Universidad Rafael Landívar

Primer Semestre 2025

Inteligencia Artificial, sección 01

Ing. Max Alejandro Cerna Flores

Proyecto Final

Reconocimiento de Letras en Lenguaje de Señas con IA

Pablo Josué Reyes Calderón - 1040621

Guatemala 19 de mayo de 2025

1. Introducción y motivación del problema.

La comunicación es un elemento fundamental en la vida cotidiana y en el desarrollo de la sociedad. Sin embargo, millones de personas en el mundo enfrentan barreras comunicativas debido a discapacidades auditivas o del habla. El lenguaje de señas es una herramienta clave para superar estas barreras, pero su comprensión está limitada a quienes la conocen, lo que puede generar exclusión y dificultades en la interacción con el entorno.

En este contexto, la inteligencia artificial (IA) y el aprendizaje automático ofrecen nuevas oportunidades para desarrollar tecnologías inclusivas que faciliten la comunicación entre personas sordas y oyentes. El presente documento describe el desarrollo de un sistema de inteligencia artificial capaz de reconocer letras en lenguaje de señas utilizando visión por computadora. El objetivo principal es facilitar la traducción de señas manuales a texto en tiempo real a través de una interfaz web, promoviendo así la accesibilidad comunicativa para personas con discapacidad auditiva o en entornos donde el lenguaje hablado no es viable.

El sistema fue desarrollado en el marco del curso de Inteligencia Artificial, aplicando técnicas de aprendizaje automático, redes neuronales y procesamiento de imágenes para lograr una solución funcional, práctica y educativa.

La motivación de este proyecto nace de la necesidad de reducir la brecha comunicativa y promover una sociedad más inclusiva mediante el uso de la tecnología. Facilitar el entendimiento de la lengua de señas a través de herramientas automáticas representa un paso importante hacia la igualdad de oportunidades para las personas con discapacidad auditiva.

2. Objetivos

Objetivo General:

Desarrollar una aplicación que permita el reconocimiento de letras en lenguaje de señas y transcribirlas automáticamente en texto mediante el uso de inteligencia artificial y visión por computadora, con el fin de facilitar la comunicación entre personas sordas y oyentes, promoviendo la inclusión social mediante tecnologías accesibles.

Objetivos Específicos:

- Capturar y procesar imágenes de señas con una cámara en tiempo real.
- Aplicar técnicas de preprocesamiento y normalización de imágenes.
- Entrenar un modelo de clasificación basado en una red neuronal convolucional (CNN).
- Implementar una interfaz web para la interacción del usuario.
- Integrar el modelo entrenado con la interfaz para mostrar texto acumulado.
- Evaluar el modelo con métricas de rendimiento como accuracy y precisión.

3. Descripción del Dataset

Se utilizó un dataset personalizado con imágenes tomadas directamente por el usuario. Cada carpeta representa una letra del alfabeto, conteniendo múltiples imágenes con variaciones de iluminación, ángulo y mano dominante.
Las clases consideradas son las letras A–Z (se excluyeron aquellas que requieren movimiento), permitiendo al usuario formar palabras letra por letra.

El dataset se organizó de la siguiente forma:

dataset/

- A/
- B/
- C/
- ...
- Z/

4. Preprocesamiento

Se aplicó el siguiente flujo de preprocesamiento para cada imagen capturada:

1. Redimensionamiento a 100x100 píxeles
2. Conversión de tipo a float32
3. Normalización de píxeles (escala 0–1)
4. Expansión de dimensiones para que coincida con la entrada del modelo CNN

Esto se implementó usando las bibliotecas OpenCV y NumPy.

5. Implementación del Modelo

Se implementó un modelo de red neuronal convolucional (CNN) utilizando TensorFlow y Keras.
La arquitectura incluyó las siguientes capas:

- Conv2D con 32 filtros y ReLU
- MaxPooling2D
- Conv2D con 64 filtros y ReLU
- MaxPooling2D
- Flatten
- Dense con 128 neuronas
- Dense final con función softmax (una clase por letra)

El modelo fue entrenado por 10 épocas con categorical_crossentropy y optimizador Adam.
Se logró una precisión de validación superior al 95%.

