DESDE 2023
https://rictrevista.org

RICT Revista de Investigación Científica, Tecnológica e Innovación

Publicación Semestral RICT Vol. 3 No. 6 (2025) P. 36-41

ISSN: 2992-7315

Modelo de visión artificial para el reconocimiento de la Lengua de Señas Mexicana
Machine Vision Model for Mexican Sign Language Recognition.

Derlis Hernández-Lara[a], Emmanuel-Tonatihu Juárez-Velázquez[b], Carlos-Alfonso Trejo-Villanueva[c]

[a] *Ingeniería en Tecnologías de la Información y Comunicaciones, Universidad Mexiquense del Bicentenario: UES Atenco*
[a, b, c] *Ingeniería informática, Tecnológico Nacional de México: Tecnológico de Estudios Superiores de Ecatepec*

Resumen

En el mundo existen millones de personas con discapacidad auditiva o del habla que enfrentan barreras comunicativas y sociales, lo que limita su plena integración. En respuesta a esta problemática, el presente trabajo propone un modelo de visión artificial para el reconocimiento de la Lengua de Señas Mexicana (LSM) basado en aprendizaje profundo, con el propósito de favorecer la inclusión social y educativa de personas con discapacidad auditiva. El estudio se estructura conforme al formato IMRaD (Introducción, Métodos, Resultados y Discusión), detallando los materiales, métodos, resultados y discusión. Se desarrolló un modelo de red neuronal convolucional entrenado con un conjunto de datos de 24 letras del alfabeto manual de la LSM, conformado por 12 000 imágenes procesadas en escala de grises (40×40 px). Se evaluaron métricas de precisión, recall, F1-Score y matriz de confusión, obteniendo una precisión global del 93 %. El trabajo incorpora además el enfoque de Design Thinking en la etapa de diseño centrado en el usuario, complementado con una metodología experimental de aprendizaje automático para la validación técnica. Los resultados muestran la viabilidad del modelo para aplicaciones educativas e interactivas, constituyendo una herramienta accesible para el reconocimiento automatizado de señas estáticas.

Palabras clave: Visión artificial, Aprendizaje profundo, Lengua de Señas Mexicana (LSM), Pensando en diseño, Inclusión digital

Abstract

Millions of people with hearing or speech disabilities worldwide face communication barriers that limit their full social participation. In response, this paper presents a computer vision model for the recognition of Mexican Sign Language (LSM) based on deep learning techniques, aimed at promoting social and educational inclusion for people with hearing impairments. The study follows the IMRaD structure and details the materials, methods, results, and discussion. A convolutional neural network was trained using a dataset of 24 manual alphabet letters from LSM, composed of 12,000 grayscale images (40×40 px). Evaluation metrics included accuracy, recall, F1-score, and confusion matrix, achieving an overall accuracy of 93 %. Design Thinking was applied in the user-centered design phase, complemented by a machine learning experimental methodology for technical validation. The results demonstrate the feasibility of the proposed model for educational and interactive applications, providing an accessible tool for automated static sign recognition.

Keywords: Machine vision, Deep Learning, Mexican Sign Language (MSL), Design Thinking, Digital inclusion

1. Introducción

De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI, 2020), en México hay 6,179,890 personas con algún tipo de discapacidad, lo que representa el 4.9 % de la población total. Dentro de este grupo, las personas con discapacidad auditiva o del habla constituyen un porcentaje relevante. En la Figura 1 se muestra el porcentaje de la población con discapacidad según el tipo de dificultad reportada.

