

Desarrollo de una App para la interpretación del alfabeto de señas a través de Machine Learning

Development of an App for the interpretation of the sign alphabet through Machine Learning

Marco Antonio Aguilar-Cortés^{1*}, Miguel Aguilar-Cortés², Jacobo Robles-Calderón³,
Nicolas Moreno-Durán⁴,

¹ Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán

² Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán

³ Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán

⁴ Tecnológico Nacional de México, Instituto Tecnológico Superior de Teziutlán

*Corresponding author: marco.ac@teziutlan.tecnm.mx

Resumen: Actualmente, en México existen aproximadamente siete millones de personas que sufren alguna discapacidad. De dicha población, al menos dos millones tienen problemas de habla, escucha o comunicación. La presente investigación muestra los resultados de crear y entrenar una red neuronal convolucional para la realización de predicciones de letras del alfabeto de señas a partir de imágenes. El modelo resultante se exportó a una versión compatible con teléfonos inteligentes con la finalidad de utilizar la cámara para realizar la captura de imágenes, así como para la realización de predicciones en tiempo real. Como parte del desarrollo de la investigación se aplicó una metodología de seis etapas que inicia con la codificación de un sitio web para la recolección de fotografías e imágenes y finaliza con la implementación de la aplicación móvil con el modelo integrado. Las tecnologías que se utilizaron fueron el Framework de Flutter, ASP.Net Core, Python, Jupyter Notebooks, Tensorflow, Keras, Opencv y MediaPipe. De esta forma se demuestra que el Machine Learning apoya al desarrollo de nuevas herramientas como el desarrollo de aplicaciones que contribuyen a que los grupos vulnerables se integren a la sociedad erradicando la discriminación y ofreciendo una mejor calidad de vida.

Palabras clave: Aplicación, Machine Learning, Lenguaje de Señas.

Abstract: Currently, in Mexico there are approximately seven million people who suffer from a disability, and of that population, at least two million have problems speaking, listening and communicating. This research shows the results of creating and training a convolutional neural network that makes predictions of images related to letters of the sign alphabet. The resulting model was exported to a version compatible with smartphones in order to use the camera to capture images, as well as to make predictions in real time. As part of the development of the research, a six-stage methodology was applied that begins with the coding of a website for the collection of photographs and images and ends with the implementation of the mobile application with the integrated model. The technologies used were the Flutter Framework, ASP.Net Core, Python, Jupyter Notebooks, Tensorflow, Keras, OpenCV and MediaPipe. In this way, it is shown that Machine Learning supports, through the development of new technologies, the integration of vulnerable groups into society, eradicating discrimination and offering them a better quality of life.

Keywords: Application, Machine Learning, Sign Language.

1 Introducción

Escuchar es uno de los cinco sentidos con los que una persona cuenta para llevar a cabo una vida normal, sin embargo, muchos seres humanos nacen o adquieren una discapacidad auditiva o de habla por diversas circunstancias.

Como menciona la Organización Mundial de la Salud (2021) la pérdida de audición y sordera es cuando el sentido del oído no es normal, es decir, su umbral de audición en ambos oídos es menos de 20dB. Ésta puede ser leve, moderada o grave. Sus causas pueden ser congénita, hereditaria o genética, siendo ésta, la más importante y poco previsible; también se adquiere por problemas de partos anormales, causa fetal o materna; por otitis media y meningitis bacteriana, éstos producen un deterioro paulatino de la audición al igual que ruidos de alta intensidad.

Con base a lo mencionado con anterioridad se puede apreciar que esta discapacidad tiene efectos importantes en las etapas tempranas de la vida de una niña o un niño, ya que afecta su pensamiento, habla, lenguaje, conducta, desarrollo social y emocional, así, como su desempeño escolar y posteriormente laboral.

En todo el mundo, más del 5% de la población, es decir, 430 millones de personas padecen de una pérdida de audición, en el cual, también se tiene un estimado para el 2050 que la cifra puede superar los 700 millones, estadísticamente 1 de cada 10 personas. (OMS, 2021).

En México, según (SSA, 2021) aproximadamente 2.3 millones de personas padecen de discapacidad auditiva, de las cuales, más del 50 por ciento son mayores de 60 años; poco más del 34 por ciento tienen entre 30 y 59 años y cerca del 2 por ciento son niñas y niños. Estadísticamente hablando, de un total de 126,014,024 personas en México, (INEGI, 2020) aproximadamente el 1.8% de la población de México sufre esta discapacidad auditiva.