6. Interfaz Web

Se desarrolló una interfaz web utilizando Flask. La página principal permite:

- Ver el video en vivo desde la cámara
- Activar y desactivar la cámara con botones
- Mostrar el texto acumulado al presionar “Guardar texto”
- Reiniciar el texto automáticamente después de guardarlo

El diseño fue hecho con HTML, CSS y JavaScript. El estilo es responsivo y limpio.

7. Integración del Modelo con la Interfaz

El modelo fue cargado desde un archivo .pkl junto con su codificador de etiquetas. El video se procesa en tiempo real, y cada letra detectada es:

- Filtrada por repetición (requiere X frames iguales)
- Regulada por un cooldown de 2 segundos para evitar duplicados

Una vez confirmada, la letra se agrega al texto acumulado que se muestra en la interfaz web.

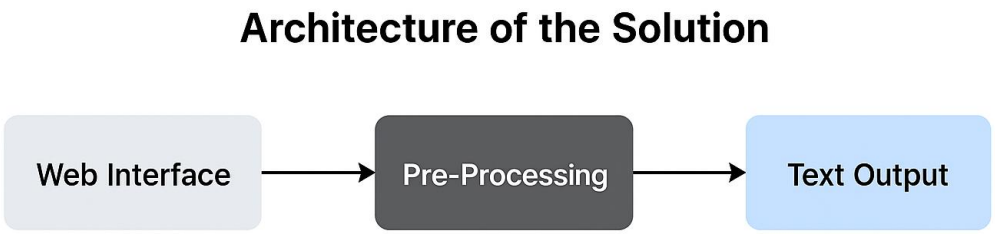
8. Evaluación del Modelo

Se usaron las siguientes métricas de evaluación:

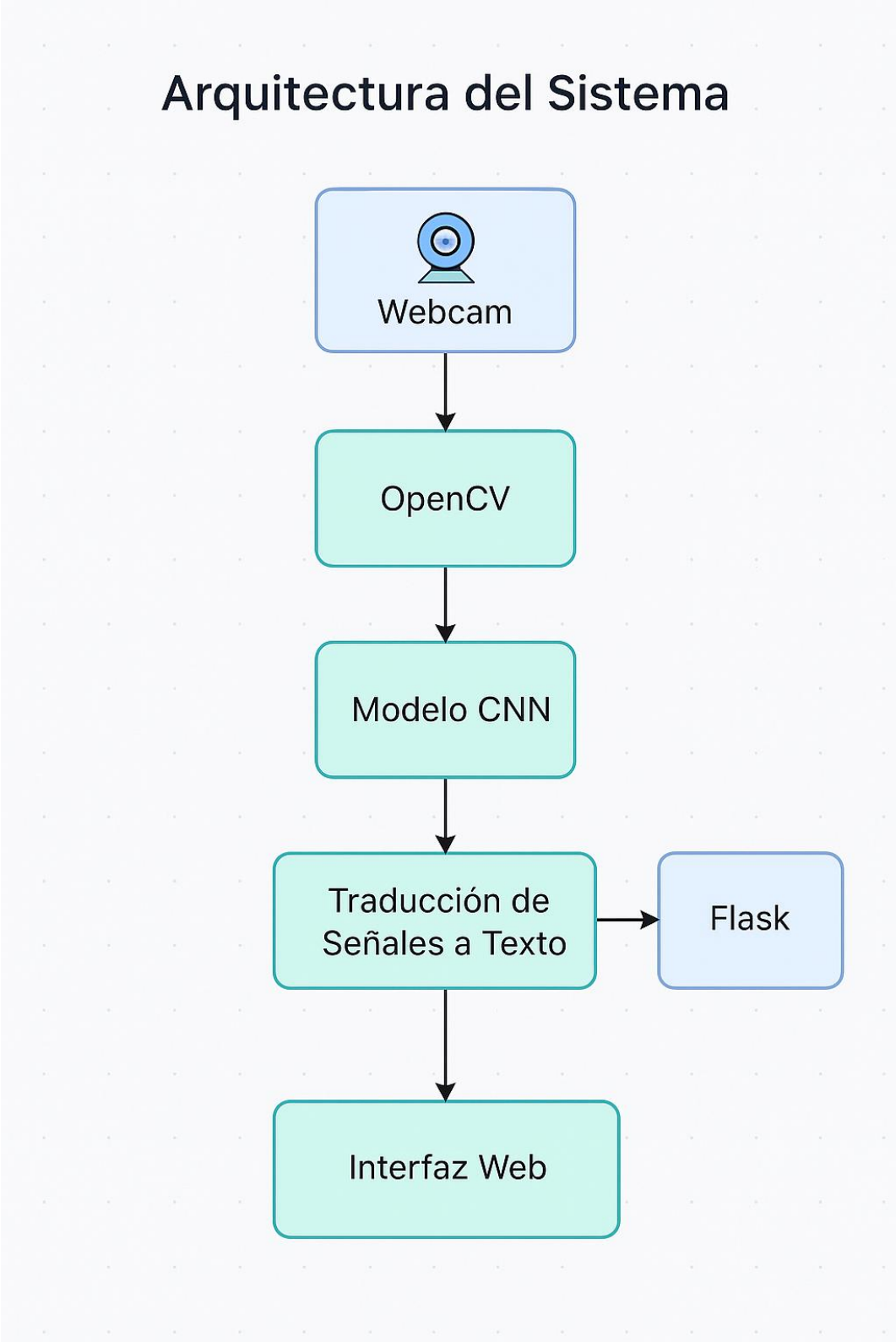
- **Accuracy** (exactitud): >95%
- **Matriz de Confusión**: mostró errores comunes entre letras visualmente similares como "M" y "N"
- **F1-score**: se utilizó en pruebas internas para verificar balance entre precisión y sensibilidad

