployment-related factors, including inference time and computational cost, remain underexplored. These insights provide guidance for future research toward more accurate, efficient, and inclusive LIBRAS recognition systems.

KEYWORDS

Sign Language Recognition, Brazilian Sign Language, LIBRAS, Artificial Intelligence, Neural Networks, Literature Review

1 INTRODUCTION

The LIBRAS (LIngua BRAsileira de Sinais – Brazilian Sign Language) stands as a fundamental visuospatial language for the deaf community in Brazil, officially recognized by law [39, 40]. Despite its critical role in individual communication, public policy, civil rights, and social inclusion, LIBRAS remains relatively undisseminated [59], with a notable absence of effective public policies promoting its teaching and use. This reality underscores a pressing need for solutions that can bridge communication gaps and enhance the inclusion of deaf and hard-of-hearing individuals.

In this context, Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Deep Learning (DL) and Computer Vision (CV) emerge as highly promising avenues for developing systems capable of recognizing and interpreting sign language gestures, converting them into text or spoken language [1]. These technologies hold the potential to empower deaf and hard-of-hearing individuals, enabling more efficient and independent communication. Practical applications, such as AI-driven solutions for real-time interpretation utilizing CV and original AI engines, demonstrate the tangible capabilities and significant potential of these technologies [67]. Such solutions are typically developed based on extensive datasets of LIBRAS videos, which are used to train algorithms capable of identifying hand positions, digital articulation points, and the flow of phrases. The application of AI, ML, and CV in LIBRAS translation transcends mere technical problem-solving; it addresses a significant societal challenge related to inclusion and communication for the deaf community in Brazil.

While several systematic reviews have explored various aspects of sign language processing with technology, a significant gap persists in the literature for a comprehensive and up-to-date synthesis of LIBRAS-specific recognition solutions. For example, a systematic review by Kahlon and Singh [32] focused exclusively on text-to-sign language translation, consciously excluding studies on sign language recognition from its analysis. Other broad reviews, such as one by ZainEldin et al. [67], offer a general overview of AI in deaf and mute communication but do not specialize in LIBRAS's unique linguistic and technical challenges. Furthermore, reviews that do address specific aspects, like the one by Prietch et al. [41], often focus on human-centric user studies and ethical considerations, rather than the technical models and performance metrics. These works, while valuable, highlight the absence of a dedicated review. Our work directly addresses this need by providing a detailed analysis of AI-based solutions specifically for recognizing and interpreting LIBRAS gestures, thereby filling a critical gap and consolidating the state of the art in this specialized domain.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.
© 2025 SBC – Brazilian Computing Society.
ISSN 2966-2753

This Systematic Literature Review (SLR) aims to fill a gap in current knowledge and provide a comprehensive and updated understanding of the state-of-the-art in the application of AI, ML, DL and CV for the translation, recognition, and interpretation of LIBRAS. By doing so, it seeks to address the need for a rigorous synthesis of available empirical evidence, informing researchers and practitioners about the advancements and challenges in this area. The use of the SLR method allows for a systematic review and identification of relevant scholarly works, where each stage adheres to established protocols. This approach ensures a rigorous, unbiased, and replicable compilation and analysis of existing knowledge, helping to avoid subjective identification or evaluation and contributing valuable insights on AI-based LIBRAS recognition. This review will specifically focus on works published within the last ten years (2015-2025) to capture the most recent advancements in this rapidly evolving field.

This paper is organized into eight sections, each addressing a key component of the systematic review. Section 2 introduces the fundamental technologies related to Brazilian Sign Language recognition and translation, with emphasis on Computer Vision and Natural Language Processing. Section 3 details the systematic review process, including the research questions, inclusion and exclusion criteria, databases consulted, and study selection strategy. Section 4 describes the main findings obtained from the selected studies, structured around the defined research questions. Section 5 analyzes trends, divergences, and methodological limitations identified across the reviewed works. Section 6 highlights current research gaps, unresolved challenges, and suggestions for future investigations. Section 7 summarizes the key contributions of the review and reinforces its relevance to the field. Finally, Section 8 discusses the potential threats to validity and the limitations of the review process.

2 BACKGROUND

To understand the application of AI in LIBRAS translation and recognition, it is important to revisit the fundamental technologies introduced earlier in this article: CV and NLP. These fields have advanced significantly in recent years, enabling robust, accurate, and, in many cases, real-time translation systems.

In the context of LIBRAS translation and recognition, CV is responsible for capturing and interpreting the visual aspects of sign language, including hand gestures, facial expressions, and body posture [12]. NLP, in turn, structures and interprets the recognized gestures into coherent linguistic units, allowing their translation into text or speech [25]. The combined use of CV and NLP forms the basis of end-to-end pipelines capable of processing LIBRAS in videos and delivering consistent results [5, 10].

The following subsections outline the main principles, methods, and applications of these technologies in the context of LIBRAS recognition and translation, providing the groundwork for subsequent analyses of models, datasets, and results.

2.1 Computer Vision

In LIBRAS recognition and translation systems, Computer Vision (CV) acts as the central component responsible for transforming raw visual input captured by cameras or sensors into representations that can be processed by artificial intelligence models. Beyond simply detecting the presence of hands or bodies, modern CV analyzes complex spatial and temporal patterns, enabling the detection of dynamic gestures, the precise segmentation of regions of interest, and the extraction of multimodal features (hands, face, body) [5, 13, 48]. These capabilities stem from the evolution of architectures and methods ranging from handcrafted descriptors and semantic segmentation to specialized deep neural networks. The following paragraphs present studies that explore different CV strategies applied to LIBRAS, highlighting recent advances, challenges, and innovations.

Numerous studies have explored various CV techniques for LIBRAS recognition. For example, da Silva et al. [15] proposed a multi-stream architecture for sign recognition in the healthcare context, integrating spatial and temporal visual cues to improve performance in domain-specific applications. Similarly, da Silva et al. [16] presented a two-stream model based on 3D convolutional neural networks (CNN), demonstrating effective recognition of dynamic gestures through the integration of motion and appearance features.

Gesture segmentation is another fundamental component of sign recognition. Sarma et al. [52] employed attention-based models for semantic hand segmentation, achieving accurate localization of hand regions in complex backgrounds, which is essential to isolate gestures from noisy visual inputs. Furthermore, Jyoti Dutta et al. [31] explored semantic segmentation techniques using deep neural networks to improve gesture recognition in continuous sequences of signs.

Skeleton-based approaches have also gained attention due to their robustness and efficiency. Alves et al. [4] proposed the use of skeletal image representations to improve recognition in LIBRAS, reducing the computational burden while preserving critical gesture information. These methods are particularly suitable for mobile or low-power devices. Similarly, Fanucchi et al. [24] perfected a video mask autoencoder for sign recognition in augmented reality applications, exemplifying the integration of CV with immersive technologies to meet real-world communication needs.

Deep learning-based gesture recognition systems often rely on large-scale datasets. In this context, Bharti et al. [9] introduced a keyframe extraction technique based on error correction to improve dynamic hand gesture recognition, aiming to reduce temporal redundancy while preserving gesture integrity. Silveira et al. [55] and Silveira et al. [56] explored generative models to synthesize gestures in LIBRAS, enabling the creation of digital signage and contributing to data augmentation in resource-poor scenarios.

Taken together, these contributions highlight the transformative role of CV in LIBRAS research. From handcrafted feature extraction [49] to advanced attention mechanisms and neural networks, the evolution of visual recognition techniques continues to shape the development of accessible communication tools for the deaf community in Brazil.

2.2 Natural Language Processing

NLP is a subfield of AI that focuses on enabling machines to understand, interpret, and generate human language [30]. In the context of LIBRAS translation, NLP is essential in the post-processing phase, where sequences of recognized signs are converted into grammatically correct and semantically coherent sentences in Brazilian Portuguese [25].

Translating between LIBRAS and spoken Portuguese presents specific challenges due to the structural differences between sign languages and spoken languages [17]. While Portuguese follows a linear Subject-Verb-Object structure, LIBRAS is based on a spatial-visual grammar, incorporating simultaneous expressions, facial cues, and flexible word order [19]. Therefore, NLP models designed for LIBRAS translation must account for these linguistic and semantic incompatibilities [17, 19].

