

Programa de Maestría en Ingeniería, UABJO. Clave SNP CONAHCYT: 007308

Nombre: López Pérez Tomas

Sexo: Masculino

Edad: 26 años

Nacionalidad: mexicana

Teléfono celular:95133031356

Email: [tomaslopezperez107@gmail.com](mailto:tomaslopezperez107@gmail.com)

Domicilio: Privada de pedro Cruz Lote 4 , Santiago Etlá 68263 San Lorenzo cacaotepec  
Oaxaca.

Numero de matrícula: 104231

Antecedentes académicos:

Licenciatura en computación

Universidad autónoma Benito Juárez de Oaxaca.



UNIVERSIDAD AUTONOMA "BENITO JAUREZ" DE OAXACA

FACULTAD DE SISTEMAS BIOLOGICOS E INNOVACION TECNOLOGICA MAESTRIA EN INGENIERIA

Título: Reconocimiento del lenguaje de señas mexicano  
mediante modelos neuronales de Inteligencia artificial.

Campo de conocimiento: Visión por computador

Nombre del aspirante: Lic. Tomas López Perez

Director de tesis: Dr. José Edgar Lara Ramírez

Vo.Bo

Dr. José Edgar Lara Ramírez



## ANTECEDENTES DE INVESTIGACIÓN

La comunicación es un aspecto fundamental de la experiencia humana, y tradicionalmente el lenguaje hablado ha sido el principal medio para expresar ideas y compartir información. Sin embargo, para aquellos cuyas habilidades de habla o audición están deterioradas, la comunicación puede convertirse en un desafío significativo. [1] Esta barrera es especialmente relevante para las personas sordas y con discapacidad auditiva, cuyas necesidades comunicativas no siempre son adecuadamente atendidas por los medios convencionales.

El lenguaje de señas ha surgido como un medio crucial para estas personas, permitiéndoles expresarse y conectarse con otros de manera efectiva. Basado en el movimiento de las manos y las expresiones faciales, el lenguaje de señas sigue reglas gramaticales específicas y presenta una rica variedad de señas que transmiten significado de manera precisa y compleja.

Desde la trinchera de las tecnologías de la información se han explorado diversas metodologías y técnicas diseñadas para superar las barreras de comunicación que enfrentan las personas sordas y con discapacidad auditiva. Estas metodologías abarcan desde el uso de tecnologías como Kinect hasta la aplicación de técnicas de aprendizaje profundo (deep learning), todas con el objetivo común de mejorar la inclusión y accesibilidad en la comunicación, algunas de ellas son descritas a continuación.

El Kinect, una innovación tecnológica desarrollada por Microsoft, ha despertado un creciente interés en la mejora de la comunicación para personas con discapacidad auditiva. Su capacidad para capturar movimientos y reconocer gestos ha sido fundamental en el desarrollo de diversas aplicaciones destinadas a mejorar la accesibilidad y la inclusión de estas personas en entornos de comunicación.

En el artículo "Hablar con sordos usando lenguaje de señas animado en 3D" [2], se presenta un innovador sistema llamado "Deaf Talk", basado en el Kinect V2 de Microsoft. Este sistema actúa como un intérprete y traductor entre usuarios de lenguaje de señas y hablantes de lenguaje natural. Utilizando la capacidad del Kinect para capturar movimientos, Deaf Talk logra una precisión notable del 87% en la conversión de habla a lenguaje de señas y del 84% en la conversión de lenguaje de señas a habla.

Por otro lado, el artículo "Reconocimiento de lengua de señas mediante Microsoft Kinect" de Anant Agarwal y Manish K Thakur [3] propone un sistema de reconocimiento del lenguaje de señas que utiliza imágenes de profundidad capturadas con el Kinect. Esta información se empleó para entrenar un clasificador SVM multiclase, cuyos resultados se compararon con técnicas existentes. El conjunto de datos utilizado incluye gestos en lengua de señas que representan los dígitos del 0 al 9. Este sistema emplea algoritmos de visión por computadora para generar un perfil característico de movimiento y profundidad para cada gesto del lenguaje de señas, logrando una mayor eficiencia en el reconocimiento en comparación con métodos tradicionales.

