

Glove-translator system for sign language based on convolutional neural networks (CNN): A systematic literature review

Mariajose Guadalupe Pedroza Garcia¹; Angelina Nereyda Silva Agape²; Evelyn Elizabeth Ayala Ñiquen³
^{1,2,3} Universidad Tecnológica del Perú, Perú, U22212450@utp.edu.pe, U22244082@utp.edu.pe, c26915@utp.edu.pe

Abstract– This systematic literature review (SLR) evaluated the design approaches, technical effectiveness, and existing gaps in the implementation of glove-based sign translator systems that employ Convolutional Neural Networks (CNNs). The PRISMA methodology was used to organize and document the study selection and evaluation process. The initial search of the Scopus and Web of Science databases identified 70 publications; after removing duplicates and applying inclusion and exclusion criteria, the final sample comprised 11 studies with thematic relevance and adequate methodological quality. Based on this selection, the most commonly used CNN architectures were examined and their performance was compared, mainly in terms of accuracy. The findings indicate the integration of these architectures with proper development. Furthermore, signal preprocessing, particularly noise filtering and normalization, improves the stability and consistency of the collected data. Despite these advances, relevant limitations persist, including the lack of hardware and evaluation metric standardization, the limited availability of public datasets for specific sign languages, and the high implementation costs. These factors reduce reproducibility and hinder application in real-world contexts. Consequently, the need for standardized datasets, common evaluation protocols, and low-cost, high-efficiency solutions is emphasized to promote practical deployment and communicative inclusion

Keywords– Translator glove, CNN, flexible sensors, gesture recognition.

Sistema guante-traductor para el lenguaje de señas basado en redes neuronales convolucionales (CNN): Una revisión sistemática de literatura

Mariajose Guadalupe Pedroza Garcia¹; Angelina Nereyda Silva Agape²; Evelyn Elizabeth Ayala Niquen³
^{1,2,3} Universidad Tecnológica del Perú, Perú, U22212450@utp.edu.pe, U22244082@utp.edu.pe, c26915@utp.edu.pe

Resumen— En esta revisión sistemática de la literatura (RSL) se evaluaron los enfoques de diseño, la eficacia técnica y las brechas existentes en la implementación de sistemas guante-traductor para el lenguaje de señas que utilizan Redes Neuronales Convolucionales (CNN). Se utilizó la metodología PRISMA para organizar y documentar el proceso de selección y evaluación de estudios. La exploración inicial de las bases de datos Scopus y Web of Science permitió localizar 70 publicaciones, y tras la depuración de duplicados y la aplicación de criterios de inclusión y exclusión, la muestra final quedó conformada por 11 estudios con pertinencia temática y adecuada calidad metodológica. A partir de esta selección, se realizó el examen de las arquitecturas de CNN de mayor uso y la comparación de su desempeño, principalmente en términos de exactitud (accuracy). Los hallazgos muestran que la integración de estas arquitecturas con sensores flexibles e inerciales correctamente configurados posibilita precisiones superiores al 95% en entono controlado, lo que respalda su viabilidad técnica para el desarrollo de prototipos. Asimismo, el preprocesamiento de señales, especialmente el filtrado de ruido y la normalización, aporta estabilidad y consistencia a los datos obtenidos. Pese a estos avances, se mantiene limitaciones relevantes, tales como la falta de estandarización en el hardware y en las métricas de evaluación, la escasa disponibilidad de bases de datos públicas para lenguas de señas específicas y los costos elevados de implementación. Estos factores reducen la reproducibilidad y dificultan su aplicación en contextos reales. En consecuencia, se destaca la necesidad de contar con datasets estandarizadas, protocolos de evaluación comunes y soluciones de bajo costo y alta eficiencia que favorezcan la aplicación práctica y la inclusión comunicativa de estos sistemas.

Palabras clave—Guante traductor, CNN, sensores flexibles, reconocimiento de gestos.

I. INTRODUCCIÓN

En los últimos años, las tecnologías vestibles (wearables) orientadas a la traducción automática de la lengua de señas han adquirido relevancia como alternativa para reducir las barreras de comunicación que afectan a las comunidades sordas [1]. Entre estas soluciones, los guantes sensoriales equipados con sensores de flexión e inerciales, acoplados a algoritmos de aprendizaje automático, entre ellos las Redes Neuronales Convolucionales (CNN), han mostrado un desempeño favorable en la identificación de señas a partir de gestos manuales tanto estáticos como dinámicos, especialmente en escenarios de prueba controlados [2], [3]. Estos sistemas capturan datos cinemáticos de la mano y los dedos, por lo que permite identificar los lenguajes de señas y los transforma en texto o voz en tiempo real, lo que representa un avance

significativo hacia una comunicación más inclusiva [4]. Es así que, la comunidad sorda enfrenta profundas desigualdades [5], debido a la escasa investigación en torno al desarrollo de soluciones tecnológicas enfocados en el Lenguaje de Señas (LS). En ese sentido, hay prototipos desarrollados para lenguas como la árabe [1], polaca [3][2] y mexicana, no obstante, persiste un vacío de conocimiento en cuanto a soluciones validadas, accesibles y pertinentes, así como la falta de una metodología estandarizada y de estudios que aborden explícitamente el equilibrio entre precisión, costo, eficiencia computacional y adaptación lingüística específica constituye la principal brecha que impide la transición efectiva del prototipo a la práctica.

A pesar de los avances tecnológicos reportados globalmente, el estudio de estas soluciones aun presenta limitaciones relevantes y cierta falta de consenso, debido a la marcada heterogeneidad en las configuraciones de hardware (tipo y número de sensores), los algoritmos de clasificación empleados (SVM, KNN, CNN, LSTM) y las métricas de evaluación utilizadas. Esta variabilidad metodológica impide establecer comparaciones directas entre resultados y dificulta la determinación de prácticas adecuadas [6], [7]. Algunos estudios reportan precisiones superiores al 95% [1], [3], [8], pero estas métricas suelen derivar de pruebas de laboratorio con conjuntos de datos limitados y configuraciones de alto costo, lo que cuestiona su escalabilidad, replicabilidad y aplicabilidad en contextos reales con recursos limitados [9].