Basado en el bajo porcentaje que hay en discapacidad auditiva en México, no se le ha dado la importancia requerida a ese 1.8% de la población, el cual debe pasar por muchos

problemas para integrarse en los entornos, educativo, social y laboral.

Algunos tipos de discapacidad como la ceguera en la cual una persona invidente va por la calle con un bastón, un perro lazarillo o presenta diferencia en la anatomía física de sus ojos, son más fáciles de identificar por las personas a su alrededor y pueden reaccionar concediéndole libertad de tránsito, ofreciendo ayuda al cruzar una vialidad o le cediéndole un asiento en el transporte público. En el caso de las personas sordas la realidad es completamente diferente pues al no presentar un síntoma fácilmente perceptible de su discapacidad la cual se manifiesta principalmente al momento de intentar establecer comunicación con personas en su entorno genera un gran problema pues un muy bajo índice de la población conoce del Lenguaje de señas. Dicha situación los obliga a aislarse y a convivir predominantemente solo con algunos miembros de su familia o con otros sordos, esto genera entre ellos una cultura apartada del resto, con su propio idioma y sus propios códigos.

Una de las principales adversidades que enfrentan las personas con problemas con sordera es el acceso a la educación, debido a que, en las escuelas no implementan políticas y estrategias de inclusión que les permita a personas discapacitadas poder participar del proceso educativo con normalidad y con ello lograr desarrollar competencias esenciales como la lectura y la escritura. Esta situación se presenta en gran medida por la falta de preparación de los docentes para cubrir las necesidades académicas de los alumnos con discapacidad auditiva. De igual manera la mayoría de las instituciones educativas no cuentan con elementos tecnológicos para contribuir a la formación académica y social de personas sordas y contribuir a que puedan ser parte activa de la sociedad.

Con relación al aprendizaje de lenguaje de señas (LSM), se han registrado un aproximado de solo 40 intérpretes certificados en Lengua de Señas Mexicana, de los cuales la mayoría corresponde a familiares de los discapacitados. En el aspecto académico, la SEP pide a las escuelas capacitación a los maestros para que aprendan a comunicarse con estudiantes que padecen discapacidad auditiva, aunque éste no es obligatorio, por lo tanto, esto no es una situación

que sea preponderante dentro de las instituciones educativas en todos los niveles.

El presente trabajo tiene como finalidad proponer una nueva herramienta tecnológica que permita realizar predicciones de la Lengua de Señas a través de una aplicación móvil que implementa una red neuronal convolucional. Como parte del trabajo uno de los factores más importantes fue la obtención de un modelo entrenado a partir de la realización de pruebas con imágenes agregadas directamente al editor. Posteriormente, el modelo se exportó una versión lite que tiene compatibilidad con el desarrollo de aplicaciones móviles. Para integrarlo se utilizó un Framework que tiene compatibilidad con el desarrollo de aplicaciones para iOS y Android y de esta forma abarcar los dos sistemas operativos con mayor uso en la actualidad a partir del desarrollo de una sola App.

Los resultados obtenidos con el desarrollo del proyecto permiten concluir que la aplicación puede brindar apoyo a las personas que sufren discapacidad auditiva o de habla a poder comunicarse con los demás logrando una integración a la sociedad y poder tener con ello una mejor calidad de vida.

2 Trabajos relacionados

Con respecto a los trabajos relacionados, existen diferentes investigaciones que han propuesto soluciones en el proceso de enseñanza - aprendizaje del lenguaje de señas, así como poder realizar la interpretación en tiempo real de las señas principales y del alfabeto.

Existe una herramienta que fue implementada a través de realidad virtual para el aprendizaje de la Lengua de Señas Mexicana (LSM). Se comienza realizando modelos de VR (realidad virtual) sobre los movimientos hechos por personas con discapacidad auditiva para crear un modelo. Una vez que se haya formado el modelo, se pueden manipular variables como tiempo, contrastes y figuras para poder obtener una interpretación más exacta de los valores que se quiere aprender, haciendo que el usuario pueda ir aprendiendo a su propio ritmo. (Borja Araiza, Araujo Díaz, & Méndez Segundo, 2019).