Además, el artículo "Reconocimiento y traducción de lengua de señas con Kinect" [4] destaca cómo el Kinect, al proporcionar datos de profundidad y color simultáneamente, permite un seguimiento más preciso y sencillo de los gestos del lenguaje de señas. Este enfoque ha sido fundamental en el desarrollo de sistemas de reconocimiento y traducción del lenguaje de señas, contribuyendo a superar las barreras comunicativas para las personas con discapacidad auditiva.

Asimismo, en el artículo "Visión por computadora mejorada con el sensor Microsoft Kinect" [5] se ofrece una revisión exhaustiva de los algoritmos y aplicaciones recientes en visión por computadora basados en el Kinect. Esta revisión destaca cómo el Kinect ha revolucionado diversos aspectos de la visión por computadora, desde el preprocesamiento hasta el análisis de gestos manuales y el mapeo 3D de interiores.

Ante el desafío de reconocer en tiempo real un amplio conjunto de gestos dinámicos, Hazari, Shihab Shahriar and Asaduzzaman [6] aplica algoritmos y modelos eficientes. Se emplea un algoritmo de vista de cuadrícula tanto para el reconocimiento como para la traducción de gestos, tanto en el entrenamiento como en la traducción. El sistema se evalúa mediante la traducción de gestos de varias personas para doce palabras diferentes, y los resultados experimentales indican una tasa de éxito de aproximadamente del 80% en la traducción de gestos.

Finalmente, el artículo "Avances en el desarrollo de un traductor básico de lenguaje de señas a voz y texto mexicano" de G. Garcia-Bautista [7] presenta un traductor automático básico de lengua de señas en tiempo real, desarrollado utilizando el sensor Kinect. Este traductor es

capaz de reconocer una variedad de señas de la Lengua de Señas Mexicana (LSM) y traducirlos al habla y texto, con una precisión significativa en la interpretación de gestos de las manos. Además, se emplea un método de reconocimiento de patrones basado en una red neuronal artificial de retropropagación (RNA) para interpretar los gestos de las manos. Por último, se utiliza el método de validación cruzada para las etapas de capacitación y prueba, logrando una precisión del 95.71% en palabras, 98.57% en números y 79.71% en letras. Como complemento, se desarrolla una interfaz de usuario interactiva que presenta los resultados en formato de voz y texto.

Otro de los trabajos relacionados es el de reconocimiento de gestos alfanuméricos en lengua de señas mexicana mediante funciones 3D de Javier Jimenez [8], el cual está basado en obtención de características 3D similares a las de Haar, extraídas de imágenes de profundidad capturadas por el sensor Microsoft Kinect. Estas características son procesadas utilizando un algoritmo de refuerzo. Para evaluar el rendimiento de nuestro método, reconocimos un conjunto de señas alfabéticas y numéricas, y comparamos los resultados con el uso de características tradicionales 2D similares a Haar. El sistema es capaz de reconocer señales estáticas de LSM con una tasa de precisión más alta que la obtenida con características 2D ampliamente utilizadas.

En conjunto, estos estudios destacan el papel fundamental del Kinect en la mejora de la comunicación y la accesibilidad para las personas con discapacidad auditiva, demostrando su versatilidad y potencial en el ámbito del lenguaje de señas.

Además del uso del Kinect, también podemos explorar el aprendizaje profundo o Deep Learning, una rama de la inteligencia artificial que, gracias a su capacidad para procesar grandes cantidades de datos y reconocer patrones complejos, ha demostrado éxito en el reconocimiento e interpretación de gestos y señas del lenguaje de señas. En esta sección, recapitularemos algunos sistemas y metodologías que contribuyen a mejorar la comunicación y la accesibilidad para las personas sordas y con discapacidad auditiva.

