



Sistema Inteligente de Traducción de Lengua de Señas mediante computadora para la Traducción en Tiempo Real: una revisión sistemática

INTELLIGENT COMPUTER-BASED SIGN LANGUAGE TRANSLATION SYSTEM FOR REAL-TIME TRANSLATION: A SYSTEMATIC REVIEW

UNICATÓLICA - Fundación Universitaria Católica Lumen Gentium

Cristhian Orobio Mendoza

cristhian.orobio01@unicatolica.edu.co

Juan David Penilla Arroyave

Juan.penilla02@unicatolica.edu.co

Cristian Felipe Marín García

cristian.marin01@unicatolica.edu.co

Julio Hernán Larrahondo Zúñiga

Julio.larrahondo01@unicatolica.edu.co

Nazly Rocio Hincapié Monsalve

nrhincapie@unicatolica.edu.co

Resumen

Las barreras en la comunicación entre las comunidades sordas y oyentes son un reto social y tecnológico actual. El lenguaje de señas, medio natural de comunicación para muchas personas sordas, requiere soluciones que permitan su traducción en tiempo real para facilitar la interacción en contextos cotidianos. Los avances en visión por computadora y el desarrollo de aplicaciones han abierto nuevas oportunidades para abordar este desafío. El presente documento expone el diseño y desarrollo de un "Sistema Inteligente de Traducción de Lengua de Señas" que, mediante la integración de diversas tecnologías, convierte gestos en datos textuales o auditivos, garantizando portabilidad y acceso en distintos entornos. Este proyecto propone el desarrollo de un sistema inteligente que integra técnicas de visión por computadora para traducir en tiempo real el lenguaje de señas a texto o voz. El sistema detecta y reconoce los gestos, ofreciendo retroalimentación inmediata y facilitando la interacción. Se fundamenta en la revisión de estudios actuales en el reconocimiento de gestos y tecnologías.

Palabras clave: Lengua de señas, reconocimiento de gestos, visión por computadora, traducción en tiempo real, inteligencia artificial

Abstract

Barriers in communication between deaf and hearing communities are a current social and technological challenge. Sign language, a natural means of communication for many deaf people, requires solutions that allow its translation in real time to facilitate interaction in everyday contexts. Advances in computer vision and application development have opened new opportunities to address this challenge. This document exposes the design and development of an "Intelligent Sign Language Translation System" that, through the integration of various technologies, converts gestures into textual or auditory data, guaranteeing portability and access in different environments. This project proposes the development of an intelligent system that integrates computer vision techniques to translate sign language into text or voice in real time. The system detects and recognizes gestures, offering immediate feedback and facilitating interaction. It is based on the review of current studies in the recognition of gestures and technologies.

Keywords: Sign language, gesture recognition, computer vision, real-time translation, artificial intelligence.

Introducción

La comunicación efectiva entre personas sordas y oyentes ha sido históricamente un desafío debido a las diferencias en los modos de expresión y comprensión. El lenguaje de señas se rige como el medio principal de comunicación para la comunidad sorda, caracterizado por su naturaleza visual y gestual. No obstante, la ausencia de un conocimiento generalizado de este lenguaje entre la población oyente ha generado barreras significativas en diversos ámbitos sociales, educativos y laborales.

En respuesta a esta problemática, la investigación y el desarrollo tecnológico han orientado sus esfuerzos hacia la creación de sistemas que faciliten la traducción del lenguaje de señas a texto o voz, y viceversa. Estos sistemas buscan aprovechar los avances en áreas como la visión por computadora, la inteligencia artificial y los sensores de movimiento para interpretar y traducir los gestos de manera precisa y en tiempo real.

El desarrollo de sistemas de traducción de lenguaje de señas se sustenta en varios pilares tecnológicos y metodológicos:

La visión por computadora permite la captura y análisis de imágenes o secuencias de video para identificar patrones correspondientes a gestos específicos del lenguaje de señas. Técnicas como la extracción de características mediante Histogramas de Gradientes Orientados (HOG) y el uso de Momentos de Hu y Zernike han demostrado ser efectivas en la representación de formas y movimientos de las manos (Mancilla-Morales et al., 2019). Estas técnicas facilitan la diferenciación de señas al resaltar características distintivas en las imágenes capturadas.

La implementación de algoritmos de aprendizaje automático, especialmente las redes neuronales, ha sido crucial para el reconocimiento y clasificación de señas. Modelos como las Redes Neuronales Convolucionales (CNN) y las Máquinas de Vectores de Soporte (SVM) se entrenan con conjuntos de datos etiquetados para identificar patrones complejos en los gestos. Por ejemplo, Mancilla-Morales et al. (2019) analizaron el desempeño de redes neuronales multicapa y SVM en el reconocimiento de señas del Lenguaje de Señas Mexicano, considerando diferentes conjuntos de características.

Además de las cámaras tradicionales, el uso de sensores de movimiento, como acelerómetros y giroscopios, integrados en dispositivos vestibles, permite la captura detallada de los movimientos tridimensionales de las manos y dedos. Estos datos complementan la información visual y mejoran la precisión en la interpretación de las señas. Benito Martínez y Alba (2024) desarrollaron un sistema vestible para la traducción del lenguaje de signos utilizando sensores de movimiento, destacando la importancia de estos dispositivos en la captura precisa de gestos.

La proliferación de dispositivos móviles ha abierto nuevas oportunidades para la implementación de sistemas portátiles de traducción de lenguaje de señas. Las aplicaciones móviles pueden integrar cámaras y sensores para capturar gestos, procesarlos en tiempo real y proporcionar retroalimentación inmediata en forma de texto o voz. Investigadores de la Universidad de Alicante desarrollaron una aplicación de traducción simultánea de la Lengua de Signos Española, facilitando la comunicación en tiempo real entre personas sordas y oyentes mediante técnicas de procesamiento de lenguaje natural y visión por computadora (Radio Alicante, 2024).

