

Estudio Avanzado del Movimiento Asociado a la Lengua de Señas Colombiana (LSC)

Emmanuel González González, Fredy Enrique Segura Quijano, Luis Felipe Giraldo, Cristian Cifuentes-De La Portilla

e. gonzalezg, fsegura, lf.giraldo404, cc.cifuentes@uniandes.edu.co

Diciembre de 2024

Resumen – La lengua de señas colombiana es parte importante de la identidad cultural de la población sorda. Este estudio se centró en analizar las particularidades de los movimientos de los señantes en video para distinguir entre el discurso de personas sordas y el de intérpretes, y cómo estas diferencias afectan el reconocimiento de las señas en la Lengua de Señas Colombiana (LSC). Para ello, se recolectaron muestras de la LSC durante interacciones de sus usuarios, así como la ejecución de palabras y su uso en frases. Además, se llevó a cabo un análisis de la base de datos LSC50 utilizando técnicas de aprendizaje automático (ML) para establecer una metodología de estudio de la LSC. Los resultados identificaron las características más importantes para la clasificación entre sordos e intérpretes y el reconocimiento de señas, destacando la eficacia de modelos como la red neuronal recurrente (RNN) y el Random Forest. *Palabras Clave* – LSC, Machine Learning, DeepLearning, Visión por computador.

I. INTRODUCCIÓN

Se estima que más del 5% de la población global, lo que equivale a unos 430 millones de personas, experimentan una pérdida auditiva considerable que afecta sus habilidades comunicativas [1]. En Colombia, la Encuesta Nacional de Calidad de Vida (ENCV) de 2019 identificó aproximadamente a 509 mil individuos con discapacidad auditiva, según informes del Departamento Administrativo Nacional de Estadística (DANE) [2]. Esta cifra es comparable con la población total de algunas ciudades, lo que resalta la relevancia de investigar las lenguas de señas no solo desde la lingüística sino desde diferentes campos del conocimiento como la sociología, biomecánica, ingeniería y ciencias de datos. Estos estudios no solo apoyan el reconocimiento de las lenguas de señas como el principal medio de comunicación para la comunidad sorda, sino que también las valoran como un elemento distintivo de su identidad cultural [3].

En el capítulo 3 del libro “Linguistics of American Sign Language”, se menciona una de las características que hacen que la lengua sea única, y es que los símbolos que lo componen se pueden descomponer en partes más pequeñas [4]. En el caso de las lenguas habladas, esas unidades contrastivas son los sonidos, y los lingüistas estudian cómo se estructuran y organizan los sonidos de una lengua. Esta característica la posee la lengua de señas y esas partes, conocidas como parámetros, incluyen la forma de la mano, el movimiento, la ubicación, la orientación y las señales no manuales (expresiones faciales).

La importancia de estos parámetros se basa en que diferencias en uno solo de ellos pueden cambiar completamente el significado de una seña. Por ejemplo, las señas “Respetar” y “Responder” comparten todos los parámetros excepto la ubicación, lo que resulta en significados diferentes. Esto muestra la necesidad de identificar y entender cada parámetro para interpretar correctamente las señas. Por otro lado, los modelos lingüísticos de Brentari [5] y Johnson [6] proponen sistemas de clasificación para las señas basándose en aspectos como el movimiento, detenciones, ubicación y complejidad de las señas. La “complejidad” de una forma de palabra en la LSC se define como la cantidad de características que se combinan para realizar una seña. Cuantas más características se superpongan, mayor será la complejidad de la seña. Brentari (1993) describe la complejidad de las señas clasificándolas en dos grupos: rasgos inherentes (posición de las manos) y rasgos prosódicos (movimientos), resumiendo las clasificaciones en la Tabla 1.

Las investigaciones lingüísticas realizadas sobre la lengua de señas colombiana (LSC) han descubierto una notable diversidad de dialectos y una amplia gama de características gestuales y manuales [7], [8]. Además, herramientas como el diccionario básico de la LSC [9] facilitan la descripción y categorización de las señas. El entendimiento de la LSC se ha profundizado mediante el análisis de su estructura gramatical, su vocabulario y la investigación de cómo los usuarios nativos e intérpretes procesan y se comunican utilizando esta lengua viso-gestual [10]. En investigaciones desde la pedagogía muestran como los patrones de atención visual son importantes, pues impactan directamente en la efectividad comunicativa. Por ello, identificar las diferencias en estos patrones entre individuos sordos nativos y sus intérpretes, permite descubrir variaciones significativas en la interpretación y uso de la LSC, lo cual puede resultar en avances importantes para la educación, la interpretación y la interacción social en la comunidad sorda [11]. Así, esta área de estudio promete mejorar las prácticas de interpretación y ampliar la comprensión de la dinámica lingüística de la LSC.