## **FUNDAMENTO TEÓRICO**

### *Visión por computadora*

Visión por computadora es el estudio de los métodos y técnicas por medio de los cuales los sistemas de visión artificial pueden ser contruidos y empleados de forma útil en aplicaciones prácticas, conjuntando la ciencia y la ingeniería del visón. [9] Es un campo de IA que permite obtener información significativa de imágenes digitales, vídeos y otras entradas visuales, con la intención de que se tomen acciones o hagan recomendaciones basadas en esa información. Si la IA permite que las computadoras piensen, la visión artificial les permite ver, observar y comprender.

La meta de la visión por computadora es crear modelos del mundo real para imágenes. Un sistema de visión por computadora recupera información útil sobre una escena de proyecciones bidimensional.

### *Inteligencia artificial*

La Inteligencia artificial (IA) es un campo de la ciencia relacionado con la creación de computadoras y máquinas que pueden razonar, aprender y actuar de una manera que normalmente requeriría inteligencia humana, o que involucren datos cuya escala exceda lo que los humanos pueden analizar. La IA es un campo amplio que abarca muchas disciplinas diferentes, incluidas la informática, el análisis de datos, las estadísticas, la ingeniería de hardware y software, la lingüística, la neurociencia y hasta la filosofía y la psicología.

El campo de la IA se compone de varias áreas de estudio, las más comunes e importantes son: [10]

- Búsqueda de soluciones.
- Sistemas expertos.
- Procesamiento del lenguaje natural.
- Reconocimiento de modelos.
- Aprendizaje computacional.
- Lógica.
- Incertidumbre y lógica difusa.

A nivel operativo para el uso empresarial, la IA es un conjunto de tecnologías que se basan principalmente en el aprendizaje automático y el aprendizaje profundo, usados entre otras cosas para el análisis de datos, la generación de predicciones y previsiones, la categorización

de objetos, el procesamiento de lenguaje natural, las recomendaciones, la recuperación inteligente de datos y mucho más.

### Aprendizaje automático

Hablar de modelado de IA nos conlleva las tareas de creación, entrenamiento e implementación de algoritmos que emulan la toma de decisiones lógica a partir de los datos disponibles, este tipo de algoritmos están ligados a los modelos de aprendizaje automático.

El aprendizaje automático (ML, por sus siglas en inglés) es el proceso mediante el cual se usan modelos matemáticos de datos para ayudar a un equipo de cómputo a aprender sin instrucciones directas. El ML se considera un subconjunto de la IA, usa algoritmos para identificar patrones en los datos, y esos patrones luego se usan para crear un modelo de datos que puede hacer predicciones. Con más experiencia y datos, los resultados del aprendizaje automático son más precisos, de forma muy similar a cómo los humanos mejoran con más práctica [11].

Mitchell en [12] define que un programa de computadora se dice que aprende a partir de una experiencia E con respecto a una clase de tareas T y medida de desempeño D, si su desempeño en las tareas T, medidas con D, mejoran con experiencia E.

Los modelos de ML se suelen clasificar como [13]:

- El aprendizaje supervisado: son algoritmos que generan una función que mapea datos de entrada (datos estructurados) a una salida deseada. la tarea más común para el aprendizaje supervisado es la clasificación.
- Aprendizaje no supervisado: son modelos de aprendizaje automático que aprenden patrones de datos no etiquetados (datos no estructurados), en esta tarea, el resultado final no se conoce con anticipación. la idea general consiste en que el algoritmo aprende de los datos y los clasifica en grupos en función de diversos atributos.
- Aprendizaje semi supervisado, en el que solo se etiquetan algunos de los datos. En el aprendizaje semi supervisado se conoce un resultado final, pero el algoritmo debe determinar cómo organizar y estructurar los datos para lograr los resultados deseados.
- Aprendizaje por refuerzo, donde el algoritmo aprende una política de como actuar dada una observación del mundo. Cada acción tiene cierto impacto en el entorno del

problema, quien a su vez retroalimenta para guiar al algoritmo. Se puede describir en términos generales como “Aprender haciendo”.