La traducción de lenguaje de señas a voz ha avanzado significativamente con el desarrollo de sistemas portátiles asistidos por aprendizaje automático. Estos sistemas, como el presentado por Zhou et al. (2020), utilizan matrices de sensores elásticos basados en hilos y circuitos impresos inalámbricos para convertir gestos manuales en señales eléctricas, que luego son traducidas a voz en tiempo real. Este enfoque ofrece una alta sensibilidad, un tiempo de respuesta rápido y una tasa de reconocimiento del 98.63%, superando limitaciones de sistemas anteriores basados en electromiografía, efectos piezoresistivos o procesamiento de imágenes, que suelen ser más complejos, costosos y susceptibles a interferencias ambientales (Zhou et al., 2020). Además, la integración de algoritmos de aprendizaje automático, como el clasificador SVM multiclase, mejora la precisión y robustez del sistema, permitiendo una comunicación más efectiva entre usuarios de lenguaje de señas y no usuarios.

La traducción de lenguaje de señas se ha enfocado en el desarrollo de sistemas que utilizan visión por computadora y redes neuronales LSTM para reconocer y traducir gestos en tiempo real. Estos sistemas buscan superar las limitaciones de los métodos tradicionales, como la dependencia de intérpretes humanos y la falta de precisión en la traducción automática. Se investiga cómo las técnicas de visión por computadora, combinadas con redes neuronales recurrentes (RNN) y LSTM, pueden capturar la dinámica temporal de los gestos, mejorando la precisión en la clasificación y traducción de señas. Sin embargo, existen limitaciones, como la necesidad de grandes conjuntos de datos etiquetados, la sensibilidad a las condiciones ambientales (iluminación, fondo) y la dificultad para procesar gestos en tiempo real con alta precisión. El propósito de estos sistemas es facilitar la comunicación entre personas sordas y oyentes, promoviendo la inclusión social y el acceso a servicios esenciales. Las referencias clave incluyen trabajos como los de Zhou et al. (2020) en sistemas portátiles, y estudios previos que utilizan CNN y LSTM para la clasificación de gestos (Cui et al., 2019; Hochreiter & Schmidhuber, 1997).

En la traducción de lenguaje de señas se ha enfocado en el desarrollo de sistemas centrados en el usuario, como "Manu", que integran aprendizaje automático y metodologías de design thinking para mejorar la accesibilidad y la precisión en la traducción de ASL. Este sistema combina algoritmos de Machine Learning, tecnologías en la nube y una interfaz React para ofrecer traducción en tiempo real, lecciones interactivas y seguimiento del progreso. Se investiga cómo estas tecnologías pueden superar las limitaciones de los sistemas tradicionales, como la falta de recursos de aprendizaje y la baja precisión en la traducción. Sin embargo, existen limitaciones, como la necesidad de mejorar la consistencia del sistema y la precisión de los algoritmos de aprendizaje automático, especialmente en entornos con variaciones en iluminación y fondos. El propósito de "Manu" es facilitar la comunicación entre personas sordas y oyentes, promoviendo la inclusión social y el acceso a la educación en ASL. Las referencias clave incluyen trabajos como los de Chong y Lee (2018) en reconocimiento de lenguaje de señas, y estudios previos que utilizan LSTM y GRU para la traducción de señas (Kothadiya et al., 2022; Jintanechaiwat et al., 2024).

En el reconocimiento de gestos manuales para la comunicación de personas sordas y mudas se ha enfocado en el desarrollo de sistemas basados en visión por computadora y redes neuronales convolucionales (CNN). Este documento presenta un prototipo de software que utiliza técnicas de reconocimiento de patrones y gestos para interpretar el lenguaje de señas en tiempo real, con el objetivo de mejorar la accesibilidad y la inclusión de personas con discapacidades auditivas o del habla. Se investiga cómo las características basadas en la forma, como la orientación, la posición del centroide y la ubicación de los dedos, pueden ser utilizadas para reconocer gestos manuales de manera eficiente. Sin embargo, existen limitaciones, como la variabilidad en los gestos, la sensibilidad a las condiciones ambientales (iluminación, fondo) y la dificultad para traducir gestos individuales en oraciones coherentes.

El propósito del sistema es facilitar la comunicación entre personas sordas/mudas y oyentes, promoviendo la inclusión social y el acceso a la información. Las referencias clave incluyen trabajos como los de Salian et al. (2017) en reconocimiento de lenguaje de señas, y estudios previos que utilizan CNN para la clasificación de gestos (Srivastava et al., 2014; Jin et al., 2016).

La traducción de lenguaje de señas mediante smartphones se ha enfocado en el desarrollo de sistemas que utilizan redes neuronales convolucionales (CNN) para reconocer y traducir gestos en tiempo real. Este documento presenta un sistema basado en smartphones que utiliza Kotlin y MediaPipe para capturar y procesar imágenes de gestos de lenguaje de señas, traducándose a texto en tiempo real. Se investiga cómo las CNN pueden mejorar la precisión y eficiencia en el reconocimiento de gestos, especialmente en el contexto del Lenguaje de Señas Filipino (FSL). Sin embargo, existen limitaciones, como la variabilidad en los gestos, las condiciones de iluminación y la necesidad de expandir el soporte a múltiples lenguajes de señas. El propósito del sistema es facilitar la comunicación entre personas sordas y oyentes, promoviendo la inclusión social y el acceso a la información. Las referencias clave incluyen trabajos como los de Goyal y Goyal (2017) en traducción automática, y estudios previos que utilizan CNN para el reconocimiento de lenguaje de señas (Seema y Sing, 2023; Ugale et al., 2023).

