



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



Departamento de Ciencias

CULAGOS/DCET/RPI2025/004



UNIVERSIDAD DE GUADALAJARA

Centro Universitario de los Lagos

División de Estudios de la Biodiversidad e Innovación Tecnológica

Departamento de Ciencias Exactas y Tecnología

INFORME TÉCNICO

PRODUCTIVIDAD INNOVADORA (CREACIONES DE LA MENTE)

**Modelo de utilidad, Marca, Diseño de Herramientas, Hardware y Software y
Denominación de origen.**

CUERPO ACADÉMICO: CALIDAD EDUCATIVA EN LAS INGENIERÍAS

Detección de LSM mediante inteligencia artificial.

AUTORES:

**JIMÉNEZ GUTIÉRREZ, AURIA LUCIA (2309726), ROJAS MENA, EVELYN
(218646366), RUIZ VÁZQUEZ, KEVIN LEONARDO (218471833)**

Lagos de Moreno, Jalisco

Agosto 2025



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



CULAGOS/DCET/RPI2025/004

DETECCIÓN DE LSM MEDIANTE INTELIGENCIA ARTIFICIAL

Jiménez Gutiérrez, Auria Lucia¹, Rojas Mena, Evelyn², Ruiz Vázquez, Kevin Leonardo³

^{1,2,3} Centro Universitario de Los Lagos, Universidad de Guadalajara. Av. Enrique Díaz de León
1144, Paseos de la Montaña, Lagos de Moreno, Jal., 47460, México

auria.jimenez@academicos.udg.mx¹; evelyn.rojas6463@alumnos.udg.mx²;

kevin.ruiz4718@alumnos.udg.mx³

RESUMEN

La convergencia entre la visión por computadora y el impacto social se materializa en el presente desarrollo técnico, el cual propone una solución robusta para la interpretación automatizada de la Lengua de Señas Mexicana (LSM). El sistema recoge 21 puntos importantes de la mano (landmarks) usando la librería MediaPipe y los procesa con una red neuronal profunda hecha en TensorFlow. A diferencia de las soluciones tradicionales que usan lenguajes de señas extranjeros (como ASL o LSF), este proyecto se basa en un conjunto de datos propio de coordenadas normalizadas. Esto asegura una precisión del 90% al 97% en la identificación de señas fijas del alfabeto en tiempo real. La creación de este banco de datos original no solo dota al sistema de una alta fidelidad cultural y postural. También constituye un capital intelectual institucional estratégico, permitiendo la autonomía técnica respecto a APIs externas. Además, sienta una base propietaria para futuras expansiones hacia gestos dinámicos y estructuras gramaticales complejas.

Palabras clave: Lengua de Señas Mexicana (Mexican Sign Language), Visión por computadora (Computer Vision), Inteligencia Artificial (Artificial Intelligence), Red Neuronal (Neural Network), Tiempo Real (Real-time).

Aplicación: El sistema está diseñado para su despliegue en entornos educativos y de servicios públicos en México, facilitando la interacción entre la comunidad sorda (aproximadamente 2.4 millones de ciudadanos) y la población oyente. En el ámbito académico, el proyecto valida el desarrollo de competencias transversales en ingeniería mecatrónica. Específicamente, en las áreas de procesamiento digital de imágenes, control de flujo y diseño de algoritmos de aprendizaje automático (machine learning). Esto demuestra la capacidad de integrar hardware de bajo costo con software de alto rendimiento.



INTRODUCCIÓN

En el panorama sociotecnológico de México, la brecha de comunicación para las personas con discapacidad auditiva persiste como un desafío estructural. A pesar de que la Lengua de Señas Mexicana (LSM) fue reconocida como lengua nacional y patrimonio lingüístico en 2005, la escasez de intérpretes y herramientas tecnológicas de soporte limita la inclusión efectiva en sectores críticos. El presente producto innovador consiste en un sistema de visión por computadora que detecta y traduce gestos manuales a texto, con el objetivo primordial de democratizar el acceso a tecnologías asistivas.

El planteamiento del problema se enfoca en la inaccesibilidad de las soluciones actuales, las cuales suelen depender de sensores costosos (como cámaras de profundidad o guantes de datos) o de modelos optimizados para lenguas extranjeras. Esta propuesta rompe dicha barrera al posicionarse como una solución de "bajo costo y fácil implementación" que utiliza una cámara web convencional. Al recuperar los antecedentes de la LSM y ajustarlos a una arquitectura de inteligencia artificial actual, el trabajo proporciona una herramienta útil que va más allá de la teoría y ofrece una aplicación práctica inmediata en lugares con recursos limitados.

MATERIALES Y MÉTODOS

Se registraron secuencias de video y capturas de pantalla para documentar los aciertos y las situaciones en las que el sistema mostró ambigüedad, especialmente en poses con dedos muy próximos entre sí. Antes de la fase experimental, los datos fueron sometidos a un proceso de limpieza (Anexo 3) y normalización (Anexo 4) con el fin de optimizar el entrenamiento del modelo. Este procedimiento permitió constatar que el prototipo es funcional para fines educativos y de divulgación, y que su rendimiento mejora sustancialmente al mantener una postura de mano estable y una iluminación consistente.