El presente trabajo de tesis está enfocado en la clasificación a partir de imágenes mediante el uso de arquitecturas neuronales, lo que cae en el aprendizaje supervisado.

El modelo del aprendizaje supervisado, que se presenta en la Figura [1] y tomado de [14], muestra que, inicialmente se requiere la existencia de un conjunto de datos de entrenamiento, los cuales son representados mediante vectores continuos o discretos de características, junto con las etiquetas, que representan la clase o categoría a la que pertenece cada uno de los ejemplos de entrenamiento, esta información alimentan un algoritmo de aprendizaje, que termina convirtiéndose en un modelo predictivo, el cual, al ser sometido a nueva información, tendrá la capacidad de asignar una etiqueta a instancias nuevas.

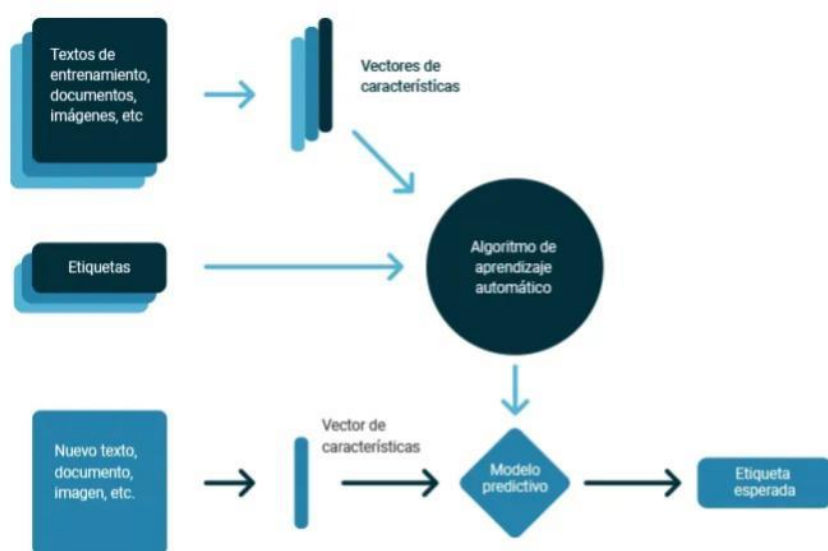


Figura 1. Modelo del aprendizaje supervisado.

### *Redes neuronales artificiales*

Una red neuronal propuesto por McCulloch y Pitts [15], puede ser definida como un modelo de razonamiento inspirado en el funcionamiento del cerebro humano. El cerebro consiste en un conjunto denso de interconexiones de células, o unidades básicas de procesamiento de información, llamadas neuronas. El cerebro humano este compuesto por una gran cantidad de neuronas, Suzana Herculano afirma que el cerebro humano tiene alrededor de 86 billones



de neuronas interconectadas. Una ANN emula las neuronas, las interconexiones y los procesos bioquímicos del cerebro, usando estructuras de datos y modelos matemáticos implementados en la computadora.

Hoy en día, gracias a las capacidades de cómputo, las ANN es uno de los modelos de aprendizaje automático más populares.

Para una ANN es necesario especificar:

1. El número de capas en la red.
2. El número de neuronas en la capa de entrada.
3. El número de capas ocultas.
4. El número de neuronas en cada capa oculta.
5. El número de neuronas en la capa de salida.
6. El conjunto inicial de datos de cada neurona artificial (pesos y sesgo).

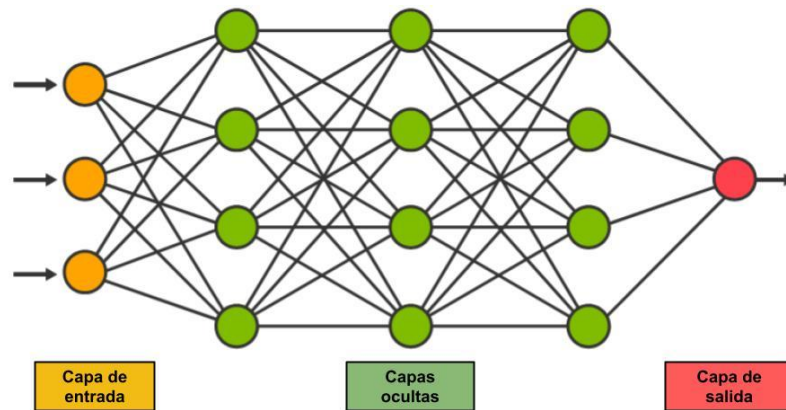


Figura 2. Arquitectura general de una Red Neuronal Artificial.