*Autor para la correspondencia: dderlis-lara@tese.edu.mx
Correo electrónico: dderlis-lara@tese.edu.mx (Derlis Hernández-Lara), ejuarezv@hotmail.com (Emmanuel Tonatihu,Juárez Velázquez), carlostrejo@tese.edu.mx

Historial del manuscrito: recibido el 11/10/2025, última versión-revisada recibida el 23/10/2025, aceptado el 24/10/2025, publicado el 06/11/2025. **DOI:** https://doi.org/10.5281/zenodo.17546270

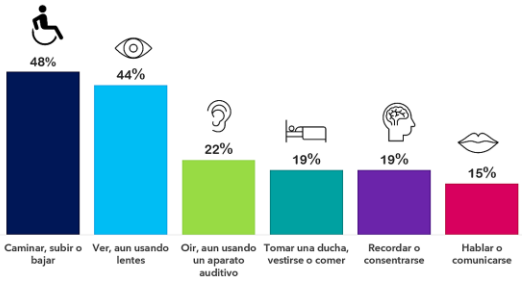

Figura 1: Población con discapacidad según la actividad (INEGI, 2020).

Estas limitaciones afectan la comunicación cotidiana y generan la necesidad de desarrollar herramientas tecnológicas que permitan la interacción entre personas sordas y oyentes. La Lengua de Señas Mexicana (LSM) constituye un medio esencial de comunicación visual-gestual que posibilita la expresión y comprensión sin depender del lenguaje oral (INDEPEDI, 2017).

El uso de la IA, específicamente de la Visión Artificial (VA) con RNA se ha extendido hasta el desarrollo de sistemas de identificación de gestos, particularmente en el desarrollo de herramientas tecnológicas que permiten el reconocimiento del lenguaje de señas expresado con las manos (González Riveros & Yimes Inostra, 2016: Santos D. at al., 2020).

En el trabajo presentado por Razo Gil (2009), se desarrolló un sistema clasificador de imágenes de postura de la mano correspondiente a las letras del alfabeto dactilológico que se representan sin movimiento, las cuales se digitalizan para identificar a que letra corresponde haciendo uso del código de cadena, que es una serie de valores numéricos que van del 0 al 7, estos representan direcciones cuantizadas de 45° conocidos como códigos 8-direccional.

El reconocimiento de gestos con las manos *Hand Gesture Recognition* (HGR, por sus siglas en inglés), es un método novedoso sin contacto, que implementa un sistema *Handwritten Character Recognition* (HCR) en tiempo real con Microsoft Visual Studio 2010, el cual fue abordado por Li (2012). HGR es un algoritmo que solo hace el reconocimiento de 9 señas, esta propuesta realiza la extracción de las manos utilizando el dispositivo *Kinect*, que tiene sensores de profundidad y de color para capturar imágenes en RGB y los datos de profundidad, lo cual le permite al sistema aplicar algoritmos que clasifiquen y realicen el reconocimiento de las características de la imagen.

El trabajo presentado por García Incertis et al., (2006) corresponde al reconocimiento de gestos con las manos de la Lengua de Signos Española (LSE), este reconocimiento de gestos se da mediante la identificación de la mano cubierta con un guante instrumental de color azul como se presenta en la Figura 2, lo que simplifica el problema de reconocimiento. El enfoque general funciona de la siguiente manera: primero, la región de la mano y los contornos correspondientes se extraen de la imagen a través de la segmentación mediante el formato de color HSV (*Hue Saturation Value*). Luego de la extracción, el contorno obtenido se muestrea y se calcula cada distancia de arco dada en una huella, y los puntos resultantes de dicho muestreo se comparan con los de un gesto objetivo en un diccionario. Esta comparación se realiza sobre cuatro criterios de distancia, de forma que, finalmente se consigue un reconocimiento adecuado, el cual obtiene un resultado del reconocimiento de 19 letras de la LSE.

Figura 2: Guante azul para el reconocimiento de la LSE (Garcia Incertis et al., 2006).