Por otra parte, se utilizó el software LESCO (el cual ayuda a crear un avatar con rasgos

humanos y con la facilidad de poder manipular tamaños, movimiento, etc.) se configuraron para poder interpretar algunas señas que fueran de ayuda para personas con esta discapacidad. Se consideraron 60 personas de edades entre 18 a 55 años.

Soto et al. (2019) crea una herramienta para la comunicación y organización de las actividades en clase como apoyo en la integración de alumnos con discapacidad en el habla. Los autores hacen alusión sobre lo difícil que puede ser para algunos estudiantes con problemas del habla, el poder participar adecuadamente en el proceso de enseñanza - aprendizaje. Derivado de ello, se propone la herramienta de aula invertida en conjunción de las tecnologías de la información y comunicación que favorezcan la participación integral y total de estos discentes con dicha discapacidad.

Otros autores han definido una aproximación a la gramática de la Lengua de Señas a través de un avatar. Dicho sistema conglobera un compendio de señas con base al lenguaje de señas costarricenses; con ayuda de un equipo de personas con diferentes discapacidades, ingenieros en software, un filósofo y un pedagogo, realizan a Juan dentro de un software: un avatar capaz de ser programado y que puede responder a las necesidades de movimientos de manos y utilización de rasgos como mano predominante y gestor articulados. aseguran que esta propuesta ha elevado la calidad de comprensión de diferentes alumnos. (Flores Solórzano & Chacón Rivas, 2019).

Jerónimo et al. (2016) crea un sistema de traducción de lenguaje de señas en tiempo real enfocado únicamente para dispositivos Android que ofrece una forma sencilla de comunicación con personas carentes de capacidad auditiva, utilizando como base la Lengua de Señas Mexicana. La aplicación está enfocada en funcionar como herramienta de consulta en el proceso de comunicación del usuario hacia la persona con limitaciones auditivas, por lo que presenta una completa compatibilidad con el Diccionario Español de Lengua de Señas mexicana publicado por la Secretaría de Educación Pública. Asimismo, la aplicación implementa la disciplina de la visión por computadora a través de la cámara del dispositivo móvil para realizar la traducción en

el proceso de comunicación de la persona con limitaciones auditivas hacia el usuario.

Trujillo (2021) implementa un sistema basado en Kinect para el reconocimiento de Lengua de Señas Mexicana con base a la trayectoria 3D y algoritmo KNN con medio de filtrado. Para la clasificación de las palabras se utiliza una red neuronal de tipo perceptrón multicapa entrenada con el algoritmo Backpropagation, obteniendo como resultado una media de 93.46% de precisión.

También existen propuestas que se desarrollaron a través de un sistema web colaborativo, tal como lo menciona Ramírez et al. (2022) que permita buscar señas con entradas de texto y encontrar videos representativos de la búsqueda. De esta forma, la investigación apoya a quienes buscan establecer comunicación con personas con discapacidad auditiva. El sitio es colaborativo ya que las personas Sordas y no Sordas pueden colaborar agregando su conocimiento sobre LSM. Con esta propuesta, se pretende generar un repositorio de contenido audiovisual que permita explotar otros aspectos de la comunicación, como la conversión de videos con señas a texto o sonido, pudiendo ser en videos grabados u en tiempo real a través de dispositivos móviles.

Como se puede observar, existen trabajos que han implementado diferentes formas de trabajar con la Lengua de Señas a través de diferentes tipos de aplicaciones, sin embargo, la propuesta de esta investigación utiliza redes neuronales convolucionales creadas con la tecnología de TensorFlow, una herramienta que facilita la creación de modelos de aprendizaje automático para aplicaciones de escritorio, así como móviles, web y nube (Tensorflow, s.f.).

Otro beneficio que se utilizó de Tensorflow es poder trabajar con un modelo más ligero que permita la compatibilidad con aplicaciones móviles. De esta forma, se pudo integrar a una aplicación desarrollada con el Framework de Flutter y una vez finalizada, poder ser exportada a los sistemas operativos más utilizados como lo son iOS y Android.

