

Modelo de interpretación de lengua de señas colombiano usando inteligencia artificial

Colombian Sign Language Interpretation Model using Artificial Intelligence

Jader Alejandro Muñoz-Galindez¹

Rubiel Vargas-Cañas²

Recibido: abril 18 de 2023

Aceptado: junio 28 de 2023

Resumen

En este trabajo se presentan dos modelos de interpretación de Lengua de Señas Colombiana (LSC), usando métodos estáticos y dinámicos que emplean inteligencia artificial. Se utilizó como referente la metodología CRISP-DM, creando una base de datos con videos de setenta participantes no expertos, siendo preprocesados y posteriormente divididos en proporciones de 70% - 30% para entrenamiento y prueba, respectivamente. El repositorio se nombró como LSC-W70 y se empleó sobre un modelo preentrenado de redes neuronales convolucionales y otro en combinación con redes LSTM. Los resultados alcanzaron un 67% y 76% accuracy para los modelos estático y dinámico, respectivamente, donde el modelo dinámico presenta mejoras en señas similares identificando la dirección del movimiento para definir el tipo de seña. En este sentido, se desarrolló una herramienta de interpretación dinámica de lengua de señas colombiano que ayuda a cerrar brechas de comunicación generando igualdad entre las personas.

Palabras clave: lengua de señas colombiano, CNN, LSTM, CRISP-DM.

Abstract

In this work, two interpretation models of Colombian Sign Language (CSL) are presented, using static and dynamic methods that employ artificial intelligence. The CRISP-DM methodology was used as a reference, creating a database with videos from seventy non-expert participants, being preprocessed and subsequently divided into proportions of 70% - 30% for training and testing, respectively. The repository was named LSC-W70 and was used on a pre-trained model of convolutional neural networks and another in combination with LSTM networks. The results reached 67% and 76% accuracy for the static and dynamic models, respectively, where the dynamic model presents improvements in similar signs by identifying the direction of movement to define the type of sign. In this sense, a dynamic Colombian sign language interpretation tool was developed that helps close communication gaps, generating equality between people.

Keywords: colombian sign language, CNN, LSTM, CRISP-DM.

¹ Ingeniero Físico, Universidad del Cauca, Popayán, Colombia. E-mail: jadermu@unicauca.edu.co
Orcid: <https://orcid.org/0009-0001-9428-4192>

² Ingeniero de Sistemas, PhD. in Biomedical Engineering, Universidad del Cauca, Popayán, Colombia. E-mail: rubiel@unicauca.edu.co
Orcid: <https://orcid.org/0000-0003-1548-942X>

1. Introducción

La discapacidad auditiva es una condición que afecta alrededor del 5% de la población mundial (Díaz et al., 2016). En Colombia, esta población es de alrededor de medio millón de personas (INSOR, 2021), quienes han desarrollado su propio canal de comunicación no verbal, por medio de señas, ademanes, gestos, entre otras. A través de este tipo de signos, las personas sordas dan a entender sus ideas y pensamientos, siendo conocido como lengua de señas y se caracteriza según las costumbres y cultura de cada región.

En Colombia, la lengua de señas ha recibido múltiples influencias por parte de lenguas de otros países, en especial de la lengua de señas española debido a migrantes y sordos educados en España en los años 50 (INSOR, 2011). Actualmente, es la lengua oficial de las personas sordas en Colombia y mediante la ley 982 de 2005 es protegida, garantizando su preservación y divulgación. Sin embargo, a pesar de los años y políticas implementadas por el gobierno nacional, no todas las personas poseen conocimiento de la Lengua de Señas Colombiana (LSC), por lo que se dificulta tener igualdad de condiciones entre las personas que poseen la discapacidad y las que no, siendo esta una brecha bastante amplia que debe ser cerrada. Para ello debe tenerse en cuenta el orden lingüístico y el espacio de conversación, que se caracteriza por el uso de todo el cuerpo, funcionando como marco de explicación de la idea y cuyas funciones deben ser de entorno dinámico para entender señas que poseen similar estructura (INSOR, 2011).

