

# Comparing different Machine Learning architectures for classifying medical terms in Colombian sign language

Javier F. Calderón-Hobaica<sup>1</sup>  
Pablo E. Caicedo-Rodríguez<sup>1</sup>

<sup>1</sup>Ingeniería Biomédica,  
Univ. Escuela Colombiana de Ingeniería - Universidad del Rosario



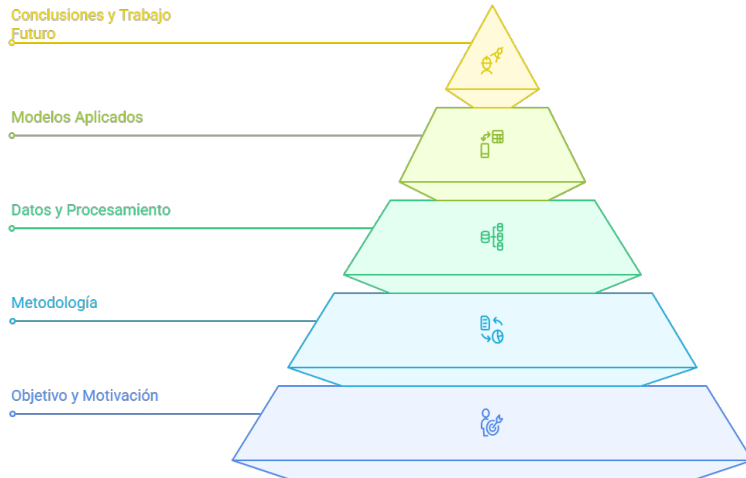
UNIVERSIDAD



Universidad del  
**Rosario**

VIGILADA MINEDUCACIÓN

Simposio de Tratamiento de Señales, Imágenes y Visión Artificial  
(27 agosto 2025)



# OBJETIVO Y MOTIVACIÓN

# Brecha comunicativa en contextos clínicos



- La disponibilidad de intérpretes de LSC en entornos sanitarios es limitada.
- Esto afecta la anamnesis, la seguridad del paciente y la continuidad del cuidado.
- Objetivo general: habilitar el reconocimiento de vocabulario médico frecuente en LSC usando cámara convencional y bibliotecas públicas.

# Brecha comunicativa en contextos clínicos



- La disponibilidad de intérpretes de LSC en entornos sanitarios es limitada.
- Esto afecta la anamnesis, la seguridad del paciente y la continuidad del cuidado.
- Objetivo general: habilitar el reconocimiento de vocabulario médico frecuente en LSC usando cámara convencional y bibliotecas públicas.

# Brecha comunicativa en contextos clínicos



- La disponibilidad de intérpretes de LSC en entornos sanitarios es limitada.
- Esto afecta la anamnesis, la seguridad del paciente y la continuidad del cuidado.
- Objetivo general: habilitar el reconocimiento de vocabulario médico frecuente en LSC usando cámara convencional y bibliotecas públicas.

# Objetivo de la investigación

## Propósito

Desarrollar y evaluar un sistema que, a partir de secuencias de puntos clave de la mano y muñeca, clasifique señas médicas frecuentes en LSC.

## Contribución

Comparación de tres arquitecturas: MLP, LSTM y LSTM mejorado con mecanismo de atención, en un corpus de señas de dominio clínico.

# METODOLOGÍA



# Flujo de trabajo

- 1 Captura de video de señas en condiciones controladas.
- 2 Extracción de puntos clave (mano y muñeca) con MediaPipe.
- 3 Construcción de series temporales (coordenadas 2D + etiqueta).
- 4 Entrenamiento de modelos (MLP / LSTM / LSTM+Atención).
- 5 Validación cruzada y comparación de métricas.



# Flujo de trabajo

- 1 Captura de video de señas en condiciones controladas.
- 2 Extracción de puntos clave (mano y muñeca) con MediaPipe.
- 3 Construcción de series temporales (coordenadas 2D + etiqueta).
- 4 Entrenamiento de modelos (MLP / LSTM / LSTM+Atención).
- 5 Validación cruzada y comparación de métricas.



# Flujo de trabajo

- 1 Captura de video de señas en condiciones controladas.
- 2 Extracción de puntos clave (mano y muñeca) con MediaPipe.
- 3 Construcción de series temporales (coordenadas 2D + etiqueta).
- 4 Entrenamiento de modelos (MLP / LSTM / LSTM+Atención).
- 5 Validación cruzada y comparación de métricas.



# Flujo de trabajo

- 1 Captura de video de señas en condiciones controladas.
- 2 Extracción de puntos clave (mano y muñeca) con MediaPipe.
- 3 Construcción de series temporales (coordenadas 2D + etiqueta).
- 4 Entrenamiento de modelos (MLP / LSTM / LSTM+Atención).
- 5 Validación cruzada y comparación de métricas.