Se presenta un avance significativo en la traducción automática del lenguaje de señas mediante un enfoque de aprendizaje profundo. La investigación se centra en superar las limitaciones de los sistemas existentes, que a menudo tienen dificultades para capturar la complejidad visual del lenguaje de señas, incluyendo la configuración de las manos, la orientación, el movimiento y el significado contextual. El propósito del estudio es desarrollar una arquitectura de aprendizaje profundo optimizada que descompone los gestos en sus componentes semióticos esenciales, permitiendo una representación más precisa y detallada del lenguaje visual. Las limitaciones identificadas incluyen la falta de estandarización en la notación del lenguaje de señas, la escasez de grandes conjuntos de datos anotados y la dificultad para capturar la variabilidad regional y cultural en los gestos. El estudio propone un decodificador semiótico visual (SLVSD) que combina un codificador semiótico visual (VSE) y un decodificador de lenguaje de señas (SLD), logrando una precisión del 99.65% en la clasificación de gestos sintéticos. Las referencias clave incluyen trabajos previos en reconocimiento de gestos, traducción automática y modelos de aprendizaje profundo, como los de Stokoe, Liddell, y Bellugi, entre otros.

Es un sistema de detección en tiempo real del Lenguaje de Señas Indio (ISL) utilizando redes LSTM (Long Short-Term Memory) y la extracción de puntos clave mediante el módulo Mediapipe. La investigación se centra en superar las limitaciones de los sistemas existentes, que en su mayoría se basan en el Lenguaje de Señas Americano (ASL) y no están adaptados a las particularidades del ISL, como el uso de ambas manos y la postura corporal. El propósito del estudio es desarrollar una herramienta de traducción de gestos manuales en tiempo real que convierte los gestos del ISL en texto, facilitando la comunicación entre personas con discapacidad auditiva y el público en general. Las limitaciones incluyen la falta de un conjunto de datos estandarizado para el ISL, la complejidad de los gestos dinámicos y la variabilidad en los ángulos de la cámara y las condiciones de iluminación. El sistema propuesto logra una precisión del 84.1% en la clasificación de gestos, con un enfoque en la creación de un conjunto de datos personalizado y el uso de LSTM para la clasificación de gestos dinámicos.

El desarrollo de un sistema de traducción en tiempo real del Lenguaje de Señas Americano (ASL) utilizando técnicas de aprendizaje profundo, específicamente la arquitectura SSD MobileNet V2 FPN Lite. El estudio se centra en mejorar la precisión del reconocimiento de gestos en condiciones de iluminación variables, como entornos con poca luz o sobreexpuestos, mediante la conversión de espacios de color (RGB a HSV y YCbCr). Las limitaciones incluyen la dependencia de la calidad de la cámara, la variabilidad en las condiciones de iluminación y la necesidad de expandir el vocabulario de señas reconocidas. El propósito del sistema es facilitar la comunicación entre personas sordas y oyentes, proporcionando una herramienta accesible y no invasiva que funcione en dispositivos de bajo consumo. El sistema logra una alta precisión en la detección de gestos bajo diversas condiciones ambientales, demostrando su potencial para mejorar la inclusión y accesibilidad de la comunidad sorda.

El desarrollo de un sistema de traducción en tiempo real del Lenguaje de Señas Americano (ASL) a texto y voz utilizando redes neuronales convolucionales (CNN). El estudio se centra en la detección de gestos manuales a través de una cámara, utilizando un modelo CNN entrenado con un conjunto de datos de 1000 imágenes de gestos de ASL. Las limitaciones incluyen la necesidad de un entorno controlado con iluminación adecuada y la dificultad para reconocer gestos complejos que involucran movimientos dinámicos y expresiones faciales. El propósito del sistema es facilitar la comunicación entre personas sordas y oyentes, eliminando la necesidad de un intérprete humano. El sistema logra una precisión del 95% en la traducción de gestos de deletreo manual (finger spelling), demostrando su eficacia en la conversión de señales estáticas a texto y voz.

Investiga el desarrollo de un sistema de reconocimiento en tiempo real del Lenguaje de Señas Indio (ISL) utilizando la arquitectura MobileNetV2 y técnicas de transfer learning. El estudio se centra en superar las limitaciones de los modelos tradicionales, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), que requieren un alto poder computacional y grandes cantidades de datos etiquetados. Las limitaciones incluyen la necesidad de un entorno controlado para la captura de gestos y la dificultad para reconocer gestos dinámicos y expresiones faciales. El propósito del sistema es facilitar la comunicación entre personas sordas y oyentes, proporcionando una herramienta accesible y eficiente que funcione en dispositivos móviles y de bajo consumo. El sistema logra una precisión del 97% con una latencia de 60 milisegundos por frame, demostrando su eficacia en la traducción de gestos estáticos a texto y voz en tiempo real.

Un sistema innovador para la traducción en tiempo real del lenguaje de señas a texto y viceversa, utilizando tecnologías avanzadas como TensorFlow, MediaPipe y procesamiento de imágenes. El propósito principal de esta investigación es cerrar la brecha de comunicación entre la comunidad sorda y muda y aquellos que no comprenden el lenguaje de señas, promoviendo la inclusión y la accesibilidad. El sistema propuesto permite la captura de gestos mediante una cámara, que luego son traducidos a texto o lenguaje hablado, y también convierte texto o audio en lenguaje de señas mediante avatares animados. Además, incluye un modo de aprendizaje para enseñar el alfabeto y los números en lenguaje de señas. Las limitaciones del sistema incluyen la necesidad de mejorar la precisión en el reconocimiento de gestos complejos, la dependencia de la calidad del video capturado y la necesidad de ampliar la base de datos de gestos para cubrir más lenguajes de señas y variaciones regionales. Aunque el sistema logra una precisión del 86% utilizando el optimizador Adam, se identificó que aumentar el número de capas convolucionales no mejora significativamente la precisión, pero sí incrementa el tiempo de procesamiento.

La convergencia de estas tecnologías ha permitido avances significativos en la creación de sistemas más precisos y accesibles para la traducción del lenguaje de señas. Sin embargo, persisten desafíos relacionados con la variabilidad en las expresiones individuales de las señas, las diferencias regionales en los lenguajes de señas y la necesidad de sistemas que operen eficazmente en diversos entornos y condiciones de iluminación. La investigación continua y el desarrollo de soluciones adaptativas son esenciales para superar estas barreras y promover una comunicación inclusiva y efectiva entre las comunidades sordas y oyentes.