Several studies [1, 64, 68] have explored sequence-to-sequence neural network architectures, often enhanced with attention mechanisms or transformer-based models, to capture complex contextual dependencies. For example, Veríssimo et al. [64] examined the use of sequence-to-sequence networks for automatic translation from Brazilian Portuguese to LIBRAS, demonstrating the potential of neural models to produce more natural and fluid translations.

Flores Brongar et al. [25] proposed a translation architecture that integrates speech-to-text conversion techniques with NLP techniques to generate LIBRAS results from spoken Portuguese, highlighting the importance of NLP in building bridges between modalities. Similarly, Rocha et al. [45] presented a translation tool that uses classic NLP methods, such as morphological analysis and semantic alignment, to convert input text into sign language sequences.

In terms of evaluation, NLP-based LIBRAS translation systems often rely on standard machine translation metrics to measure how closely the system output matches human translations. Among these, the BLEU (Bilingual Evaluation Understudy) score is one of the most common, used to assess the similarity between machine-generated translations and human references, as in Veríssimo et al. [64] and Flores Brongar et al. [25]. Other metrics, such as ROUGE [34], METEOR [7], and classification-based measures [58] such as accuracy and F1 score, are also used in tasks involving predefined phrase matching.

The effectiveness of these evaluation metrics, however, is closely tied to the availability and quality of annotated datasets. One of the recurring challenges in the literature is the lack of publicly available standardized datasets for the translation of LIBRAS into Portuguese. As a result, many studies rely on custom or collaborative corpora. For example, Silva et al. [53] developed a LIBRAS translator based on a collaboratively constructed corpus, emphasizing the need for domain-specific data in this research area.

These studies highlight the essential role of NLP, when combined with CV and ML, in building effective LIBRAS translation systems. Such systems not only address a technical challenge but also represent a step toward greater inclusion and accessibility for the Brazilian deaf community.

3 METHODOLOGY

A Systematic Literature Review (SLR) is a structured and standardized approach to identifying, classifying, and synthesizing studies on a specific topic [38]. To further standardize this research method, we adopted the protocol PRISMA [35] (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which provides guidelines to help plan, conduct, and report systematic literature reviews. Additionally, we employed the Parsifal tool, an online platform designed to support researchers in the planning and execution of SLRs by facilitating protocol definition, search string management, study selection, and data extraction in a collaborative and organized manner [37].

3.1 Research Questions

Throughout this RSL, our aim is to identify studies that investigate artificial intelligence-based approaches — including, but not limited to, machine learning, deep learning, computer vision, and natural language processing — for the translation, recognition, and interpretation of LIBRAS.

Following the PRISMA protocol, we defined the following Research Questions (RQs) for the RSL:

RQ1: What Artificial Intelligence-based solutions are currently employed in published works for the recognition of LIBRAS signs?

RQ2: What are the performance metrics and reported results of existing models in LIBRAS translation tasks?

RQ3: What datasets are used to train and evaluate these models, and what are their main characteristics and limitations?

3.2 Data Sources and Digital Libraries

Six scientific databases were selected as primary research sources for conducting the SLR. The selection was based on the relevance and coverage of these digital libraries in Artificial Intelligence, Machine Learning, Computer Vision, and related areas.

Table 1 presents the selected databases, including both international and national repositories, aiming to ensure diversity and quality in the retrieved publications. The searches were carried out using the specific search strings presented in section 3.1.

It is worth mentioning that, in addition to international sources such as ACM, IEEE, and Springer, national databases such as CAPES Journals and Sol SBC were included, aiming to cover relevant studies in the Brazilian context, particularly those related to LIBRAS.

Table 1: Scientific databases considered for the SLR

Database	URL
ACM	https://dl.acm.org
CAPES	https://periodicos.capes.gov.br
IEEE Xplore	http://ieeexplore.ieee.org
ScienceDirect	http://www.sciencedirect.com
Sol SBC	https://sol.sbc.org.br/busca
Springer Link	http://link.springer.com

3.3 Search Strategy

After defining the RQs and selecting the digital libraries (Table 1), the next step was to formulate keywords and synonyms to build the search string in Parsifal. These keywords were organized to align with the components of each research question, enabling Parsifal to generate systematic and reproducible search strings, thus ensuring traceability and consistency throughout the review process.

- **Brazilian Sign Language**: synonyms include *"LIBRAS"* and *"Língua Brasileira de Sinais"*.
- **Artificial Intelligence**: synonyms include *"Inteligência Artificial"* and related terms, such as *"Machine Learning"* and *"Neural Networks"*.
- **Recognition**: synonyms include *"Reconhecimento"* and related terms, such as *"Translation"* and *"Detection"*.

These terms were combined using Boolean operators (`AND`, `OR`) to construct the final search string. This strategy aimed to retrieve studies focusing on artificial intelligence techniques applied to LIBRAS translation, including model evaluation, dataset analysis, and performance metrics.

The final search string used in Parsifal can be summarized as:

```
("Libras" OR "Brazilian Sign Language" OR
"língua brasileira de sinais")AND("artificial
intelligence" OR "inteligência artificial"
OR "machine learning" OR "aprendizado de
máquina" OR "deep learning" OR "aprendizado
profundo" OR "neural networks" OR "redes
neurais")AND("recognition" OR "reconhecimento"
OR "translation" OR "tradução" OR "detection"
OR "detecção")
```

3.4 Selection Criteria

The selection of studies was based on predefined inclusion and exclusion criteria, as described below.

3.4.1 Inclusion Criteria.

- Articles written in English or Portuguese.
- Complete articles (full text available).
- Primary or secondary studies.
- Studies involving translation from sign language (LIBRAS).

3.4.2 Exclusion Criteria.

- Studies involving other sign languages (Non-LIBRAS).
- Approaches not employing Artificial Intelligence techniques.
- Proceedings, abstracts, or short papers.
- Studies published before 2015.
- Works focused on translation *to* sign language (instead of *from* sign language).
- Studies addressing only alphanumeric sign translation.

3.5 Quality Assessment Checklist

To ensure the methodological rigor of the systematic review, a Quality Assessment Checklist was applied to evaluate each selected study. Five assessment questions were defined to analyze the relevance, transparency, and methodological quality of the studies, as follows:

(1) Does the article describe a solution using Artificial Intelligence, Machine Learning, or Computer Vision applied to LIBRAS signal identification?
(2) Is the dataset used public and accessible?
(3) Does the article describe the dataset used for training and evaluation?
(4) Does the article define and apply quantitative performance metrics for translation or recognition tasks?
(5) Does the article present and discuss results?

Each question was scored according to the following criteria:

- **Yes**: 1.0 point
- **Partially**: 0.5 point
- **No**: 0.0 point

The **maximum possible score** for each study was 5.0 points, and a **cutoff score** of 3.0 points was defined. Only studies scoring equal to or above this threshold were considered as valid for inclusion in the result analysis. This assessment allowed the exclusion of low-quality studies and ensured that only relevant and methodologically sound works contributed to the synthesis of evidence.

3.6 Data Extraction Form

A data extraction form was designed to systematically collect relevant information from each selected study. The extracted data included bibliographic details, methodological characteristics, and evaluation metrics. Table 2 summarizes the fields used in the extraction process.

Table 2: Data extraction fields

Field	Type	Values
Repository	Multi-select field	ACM, CAPES Periódicos, IEEE, Science@Direct, Sol SBC, Springer Link
Year	Integer	Publication year
Task Type	Multi-select field	Signal Generation, Continuous Sign Recognition, Isolated Sign Recognition, Sign-to-Text Translation, Others
Approach	String	Method or technique used
Dataset Used	String	Dataset name
Input Type	Multi-select field	Image, Pose, Audio, Text, Video
F1-score	Float	n/a
Accuracy	Float	n/a
Precision	Float	n/a
Recall	Float	n/a
Inference Time	Float	n/a
BLEU Score	Float	n/a

3.7 Import Studies

After performing the search in the selected databases, the retrieved studies were imported into the Parsifal tool, as shown in Figure 1. Each source represents a database used, along with the number of studies imported from it.