# Flujo de trabajo

- 1 Captura de video de señas en condiciones controladas.
- 2 Extracción de puntos clave (mano y muñeca) con MediaPipe.
- 3 Construcción de series temporales (coordenadas 2D + etiqueta).
- 4 Entrenamiento de modelos (MLP / LSTM / LSTM+Atención).
- 5 Validación cruzada y comparación de métricas.



# DATOS Y PROCESAMIENTO

# Conjunto de datos

- Vocabulario: señas de uso clínico habitual (p. ej., síntomas frecuentes) y alfabeto dactilológico.
- Representación: por muestra se obtiene una secuencia *temporal* de coordenadas por articulación.
- Formato final: matriz  $T \times F$  (tiempo  $\times$  características) con etiqueta de clase.

## Observación

El uso de puntos clave reduce sensibilidad a iluminación y tono de piel, favoreciendo la generalización.

## Enfermedades

- Diabetes tipo 2
- Cansancio
- Gastritis o ERGE
- Ardor
- Hipertension arterial
- Infeccion urinaria no complicada

## Síntomas

- Abecedario
- Actividad
- Alta
- Borrosa
- Cabeza
- Cardiovascular

# Conjunto de datos

- Vocabulario: señas de uso clínico habitual (p. ej., síntomas frecuentes) y alfabeto dactilológico.
- Representación: por muestra se obtiene una secuencia *temporal* de coordenadas por articulación.
- Formato final: matriz  $T \times F$  (tiempo  $\times$  características) con etiqueta de clase.

## Observación

El uso de puntos clave reduce sensibilidad a iluminación y tono de piel, favoreciendo la generalización.

## Enfermedades

- Diabetes tipo 2
- Cansancio
- Gastritis o ERGE
- Ardor
- Hipertension arterial
- Infeccion urinaria no complicada

## Síntomas

- Abecedario
- Actividad
- Alta
- Borrosa
- Cabeza
- Cardiovascular



# Conjunto de datos

- Vocabulario: señas de uso clínico habitual (p. ej., síntomas frecuentes) y alfabeto dactilológico.
- Representación: por muestra se obtiene una secuencia *temporal* de coordenadas por articulación.
- Formato final: matriz  $T \times F$  (tiempo  $\times$  características) con etiqueta de clase.

## Observación

El uso de puntos clave reduce sensibilidad a iluminación y tono de piel, favoreciendo la generalización.

## Enfermedades

- Diabetes tipo 2
- Cansancio
- Gastritis o ERGE
- Ardor
- Hipertension arterial
- Infeccion urinaria no complicada

## Síntomas

- Abecedario
- Actividad
- Alta
- Borrosa
- Cabeza
- Cardiovascular

# Preprocesamiento y particionado

- Limpieza y normalización de coordenadas.
- Aumento leve (ruido gaussiano) para robustez.
- Balance de clases en entrenamiento.
- División en entrenamiento (49 %)/validación(21 %)/prueba(30 %) para vigilar sobreajuste.

# Preprocesamiento y particionado

- Limpieza y normalización de coordenadas.
- Aumento leve (ruido gaussiano) para robustez.
- Balance de clases en entrenamiento.
- División en entrenamiento (49 %)/validación(21 %)/prueba(30 %) para vigilar sobreajuste.

# Preprocesamiento y particionado

- Limpieza y normalización de coordenadas.
- Aumento leve (ruido gaussiano) para robustez.
- Balance de clases en entrenamiento.
- División en entrenamiento (49 %)/validación(21 %)/prueba(30 %) para vigilar sobreajuste.

# Preprocesamiento y particionado

- Limpieza y normalización de coordenadas.
- Aumento leve (ruido gaussiano) para robustez.
- Balance de clases en entrenamiento.
- División en entrenamiento (49 %)/validación(21 %)/prueba(30 %) para vigilar sobreajuste.

# MODELOS

# Perceptrón Multicapa (MLP)

- Arquitectura densa aplicada a *ventanas temporales* (secuencia a vector).
- Funciones de activación no lineales y regularización (*dropout*).
- Ventaja: simplicidad y eficiencia; riesgo: menor modelado de dinámica temporal.

# Perceptrón Multicapa (MLP)

- Arquitectura densa aplicada a *ventanas temporales* (secuencia a vector).
- Funciones de activación no lineales y regularización (*dropout*).
- Ventaja: simplicidad y eficiencia; riesgo: menor modelado de dinámica temporal.



# Perceptrón Multicapa (MLP)

- Arquitectura densa aplicada a *ventanas temporales* (secuencia a vector).
- Funciones de activación no lineales y regularización (*dropout*).
- Ventaja: simplicidad y eficiencia; riesgo: menor modelado de dinámica temporal.

