

Lentes inteligentes, una solución para la discapacidad auditiva y enseñanza de la lengua de señas mexicana

J.A Chin Cantillo
Escuela de Ingeniería
Universidad Modelo
Mérida, México
12112846@modelo.edu.mx
ORCID: 0009-0006-7094-2716

M.A Ulín Mena
Escuela de Ingeniería
Universidad Modelo
Mérida, México
15233697@modelo.edu.mx
ORCID: 0009-0000-3488-5300

P.I Pérez Rosas
Escuela de Ingeniería
Universidad Modelo
Mérida, México
19198549@modelo.edu.mx
ORCID:

A.E Pacheco Álvarez
Escuela de Ingeniería
Universidad Modelo
Mérida, México
15235334@modelo.edu.mx
ORCID:

M.R Ruiz Pinto
Escuela de Ingeniería
Universidad Modelo
Mérida, México
15234184@modelo.edu.mx
ORCID: 0009-0001-6790-8447

Abstract— La lengua de señas es una de las herramientas más fundamentales para la comunidad sorda, sin embargo, las barreras sociales limitan su interacción social e integración. Este artículo propone una solución basada en un prototipo de lentes inteligentes que permiten la traducción bidireccional de la dactilología mexicana a texto y el audio a gestos y promueve el aprendizaje interactivo a través de estímulos visuales. El prototipo consta de una interfaz digital fácil de usar, una arquitectura de red neuronal MobileNetV2 para el reconocimiento de gestos y un sistema de proyección óptica para que el usuario pueda ver la interfaz digital. Los componentes fueron variados, como una pantalla OLED, sensores táctiles y características ergonómicas para mejorar la comodidad y la portabilidad. Este enfoque innovador tiene como objetivo reducir las barreras de comunicación e integración de las comunidades sordas y oyentes.

Keywords—Visión por computadora, Dactilología, Redes neuronales

I. INTRODUCCIÓN

La lengua de señas es una de las principales formas de comunicación para personas sordas. El desarrollo integral de una persona depende en gran medida de su capacidad de comunicación, lo que influye en sus habilidades y personalidad. Sin embargo, gran parte del mundo ha adoptado un método de enseñanza enfocado en el habla, lo que, de acuerdo con Herrera Fernández [1], ha provocado mucho analfabetismo, rechazo social y mala calidad de vida para los sordos.

Según la Organización Mundial de la Salud (OMS) [2], más del 5% de la población mundial, lo que equivale a 430 millones de personas, padece alguna pérdida auditiva discapacitante y requiere rehabilitación. Se estima que para el año 2050, esta cifra aumente a 700 millones, afectando a 1 de cada 10 personas. La

OMS también señala que la pérdida de audición aumenta con la edad, y que más del 25% de las personas mayores de 60 años padecen de una pérdida de audición discapacitante. En México, aproximadamente 2.3 millones de personas sufren de una discapacidad auditiva. Las personas mayores de 60 años representan a más del 50%; las personas entre 30 y 59 años un poco más del 34% y cerca de 2% representa a niños y niñas [3].

En el país, actualmente menos del 2% de la población oyente sabe comunicarse en Lengua de Señas Mexicana (LSM). Esta situación es inquietante, ya que la falta de una cultura de empatía y conciencia en nuestra sociedad se debe en gran parte a la escasez de políticas que promuevan su aprendizaje. Como afirma Pérez de la Fuente [4], "Existen concepciones de la sordera, socialmente construidas, que están en tensión, y tienen su traslación específica en términos de inclusión, redistribución y reconocimiento de las personas pertenecientes a esta minoría".

En trabajos como el de Bantupalli & Xie [5], un prototipo de traductor de Lengua de Señas Americano fue desarrollado usando algoritmos de Redes Neuronales Convolucionales (RNC), algoritmos de Memoria Larga a Corto Plazo (MLCP) y una Red Neuronal Recurrente.