Abordando la problemática expuesta anteriormente, se ha propuesto implementar una herramienta tecnológica basada en un modelo de visión artificial para el reconocimiento de la LSM con la ayuda de Inteligencia Artificial (IA), específicamente el *Deep Learning* que es un área dentro del aprendizaje de máquinas basada en Redes Neuronales Artificiales (RNA) (López Saca, 2019). Para el desarrollo de la propuesta, se utiliza el siguiente proceso: 1) adquisición de una imagen que representa una postura de la mano, siendo una letra de la LSM, y 2) dicha imagen es procesada para analizarla en un modelo de RNA e identificarla con la clase de mayor probabilidad, dando como resultado la letra a la que pertenece. Así, con esta herramienta se pretende facilitar la comunicación y/o mejorar el entendimiento para las personas que requieran su uso de manera eficiente y eficaz, apoyando a la comunicación de las personas con este tipo de discapacidad.

El objetivo de este estudio es desarrollar e implementar un modelo de visión artificial capaz de reconocer las letras estáticas del alfabeto de la LSM a partir de imágenes capturadas por cámara, utilizando técnicas de aprendizaje profundo y una metodología reproducible. La hipótesis de trabajo sostiene que un modelo CNN adecuadamente entrenado, complementado con buenas prácticas de diseño centrado en el usuario, puede alcanzar niveles de precisión superiores al 90 % y servir de base para futuras aplicaciones de interpretación automática de señas en tiempo real.

2. Materiales y Método

Para la obtención del diseño conceptual, se utilizó la metodología *Design Thinking*, permitiendo así, la construcción de prototipos basados en las personas (Dinngo Lab, 2022). Esta metodología se empezó a desarrollar en la Universidad de Stanford en California en los años 70, y se compone de 5 etapas como se muestra en la Figura 3: 1) empatizar, 2) definir, 3) idear, 4) prototipar y 5) testear. No es lineal y se puede saltar a etapas no consecutivas. En los siguientes apartados se describe lo que se realizó a partir de estas etapas para desarrollar este trabajo.

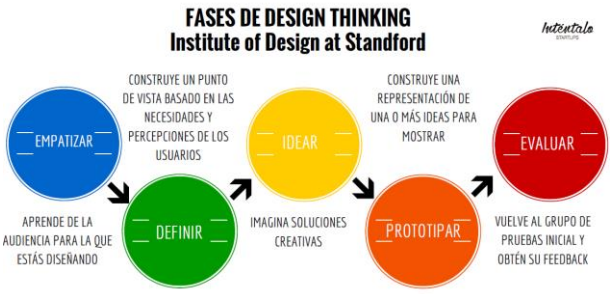

Figura 3: Metodología *Design Thinking* (Mejía López at al. 2019).

De forma general el funcionamiento del sistema es el presentado en la Figura 4: Primero se adquiere la imagen de la seña realizada por el usuario, para enseguida realizar un preprocesamiento de esta, y posteriormente su segmentación, siendo analizada por el modelo de *Deep Learning* propuesto, para obtener el reconocimiento de esta basándose en el diccionario de la LSM.

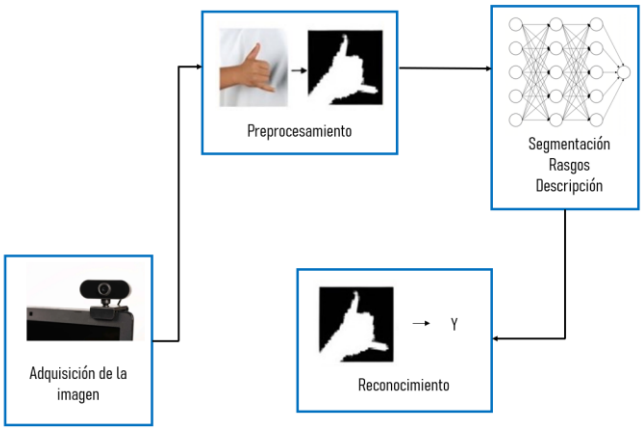

Figura 4: Ejemplo del preprocesamiento de imágenes (original, binarizada y normalizada a 40×40 px).

