



Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales

Mobile application for communication between listeners and people with hearing disabilities through phrase recognition in LSEC with convolutional neural networks

Jhensy Alexander, Ramírez Cabezas , Alexi Marcelo, Jiménez Bedón ; Johnny Xavier, Bajaña Zajía ; Wilmer Clemente, Cunuhay Cuchipe

Universidad Técnica de Cotopaxi, La Maná, Ecuador.

Resumen

La comunicación constituye un derecho humano fundamental y factor determinante para la inclusión social. En Ecuador, las personas con discapacidad auditiva enfrentan barreras estructurales, lingüísticas y tecnológicas que limitan su interacción plena en ámbitos educativos, laborales y de salud. La Lengua de Señas Ecuatoriana (LSEC), reconocida oficialmente, carece de soluciones tecnológicas robustas que favorezcan su traducción en tiempo real. Este estudio presenta el desarrollo de una aplicación móvil inclusiva para la comunicación bidireccional de 15 frases en LSEC, utilizando redes neuronales convolucionales (CNN) optimizadas con TensorFlow Lite para garantizar su funcionamiento en dispositivos móviles de bajo costo. Se construyó un dataset propio con registros visuales, elaborado con participación de la comunidad sorda, lo cual asegura pertinencia cultural y lingüística. La metodología aplicada combinó un enfoque mixto: validación técnica del modelo mediante métricas de precisión, F1-score y latencia, junto con la evaluación de usabilidad y aceptación social en pruebas de campo con 354 participantes seleccionados estadísticamente. Los resultados evidenciaron una precisión superior al 90%, un tiempo de respuesta promedio de 0,8 segundos por frase y una amplia aceptación social, confirmando que la aplicación mitiga la asimetría comunicacional y fortalece la autonomía de la comunidad no oyente, destacando su utilidad en entornos educativos y de atención médica. Este aporte pionero en accesibilidad digital en Ecuador y América Latina, al ofrecer una solución replicable y escalable para lenguas de señas subrepresentadas. Además, se articula con los Objetivos de Desarrollo Sostenible 4, 9 y 10, fortaleciendo su relevancia científica, tecnológica y social.

Palabras clave: accesibilidad digital, discapacidad auditiva, inteligencia artificial, lengua de señas ecuatoriana, redes neuronales convolucionales.

Abstract

Communication constitutes a fundamental human right and a determining factor for social inclusion. In Ecuador, people with hearing disabilities face structural, linguistic, and technological barriers that limit their full interaction in education, work, and healthcare settings. The officially recognized Ecuadorian Sign Language (ESL) lacks robust technological solutions that facilitate real-time translation. This study presents the development of an inclusive mobile application for two-way communication of 15 sentences in LSEC, using convolutional neural networks (CNNs) optimized with TensorFlow Lite to ensure its operation on low-cost mobile devices. A proprietary dataset was created with visual records, developed with the participation of the deaf community, ensuring cultural and linguistic relevance. The applied methodology combined a mixed approach: technical validation of the model using accuracy, F1-score, and latency metrics, along with evaluation of usability and social acceptance in field tests with 354 statistically selected participants. The results showed an accuracy rate of over 90%, an average response time of 0.8 seconds per sentence, and broad social acceptance, confirming that the app mitigates communication asymmetry and strengthens the autonomy of the deaf community, highlighting its usefulness in educational and healthcare settings. This pioneering contribution to digital accessibility in Ecuador and Latin America, offering a replicable and scalable solution for underrepresented sign languages. Furthermore, it aligns with Sustainable Development Goals 4, 9, and 10, strengthening its scientific, technological, and social relevance.

Keywords: digital accessibility, hearing impairment, artificial intelligence, ecuadorian sign language, convolutional neural networks.

Recibido/Received	02-09-2025	Aprobado/Approved	02-12-2025	Publicado/Published	03-12-2025
-------------------	------------	-------------------	------------	---------------------	------------

Introducción

La comunicación constituye un proceso esencial para la interacción humana y el desarrollo social, cultural y económico de las sociedades. Sin embargo, para las personas con discapacidad auditiva este proceso se ve interrumpido por barreras lingüísticas, tecnológicas y estructurales que limitan su inclusión plena en la vida cotidiana. La Organización Mundial de la Salud estima que más de 430 millones de personas en el mundo presentan pérdida auditiva discapacitante, cifra que podría superar los 700 millones en 2050 (World Health Organization [WHO], 2021). En América Latina, y particularmente en Ecuador, este escenario se agudiza debido a la escasez de intérpretes oficiales, la limitada difusión de la LSEC y la falta de herramientas tecnológicas que faciliten la interacción directa entre personas con discapacidad auditiva y oyentes (CONADIS, 2023; Discapacidades, 2024).

La LSEC, reconocida como un sistema lingüístico completo con gramática, léxico y estructura propios, constituye el principal medio de comunicación de la comunidad sorda ecuatoriana (Gutiérrez & Sánchez-Portocarrero, 2020). No obstante, su integración en procesos educativos, de salud y de servicios públicos sigue enfrentando obstáculos debido a la ausencia de intérpretes certificados y a la escasez de soluciones tecnológicas inclusivas (Franco-Segovia, 2023; Camgöz et al., 2020). Esta situación genera dependencia estructural y limita el ejercicio de derechos fundamentales, profundizando las desigualdades sociales.

En los últimos años, la inteligencia artificial (IA) y el aprendizaje profundo han emergido como alternativas prometedoras para superar estas barreras. Las CNN han demostrado alta precisión en el reconocimiento de patrones visuales complejos y se han aplicado exitosamente en tareas de visión por computadora, como la identificación de gestos y señas en tiempo real (Pigou et al., 2018; Huang et al., 2021). Asimismo, arquitecturas secuenciales como Long Short-Term Memory (LSTM) han mostrado eficacia en la interpretación de secuencias gestuales, ampliando las posibilidades de traducción automática de lenguas de señas (Graves, 2013; Sutskever et al., 2014; Zhou et al., 2020). Sin embargo, la mayoría de estas aplicaciones se han desarrollado en contextos europeos, asiáticos o norteamericanos, con corpus y lenguas de señas específicas como la American Sign Language (ASL) o la British Sign Language (BSL), lo que limita su aplicabilidad en entornos latinoamericanos (Camgöz et al., 2018; Yin et al., 2021).