Metodología de Investigación

Esta revisión sistemática tiene como objetivo resumir los hallazgos y conclusiones reportados en la literatura sobre los sistemas de traducción de lenguaje de señas utilizando técnicas de inteligencia artificial y aprendizaje profundo. Se realiza una búsqueda bibliográfica con el fin de identificar estudios relevantes en este campo. Con respecto a la literatura reciente sobre la traducción de lenguaje de señas mediante técnicas de aprendizaje profundo, surge la siguiente pregunta de investigación: ¿Qué tecnologías se están utilizando en los sistemas inteligentes de traducción de lenguaje de señas y cuáles son sus principales limitaciones?

Los objetivos de esta revisión sistemática son los siguientes:

- Analizar las tecnologías predominantes en los sistemas actuales de traducción de lenguaje de señas.
- Evaluar las principales barreras tecnológicas y desafíos en dichos sistemas.
- Proponer áreas clave de investigación y desarrollo para superar las limitaciones tecnológicas detectadas.

Esta revisión sistemática busca proporcionar una visión integral sobre el estado actual de la investigación en este campo, facilitando la identificación de oportunidades para el desarrollo de sistemas más precisos, accesibles y eficientes en la traducción del lenguaje de señas.

Análisis bibliométrico: Se llevó a cabo un análisis bibliométrico con el apoyo de la herramienta Biblioshiny (interfaz gráfica de la librería bibliometrix de R), la cual facilitó la organización y el procesamiento de los datos bibliográficos. Esta herramienta permitió obtener indicadores cuantitativos, realizar mapas de coocurrencia de palabras clave y visualizar la evolución temática de las publicaciones incluidas en la revisión.

Estrategia de búsqueda: Se lleva a cabo siguiendo las pautas establecidas por PRISMA como se observa en la Figura 1. Se utilizaron operadores booleanos para combinar términos clave como "lenguaje de señas", "reconocimiento de gestos", "inteligencia artificial" y "visión por computadora" en las bases de datos, scholar google, semantic scholar y lens. La búsqueda se limitó a estudios publicados en inglés y en revistas de acceso abierto, comprendiendo el período desde el año 2020 hasta la fecha usando la ecuación de búsqueda ("mobile application" OR "mobile app" OR "mobile device") AND ("sign language" OR "sign language" OR "sign language") AND ("machine learning" OR "machine learning" OR "deep learning" OR "deep learning") AND ("gesture recognition" OR "gesture recognition") AND ("real-time translation" OR "real-time translation"). Tras la eliminación de duplicados, se realizó una revisión de títulos y resúmenes para descartar estudios irrelevantes, utilizando la herramienta Rayyan AI para facilitar la organización y el filtrado de los artículos. Finalmente, la selección de los estudios se efectuó mediante la aplicación de criterios de inclusión y exclusión, siendo el proceso realizado de manera independiente por tres revisores.

Procedimientos de inclusión y exclusión: Se definieron los siguientes criterios de inclusión y exclusión para la selección de estudios: se incluyeron únicamente documentos disponibles en Google Scholar, Semantic Scholar y Lens, publicados a partir de 2020, en inglés y que fueran de acceso abierto o gratuitos. Además, se consideraron exclusivamente aquellos estudios que abordaran sistemas de traducción de lenguaje de señas y que implementaran técnicas de inteligencia artificial, dado el enfoque de la revisión en tecnologías inteligentes. En contraste, se excluyeron aquellos estudios que no se relacionarán directamente con el lenguaje de señas y aquellos que no incorporan la inteligencia artificial en sus metodologías de traducción.

Evaluación de la Calidad: Para analizar la calidad metodológica de esta revisión, cuatro revisores realizaron el análisis de forma independiente siguiendo las directrices establecidas por PRISMA. Además, se utilizó Rayyan, una aplicación web y móvil para revisiones sistemáticas, que permitió identificar el riesgo de sesgo en cada estudio, considerando parámetros como la aleatorización y el cegamiento. En caso de desacuerdo sobre la calidad de los estudios, se consultó a un quinto revisor, quien resolvió las discrepancias.

Extracción y Análisis de Datos: En nuestro proyecto de traducción de lenguaje de señas, se extrajeron y analizaron variables clave de los estudios incluidos en la revisión sistemática. A continuación, se presenta una breve descripción de los datos relevantes que se extrajeron:

Título: El título del estudio.

Objetivo : El propósito del artículo o investigación .

Año: El año en el que se realizó o publicó el estudio.

Integración con inteligencia artificial : arquitectura o red neuronal empleada en el estudio

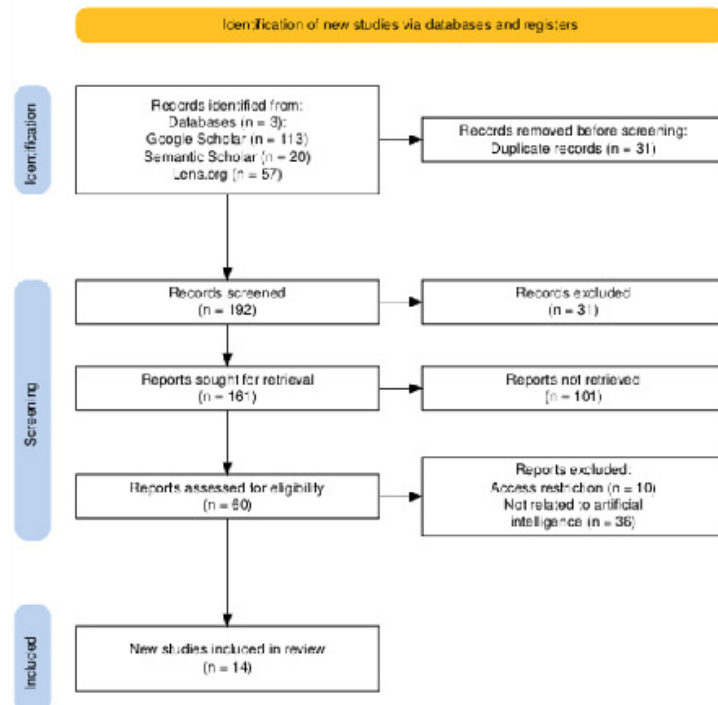
Limitaciones : principales limitación en el proyecto

Avances: principales resultados en el avance tecnológico

Como resultado de esta investigación, se espera identificar las tecnologías más utilizadas en los sistemas de traducción de lenguaje de señas, además de proporcionar una visión detallada de las innovaciones actuales en este campo y sus principales desafíos.