Actualmente, existen dispositivos móviles que utilizan procesamientos en la nube para convertir letras de dactilología en texto. Hays, del Instituto de Tecnología de Rochester en Nueva York [6] ha desarrollado un algoritmo en C++ que convierte signos de Lengua de Señas Americano (LSA) en texto. Este enfoque integra el procesamiento local con asistencia en la nube para mejorar la precisión en el reconocimiento de lengua de señas.

Por otra parte, en 2022 se llevó a cabo un prototipo digital de un traductor de lengua de señas en tiempo real [7] enfocado

en su uso en el sector hotelero, desarrollando una ampliación de un algoritmo de aprendizaje profundo enfocado en el modelo de predicción a partir del conjunto de datos haciendo uso de Python, OpenCV, Tensorflow y Keras como base para la construcción y entrenamiento del modelo predictivo.

En el desarrollo de las tecnologías relacionadas se explora cómo el aprendizaje profundo puede mejorar la precisión y la eficiencia del reconocimiento de la lengua de señas en tiempo real [8]. Su enfoque se basa en redes neuronales profundas para interpretar los movimientos de las manos y las expresiones faciales, facilitando una comunicación más fluida y natural para los usuarios que utilizan la lengua de señas en contextos cotidianos. Por ese lado se introduce una solución innovadora mediante el uso de gafas de realidad aumentada (AR) para la traducción del lenguaje de señas [9].

Uno de los principales desafíos es la complejidad técnica involucrada en la traducción automática de lenguaje hablado a LSM. Esta tarea no solo implica la traducción de palabras, sino también la interpretación de gestos, expresiones faciales y movimientos corporales que varían según el contexto y la gramática visual de la lengua de señas. El desarrollo de algoritmos y hardware capaces de realizar esta tarea con precisión es complejo y aún está en fases experimentales [10].

Poder vencer estos desafíos es importante para mejorar la calidad de vida de las personas sordas e impulsar una mayor integración social. Una solución para abordar estas barreras es el desarrollo de un prototipo de lentes inteligentes accesibles que ofrecería una traducción bidireccional de dactilología y entrenamiento interactivo. Este dispositivo aparte de brindar una traducción instantánea y precisa entre lenguaje hablado y dactilología de la LSM; también facilitará el aprendizaje y la práctica de la dactilología en tiempo real. La implementación de esta tecnología podría transformar la enseñanza y el aprendizaje de la LSM, reduciendo las barreras comunicativas y fomentando una integración más efectiva entre comunidad sorda y oyentes.

II. METODOLOGÍA

La lista de materiales principales usados en el desarrollo del proyecto se muestra en la Tabla 1.

TABLA I. MATERIALES DE ELABORACIÓN

Cantidad	Componente
1	Arduino Nicla Vision
1	Arduino Nano ESP32
1	Pantalla OLED
3	Botones touch
3	Lentes focales
1	Pila litio 3.7V 2000mA
1	Pila litio 9V 200mA
1	Módulo de carga

Para la creación de los lentes traductores el proyecto se dividió en cuatro secciones: Creación de una interfaz digital, implementación de una red neuronal para clasificación de

imágenes, principios de óptica y, modelado 3D y circuito físico. Observe la Fig. 1 para el funcionamiento de los lentes inteligentes.

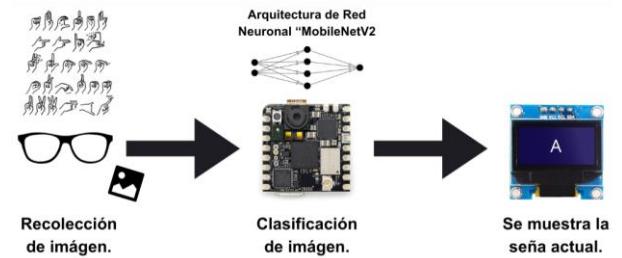


Fig. 1. Diagrama de funcionamiento.