The search strings were adapted according to the specifications of each database. It is important to note that some sources, such as CAPES Journals, did not result in any imported studies due to a lack of relevant results or technical limitations in the extraction process.

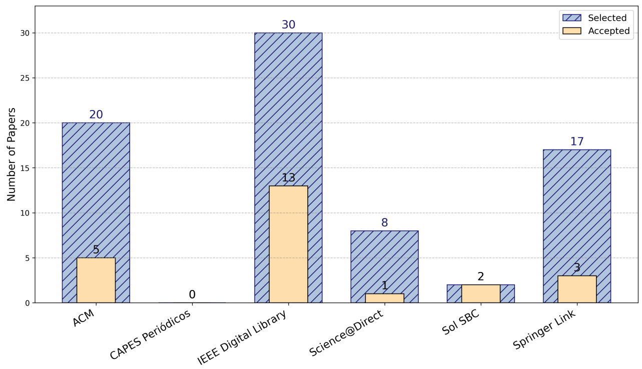

Figure 1: Selected vs Accepted papers by source

Studies imported from different sources proceeded to the selection phase, where those meeting the predefined inclusion criteria were retained.

3.8 Study Selection

During the Study Selection phase, all imported papers were analyzed and classified as *Accepted*, *Rejected*, *Unclassified*, or *Duplicated*, according to predefined inclusion and exclusion criteria. In total, 77 studies were evaluated using the Parsifal tool. This process was carried out manually by analyzing the title and abstract of each paper to determine its inclusion.

Figure 1 presents the number of *Selected* versus *Accepted* papers by source, where each column represents a different digital library or database. It can be observed that most of the selected papers were published in IEEE Xplore and the ACM Digital Library.

Figure 2 presents the PRISMA flow diagram summarizing the study selection process. From the 77 identified records, 12 duplicates were removed and 65 records remained for screening. Based on title and abstract analysis, 41 studies were excluded, leaving 24 for eligibility assessment. One study was excluded for not being related to LIBRAS, resulting in 23 studies included in the final review.

Each study was manually classified by three independent reviewers according to the predefined inclusion and exclusion criteria. Any disagreements were discussed and resolved by consensus among the reviewers. The classification categories were as follows:

- **Accepted:** Studies that met all inclusion criteria and addressed the research objective.
- **Rejected:** Studies that did not meet the inclusion criteria or clearly fell under the exclusion criteria.
- **Unclassified:** Studies pending a decision (none remained unclassified at the end of the process).
- **Duplicated:** Studies identified as duplicates from different sources.

Figure 3 shows the distribution of accepted papers by publication year. While publications were relatively sparse between 2015 and 2022, a noticeable increase occurred in recent years, with 2023 and 2024 accounting for nearly half of all accepted studies. In particular, 2024 alone contributed seven papers, suggesting a growing research interest in the topic and possibly reflecting the recent advances and heightened attention in the field.

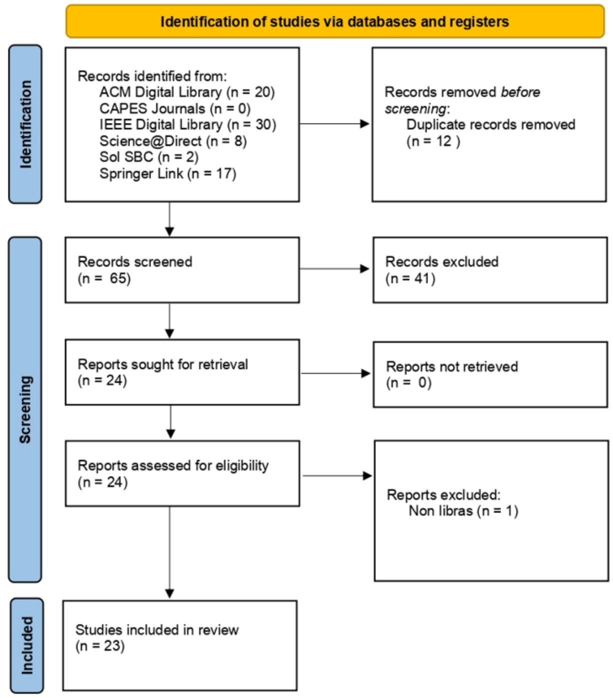

Figure 2: PRISMA flow diagram

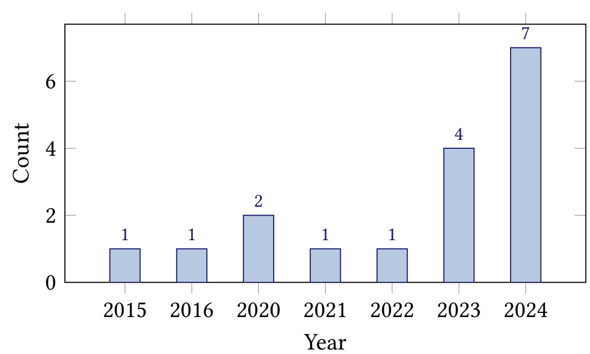

Figure 3: Papers by year

4 LITERATURE REVIEW RESULTS

This section presents the main findings obtained from the analysis of the studies included in the systematic review. The results are organized according to the research questions, focusing on model performance, dataset usage, and computational efficiency.

The following subsections provide a structured summary of the selected studies, presenting the approaches, evaluation practices, and resources employed. Rather than jumping to conclusions, each RQ is addressed individually, allowing for a clearer understanding of the technological strategies adopted, the evaluation methods applied, and the characteristics of the datasets used.

Table 3: Summary of AI-Based Approaches for LIBRAS Sign Recognition (RQ1)

Reference	Main AI Technique	Input Data and Modality
Furtado et al. [26]	CNN (Inception-v3) [36] with Transfer Learning	Markerless visual images of the LIBRAS alphabet.
Sarma et al. [52]	CNNs (UNET [48] for segmentation, VGG16 [57] for classification)	Images for semantic segmentation of hands.
da Silva et al. [16]	Two-stream 3D CNN (I3D)	Raw RGB images and optical flow from sign videos.
Alves et al. [4]	CNN (ResNet-18) [27]	2D skeleton images generated from OpenPose [11] landmarks.
de Avellar Sarmento and Ponti [17]	CNNs for spatial features, combined with LSTMs for temporal tasks	Landmarks extracted with MediaPipe [47] or video data.
Bharti et al. [9]	CNN (Inception-ResNet-v2) for image classification	Key frames extracted from dynamic gesture videos.
De Brito et al. [18]	Long Short-Term Memory (LSTM) [28]	Sequential hand landmark data extracted via MediaPipe [47].
Ruiz et al. [49]	BiLSTM and Transformer [62] for classification	Handcrafted features from MediaPipe [47], YOLOv5 [29], and OpenFace [6].
Bastos et al. [8]	LSTM and Dense layers	Full body detection data from MediaPipe Holistic [47].
Flores Brongar et al. [25]	Transformer (T5) [42] in a Sequence-to-Sequence model	Portuguese text from speech conversion, used for sign generation.
Silva et al. [53]	Transformer [62] and LightConv [66] for machine translation	Data for the incremental training of the VLibras translator.
Fanucchi et al. [24]	Vision Transformer (VideoMAE)	Sign language videos for augmented reality applications.
Silveira et al. [56]	Conditional VAE-GAN [33] (adapted SynLibras)	Pose keypoints ("stick figures") for sign synthesis on mobile devices.
da Silva et al. [15]	Multi-stream architecture with an MLP for fusion	Multiple streams: RGB image, body pose, and hand segmentation.
Silveira et al. [55]	Conditional VAE-GAN [33] (SynLibras model)	Pose data to synthesize (generate) the signer and gesture.
Abreu and Dias [2], Abreu et al. [3]	Classic Machine Learning (KNN, SVM)	EMG (Electromyography) signals and other data from wearable sensors.