Para afrontar este reto, se han adelantado diferentes investigaciones que hacen uso de diversas herramientas, entre ellas la inteligencia artificial, que buscan entender de manera automática dichas señas. Ejemplo de ello son investigaciones con el uso de sensores electromiográficos, donde su objetivo es la clasificación de 27 gestos de LSC con voluntarios que presentan discapacidad auditiva (Galvis-Serrano et al., 2019), sin embargo, estos métodos tienen implicaciones invasivas sobre las personas participantes. Para evitar estas dificultades, otras investigaciones efectúan la

interpretación de la lengua de señas a partir de imágenes y video. La mayoría de estas investigaciones se centra en la detección, principalmente, del alfabeto en forma estática enfocado sobre la mano, extrayendo características durante la etapa de pre-procesamiento (Jiménez et al., 2019; Monsalve-Pineda & Polo-Álvarez, 2016).

También se ha probado con redes neuronales convolucionales (CNN), que extraen en su etapa de entrenamiento las características de las imágenes y las clasifica directamente (Martínez, et al., 2020; Suat-Rojas et al., 2021). Asimismo, se hace uso de redes convolucionales pre-entrenadas (Ortiz-Farfán & Camargo-Mendoza, 2020), y modelos de combinación de redes convolucionales con redes recurrentes LSTM (Vaghasiya, 2021), o una combinación con los denominados transformer (Boháček & Hruz, 2022), para la identificación dinámica de la lengua de señas.

En el caso de Colombia, las bases de datos dinámicas son muy pocas, aunque Ortiz-Farfán y Camargo-Mendoza (2020), intentan realizar una base de datos dinámica, aún faltan muchos datos para poder realizar un clasificador dinámico. En otras palabras, no se tiene un referente en este tema ni lo principal para realizar clasificación de lengua de señas: una base de datos propia. En este artículo se presenta el desarrollo de dos modelos de inteligencia artificial, que puedan resolver la interpretación de la lengua de señas colombiana de forma dinámica, respetando la lingüística y el espacio de conversación caracterizado por su clasificación en torso o cuerpo completo.

2. Metodología

Para la realización de esta investigación se empleó la metodología CRISP-DM referencia, que es una metodología aplicada en la minería de datos y cuenta con seis pasos (Figura 1). El primero hace referencia al entendimiento del negocio, los dos siguientes al tratamiento de los datos, continúa con el modelamiento y evaluación, para finalmente llegar al despliegue. Los pasos de esta metodología no son secuenciales, pudiendo retroceder entre cada uno de ellos para realizar una mejor adaptación de los datos.

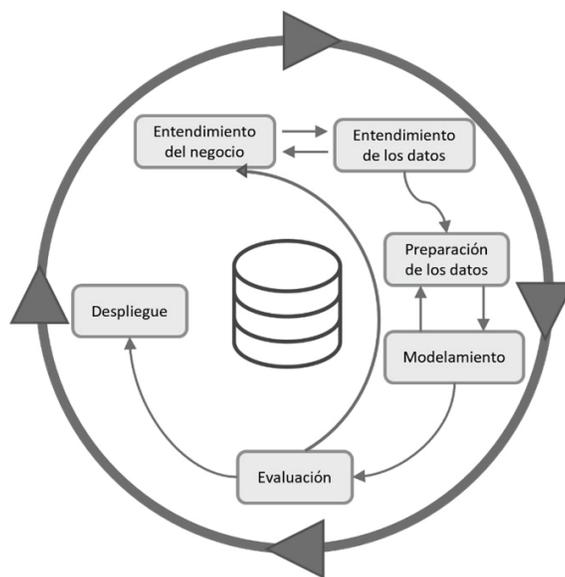


Figura 1. Metodología CRISP-DM.

2.1 Creación del dataset

La parte significativa del desarrollo de un modelo de interpretación empieza con la creación del dataset. En este caso, se eligieron diez palabras de uso común esenciales para establecer una comunicación en tono amigable e interactuar con las personas durante diferentes momentos del día. Estas palabras se dividieron en dos conjuntos: Saludos e identificación (Tabla 1), que incluyen las palabras “Hola”, “Buenos” y sus derivaciones en “Días”, “Tardes” y “Noches”; junto con las palabras “Yo”, “Nombre”, “Años”, “Gustar” y “Licor”.