A. Creación de una interfaz digital

La creación de una interfaz digital le permite al usuario adaptar el dispositivo a sus necesidades. La interfaz digital fue diseñada de manera que sea fácil de navegar y visualmente atractiva para el usuario. Para ello, la interfaz fue animada con el programa de edición After Effects, posteriormente se exportó cada fotograma como una imagen para ser visualizada en la pantalla OLED.

La interfaz cuenta con tres opciones: "Traducir", "Ajustes" y "Aprendiendo". Cada opción posee un submenú específico para cada una. Los usuarios pueden navegar a través de la interfaz con ayuda de tres botones sensibles al tacto. El primer botón desplaza al usuario a la derecha, el segundo confirma la selección actual y el tercero desplaza hacia la izquierda. Dependiendo de la opción seleccionada el usuario es capaz de realizar distintas acciones: En el submenú de "Traducir", el usuario convierte señas a texto, gracias a la cámara del microcontrolador. En la sección de "Ajustes", el usuario configura sus preferencias y en la última sección de "Aprendiendo", es posible jugar un minijuego que fomente el aprendizaje de la lengua de señas mexicana al proporcionar un estímulo visual que proporcione retroalimentación constante.

B. Implementación de una red neuronal.

La detección y clasificación de imágenes para el funcionamiento de los lentes inteligentes funciona por la cámara y procesador del microcontrolador Nicla Vision. Ya que la cámara captura imágenes en tiempo real.

Para la clasificación de los datos recolectados se hizo uso de una arquitectura de red neuronal profunda conocida como "MobileNetV2", la cual se caracteriza por ofrecer un equilibrio entre poder y eficiencia, lo que logra que sea usada en dispositivos de bajo poder como teléfonos móviles y microcontroladores. Esta arquitectura introduce dos nuevos conceptos: un residual inverso y un cuello de botella lineal. Este bloque captura características finas de manera eficiente mientras se mantiene una buena eficiencia computacional [11].

La recolección de datos y entrenamiento de la red neuronal fue realizada en la plataforma "Edge Impulse". Esta plataforma fue creada específicamente para facilitar el desarrollo, implementación y escalabilidad de aplicaciones de Machine Learning embebidas. Posterior a la recolección y entrenamiento de datos, la red neuronal con la arquitectura MobileNetV2, apta

para microcontroladores, puede ser fácilmente exportada como una librería para Arduino gracias a Edge Impulse. La arquitectura de la red neuronal se presenta en la Fig. 2.

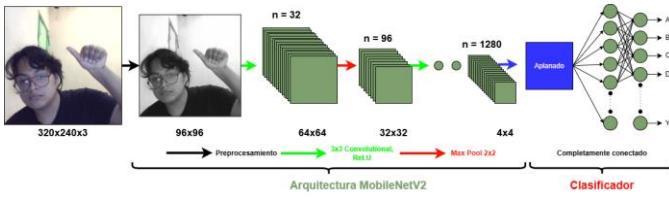


Fig. 2. Arquitectura de MobileNetV2.

C. Principios de óptica

El primer paso en el diseño del sistema óptico es elegir un cristal que refleje la luz proveniente de la pantalla OLED, para ello el cristal debe cumplir con los principios de la ley de reflexión como se aprecia en la Fig. 3, que establece que el ángulo de incidencia es igual al ángulo de reflexión. La incidencia es la onda de luz que llega a la superficie reflectante, mientras que la reflexión es la onda de luz que rebota en la superficie reflectante, este reflejo redirige la luz de la pantalla OLED hacia el cristal de los lentes, proyectando la imagen para que el usuario la perciba de manera clara.

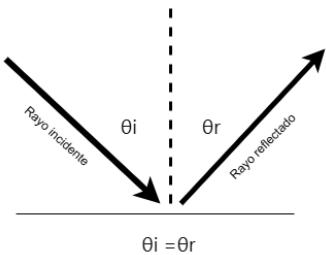


Fig. 3. Diagrama de principio de reflexión.

Para ello se optó por buscar un cristal con revestimiento semitransparente, como un vidrio de alta transparencia con una capa dieléctrica o un revestimiento de espejo parcial. Este tipo de cristal logra que una parte de la luz se refleje mientras que el resto se transmite, logrando un efecto de proyección.