3 Metodología

Como menciona Hernández et al. (2010) es necesario realizar el diseño de investigación con el fin de poder visualizar la manera práctica y concreta de poder resolver las preguntas de investigación y de demostrar los objetivos: derivado a la naturaleza de la investigación, la metodología que acompaña dicho trabajo es una metodología descriptiva, esto debido a la descripción que se realiza en la ejecución de la solución el proyecto: con respeto al diseño de investigación se considera una investigación experimental (debido a que se realiza la manipulación deliberada de las variables), longitudinal (debido a que se reúne información en diferentes momentos de manera que considera tendencias o aspectos relacionados al tiempo) y prospectivos (debido a que se observan diferentes causas presumibles y se avanzan longitudinalmente con el fin de observar sus consecuencias).

En lo que se refiere a la metodología tecnología aplicada al desarrollo de la herramienta de recolección e interpretación, se propuso un modelo basado en seis etapas donde se inicia con la recolección de imágenes y finaliza en la implementación del modelo obtenido de la red neuronal en la aplicación en un dispositivo móvil. En la figura 1, se muestra de manera gráfica el proceso realizado.



Figura 1. Proceso de implementación de una red neuronal convolucional a un dispositivo móvil

3.1 Recolección de imágenes

Durante la primera etapa de la metodología se desarrolló una aplicación web para que un

conjunto de personas realizaran el proceso de captura y envío de imágenes al servidor, ya sea desde una webcam o de manera tradicional por medio de un formulario de registro dentro de la aplicación.

Para cada usuario registrado, se crea un espacio en el servidor con diferentes carpetas donde cada una representa cada una de las letras. No existe límite para el número de imágenes ni tampoco se restringe el tamaño. En la figura 2, se visualiza como el usuario registra la letra en el servidor.

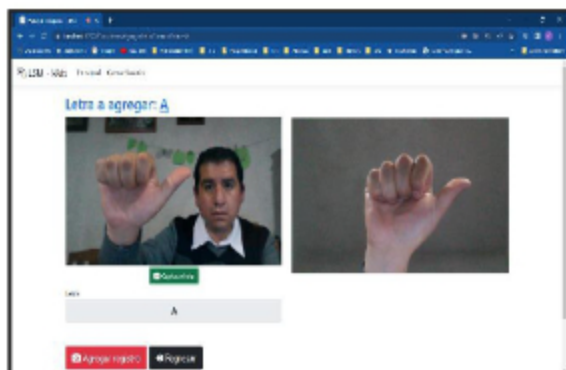


Figura 2. Recolección de imágenes por parte del usuario

Un punto importante para considerar es que para entrenar una red neuronal y garantizar que el modelo resultante sea confiable, se requiere de una gran cantidad de imágenes para que las predicciones se aproximen a los resultados esperados. En la figura 3, se muestra la interfaz gráfica de la aplicación que el usuario visualiza en la aplicación web y en la figura 4 se muestra un esquema de como las imágenes son almacenadas y organizadas.

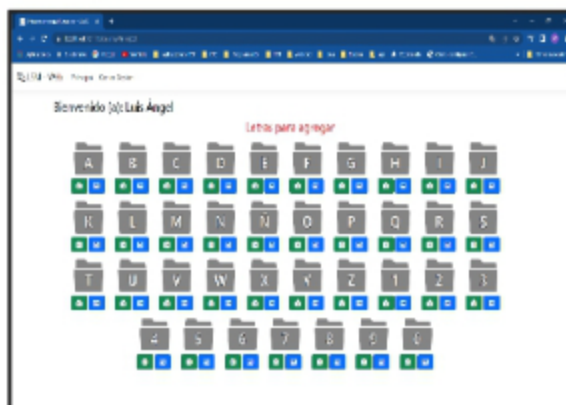


Figura 3. Interfaz principal de la aplicación web para la recolección de imágenes de señas.