4.1 AI-Based Solutions for LIBRAS Sign Recognition (RQ1)

The recognition and interpretation of LIBRAS is significantly shaped by the application of AI, ML, DL and CV. These technologies are pivotal in developing systems that translate sign language gestures into text or spoken language, thereby enhancing communication accessibility for the deaf and hard-of-hearing community. The studies reviewed employ a diverse range of AI-based solutions, from sophisticated neural network architectures to specialized feature extraction and multi-modal integration strategies. Table 3 a summary of these methodologies, detailing the primary AI techniques and input modalities used in the referenced works.

Deep Learning Architectures for Sign Recognition and Translation. A prominent trend in LIBRAS research involves the use of various deep neural network architectures, primarily for isolated and continuous sign recognition, as well as for sign generation and translation:

- **Convolutional Neural Networks (CNNs):** CNNs are foundational for processing visual data in LIBRAS.
  - **Furtado et al. [26]** utilized the Inception-v3 model [36], a CNN, with transfer learning for markerless visual recognition of the LIBRAS alphabet.
  - **Sarma et al. [52]** employed the UNET architecture [48] for semantic segmentation of hands, followed by a pre-trained VGG16 model [57] for gesture classification.
  - For isolated sign recognition in a health context, **da Silva et al. [16]** proposed a two-stream architecture based on 3D Convolutional Neural Networks (specifically I3D) [13], processing raw RGB images and optical flow.
  - **Alves et al. [4]** used a ResNet-18 [27] CNN to classify 2D skeleton images derived from OpenPose [11] landmarks for isolated sign recognition.
  - **de Avellar Sarmento and Ponti [17]** investigated CNNs (MobileNetV2 [50], InceptionResNetV2 [60], ResNet50V2 [27]) for spatial feature extraction, combined with LSTMs for temporal classification in a cross-dataset study.
  - **Bharti et al. [9]** used CNNs like Inception-ResNet-v2 [60] for image classification within a novel key frame extraction technique for dynamic hand gesture recognition.
  - **Sarma et al. [52]** also proposed a system using CNNs combined with attention modules for semantic hand segmentation and gesture recognition, adapting VGG16 [57] for classification.
- **Recurrent Neural Network (RNNs) and Long Short-Term Memory (LSTM):** These networks are crucial for handling the temporal dynamics inherent in sign language.
  - **De Brito et al. [18]** used an LSTM model, a type of RNN, to process sequential hand landmark data extracted via MediaPipe [47] for isolated sign recognition.
  - **de Avellar Sarmento and Ponti [17]** combined LSTMs with CNNs or MediaPipe landmarks for temporal classification of isolated signs.

- **Ruiz et al. [49]** utilized a Bidirectional Long Short-Term Memory (BiLSTM) network for word-level sign language recognition based on handcrafted features.
- **Bastos et al. [8]** developed a system using LSTM and Dense layers for continuous and isolated sign recognition, leveraging MediaPipe Holistic [47] for body detection.

- **Transformers and Sequence-to-Sequence Models:** These advanced architectures are increasingly used for their ability to capture long-range dependencies and complex contextual information, particularly in translation tasks.
  - **Flores Brongar et al. [25]** employed a NLP model called Sequence-to-Sequence (Seq2Seq) based on the Transformer architecture for Portuguese-to-LIBRAS sign generation, integrating speech-to-text conversion.
  - **Ruiz et al. [49]** also used a Transformer model [62] for classification in their word-level sign language recognition approach.
  - **Silva et al. [53]** proposed a tool called "Automation" that manages incremental training for the VLibras [65] translator, utilizing Transformer and LightConv neural networks with the Fairseq framework.
  - **Fanucchi et al. [24]** fine-tuned a Video Masked Autoencoder (VideoMAE) [61], an architecture based on Vision Transformers (ViT), for LIBRAS sign recognition aimed at augmented reality applications.

Feature Extraction and Other Input Modalities. Beyond raw video input, studies employ various techniques to extract relevant features and utilize different input modalities:

- **Landmark-Based and Skeleton-Based Features:**
  - **De Brito et al. [18]** and **de Avellar Sarmento and Ponti [17]** used MediaPipe [47] to extract key points (landmarks) of hands, body, and pose, which then serve as features for RNN/LSTM models.
  - **Ruiz et al. [49]** extracted handcrafted features using MediaPipe [47], YOLOv5 [29] (for hand detection), and OpenFace [6] (for head pose and gaze direction), followed by Principal Component Analysis (PCA) for dimensionality reduction.
  - **Alves et al. [4]** converted OpenPose [11] landmarks into 2D skeleton images using the Skeleton-DML algorithm, which were then classified by a CNN.
  - **Silveira et al. [56]** adapted the SynLibras model for mobile devices by changing pose representation from dense heatmaps to "stick figures" derived from Google MediaPipe key points.
- **Multi-Stream Architectures:** These approaches combine different input streams or feature types for a more comprehensive understanding.
  - **da Silva et al. [16]** used a two-stream 3D CNN, processing raw RGB images and optical flow.
  - **da Silva et al. [15]** proposed a multiple stream architecture where each stream processes a distinct input (RGB image, body pose, and semantic hand segmentation), with outputs combined by a Multilayer Perceptron (MLP) for final classification.
- **Generative Models for Sign Synthesis:**
  - **Silveira et al. [55]** introduced SynLibras, a Conditional Variational Autoencoder - Generative Adversarial Network (CVAE-GAN) model [33], designed to disentangle the interpreter's appearance from their gestural pose for LIBRAS synthesis.
  - **Silveira et al. [56]** further adapted the SynLibras CVAE-GAN for generating digital LIBRAS signers on mobile devices.

Non-Vision Based Approaches. While most recent advancements rely on computer vision, some studies explore alternative input modalities:

- **Sensor-Based Technologies:**
  - **Abreu and Dias [2]** and **Abreu et al. [3]** explored the use of EMG signals captured by wearable devices like the Myo Armband, along with accelerometer, gyroscope, and magnetometer data. These signals were used to extract statistical features, which were then classified by algorithms such as K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) for dactylology (fingerspelling) recognition.

The integration of these diverse AI, ML, DL, and CV techniques, often within end-to-end pipelines, is crucial for building effective LIBRAS translation systems that address both technical challenges and the broader societal need for inclusion.

4.2 Reported Performance and Evaluation Metrics (RQ2)

Evaluating the performance of AI-based systems for Brazilian Sign Language (LIBRAS) recognition and translation is crucial to ensuring their reliability and practical applicability. As one can note in Table 4 the selected studies employ a diverse set of evaluation metrics, reflecting the multimodal and multi-step nature of the approaches.

For recognition tasks, common classification metrics such as **accuracy**, **precision**, **recall**, and **F1-score** are widely adopted. These metrics provide insights into the correctness and robustness of models, especially when dealing with imbalanced datasets or multi-class scenarios. Yet, in translation tasks, metrics from the machine translation domain are employed, primarily the **BLEU** (Bilingual Evaluation Understudy) score, which assesses the overlap between machine-generated results and reference human translations. Some studies also report **inference time**, an important aspect for real-time applications.

Accuracy is the most frequently reported metric among the analyzed works. It reflects the proportion of correct predictions and is particularly useful in balanced classification tasks. For example, in a balanced dataset scenario, da Silva et al. [15] reports 90.0% accuracy using a multi-stream architecture for sign recognition in healthcare settings. Similarly, Furtado et al. [26] achieved 97.0% accuracy with a CNN-based model (Inception-v3) using transfer learning for alphabet recognition in LIBRAS, also with a balanced dataset. In contrast, performance tends to be lower in more challenging evaluation contexts, such as cross-dataset testing. As demonstrated by de Avellar Sarmento and Ponti [17], a model combining CNNs and LSTMs achieved 66.7% accuracy when evaluated on different datasets.

The F1-score, representing the harmonic mean between precision and recall, is widely used to evaluate classification models. It is particularly relevant in scenarios with imbalanced datasets, as it balances the trade-off between false positives and false negatives, providing a more representative measure of model performance than precision alone. However, when the dataset is balanced, the F1-score often produces values very close to precision, since precision and recall tend to be similar. One can observe this behavior in da Silva et al. [16], where a two-stream 3D CNN model for isolated signal recognition achieved an F1-score of 0.95, closely matching the reported accuracy. A similar pattern is found in Alves et al. [4], where the authors obtained an F1-score of 0.93 and precision and recall values of 0.94 and 0.93, respectively, in a balanced dataset.