# LSTM y Enhanced LSTM

- **LSTM**: captura dependencias de largo alcance en secuencias.
- **ELSTM**: integra contexto pasado y futuro; la atención focaliza instantes informativos.
- Ventaja: mejor discriminación en transiciones sutiles; costo: mayor complejidad computacional.

# LSTM y Enhanced LSTM

- **LSTM**: captura dependencias de largo alcance en secuencias.
- **ELSTM**: integra contexto pasado y futuro; la atención focaliza instantes informativos.
- Ventaja: mejor discriminación en transiciones sutiles; costo: mayor complejidad computacional.

# LSTM y Enhanced LSTM

- **LSTM**: captura dependencias de largo alcance en secuencias.
- **ELSTM**: integra contexto pasado y futuro; la atención focaliza instantes informativos.
- Ventaja: mejor discriminación en transiciones sutiles; costo: mayor complejidad computacional.

# Protocolo y métricas

- Validación cruzada  $k$ -fold para estimar capacidad de generalización.
- Métricas: exactitud, precisión, exhaustividad (*recall*) y F1.

## Pérdida de clasificación (ejemplo)

Para clasificación multiclase con *softmax*:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log \hat{y}_{ic}$$

## F1 (por clase)

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Protocolo y métricas

- Validación cruzada  $k$ -fold para estimar capacidad de generalización.
- Métricas: exactitud, precisión, exhaustividad (*recall*) y F1.

## Pérdida de clasificación (ejemplo)

Para clasificación multiclase con *softmax*:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log \hat{y}_{ic}$$

## F1 (por clase)

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Hallazgos Principales

**Cuadro 1:** Resultados comparativos de LSTM, MLP y ELSTM (validación cruzada).

Métrica	LSTM	MLP	ELSTM
Pérdida de entrenamiento	0.7012	0.8554	0.0714
Pérdida de validación	0.6958	0.8131	0.2252
<b>Validación cruzada</b>			
Exactitud promedio (entrenamiento)	$0,8409 \pm 0,0095$	$0,9937 \pm 0,0004$	$0,9701 \pm 0,0054$
Exactitud promedio (prueba)	$0,7975 \pm 0,0092$	$0,9687 \pm 0,0006$	$0,9483 \pm 0,0066$
Exactitud promedio	$0,8006 \pm 0,0093$	$0,9693 \pm 0,0006$	$0,9592 \pm 0,0054$
<i>Recall</i> promedio	$0,7975 \pm 0,0092$	$0,9687 \pm 0,0006$	$0,9483 \pm 0,0066$
F1 promedio	$0,7964 \pm 0,0092$	$0,9687 \pm 0,0006$	$0,9462 \pm 0,0074$

# Hallazgos principales

- Los modelos recurrentes modelan de forma más estable la dinámica de señas.
- El mecanismo de atención tiende a reducir la pérdida y a mejorar la separación entre categorías cercanas.
- El MLP puede alcanzar alta exactitud con mayor propensión al sobreajuste en ciertos escenario.
- El sobre-ajuste del mecanismo de atención es altamente dependiente de la cantidad de datos de entrenamiento.



# Hallazgos principales

- Los modelos recurrentes modelan de forma más estable la dinámica de señas.
- El mecanismo de atención tiende a reducir la pérdida y a mejorar la separación entre categorías cercanas.
- El MLP puede alcanzar alta exactitud con mayor propensión al sobreajuste en ciertos escenario.
- El sobre-ajuste del mecanismo de atención es altamente dependiente de la cantidad de datos de entrenamiento.

# Hallazgos principales

- Los modelos recurrentes modelan de forma más estable la dinámica de señas.
- El mecanismo de atención tiende a reducir la pérdida y a mejorar la separación entre categorías cercanas.
- El MLP puede alcanzar alta exactitud con mayor propensión al sobreajuste en ciertos escenario.
- El sobre-ajuste del mecanismo de atención es altamente dependiente de la cantidad de datos de entrenamiento.

# Hallazgos principales

- Los modelos recurrentes modelan de forma más estable la dinámica de señas.
- El mecanismo de atención tiende a reducir la pérdida y a mejorar la separación entre categorías cercanas.
- El MLP puede alcanzar alta exactitud con mayor propensión al sobreajuste en ciertos escenario.
- El sobre-ajuste del mecanismo de atención es altamente dependiente de la cantidad de datos de entrenamiento.

# CONCLUSIONES Y TRABAJOS FUTUROS

# Relación con trabajos previos

- Tendencia favorable al uso de puntos clave con modelos secuenciales.
- Vocabularios especializados (médico) se benefician de atención.
- Aporte: enfoque específico en LSC y terminología clínica.