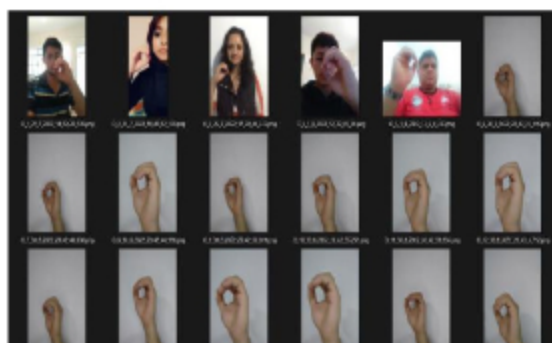


Figura 4. Imágenes recolectadas por diferentes usuarios que utilizan la aplicación

3.2 Procesamiento y agrupamiento

En lo que concierne al agrupamiento, si la persona encargada reunía todas las imágenes de todos los usuarios que registran fotografías se haría un proceso muy complejo que requeriría una mayor cantidad de tiempo para la conformación del banco de datos. Por esa razón, se desarrollaron funciones recursivas que permitieran la recolección de todas las imágenes de todos los usuarios y las agrupara en carpetas por letras. De esta forma, se generó el dataset a utilizar para el entrenamiento del modelo de la red neuronal y se pudiera proceder a la realización de las predicciones. Es importante comentar que algunos usuarios tomaron fotografías desde otros ángulos o con algún objeto que pudiera afectar al entrenamiento del modelo. Por esa razón se realizó una depuración de las ilustraciones que no cumplieran con lo requerido. Este proceso si se tuvo que realizar manualmente debido a que se tuvo que verificar imagen por imagen para tomar la decisión si se agregaba al dataset o se desechaba.

3.3 Entrenamiento

Al finalizar el proceso de agrupamiento de imágenes, se procedió a realizar el desarrollo y entrenamiento de la red neuronal. Para este procedimiento se utilizó como lenguaje de programación Python, las herramientas de Tensorflow y Keras para el desarrollo de la red neuronal y como entorno de desarrollo al Proyecto Jupyter Notebook (por cuestiones de la gestión de archivos, no se utilizó Google Colab).

3.4 Salida e implementación

Al finalizar el entrenamiento de la red neuronal, se obtiene como resultado un modelo que ya permite realizar predicciones. Para realizar las pruebas correspondientes, se codificó un

programa en Python, que, apoyado por la tecnología de OpenCV y Mediapipe, se logró detectar la mano. Los resultados permitieron comprobar que el modelo estaba funcionando correctamente. En la figura 5 se muestra cómo se realiza la predicción.

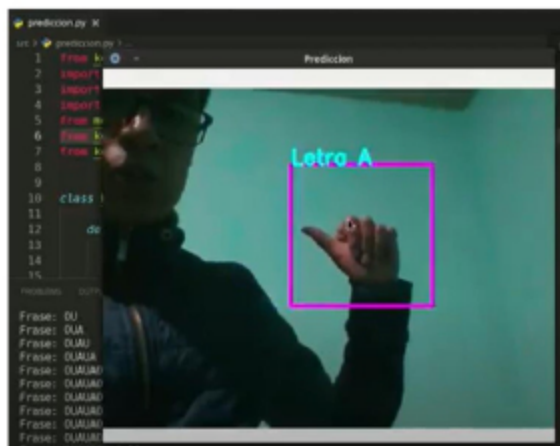


Figura 5. Primeras predicciones por parte del modelo exportado

El objetivo del modelo no fue utilizarlo solamente en Python, era poder implementarlo en dispositivos móviles. Por esa razón, se utilizó Tensorflow Lite para exportar el modelo a una versión más ligera. Dicha versión se puede utilizar en aplicaciones nativas, sin embargo, se buscaba algo más ambicioso: utilizar el modelo en un Framework para el desarrollo de aplicaciones móviles híbridas. Una de las tecnologías que permitía esta implementación es Flutter con un paquete llamado tflite versión 1.1.2.

4 Resultados

Una vez finalizado la codificación, se procede a generar el archivo de instalación con extensión .apk (no se realizaron pruebas en un dispositivo con sistema operativo iOS debido a que no se cuenta con equipo Mac, solo se realizaron las pruebas en Android).

En la figura 6 se puede visualizar la pantalla de inicio de sesión, así como la de información del usuario.



Figura 6. Pantalla inicial e información del perfil de usuario

Con la integración del modelo, la aplicación pudo lograr con éxito la interpretación de algunas letras como las vocales. Así mismo, se desarrolló otros módulos para continuar con la generación de imágenes, pero ahora con la aplicación móvil. En la figura 7 se muestra el proceso de recolección de imágenes donde se puede utilizar la cámara del dispositivo o también se puede seleccionar algunas fotografías de la galería de imágenes.