Estudiar cómo las particularidades de movimiento y contexto afectan la distribución de los parámetros formacionales de la LSC proporcionan una comprensión más detallada de sus características distintivas y podrían permitir el reconocimiento entre sordos nativos e intérpretes de la LSC [11]. Entender estas diferencias es crucial para mejorar la calidad interpretativa y garantizar la transmisión precisa y fiel de los mensajes. En trabajos previos como el de Andrés Flores con su base de datos

LSC50 [12], muestran un análisis semántico de señas con ayuda de algoritmos recientes de Machine Learning (ML) usando modelos transformes visuales los cuales ofrecen una extracción de características adaptables a los datos con menos parámetros [13], en el cual logra agrupar un conjunto de 50 señas dada la energía cinética obtenida con sensores inerciales (IMUs) y movimiento observado en video [14]. Edison en su estudio de reconocimiento en tiempo real de la LSC logra la identificación del alfabeto en base a video 3D con tecnología Real Sense [15]. Tecnologías como HETAH realiza la tarea de traducción de LSC a español hablado y escrito, sin embargo, el proceso contrario sigue presentándose como un reto tecnológico.

En este estudio se buscó evaluar cómo las particularidades de los movimientos de los señantes en video pueden distinguir el discurso de personas sordas del de los intérpretes y cómo estas diferencias afectan el reconocimiento de las señas. Para esto se propusieron dos trabajos realizados en paralelo: en primer lugar, la elaboración de una base de datos multimodal, única en su tipo, con información relevante para el análisis de la LSC desde distintas disciplinas como la lingüística, la pedagogía, la visión por computadora, la biomecánica y la ciencia de datos; en segundo lugar, el desarrollo de un protocolo para el futuro análisis de la base creada, el cual se probará con la base de datos LSC50. Con el fin de brindar al lector una perspectiva de lo que encontrará en este documento, en la Figura 1 se muestra el flujo de trabajo realizado.

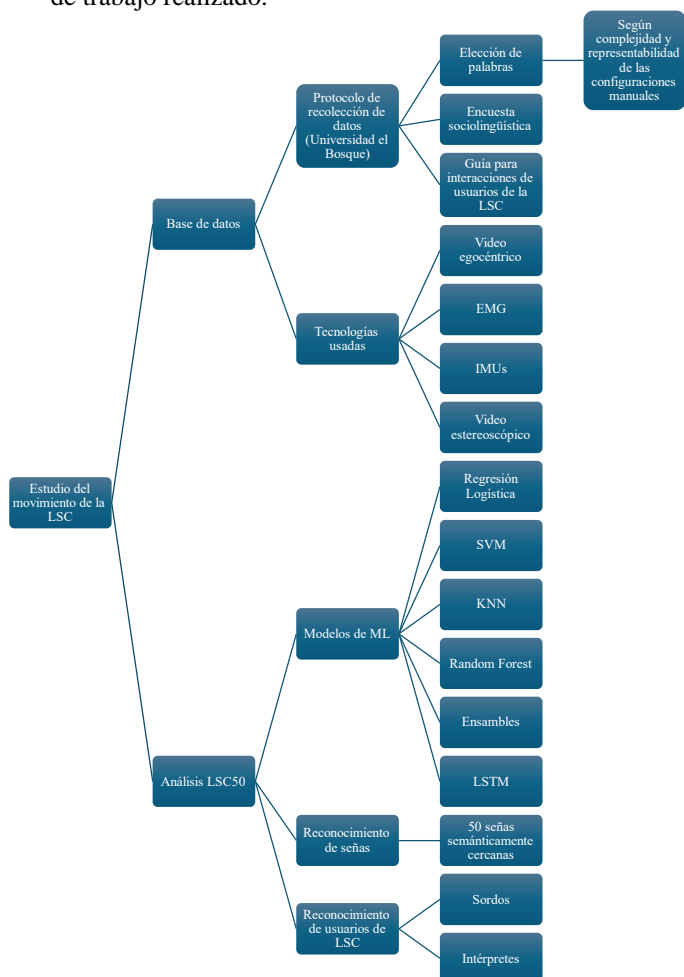


Figura 1. Guía de trabajo para el estudio del movimiento relacionado a la lengua de señas colombiana.

Tipo	Manos	Detalle
Tipo 0	Una mano	Sin contacto con el cuerpo
Tipo X	Una mano	Con contacto con el cuerpo
Tipo I	Dos manos	Simétricas y alternantes
Tipo II	Dos manos	Misma forma de mano, diferente movimiento en ellas
Tipo III	Dos manos	Diferentes formas de manos
Tipo C	Combinación de secuencias	Combinación de tipos

Tabla 1. Clasificación de la complejidad de las señas según modelo prosódico de Brentari.