En las grabaciones se contó con la participación de 70 voluntarios; 40 hombres y 30 mujeres, no expertos en Lengua de Señas Colombiana, con edades entre 18 y 25 años. A los participantes se les instruyó el procedimiento durante 10 minutos, además de solicitar su consentimiento informado para ser grabados. La grabación se efectuó con un dispositivo móvil Huawei P20 lite, a una resolución de 640x480 píxeles y 30 fotogramas por segundo, en un fondo claro sin control de iluminación ni vestimenta. La grabación se realizó continuamente dando un tiempo de descanso entre seña y seña de dos

segundos, para evitar cansancio y confusión en los participantes.

Tabla 1. Palabras para la comunicación LSC.

Saludos	Identificación
Hola	Yo
Buenos	Nombre
Días	Años
Tardes	Gustar
Noches	Licor

Se obtuvieron 70 videos de aproximadamente un minuto, que fueron previsualizados para determinar la calidad de los datos. Mediante el uso de la librería OpenCV, se realizó la extracción de seis cuadros en formato “.jpg”, lo cual permitió captar la transición de las señas. Para evitar el exceso de información no relevante, se recortó la imagen alrededor del cuerpo y redimensionó a un formato más accesible para los modelos de inteligencia artificial, con tamaño de 255x255 píxeles, sin tener en cuenta el aspecto radial de la imagen. Esta colección de imágenes se

codificó y almacenó en carpetas donde cada una lleva el nombre de "Per+Numero", mientras que las imágenes fueron nombradas como "Per + Numero + Etiqueta + Numero de cuadro".

2.2 Modelamiento

Las imágenes pueden alimentar modelos estáticos directamente. En contraste, los modelos dinámicos deben ser alimentados como secuencia de tiempo, disminuyendo la cantidad de información y reemplazándola por características que se adjuntan a un tensor de entrada. La extracción de características se realiza mediante redes neuronales convolucionales pre-entrenadas para evitar el trabajo de ajuste fino de la red, facilitando el entrenamiento y obtención de características. Lo anterior teniendo en cuenta que en previos entrenamientos se han grabado pesos en la red, que son útiles en la disminución del tiempo de entrenamiento y adaptación a nueva información.

La técnica anterior se conoce como aprendizaje por transferencia y existen diversas arquitecturas de donde se pueden obtener características de imágenes. En este caso se eligió VGG16 por su simplicidad y buen desempeño ante repositorios de información de gran magnitud. Es así como se probaron las dos condiciones, un modelamiento estático compuesto por la red VGG16, que recibe las imágenes RGB en tamaño 255x255 pixeles y se modifica la última capa densa a un total de diez neuronas. El modelo dinámico es una combinación de la red pre-entrenada, a la cual se le eliminó las capas densas y se estableció en la capa de aplanamiento, que cuenta con las características de las imágenes, obteniendo 25.088 características que alimentan a un conjunto de 128 módulos LSTM que manejan la información como serie de tiempo, conectada a una capa densa de diez neuronas de salida. El resumen de estos enfoques se visualiza en las Figuras 2 y 3.