En este contexto, la presente RSL se sustenta en la necesidad de integrar y examinar de manera crítica la evidencia disponible sobre guantes traductores y sistemas sensoriales basados en CNN, con el fin de consolidar un marco técnico de referencia que oriente el diseño de soluciones viables orientadas a el reconocimiento automático de la lengua de señas. Asimismo, el estudio surge del interés por evaluar configuraciones optimizadas de bajo costo que favorezcan su implementación práctica. A partir de los hallazgos obtenidos, se efectuó la identificación de las arquitecturas de sensores y modelos algorítmicos de mayor eficiencia, junto con la delimitación de parámetros clave de desempeño (precisión, latencia, costo) y la caracterización de limitaciones metodológicas recurrentes. De esta manera, el trabajo se plantea como una guía técnica para futuras investigaciones y desarrollos de ingeniería enfocados en la construcción de prototipos accesibles y adaptados a las particularidades del LS.

El objetivo de esta RSL es analizar la evidencia científica disponible sobre la arquitectura y el rendimiento de guantes traductores de lengua de señas que utilicen sensores flexibles/inerciales y Redes Neuronales Convolucionales (CNN) para la clasificación de gestos, con el fin de identificar tendencias, efectividad comparativa y brechas críticas que informen el desarrollo de un prototipo optimizado y de bajo costo.

La revisión se organiza en diferentes secciones. En una primera etapa, se detalla la estrategia de búsqueda sistemática y el procedimiento de selección de estudios conforme a la metodología PRISMA. Posteriormente, se examinan los componentes técnicos esenciales de los sistemas guante-traductor, entre ellos las técnicas de preprocesamiento de señales más empleadas, las arquitecturas predominantes de Redes Neuronales Convolucionales (CNN) y las métricas de evaluación de rendimiento reportadas. Por último, se presentan las principales limitaciones y brechas identificadas en la literatura, así como recomendaciones orientadas al desarrollo de soluciones y a futuras líneas investigaciones. El documento concluye con una síntesis de los hallazgos más relevantes.

II. METODOLOGÍA

A. Estrategia de búsqueda

La presente revisión sistemática se centró en el análisis de la literatura relacionada con sistemas de traducción de lengua de señas basados en guantes sensoriales y Redes Neuronales Convolucionales (CNN). El interés principal correspondió a estudios orientados al reconocimiento del alfabeto dactilológico mediante sensores flexibles e inerciales integrados con modelos de aprendizaje profundo. En consecuencia, se consideraron trabajos que emplearan arquitecturas CNN para el procesamiento y clasificación de señales generadas por dispositivos vestibles aplicados a diferentes lenguas de señas.

El proceso de búsqueda y selección se desarrolló conforme a las directrices de la metodología PRISMA y complemento con el esquema PICO, utilizado para delimitar el problema de investigación, las tecnologías analizadas, los criterios de comparación y los resultados de interés. Este marco metodológico permitió establecer una pregunta de investigación concreta y definir criterios consistentes para la inclusión de estudios, con énfasis en la evaluación del desempeño de las CNN en términos de precisión, robustez y viabilidad de implementación.

Las búsquedas se realizaron en bases de datos científicas indexadas de alto impacto, principalmente Scopus y Web of Science. Se incluyeron artículos que describieran experimentos con guantes instrumentados, sensores de flexión, unidades de medición inercial y modelos de Redes Neuronales Convolucionales (CNN) aplicados al reconocimiento de gestos. Asimismo, se dio prioridad a las investigaciones con validación experimental, métricas cuantitativas de desempeño y descripciones claras de las arquitecturas empleadas.

La estrategia de búsqueda se estructuró según los componentes del modelo PICO: Problema (P), Intervención (I),

Comparación (C) y Resultado (O). Esta estructura facilita la identificación, clasificación y selección de estudios relevantes para la revisión.

TABLA I
COMPONENTES DE LA ESTRATEGIA PIOC

Acrónimo	Componente
(P)roblema	Reconocimiento del alfabeto de la lengua de señas
(I)ntervención	Redes neuronales convolucionales (CNN)
(O)utcome/Salida	Precisión y desempeño en la identificación de gestos.
(C)ontexto	Guantes sensoriales de traducción

Con base a estos componentes, se formula la siguiente pregunta de investigación: ¿Cómo influyen las arquitecturas de redes neuronales convolucionales (I) en el reconocimiento del alfabeto del lenguaje de señas (P) en términos de precisión y desempeño en la identificación de gestos (O) mediante guantes sensoriales de traducción (C)?

Dado que la pregunta PICO abarca distintos aspectos del problema de investigación, se definieron las siguientes subpreguntas para analizar cada componente de manera específica:

TABLA II
SUBPREGUNTAS PIOC

Acrónimo	RQ
P	¿Qué limitaciones de implementación presentan los sistemas de guantes sensoriales para el reconocimiento del alfabeto de la lengua de señas?
I	¿Qué arquitecturas de CNN se emplean con mayor frecuencia para la clasificación de gestos capturados por guantes sensoriales?
O	¿Qué métricas se utilizan para evaluar el desempeño de las CNN en el reconocimiento de gestos?
C	¿Cuáles son los factores y qué configuraciones de hardware presentan los sistemas de guantes sensoriales para el reconocimiento del alfabeto de la lengua de señas?