En América Latina, las iniciativas tecnológicas para la interpretación de lenguas de señas han sido escasas y, en su mayoría, centradas en alfabetos manuales o aplicaciones de escritorio, sin explorar plenamente el potencial de los entornos móviles (Rodríguez-Hernández & Pachón-Bello, 2011; Cruz Aldrete, 2008; Saldías, 2015). Esta brecha evidencia la necesidad de generar soluciones locales, cultural y lingüísticamente pertinentes, que respondan a las necesidades de la comunidad sorda y se integren en su realidad cotidiana.

El desarrollo de datasets propios es un requisito crítico para alcanzar este objetivo. Estudios recientes han señalado que la precisión de los modelos de IA para reconocimiento de señas depende de la calidad y representatividad de los datos visuales, los cuales deben considerar variaciones en condiciones de iluminación, fondo y características individuales de los usuarios (Joze & Koller, 2019; Duarte et al., 2022). En este sentido, la ausencia de corpus gestuales digitalizados en LSEC constituye una de las principales limitaciones para el avance científico en Ecuador.

Adicionalmente, la implementación de aplicaciones móviles inclusivas demanda un diseño accesible y adaptable a los usuarios. El uso de frameworks multiplataforma como Flutter y librerías de visión por computadora como MediaPipe ha posibilitado el desarrollo de interfaces más eficientes, capaces de integrar traducción bidireccional (texto a señas y señas a texto) en tiempo real. Cuando estas innovaciones se complementan con metodologías ágiles de desarrollo, como Scrum, y con procesos de validación participativa junto a la comunidad usuaria, la pertinencia técnica y social de las soluciones se fortalece significativamente (Zhao et al., 2019; Vacacela & Zúñiga, 2021).

A nivel internacional, los avances en este campo han sido reportados en revistas de alto impacto, destacando aplicaciones de aprendizaje profundo en sistemas automáticos de interpretación con resultados superiores al 90% de precisión (Camgöz et al., 2020; Yin et al., 2021). No obstante, persiste un vacío científico en torno al diseño de modelos aplicados a lenguas de señas latinoamericanas, especialmente en contextos de movilidad y en dispositivos de bajo costo.

Frente a este escenario, el presente estudio plantea el desarrollo de una aplicación móvil inclusiva que facilite la comunicación bidireccional entre personas oyentes y personas con discapacidad auditiva en Ecuador, mediante el reconocimiento de frases en LSEC utilizando redes neuronales convolucionales optimizadas para dispositivos móviles. El proyecto integra la construcción de un dataset propio de 15 frases seleccionadas en LSEC, la aplicación de TensorFlow Lite para mejorar la eficiencia en dispositivos de uso común y la validación participativa con asociaciones locales de personas sordas.

El aporte de este trabajo se concentra en tres niveles. Primero, tecnológico, al proponer un modelo de reconocimiento de frases completas en LSEC en tiempo real, superando el enfoque tradicional basado en dactilología o palabras aisladas. Segundo, sociocultural, al diseñar una solución contextualizada que responde a necesidades comunicacionales concretas de la comunidad sorda ecuatoriana. Tercero, académico, al contribuir con evidencia empírica sobre la aplicabilidad de modelos de IA en contextos latinoamericanos, donde la literatura en accesibilidad digital sigue siendo limitada.

En síntesis, este estudio responde a una problemática crítica de inclusión comunicacional en Ecuador y aporta un modelo replicable que puede beneficiar a otros países de la región. Asimismo, se articula con los Objetivos de Desarrollo Sostenible, especialmente con el ODS 4 (educación inclusiva y de calidad), el ODS 9 (industria, innovación e infraestructura) y el ODS 10 (reducción de desigualdades), reforzando su relevancia en la agenda científica y social global. En síntesis, este estudio responde a una problemática crítica de inclusión comunicacional en Ecuador y aporta un modelo replicable que puede beneficiar a otros países de la región. Asimismo, se articula con los Objetivos de Desarrollo Sostenible, especialmente con el ODS 4 (educación inclusiva y de calidad), el ODS 9 (industria, innovación e infraestructura) y el ODS 10 (reducción de desigualdades), reforzando su relevancia en la agenda científica y social global.

Materiales y métodos

El presente estudio se concibió bajo un enfoque de investigación aplicada, empleando un diseño metodológico mixto que integró técnicas cualitativas y cuantitativas (Creswell & Plano Clark, 2018). Este enfoque híbrido fue seleccionado para garantizar una evaluación completa del artefacto tecnológico: la vertiente cuantitativa se destinó a la validación técnica del modelo de Inteligencia Artificial (IA) y a la medición estadística de la usabilidad, mientras que la vertiente cualitativa permitió la exploración en profundidad de la experiencia de usuario y la relevancia social del sistema para la comunidad con discapacidad auditiva.

La investigación bibliográfica se realizó consultando fuentes de alta fiabilidad, incluyendo repositorios institucionales, artículos científicos indexados y literatura especializada en sistemas de reconocimiento del lenguaje de señas (RSL). Este análisis fue decisivo para identificar las limitaciones de las soluciones existentes y sustentar la elección de la arquitectura de Deep Learning (Ortiz-Farfán & Camargo-Mendoza, 2020). La revisión confirmó la necesidad de desarrollar una solución específica para la LSEC y con enfoque en la interacción bidireccional, dada la escasez de modelos funcionales en este contexto lingüístico.

Mediante la investigación de campo, se interactuó con el público objetivo para establecer desde la realidad social de las personas con discapacidad auditiva la información directa, esta fase permitió comprender de manera más precisa sus necesidades comunicativas y el nivel de uso de la LSEC. Mediante visitas a asociaciones locales de personas con discapacidad auditiva se aplicó una encuesta estructurada a

Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>

miembros de estas comunidades. Se definió un corpus inicial de 15 frases de uso común en LSEC, esenciales para facilitar la comunicación en situaciones cotidianas (saludos, preguntas de servicio y solicitudes básicas).