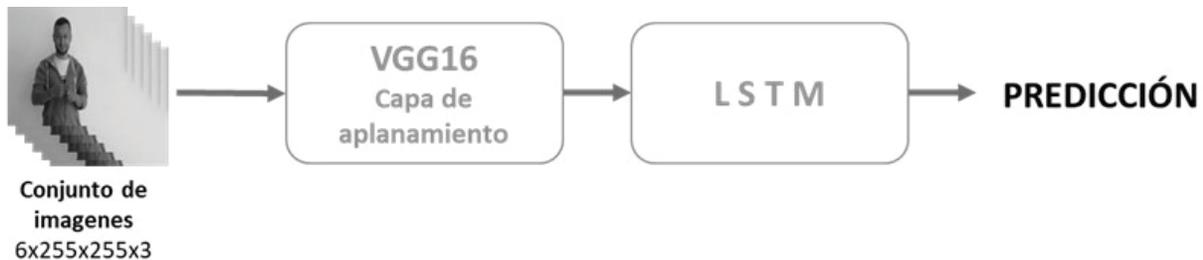


Figura 2. Metodología CRISP-DM.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 255, 255, 3)]	0
block1_conv1 (Conv2D)	(None, 255, 255, 64)	1792
block1_conv2 (Conv2D)	(None, 255, 255, 64)	36928
block1_pool (MaxPooling2D)	(None, 127, 127, 64)	0
block2_conv1 (Conv2D)	(None, 127, 127, 128)	73856
block2_conv2 (Conv2D)	(None, 127, 127, 128)	147584
block2_pool (MaxPooling2D)	(None, 63, 63, 128)	0
block3_conv1 (Conv2D)	(None, 63, 63, 256)	295168
block3_conv2 (Conv2D)	(None, 63, 63, 256)	590880
block3_conv3 (Conv2D)	(None, 63, 63, 256)	590880
block3_pool (MaxPooling2D)	(None, 31, 31, 256)	0
block4_conv1 (Conv2D)	(None, 31, 31, 512)	1180160
block4_conv2 (Conv2D)	(None, 31, 31, 512)	2359808
block4_conv3 (Conv2D)	(None, 31, 31, 512)	2359808
block4_pool (MaxPooling2D)	(None, 15, 15, 512)	0
block5_conv1 (Conv2D)	(None, 15, 15, 512)	2359808
block5_conv2 (Conv2D)	(None, 15, 15, 512)	2359808
block5_conv3 (Conv2D)	(None, 15, 15, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 512)	12845568
dropout (Dropout)	(None, 512)	0
fc2 (Dense)	(None, 512)	262656
dropout_1 (Dropout)	(None, 512)	0
output (Dense)	(None, 10)	5130

=====
 Total params: 27,828,042
 Trainable params: 13,113,354
 Non-trainable params: 14,714,688

a.

Layer (type)	Output Shape	Param #
input_6 (InputLayer)	[(None, 6, 255, 255, 3)]	0
time_distributed_2 (TimeDistributed)	(None, 6, 25088)	14714688
lstm_1 (LSTM)	(None, 128)	12911104
dense_1 (Dense)	(None, 10)	1290

=====
 Total params: 27,627,082
 Trainable params: 15,272,202
 Non-trainable params: 12,354,880

b.

Figura 3. Resumen de los parámetros: (a) Modelos VGG16, (b) VGG16+LSTM.

Para evitar las condiciones de sobre-entrenamiento, se utilizaron técnicas de regularización para obtener el mejor modelo posible considerando la estructura del modelo. En el desarrollo estático se empleó la técnica de "DataAugmentation", para crecer el número de datos empleando las siguientes configuraciones: rotación de 15 grados, desplazamientos horizontales y verticales, acercamiento de un 30% y un efecto espejo que mejora las condiciones de interpretación sin importar la mano dominante. Para el enfoque dinámico, no se consideró la generación adicional de datos para preservar la coherencia y

sentido lógico de las transiciones entre cuadros; sin embargo, se empleó la regularización L2 para aprovechar sus beneficios de penalización de los pesos más grandes que componen el modelo, permitiendo que no haya un sobreajuste a los datos de entrenamiento. Los entrenamientos se realizaron a través de una máquina virtual, proporcionado por la herramienta en línea Google Colab en su versión paga; los recursos obtenidos fueron un disco 150Gb, una memoria RAM de 80Gb y una GPU Nvidia V100, recursos suficientes para realizar el entrenamiento de los modelos.

2.3 Evaluación

La evaluación de los modelos se realizó con las métricas empleadas en el estado del arte, que exponen el desempeño ante datos desconocidos por la red.

Matriz de confusión: Contiene información sobre la predicción de los datos de prueba y establece los errores del modelo al momento de realizar clasificaciones, muestra los aciertos y desaciertos cometidos en cada una de las categorías. Las mejores predicciones se dan al tener la gran mayoría de datos sobre la diagonal de la matriz (Mindlin, 2021).

Accuracy: Es una métrica que hace una descripción general del rendimiento en comparación a todas las clases, su cálculo se realiza respecto al valor de la suma de los verdaderos positivos y negativos con respecto a todas las respuestas de los datos de prueba (IBM, 2021).

Precisión: Es la relación entre el número de muestras positivas clasificadas correctamente sobre el número total de muestras clasificadas como positivas, ya sean correctas o incorrectas (IBM, 2021).