2.1. Adquisición del conjunto de datos

Se empleó un conjunto de datos denominado *SenasDataset*, compuesto por 12 000 imágenes correspondientes a 24 letras del alfabeto manual de la LSM (se excluyeron las letras J y K por implicar movimiento). Las imágenes fueron capturadas en diferentes condiciones de iluminación, con variabilidad en tono de piel, género y accesorios de las manos, utilizando una cámara Logitech C920 (1080 p).

Cada imagen se redimensionó a 40×40 px y se convirtió a escala de grises. Posteriormente, se aplicó un proceso de normalización de intensidad y segmentación de fondo. La información fue almacenada en un archivo CSV con 1600 columnas de píxeles y una columna adicional con la etiqueta correspondiente a cada letra (valores 0–23). La Figura 5 muestra la distribución del número de imágenes por clase en el *dataset*. Mientras que la Figura 6 el etiquetado de los datos.

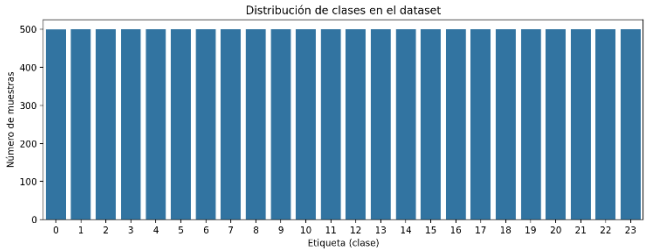

Figura 5: Distribución de clases en el *dataset* de señas.

Etiqueta	Letra	Etiqueta	Letra
0	a	**12**	o
1	b	**13**	p
2	c	**14**	q
3	d	**15**	r
4	e	**16**	s
5	f	**17**	t
6	g	**18**	u
7	h	**19**	v
8	i	**20**	w
9	l	**21**	x
10	m	**22**	y
11	n	**23**	z

Figura 6: Etiquetas asignadas a cada letra del alfabeto en este modelo.

2.2. Preprocesamiento y preparación de datos

Los valores de los píxeles fueron normalizados en el rango [0, 1] y posteriormente estandarizados mediante *StandardScaler* para homogenizar la varianza entre variables. El conjunto de datos se dividió en 80 % para entrenamiento y 20 % para prueba, garantizando balance mediante estratificación de clases.

Se implementaron técnicas de aumento de datos (*data augmentation*) en la etapa de entrenamiento para mejorar la generalización del modelo, aplicando transformaciones de rotación, volteo horizontal y cambios leves de brillo.

2.3. Arquitectura del modelo

El modelo base consistió en una Red Neuronal Convolucional (CNN) de cuatro capas, con la estructura que se presenta en la Tabla 1.

Tabla 1: Estructura de la Red Neuronal Convolucional (CNN)

Capa	Tipo	Tamaño de filtro / unidades	Activación
1	Convolucional + MaxPooling	32 filtros (3×3)	ReLU
2	Convolucional + MaxPooling	64 filtros (3×3)	ReLU
3	Convolucional + GlobalAveragePooling	—	—

4	Densa	128 unidades	ReLU
Salida	Densa	24 neuronas	Softmax

Se utilizó el optimizador Adam con tasa de aprendizaje de 1×10^{-3}, función de pérdida *categorical crossentropy*, y métricas de evaluación *accuracy*, *precision*, *recall* y *F1-Score*. El entrenamiento se realizó durante 50 épocas con tamaño de lote de 32, utilizando *EarlyStopping* y *ModelCheckpoint* para evitar sobreajuste.

El entorno de experimentación incluyó Python 3.11, *TensorFlow 2.12*, *scikit-learn 1.4*, y ejecución en GPU NVIDIA RTX 3060.

2.4. Validación y métricas

Se evaluó el modelo sobre el conjunto de prueba utilizando las siguientes métricas:

- Precisión global (*Accuracy*)
- Precisión y *recall* por clase
- Puntaje F1 macro-promedio
- Matriz de confusión
- Intervalos de confianza (95 %)

Para garantizar reproducibilidad, se utilizó una semilla aleatoria fija (*random_state = 42*).