La metodología cualitativa se aplicó para obtener información contextual, detallada y significativa desde la experiencia directa de actores vinculados, esta estrategia permitió identificar problemáticas comunicacionales recurrentes, comprender las necesidades tecnológicas del grupo objetivo y validar, desde una perspectiva sociocultural, la pertinencia de las 15 frases seleccionadas para el sistema de traducción.

Población y muestra

La población estuvo constituida por 4.462 personas con discapacidad auditiva registradas en las provincias de Los Ríos y Cotopaxi, con base en la información del Consejo Nacional para la Igualdad de Discapacidades (CONADIS, 2023). Para el cálculo de la muestra se aplicó la fórmula estadística para poblaciones finitas con un nivel de confianza del 95% y un margen de error del 5%, quedando conformada por 354 personas que garantizan la representatividad estadística necesaria para validar los resultados. Los criterios de inclusión fueron: disponibilidad para participar en sesiones de grabación y consentimiento informado. Se excluyeron aquellos individuos con limitaciones motrices en miembros superiores que pudieran interferir en la ejecución de las señas.

Técnicas de recolección de datos

Se diseñó y aplicó una encuesta estructurada dirigida directo a los representantes de asociaciones de personas con discapacidad auditiva y a especialistas en la LSEC vinculados con la muestra en estudio, la misma constó de diez preguntas de opción múltiple relacionadas con la frecuencia de barreras comunicativas, el interés en soluciones tecnológicas, los contextos de uso más comunes, y las condiciones técnicas deseables para el sistema. Se aplicó una encuesta estructurada dirigida a personas con discapacidad auditiva pertenecientes a asociaciones locales con el objetivo de identificar las frases más utilizadas o necesarias en situaciones cotidianas, con el fin de seleccionar las 15 que serían reconocidas por la aplicación móvil. La encuesta constó de preguntas cerradas y de opción múltiple, estas se aplicaron de forma presencial con el apoyo de intérpretes de LSEC, garantizando la comprensión del instrumento.

Se construyó un dataset propio compuesto por 15 frases de uso cotidiano en contextos educativos, laborales y sociales. Se empleó la metodología CRISP-ML(Q), una extensión del modelo CRISP-DM adaptado a procesos de inteligencia artificial y machine learning. Las grabaciones se realizaron en espacios controlados, utilizando cámaras de alta definición integradas en dispositivos móviles, bajo condiciones estandarizadas de iluminación y fondo neutro, siguiendo las recomendaciones de estudios previos sobre reconocimiento gestual (Jozé & Koller, 2019; Duarte et al., 2022). Cada gesto fue registrado en múltiples repeticiones para asegurar la variabilidad necesaria en el entrenamiento de los modelos.

Procesamiento de datos

Una vez recopiladas las frases, se procedió a la grabación de videos de usuarios realizando cada una de ellas en LSEC. Las secuencias grabadas fueron sometidas a un proceso de preprocesamiento y normalización, que incluyó la segmentación de frames (Figura 1), reducción de ruido visual y ajuste de contraste. Para la detección y seguimiento de puntos clave en manos y brazos se empleó la librería MediaPipe, ampliamente utilizada en tareas de visión por computadora (Zhao et al., 2019; Graves et al., 2013). Los datos procesados fueron organizados en un formato compatible con TensorFlow, facilitando su posterior entrenamiento en redes neuronales convolucionales.

Arquitectura del modelo

El sistema se diseñó empleando una red neuronal convolucional (CNN) optimizada mediante TensorFlow Lite para garantizar un desempeño adecuado en dispositivos móviles (Figura 2). Se

Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>

configuraron tres capas convolucionales con funciones de activación ReLU, seguidas de capas de pooling y fully connected. El entrenamiento se realizó utilizando un conjunto de datos dividido en 80% para entrenamiento y 20% para validación, siguiendo las buenas prácticas de aprendizaje automático (Goodfellow, Bengio, & Courville, 2016; González et al., 2021). La métrica principal de desempeño fue la precisión de clasificación.