Translation tasks adopt specific evaluation criteria. The BLEU score is the most common metric in this domain. For example, Flores Brongar et al. [25] employed a Transformer-based architecture (T5) and obtained a BLEU score of 75.3 and 0.91% inference time in a system that integrates speech-to-text conversion with LIBRAS generation. This demonstrates the potential of NLP-based models to handle the complexity and semantic variability of sign language generation.

In contrast to these higher results, studies that evaluated models on multiple datasets or used publicly available corpora tend to present lower or more variable results, reflecting the challenges of generalization. de Avellar Sarmento and Ponti [17], for example, tested various combinations of CNN and LSTM in a cross-dataset scenario and reported an accuracy of only 66.7%, demonstrating the domain shift between training and evaluation conditions. Similarly, Alves et al. [4], who used the Libras and SynLibras corpus datasets in different evaluation contexts, reported an accuracy of 93.0% and an F1-score of 0.93, highlighting the effectiveness of skeleton-based features but also demonstrating the sensitivity of performance to the evaluation set.

Conversely, studies that evaluated models on multiple datasets or used publicly available corpora tend to present lower or more variable results, reflecting the challenges of generalization. de Avellar Sarmento and Ponti [17], for example, tested various combinations of CNN and LSTM in a cross-dataset scenario and reported an accuracy of only 66.7%, demonstrating the domain shift between training and evaluation conditions. Similarly, Alves et al. [4], who used the Libras and SynLibras corpus datasets in different evaluation contexts, reported an accuracy of 93.0% and an F1-score of 0.93, highlighting the effectiveness of skeleton-based features but also demonstrating the sensitivity of performance to the evaluation set.

Precision and recall are reported less consistently across studies. When available, these metrics help clarify the balance between correctly identifying signals and avoiding false positives. For example, Alves et al. [4] reported precision (0.94) and recall (0.93), demonstrating a well-balanced performance of their classifier. However, many studies omit these values, focusing instead on aggregate metrics such as accuracy or F1 scores, which can obscure performance on imbalanced datasets or multi-class tasks.

A smaller subset of studies report inference time, which is relevant for real-time or embedded applications. Alves et al. [4] mentions an average inference time of 4.58 seconds per sample, which may limit deployment in mobile or low-latency environments. In contrast, Silveira et al. [56] explored mobile deployment strategies by modifying pose representation methods in his generative model in LIBRAS, and get a value of 2.8 seconds.

Despite the adoption of standard evaluation metrics, there is no consensus on benchmarking practices in the field. Different datasets, task types (e.g., isolated sign recognition vs. continuous gesture translation), and evaluation conditions (e.g., independent sign splits, real-world scenarios) make direct comparison of reported results difficult. For example, while Flores Brongar et al. [25] obtained a BLEU score of 75.3% on a translation task, Silva et al. [53] reported a lower BLEU score of 39.26 using the VLibras corpus and a collaborative training strategy. These variations highlight the importance of contextualizing performance metrics within the experimental setup of each study.

Another recurring methodological challenge is the lack of standardized evaluation protocols across studies. As seen in Table 4, authors employ different datasets, model architectures, and even performance metrics, often without specifying important details such as data splits, signer variability, or separation between training and validation. This heterogeneity limits the comparability and reproducibility of results, even when common metrics such as accuracy or BLEU are used.

For example, although several studies report high accuracy values, such as 99.87% in Abreu et al. [3] or 98.97% in Sarma et al. [51], the lack of detailed methodological descriptions, such as the number of signers, signal classes, or test scenarios, makes it difficult to assess whether these results result from model generalization or overfitting to restricted datasets. On the other hand, studies like Ruiz et al. [49], which employ handcrafted resources and test their models on public datasets like the Libras Corpus, provide clearer experimental protocols, contributing to transparency in evaluation.

Although widely used in translation tasks, using BLEU as a primary metric has its limitations. As demonstrated by Flores Brongar et al. [25] and Silva et al. [53], BLEU does not capture semantic nuances, signer variability, or syntactic differences between LIBRAS and Portuguese. Furthermore, the use of custom parallel corpora, as in Flores Brongar et al. [25], introduces additional uncertainty regarding the consistency and linguistic diversity of the references.

Finally, a critical but underreported aspect is the consideration of real-world deployment constraints. Few studies evaluate model performance in signer-independent scenarios or natural environments with varying lighting, camera angles, or backgrounds. Although Fanucchi et al. [24] takes a step toward realistic implementation by integrating their model into an augmented reality system, they do not provide detailed performance metrics beyond classification accuracy.

Although a wide range of evaluation metrics are used in LIBRAS recognition and translation studies, the diversity of experimental settings, the lack of standardized references, and the limited reporting of methodological details represent barriers to replicability and generalization. Future research would benefit from shared evaluation protocols, open datasets with diverse signatories and contexts, and the broader adoption of transparent reporting practices.

Table 4: Literature review results summary (RQ2)

Reference	Year	Dataset	F1-score	Accuracy	Precision	Recall	Inference Time (seconds)	BLEU
Abreu et al. [3]	2016	Own Dataset	-	0.99	-	-	-	-
Abreu and Dias [2]	2018	Own Dataset	-	-	-	-	-	-
da Silva et al. [16]	2020	Own Dataset	0.95	0.96	0.95	0.96	-	-
da Silva et al. [15]	2023	Own Dataset	-	0.90	-	-	-	-
Furtado et al. [26]	2023	Own Dataset	0.97	0.97	0.97	0.97	-	-
Ruiz et al. [49]	2023	Corpus Libras [20]	-	0.94	-	-	-	-
de Avellar Sarmento and Ponti [17]	2023	Cross-Dataset	-	0.66	-	-	-	-
Bharti et al. [9]	2023	Cross-Dataset	-	0.96	-	-	-	-
Flores Brongar et al. [25]	2024	Own Dataset	-	-	-	-	0.91	75.3
De Brito et al. [18]	2024	Own Dataset	-	0.85	-	-	-	-
Alves et al. [4]	2024	Cross-Dataset	0.93	0.93	0.94	0.93	4.58	-
Silva et al. [53]	2021	Corpus VLibras [46]	-	-	-	-	-	39.26
Bastos et al. [8]	2015	BrSL [8]	-	0.90	-	-	-	-
Sarma et al. [51]	2021	BrSL [8]	-	0.98	-	-	-	-
Sarma et al. [52]	2024	BrSL [8]	-	0.98	-	-	-	-
Silveira et al. [55]	2022	SynLibras [54]	-	-	-	-	-	-
Silveira et al. [56]	2023a	SynLibras [54]	-	-	-	-	2.80	-
Veríssimo and Costa [63]	2020	Emo-PT [23]	0.94	0.94	-	-	-	-
Fanucchi et al. [24]	2024	MINDS-Libras [44]	-	0.84	-	-	-	-

4.3 Datasets: Characteristics and Limitations (RQ3)

Most Brazilian Sign Language recognition and translation research relies on specific datasets for training and evaluating models. These datasets vary significantly in scope and complexity: some are limited to fingerspelling (manual alphabet), others focus solely on hand gestures, while more comprehensive datasets incorporate both hand and full-body motion, including facial expressions. This diversity reflects the differing goals of each study—ranging from isolated gesture classification to robust multimodal translation approaches. Additionally, the datasets differ in collection methodology, annotation quality, number of classes, and public availability.

Based on the findings of the systematic review, the following datasets emerged as the most relevant in the literature, given their frequent use, reported performance, and alignment with the research objectives of the analyzed works:

V-LIBRAS [46] is a parallel corpus comprising 1,364 Portuguese expressions translated into Libras, totaling 4,089 signs. Up to three different signers recorded the signs using a *chroma key* background, allowing for scene replacement. The corpus is publicly available through a specific online portal and is widely used in automatic translation systems.

MINDS-Libras [44] is a multimodal dataset developed at UFMG, containing 20 signs recorded five times by 12 signers (approximately 1,200 samples). Recordings include RGB video, depth, and body and facial landmarks extraction. It is used for dynamic sign recognition and allows robust analysis of inter-signer variability.