### **PLANTEAMIENTO DEL PROBLEMA:**

El lenguaje de señas mexicana (LSM) es el principal medio de comunicación para la comunidad sorda en México. Sin embargo, a pesar de su importancia, las herramientas y tecnológicas que facilitan la traducción y comprensión de LSM son limitadas. La falta de recursos adecuados para interpretar LSM en tiempo real crea barreras significativas para la inclusión de las personas sordas en diversos ámbitos de la vida cotidiana, como la educación, salud publica entre otras.

## **HIPÓTESIS**

Hipótesis: La implementación de una arquitectura neuronal aumentará significativamente la precisión en la interpretación del lenguaje de señas mexicano (LSM) en videos, en comparación con los métodos tradicionales de reconocimiento de patrones

Variable independiente: Arquitectura neuronal.

Variable dependiente: Precisión en la interpretación del lenguaje de señal mexicano.

## **JUSTIFICACIÓN**

El lenguaje de señal mexicana (LSM) es el medio de comunicación primordial para las personas sordas en México, permitiéndoles expresar ideas, emociones y necesidades de manera efectiva, Sin embargo, a pesar de su importancia, las barreras de comunicación entre la comunidad sorda y la oyente siguen siendo un obstáculo significativo para la inclusión social, educativa y laboral.

El avance en tecnología ofrece una oportunidad para desarrollar soluciones innovadoras que faciliten la interpretación y traducción del LSM, específicamente la inteligencia artificial nos permite desarrollar e implementar mecanismos para la interpretación de lenguaje de señas.

La falta de herramientas accesibles para la interpretación de LSM limita la participación de personas sordas en la sociedad. En contextos educativos, la ausencia de intérpretes de LSM o tecnológicas adecuadas impide que los estudiantes sordos accedan a una educación de calidad limitando sus oportunidades de desarrollo personal y profesional.

## **OBJETIVOS**

En esta sección se presentan las metas que se plantearon alcanzar en el desarrollo de este proyecto de tesis.

### **OBJETIVO GENERAL**

Proponer una metodología para la interpretar gestos individuales del lenguaje de señas mexicano, para presentar una alternativa base para el futuro desarrollo de herramientas para disminuir las barreras de la inclusión, mediante la aplicación de tecnologías de visión usando Deep learning.

## OBJETIVOS ESPECÍFICOS.

1. Reconocer las propuestas y alternativas existentes para el reconocimiento del lenguaje de señas para tener un panorama más específico de la problemática identificar los elementos necesarios para abordarla, mediante la revisión de literatura existente.
2. Construir una base de conocimiento sobre el lenguaje de señas mexicano que permitan entrenar modelos de aprendizaje para la tarea de reconocimiento, con el apoyo de expertos en el LSM.
3. Determinar el modelo de aprendizaje a utilizar para la tarea de reconocimiento de LSM, mediante el analizar las características de los datos recolectados.
4. Desarrollar e implementar una metodología basada en modelos de aprendizaje automático para interpretar LSM. Mediante el uso de técnicas y procesos del aprendizaje automático.
5. Evaluar el desempeño de la metodología propuesta para realizar mejoras continuas basadas en los resultados obtenidos, mediante el uso de métricas apropiadas.

## METODOLOGÍA

La metodología propuesta en este trabajo consta de una serie de procesos y elementos diseñados para procesar videos e interpretar LSM. Este proceso se llevará a cabo utilizando una arquitectura neuronal. La metodología se resume en el diagrama que se muestra en la figura [2].