Una vez con el cristal seleccionado se propone un ángulo de incidencia y se alinea la pantalla OLED basado en el área donde queremos que se proyecte la imagen Fig. 4.

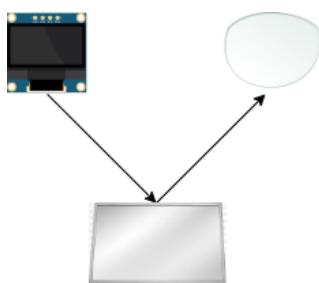


Fig. 4. Diagrama de sistema óptico de proyección.

D. Modelado 3D y circuito físico

En el diseño digital del adaptador (Fig. 5) se observan distintas aperturas, las cuales ayudan a integrar los componentes necesarios como: la cámara del Nicla Vision, la pantalla OLED, los botones touch, el adaptador de carga, etc. Además, el diseño incluye ganchos para asegurar el dispositivo a la montura de las gafas, garantizando una mejor estabilidad.

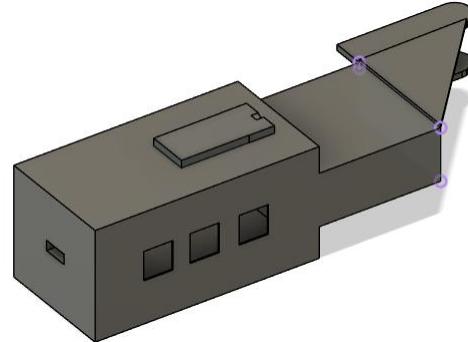


Fig. 5. Diseño digital del adaptador.

El circuito (Fig. 6) consistió en el microcontrolador “Arduino Nicla Vision” y “Arduino Nano ESP32”, tres sensores touch, dos pilas de litio, 1 módulo de carga y la pantalla OLED. En la Fig 6. es posible observar la conexión de los componentes electrónicos.

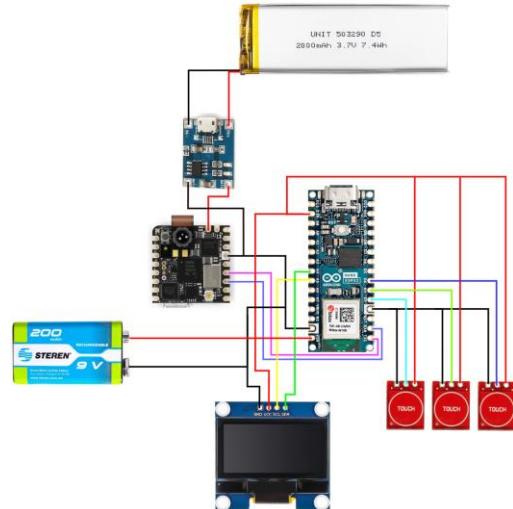
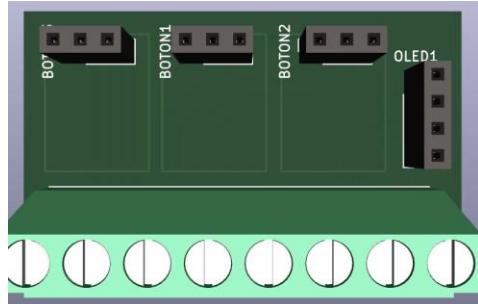
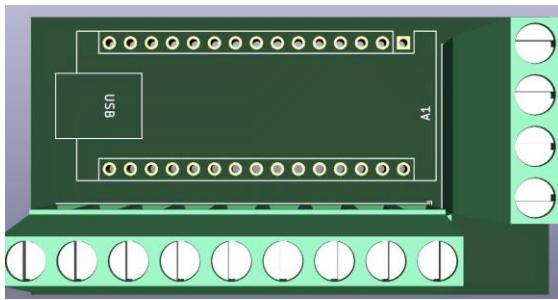


Fig. 6. Diseño de circuito digital.