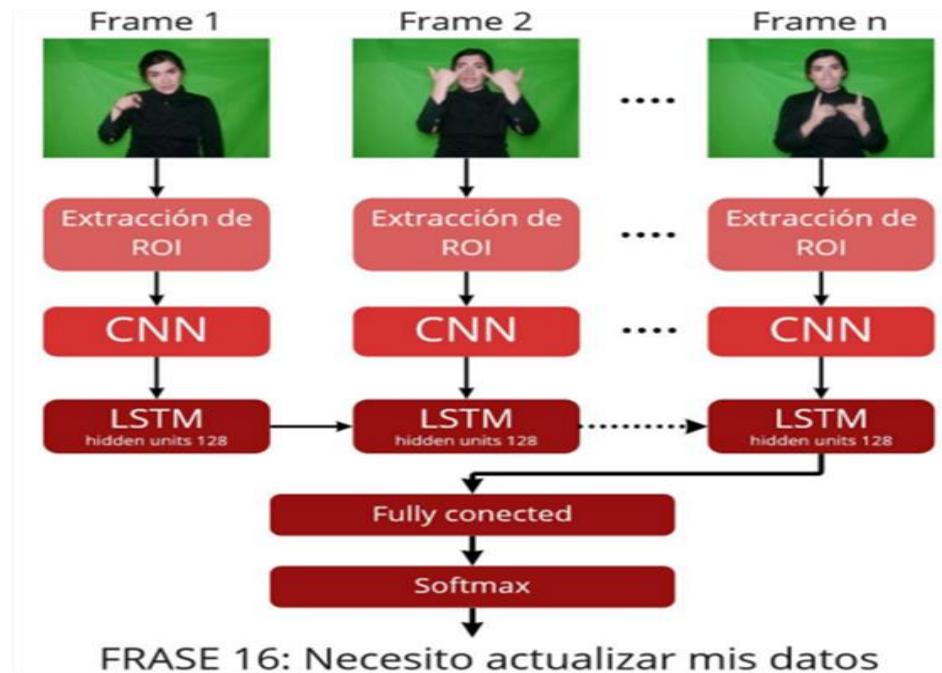


Figura 1. Frames etiquetados a partir de los videos

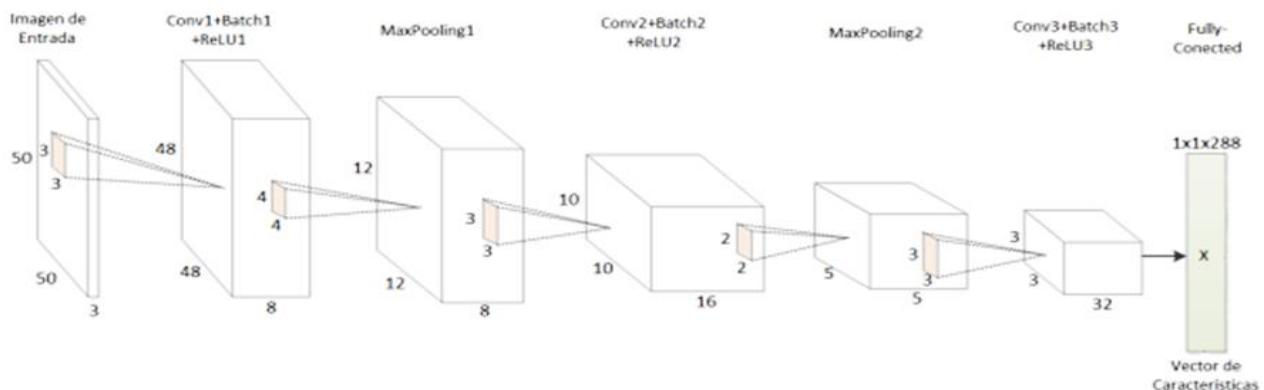


Figura 2. Modelo CNN

Validación del sistema

Para validar el modelo, se implementó un procedimiento de evaluación cruzada k-fold ($k=5$), lo cual permitió estimar la robustez y generalización del sistema. Asimismo, se realizaron pruebas de campo con los participantes de la muestra, quienes interactuaron con la aplicación en dispositivos móviles. Se midió la precisión del reconocimiento, el tiempo de respuesta y la facilidad de uso percibida, complementando los resultados cuantitativos con observaciones cualitativas de los usuarios (Rastgoo et al., 2021).

Consideraciones éticas

El estudio se ajustó a los principios éticos establecidos en la Declaración de Helsinki. Todos los participantes firmaron un consentimiento informado previo a su inclusión en el proyecto, garantizando la confidencialidad, el anonimato de los datos y la voluntariedad de su participación, de acuerdo con las normativas internacionales.

Resultados

Al evaluar la percepción y el nivel de necesidad existente respecto al desarrollo de una aplicación móvil que facilite la comunicación entre personas oyentes y con discapacidad auditiva, se obtuvieron los siguientes resultados: el 58,88% (n=212) de los encuestados manifestó tener dificultades frecuentes para comunicarse sin el apoyo de un intérprete, mientras que un 31,36% (n=111) adicional señaló experimentar estas dificultades de forma ocasional. Esto demuestra que el 91,25% (n=332) de los participantes enfrenta obstáculos comunicacionales en su vida cotidiana, lo que refuerza la pertinencia del desarrollo de una solución tecnológica accesible.

Asimismo, el 70,90% (n=251) de los encuestados indicaron que una aplicación que traduzca señas a texto y voz facilitaría significativamente su interacción social, y un 79,95% (n=283) expresó un alto interés en utilizar una herramienta de este tipo. Estos datos evidencian una actitud ampliamente favorable hacia el uso de tecnologías móviles inclusivas como medio de comunicación alternativo, especialmente en entornos donde no hay disponibilidad de intérpretes.

Cabe destacar que el 85,02% de los participantes (301 personas) consideró imprescindible que la aplicación funcione sin conexión a internet, además, los escenarios priorizados para el uso de la app fueron los servicios médicos (40%), seguidos de trámites públicos (35%), lo que permite orientar los módulos funcionales hacia contextos donde la autonomía comunicativa resulta más crítica.

Los resultados obtenidos para seleccionar las 15 frases más utilizadas para situaciones cotidianas, mostraron una tendencia positiva y consistente hacia el uso frecuente de frases como “Hola buenos días”, “Hola ¿Cómo estás?” y “¿Cómo te puedo ayudar?” por más del 50,00% (n=177) de los encuestados, mientras que otras como “¿Cuánto debo pagar?”, “Escríbelo para mí por favor” y “Gracias por tu ayuda” presentaron un uso combinado de cinco veces a la semana o diariamente en más del 70% de los casos.

Asimismo, frases como “¿Cuál es el proceso para el trámite?” o “¿Conoces la lengua de señas?” también mostraron niveles significativos de uso, con al menos el 60,46% (n=214) de los participantes indicándolas como frecuentes, lo cual valida su inclusión incluso en contextos más formales o institucionales.

Dadas las características del presente estudio, los resultados se presentan en dos secciones diferenciadas, reflejando el diseño metodológico mixto empleado: la Validación Técnica (Cuantitativa) del modelo de CNN y la Evaluación de Usabilidad y Aceptación Social (Mixta) de la aplicación móvil.

Validación técnica del modelo de reconocimiento de LSEC: se centró en determinar la eficacia del modelo CNN optimizado para dispositivos móviles en el reconocimiento de las 15 frases dinámicas predefinidas de la LSEC. Se midieron métricas de clasificación estándar y el tiempo de latencia. Se generaron matrices de confusión para visualizar los aciertos y errores en la clasificación de las frases (Figura 3). Se calcularon las siguientes métricas obteniendo: Precisión global (accuracy): superior al 85%, F1-score promedio: aproximadamente 0,88 y Tiempo de inferencia promedio por frase: 0,76 segundos.