SynLibras-Pose [54] consists of 1,133 videos of Libras words with automatic (and editable) annotations of 30 keypoints: 12 from the body, 10 from the hands, and 8 from the face. The data were extracted from the Libras–Portuguese dictionary developed by UFV [22]. Its primary focus is pose synthesis and transfer between signers, practical for computer vision-based methods and image generation.

Corpus Libras (UFSC) [20] is an open linguistic corpus with spontaneous and elicited Libras videos. It is intended for linguistic studies, documentation, and sociolinguistic analysis. It has broad representativeness, although data collection and annotation protocols vary.

RBMT, **NMT**, and **VLibrasBD** [65] refer to translation approaches and corpora used in automatic Portuguese–Libras translation. VLibrasBD is a parallel corpus that feeds the VLibras system, supporting rule-based (RBMT) and neural machine translation (NMT). Access may be restricted or subject to institutional agreements [21].

SignBank (Brazil) [14] is a lexical database inspired by the international SignBank model, recording Libras signs linked to glosses, videos, and linguistic features such as handshape and location. It is widely used in linguistic description studies, teaching, and Libras lexicography research.

Corpus Libras-Movement (UCI) [43] is a classic dataset with 15 classes of manual movements, each with 24 instances. Data were extracted from videos and converted into two-dimensional time series (x, y) with 45 samples per instance, which were used in classification and clustering tasks.

Table 5 presents a comparative summary of the main datasets discussed in this section, including the number of classes, annotation type, collection method, supported modalities, and access type.

5 DISCUSSION

The comprehensive technical analysis of the selected studies reveals a progressive evolution in the application of Artificial Intelligence-based approaches for LIBRAS recognition and translation. A firm reliance on CNN-based architectures is observed across most works, particularly for processing visual information such as hand gestures and body movement. These architectures are often paired with transfer learning techniques to improve generalization, even in limited-data scenarios. Recurrent models, such as LSTMs and BiLSTMs, are widely employed to model temporal dependencies. In contrast, Transformers and Sequence-to-Sequence models have recently gained prominence due to their superior performance in capturing contextual and sequential nuances, especially in translation tasks.

Table 5: Comparative summary of major Brazilian Sign Language datasets (RQ3)

Dataset	Number of Classes	Number of Samples	Annotation Type	Collection Method	Modalities	Access
V-LIBRAS	1,364 signs	4,089 videos	Manual (video + gloss)	Studio, chroma key, 3 signers	Hands only	Public [46]
MINDS-Libras	20 signs	~1,200 videos	Landmarks (RGB + depth)	RGB + depth + pose, 12 signers	Hands, face, body	Request [44]
SynLibras-Pose	Not specified	1,133 videos	Editable keypoints (30 pts)	From dictionary videos	Pose: hands, face, body	Public [54]
Corpus UFSC	Not specified	Not specified	Linguistic glosses	Spontaneous + elicited	Natural variation	Public [20]
VLibras-BD	Not specified	Not specified	Parallel text–video	Text–sign alignment	Manual signs	Restricted [65]
SignBank	Not specified	Lexical entries	Glosses + metadata	Linguistic registry	Manual + phonology	Public [14]
Libras Movement (UCI)	15 classes	24 instances	2D time series (x,y)	From video to time series	Hand traj. only	Public [43]

Multimodal strategies are increasingly explored, integrating hand motion, facial expressions, and body pose data to enhance recognition robustness. Frameworks like MediaPipe, OpenPose, and YOLOv5 are frequently used for landmark and skeleton extraction, providing structured inputs to CNNs and RNNs. Multi-stream architectures also appear as a key design pattern, enabling the combination of different input types (e.g., RGB images, pose estimation, optical flow) for comprehensive feature fusion. A few studies have investigated generative models (e.g., CVAE-GANs [33]) for sign synthesis and avatar animation, particularly targeting mobile and augmented reality applications.

Despite these technological advances, specific patterns of divergence are apparent. While some studies prioritize isolated sign recognition, others focus on continuous translation, signer-independent learning, or real-time interaction. Annotation strategies differ considerably, with some authors developing handcrafted datasets while others leverage public corpora like V-LIBRAS, MINDS-Libras, and SynLibras-Pose [54]. Evaluation protocols also vary, ranging from accuracy and F1-score to confusion matrices and qualitative assessments, making direct comparisons between studies difficult.

In terms of performance, recent studies (2022–2023) reported higher accuracy rates, often exceeding 90%—compared to earlier efforts. This improvement is likely attributed to the increasing availability of annotated datasets and the widespread adoption of transfer learning. However, performance remains sensitive to task complexity (e.g., fingerspelling vs. full-sentence translation), data quality, and the modalities considered.

The field also exhibits fragmentation, including facial cues and non-manual components. Although multimodal integration is recognized as beneficial, many models still limit themselves to manual signs, overlooking crucial grammatical elements conveyed through facial expressions and head movement. This inconsistency hampers both the generalizability and completeness of the proposed systems.

In light of these findings, this review highlights the significant progress and methodological heterogeneity in this research domain. The diversity of approaches demonstrates the field's richness and underscores the need for standardization. Shared datasets, unified evaluation benchmarks, and typical annotation schemes facilitate reproducibility and enable fairer comparison across studies. Addressing these challenges is essential for advancing LIBRAS recognition systems toward real-world deployment and broader social impact.

6 OPEN ISSUES

Despite recent advances driven by deep learning architectures and multimodal strategies, several open issues remain in LIBRAS recognition and translation. One recurring challenge is the limited integration of non-manual features such as facial expressions, gaze direction, and head movements—elements that convey essential grammatical and emotional information in LIBRAS. Although tools exist for extracting such signals, their application in recognition and translation models is still underexplored, slowing progress toward more natural and linguistically accurate systems.

Another important aspect is the need to expand the development of large-scale, balanced, and multimodal datasets. Existing initiatives such as V-LIBRAS [46], MINDS-Libras [44], and SynLibras-Pose [54] represent progress but do not fully encompass the diversity of classes, signer representation, and modality coverage required for broader applications. This scenario challenges the creation of models capable of handling diverse usage contexts, especially in environments demanding robustness to linguistic and visual variations. Moreover, continuous sign recognition, contextual discourse processing, and interpreting complex interactions remain largely unexplored areas.

The absence of methodological standardization also persists as a challenge. Heterogeneous annotation schemes and the lack of unified evaluation protocols hinder result comparability and the consolidation of advances. Metrics such as accuracy, F1-score, and BLEU are used inconsistently across studies, often without consideration for the linguistic characteristics of LIBRAS. In particular, despite being widely adopted in machine translation, BLEU is ill-suited for capturing the syntactic and semantic differences between LIBRAS and Portuguese.

Aspects related to temporal and signer variability also receive limited attention. Longitudinal and signer-independent protocols are crucial for assessing model consistency in real-world scenarios but remain rare. Issues related to system deployment—such as inference time, computational cost, and performance under varying environmental conditions (e.g., lighting, camera positioning)—also require further investigation to enable effective use in mobile and embedded applications.

Sensor-based approaches (e.g., EMG, accelerometers) show potential to complement computer vision systems, particularly in constrained environments, yet remain underexplored. Generative models for sign synthesis, with animation and training data augmentation applications, are emerging but require further development to achieve greater realism, flexibility, and user acceptance.

To advance in these areas, future research should:

- Develop large-scale, multimodal, and publicly accessible LIBRAS datasets;
- Establish standardized annotation and evaluation frameworks, including task-specific and signer-aware protocols;
- Explore advanced architectures (e.g., Transformers, VideoMAE) in signer-independent and real-time contexts;
- Adopt participatory design methodologies with active involvement of the deaf community;
- Investigate low-resource learning techniques and multimodal integration strategies that go beyond vision-only input.

Addressing these open issues will be decisive for transforming experimental advances into concrete applications capable of promoting more accessible and inclusive communication for the Brazilian deaf community.