Resultados y discusiones: Catorce artículos fueron incluidos en el análisis de los sistemas de traducción de lenguaje de señas mediante técnicas de inteligencia artificial y aprendizaje profundo. Los estudios revisados exploran diversas metodologías, incluyendo redes neuronales convolucionales (CNN), modelos recurrentes (LSTM) y el uso de sensores vestibulares, evidenciando avances en la precisión y eficiencia de estos sistemas. Sin embargo, persisten desafíos como la variabilidad en la ejecución de gestos, las condiciones ambientales y la disponibilidad de datos de entrenamiento representativos. Para una mejor comprensión del proceso de selección de estudios, se proporciona un diagrama de flujo que detalla la identificación, evaluación y exclusión de artículos según los criterios definidos.

Figura 1. Identification of new studies via databases and registers



Fuente. Diagrama de flujo de Prisma para la identificación de estudios creado en estech.shinyapps.io/prisma_flowdiagram

Se llevó a cabo un análisis bibliométrico con el apoyo de la herramienta Biblioshiny (interfaz gráfica de la librería bibliometrix de R), la cual facilitó la organización y el procesamiento de los datos bibliográficos. Esta herramienta permitió obtener indicadores cuantitativos, realizar mapas de coocurrencia de palabras clave y visualizar la evolución temática de las publicaciones incluidas en la revisión.

Una representación visual de términos clave obtenida a partir de datos, como se observa en la Figura 2, resalta los términos clave más relevantes en el ámbito de la investigación sobre reconocimiento y traducción del lenguaje de señas. Conceptos como "aprendizaje", "reconocimiento", "traducción", "sistema" y "gesto" emergen con gran relevancia, evidenciando su frecuente coexistencia en estudios recientes. Además, términos como "red neuronal", "detección facial", "micro-movimientos" y "real-time" sugieren el papel crucial de la inteligencia artificial y las tecnologías avanzadas en el desarrollo de sistemas de accesibilidad para personas con discapacidad auditiva. La intersección de estas palabras indica un enfoque creciente en la automatización del reconocimiento de señas mediante algoritmos de aprendizaje profundo, con aplicaciones en dispositivos portátiles y asistentes inteligentes. Así, la investigación en este campo continúa evolucionando hacia sistemas más precisos y eficientes, facilitando la comunicación inclusiva mediante innovaciones tecnológicas.

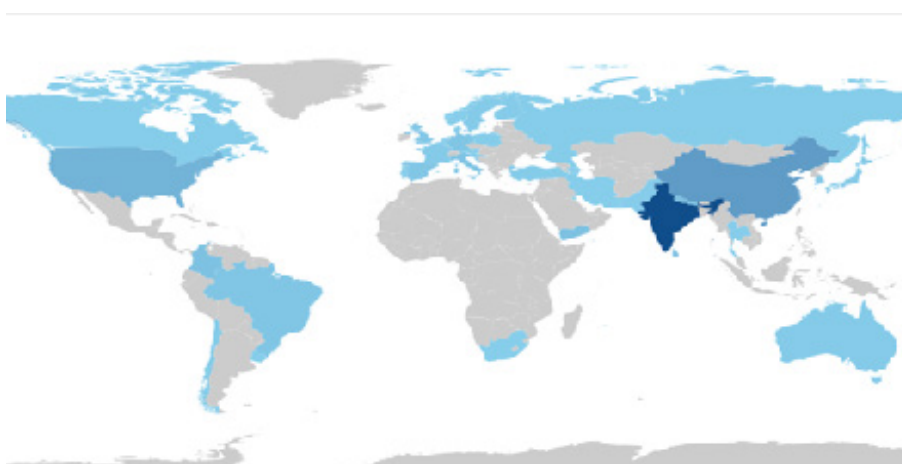
Figura 2. Diagrama de Nube de palabras que representa las palabras clave más utilizadas



Fuente. Creación propia. Creado con Biblioshiny

En el panorama global de la investigación científica en este campo, se destaca la contribución activa de diversos países al avance del conocimiento. En particular, India, China y Estados Unidos emergen como los principales actores en el desarrollo de estudios sobre este tema, como se muestra en la Figura 3.

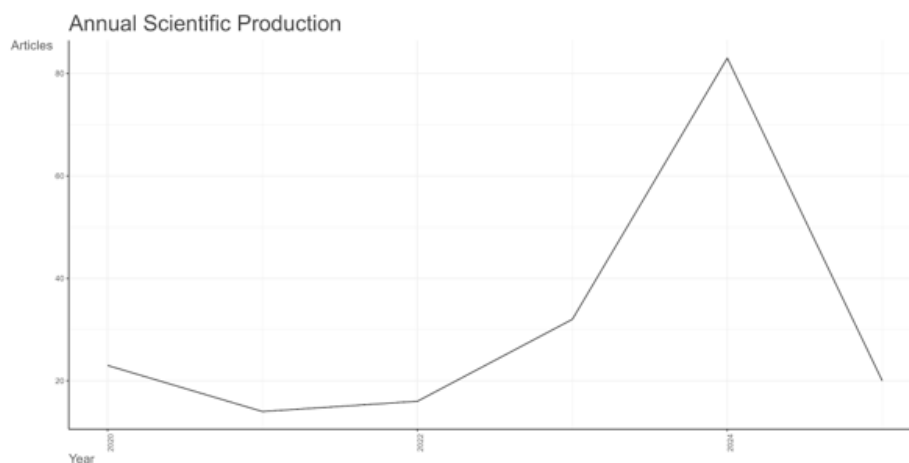
Figura 3. Diagrama de Producción científica de los países que representa los países con más aportes a esta investigación