A partir de las subpreguntas PICO formuladas, se identificaron las siguientes palabras clave que guiaron la búsqueda y selección en las bases de datos científicas:

TABLA III
PALABRAS CLAVE DE LOS COMPONENTES DE LA PREGUNTA PICO

Acrónimo	Palabras clave
P	“sign language”, “gestual language”, “manual language”, “non-verbal communication”
I	“CNN”, “convolutional neural network”
O	“accuracy”, “recovery”, “average latency”, “ms”, “F1 score”, “F1-score”
C	“Translator glove”, “glove”, “wearable”

B. Ecuación de búsqueda

A partir de las palabras claves definidas, se construyeron las ecuaciones de búsqueda con el fin de recuperar estudios relevantes relacionados con el reconocimiento de la lengua de señas mediante guantes sensoriales y redes neuronales convolucionales. Las búsquedas se realizaron en las bases de datos Scopus y Web of Science, mediante el uso de operadores booleanos (AND, OR) para combinar sinónimos y términos

equivalentes, y se restringieron los resultados a artículos científicos. La búsqueda se realizó el 11 de enero de 2026.

TABLA IV
 ECUACIÓN DE BÚSQUEDA

Scopus	Web of Science
(TITLE-ABS-KEY ("sign language" OR "gestural language" OR "manual language" OR "non-verbal communication") AND TITLE-ABS-KEY ("CNN" OR "convolutional neural network") AND TITLE-ABS-KEY ("Translator glove" OR "glove" OR "wearable") AND TITLE-ABS-KEY ("Accuracy" OR "recovery" OR "average latency" OR "ms" OR "F1 score" OR "F1-score")) AND (LIMIT-TO (DOCTYPE , "ar"))	"sign language" OR "gestural language" OR "manual language" OR "non-verbal communication" (All Fields) and "CNN" OR "convolutional neural network" (All Fields) and "Translator glove" OR "glove" OR "wearable" (All Fields) and "Accuracy" OR "recovery" OR "average latency" OR "ms" OR "F1 score" OR "F1-score" (All Fields) and Article (Document Types)

C. Criterios de inclusión y exclusión

Con el fin de garantizar la relevancia y calidad de los estudios seleccionados, se establecieron criterios de inclusión y exclusión. Estos criterios se diseñaron el objetivo de identificar los trabajos que emplearan redes convolucionales con guantes sensoriales para el reconocimiento de lengua de señas y garantizar la coherencia en el análisis de resultados.

TABLA V
 CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

Criterios de Inclusión	Criterios de Exclusión
CI1: Artículos científicos de acceso abierto (Open Access). CI2: Estudios que empleen redes neuronales convolucionales (CNN) CI3: Sistemas basados en guantes sensoriales para reconocimiento de lengua de señas.	CE1: Trabajos que utilicen modelos distintos a CNN. CE2: Enfoques que no empleen dispositivos tipo guante. CE3: Artículos sin validación experimental.

D. Proceso de selección de estudios

Para la conducción de la revisión se adopto el marco PRISMA, reconocido por proporcionar criterios sistemáticos para la identificación, depuración y evaluación de estudios en revisiones de literatura. durante la búsqueda inicial se localizaron un total de 40 artículos en Scopus y 30 en Web of Sciente, lo que dio lugar a un total de 70 documentos potencialmente relevantes al tema de investigación.

En la primera etapa, se identificaron y eliminaron 23 documentos duplicados, lo que resultó en 47 registros para su evaluación.

Posteriormente, se examinó el título y resumen de los 47 artículos restantes para determinar su pertinencia con la pregunta de investigación. Durante este proceso se excluyeron 27 registros que no cumplían con los criterios de elegibilidad temática lo que redujo así el número de estudios a 20, los cuales fueron evaluados en su totalidad.

De las 20 publicaciones evaluadas, todas fueron recuperadas con éxito para su evaluación final. Acto seguido, se aplicaron los criterios de exclusión, lo que resultó en 11

estudios. El proceso completo de selección y filtrado se describe en detalle en la figura 1.

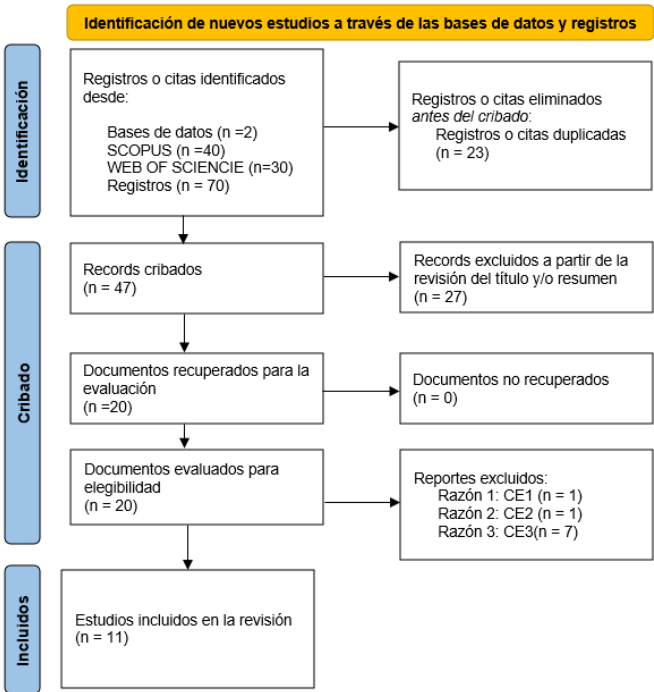


Fig. 1 DIAGRAMA DE ESTUDIO PRISMA

III. RESULTADOS

Esta sección expone los resultados derivados del análisis de los estudios seleccionados. La información se organiza en cinco bloques. Se inicia con una caracterización general de los artículos revisados, seguida de las limitaciones técnicas identificadas en los sistemas de guantes sensoriales, las arquitecturas de redes neuronales convolucionales empleadas para la clasificación de gestos, las métricas de evaluación del desempeño de los modelos y, finalmente, las configuraciones de hardware reportadas en los sistemas desarrollados. Esta organización permite ofrecer una visión sistemática y coherente de los elementos clave identificados en la literatura.