Figura 7. Recolección de imágenes desde celular

En la figura 8, se pueden visualizar ejemplos de cómo el modelo, a través de la cámara del dispositivo móvil, puede realizar las predicciones. Cabe mencionar que las letras con las que primero se realizaron las primeras pruebas fueron las vocales.

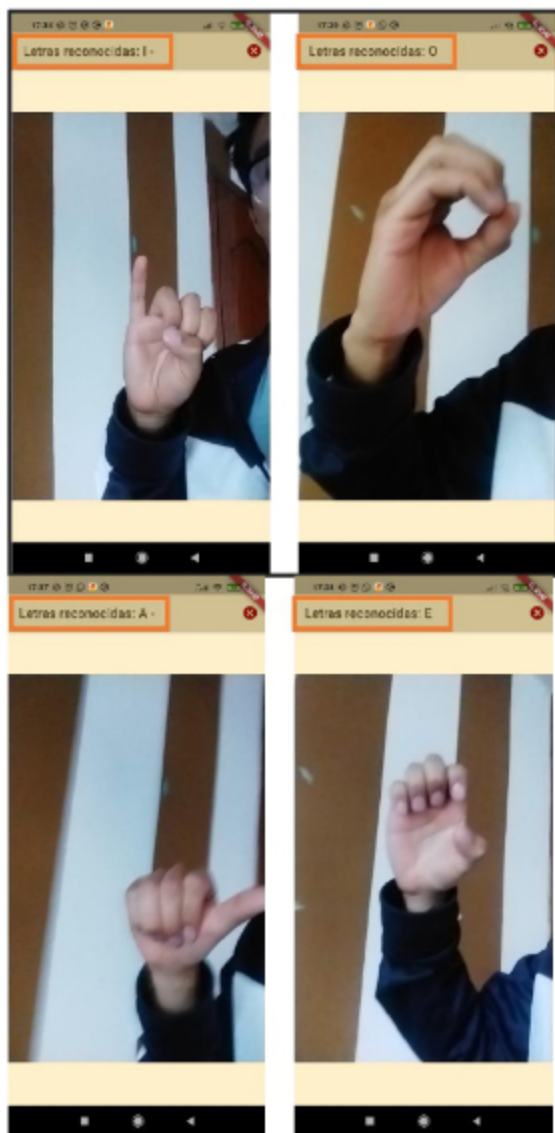


Figura 8. Predicción correcta de las vocales del abecedario en Lengua de Señas

5 Discusión

Como se puede observar, inicialmente las pruebas se realizaron con las vocales (a,e,i,o,u) teniendo como resultado predicciones exactas sobre la letra que se requería interpretar. Es importante señalar que las imágenes fueron sometidas a un filtro para garantizar la calidad de las mismas, así como a seleccionar aquellas que no presentaran problemas con relación al fondo y claridad.

Ahora, teniendo una base más exacta de lo realizado, se procedió a entrenar un modelo más completo que permitió incluir más letras y poder integrarlas a la aplicación móvil y tener la capacidad de interpretar un mensaje en tiempo real que se encuentra en Lengua de Señas.

Desafortunadamente no es posible integrar en estos momentos las letras que requieren movimiento como lo son J, K, Q, X y Z. Para eso se requiere trabajar con otra tecnología debido a que la red neuronal desarrollada solo se encuentra entrenada con imágenes. Es posible que, si se trata de interpretar una letra con movimiento, la predicción presente otros resultados fuera de los rangos esperados.

6 Conclusiones

En función del desarrollo de la propuesta planteada, la creación de nuevas soluciones tecnológicas puede favorecer a grupos vulnerables como lo son las personas que sufren una discapacidad, en especial las que tienen dificultades para hablar y escuchar. La conjunción entre los métodos pedagógicos y las TIC's pueden facilitar los procesos de enseñanza-aprendizaje para los discentes que padecen las deficiencias antes mencionadas. Por otro lado, el esfuerzo de los docentes tiene que motivarlos a poder acercarse cada vez más a la tecnología y a las bondades que esta ofrece en cuanto a inclusión se refiere. Esta propuesta es de suma importancia debido a que cumple con la necesidad de hacer cada vez más inclusivos los centros educativos en México.