La realización de una placa de circuito impreso (PCB) se dividió en dos partes: “circuito derecho (Fig. 7a)” y “circuito izquierdo (Fig. 7b)”. Esto representa el lugar donde se ubican en el adaptador de la impresión 3D. Para esto, se diseñaron dos PCBs en el software KiCAD.



a.



b.

Fig. 7. a. Diseño de PCB derecho. b. Diseño de PCB izquierdo.

La PCB del lado derecho se encarga de alimentar al Nicla Vision para el reconocimiento de gestos en tiempo real, mientras que la del lado izquierdo se encarga de manejar el Arduino Nano ESP32 para controlar el menú. Sin embargo, ambas son conectadas a través de una conexión llamada “Universal Asynchronous Receiver/Transmitter” (UART), el cuál es un protocolo para intercambiar información entre dos dispositivos usando una comunicación serial. Esta conexión UART permite comunicar ambos dispositivos a pesar de la diferencia en lenguajes de programación, debido a que el Nicla Vision usa MicroPython y el Arduino Nano ESP32 C/C++.

Cuando el usuario entra en el apartado de “Traducir”, se manda una señal hacia el Nicla Vision que permite la inicialización del modelo de detección de gestos en tiempo real. Cuando el usuario sale de este apartado, se manda otra señal para detener el proceso del Nicla Vision y así evitar que siempre esté capturando, lo que puede ocasionar un sobrecalentamiento.

III. RESULTADOS

A. Interfaz digital

Los resultados de este proyecto incluyen la implementación de una interfaz en una pantalla OLED (Fig. 8).



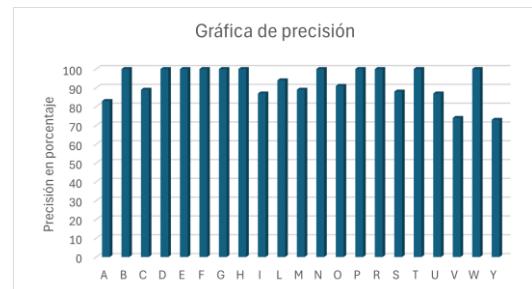
Fig. 8. Interfaz principal en pantalla OLED.

En el apartado de “Traducir” el usuario es capaz de captar los gestos de la dactilología mexicana en tiempo real y convertirlos a texto. El submenú de “Ajustes” ajusta los lentes a las necesidades del usuario, ya sea el usuario un oyente o no-oyente. El último submenú “Aprendiendo” funciona como un minijuego, en el que se irá retroalimentando el conocimiento del usuario acerca de la dactilología mexicana. El juego consiste en que el usuario debe de realizar la seña correspondiente a la letra mostrada en pantalla y mostrar el gesto hacia la cámara.

B. Red neuronal

La red neuronal clasifica 21 etiquetas, las cuales corresponden a las letras del abecedario español, sin contar las que requieren de movimiento. Se logró obtener un porcentaje de precisión de 92% sin contar el fondo.

La red neuronal fue entrenada como un modelo cuantizado, lo que mejora el uso de memoria en el microcontrolador al disminuir la precisión de los pesos y sesgos de la red neuronal. Además, se alcanzó un porcentaje en la métrica de Recall del 91%, lo cual demuestra que, en la mayoría de los casos, se detectan los casos positivos a diferencia de falsos negativos. Mientras que el F1-Score alcanzó un total de 91.4%, el cual es una mezcla entre la precisión y Recall.



a.