Con el fin de cuantificar el desempeño del modelo, se llevó a cabo un análisis de métricas de clasificación considerando los indicadores de precisión, exhaustividad (recall) y F1-Score. Los resultados obtenidos se resumen en la Tabla 1, donde se observa una precisión promedio de 97.39%, un recall promedio de 97.09% y un F1-Score promedio de 97%, lo que confirma la



UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



CULAGOS/DCET/RPI2025/004

capacidad del sistema para reconocer las señales del alfabeto.

Tabla1. Métricas de clasificación del modelo de reconocimiento de señas LSM.

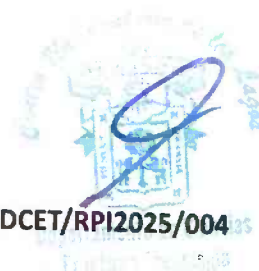
Letra	Precisión	Recall	F1-Score	Support
A	1.00	1.00	1.00	22
B	0.96	1.00	0.98	22
C	0.96	1.00	0.98	22
D	0.92	1.00	0.96	22
E	0.96	0.96	0.96	23
F	1.00	0.96	0.98	23
G	1.00	1.00	1.00	23
H	1.00	1.00	1.00	23
I	1.00	0.91	0.95	23
J	1.00	1.00	1.00	22
K	1.00	1.00	1.00	22
L	1.00	0.91	0.95	23
M	0.88	1.00	0.94	23
N	0.82	1.00	0.94	23
O	1.00	0.96	0.98	23
P	1.00	1.00	1.00	22
Q	1.00	1.00	1.00	23
R	1.00	1.00	1.00	22
S	1.00	0.65	0.79	23
T	0.96	1.00	0.98	22
U	1.00	0.91	0.95	22
V	0.96	1.00	0.98	23
W	1.00	1.00	1.00	23
X	1.00	1.00	1.00	22
Y	0.92	1.00	0.96	22
Z	1.00	1.00	1.00	22

Pra identificar las principales áreas de mejora, se elaboró una matriz de confusión que permite visualizar las clases en las que el sistema presenta mayor confusión. En dicha representación gráfica (Figura 1) se observa que las confusiones más comunes se producen entre las letras con configuraciones manuales similares, como M/N.



UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



CULAGOS/DCET/RPI2025/004

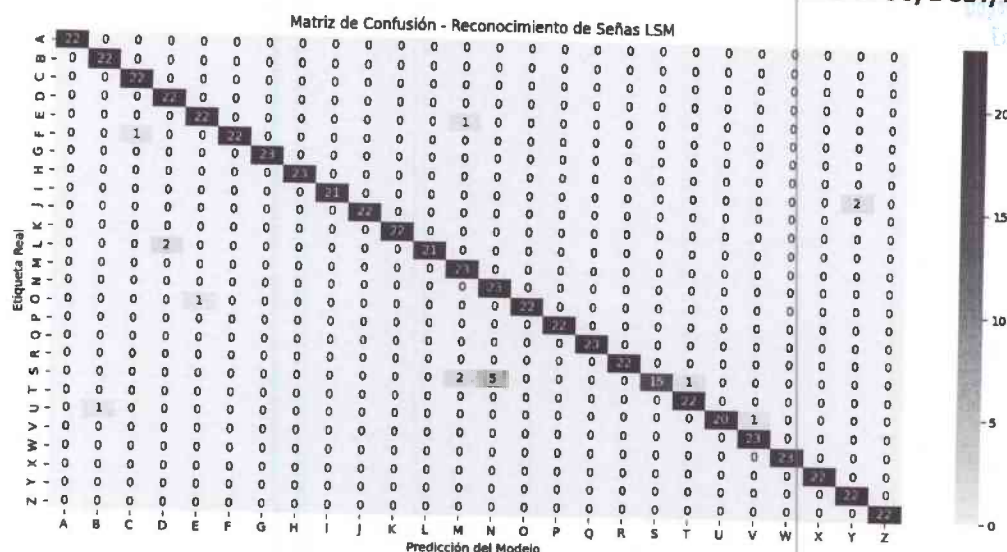


Figura 1. Matriz de confusión del modelado de reconocimiento de señas LSM.

Adicionalmente, se evaluó el rendimiento en tiempo real del prototipo mediante la medición de la latencia de predicción y la tasa de fotogramas por segundo (FPS), con el objetivo de determinar la capacidad del sistema para operar de manera fluida durante la detección continua de señas. Una muestra de resultados obtenidos se presentan en la Tabla 2, donde se registraron valores de latencia que oscilan entre 96.9 ms y 118.4 ms, con un promedio de 108.6 ms y una tasa de procesamiento comprendida entre 8.4 FPS y 10.3 FPS, con un promedio aproximado de 9.2 FPS. Estos valores reflejan un rendimiento estable del sistema, ideal para aplicaciones interactivas en tiempo real.