Fuente. Creación propia. Creado con Biblioshiny

La Figura 4 presenta un diagrama de producción científica anual que muestra la cantidad de artículos publicados por año desde el año 2020. Se observa una tendencia variable, con un incremento significativo entre 2022 y 2024. lo que nos permiten analizar la evolución de la investigación y detectar posibles factores que influyen en la variación de la producción académica a lo largo del tiempo

Figura 4. Producción científica anual que representa cantidad de artículos producidos por año desde el 2020



Fuente. Creación propia. Creado con Biblioshiny

Para concluir, se presenta un análisis detallado de los estudios más relevantes incluidos en la revisión sistemática. La Tabla I, en donde estudia la aplicación de la Inteligencia Artificial (IA) en la traducción del lenguaje de señas, un campo que ha experimentado avances significativos gracias a la integración de modelos de aprendizaje profundo. La sistematización de diversos proyectos evidencia que el uso de Redes Neuronales Convolucionales (CNN), modelos Long Short-Term Memory (LSTM) y frameworks como TensorFlow han permitido desarrollar sistemas de traducción con altos niveles de precisión, alcanzando hasta un 98.63% en algunos casos. Los resultados reflejan la efectividad de estos modelos en la interpretación de señas en tiempo real, facilitando la comunicación para personas con discapacidad auditiva. No obstante, se identificaron diversas limitaciones que aún representan retos para su implementación generalizada. Entre ellas, destacan la dependencia de conjuntos de datos de calidad, la sensibilidad a condiciones ambientales como iluminación y variabilidad gestual, así como la necesidad de recursos computacionales elevados para ejecutar los modelos con eficiencia. Desde una perspectiva de sistematización, los hallazgos sugieren que futuras investigaciones deben centrarse en la ampliación y diversificación de bases de datos de lenguaje de señas, la optimización de modelos para dispositivos de bajo rendimiento y el desarrollo de soluciones más robustas ante variaciones en el entorno.

A continuación, resume los trabajos seleccionados para el análisis tras la aplicación de la metodología PRISMA.

Tabla 1. Análisis tras la aplicación de la metodología PRISMA

Título	Año	Objetivo del Artículo	Resultados Principales	Integración con IA	Principales Limitaciones
Sign Language Translator System Using Computer Vision And LSTM Neural Networks (Wang et al., 2024)	2024	Desarrollar un sistema que reconozca y traduzca gestos del lenguaje de señas a texto (y voz) en tiempo real mediante visión por computadora y redes neuronales.	Alcanzó una precisión del 96% en el reconocimiento de gestos, entrenado con un conjunto de 10,000 imágenes.	Utiliza modelos de deep learning (CNN y LSTM) para procesar imágenes y secuencias de gestos.	Dependencia de la calidad del conjunto de datos y sensibilidad a condiciones ambientales (iluminación, ángulos de captura, etc.).
Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays (Zhou et al., 2024)	2024	Desarrollar un dispositivo wearable que traduzca el lenguaje de señas a voz mediante sensores flexibles y procesamiento de señales.	Logró un reconocimiento del 98.63% con un tiempo de respuesta inferior a 1 segundo.	Emplea algoritmos de machine learning para clasificar señales electromagnéticas obtenidas por sensores vestibles.	Complejidad estructural, requerimientos de materiales de alta calidad y vulnerabilidad a interferencias ambientales.
EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and AI (Kumar et al., 2020)	2020	Crear un sistema interactivo que traduzca el lenguaje de señas de Sri Lanka a texto y voz, y que convierte texto a señas animadas en una interfaz digital.	Demostró eficacia en la traducción bidireccional (señas a texto/voz y viceversa) con precisión aceptable (valor numérico no especificado).	Incorpora algoritmos de machine learning y técnicas de procesamiento de lenguaje natural para mejorar la contextualización de la traducción.	Limitaciones relacionadas con la disponibilidad de conjuntos de datos robustos y la adaptación a variaciones dialectales del lenguaje de señas.
Sign Language Translator via Smartphone Image Analysis using Convolutional Neural Network (Torres & Alday, 2024)	2024	Desarrollar una aplicación móvil que traduzca gestos del lenguaje de señas capturados mediante la cámara del smartphone a texto en tiempo real.	Obtuvo una precisión del 92% utilizando una base de datos de 5,000 gestos del lenguaje de señas filipino; la aplicación presenta una interfaz intuitiva.	Emplea CNN y técnicas de procesamiento de imágenes (con MediaPipe y Kotlin) para el reconocimiento en tiempo real.	Variabilidad en la ejecución de gestos, condiciones de iluminación y un tamaño limitado del conjunto de datos de entrenamiento.