7 Trabajo a futuro

Después de realizar las pruebas con un conjunto de letras, se continuará trabajando en el entrenamiento de las demás, así como de aplicar la interpretación de las letras que se requieren movimiento. Así mismo, se analizarán otras tecnologías que utilicen la versión de Tensorflow lite para poder integrarlo con tecnologías más recientes y tenga mayor compatibilidad con los sistemas operativos actuales con los que cuentan los dispositivos móviles. De igual modo cuando se tenga un conjunto de datos depurado y validado por la evaluación del modelo de datos se pretende hacer público el dataset para que sea accesible por investigadores y personas interesadas en el desarrollo de este tipo de herramientas.

Agradecimientos

A las personas que apoyaron en la recolección de imágenes y a los alumnos que participaron en el desarrollo de las diferentes aplicaciones que se realizaron.

Referencias bibliográficas

- Borja Araiza, A., Araujo Díaz, D., & Méndez Segundo, L. (2019). Herramienta en realidad virtual para el aprendizaje del lenguaje de señas mexicano. *Computing Science*, 148(8), 55-61.
- Chávez Sánchez, J., Ramírez Trejo, C., & Chacón Rivas, M. (2019). Interacción con avatar 3D para la creación y edición de poses corporales y faciales en la lengua de señas costarricense LESCO. *Avances sobre reflexiones, aplicaciones y tecnologías inclusivas*, 34-40.
- Cuji, B., Gavilanes, W., & Silva, A. (2018). Aprendizaje del lenguaje de señas mediado por las TIC. *Espacios*, 39(29), 24.
- Flores Solórzano, S., & Chacón Rivas, M. (2019). Aproximación a la Gramática de la Lengua de Señas Costarricense a través de un Avatar. *Avances sobre reflexiones, aplicaciones y tecnologías inclusivas*, 83-87.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la Investigación*. México D.F.: McGRAW-HILL.
- INEGI. (2020). *INEGI*. Obtenido de <https://www.inegi.org.mx/temas/estructura/>
- INEGI. (3 de Diciembre de 2021). *Estadísticas a propósito del Día Internacional de las Personas con discapacidad (Datos Nacionales)*. Obtenido de https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP_PersDiscap21.pdf
- Jerónimo Bañuelos, J., Domínguez García, M. A., & Betanzos Martínez, O. E. (2016). *Destellos de la Ciencia Iluminando el Conocimiento*. Puebla: Universidad Popular Autónoma del Estado de Puebla.
- OMS. (2 de Marzo de 2021). *Mejorar la salud para todos, en todas partes*. Obtenido de Sordera y pérdida de la audición: <https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- Ortiz Farfán, N., & Camargo Mendoza, J. E. (2020). Modelo computacional para reconocimiento de lenguaje de señas en un contexto colombiano. *Tecnológicas*, 23(48), pp. 197 - 232.
- Ramírez Noriega, A., Figueroa Pérez, J. F., Mendivil Torres, J., Ayala Zuñiga, G., & Martínez Ramírez, Y. (2022). Una Propuesta para un Sistema Web Colaborativo para el Lenguaje de Señas Mexicano. *Investigación Básica y Aplicada*, 261-267.
- Soto Muñoz, J. G., Carreño León, M. A., Sandoval Bringas, J. A., Durán Encinas, I., & Sánchez Paz Rubio, A. (2019). Experiencia en la utilización de herramientas tecnológicas para la comunicación y organización de las actividades de clase como apoyo en la integración de alumnos con discapacidad en el habla. *Avances sobre reflexiones, aplicaciones y tecnologías inclusivas*, 77-80.
- SSA. (28 de Noviembre de 2021). *Gobierno de México*. Obtenido de <https://www.gob.mx/salud/prensa/530-con-discapacidad-auditiva-2-3-millones-de-personas-instituto-nacional-de-rehabilitacion?idiom=es#:~:text=En%20M%C3%A9xico%2C%20aproximadamente%202.3%20millones,ciento%20son%20ni%C3%B1as%20y%20ni%C3%B1os>
- Tensorflow. (s.f.). *Introducción a TensorFlow*. Obtenido de <https://www.tensorflow.org/learn?hl=es-419>
- Trujillo Romero, F., & García Bautista, G. (2021). Reconocimiento de palabras de la Lengua de Señas Mexicana utilizando información RGB-D. *Computation e Informática*.