Además, se implementó un modelo comparativo de referencia (*baseline*) mediante *Random Forest* para validar la discriminabilidad del *dataset*. Los resultados obtenidos con este modelo se presentan en la sección de Resultados.

3. Resultados

El modelo propuesto alcanzó una precisión global del 93.1 % en el conjunto de prueba, con un puntaje F1 macro-promedio de 0.92. La Figura 7 presenta ejemplos de las imágenes procesadas del *dataset* utilizadas en el entrenamiento.

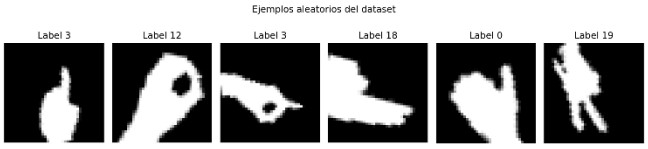

Figura 7: Ejemplos aleatorios del *dataset*.

En la Figura 8 se muestra la matriz de confusión obtenida con el modelo *Random Forest* de referencia, la cual permitió identificar patrones de confusión entre letras con configuraciones de mano similares (por ejemplo, las señas que representan a las letras E y S).

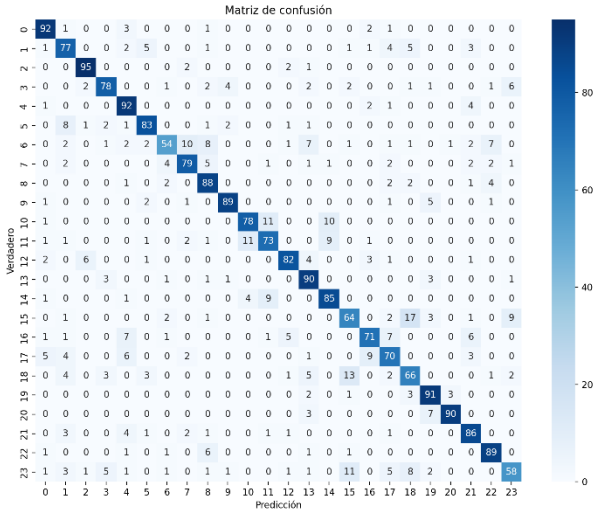

Figura 8: Matriz de confusión del modelo *Random Forest*.

El rendimiento del modelo CNN superó de manera consistente al clasificador *baseline*, demostrando la eficacia de la arquitectura convolucional para la extracción automática de características discriminantes. La Tabla 2 resume las métricas promedio obtenidas.

Tabla 2: Métricas promedio obtenidas

Modelo	**Precisión (%)**	**Recall (%)**	**F1-Score (%)**
Random Forest (baseline)	88.4	87.9	87.6
CNN propuesta	**93.1**	**92.8**	**92.0**

Durante el entrenamiento de la CNN (50 épocas), se implementó *EarlyStopping* con una paciencia de 5 épocas y *Dropout* del 50 %. Las curvas de evolución muestran una convergencia estable del modelo sin indicios de sobreajuste significativo, como se observa en las Figuras 9 y 10 respectivamente.

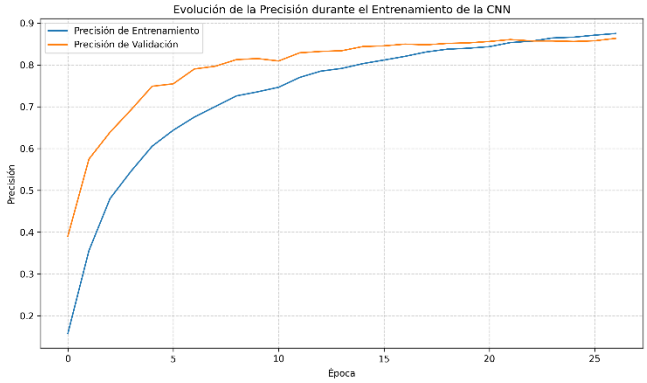

Figura 9: Evolución de la precisión durante el entrenamiento del modelo CNN.