*Sistema Inteligente de Traducción de Lengua de Señas
mediante computadora para la Traducción en Tiempo
Real: una revisión sistemática*

Real Time Sign Language Translator Using Machine Learning (Repal et al., 2024)	2024	Implementar una aplicación web para la traducción en tiempo real del lenguaje de señas a texto y voz, facilitando la comunicación bidireccional	Presenta resultados prometedores en precisión y tiempos de respuesta, aprovechando técnicas de reconocimiento basadas en TensorFlow.	Utiliza TensorFlow y técnicas avanzadas de machine learning para el procesamiento de imágenes y reconocimiento de gestos.	Dificultades relacionadas con condiciones de iluminación, movimientos variables de los usuarios y la necesidad de grandes bases de datos para entrenar el modelo.
Sign Language Visual Semiotic Decoder Using Deep Learning Approach (Rodríguez-Rodríguez & Alfonso-Morales, 2024)	2024	Proponer un método innovador que modele los gestos de lenguaje de señas mediante análisis semiótico y deep learning.	Alcanzó una precisión del 99.65% en medidas de correlación (Cohen's Kappa y Matthews) utilizando gestos sintéticos.	Emplea deep learning para modelar la semiótica de los gestos a través de redes neuronales optimizadas.	Dependencia de gestos sintéticos, validación pendiente con datos reales y adaptación a la diversidad cultural.
Enhancing Communication Accessibility: A Hand Gesture Recognition System for Deaf and Mute Individuals (Ramya et al., 2024)	2024	Desarrollar un sistema de reconocimiento de gestos manuales que mejore la comunicación para personas sordas y mudas.	Demostró eficacia en el reconocimiento en tiempo real, utilizando características basadas en la forma, orientación y posición de los dedos.	Utiliza algoritmos de reconocimiento de patrones y machine learning para clasificar gestos.	Sensibilidad a condiciones de iluminación, variabilidad en la ejecución de gestos y dependencia de la calidad de la imagen.
SIGN LANGUAGE TO TEXT IN REAL TIME USING CONVOLUTIONAL NEURAL NETWORK (Tripathi et al., 2023)	2023	Convertir los movimientos de la ASL a texto y voz en tiempo real mediante un programa de escritorio basado en CNN.	Traducción en tiempo real con una precisión mínima reportada del 75%, utilizando un conjunto de datos de 1,000 imágenes.	Emplea CNN para detección y procesamiento de gestos mediante OpenCV y Python.	Requiere hardware de calidad, es sensible a condiciones ambientales y muestra variabilidad en la precisión según el conjunto de datos.

Sign Language Visual Semiotic Decoder Using Deep Learning Approach	2024	Desarrollar una metodología innovadora para la traducción del lenguaje de señas utilizando un enfoque basado en deep learning. El estudio busca mejorar la interpretación automática de señas mediante un análisis semiótico detallado, descomponiendo los gestos en sus componentes fundamentales.	Se presentó un sistema estructurado en dos partes: el Visual Semiotic Encoder (VSE), que convierte gestos en representaciones textuales, y el Sign Language Decoder (SLD), que realiza la interpretación mediante redes neuronales profundas.	Se emplean redes neuronales profundas basadas en arquitecturas como Inception, ResNet y redes Siamese, optimizadas para analizar y clasificar signos	Dependencia de datos sintéticos. Falta de estandarización Limitaciones en la variabilidad de los gestos.
Designing a User-Centric American Sign Language Translation System: Integrating Machine Learning with Design Thinking (Surachettapong et al., NR)	2024	Desarrollar un sistema centrado en el usuario para la traducción y aprendizaje de ASL, integrando machine learning y design thinking.	Validación de usabilidad con un SUS score de 74.69 y precisión de traducción del 81.24%.	Combina algoritmos de machine learning con metodologías de design thinking y plataformas cloud.	Requiere mejoras en la velocidad de traducción, interfaz de usuario y acceso móvil.
Enhancing Gesture Recognition for Sign Language Interpretation in Challenging Environment Conditions: A Deep Learning Approach (Amalfitano et al., 2023)	2023	Mejorar el reconocimiento de gestos en condiciones ambientales difíciles mediante deep learning.	Alta precisión en reconocimiento y adaptabilidad a condiciones de iluminación variable mediante técnicas de cambio de espacio de color.	Utiliza deep learning (SSD MobileNet FPN), transfer learning y data augmentation.	Requiere adaptaciones específicas para diversas condiciones de iluminación y exposición.

*Sistema Inteligente de Traducción de Lengua de Señas
mediante computadora para la Traducción en Tiempo
Real: una revisión sistemática*

HAND SIGN LANGUAGE RECOGNITION USING AUGMENTED REALITY & MACHINE LEARNING (Asif et al., 2024)	2024	Desarrollar un sistema basado en AR y ML para el reconocimiento en tiempo real del lenguaje de señas y su traducción a voz.	Demostró alta precisión en tiempo real y mejoró la experiencia del usuario mediante AR.	Integra técnicas de machine learning y reconocimiento de patrones con AR y MediaPipe.	Sensible a condiciones de iluminación y dependencia de la calidad de imagen.
Real Time Indian Sign Language Detection Using LSTM and Keypoint Extraction (Tharsan et al., 2022)	2022	Traducir gestos del Indian Sign Language a texto en tiempo real mediante la extracción de keypoints y un modelo LSTM.	Implementación exitosa en tiempo real utilizando Mediapipe para keypoint extraction y LSTM para clasificación.	Utiliza un modelo LSTM combinado con técnicas de extracción de keypoints.	Problemas con variabilidad en gestos, condiciones ambientales y disponibilidad limitada de datos.
Real-time Sign Language Recognition Using MobileNetV2 and Transfer Learning (Temkar et al., 2023)	2023	Desarrollar un sistema de reconocimiento en tiempo real para ISL, eficiente para dispositivos móviles mediante transfer learning.	Modelo preciso y computacionalmente eficiente, apto para dispositivos con recursos limitados.	Utiliza MobileNetV2 preentrenado junto con técnicas de transfer learning para adaptación a ISL.	Posible latencia en tiempo real y dependencia de datos etiquetados de alta calidad.

Fuente. Creación propia

Se recomienda para potenciar futuras investigaciones en la traducción automática de la lengua de señas mediante inteligencia artificial, se recomienda emplear herramientas avanzadas como TensorFlow, que han demostrado alta precisión en el reconocimiento de señas. Además, es fundamental compartir públicamente los conjuntos de datos de gestos utilizados, facilitando así el desarrollo de modelos más precisos. Asimismo, se sugiere fomentar la colaboración internacional entre investigadores, lo que enriquecerá las perspectivas y soluciones en este ámbito. Dado que existen diferencias significativas entre los lenguajes de señas de cada país, la cooperación global permitirá abordar estas variaciones y desarrollar sistemas más inclusivos y efectivos.