Sensibilidad: Esta métrica de evaluación mide la capacidad del modelo de detectar muestras positivas, se calcula como la relación entre el número de verdaderos positivos entre la suma del total de verdaderos positivos y falsos negativos (IBM, 2021).

Puntuación F1: Es una combinación de las métricas Precisión y sensibilidad, sirviendo para entender el rendimiento combinado entre varias soluciones, su decisión se establece como el doble producto de la precisión por la sensibilidad sobre la suma de estos mismos (IBM, 2021).

3. Resultados y discusión

La base de datos se nombró como LSC-W70 y está compuesta por 4.200 imágenes de tamaño 640x480 píxeles en formato "jpg", correspondientes a seis cuadros por cada clase. También se incluye un subconjunto de tamaño 255x255 píxeles bajo las mismas condiciones, donde se capta la transición de cuadros en algunas de las señas realizadas por los voluntarios (Figura 4).



Figura 4. Imágenes de LSC_W70.

Para alimentar los modelos de interpretación, se realizó una disposición de los datos de la siguiente manera: 50 personas para entrenamiento, diez para validación y diez para prueba, cuyos participantes son diferentes a los empleados en la etapa de entrenamiento. Con esta división de datos se establece un porcentaje entre entrenamiento y prueba del 70% y 30%, respectivamente. Para evitar las condiciones de sobre-entrenamiento y así obtener los mejores estimados de las redes, se empleó la técnica de regularización "early stopping". En el enfoque estático se utilizaron las seis imágenes dispuestas directamente, mientras que para el modelo dinámico las imágenes se adjuntaron en un tensor, teniendo la forma (6 Fotogramas, Ancho, Alto, 3 canales RGB), necesario para trabajar la información como una serie de

tiempo. Estos formatos deben conservarse para realizar el despliegue en vivo, considerándose las técnicas de "early" y "late fusión" para los modelos dinámico y estático, respectivamente.

El accuracy de ambos entrenamientos convergieron a valores específicos (Figura 5), donde las curvas de entrenamiento y prueba se aproximan. En el caso dinámico, la diferencia máxima corresponde a un 10%, siendo la mejor aproximación del entrenamiento del modelo. Además, se destacan las paradas de entrenamiento para el modelo estático en 56 épocas y 35 épocas para el modelo dinámico, logrando una estabilización del entrenamiento en pocos pasos gracias a la riqueza de la información con la que se alimentan los modelos.