Figura 3. Metodología

La primera fase, *adquisición de la información*, implica la grabación de los videos en donde las personas realizan los movimientos de las unidades comunicativas, esto se realizará con el apoyo de expertos en el tema y personas que maneja el LSM como lenguaje de comunicación cotidiano, lo que nos permite tener información de calidad y relevancia.

La *preparación de la información* constituye la segunda fase de la metodología, donde se adecuará la información a la forma en la que se utiliza en los modelos de aprendizaje supervisado, es decir, se realiza el etiquetado de la información, así como la extracción del contenido relevante.

El *aprendizaje*, es crucial ya que en esta etapa se definirá el modelo de aprendizaje a utilizar, así como sus características, a partir de la naturaleza de la información, lo que nos llevará alcanzar la tarea de reconocimiento de LSM.

Por último, la *evaluación* del modelo nos permite medir el impacto y la efectividad del modelo aprendido, esto mediante el uso de métricas estandarizadas aplicadas al aprendizaje automático. Así mismo se realizará un análisis de los resultados para entender la correlación que existe entre los datos y los resultados.

CRONOGRAMA

Tarea	Cronograma 2024						
	JUL	AG	SE	OC	NO	DI	EN
Adquisición de la información							
Preparación de la información							
Construcción y entrenamiento del modelo de aprendizaje							
Evaluación de resultados e implementación de mejoras y correcciones							

Tarea	Cronograma 2025				
	FEB	MAR	ABRIL	MAYO	JUN
Interpretación de los resultados					
Redacción de documento de reporte de tesis					

## REFERENCIAS BIBLIOGRÁFICAS

- [1] K. Deep, «Deepsign: Sign Language Detection and Recognition Using Deep Learning; Deepsign: detección y reconocimiento de lenguaje de señas mediante aprendizaje profundo», MDPI, 2022, doi: <https://doi.org/10.3390/electronics11111780>.
- [2] M. Ahmed, «Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2, Hablar con sordos usando lenguaje de señas animado en 3D», IEEE, 2016, doi: 10.1109/SAI.2016.7556002.
- [3] «Sign language recognition using Microsoft Kinect», doi: 10.1109/IC3.2013.6612186.
- [4] «Sign Language Recognition and Translation with Kinect», [En línea]. Disponible en: [https://vipl.ict.ac.cn/homepage/ksl/document/DEVISIGN-TR\\_V3-R1.pdf](https://vipl.ict.ac.cn/homepage/ksl/document/DEVISIGN-TR_V3-R1.pdf)
- [5] «Enhanced Computer Vision With Microsoft Kinect Sensor: A Review», doi: 10.1109/TCYB.2013.2265378.
- [6] «Designing a sign language translation system using kinect motion sensor device», doi: 10.1109/ECACE.2017.7912929.
- [7] «Advances to the development of a basic Mexican sign-to-speech and text language translator», doi: 10.1117/12.2238281.
- [8] J. Jimenez, «Mexican Sign Language Alphanumeric Gestures Recognition using 3D Haar-like Features», 2017, doi: 10.1109/TLA.2017.8071247.
- [9] E. R. Davies, Machine vision: theory, algorithms, practicalities. Elsevier, 2004
- [10] P. Ponce, Inteligencia artificial: con aplicaciones a la ingeniería. Alpha Editorial, 2010.
- [11] J. Luna-Gonzalez, "Tipos de aprendizaje automático," Medium, 2018. [Online]. Available: <https://medium.com/soldai/tipos-de-aprendizaje-automático-6413e3c615e2>. [Accessed: Jul. 01, 2024].
- [12] T. M. Mitchell, Machine learning, vol. 1. McGraw-hill New York, 2007.
- [13] T. O. Ayodele, "Types of machine learning algorithms," New advances in machine learning, vol. 3, pp. 19–48, 2010
- [14] M. Negnevitsky, *Artificial intelligence: a guide to intelligent systems*. Pearson education, 2005.
- [15] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943.