A. Perfil bibliométrico de los estudios seleccionados

Con el fin de contextualizar la literatura analizada, se ha elaborado una tabla bibliométrica (Tabla VI), que resume los aspectos claves de los once artículos incluidos. Esta va a presentar información como los autores, el título del artículo, la revista donde fue publicada, el país de origen, el año de publicación y la cantidad de citas registradas hasta la fecha. Se observa que la mayoría de los trabajos fueron publicados en el año 2021 y 2024, el país con mayor representación en la producción científica revisada es China. Las citas acumuladas varían considerablemente, destacándose algunos estudios de alto impacto en la comunidad académica, como el artículo de

Don Y. [10] con 95 citas, mientras que otras publicaciones recientes presentan una visibilidad e impacto aún en desarrollo.

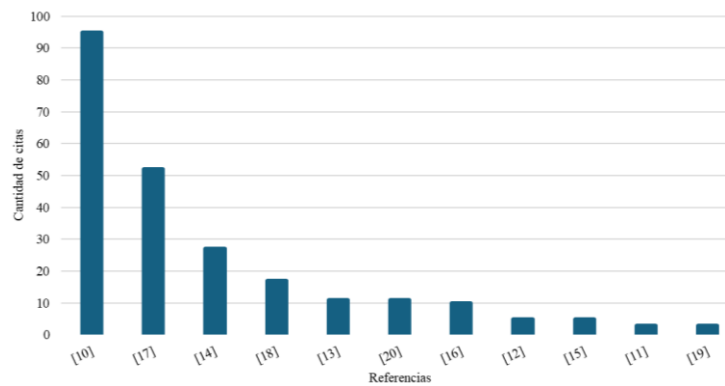


Fig. 2 CANTIDAD DE CITAS POR ESTUDIO INCLUIDO EN LA REVISIÓN

TABLA VI
DATOS BIBLIOMÉTRICOS DE LOS ESTUDIOS SELECCIONADOS

Ref.	Autores	Título	Revista	País	Año	Nº Citas
[11]	Begum, H.; Chowdhury, O.; Hridoy, M.S.R.; Islam, M.M.	AI-Based Sensory Glove System to Recognize Bengali Sign Language (BaSL)	IEEE Access	Bangladesh	2024	3
[12]	Wajahat, M.; Kouzani, A.Z.; Khoo, S.Y.; Mahmud, M.A.P.	Development of AI-Enabled Sign Language Predicting Glove Using 3-D Printed Triboelectric Sensors	IEEE Journal on Flexible Electronics	Australia	2024	5
[10]	Dong, Y.; Liu, J.; Yan, W.	Dynamic Hand Gesture Recognition Based on Signals from Specialized Data Glove and Deep Learning Algorithms	IEEE Transactions on Instrumentation and Measurement	China	2021	95
[13]	Rwelli, R.E.; Shahin, O.R.; Taloba, A.I.	Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network	International Journal of Advanced Computer Science and Applications	Arabia Saudita	2021	11
[14]	Ayodele, E.; Bao, T.Z.; Zaidi, S.A.R.; Hayajneh, A.M.A.; Scott, J.; Zhang, Z.Q.; McLemon, D.	Grasp Classification With Weft Knit Data Glove Using a Convolutional Neural Network	IEEE SENSORS JOURNAL	Inglaterra	2021	27
[15]	Liang, Y.; Jettanasen, C.; Chiradeja, P.	Progression Learning Convolution Neural Model-Based Sign Language Recognition Using Wearable Glove Devices	Computation	Tailandia	2024	5
[16]	Zhang, Y.; Xu, W.; Zhang, X.; Li, L.	Sign annotation generation to alphabets via integrating visual data with somatosensory data from flexible strain sensor-based data glove	Measurement: Journal of the International Measurement Confederation	China	2022	10
[17]	Saggio, G.; Cavallo, P.; Ricci, M.; Errico, V.; Zea, J.; Benalcázar, M.	Sign language recognition using wearable electronics: Implementing K-nearest neighbors with dynamic time warping and convolutional neural network algorithms	Sensors (Switzerland)	Italia	2020	52
[18]	Lu, C.; Kozakai, M.; Jing, L.	Sign Language Recognition with Multimodal Sensors and Deep Learning Methods	Electronics (Switzerland)	Japón	2023	17
[19]	Sonugur, G.; Cayli, A.	Turkish sign language recognition using fuzzy logic asisted ELM and CNN methods	Journal of Intelligent and Fuzzy Systems	Turkia	2023	3
[20]	Wang, D.; Wang, M.; Zhang, Z.; Liu, T.; Meng, C.; Guo, S.	Wearable Electronic Glove and Multilayer Para-LSTM-CNN-Based Method for Sign Language Recognition	IEEE Internet of Things Journal	China	2024	11

B. Limitaciones de los sistemas de guantes sensoriales.

Los sistemas de guantes sensoriales orientados al reconocimiento del alfabeto de la lengua de señas evidencian diversas limitaciones que repercuten en su precisión y confiabilidad. Diversos estudios señalan problemas asociados a la exactitud de los sensores, la necesidad de calibraciones frecuentes y la dificultad para registrar movimientos finos de los dedos. Estas condiciones afectan la diferenciación de gestos similares dentro deficiencias del

alfabeto manual y, en consecuencia, disminuyen el desempeño del reconocimiento automático [10], [15], [16], [18].

Asimismo, se observan obstáculos vinculados a la implementación práctica, entre ellos el costo del hardware, la complejidad del sistema y los retos de integración en entornos reales de uso. tales factores limitan la escalabilidad de los guantes sensoriales y restringen su adopción en contextos

educativas o de asistencia cotidiana para personas usuarias de la lengua de señas [13], [17].