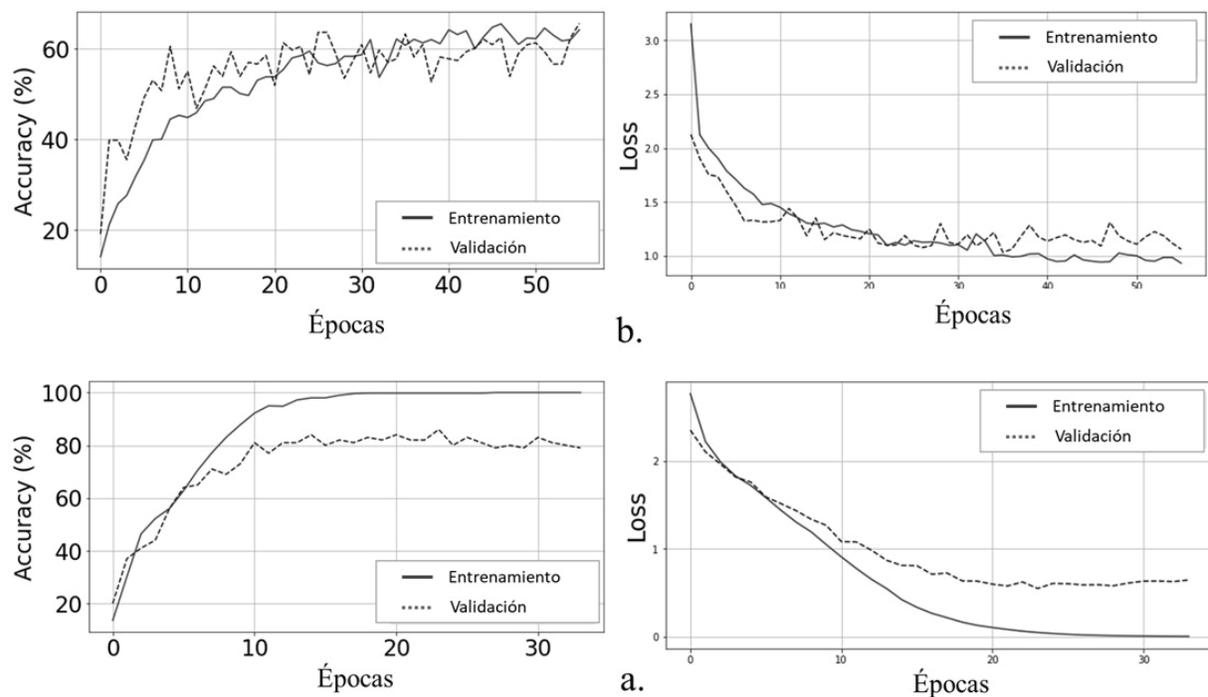


Figura 5. Entrenamiento y validación de los modelos: (a) Estático, (b) dinámico.

Estos resultados son influenciados directamente por la metodología, donde el proceso de desarrollo se realizó varias veces hasta obtener este resultado. Lo anterior permitió probar distintos prototipos de modelos para obtener el mejor desarrollo, haciendo énfasis en el estado del arte y ajuste de los datos para mejorar los

rendimientos de modelos siguientes, con lo cual, el proceso se enfoca no solo en el mejor desarrollo, sino en la supervisión, estructura, planificación del proyecto y mejora respecto al estado del arte. Los resultados de ambos entrenamientos se resumen en la Tabla 2, donde se especifican las métricas de evaluación.

Tabla 2. Resultados de las métricas de evaluación.

Modelo	Estático: VGG16	Dinámico: VGG16+LSTM
Épocas	56	35
Puntuación F1	65,2	77,3
Precisión	66,6	82,3
Sensibilidad	66,0	76,0
Accuracy	66,5	76,0

En la figura 6 se presentan las matrices de confusión, donde la gran mayoría de los datos están presentes sobre la diagonal principal de la matriz, logrando un buen resultado promedio de la red mayor que el 65% de verdaderos positivos en la mayoría de los casos. En la tabla 3 se compara el accuracy de varios modelos reportados en la literatura con el modelo propuesto en esta investigación, que arrojó un valor de 76% frente al 68% reportado por Ortiz-Farfán y Camargo-Mendoza (2020). Esto indica una mejora en el modelo para la detección del LSC, empleando sistemas dinámicos pre-entrenados.