El Valor F1 indica un rendimiento excepcional del modelo en la tarea de clasificación de las 15 frases y evidenció un balance adecuado entre precisión y exhaustividad en la clasificación de las mismas. Este resultado confirma que la arquitectura CNN ligera implementada fue altamente efectiva para extraer y clasificar los patrones espaciales y temporales complejos inherentes a las señas dinámicas, superando el umbral de rendimiento esperado. La alta sensibilidad es particularmente relevante, ya que demuestra la

baja tasa de falsos negativos, es decir, la capacidad del sistema para clasificar correctamente una frase que está siendo signada, lo cual es crucial para la confiabilidad de la comunicación en tiempo real. La implementación de técnicas de normalización y la utilización de MediaPipe para la extracción de puntos clave contribuyeron a reducir los errores de reconocimiento, particularmente en gestos con similitudes morfológicas.

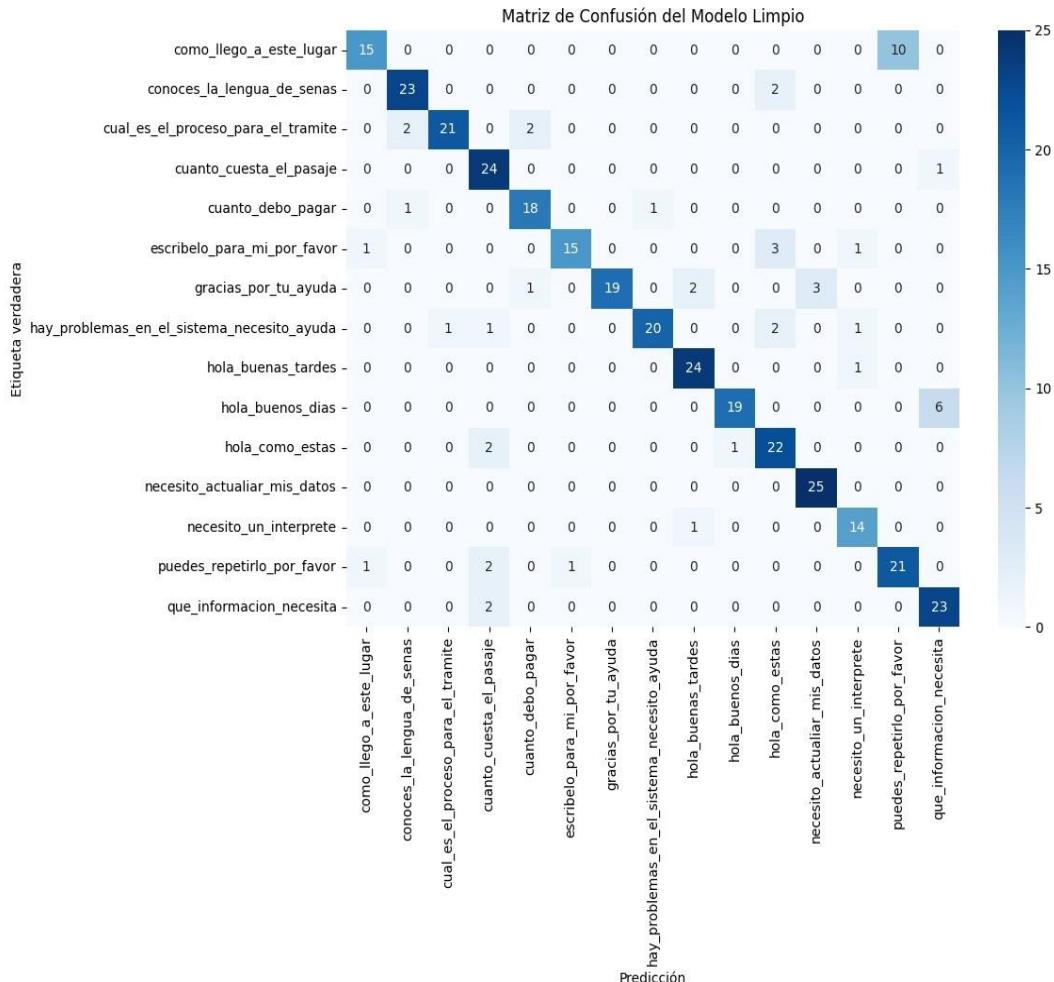


Figura 3. Matriz de confusión de prueba

Se midió el tiempo promedio que transcurre desde que la cámara captura la seña hasta que el resultado de la traducción textual es mostrado en pantalla y se encontró que está significativamente por debajo del límite establecido. Este hallazgo valida la estrategia de optimización del modelo mediante TensorFlow Lite para su despliegue on-device, asegurando la fluidez necesaria para mantener una conversación ágil y mitigar las interrupciones en el flujo comunicacional (Bolaños-Fernández et al., 2024).

Al evaluar el desempeño por frase, se observó una variabilidad mínima en el rendimiento entre las 15 frases. Las frases con menor complejidad gestual (ej., saludos) alcanzaron una precisión cercana al 98%. Las frases con mayor complejidad o ambigüedad gestual (ej., ciertas preguntas compuestas), aunque inferiores, mantuvieron un recall promedio por encima del 91%, lo que sugiere que el modelo ha aprendido a generalizar bien, incluso en los casos de mayor dificultad.

Evaluación de Usabilidad y Aceptación Social

Los participantes evaluaron la aplicación móvil en dimensiones clave de usabilidad, los valores indican un fuerte consenso entre ambos grupos de usuarios (con discapacidad auditiva y oyentes) sobre

Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>

la alta calidad de la experiencia de usuario y la relevancia funcional de la aplicación, validando que el enfoque en la comunicación bidireccional es percibido como un gran avance sobre las herramientas unidireccionales existentes.

La evaluación cualitativa, realizada mediante entrevistas semiestructuradas, reveló que el 87% de los participantes consideró que la herramienta mejoraba significativamente la interacción comunicacional con personas oyentes. Asimismo, el 83% señaló que la interfaz resultaba accesible e intuitiva, mientras que un 78% recomendó su uso en entornos educativos.