Otra limitación es la dependencia del usuario, ya que el desempeño del sistema varía según características individuales como la forma de la mano, la velocidad del gesto y el nivel de entrenamiento. Además, varios estudios reportan un vocabulario reducido, enfocado principalmente en letras aisladas del alfabeto, lo que restringe su aplicabilidad en contextos de comunicación continua y natural [11], [14], [19], [10].

La Fig. 3 muestra la distribución de artículos que abordan las principales limitaciones de los sistemas de guantes sensoriales, y destaca a las limitaciones técnicas como las más frecuentes, seguidas por las barreras de implementación, la dependencia del usuario, el vocabulario reducido y los problemas de ergonomía [12].

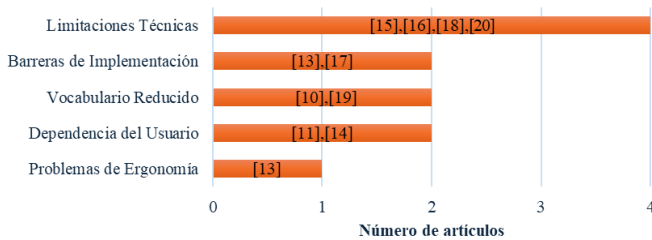


Fig. 3 DISTRIBUCIÓN DE LAS PRINCIPALES LIMITACIONES DE LOS SISTEMAS DE GUANTES SENSORIALES

C. Arquitecturas de CNN empleadas en la clasificación.

La literatura reporta el uso de diversas arquitecturas de Redes Neuronales Convolucionales (CNN) para la clasificación de gestos capturados mediante guantes sensoriales. Una de las configuraciones más frecuentes corresponde a modelos CNN 1D personalizados, desarrollados para el procesamiento de señales temporales generadas por sensores flexibles e inerciales. Este tipo de diseño facilita la extracción de características relevantes a partir de las secuencias de datos, lo que favorece el reconocimiento del alfabeto manual y disminuye la necesidad de etapas extensas de preprocesamiento [11], [14], [18], [19].

Asimismo, se registran arquitecturas más profundas o especializadas, con capas adicionales y esquemas de optimización ajustados a la naturaleza multicanal de las señales sensoriales. Estas configuraciones incrementan la capacidad de generalización del modelo y mejoran la discriminación entre gestos similares, lo que se traduce en mayores niveles de precisión en escenarios controlados y semiestructurados [13], [15], [16].

En contraste, algunos estudios mantienen el uso de modelos CNN convencionales, caracterizados por estructuras más simples y menor complejidad de implementación. Aunque esta alternativa reduce el costo computacional, el desempeño puede verse afectado ante gestos complejos o variaciones significativas entre usuarios [10], [17]. Asimismo, se describen propuestas híbridas que integran CNN con otros enfoques de aprendizaje profundo para

abarcar tanto información espacial como temporal, con mejoras reportadas en la exactitud del reconocimiento [20], [12].

La Fig. 4 muestra la distribución de artículos según las arquitecturas de CNN empleadas, descritas anteriormente, en la clasificación de gestos mediante guantes sensoriales, destaca una mayor presencia de CNN 1D personalizadas, seguidas por arquitecturas avanzadas, modelos estándar e híbridos.

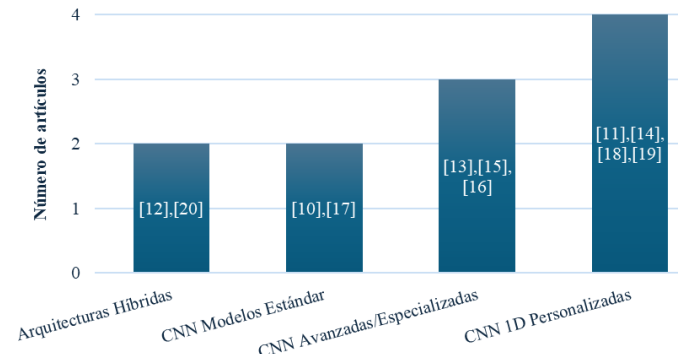


Fig. 4 ARQUITECTURAS CNN MÁS EMPLEADAS

D. Métricas de evaluación del desempeño.

La evaluación del desempeño de las CNN empleadas en el reconocimiento de gestos mediante guantes sensoriales se basa principalmente en métricas cuantitativas que permiten medir la eficacia del proceso de clasificación. La métrica más utilizada es la exactitud (accuracy), debido a su simplicidad y capacidad para reflejar el porcentaje global de gestos correctamente reconocidos. Los estudios analizados reportan valores elevados de accuracy, lo que evidencia el alto potencial de las CNN en este tipo de aplicaciones [10], [14], [16], [17], [19].

Entre los resultados más destacados, sobresalen tres arquitecturas que alcanzan los mayores porcentajes de exactitud: la ResNet-18 modificada, con un accuracy de 99.1% [16]; la CNN 3D multimodal, con 98.9% [19]; y la CNN 1D personalizada, con 98.7% [10]. Estos valores reflejan la capacidad de las arquitecturas profundas y multimodales para capturar patrones complejos en las señales sensoriales y diferenciar con mayor precisión los gestos del alfabeto manual.

En conjunto, los resultados muestran que arquitecturas CNN profundas, híbridas y especializadas alcanzan valores de accuracy superiores al 94%, lo que confirma la efectividad de estas métricas para comparar el desempeño de diferentes enfoques y seleccionar configuraciones adecuadas según los requerimientos del sistema [12], [13], [15].

Además de la exactitud, algunos trabajos incorporan métricas complementarias como precisión, sensibilidad (recall) y F1-score, con el fin de evaluar de manera más integral el rendimiento del modelo, especialmente en escenarios con gestos similares o clases desbalanceadas. Estas métricas permiten analizar la estabilidad del sistema y

su capacidad para minimizar errores de clasificación específicos [20], [18], [11].