7 CONCLUSION

This systematic review of the literature examined the use of AI-based approaches in the translation and recognition of Brazilian Sign Language. The review identified a strong prevalence of CNN-based architectures for gesture classification, with growing integration of recurrent networks (LSTMs), Transformer-based models, and multi-stream or multimodal approaches.

In all reviewed studies, model performance varied depending on the complexity of the dataset, the coverage of modality, and the evaluation context. Tasks such as fingerspelling with self-collected data achieved accuracy above 95%, while cross-dataset and signer-independent scenarios often reported lower results. A wide range of evaluation metrics—including accuracy, F1-score, BLEU, and inference time—were used, though benchmarking protocols remain inconsistent and often underreported.

The analysis also revealed a gradual shift toward the use of public datasets such as Corpus Libras [20], VLibras [46], SynLibras-Pose [54], and MINDS-Libras [44], although many models still rely on private or limited datasets. The lack of standardized corpora and annotations, along with the limited modeling of non-manual signals such as facial expressions and body pose, remains a significant barrier to generalization and real-world deployment.

This review contributes by mapping the state of the art, identifying methodological limitations, and highlighting future research directions—particularly in the areas of dataset quality, evaluation standardization, and inclusive system design. Moving forward, interdisciplinary collaboration with the deaf community will be essential to ensure that AI-based LIBRAS solutions are accessible, context-aware, and linguistically faithful.

8 LIMITATIONS, WEAKNESSES AND THREATS TO VALIDITY

This systematic review was limited in scope to the literature on LIBRAS, using a search strategy focused explicitly on this sign language. Although the data extraction process was conducted carefully, some relevant studies may have been omitted due to search criteria and database limitations.

The available dataset descriptions were predominantly qualitative, lacking detailed quantitative metrics that could deepen comparative analysis. Additionally, the absence of complete information on data collection and annotation methods limited the assessment of potential biases. Finally, as with any literature review, publication bias remains a concern, since studies reporting less favorable results may be underrepresented.

Disagreements among the authors regarding study inclusion and exclusion were resolved through discussion, ensuring consensus. The experience of one author with systematic reviews was instrumental throughout the process. Study categories were defined based on the literature available in the selected works.

ACKNOWLEDGMENTS

The authors of this work would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001, CNPq, and FAPERGS for funding this research.

REFERENCES

[1] Bashaer Abdullah, Ghada Amoudi, and Hanan Alghamdi. 2024. Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects. *IEEE Access* PP (01 2024), 1–1. https://doi.org/10.1109/ACCESS.2024.3457692

[2] Filipe Abreu and Jorge Dias. 2018. A Dataset for Electromyography-Based Dactylology Recognition. In *Proceedings of the 9th International Conference on Advances in Computer-Human Interactions (ACHI 2016)*. IARIA, Lisbon, Portugal, 325–330. https://www.thinkmind.org/download.php?articleid=achi_2016_12_30_10079

[3] João Gabriel Abreu, João Marcelo Teixeira, Lucas Silva Figueiredo, and Veronica Teichrieb. 2016. Evaluating Sign Language Recognition Using the Myo Armband. (2016).

[4] Carlos Eduardo G. R. Alves, Francisco De A. Boldt, and Thiago M. Paixão. 2024. Enhancing Brazilian Sign Language Recognition Through Skeleton Image Representation. (2024).

[5] Tejaswini Ananthanarayana, Priyanshu Srivastava, Akash Chintha, Akhil Santha, Brian Landy, Joseph Panaro, Andre Webster, Nikunj Kotecha, Shagan Sah, Thomastine Sarchet, Raymond Ptucha, and Ifeoma Nwogu. 2021. Deep Learning Methods for Sign Language Translation. *ACM Trans. Access. Comput.* 14, 4, Article 22 (Oct. 2021), 30 pages. https://doi.org/10.1145/3477498

[6] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 59–66. https://doi.org/10.1109/FG.2018.00019

[7] Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. 65–72.

[8] Igor L.O. Bastos, Michele F. Angelo, and Angelo C. Loula. 2015. Recognition of Static Gestures Applied to Brazilian Sign Language (Libras). (2015).

[9] S. Bharti, A. Balmik, and A. Nandy. 2023. Novel error correction-based key frame extraction technique for dynamic hand gesture recognition. *Neural Computing and Applications* (2023).

[10] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 16–31. https://doi.org/10.1145/3308561.3353774

[11] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields.

IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 1 (2021), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257
[12] Fabio Cappabianco, Igor Ribeiro Ferreira de Matos, and Luiz Eduardo Galvão Martins. 2024. A Systematic Literature Review on Computer Vision Applied to Brazilian Sign Language. *Luiz Eduardo Galvão, A Systematic Literature Review on Computer Vision Applied to Brazilian Sign Language* (2024).
[13] João Carreira and Andrew Zisserman. 2017. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4724–4733. https://doi.org/10.1109/CVPR.2017.502
[14] Omdena São Paulo Chapter. 2023. Brazilian Sign Language Datasets: UFV, INES, SignBank, V-Libras. https://github.com/OmdenaAI/SaoPauloBrazilChapter_BrazilianSignLanguage. Acesso em: jul. 2025.
[15] D.R.B. da Silva, T.M.U. de Araújo, T.G. do Rêgo, L.C.S. Florêncio, A.F. Neto, and A.C. de Paiva. 2023. A multiple stream architecture for the recognition of signs in Brazilian sign language in the context of health. *Multimedia Tools and Applications* (2023).
[16] Diego R. B. da Silva, Tiago Maritan U. Araujo, Thais Gaudencio do Rêgo, and Manuella Aschoff Cavalcanti Brandão. 2020. A Two-Stream Model Based on 3D Convolutional Neural Networks for the Recognition of Brazilian Sign Language in the Health Context. (2020).
[17] Amanda Hellen de Avellar Sarmento and Moacir Antonelli Ponti. 2023. A Cross-Dataset Study on the Brazilian Sign Language Translation. (2023).
[18] Aron Caiuá V. De Brito, Ana Patrícia F. M. Mascarenhas, Marco A.C. Simões, Ivanoé J. Rodowanski, Jorge Alberto P. De Campos, Josemar Rodrigues De Souza, and Jose Grimaldo Da Silva Filho. 2024. RoboSign: Improving Interaction Between Service-Robot and Hearing Impairments People. (2024).
[19] Ronice Müller de Quadros. 2006. *Gramática da Língua de Sinais Brasileira*. Editora Bookman, Porto Alegre, Brasil. Obra de referência sobre a estrutura linguística da LIBRAS.
[20] Universidade Federal de Santa Catarina (UFSC). 2020. Corpus Libras UFSC. https://corpuslibras.ufsc.br. Acesso em: jul. 2025.
[21] Giely M. C. de Souza. 2019. *Tradução Automática Português-Libras: Uma abordagem híbrida baseada em RBMT e NMT*. Ph. D. Dissertation. Universidade Federal de Santa Catarina. https://repositorio.ufsc.br/xmlui/handle/123456789/187888
[22] Universidade Federal de Viçosa. 2020. Dicionário Online Bilíngue Libras/Português. https://sistemas.cead.ufv.br/dicionariolibras. Acesso em: 11 ago. 2025.
[23] Mariza Miola Dosciatti, Lohann Paterno Coutinho Ferreira, and Emerson Cabrera Paraiso. 2013. Identificando Emoções em Textos em Português do Brasil usando Máquina de Vetores de Suporte em Solução Multiclasse. In *Congresso Nacional de Iniciação Científica - CONIC*. Brasil. https://www.researchgate.net/publication/277813389 Disponível em: https://www.researchgate.net/publication/277813389.
[24] Rodrigo Zempulski Fanucchi, Arlindo Rodrigues Galvão Junior, Gabriel da Mata Marques, Lucas Brandão Rodrigues, Anderson da Silva Soares, and Telma Woerle Lima Soares. 2024. Fine-Tuning a Video Masked Autoencoder to Develop an Augmented Reality Application for Brazilian Sign Language Interpretation. (2024).
[25] Andrew Flores Brongar, Jordan Zitzke Pinho, Rosa Vicari, Gisele Moraes Simas, and Regina Barwaldt. 2024. Improving Translation from Portuguese to Brazilian Sign Language with Speech-to-Text Integration Through Natural Language Processing. (2024).
[26] Silas Luiz Furtado, Jauvane C. De Oliveira, and Shervin Shirmohammadi. 2023. Interactive and Markerless Visual Recognition of Brazilian Sign Language Alphabet. (2023).
[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. https://doi.org/10.1109/CVPR.2016.90
[28] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[29] Glenn Jocher. 2020. *YOLOv5 by Ultralytics*. https://doi.org/10.5281/zenodo.3908559
[30] Daniel Jurafsky and James H. Martin. 2019. *Speech and Language Processing* (3 ed.). Pearson. Draft available at https://web.stanford.edu/ jurafsky/slp3/.
[31] H Pallab Jyoti Dutta, Debajit Sarma, M.K. Bhuyan, and R. H. Laskar. 2020. Semantic Segmentation based Hand Gesture Recognition using Deep Neural Networks. (2020).
[32] Navroz Kaur Kahlon and Williamjeet Singh. 2023. Machine translation from text to sign language: a systematic review. *Universal Access in the Information Society* 22, 1 (2023), 1–35.
[33] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. 2016. Autoencoding beyond pixels using a learned similarity metric. In *International conference on machine learning*. PMLR, 1558–1566.
[34] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*. Association for Computational Linguistics.
[35] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, and Vivian A. Welch. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372 (2021), n71. https://doi.org/10.1136/bmj.n71
[36] Papers with Code. 2025. *Inception-v3*. https://paperswithcode.com/method/inception-v3 Licença: CC-BY-SA.
[37] Parsifal. n.d.. About Parsifal: Perform Systematic Literature Reviews. https://parsif.al/about/. Accessed: 2025-08-11.
[38] Michele D. Peters, Christina M. Godfrey, Deborah McInerney, Patricia Soares, Hanan Khalil, and Beverley J. Parker. 2015. The Joanna Briggs Institute Reviewers' Manual: Methodology for JBI Scoping Reviews. *The Joanna Briggs Institute* (2015). Available online: https://nursing.lsuhsc.edu/JBI/docs/ReviewersManuals/Scoping-.pdf.
[39] Presidência da República do Brasil. 2002. Lei nº 10.436, de 24 de abril de 2002. http://www.planalto.gov.br/ccivil_03/leis/2002/l10436.htm. Recognizes the Brazilian Sign Language (Libras) as a legal means of communication and expression..
[40] Presidência da República do Brasil. 2005. Decreto nº 5.626, de 22 de dezembro de 2005. http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Decreto/D5626.htm. Regulates Lei nº 10.436/2002 and mandates the provision of accessibility for deaf people in education, media, and public services..
[41] Soraia Prietch, J. Alfredo Sánchez, and Josefina Guerrero. 2022. A Systematic Review of User Studies as a Basis for the Design of Systems for Automatic Sign Language Processing. *ACM Trans. Access. Comput.* 15, 4, Article 36 (Nov. 2022), 33 pages. https://doi.org/10.1145/3563395
[42] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 1, Article 140 (Jan. 2020), 67 pages.
[43] UCI Machine Learning Repository. 2008. Libras Movement Dataset. https://archive.ics.uci.edu/ml/datasets/Libras+Movement. 15 classes de movimentos manuais. Acesso em: jul. 2025.
[44] Tamires et al. Rezende. 2022. MINDS-Libras: Um Conjunto Multimodal para Sinais em Libras. https://gitlab.com/tamiresrezende/minds-libras-database-analise-descritiva. Acesso mediante solicitação. UFMG. Acesso em: jul. 2025.
[45] Jampierre Rocha, Jeniffer Lensk, Taís Ferreira, and Marcelo Ferreira. 2020. Towards a Tool to Translate Brazilian Sign Language (Libras) to Brazilian Portuguese and Improve Communication with Deaf. (2020).
[46] Ailton José Rodrigues and Emmanuel P. Barbosa. 2020. Corpus paralelo Português-Libras para tradução automática baseado em avatares 3D. https://libras.cin.ufpe.br. Acesso em: jul. 2025.
[47] Karl Rohr. 2001. *Landmark-based image analysis: using geometric and intensity models*. Vol. 21. Springer Science & Business Media.
[48] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
[49] Daniel Sánchez Ruiz, J. Arturo Olvera-López, and Ivan Olmos-Pineda. 2023. Word Level Sign Language Recognition via Handcrafted Features. *IEEE Latin America Transactions* (2023).
[50] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
[51] Debajit Sarma, Hridoy Jyoti Prasad Dutta, Krishna Singh Yadav, Sahin Ahmed, and Manash Hazarika. 2021. Attention-based hand semantic segmentation and gesture recognition using deep networks. *Evolving Systems* (2021).
[52] Saurav Kumar Sarma, Hemanta Kalita, and Rumi Devi. 2024. Semantic Segmentation based Hand Gesture Recognition using Deep Neural Networks. In *Proceedings of the 2nd International Conference on Computer Vision and Robotics (CVR 2024)*. Springer, 102–109. https://doi.org/10.1007/978-981-99-9053-5_9
[53] Marcos Silva, Tiago Araújo, Rostand Costa, Samuel Moreira, and Lucas Alves. 2021. Automation: Uma Ferramenta para a Evolução de um Tradutor Automático Português-Libras com Corpus Colaborativo. (2021).
[54] José Silveira, Rodrigo Vieira, and Adriano Bastos. 2023. *SynLibras-Pose Dataset*. https://doi.org/10.5281/zenodo.8415850 Acesso em: jul. 2025.
[55] Wellington Silveira, Andrew Alaniz, Marina Hurtado, Bernardo Castello Da Silva, and Rodrigo De Bem. 2022. SynLibras: A Disentangled Deep Generative Model for Brazilian Sign Language Synthesis. (2022).
[56] Wellington Silveira, Andrew Alaniz, Marina Hurtado, Luca Mendonça, and Rodrigo de Bem. 2023. Towards Generating Digital LIBRAS Signers on Mobile Devices. (2023).
[57] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
[58] Marina Sokolova and Guy Lapalme. 2009. A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management* 45, 4

(2009), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

[59] Esther de Oliveira Souza, Juliana Guimarães Faria, Layane Rodrigues de Lima, and Inés Martins. 2023. Digital Content in Brazilian Sign Language (Libras): A Study of Brazilian Channels of Deaf People on YouTube. *Comunicação e sociedade* 43 (2023), e023001.

[60] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

[61] Zhan Tong, Yibing Song, Jue Wang, and Limin Wang. 2022. VideoMAE: masked autoencoders are data-efficient learners for self-supervised video pre-training. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '22)*. Curran Associates Inc., Red Hook, NY, USA, Article 732, 16 pages.

[62] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[63] Vinícius Veríssimo and Rostand Costa. 2020. Using Data Augmentation and Neural Networks to Improve the Emotion Analysis of Brazilian Portuguese Texts. (2020).

[64] Vinícius Veríssimo, Cecília Silva, Vitor Hanael, Caio Moraes, Rostand Costa, Tiago Maritan, Manuella Aschoff, and Thaís Gaudêncio. 2019. A study on the use of sequence-to-sequence neural networks for automatic translation of brazilian portuguese to LIBRAS. (2019).

[65] Equipe VLibras. 2020. VLibrasBD - Corpus paralelo para tradução automática Português–Libras. Disponível mediante solicitação institucional ao projeto VLibras..

[66] Felix Wu, Angela Fan, Alexei Baevski, Yann N Dauphin, and Michael Auli. 2019. Pay less attention with lightweight and dynamic convolutions. *arXiv preprint arXiv:1901.10430* (2019).

[67] Hanaa ZainEldin, Samah A Gamel, Fatma M Talaat, Mansourah Aljohani, Nadiah A Baghdadi, Amer Malki, Mahmoud Badawy, and Mostafa A Elhosseini. 2024. Silent no more: a comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. *Artificial Intelligence Review* 57, 7 (2024), 188.

[68] Yanqiong Zhang and Xianwei Jiang. 2023. Recent Advances on Deep Learning for Sign Language Recognition. *CMES - Computer Modeling in Engineering and Sciences* (2023).