II. BASE DE DATOS EgoLSC50

EgoLSC50 es una base de datos multimodal que recopila información sobre interacciones de usuarios de la LSC en diversos contextos. Incluye la ejecución de señas individuales y frases breves dentro de un ambiente controlado. Esta base de datos alberga contenido en formatos de video y video egocéntrico, así como datos recogidos por sensores inerciales y de electromiografía. Hasta el momento, cuenta con la participación de 6 voluntarios y registra 3 interacciones distintas.

A. Estrategia de recolección y herramientas tecnológicas:

Para la recolección de datos, se buscó a voluntarios sordos y a intérpretes de LSC, todos ellos mayores de edad y residentes en Bogotá. Se les aplicó una encuesta para recabar su información sociolingüística y se les presentó un consentimiento informado. Este documento estipula su aceptación para ser grabados y que su imagen sea de acceso público. Los documentos de consentimiento informado y las instrucciones se proporcionaron en español y en LSC a través de videos. Además, durante la recolección de datos, un intérprete estuvo presente en todo momento para facilitar la comunicación entre los investigadores y los voluntarios.

Las herramientas tecnológicas utilizadas en la elaboración de la base de datos fueron:

- Cámara estereoscópica Zed 2i.
- Cámaras GoPro Hero 10 Black.
- Gafas Aria de Meta.
- Sensores inerciales Movella Xdot.
- Sensores de electromiografía Delsys Trigno Avanti.

La ZED 2i es una avanzada cámara estéreo diseñada para la percepción espacial y desarrollo de inteligencia artificial como se muestra en la Figura 2. La cámara Zed 2i empleada para registrar videos en HD 720p a 60 fps durante la toma de datos de señas y frases aisladas. Por otro lado, los IMUs Movella Xdot, que son sensores especializados en el análisis y reporte de cinemática humana, se utilizaron para almacenar datos de cuaterniones, ángulos de Euler y velocidades relativas a los ejes de los sensores. Por la parte de Las gafas inteligentes Aria de Meta, en colaboración con Ray-Ban, son unas gafas de realidad aumentada impulsadas por inteligencia artificial. Estas gafas cuentan con varias cámaras para la grabación y reconstrucción del entorno junto a una cámara interna para eye-tracking. Por último, los sensores de electromiografía Delsys Trigno Avanti permiten grabaciones simultaneas de EMG e IMU.

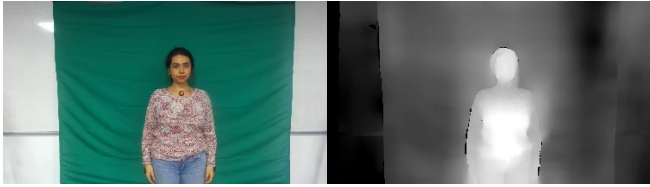


Figura 2 Visión Cámara Zed 2i

B. Protocolo:

La elaboración del protocolo fue realizada en colaboración de profesores y directivos del programa de formación de intérpretes de la Universidad el Bosque. La recolección de datos se estructuró en dos fases distintas. En la primera, se capturaron vídeos de la interacción de parejas de usuarios de la Lengua de Señas Colombiana (LSC). Durante esta etapa, se requirió a los participantes el uso de las gafas Aria, mientras se posicionaban cámaras GoPro en su entorno. Posteriormente, con las gafas Aria y las cámaras ya grabando, se hizo sonar durante 1 segundo una señal sinusoidal de 20 kHz para la sincronización de los videos y se instó a los voluntarios a realizar un ejercicio de actuación. A cada uno se le asignó un rol específico dentro de dos escenarios predeterminados, desarrollando cada uno por un lapso de entre cinco y diez minutos. En caso de que no se completaran grabaciones de al menos 5 minutos se les daba escenarios y roles extra [16].

Una vez finalizada la tarea de actuación, se proyectaron ante los voluntarios dos vídeos educativos del INSOR, estos narraban las historias de Rodolfo el reno y Eurípides. Tras la visualización, se formularon a los voluntarios una serie de preguntas orientadoras con el fin de suscitar un debate [16]. Con la conclusión de esta actividad, se dio por terminada la primera fase y se hizo una pausa en la grabación antes de proceder a la siguiente etapa.

Al comienzo de la segunda etapa de recolección de datos, se colocaron siete sensores inerciales en los centros de masa de la mano, antebrazo, brazo y torso de los voluntarios. Estos centros fueron calculados utilizando las medidas y proporciones individuales, tal como se ilustra en la Figura 3. Adicionalmente, se ubicaron sensores de electromiografía alrededor del bíceps braquial y los flexores profundos de los dedos, también mostrados en la Figura 4. Tras la colocación de los sensores de EMG, se procedió a realizar una prueba de contracción voluntaria máxima a los sujetos de estudio.