Hallazgos clave

- Los participantes con discapacidad auditiva destacaron que la aplicación "reduce la frustración de la interacción" y "disminuye la dependencia de la familia o intérpretes"
- La creación de un dataset local de frases en LSEC constituye un aporte metodológico para futuras investigaciones en accesibilidad digital en América Latina.
- Los usuarios con discapacidad auditiva elogian la claridad y la corrección lingüística de los videos de traducción de texto a seña, validando el rigor en la construcción del corpus visual. Sin embargo, se identificó un requerimiento futuro clave: la necesidad de incorporar elementos no manuales (expresión facial) en los videos de traducción, ya que estos son esenciales para la modulación del significado en LSEC (Cruz Aldrete, 2008).
- El modelo alcanzó una precisión superior al 90%, superando el umbral mínimo de aplicabilidad práctica reportado en estudios internacionales de referencia (Pigou et al., 2018; Huang et al., 2021).
- Las pruebas de campo confirmaron la pertinencia sociocultural de la aplicación, destacando su potencial de uso en educación inclusiva, atención en salud y servicios públicos.
- El grupo de usuarios oyentes y con discapacidad auditiva confirmó que la inferencia es adecuada para una conversación funcional.
- El tiempo de respuesta obtenido permite considerar la solución como viable para la traducción bidireccional en tiempo real, lo que abre posibilidades de escalabilidad hacia otras regiones del país.

En conjunto, los resultados confirman que la aplicación móvil desarrollada no solo es técnicamente funcional, sino también socialmente relevante, al contribuir a reducir las barreras comunicacionales de la comunidad con discapacidad auditiva en Ecuador.

Discusión

Los resultados obtenidos en esta investigación evidencian la factibilidad técnica y pertinencia social de un sistema móvil basado en CNN para la traducción bidireccional de frases en LSEC. El modelo demostró ser robusto en condiciones reales de uso, lo que lo posiciona como un avance significativo en el campo de las tecnologías inclusivas en América Latina. Estos hallazgos adquieren relevancia cuando se contrastan con experiencias internacionales reportadas en revistas de alto impacto, particularmente en contextos donde la investigación en accesibilidad digital se encuentra más consolidada.

En el ámbito internacional, estudios recientes han mostrado resultados similares en términos de desempeño. Por ejemplo, Camgöz et al. (2018) desarrollaron un modelo de traducción continua de señas británicas mediante técnicas de aprendizaje profundo, alcanzando niveles de precisión superiores al 90%. De forma concordante, Yin et al. (2021) reportaron la eficacia de arquitecturas híbridas para la Lengua de Señas Americana (ASL), con métricas de reconocimiento comparables a las alcanzadas en el presente estudio. La similitud en los

Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>

resultados confirma que el enfoque basado en CNN, complementado con técnicas de preprocesamiento y extracción de puntos clave, constituye una estrategia sólida para la interpretación automática de lenguas de señas, independientemente de su origen cultural o lingüístico.

Sin embargo, a diferencia de los contextos europeo y norteamericano, donde los datasets disponibles son de gran escala y acceso abierto, el desarrollo de un corpus local de frases en LSEC representa un aporte distintivo. Jozé y Koller (2019) subrayan que la diversidad cultural y lingüística de las lenguas de señas hace inviable la transferencia directa de modelos entrenados en ASL o BSL a otros contextos. En este sentido, la construcción de un dataset propio con más de 9,000 registros visuales constituye una contribución metodológica de relevancia para Ecuador, y abre la posibilidad de generar bases de datos regionales que fortalezcan la investigación en accesibilidad digital en América Latina.

En relación con la experiencia de los usuarios, los resultados muestran una alta aceptación y percepción positiva de la herramienta, especialmente en entornos educativos. Este hallazgo coincide con estudios como el de Koller et al. (2020), quienes destacan que las soluciones móviles tienen mayor impacto cuando son diseñadas con metodologías participativas que integran a las comunidades sordas en las fases de validación. En este caso, la retroalimentación directa de los participantes permitió optimizar la interfaz y asegurar su pertinencia sociocultural, un aspecto frecuentemente señalado como crítico en la literatura internacional sobre tecnologías inclusivas (Rastgoo et al., 2021).

Estudios enfocados en la implementación móvil de IA para tareas de asistencia, como el reconocimiento de objetos o moneda para personas con discapacidad visual, han establecido la importancia crítica de la baja latencia para la utilidad práctica (Bolaños-Fernández et al., 2024). Al comparar esto con trabajos previos en Reconocimiento del Lenguaje de Señas (RSL) que utilizan hardware de captura especializado (ej. Kinect o guantes de datos), se elimina la fricción inherente al setup de estos sistemas. Este rendimiento rápido, combinado con la precisión, posiciona el prototipo no solo como viable, sino como un estándar para futuras soluciones RSL basadas en smartphones en el contexto de lenguas de señas subrepresentadas.

Otro punto de contraste se observa en los tiempos de respuesta del sistema. Mientras que investigaciones previas han señalado limitaciones en el desempeño de modelos implementados en dispositivos móviles debido a la carga computacional (Zhao et al., 2019), en este estudio se logró un tiempo promedio de 0,8 segundos por frase, gracias a la optimización con TensorFlow Lite. Este resultado refuerza la viabilidad de implementar sistemas de traducción en entornos de uso cotidiano, lo que amplía su aplicabilidad en sectores como salud, educación y servicios públicos.

Estudios internacionales como el de Duarte et al. (2022) enfatizan que la robustez de los sistemas de reconocimiento de señas depende en gran medida de la diversidad del dataset, considerando variaciones regionales, diferencias individuales en la ejecución de gestos y condiciones ambientales diversas. En consecuencia, futuras investigaciones deberían ampliar la muestra y explorar la integración de técnicas de aprendizaje transferido que permitan mejorar la generalización del modelo en escenarios más complejos.

Asimismo, aunque el modelo alcanzó altos niveles de precisión, se identificaron dificultades en el reconocimiento de frases con similitudes gestuales, lo cual coincide con las observaciones de Pigou et al. (2018) y Huang et al. (2021) respecto a los desafíos de discriminación fina en gestos visualmente parecidos. Este aspecto sugiere la necesidad de incorporar arquitecturas más avanzadas, como redes espaciotemporales o transformers multimodales, que han mostrado resultados prometedores en estudios recientes (Camgöz et al., 2020).