Tabla 2. Resultados experimentales de latencia y FPS.

Muestra	Latencia (ms)	FPS
1	114.66	8.72
2	109.91	9.10
3	111.44	8.97
4	98.31	10.17
5	110.70	9.03
6	107.19	9.33
7	112.95	8.85
8	106.11	9.42



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



CULAGOS/DCET/RPI2025/004

9	112.67	8.88
10	105.21	9.51
11	118.29	8.82
12	113.35	8.45
13	118.44	8.44
14	96.95	10.31
15	101.27	9.87
16	111.93	8.93
17	107.84	9.27
18	101.06	9.90
19	112.31	8.90
20	98.85	10.12

Este procedimiento demostró que el prototipo funciona bien para la educación y la divulgación, y que su rendimiento mejora mucho al mantener una postura de mano estable y una buena iluminación.

RESULTADOS

El proyecto mostró que es técnicamente viable construir un sistema de reconocimiento de señas estáticas de la Lengua de Señas Mexicana utilizando herramientas de libre acceso y recursos de cómputo convencionales.

La combinación de MediaPipe Hands para la extracción de características y una red neuronal densa en TensorFlow permitió alcanzar predicciones confiables en tiempo real, siempre que las condiciones de captura se mantuvieran dentro de rangos razonables de iluminación y encuadre. La elaboración de un dataset propio resultó determinante para ajustar el modelo a las particularidades de las señas empleadas. Esto, a su vez, incrementó la precisión y redujo la sensibilidad al ruido del entorno.

Desde el punto de vista de la ingeniería mecatrónica, el sistema reúne partes para recoger datos, procesarlos y controlar el flujo, además de incluir un módulo de decisión que utiliza aprendizaje automático, lo que refleja la interdisciplinariedad de esta área.

En términos de alcance, el prototipo ya es presentable y útil para actividades de enseñanza y divulgación; este sistema sirve como base para incorporar más señas, números y gestos dinámicos,



además de su posible adopción en aplicaciones móviles o plataformas educativas.

Finalmente, este prototipo no solo valida un enfoque técnico, sino que representa un avance real hacia herramientas accesibles, inclusivas y de bajo costo, respondiendo directamente al problema social identificado: la necesidad de facilitar la comunicación entre personas sordas y oyentes.

CONTRIBUCIONES

La sistematización de este desarrollo permite identificar hitos fundamentales para la transferencia de conocimiento y la evolución de tecnologías asistivas en la región:

Dataset Institucional Original: Generación de un conjunto de datos específico para LSM que captura variaciones regionales y culturales, eliminando la dependencia de modelos anglosajones.

Eficiencia Computacional: Validación de una red neuronal de alta precisión (>90%) con bajo consumo energético, apta para hardware convencional.

Accesibilidad Económica: Eliminación de barreras de entrada mediante el uso de periféricos estándar (webcam) y software libre.

Escalabilidad Modular: Diseño de una arquitectura capaz de integrarse en sistemas complejos de interpretación de gestos dinámicos y frases completas.

REFERENCIAS

- [1] H. Caballero Hernández, V. Muñoz Jiménez, y M. A. Ramos Corchado, "Desarrollo de software dedicado a la traducción de la Lengua Mexicana de Señas mediante Deep Learning y Machine Learning," Paakat: Revista de Tecnología y Sociedad, vol. 15, no. 28, 2025. [En línea]. Disponible: <https://doi.org/10.32870/Pk.a15n28.897>
- [2] M. E. Rodríguez Trejo, "Modelos computacionales aplicados a la lengua de señas mexicana," Tesis de maestría, Universidad Autónoma del Estado de Morelos, 2023.
- [3] A. Pérez Paredes, "Propuesta al Consejo para Prevenir y Eliminar la Discriminación de la Ciudad de México (COPRED). Reconocimiento oficial de la Lengua de Señas Mexicana como lengua nacional y patrimonio lingüístico de México," 2021.
- [4] Google AI, "Guía de soluciones de MediaPipe," 2025. [En línea]. Disponible: developers.google.com/mediapipe



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LOS LAGOS

División de Estudios de la Biodiversidad e Innovación Tecnológica
Departamento de Ciencias Exactas y Tecnología



CULAGOS/DCET/RPI2025/004

[5]

TensorFlow, "Keras API Documentation," The TensorFlow Authors, 2024. [En línea]. Disponible: www.tensorflow.org/api_docs

Glosario técnico

MediaPipe: Framework de Google para visión por computadora que permite el rastreo de landmarks anatómicos en tiempo real.

Landmarks: Coordenadas tridimensionales (x, y, z) que representan los puntos de articulación y extremos de la mano (21 puntos en total).

ReLU (Rectified Linear Unit): Función de activación que introduce no linealidad en la red neuronal, acelerando el entrenamiento al evitar la saturación de gradientes.

Softmax: Función de activación final que convierte el vector de salida en una distribución de probabilidades para cada clase posible.

Overfitting (Sobreajuste): Error de generalización donde el modelo memoriza el ruido de los datos de entrenamiento en lugar de aprender el patrón subyacente.