Posteriormente, se posicionó a los voluntarios frente a una pantalla verde para capturar su imagen lateralmente con cámaras GoPro y frontalmente con la cámara Zed 2i. Una vez en posición, se inició la grabación simultánea de las cámaras y sensores, y se solicitó a los voluntarios elevar los brazos durante diez segundos; este procedimiento se utiliza para sincronizar los datos de los sensores con los videos. Concluida la fase de preparación, se instruyó a los participantes para que repitieran tres veces una serie de cincuenta señas mostradas individualmente en un video, indicándoles que mantuvieran la mirada en la cámara mientras ejecutaban cada señal y que iniciaran y finalizaran cada una con los brazos en posición baja. Por último, se repitió el proceso con once frases cortas [16].

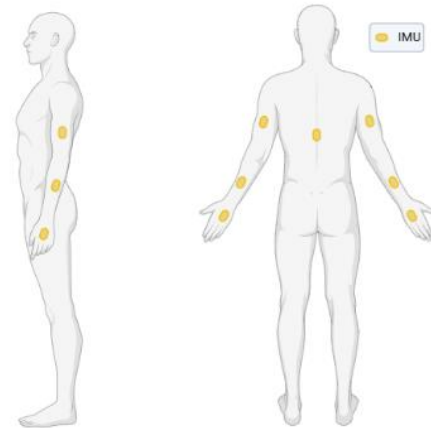


Figura 3 Ubicación de los sensores inerciales creado en BioRender.com

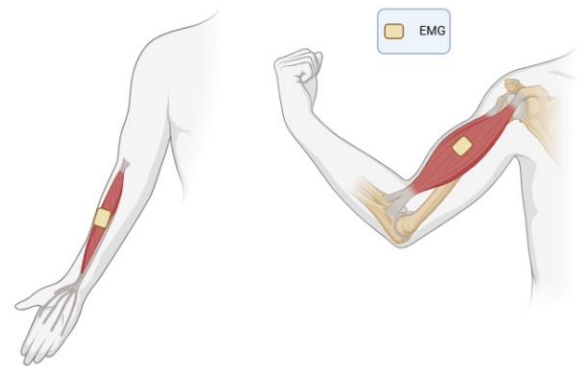


Figura 4. Ubicación de los sensores de EMG en flexor profundo de los dedos y bíceps braquial



Figura 5 Escenario de recolección de las palabras y frases

C. Elección de palabras.

Para la construcción de la base de datos, se seleccionaron cincuenta términos del diccionario básico de la lengua de señas colombiana [9], con el objetivo de incluir señas que representen todas las configuraciones manuales y niveles de complejidad de acuerdo con el modelo prosódico de la fonología de las lenguas

de señas [5]. La clasificación de la complejidad, realizada en conjunto con la lista de palabras, está detallada en el documento denominado “PROTOCOLO LSCPRO”.

50 señas EgoLSC50

Pase Licencia	Pecado	Carretera	Depender	Vida
Lejos	Día	Miércoles	Crear	Norte
Puntual	Sacerdote	Hijo	Seguro	Tener
Comedor	Ocho	Lavadora	Flota vehículo	Camiseta
Internet	Museo	Tres	Armario	Besar
Cien	Barrio	Agosto	Ignorar	Chaqueta
Trancón	Bus	Ayuda	Pasta	Ahora
Gato	Molestar	Flaco	Fe	Cama
herramienta				
Cámara	Fabrica	Ajo	Triste	Cuidar
Fotográfica				
Parque	Nadar	Celoso	Poco	Generoso

Tabla 2. Lista de palabras base de datos EgoLSC50

III. PROCESAMIENTO BASE DE DATOS LSC50

En el marco del proceso de recopilación de datos en curso, se efectuó un análisis preliminar de la base de datos LSC50, con el propósito de desarrollar un protocolo metodológico para futuras evaluaciones cuando el conjunto de datos se encuentre más consolidado.

A. Preprocesamiento videos LSC50.

La base de datos LSC50 se compone de registros en video de cinco voluntarios, quienes ejecutaron cuatro repeticiones de cincuenta señas semánticamente afines. El grupo de voluntarios estuvo conformado por dos hombres y tres mujeres, incluyendo a dos intérpretes de lengua de señas y tres personas con sordera. En cada fotograma correspondiente a las repeticiones de señas, se extrajeron puntos de referencia anatómicos del cuerpo de los señantes utilizando las herramientas MediaPipe Pose y MediaPipe Hand [17]. Estas aplicaciones permitieron capturar la información posicional de 75 puntos distintos del cuerpo (21 por cada mano y 33 entre cuerpo y rostro), sin embargo, se trabajó únicamente con la ubicación en x y y de cada marcador anatómico omitiendo la coordenada z . Tras completar la adquisición de los marcadores, se procedió a igualar la longitud de las señales interpolando hasta alcanzar 134 fotogramas por señal. De esta manera, se conformó un arreglo tridimensional con dimensiones (1000,134,150), que servirá como entrada para los modelos de redes neuronales recurrentes (RNN). En una etapa posterior, para el entrenamiento de los métodos clásicos de aprendizaje automático (ML), se extrajeron 156 características de cada señal utilizando la librería TSFEL. Estas características comprenden métricas de dominio temporal, frecuencial, estadístico y fractal, resultando en un arreglo final de dimensiones (1000,23400) [18].