TABLA VII
COMPARACIÓN DE ARQUITECTURAS Y MÉTRICA REPORTADA

Ref.	Arquitectura CNN Principal	Accuracy (%)
[16]	ResNet-18 modificada	99.1%
[19]	CNN 3D Multi-modal	98.9%
[10]	CNN 1D personalizada	98.7%
[14]	CNN profunda (7 capas)	98.5%
[17]	CNN 1D profunda	98%
[20]	Edge-Cloud CNN Híbrida	97.6%
[18]	CNN-LSTM Híbrida	97.3%
[11]	CNN 1D Multibranch con Atención	96.8%
[15]	CNN 2D (VGG-like)	96.2%
[12]	U-Net modificada	95.4%
[13]	CNN 1D simple	94.8%

E. Factores y configuración del hardware

Los sistemas de guantes sensoriales para el reconocimiento del alfabeto de la lengua de señas consideran diversos factores y configuraciones. Varios trabajos realizan un análisis por clase, para poder evaluar el desempeño del sistema para cada gesto individual del alfabeto, lo que permite identificar fortalezas y debilidades específicas del hardware empleado [10], [11], [12], [17], [19].

Otra tendencia importante se centra en la robustez y generalización del sistema, mediante configuraciones de hardware capaces de mantener un rendimiento estable frente a variaciones entre usuarios y condiciones de uso. Estos enfoques suelen integrar múltiples sensores y estrategias de adquisición para mejorar la consistencia de los datos [14], [16], [18].

Asimismo, se han propuesto configuraciones que incorporan métricas especializadas, las cuales permiten evaluar de manera más detallada el desempeño del sistema según el tipo de gesto o sensor utilizado [13], [15]. Además, algunos estudios priorizan la eficiencia computacional, para poder reducir el consumo energético y la latencia del sistema, especialmente en aplicaciones portátiles o en tiempo real [20].

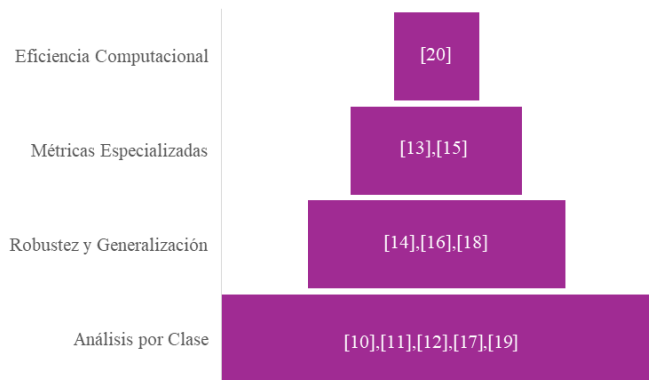


Fig. 5 FACTORES EN EL DISEÑO DE LOS GUANTES SENSORIALES

En cuanto a la implementación física, se observa una amplia variedad de plataformas de hardware. Los sistemas basados en Arduino con comunicación Bluetooth son ampliamente utilizados debido a su bajo costo y facilidad de integración, lo que los hace adecuados para prototipos y aplicaciones educativas [11], [19], [13]. Por otro lado, algunos trabajos incorporan características especiales, como sensores de alta resolución o módulos de procesamiento dedicados, para mejorar la calidad de la señal adquirida [10], [15], [16].

Finalmente, se identifican configuraciones más avanzadas, como los sistemas IoT basados en ESP32, que permiten conectividad inalámbrica y procesamiento distribuido [20], [18], así como plataformas con procesamiento avanzado, orientadas a aplicaciones de mayor complejidad computacional [12], [14]. También se reportan sistemas de adquisición de alta precisión, diseñados para capturar movimientos finos del alfabeto manual con mayor fidelidad, y así mejorar la calidad del reconocimiento [17].

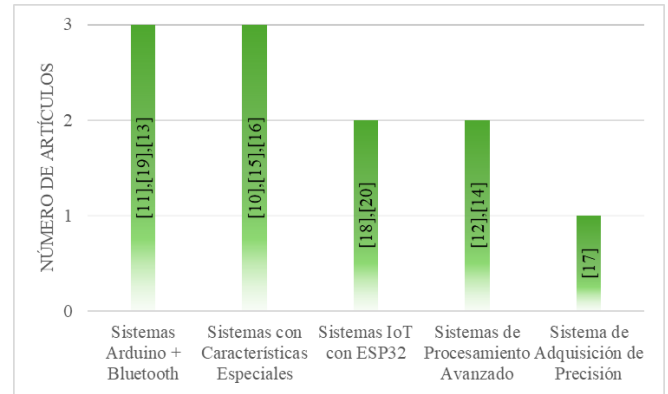


Fig. 6 HARDWARE MÁS EMPLEADO EN LOS EN SISTEMAS DE GUANTES SENSORIALES

IV. DISCUSIÓN

El análisis comparativo de los estudios incluidos en esta revisión sistemática evidencia que las redes neuronales convolucionales (CNN) constituyen el enfoque predominante para el reconocimiento de gestos mediante guantes sensoriales, en concordancia con lo reportado en [21] y [22], donde indican que las arquitecturas basadas en este modelo alcanzan mayores tasas de reconocimiento que clasificadores tradicionales como Support Vector Machine (SVM), k-Nearest Neighbors (KNN) o Hidden Markov Model (HMM), especialmente al emplearse configuraciones profundas o esquemas híbridos.

En relación con los dispositivos, los resultados de esta RSL muestran una marcada dependencia de los datos para el desempeño de los modelos, en base a la configuración del hardware, particularmente al número y tipo de sensores integrados en los guantes. Esta observación es coherente con lo documentado en [21], donde se destaca que la combinación de sensores de flexión e IMU mejora la discriminación de

gestos complejos, aunque introduce mayor variabilidad interusuario y mayores requerimientos de calibración. De forma complementaria, el estudio [22] resalta la heterogeneidad en las fuentes de datos y en los protocolos de adquisición dado que limita la comparabilidad directa entre investigaciones y condiciona la capacidad de generalización de los modelos, aspecto que también se evidencia en los trabajos analizados en la presente revisión.