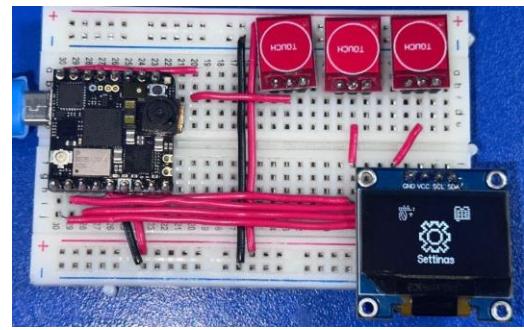
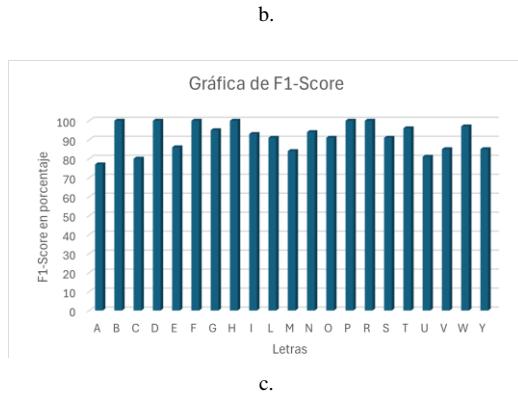


Fig. 11. Circuito físico final.

IV. CONCLUSIONES

C. Principios de óptica

Se aplicaron principios de óptica para lograr una proyección funcional. La pantalla OLED genera la representación inicial, cuya luz, al reflejarse en un espejo posicionado con un ángulo de 45°, sigue la ley de reflexión para redirigir la imagen hacia el lente de Fresnel. Este lente utiliza el principio de refracción para desviar y amplificar los rayos luminosos, haciendo que se aumente el tamaño y la claridad de la proyección. Posteriormente, la representación ampliada se dirige a una hoja de acrílico, permitiendo su visualización a través de los lentes.

Este diseño fue seleccionado de esa manera porque optimiza el uso de componentes compactos y garantiza una correcta visualización, ampliando la información proyectada sin distorsiones significativas, por ende, mejora la interacción del usuario con el sistema.

D. Diseño 3D y circuito físico

El diseño de la carcasa se desarrolló en Fusion 360 (Fig. 10), en el cual se integrará, tanto el software, como el hardware necesario para cumplir con el propósito principal del prototipo. El adaptador tiene una estructura compacta y ergonómica, que está diseñada para adaptarse de manera segura a distintos tipos de lentes, fue pensado de esa manera para brindarle al usuario accesibilidad y comodidad.

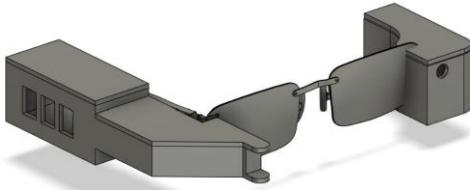


Fig. 10. Diseño digital del adaptador.

El circuito en físico final puede ser observado en la Fig. 11.

El presente artículo demuestra la viabilidad de utilizar tecnologías de visión por computadora y redes neuronales para abordar problemas complejos como el reconocimiento de gestos en tiempo real. La implementación de un modelo de clasificación basado en Edge Impulse, junto con la conexión UART para la interacción entre dos microcontroladores fue clave para optimizar los procesos y evitar la saturación del Nicla Vision.

El prototipo desarrollado logró traducir correctamente las señas de la dactilología mexicana, haciendo uso de la arquitectura MobileNetV2 para optimizar el espacio ocupado en la memoria del Nicla Vision. Además, se diseñó una interfaz digital interactiva para mejorar la comodidad del usuario. Sin embargo, las letras J, K, Ñ, Q, X y Z no fueron incluidas debido a que requieren de movimiento para poder ser identificadas, lo que requeriría de algoritmos de series de tiempo.

Este proyecto, a pesar de estar en etapas tempranas de desarrollo, tiene un gran potencial para eliminar las barreras de comunicación entre las personas oyentes y no oyentes. Así como también ser una gran herramienta para fomentar el aprendizaje de la lengua de señas mexicana. Se identificaron áreas de mejora como la implementación de otras lenguas de señas para ampliar el alcance del prototipo. Además, también está la posibilidad de convertir los lentes inteligentes en un traductor universal capaz de procesar tanto señas como lenguaje hablado en tiempo real.