B. Modelos de Aprendizaje Automático.

El modelo seleccionado consiste en una red neuronal recurrente con seis capas ocultas. Las tres primeras capas contienen 134 unidades de memoria a corto y largo plazo (LSTM por sus siglas en inglés), seguidas de dos capas completamente conectadas con función de activación ReLu. La capa final es una capa de salida completamente conectada que utiliza una función de activación softmax, tal como se ilustra en la Figura 6. Esta configuración se diseñó para el procesamiento y análisis de los datos secuenciales que necesita manejar el

modelo. Adicionalmente, para clasificación de señas se entrenaron modelos clásicos de ML, los modelos elegidos fueron SVM, KNN, Random Forest y un ensamble de estos.

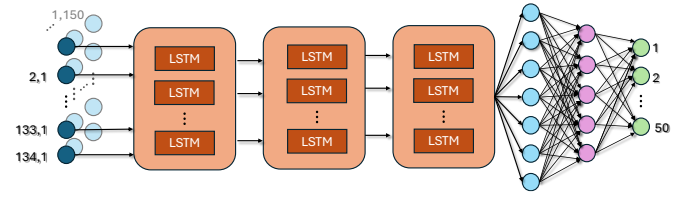


Figura 6. Estructura red neuronal recurrente tipo LSTM

C. Clasificación de señas.

Para el entrenamiento de los modelos, se propusieron inicialmente particiones aleatorias y balanceadas de los datos, asignando un ochenta por ciento para el entrenamiento y un veinte por ciento para la validación. Con cada modelo se realizó una búsqueda de maya para encontrar los mejores hiper parámetros y se entrenó el mejor de cada uno utilizando diez particiones diferentes, después se calculó el promedio de la precisión de cada modelo, como se muestra en la Tabla 3.

Con el objetivo de evaluar si los modelos pueden reconocer las señas realizadas por los señantes vistos durante el entrenamiento, se reservó una repetición de las señas de cada voluntario para la validación, utilizando las restantes para el entrenamiento de los modelos. Esto resultó en un veinticinco por ciento de los datos destinados a la validación y un setenta y cinco por ciento para el entrenamiento. Se rotaron las cuatro repeticiones de los señantes y se calculó el promedio de las precisiones, como se muestra en la Tabla 5.

Para evaluar si los modelos son capaces de generalizar a señantes nunca vistos, se entrena con las señas de 4 voluntarios y se reserva uno para validación. Se realizan diferentes entrenamientos rotando el voluntario de validación, obteniendo las precisiones mostradas en la Tabla 3.

Resultados de clasificación de 50 señas con una partición balanceada aleatoria

Resultados reservando una repetición de las señas de cada voluntario

Modelo	Accuracy	Accuracy
SVM	0.565	0.507
KNN	0.540	0.673
Random Forest	0.600	0.824
Ensamble	0.675	0.848
LSTM	0.245	0.451

Tabla 3. Precisión de modelos reservando una repetición de las señas de cada voluntario.

B. Reconocimiento entre nativos e intérpretes.

Con el propósito de diferenciar el discurso entre hablantes nativos e intérpretes, se cambió la capa de salida de la RNN de 50 neuronas con activación softmax a una con activación sigmoide y se adicionó la regresión logística a los métodos clásicos. Al tener un total de 24,300 características para la clasificación binaria y solo mil muestras, se realizó una reducción de dimensionalidad con PCA, como se muestra en la

Figura 9. Los modelos se entrenaron siguiendo el proceso para reconocimiento de señas, realizando particiones aleatorias, una búsqueda de los mejores hiperparámetros y promediando la precisión de cada modelo (Tabla 4).

Resultados clasificación Nativo – Intérprete con 3 componentes principales

Modelo	Accuracy	Accuracy PCA
SVM	0.995	0.995
KNN	0.985	0.995
Random Forest	0.995	0.985
Ensamble	0.995	0.985
Regresión Logística	0.995	0.985
LSTM	0.731*	0.856

Tabla 4. Precisión de los modelos con los 3 principales componentes de PCA.

El modelo de RNN presentaba un comportamiento variable dependiendo de la partición realizada, lo que provocaba que cerca de un tercio de los entrenamientos siempre clasificara como sordos. Al analizar la importancia normalizada de las características ofrecidas por el algoritmo de Random Forest, se observó que el noventa y ocho por ciento de la distribución se encuentra en solo tres características (Figura 7). Por lo tanto, se procedió a entrenar 48,000 modelos de árboles de decisión, cada uno con una partición de datos aleatoria y balanceada.