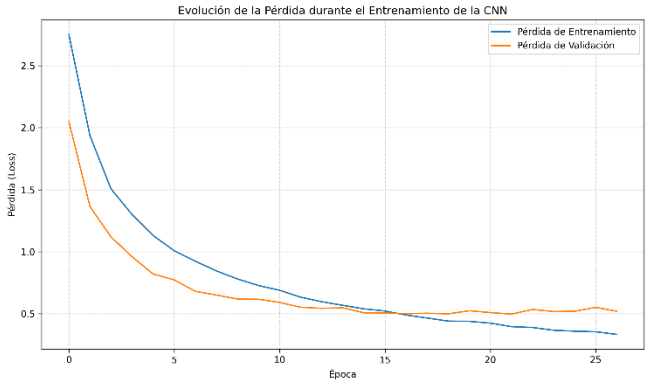

Figura 10: Evolución del error durante el entrenamiento del modelo CNN.

3.1 Evaluación e implementación del modelo

Además de las métricas cuantitativas previamente presentadas, se realizaron pruebas de validación funcional para ilustrar el desempeño del modelo en escenarios reales de reconocimiento de señas. En la Figura 11, se muestra un ejemplo del funcionamiento del modelo con una seña de entrada correspondiente a la letra «O». La imagen original fue procesada, redimensionada a 40×40 píxeles y binarizada antes de ser ingresada a la red neuronal. Posteriormente, el modelo realizó la predicción automática de la clase asociada, identificando correctamente la seña representada.

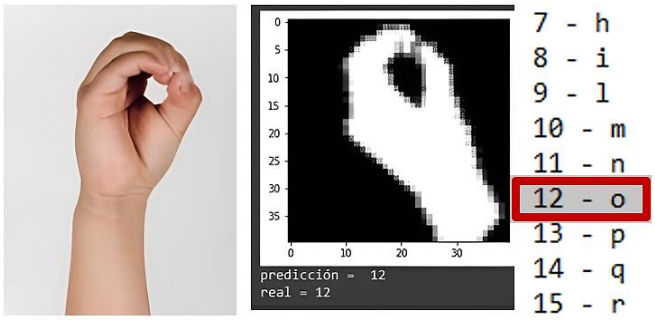

Figura 11: Ejemplo de implementación del modelo con la seña correspondiente a la letra «O» en la LSM.

De manera complementaria, la Figura 12 presenta una segunda prueba en la que un usuario ejecuta una seña distinta, en este caso, la letra «V», utilizando la mano derecha, siguiendo las condiciones de captura definidas en el protocolo experimental. La imagen capturada es preprocesada por el sistema, binarizada y normalizada antes de ingresar a la CNN, la cual emite como salida la etiqueta de clase con la mayor probabilidad. Este procedimiento evidencia la capacidad del modelo para generalizar correctamente a muestras nuevas y reconocer de forma precisa las letras del alfabeto dactilológico mexicano.

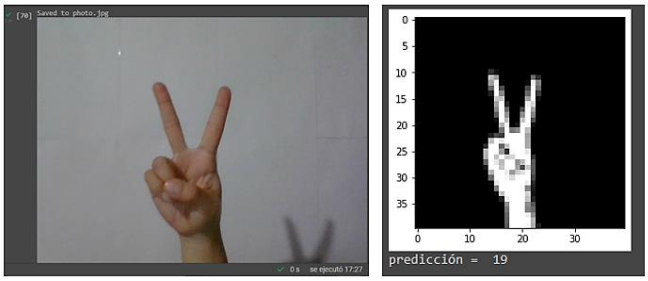

Figura 12: Ejemplo de reconocimiento final del modelo con la seña correspondiente a la letra «V» en la LSM.