Referencias bibliográficas

- Camgoz, N. C., Hadfield, S., Koller, O., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10023-10033. <https://doi.org/10.1109/CVPR42600.2020.01004>
- Dhiman, R., Joshi, G., & Rama Krishna, C. (2021). A deep learning approach for Indian sign language gestures classification with different backgrounds. *Journal of Physics: Conference Series*, 1742(1), 012034. <https://doi.org/10.1088/1742-6596/1742/1/012034>
- Fernandes, L., Dalvi, P., Junnarkar, & Bansode, M. (2020). Convolutional Neural Network based Bidirectional Sign Language Translation System. *Proceedings of the Third International Conference on Smart Systems and Inventive Technology*, 769-775. <https://doi.org/10.1109/ICS-SIT48917.2020.9214189>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- Jintanechaiwat, W., Jongsathtipahubil, K., Pimsan, N., et al. (2024). Using LSTM to translate Thai sign language to text in real time. *Discover Artificial Intelligence*, 4(17). <https://doi.org/10.1007/s44163-024-00113-8>
- Kim, C. J., & Park, H. M. (2021). Per-frame Sign Language Gloss Recognition. *Proceedings of the International Conference on Information and Communication Technology Convergence*, 1125-1127. <https://doi.org/10.1109/ICTC52510.2021.9620911>
- Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-Gonzalez, A.-B., & Corchado, J. M. (2022). Deep-sign: Sign Language Detection and Recognition Using Deep Learning. *Electronics*, 11(11), 1780. <https://doi.org/10.3390/electronics11111780>
- Kumar, D. M., Bastiansz, G. M. A. S., & Bavanraj, K. (2020). EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and AI. *Sri Lanka Institute of Information Technology*. <https://ieeexplore.ieee.org/document/9357154>
- Madani, K., et al. (2023). A combined approach for improving humanoid robots autonomous cognitive capabilities. *Knowledge and Information Systems*, 65(8), 3197-3221. <https://doi.org/10.1007/s10115-023-01854-1>
- Muscetti, M., et al. (2022). Multimedia ontology population through semantic analysis and hierarchical deep features extraction techniques. *Knowledge and Information Systems*, 64(5), 1283-1303. <https://doi.org/10.1007/s10115-022-01667-8>
- Pahuja, D., & Jain, S. (2020). Recognition of Sign Language Symbols using Templates. *Proceedings of the International Conference on Reliability*, 1157-1160. <https://ieeexplore.ieee.org/document/9198001>
- Rastgoo, R., et al. (2021). Sign language recognition: a deep survey. *Expert Systems with Applications*, 164, 113794. <https://doi.org/10.1016/j.eswa.2020.113794>
- Rinaldi, A. M., et al. (2021). Visual query posing in multimedia web document retrieval. *ICSC*, 415-420. <https://ieeexplore.ieee.org/document/9364496>
- Rodríguez-Rodríguez, J. J., & Alfonso-Morales, W. (2024). Sign Language Visual Semiotic Decoder Using Deep Learning Approach. *Preprints.org*. <https://doi.org/10.20944/preprints202411.0933.v1>

- Saggio, G., Cavallo, P., Ricci, M., Errico, V., Zea, J., & Benalc'azar, M. E. (2020). Sign language recognition using wearable electronics: Implementing k-nearest neighbors with dynamic time warping and convolutional neural network algorithms. *Sensors*, 20(14), 3879. <https://doi.org/10.3390/s20143879>
- Seema, S., & Sing, P. (2023). A Comprehensive Review of CNN-Based Sign Language Translation System. *Proceedings of Data Analytics and Management*, 572, 347-362. https://doi.org/10.1007/978-981-19-7615-5_32
- Ugale, M., Rodrigues, O., Shinde, A., Desle, K., & Yadav, S. (2023). A Review on Sign Language Recognition Using CNN. *ResearchGate*. https://www.researchgate.net/publication/370425420_A_Review_on_Sign_Language_Recognition_Using_CNN
- Wang, C., et al. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*. <https://arxiv.org/abs/2207.02696>
- Zhou, Z., Chen, K., Li, X., Zhang, S., Wu, Y., Zhou, Y., Meng, K., Sun, C., He, Q., Fan, W., Fan, E., Lin, Z., Tan, X., Deng, W., Yang, J., & Chen, J. (2020). Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nature Electronics*, 3(9), 571-580. <https://doi.org/10.1038/s41928-020-0428-6>
- Al-Qurishi, M., Khalid, T., & Souissi, R. (2021). Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. *IEEE Access*, 9, 126917-126951. <https://ieeexplore.ieee.org/document/9530569>
- Bencherif, M. A. (2021). Arabic Sign Language Recognition System Using 2D Hands and Body Skeleton Data. *IEEE Access*, 9, 59612-59627. <https://ieeexplore.ieee.org/document/9389720>
- Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to Mask R-CNN: a survey. *Computational Intelligence in Pattern Recognition*, 657-668. <https://www.sciencedirect.com/topics/computer-science/deep-learning-technique>

PODCAST



SISTEMA INTELIGENTE DE TRADUCCIÓN DE LENGUA DE SEÑAS MEDIANTE COMPUTADORA PARA LA TRADUCCIÓN EN TIEMPO REAL: UNA REVISIÓN SISTEMÁTICA

Cristhian Orobio Mendoza

cristhian.orobio01@unicatolica.edu.co

Juan David Penilla Arroyave

Juan.penilla02@unicatolica.edu.co

Cristian Felipe Marín García

cristian.marin01@unicatolica.edu.co

Julio Hernán Larrahondo Zúñiga

Julio.larrahondo01@unicatolica.edu.co

Nazly Rocio Hincapié Monsalve

nrhincapie@unicatolica.edu.co