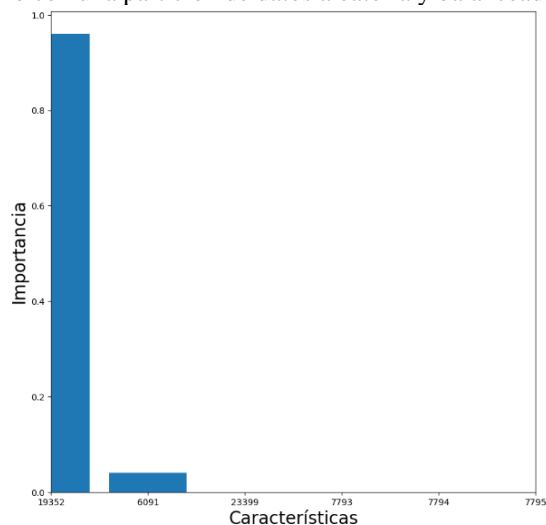


Figura 7. Importancia de las características según árboles de decisión.

Resultados de clasificación de 50 señas: accuracy tomando como datos de validación un voluntario y entrenando con los demás.

Modelo	Voluntario 1	Voluntario 2	Voluntario 3	Voluntario 4	Voluntario 5
SVM	0.030	0.130	0.385	0.315	0.415
Random Forest	0.070	0.235	0.460	0.420	0.420
Knn	0.135	0.080	0.250	0.190	0.240
Ensamble	0.075	0.200	0.415	0.355	0.380
LSTM	0.015	0.025	0.079	0.105	0.549

Tabla 5. Precisión reservando un voluntario para validación.

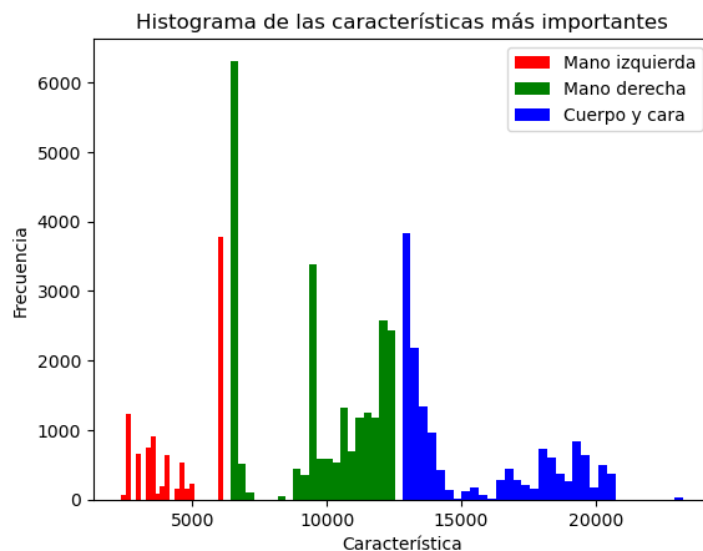


Figura 8. Histograma obtenido al entrenar 48,000 árboles de decisión y obtener el índice de la característica más importante.

Al analizar las características más importantes, se observa que, de las 24,300 características, solo 234 fueron identificadas como las más relevantes, y el percentil 90 incluye 67 de ellas. Al ubicar los marcadores, el eje y la característica a los cuales pertenecen estas 67 características se encontró que 47 corresponden al movimiento en el eje x, 45 a los marcadores de la parte externa de ambas manos, 22 a los marcadores del rostro y 32 a la característica "percentil 1 de la función acumulada de probabilidad empírica" (ECDF_1). Al entrenar los modelos utilizando la característica ECDF_1 de los 3 marcadores del meñique derecho y el marcador del centro del rostro (8 características) con cinco particiones aleatorias y balanceadas, se obtuvo el promedio de las precisiones de la Tabla 6. En esta, se destaca la RNN por lograr una precisión casi perfecta con un promedio de 0.998.

IV. ANÁLISIS

Durante el procesamiento de la base de datos y la evaluación inicial de clasificación de señas y usuarios de LSC, se evidenciaron limitaciones metodológicas relacionadas con la distribución asimétrica de las realizaciones (sordo – intérprete) y el número restringido de muestras (20 por seña) que comprometen el análisis y la generalización de los modelos,

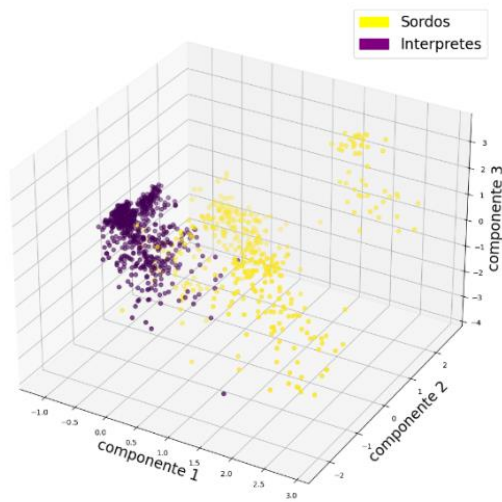


Figura 9. Visualización de los 3 primeros componentes principales.