9. Diagramas

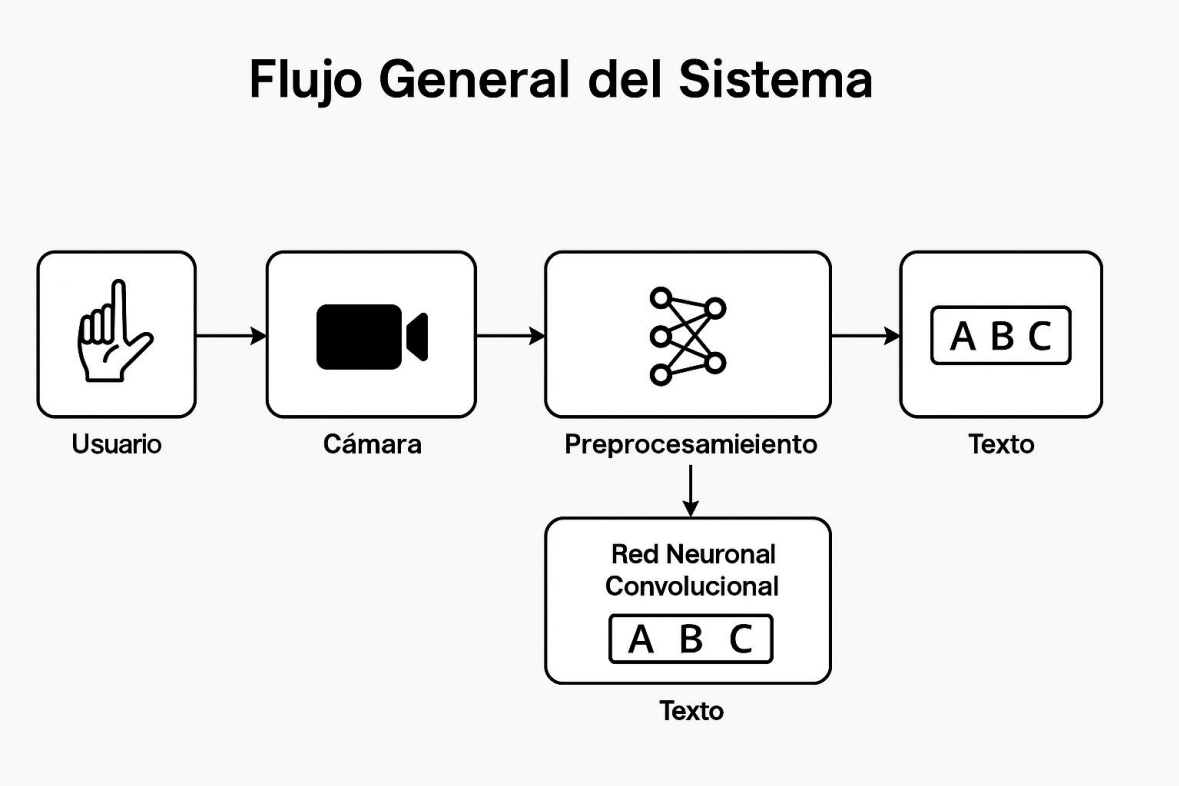
- Arquitectura de la solución



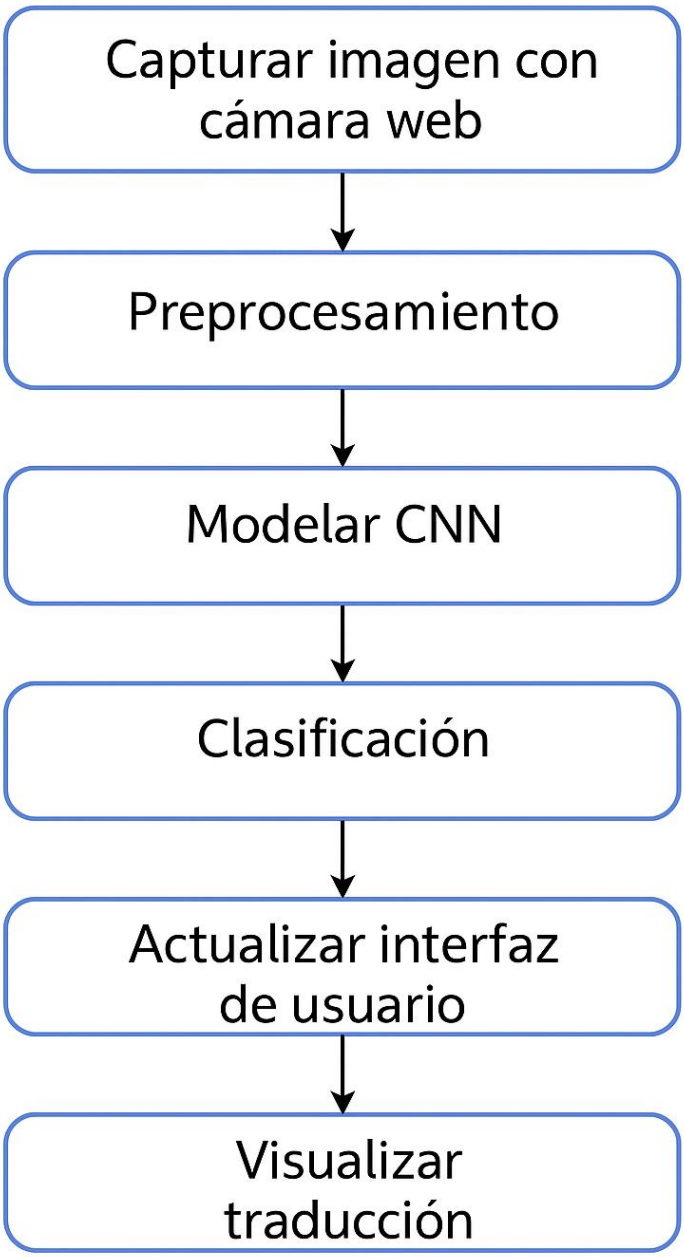
- Casos de uso.