Desde el punto de vista metodológico, se identificó una diversidad significativa en las arquitecturas empleadas, entre las que se encuentran CNN puras y modelos híbridos CNN-LSTM orientados al reconocimiento de secuencias dinámicas. Este patrón concuerda con los hallazgos de [22], donde se señala que las arquitecturas híbridas permiten capturar dependencias temporales de largo alcance, mientras que las CNN se especializan en la extracción automática de características especiales. Asimismo, en [21] se reporta una tendencia similar en sistemas basados en guantes, en los que la integración de modelos secuenciales mejora el reconocimiento de gestos continuos frente a enfoques que consideran únicamente posturas estáticas.

En cuanto a las métricas de evaluación, los estudios revisados priorizan indicadores como accuracy, precisión, recall y F1-score, lo cual es consistente con las prácticas reportadas en [21] y [22]. No obstante, en concordancia con dichas revisiones, los estudios analizados evidencian un predominio de accuracy como métrica principal de evaluación; sin embargo, los estudios [10], [11], [17], [18] y [20] reportan adicionalmente indicadores asociados a estabilidad temporal, análisis de clase, sensibilidad y eficiencia computacional en la evaluación del desempeño. Este comportamiento coincide con la tendencia reportada en [21] y [22], donde se reconoce una transición gradual hacia evaluaciones más integrales, aunque con diferencias en la frecuencia de adopción y en los criterios empleados para reportarlas.

Respecto a los entornos experimentales, la mayoría de los estudios analizados fueron validados en condiciones controladas o simuladas, con el uso de conjuntos de datos propios o bases limitadas de gestos aislados. Esta situación coincide con lo señalado en [21], donde destacan que gran parte de los sistemas basados en guantes son evaluados en laboratorio, y con lo reportado en [22], donde se advierte que los altos niveles de desempeño reportados en escenarios controlados no necesariamente reflejan el comportamiento de los modelos en contextos reales de comunicación continua y con variabilidad significativa entre usuarios.

La confrontación de los hallazgos de esta RSL con los resultados reportados en [21] y [22] evidencian que el avance en sistemas de traducción mediante guantes sensoriales y CNN depende, de manera prioritaria, de la estandarización de protocolos de adquisición de datos, de la disponibilidad de datasets representativos y de la evaluación en entornos más cercanos a escenarios reales. Estas consideraciones permiten interpretar los resultados dentro del panorama actual del

reconocimiento de automático de la lengua de señas y, a su vez, plantean oportunidades para futuras investigaciones. En este sentido, en concordancia con [21] y [22], resulta relevante impulsar el desarrollo de conjuntos de datos públicos y estandarizados, particularmente para lenguas de señas distintas a ASL y CSL, ya que ello favorecería la validación cruzada de modelos, una comparación más sistemática entre investigaciones y el desarrollo de soluciones con mayor robustez y capacidad de generalización.

V. CONCLUSIONES

En síntesis, la revisión realizada muestra que las Redes Neuronales Convolucionales (CNN), junto con sus variables híbridas, son las arquitecturas más empleadas en sistemas de reconocimiento de gestos basados en guantes sensoriales. Su uso se relaciona con la capacidad para procesar tanto la información espacial como temporal, generada por los movimientos de la mano. Asimismo, se observó que el rendimiento del sistema esta influenciado por la cantidad y el tipo de sensores integrados, siendo frecuente la combinación de sensores de flexión e inerciales para mejorar la diferenciación de gestos dinámicos. La métrica de evaluación más reportada es la precisión (accuracy), lo que evidencia que la mayoría de los estudios priorizan pruebas en condiciones controladas.

Esta revisión aporta una integración organizada de los enfoques de hardware y de aprendizaje profundo utilizados en guantes traductores, permitiendo reconocer tendencias tecnológicas, arquitecturas recurrente y criterios de evaluación comunes. En conjunto, esta información ofrece una visión clara del estado actual del área y facilita la comparación entre propuestas.

No obstante, persisten algunas limitaciones relevantes. Entre ellas destacan la falta de conjuntos de datos públicos y estandarizados, la diversidad de protocolos experimentales y la predominancia de pruebas en laboratorio. Estas condiciones dificultan la comparación objetiva de resultados y reducen la posibilidad de trasladar los sistemas a escenarios reales de uso.

A partir de los hallazgos obtenidos, se establecen como prioridades para futuras investigaciones el desarrollo de conjuntos de datos representativos de distintas lenguas de señas, la optimización de arquitecturas CNN para su implementación en dispositivos portables de bajo costo y la validación de los sistemas en entornos reales con variabilidad interusuario. Estas acciones permitirán reducir las brechas identificadas y fortalecer el diseño de soluciones de traducción con mayor aplicabilidad práctica.