Resultados Clasificación Sordo - Interprete		
Modelo	Accuracy	Accuracy 8 Características
SVM	0.995	0.995
KNN	0.985	0.990
Random Forest	0.995	0.985
Ensamble	0.995	0.990
Regresión Logística	0.995	0.990
LSTM	0.731*	0.998

Tabla 6. Precisión teniendo en cuenta solo las 8 características más importantes.

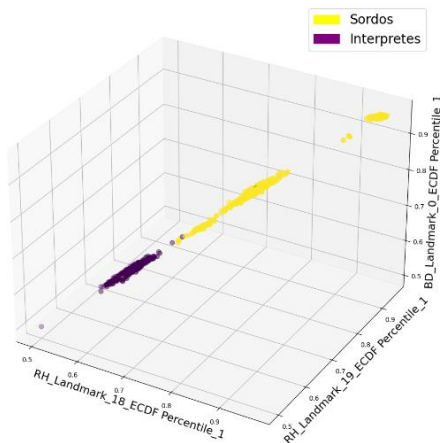


Figura 10. Visualización de las 3 características más importantes.

especialmente la RNN. Sin embargo, se observó que los modelos, al realizar la partición aleatoria, tuvieron un rendimiento mínimo diez veces superior al aleatorio. Cuando se rotaron los voluntarios para el entrenamiento y validación se encontró que, al usar usuarios sordos para el entrenamiento y los intérpretes para la validación, la precisión de los modelos decae considerablemente, llegando casi a una precisión aleatoria. Por el contrario, al realizar el entrenamiento incluyendo los datos y características de los voluntarios intérpretes, los modelos son capaces de generalizar para los sordos. Esto sugiere que los movimientos realizados por intérpretes ofrecen una mayor cantidad de información que permite generalizar el reconocimiento de las señas para todo tipo de usuarios, y que existe una diferencia notable en el discurso entre sordos e intérpretes. Por último, respecto al reconocimiento de señas, cuando se presentan a los modelos al menos una vez las señas de todos los voluntarios y reservar una de las repeticiones para validación, estos son capaces de reconocerlas con una precisión superior a las particiones anteriores, al lograr mejoras de hasta el veinticuatro por ciento. Esto sugiere que los modelos reconocen las particularidades de cada usuario de la LSC que observa.

En el intento por identificar el tipo de usuario de la LSC a partir de las señas ejecutadas, los modelos clásicos de aprendizaje automático (ML) lograron diferenciar a los usuarios con una precisión promedio del 98% al 99%. Esto respalda la hipótesis de que existen diferencias distintivas entre los discursos de personas sordas y los de intérpretes. No obstante, la dependencia del modelo de redes neuronales recurrentes (RNN) en los datos de entrenamiento sugiere que el patrón que permite diferenciar usuarios es complejo de extraer en la serie temporal. Al realizar la reducción de dimensionalidad de las características utilizadas en los modelos clásicos de ML y graficar los tres componentes que más explican la varianza de los datos en relación con el tipo de usuario se observó que, podrían ser fácilmente diferenciables con tan solo esas tres componentes sin perder una cantidad significativa de precisión, siendo el máximo un uno por ciento e incluso mejorando en algunos modelos como la KNN y RNN. Al utilizar los árboles de decisión para estimar la importancia de cada una de las características en la tarea de reconocimiento de usuarios, se encontró una relación entre la amplitud de los movimientos de la cabeza y parte distal de las manos con el tipo de usuario de la LSC observando que los sordos tienden a tener una amplitud mayor, lo que permitió una mejora en los desempeños de los modelos logrando clasificar perfectamente los usuarios de la LSC con el modelo LSTM.

V. CONCLUSIONES

El presente estudio ha permitido identificar y analizar características de los movimientos asociados a las señas en la Lengua de Señas Colombiana utilizando técnicas avanzadas de ML. A pesar de las limitaciones metodológicas, como el tamaño reducido de la muestra y el desequilibrio de datos, los modelos de clasificación han demostrado ser capaces de identificar patrones significativos en las señas. En particular, dos modelos se destacaron: la red neuronal recurrente (RNN), que ha mostrado un rendimiento

notable al reconocer el discurso entre usuarios sordos e intérpretes, y el Random Forest, eficaz para el reconocimiento de señas. Este trabajo muestra la necesidad de continuar desarrollando la base de datos, extendiendo la cantidad de voluntarios de forma equilibrada entre usuarios, para mejorar la precisión y generalización de los modelos de aprendizaje automático en el análisis de la LSC. También, los hallazgos sugieren que los movimientos realizados por intérpretes ofrecen una mayor cantidad de información para el reconocimiento de señas. Además, el que los marcadores de la parte exterior de las manos y centro de la cabeza nos ofrezcan información suficiente para el reconocimiento de usuarios ofrece información relevante para el diseño de herramientas pedagógicas para la formación de futuros intérpretes.