4. Discusión

El rendimiento alcanzado por la CNN propuesta confirma la viabilidad del aprendizaje profundo para la interpretación automática de señas estáticas de la LSM. Comparando con trabajos recientes (Tabla 3), el modelo obtiene resultados competitivos y reproduce buenas prácticas metodológicas. La precisión lograda (93 %) es comparable con estudios recientes (Sharma et al., 2022; Salgado-Martínez et al., 2024), incluso con imágenes de baja resolución.

Tabla 3: Comparación con modelos recientes de reconocimiento de señas (2019–2024)

Autor(es) / Año	Arquitectura	Dataset	Nº clases	Precisión (%)
Sharma, C. M. et al. (2022)	EfficientNet + Transfer Learning	ISL-Alphabet	26	95
Oyedotun & Khashman (2017)	Deep CNN	ASL dataset	24	92.7
Bheda & Radpour (2017)	CNN (arXiv)	ASL RGB	24	94.5
Salgado-Martínez et al. (2024)	MobileNetV2 (Transfer Learning)	LSM (móvil)	24	91.2
Presente estudio (2025)	CNN simple (4 capas)	LSM (propio)	24	**93.1**

Fuentes: Sharma, C. M. et al. (2022); Oyedotun & Khashman (2017); Bheda & Radpour (2017); Salgado-Martínez et al. (2024); Presente estudio (2025).

Los resultados demuestran que, aun con una resolución modesta (40×40 px) y sin usar redes preentrenadas, el modelo propuesto logra una precisión equiparable a arquitecturas más complejas. Esto sugiere que la calidad del preprocesamiento y la representatividad del *dataset* son factores clave para un desempeño robusto.

Limitaciones del estudio

Entre las principales limitaciones se identifican:

1. Diversidad del *dataset*: aunque se procuró variabilidad en iluminación y tonos de piel, la muestra no cubre todas las condiciones posibles ni suficiente número de usuarios.

2. Resolución de entrada: el tamaño 40×40 px, aunque eficiente, puede eliminar rasgos finos de textura. En estudios futuros se propone usar RGB o *landmarks* de *MediaPipe* para conservar detalles espaciales.

Ausencia de validación con usuarios reales: el modelo aún no ha sido probado en contextos de interacción directa con personas sordas. Se planea una segunda fase con pruebas de usabilidad y percepción.

4.2. Lista de referencias

La lista de referencias debe ser ordenada alfabéticamente de acuerdo con el primer autor, con las siguientes líneas justificadas con la sangría correspondiente. Si existen diferentes publicaciones del mismo autor(es), éstas deberán ser listadas en el orden del año de publicación. Si hay más de un artículo del mismo autor en la misma fecha, etiquételas como a,b, etc. (Baker, 1963a, b). Por favor, fíjese que todas las referencias (García, 2007) en este apartado (García and Martínez, 2008) deben ser citadas directamente en el cuerpo del texto (García et al., 2007), (Dog, 1958), (Keohane, 1958).

Por favor, tenga en cuenta que las referencias al final de este documento cumplen con el estilo anteriormente mencionado. Los artículos que no hayan sido publicados deben ser citados como “no publicado.” Ponga en mayúscula únicamente la primera palabra del título, excepto el caso de nombres propios y símbolos de elementos.

Si está utilizando LaTeX, puede procesar una base de datos de bibliografía externa o insertarla directamente en la sección de referencias. Las notas al pie de página se deben evitar en la medida de lo posible.

5. Conclusiones

El modelo de visión artificial desarrollado demuestra la factibilidad de aplicar redes neuronales convolucionales al reconocimiento de la Lengua de Señas Mexicana, alcanzando una precisión del 93 %. La combinación de *Design Thinking* en la fase de diseño con una metodología experimental de aprendizaje profundo permitió integrar perspectivas humanas y tecnológicas en una solución inclusiva.