El circuito y comunicación entre los microcontroladores también puede ser mejorado para eliminar la conexión física de los microcontroladores por la conexión UART. Para esto, hacer uso del bluetooth integrado del Nicla Vision y Arduino Nano ESP32 es esencial para lograr una conexión inalámbrica. El uso de bluetooth también permitiría conectar una aplicación móvil, expandiendo las capacidades del dispositivo con algoritmos avanzados para reconocimiento de audio y procesamiento de imágenes.

Para asegurar la implementación del prototipo en el mundo real, será fundamental colaborar con la comunidad Sorda y con instituciones educativas especializadas en la enseñanza de la lengua de señas. Este enfoque garantizará que el diseño y las funcionalidades del prototipo respondan a las necesidades específicas y que las pruebas se realicen en un entorno inmersivo.

Este proyecto tiene el potencial de cambiar la vida de la comunidad sorda al eliminar las barreras de comunicación y promover un entorno más inclusivo para todas las personas.

RECONOCIMIENTOS

Los autores desean expresar su agradecimiento al Profesor R. Nava Martínez por su valioso apoyo en la revisión de este artículo. Sus observaciones y sugerencias fueron fundamentales para mejorar las ideas y la calidad de este trabajo.

REFERENCIAS

- [1] Herrera Fernández, V. (2014). Alfabetización y bilingüismo en aprendices visuales. Aportes desde las epistemologías de sordos. *Educación y Educadores*, XVII(1), 135-148. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] Organización Mundial de la Salud. (2024). Sordera y pérdida de la audición. Obtenido de Organización Mundial de la Salud: <https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss#:~:text=M%C3%A1s%20de%205%25%20de%20la,una%20de%20cada%20diez%20personas>. K. Elissa, "Title of paper if known," unpublished.
- [3] Gobierno de México. (2021). Con discapacidad auditiva, 2.3 millones de personas: Instituto Nacional de Rehabilitación. Obtenido de Gobierno de México: <https://www.gob.mx/salud/prensa/530-con-discapacidad-auditiva-2-3-millones-de-personas-instituto-nacionalderehabilitacionidiom=es#:~:text=En%20M%C3%A9jico%20>
- %20aproximadamente%202.3%20millones,ciento%20son%20ni%C3%B3n%C3%A1s%20y%20ni%C3%81os
- [4] Pérez, O. (2014). Las personas sordas como minoría cultural y lingüística. Universidad Carlos III de Madrid, 15. ISSN 1989-7022
- [5] Bantupalli, K., & Xie, Y. (2018). American Sign Language Recognition using Deep. 92018 IEEE International Conference on Big Data (Big Data). doi:10.1109/BigData.2018.8622141
- [6] Hays, P., Ptucha, R., & Melton, R. (2013). Mobile device to cloud co-processing of ASL finger spelling to text conversion. IEEE. doi:10.1109/WNYIPW.2013.6890987
- [7] Plazas López, J. A., Gutiérrez Leguizamón, J. J., Suárez Barón, M. J., & González Sanabria, J. S. (2022). Reconocimiento de lengua de señas colombiana mediante redes neuronales convolucionales y captura de movimiento. Redalyc. doi:10.14483/22487638.19213
- [8] C. H. Wang, H. Y. Lee, and J. M. Hu(2021) “Real-time sign language recognition with the help of deep learning,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 45-56. doi: <https://doi.org/10.1007/s12652-020-02394-1>
- [9] Wang, L., Zhao, Z., & Zhang, Y. (2022). Augmented reality glasses for sign language translation. *IEEE Transactions on Human-Machine Systems*, 52(4), 621-630. <https://doi.org/10.1109/THMS.2022.3167274>
- [10] Papatsimouli, M., Sarigiannidis, P., & Fragulis, G. F. (2023). A Survey of Advancements in Real-Time Sign Language Translators: Integration with IoT Technology. *Technologies*, 11(4), 83. DOI: <https://doi.org/10.3390/technologies11040083>
- [11] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 4510-4520. doi:10.1109/CVPR.2018.004