En trabajos futuros se plantea la inclusión y comparación de modelos con las señales recolectadas de EMG e IMUs, junto al entrenamiento de modelos más reciente como Transformers de video o redes GRU ya que, estos modelos Gated Recurrent Unit y LSTM son considerados mejores que los LSTM en ciertos contextos debido a su arquitectura más simple, que reduce el número de parámetros y hace que sean más rápidos de entrenar y menos costosos computacionalmente. Además, los GRU tienden a necesitar menos datos de entrenamiento y pueden alcanzar un rendimiento comparable o superior al de los LSTM en secuencias de datos más cortas o medianas.

VI. REFERENCIAS

- [1] “Sordera y pérdida de la audición”. Consultado: el 15 de septiembre de 2024. [En línea]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- [2] L. M. Niño Hernández y K. P. Jaraba Carmona, “Caracterización de Ciudadanos Usuarios y Grupos de Interés INSOR 2021”, <https://www.insor.gov.co/home/descargar/Caracterizacion-de-Ciudadanos-Usuarios-y-Grupos-de-Interes-INSOR-2021.pdf>.
- [3] L. A. Tovar, “La importancia del estudio de las lenguas de señas”, *Lenguaje*, sep. 2020, doi: 10.25100/lenguaje.v0i28.14117.
- [4] Robbin Battison, “Signs Have Parts: A Simple Idea”, en *Linguistics of American Sign Language, 5th Ed*, Gallaudet University Press, 2020, p. 242. doi: 10.2307/j.ctv2rh2b9v.37.
- [5] D. Brentari, *A Prosodic Model of Sign Language Phonology*. The MIT Press, 1999. doi: 10.7551/mitpress/5644.001.0001.
- [6] R. E. Johnson y S. K. Liddell, “Toward a Phonetic Representation of Signs: Sequentiality and Contrast”, *Sign Lang Stud*, vol. 11, núm. 2, pp. 241–274, 2010.
- [7] C. B. Y. Milena y B. M. A. Giovanni, “Variación sociolingüística en la lengua de señas colombiana: observaciones sobre el vocabulario deportivo, en el marco de la planificación lingüística”, 2013. [En línea]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/49156>
- [8] E. Marely y P. Bello, “ESTUDIO DESCRIPTIVO DE LA LENGUA DE SEÑAS COLOMBIANA (LSC) INVESTIGACIÓN DOCENTE YENNY RODRÍGUEZ HERNÁNDEZ”, 2011, Consultado: el 15 de septiembre de 2024. [En línea]. Disponible en: <https://repositorio.iberu.edu.co/handle/001/539>
- [9] I. Caro, C. Departamento, y D. E. Lexicografía, “DICCIONARIO BÁSICO DE LA LENGUA DE SEÑAS COLOMBIANA”.
- [10] L. A. Tovar, “La definición en la lengua de señas colombiana (LSC)”, *Lenguaje*, vol. 45, núm. 2, pp. 383–418, jul. 2017, doi: 10.25100/lenguaje.v45i2.5277.
- [11] E. Hirshorn, “Visual Attention and Deafness”, Washington, DC, may 2011.
- [12] A. Flórez, L. F. Giraldo, C. Cifuentes-De, y L. Portilla, “Colombian sign language analysis and recognition”, 2022, Consultado: el 16 de septiembre de 2024. [En línea]. Disponible en: <http://hdl.handle.net/1992/64165>
- [13] A. Vaswani *et al.*, “Attention Is All You Need”, 2023. [En línea]. Disponible en: <https://arxiv.org/abs/1706.03762>
- [14] C. Cifuentes, L. Giraldo, y B. Solórzano, “LSC50: Colombian Sign Language Video and Inertial Measurement Dataset”, oct. 2024, doi: 10.6084/m9.figshare.27383016.v1.
- [15] E. Orlando, M. López, y F. Lozano, “Universidad de los Andes”, 2016.
- [16] E. González, “Protocolo LSCPRO”, BOGOTA D.C., jun. 2024.
- [17] “Framework de MediaPipe | Google AI Edge | Google AI for Developers”. Consultado: el 10 de noviembre de 2024. [En línea]. Disponible en: <https://ai.google.dev/edge/mediapipe/framework?hl=es-419>
- [18] M. Barandas *et al.*, “TSFEL: Time Series Feature Extraction Library”, *SoftwareX*, vol. 11, ene. 2020, doi: 10.1016/j.softx.2020.100456.