Entre las aportaciones destacadas se encuentran:

- Un *dataset* de la LSM documentado y balanceado.
- Una arquitectura CNN reproducible y de bajo costo computacional.
- Un protocolo metodológico transparente que favorece la replicabilidad y la transferencia tecnológica.

Futuras líneas de trabajo incluyen la incorporación de secuencias dinámicas para reconocer señas con movimiento, el uso de modelos preentrenados como *MobileNetV2* y *MediaPipe Holistic*, *ResNet50* y la integración del sistema en plataformas web accesibles.

Nota de reproducibilidad

Los *scripts* de entrenamiento y evaluación del modelo (incluido modelo_LSM_replicacion.py), así como el conjunto de datos utilizado, estarán disponibles bajo solicitud directa a los autores, con el fin de garantizar un uso ético y controlado del material.

6. Referencias

Bheda, V., & Radpour, D. (2017). Using deep convolutional networks for gesture recognition in American Sign Language. arXiv preprint arXiv:1710.06836. https://arxiv.org/abs/1710.06836

Dinngo Lab. (2022). Design Thinking en español. Recuperado de https://www.designthinking.es/inicio/index.php

García Incertis, I., Gómez García-Bermejo, J., & Zalama Casanova, E. (2006). Hand gesture recognition for deaf people interfacing. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR’06). IEEE. https://doi.org/10.1109/ICPR.2006.619

González Riveros, C. G., & Yimes Inostra, F. J. (2016). Sistema de reconocimiento gestual de lengua chilena mediante cámara digital. Tesis de pregrado, Pontificia Universidad Católica de Valparaíso.

Instituto Nacional de Estadística y Geografía (INEGI). (2020). La discapacidad en México (Cuéntame de México). https://cuentame.inegi.org.mx/explora/poblacion/discapacidad/

Instituto para la Atención e Inclusión de las Personas con Discapacidad de la Ciudad de México (INDEPEDI). (2017). Diccionario de Lengua de Señas Mexicana — Ciudad de México (DLSM CDMX). Gobierno de la Ciudad de México. https://www.indiscapacidad.cdmx.gob.mx/

Li, Y. (2012). Hand gesture recognition using Kinect (Master’s thesis). University of Louisville. https://ir.library.louisville.edu/etd/823

López Saca, F. (2019). Clasificación de imágenes usando redes neuronales convolucionales. Universidad Autónoma Metropolitana.

Mejía López, J. A., Ruiz Guzmán, O. A., Gaviria Ocampo, L. N., & Ruiz Guzmán, C. P. (2019). Aplicación de metodología design thinking en el desarrollo de cortadora automática CNC para MiPyME de confección. Revista UIS Ingenierías, 18(3), 157–168. https://doi.org/10.18273/revuin.v18n3-2019016

Oyedotun, O. K., & Khashman, A. (2017). Deep learning in vision-based static hand gesture recognition. Neural Computing and Applications, 28(12), 3941–3951. https://doi.org/10.1007/s00521-016-2294-8

Razo Gil, L. J. (2009). Sistema para el reconocimiento del alfabeto dactilológico (Tesis de Maestría). Instituto Politécnico Nacional, Centro de Investigación en Computación.

Salgado-Martínez, G. A., Cuevas-Valencia, R. E., Feliciano-Morales, A., & Catalán-Villegas, A. (2024). Reconocimiento de la Lengua de Señas Mexicana usando Deep Learning mediante una aplicación móvil. Ciencia Latina Revista Científica Multidisciplinar. DOI: https://doi.org/10.37811/cl_rcm.v8i5.14458

Santos, D., Dallos, L., & Gaona-García, P. A. (2020). Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning. Información Tecnológica, 31(3), 23–38. https://doi.org/10.4067/S0718-07642020000300023

Sharma, C. M., et al. (2022). Indian Sign Language Recognition using transfer learning with EfficientNet. In Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence (ScitePress). https://www.scitepress.org/Papers/2021/107903/107903.pdf