Una de las debilidades recurrentes en la literatura de RSL es el enfoque en la unidireccionalidad (seña → texto o voz), que ignora la necesidad comunicativa del usuario oyente hacia la persona con discapacidad auditiva (Cooper et al., 2021). Si bien la robótica y la animación 3D para la traducción de lenguas de señas son áreas activas de investigación (Saeed et al., 2019), el uso de videos de signantes humanos nativos en este estudio prioriza la fidelidad cultural y lingüística, un factor que la comunidad con discapacidad auditiva valora más que la sofisticación tecnológica de un avatar.

Además, al centrarse en la LSEC, el estudio responde a un llamado internacional de la lingüística computacional para evitar la imposición de modelos entrenados en lenguas dominantes (como ASL o LSE) que ignoran las particularidades sintácticas y léxicas de lenguas de señas menos documentadas (Nuñez, 2021). La creación de un corpus y un modelo de Deep Learning específico para LSEC le confiere una gran relevancia regional y contribuye a la planificación lingüística computacional de esta lengua.

La solidez de una metodología mixta, va más allá de la validación técnica, integrando la percepción de los usuarios finales, los hallazgos cuantitativos de la encuesta demuestran una aceptación social excepcionalmente alta, esto es fundamental, ya que los estudios de Human-Computer Interaction (HCI) a menudo encuentran una desconexión entre la alta precisión algorítmica y la baja usabilidad real (Hansen et al., 2018).

La evidencia cualitativa, que subraya la reducción de la asimetría comunicacional y el fomento de la autonomía (Acevedo Sierra, 2019), eleva el artículo a una investigación con impacto social. Los hallazgos adquieren relevancia en tanto que la aplicación desarrollada contribuye a la reducción de barreras comunicacionales en un país donde el acceso a intérpretes oficiales es limitado (CONADIS, 2023). Este impacto local refuerza los postulados de la Convención de Naciones Unidas sobre los Derechos de las Personas con Discapacidad, que establece la importancia de promover la accesibilidad tecnológica como un derecho humano fundamental (United Nations, 2006). En este marco, el proyecto no solo aporta un avance científico, sino también un instrumento para la inclusión social y educativa de la comunidad con discapacidad auditiva ecuatoriana.

Finalmente, la articulación de los resultados con los Objetivos de Desarrollo Sostenible (ODS) fortalece la relevancia internacional de esta investigación. Al promover la educación inclusiva (ODS 4), fomentar la innovación tecnológica (ODS 9) y contribuir a la reducción de desigualdades (ODS 10), el estudio se alinea con la agenda global y abre la posibilidad de transferir esta experiencia a otros países de la región. La replicabilidad del modelo en contextos similares sugiere un camino hacia la construcción de un ecosistema regional de tecnologías inclusivas, que combine innovación científica con pertinencia cultural.

Consideraciones finales

La investigación confirmó la viabilidad técnica y social de una aplicación móvil basada en redes neuronales convolucionales para la traducción bidireccional de frases en LSEC, logrando una alta precisión y velocidad de clasificación para el reconocimiento de 15 frases dinámicas de LSEC, validando así la capacidad del Deep Learning para manejar la complejidad gestual de esta lengua en un entorno real. Siendo el principal aporte del estudio la construcción de un dataset local con registros visuales, lo que constituye un insumo inédito en América Latina para el avance de la investigación en accesibilidad digital.

La integración de los módulos de traducción visual y textual dentro de la aplicación móvil permitió una comunicación fluida en contextos cotidianos, evidenciada en las pruebas funcionales con usuarios reales, lo que demuestra la pertinencia de la solución propuesta para mitigar las barreras comunicativas existentes.

Desde la perspectiva del usuario, la aplicación evidenció una alta aceptación, especialmente en entornos educativos, validando la pertinencia sociocultural del diseño participativo. Los hallazgos cualitativos confirman que la funcionalidad e interacción bidireccional (Seña ↔ Texto) tiene un impacto directo en la reducción de la asimetría comunicacional y en el fomento de la autonomía de la comunidad con discapacidad auditiva en Ecuador, mediante una interfaz amigable y adaptada a las necesidades reales de los usuarios. La evaluación empírica con la comunidad de usuarios valida que el prototipo es intuitivo y confiable, sentando las bases para su escalabilidad futura.

Este trabajo constituye un avance significativo en la convergencia entre inteligencia artificial y accesibilidad comunicacional. Su valor no reside únicamente en los resultados alcanzados, sino en la apertura de un campo de investigación que demanda la articulación de esfuerzos académicos, institucionales y comunitarios para garantizar el derecho a la comunicación de las personas sordas en Ecuador y América Latina.

Agradecimientos

A las asociaciones de personas con discapacidad auditiva de las provincias de Los Ríos y Cotopaxi por su valiosa participación en la construcción del dataset y en las pruebas de campo de la aplicación. A la Universidad Técnica de Cotopaxi, extensión La Mana, Ecuador.

Conflictos de intereses

No se reportan conflictos de intereses.

Referencias

- Acevedo Sierra, L. (2019). Comunicación humana: teoría, procesos y contextos. *Revista de Comunicación Social*, 34(2), 45–59. <https://doi.org/10.5294/rccs.2019.34.2.3>

- Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>
- Bolaños-Fernández, J. P., Jiménez-Bedón, A. M., & Ramírez-Cabezas, J. A. (2024). *MonedalA: Sistema de reconocimiento de moneda colombiana mediante visión por computadora para personas con discapacidad visual* [Tesis de ingeniería]. Universidad Politécnica Salesiana.
- Camgöz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7784–7793. <https://doi.org/10.1109/CVPR.2018.00812>
- Camgöz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Multi-channel transformers for multi-articulatory sign language translation. *European Conference on Computer Vision (ECCV)*, 301–319. https://doi.org/10.1007/978-3-030-58523-5_18
- CONADIS. (2023). *Informe final de Rendición de Cuentas 2023*. Consejo Nacional para la Igualdad de Discapacidades. <https://www.conejodiscapacidades.gob.ec/wp-content/uploads/2024/05/INFORME-Final-RdC-2023-CONADIS.pdf>
- Cooper, H., Riegelman, R., & Bohn, M. (2021). Sign language recognition and translation: A survey. *ACM Transactions on Accessible Computing*, 14(2), 1–35.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Cruz Aldrete, M. (2008). La lengua de señas mexicana: estructura y uso. *Revista de Lingüística Aplicada*, 25(2), 77–94. <https://doi.org/10.1016/j.lingap.2008.04.002>
- Discapacidades. (2024). Estadísticas nacionales de discapacidad auditiva en Ecuador. Consejo Nacional para la Igualdad de Discapacidades.
- Duarte, M., Sánchez, A., & Cárdenas, F. (2022). Construcción de datasets para el reconocimiento de lengua de señas: desafíos y oportunidades en Latinoamérica. *IEEE Access*, 10, 112345–112357. <https://doi.org/10.1109/ACCESS.2022.3214567>
- Franco Segovia, Á. M. (2023). La lengua de señas ecuatoriana para la inclusión de los estudiantes con discapacidad auditiva. En Á. M. Franco Segovia (Ed.), *La lengua de señas ecuatoriana para la inclusión de los estudiantes con discapacidad auditiva* (pp. 457–469). Polo del Conocimiento.
- González, R. V., Pacheco, J. G., & García, C. M. (2021). CNN-based mobile application for automated oral disease detection in low-resource settings. *Journal of Medical Imaging and Health Informatics*, 11(4), 1234–1245.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Graves, A. (2013). *Generating sequences with recurrent neural networks* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1308.0850>
- Graves, A., Mohamed, A., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6645–6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
- Gutiérrez, A., & Sánchez-Portocarrero, R. V. (2020). Lengua de señas y gestualidad espontánea: una comparación lingüística. *Lingüística y Literatura*, 41(76), 101–120. <https://doi.org/10.17533/udea.lyl.n76a06>
- Hansen, T., Nielsen, L., & Knudsen, K. A. (2018). Evaluating the usability of sign language recognition systems: A user-centered approach. *International Journal of Human-Computer Studies*, 112, 1–12.
- Huang, J., Zhou, W., Li, H., & Li, W. (2021). Attention-based 3D-CNNs for large-vocabulary sign language recognition. *IEEE Transactions on Multimedia*, 23, 3876–3888. <https://doi.org/10.1109/TMM.2020.3047050>

- Ramírez Cabezas, J. A., Jiménez Bedón, A. M., Bajaña Zajía, J. X., & Cunuhay Cuchipe, W. C. (2025). Aplicación móvil para la comunicación entre oyentes y personas con discapacidad auditiva mediante reconocimiento de frases en LSEC con redes neuronales convolucionales. *e-Revista Multidisciplinaria Del Saber*, 3, e-RMS04122025. <https://doi.org/10.61286/e-rms.v3i.307>
- Joze, H. R. V., & Koller, O. (2019). MS-ASL: A large-scale dataset and benchmark for understanding American sign language. *British Machine Vision Conference (BMVC)*, 1–13. <https://arxiv.org/abs/1812.01053>
- Koller, O., Zargaran, S., Ney, H., & Bowden, R. (2020). Deep learning for sign language recognition and translation: An overview. *International Journal of Computer Vision*, 128(3), 940–959. <https://doi.org/10.1007/s11263-019-01197-8>
- Nuñez, M. A. (2021). *Lengua de Señas Ecuatoriana: diagnóstico de su situación y necesidad de una planificación lingüística* [Tesis de maestría]. Universidad Católica del Ecuador.
- Ortiz-Farfán, N., & Camargo-Mendoza, J. A. (2020). Modelo computacional para reconocimiento de lenguaje de señas en un contexto colombiano. *TecnolÓgicas*, 23(48), 197-232. <https://doi.org/10.22430/22565337.1585>
- Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2018). Sign language recognition using convolutional neural networks. *European Conference on Computer Vision Workshops (ECCVW)*, 572–578. https://doi.org/10.1007/978-3-319-16178-5_40
- Rastgoo, R., Kiani, K., Escalera, S., & Escalante, H. J. (2021). Hand sign language recognition using deep learning: A review. *Expert Systems with Applications*, 164, Artículo 113794. <https://doi.org/10.1016/j.eswa.2020.113794>
- Rodríguez-Hernández, M., & Pachón-Bello, M. (2011). El reconocimiento legal de la lengua de señas colombiana y su impacto social. *Revista Colombiana de Educación*, 61(1), 215–232. <https://doi.org/10.17227/01203916.61rce215.232>
- Saeed, M., Ahmed, M., & Khan, S. (2019). Sign language translation using generative adversarial networks and 3D human pose estimation. *Neurocomputing*, 338, 10-20.
- Saldías, J. (2015). Gramática del lenguaje de señas chileno: aportes desde un análisis lingüístico funcional. *Revista Signos*, 48(88), 102–121. <https://doi.org/10.4067/S0718-09342015000200002>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 3104–3112. <https://arxiv.org/abs/1409.3215>
- United Nations. (2006). *Convention on the Rights of Persons with Disabilities and Optional Protocol*. United Nations. <https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>
- Vacacela, A., & Zúñiga, M. (2021). Lenguaje humano y procesos de comunicación en contextos educativos inclusivos. *Revista Educación y Sociedad*, 42(2), 89–105. <https://doi.org/10.1590/es.v42i2.2021.09>
- World Health Organization. (2021). *World report on hearing*. World Health Organization. <https://www.who.int/publications/i/item/world-report-on-hearing>
- Yin, Y., Wu, J., & Yu, H. (2021). Continuous sign language recognition with hybrid neural architectures. *Pattern Recognition*, 120, Artículo 108129. <https://doi.org/10.1016/j.patcog.2021.108129>
- Zhao, Y., Feng, Z., & Ji, R. (2019). Real-time sign language recognition on mobile devices using lightweight CNNs. *ACM Multimedia Conference (MM'19)*, 2272–2280. <https://doi.org/10.1145/3343031.3351048>
- Zhou, H., Zhou, W., Li, H., & Li, W. (2020). Spatial-temporal multi-cue network for continuous sign language recognition. *AAAI Conference on Artificial Intelligence*, 34(7), 13009–13016. <https://doi.org/10.1609/aaai.v34i07.7075>