# Relación con trabajos previos

- Tendencia favorable al uso de puntos clave con modelos secuenciales.
- Vocabularios especializados (médico) se benefician de atención.
- Aporte: enfoque específico en LSC y terminología clínica.

# Relación con trabajos previos

- Tendencia favorable al uso de puntos clave con modelos secuenciales.
- Vocabularios especializados (médico) se benefician de atención.
- Aporte: enfoque específico en LSC y terminología clínica.

# Limitaciones del estudio

- Diversidad limitada de signantes y estilos personales.
- Posibles variaciones dialectales no cubiertas.
- Tamaño muestral acotado frente a la complejidad gestual de la LSC.



# Limitaciones del estudio

- Diversidad limitada de signantes y estilos personales.
- Posibles variaciones dialectales no cubiertas.
- Tamaño muestral acotado frente a la complejidad gestual de la LSC.

# Limitaciones del estudio

- Diversidad limitada de signantes y estilos personales.
- Posibles variaciones dialectales no cubiertas.
- Tamaño muestral acotado frente a la complejidad gestual de la LSC.

# Implicaciones y casos de uso

- Apoyo en triaje y anamnesis básica.
- Orientación de la consulta en entornos con recursos limitados.
- Potencial educativo para formación en accesibilidad y salud digital.

# Implicaciones y casos de uso

- Apoyo en triaje y anamnesis básica.
- Orientación de la consulta en entornos con recursos limitados.
- Potencial educativo para formación en accesibilidad y salud digital.

# Implicaciones y casos de uso

- Apoyo en triaje y anamnesis básica.
- Orientación de la consulta en entornos con recursos limitados.
- Potencial educativo para formación en accesibilidad y salud digital.

# Consideraciones tecnológicas

- Bibliotecas abiertas (p. ej., PyTorch, MediaPipe) y hardware de gama media.
- Pipeline reproducible: extracción de *keypoints*, preparación de secuencias y entrenamiento.
- Despliegue potencial en equipos convencionales (portátiles y PC de escritorio).

# Consideraciones tecnológicas

- Bibliotecas abiertas (p. ej., PyTorch, MediaPipe) y hardware de gama media.
- Pipeline reproducible: extracción de *keypoints*, preparación de secuencias y entrenamiento.
- Despliegue potencial en equipos convencionales (portátiles y PC de escritorio).

# Consideraciones tecnológicas

- Bibliotecas abiertas (p. ej., PyTorch, MediaPipe) y hardware de gama media.
- Pipeline reproducible: extracción de *keypoints*, preparación de secuencias y entrenamiento.
- Despliegue potencial en equipos convencionales (portátiles y PC de escritorio).



# Conclusiones y trabajo futuro

- La modelación secuencial con atención ofrece un balance favorable entre ajuste y generalización.
- Próximos pasos: ampliar corpus (múltiples intérpretes/dialectos), extender vocabulario y evaluar en clínica.
- Evaluar el efecto del aumento del corpus sobre los procesos de atención.
- Contribución a la equidad: comunicación más fluida paciente–equipo de salud.

# Conclusiones y trabajo futuro

- La modelación secuencial con atención ofrece un balance favorable entre ajuste y generalización.
- Próximos pasos: ampliar corpus (múltiples intérpretes/dialectos), extender vocabulario y evaluar en clínica.
- Evaluar el efecto del aumento del corpus sobre los procesos de atención.
- Contribución a la equidad: comunicación más fluida paciente–equipo de salud.

# Conclusiones y trabajo futuro

- La modelación secuencial con atención ofrece un balance favorable entre ajuste y generalización.
- Próximos pasos: ampliar corpus (múltiples intérpretes/dialectos), extender vocabulario y evaluar en clínica.
- Evaluar el efecto del aumento del corpus sobre los procesos de atención.
- Contribución a la equidad: comunicación más fluida paciente–equipo de salud.

# Conclusiones y trabajo futuro

- La modelación secuencial con atención ofrece un balance favorable entre ajuste y generalización.
- Próximos pasos: ampliar corpus (múltiples intérpretes/dialectos), extender vocabulario y evaluar en clínica.
- Evaluar el efecto del aumento del corpus sobre los procesos de atención.
- Contribución a la equidad: comunicación más fluida paciente–equipo de salud.

# Agradecimientos y contacto

- Agradecimientos a:
  - Ingeniería Biomédica de las Universidad Escuela Colombiana de Ingeniería y Universidad del Rosario.
  - MD, Anest. Camila Guarín.
- Contacto: [pablo.caicedo@escuelaing.edu.co](mailto:pablo.caicedo@escuelaing.edu.co)

**GRACIAS**