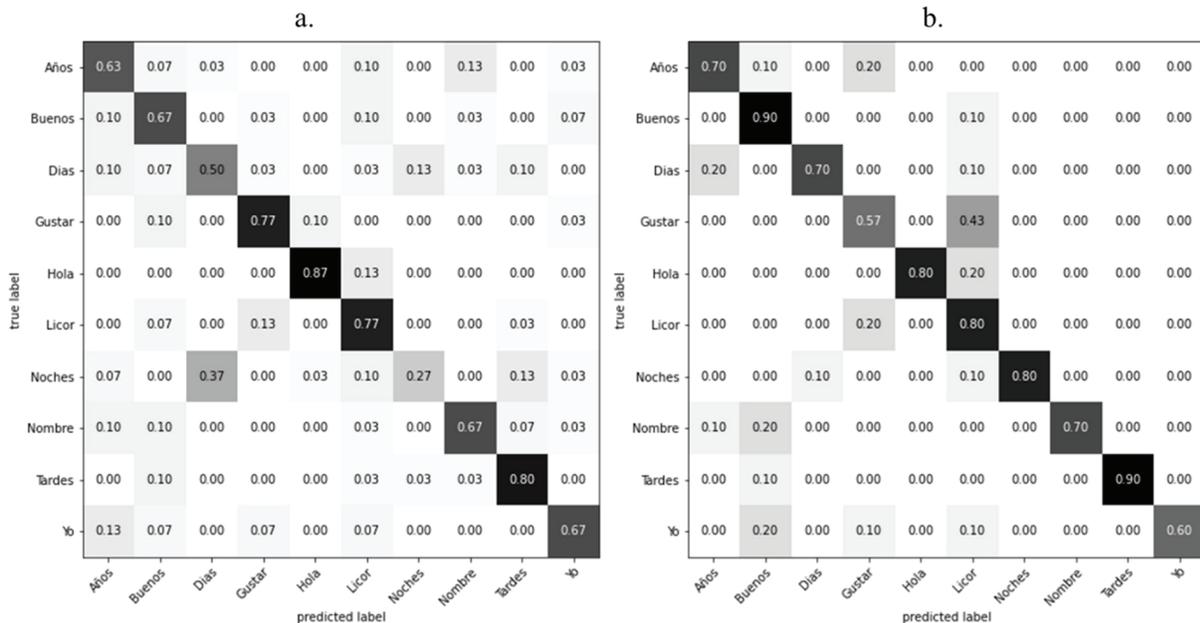


Figura 6. Matriz de confusión de los modelos: (a) Estático, (b) Dinámico.

Los resultados indican que el modelo aprendió a generalizar ante datos desconocidos, y se aprecia la diferencia de mejora del modelo dinámico en comparación al modelo estático, no solo en la cantidad de épocas en las cuales converge, sino que logra una clasificación más apropiada de señas similares, como “Días” y “Noches”, donde la secuencia y sentido del movimiento define la palabra. En este sentido, el enfoque dinámico posee una ventaja superior al desarrollo estático, debido a que resuelve en

un solo modelo la interpretación de todo tipo de señas, permitiendo una comunicación más precisa y fluida de la lengua de señas, mejorando así la comunicación y la comprensión entre personas sordas y oyentes. No obstante, se presentan algunos problemas en reconocer la palabra “gustar” confundiéndola con “Licor”, debido a que la seña se realiza en posiciones similares donde la transición de cuadros es difícil de identificar.

En cuanto al estado del arte, se han superado resultados que manejan la misma lengua de señas (LSC) y de la base de datos estadounidense WLASL (Vaghasiya, 2021). No obstante, en (Mindlin, 2021), el valor precisión de su clasificación es superior al aquí desarrollado, lo cual es consecuencia de la base de datos que emplearon los autores, en donde se usaron guantes de colores que ayudan al reconocimiento de sus gestos. Por tanto, en investigaciones que emplean dos dataset para verificar sus modelos (Boháček & Hruz, 2022; Konstantinidis et al., 2018; Vaghasiya, 2021), estas reportan resultados casi perfectos con el repositorio de lengua de señas argentino.

Si se tiene en cuenta el número de etiquetas, hay una diferencia significativa para realizar una

comparación por la cantidad de estas mismas que corresponden a cada investigación, pero se debe considerar que estos trabajos se realizan sobre un ambiente controlado, lo que facilita el reconocimiento y mejora de los resultados. Es por ello por lo que se ha intentado compensar trabajando en un ambiente semi-controlado, que representa un reto mayor para realizar interpretaciones de lengua de señas. En este sentido, se ha logrado que los modelos dinámicos se establezcan como una alternativa de interpretación más correcta para las palabras que presentan movimientos y gran condensación de información. La comparación con el estado del arte se resume en la Tabla 3.

Tabla 3. Comparación con el estado del arte.

Referencia	Modelo	País	Voluntarios	Etiquetas	Accuracy %
(Ortiz-Farfán & Camargo-Mendoza, 2020)	CNN - Transfer learning	Colombia	5	22	68.00
(Konstantinidis et al., 2018)	Transfer learning-optical Flow y Pose	Estados Unidos	97	100	69.33
(Mindlin, 2021)	CNN + LSTM	Argentina	10	64	99.40
(Vaghasiya, 2021)	Transfer Learning + LSTM	Estados Unidos	97	100	71,11
(Boháček & Hruz, 2022)	Pose + Transformer	Estados Unidos	97	100	63.18
Propio	VGG16 + LSTM	Colombia	70	10	76.00

4. Conclusiones

Se demostró que algunos modelos pre-entrenados en combinación con redes recurrentes LSTM son útiles para interpretar la lengua de señas colombiana. El resultado de los modelos probados en esta investigación supera a los reportados previamente en la literatura, que ma-

nejan la misma lengua de señas y características de interpretación. En este sentido, la aplicación de estos algoritmos permitirá el desarrollo de herramientas de interpretación que se adapten al entorno, reduciendo las brechas en la comunicación, brindando una inclusión asertiva, caracterizada por el uso de nuevas tecnologías.