Arquitectura del Sistema
Webcam
OpenCV
Modelo CNN
Traducción de Señales a Texto
Flask
Interfaz Web

- Flujo general

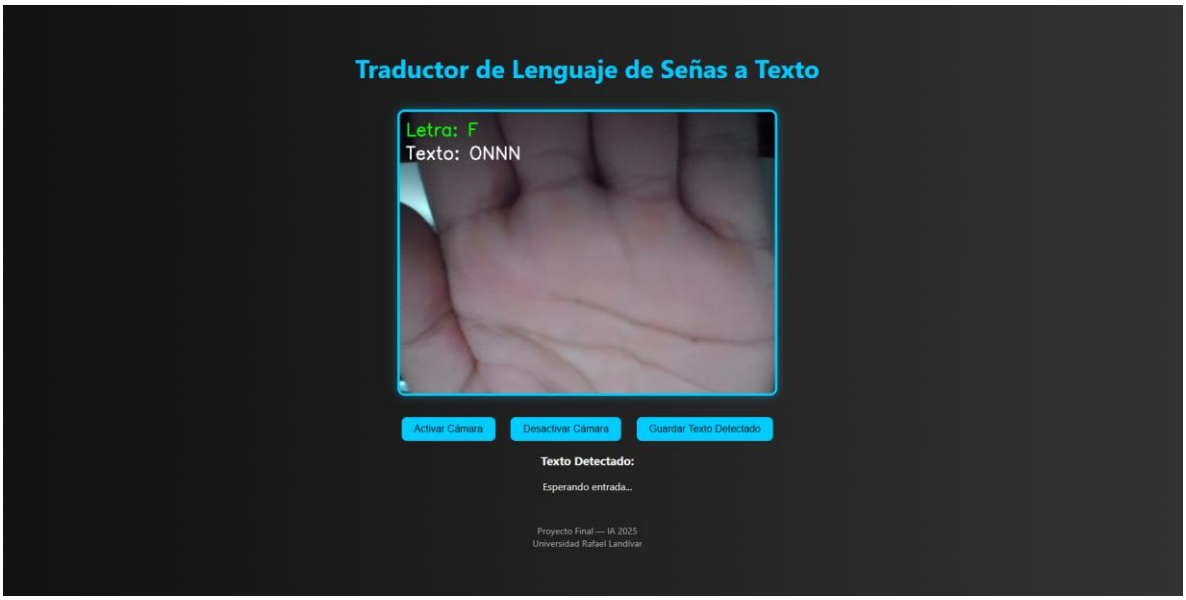


- Componentes y secuencia de interacción.



10. Evidencias de Funcionamiento

Podemos observar la página Web donde tiene 3 botones uno para activar la cámara, otro para desactivarla y por último el botón donde guarda el texto detectado y reinicia el texto de la cámara.



11. Conclusiones y Aprendizajes

- Este proyecto permitió aplicar de manera práctica los conocimientos adquiridos en inteligencia artificial, especialmente en clasificación, visión por computadora y aprendizaje supervisado.
- El sistema demuestra que es posible implementar soluciones accesibles utilizando IA, con resultados precisos y funcionales. También se identificaron desafíos como la similitud visual entre algunas letras y la importancia del preprocesamiento y estabilización para lograr predicciones confiables.
- La combinación de técnicas de visión por computadora y aprendizaje profundo permite identificar patrones en los gestos manuales con alta eficiencia, aunque factores como la iluminación, el fondo y la velocidad de ejecución de las señas pueden afectar el rendimiento del modelo.