REFERENCIAS

- [1] M. Halabi and Y. Harkouss, "Real-time arabic sign language recognition system using sensory glove and machine learning," *Neural Comput. Appl.*, vol. 37, no. 9, pp. 6977–6993, Mar. 2025, doi: 10.1007/s00521-025-11010-1.
- [2] L. M. Valverde-Cedillo, G. I. Ochoa-Rodriguez, and C. V. Salazar-Hernandez, "Diseño de un dispositivo wearable para la traducción

- del lenguaje de señas mexicano,” *Científica*, vol. 25, no. 2, pp. 1–15, Jul. 2021, doi: 10.46842/ipn.cien.v25n2a10.
- [3] J. Piskozub and P. Strumillo, “A Sensory Glove With a Limited Number of Sensors for Recognition of the Finger Alphabet of Polish Sign Language,” *IEEE Access*, vol. 13, pp. 28408–28418, 2025, doi: 10.1109/ACCESS.2025.3540578.
- [4] instituto Nacional de Estadística e Informática, “Perú: Caracterización de las Condiciones de Vida de la Población con Discapacidad, 2022,” Lima, Mar. 2024. Accessed: Nov. 04, 2025. [Online]. Available: <https://www.gob.pe/institucion/inei/informes-publicaciones/5357753-peru-caracterizacion-de-las-condiciones-de-vida-de-la-poblacion-con-discapacidad-2022>
- [5] Consejo Nacional para la Integración de la Persona con Discapacidad, “Informe del proceso de consulta de la Política Nacional Multisectorial en Discapacidad para el Desarrollo al 2030,” Lima, May 2021. Accessed: Nov. 04, 2025. [Online]. Available: <https://www.gob.pe/institucion/conadis/informes-publicaciones/1925407-informe-del-proceso-de-consulta-de-la-politica-nacional-multisectorial-en-discapacidad-para-el-desarrollo-al-2030>
- [6] J. Delpreto, J. Hughes, M. D’Aria, M. De Fazio, and D. Rus, “A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10589–10596, Oct. 2022, doi: 10.1109/LRA.2022.3191232.
- [7] M. R. Cassim, J. Parry, A. Pantanowitz, and D. M. Rubin, “Design and construction of a cost-effective, portable sign language to speech translator,” *Inform. Med. Unlocked*, vol. 30, Jan. 2022, doi: 10.1016/j.imu.2022.100927.
- [8] A. I. M. Thaim, N. Sazali, K. Kadirgama, A. S. Jamaludin, F. M. Turan, and N. A. Razak, “Smart Glove for Sign Language Translation,” *Journal of Advanced Research in Applied Mechanics*, vol. 112, no. 1, pp. 80–87, Dec. 2023, doi: 10.37934/aram.112.1.8087.
- [9] S. G. L., V. B. Kumar P, T. M. R, and T. Ann Tharakan, “Enhancing Collaborative Interaction with the Augmentation of Sign Language for the Vocally Challenged,” 2023. [Online]. Available: www.ijacsa.thesai.org
- [10] Y. Dong, J. Liu, and W. Yan, “Dynamic Hand Gesture Recognition Based on Signals From Specialized Data Glove and Deep Learning Algorithms,” *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021, doi: 10.1109/TIM.2021.3077967.
- [11] H. Begum, O. Chowdhury, Md. S. R. Hridoy, and M. M. Islam, “AI-Based Sensory Glove System to Recognize Bengali Sign Language (BaSL),” *IEEE Access*, vol. 12, pp. 145003–145017, 2024, doi: 10.1109/ACCESS.2024.3472469.
- [12] M. Wajahat, A. Z. Kouzani, S. Y. Khoo, and M. A. P. Mahmud, “Development of AI-Enabled Sign Language Predicting Glove Using 3-D Printed Triboelectric Sensors,” *IEEE Journal on Flexible Electronics*, vol. 3, no. 6, pp. 266–273, Jun. 2024, doi: 10.1109/JFLEX.2024.3419078.
- [13] R. El Rwelli, O. R. Shahin, and A. I. Taloba, “Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021, doi: 10.14569/IJACSA.2021.0121273.
- [14] E. Ayodele *et al.*, “Grasp Classification With Weft Knit Data Glove Using a Convolutional Neural Network,” *IEEE Sens. J.*, vol. 21, no. 9, pp. 10824–10833, May 2021, doi: 10.1109/JSEN.2021.3059028.
- [15] Y. Liang, C. Jettanasen, and P. Chiradeja, “Progression Learning Convolution Neural Model-Based Sign Language Recognition Using Wearable Glove Devices,” *Computation*, vol. 12, no. 4, p. 72, Apr. 2024, doi: 10.3390/computation12040072.
- [16] Y. Zhang, W. Xu, X. Zhang, and L. Li, “Sign annotation generation to alphabets via integrating visual data with somatosensory data from flexible strain sensor-based data glove,” *Measurement*, vol. 202, p. 111700, Oct. 2022, doi: 10.1016/j.measurement.2022.111700.
- [17] G. Saggio, P. Cavallo, M. Ricci, V. Errico, J. Zea, and M. E. Benalcázar, “Sign Language Recognition Using Wearable Electronics: Implementing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network Algorithms,” *Sensors*, vol. 20, no. 14, p. 3879, Jul. 2020, doi: 10.3390/s20143879.
- [18] C. Lu, M. Kozakai, and L. Jing, “Sign Language Recognition with Multimodal Sensors and Deep Learning Methods,” *Electronics (Basel)*, vol. 12, no. 23, p. 4827, Nov. 2023, doi: 10.3390/electronics12234827.
- [19] G. Sonugür and A. Çaylı, “Turkish sign language recognition using fuzzy logic asisted ELM and CNN methods,” *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 5, pp. 8553–8565, Nov. 2023, doi: 10.3233/JIFS-231601.
- [20] D. Wang, M. Wang, Z. Zhang, T. Liu, C. Meng, and S. Guo, “Wearable Electronic Glove and Multilayer Para-LSTM-CNN-Based Method for Sign Language Recognition,” *IEEE Internet Things J.*, vol. 11, no. 24, pp. 40787–40799, Dec. 2024, doi: 10.1109/IJOT.2024.3454215.
- [21] Z. R. Saeed, Z. B. Zainol, B. B. Zaidan, and A. H. Alamoodi, “A Systematic Review on Systems-Based Sensory Gloves for Sign Language Pattern Recognition: An Update from 2017 to 2022,” 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3219430.
- [22] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, “Machine learning methods for sign language recognition: A critical review and analysis,” *Intelligent Systems with Applications*, vol. 12, p. 56, 2021, doi: 10.1016/j.iswa.2021.20.