Como trabajo futuro, se debería incluir una persona experta en el tema para que de su concepto y retroalimentación en el desarrollo de los modelos. Además, se debería verificar la generalización de la información con nuevos participantes, lo cual permitirá proponer un modelo de mayor robustez y adaptabilidad, que pueda ser desplegado en plataformas de acceso móvil.

Referencias

- Boháček, M., & Hruz, M. (2022). Sign Pose-based Transformer for Word-level Sign Language Recognition. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 182-191. <https://doi.org/10.1109/WACVW54805.2022.00024>
- Díaz, C., Goycoolea, M., & Cardemil, F. (2016). HIPOA-CUSIA: TRASCENDENCIA, INCIDENCIA Y PREVALENCIA. *Revista Médica Clínica Las Condes*, 27 (6), 731-739. <https://doi.org/10.1016/j.rmcl.2016.11.003>
- Galvis-Serrano, E. H., Sánchez-Galvis, I., Flórez, N., & Zabala-Vargas, S. (2019). Clasificación de Gestos de la Lengua de Señas Colombiana a partir del Análisis de Señales Electromiográficas utilizando Redes Neuronales Artificiales. *Información Tecnológica*, 30 (2), 171-180. <https://doi.org/10.4067/S0718-07642019000200171>
- IBM. (febrero 27, 2021). Documentation. <https://www.ibm.com/docs/en/cloud-paks/cp-data/3.0.1?topic=overview-accuracy>
- INSOR, (2021). Informe técnico Estado Goce en Derechos de la Población sorda 2019.
- INSOR. (2011). Diccionario básico de la lengua de señas colombiana. Imprenta nacional de Colombia. http://www.insor.gov.co/descargar/diccionario_basico_completo.pdf
- Jimenez, G., Moreno, E., Guzman, R., & Barrero, J. (2019). Automatic method for Recognition of Colombian Sign Language for vowels and numbers from zero to five by using SVM and KNN. *2019 Congreso Internacional de Innovación y Tendencias En Ingeniería (CONIITI)*. <https://doi.org/10.1109/coniiti48476.2019.8960695>
- Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2018). A deep learning approach for analyzing video and skeletal features in sign language recognition. *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. <https://doi.org/10.1109/ist.2018.8577085>
- Martínez, F., Robayo-Betancourt, F., & Arbulú, M. (2020). A gesture recognition system for the Colombian sign language based on convolutional neural networks. *Bulletin of Electrical Engineering and Informatics*, 9 (5), 2082-2089. <https://doi.org/10.11591/eiei.v9i5.2440>
- Mindlin, I. (2021). *Reconocimiento de Lengua de Señas con redes neuronales recurrentes*. Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/129853>
- Monsalve-Pineda, K. V., & Polo-Álvarez, J. E. (2016). *Traductor de símbolos de alfabeto del lenguaje de signos colombiano al lenguaje escrito*. Universidad del Valle. <https://bibliotecadigital.univalle.edu.co/handle/10893/17348>
- Ortiz-Farfán, N., & Camargo-Mendoza, J. E. (2020). Modelo computacional para reconocimiento de lenguaje de señas en un contexto colombiano. *Tecnológicas*, 23 (48), 197-232. <https://doi.org/10.22430/22565337.1585>
- Suat-Rojas, N. E., Montoya-Serna, B. S., Pinzón-Velásquez, E. M., & Rodríguez-Galeano, O. S. (2021). Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales. *Orinoquia*, 25 (1), 25-30. <https://doi.org/10.22579/20112629.680>
- Vaghasiya, D. (2021). *Dynamic gesture classification of American Sign Language using deep learning*. Laurentian University of Sudbury. <https://zone.biblio.laurentian.ca/jspui/handle/10219/3843>
- Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). *Python Deep Learning: Exploring Deep Learning Techniques and Neural Network Architectures with PyTorch, Keras, and TensorFlow*. Packt Publishing.