Apoyo a la comunicación y acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG y UNIMINUTO mediante un aplicativo web

Yeferson Audelo Cordoba Gomez, Juan Carlos Cordoba Rodriguez

Karen Yisenia Rodriguez Rosero

Programa de Ingeniería, Facultad de Sistemas, Universidad CESMAG

Nota del autor:

El presente trabajo de grado, desarrollado en el marco de la Vicerrectoría de Investigación de la Universidad CESMAG, tiene como propósito desarrollar un aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana, como apoyo a la comunicación y al acceso a la información de estudiantes con discapacidad auditiva. La información y los resultados podrán ser utilizados con fines académicos y de investigación. Los autores declaran no presentar conflictos de intereses relacionados con el proyecto.

Para correspondencia, puede contactarse al Programa de Ingeniería de Sistemas de la Universidad CESMAG mediante el correo: **ingenieriadesistemas@unicesmag.edu.co**

Apoyo a la comunicación y acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG y UNIMINUTO mediante un aplicativo web

Yeferson Audelo Cordoba Gomez, Juan Carlos Cordoba Rodriguez

Karen Yisenia Rodriguez Rosero

Programa de Ingeniería, Facultad de Sistemas, Universidad CESMAG

Asesor: Mg. Diego Javier Villarreal Moreno

10 de agosto de 2026

Nota de aceptación

"Va la aceptación definida desde la universidad".

Nota de exclusión

DEDICATORIA

Dedico este trabajo a Dios, y reconozco que toda la gloria es para Él, por brindarme la fortaleza, la sabiduría y la perseverancia necesarias para culminar esta etapa tan importante de mi vida.

A mi familia y conocidos que me ayudaron incondicionalmente, en especial a mi padre, Rodrigo Córdoba Guerrero, quien fue el promotor de todo este proceso y un pilar fundamental para alcanzar mi título universitario. Así mismo, a la familia Castillo y Rodríguez Rosero, por su apoyo, confianza y acompañamiento a lo largo de este camino.

A una persona muy especial en mi vida, mi novia Karen Rodríguez Rosero, por su amor, apoyo constante, comprensión y motivación en los momentos más desafiantes de este proceso.

A mis docentes y mentores, por compartir su conocimiento, orientación y compromiso durante todo este proceso académico.

Y finalmente, a todas aquellas personas que, de una u otra manera, contribuyeron a la realización de este proyecto, brindándome ánimo, confianza y motivación para no rendirme.

Este logro no es solo mío, sino de todos quienes hicieron parte de este camino.

Yeferson C.

Este proyecto de grado representa mucho más que un requisito académico; es el resultado de un camino lleno de esfuerzo, aprendizaje y crecimiento personal. Por ello, quiero dedicar estas líneas a quienes han sido parte fundamental de este logro.

En primer lugar, a Dios, por brindarme la fortaleza, la sabiduría y la oportunidad de llegar hasta este punto. Por acompañarme en cada paso y darme la capacidad de superar los desafíos que se presentaron en el camino.

A mis padres, quienes han sido el pilar más importante en mi vida. Gracias por su apoyo incondicional, por sus consejos, su paciencia y por creer en mí incluso en los momentos en los que yo dudaba. Este logro también es suyo.

A mi familia, que de una u otra forma estuvo presente, brindándome ánimo, comprensión y motivación durante todo este proceso. Cada palabra de aliento y cada gesto de apoyo fueron esenciales para llegar hasta aquí.

A mi pareja, por acompañarme en este camino, por su comprensión en los momentos de ausencia, por su apoyo constante y por impulsarme a seguir adelante incluso en los momentos más difíciles.

A todas las personas que hicieron parte de este proceso, docentes, compañeros y quienes aportaron de manera directa o indirecta a la realización de este proyecto. Su conocimiento, guía y apoyo fueron fundamentales para hacerlo posible.

Y finalmente, me dedico este logro a mí mismo, por la constancia, la disciplina y la resiliencia demostrada a lo largo de este proceso. Por no rendirme, por seguir adelante a pesar de los obstáculos y por creer que era posible alcanzar esta meta.

Este proyecto no solo marca el final de una etapa, sino también el inicio de nuevos retos y oportunidades.

Juan C.

A Dios, por ser la luz que guió cada uno de mis pasos, incluso en los momentos en los que el camino parecía incierto. Por darme la fortaleza para continuar cuando el cansancio pesaba, la serenidad para afrontar cada desafío y la fe necesaria para no rendirme. Por sostenerme en silencio cuando sentía que no podía más y recordarme que cada esfuerzo tenía un propósito. Este logro es, en gran parte, reflejo de su infinita gracia en mi vida.

A mis padres, Roberto Rodriguez y Jhaneth Rosero, quienes son el verdadero fundamento de este logro. No existen palabras suficientes para describir la magnitud de su amor, ni el valor de todo lo que han hecho por mí. Este proyecto no solo representa mis horas de estudio y dedicación, sino también sus sacrificios silenciosos, sus renuncias, su trabajo constante y su fe inquebrantable en mi futuro. Detrás de cada avance estuvieron ustedes, sosteniéndome incluso cuando no lo notaba, dando más de lo que tenían para que yo pudiera seguir adelante. Gracias por darme oportunidades que ustedes quizá no tuvieron, por acompañarme en cada etapa, por impulsarme cuando sentía miedo y por enseñarme, con su ejemplo, que los sueños se construyen con esfuerzo, disciplina y amor. Cada paso que doy lleva su huella, cada logro tiene su esencia, y este título es, sin duda, también suyo.

A Yunior, mi compañero felino, mi guardián de madrugadas y testigo silencioso de incontables noches de desvelo. Con su presencia tranquila y su compañía constante, estuvo ahí en cada etapa de este proceso, acompañándome incluso en los momentos más exigentes, como si entendiera cada reto que enfrentaba. Fue refugio en medio del estrés, calma en los días difíciles y compañía fiel en largas jornadas donde el cansancio parecía ganar. Hoy sé que sigue estando, de otra manera, pero con la misma esencia, acompañando cada uno de mis logros. Su amor no se ha ido, solo ha cambiado de forma, y permanece conmigo en cada paso que doy. Este triunfo también le pertenece, porque su huella es imborrable en este camino.

A Armandito y Copito, mis compañeros felinos, quienes llegaron en la etapa final, pero lograron convertirse en un apoyo inesperado y profundamente valioso. En medio del estrés, las responsabilidades y los días de mayor exigencia, su compañía, su ternura y esos pequeños momentos de tranquilidad que brindan, llenaron de calma y alegría mi entorno. Me recordaron, incluso en medio de la presión, que siempre hay espacio para respirar, para sonreír y para seguir adelante.

A Yeferson, por ser un apoyo constante en este camino, por su compañía, su paciencia y por estar presente en los momentos en los que más lo necesité. Gracias por escucharme, por animarme cuando el cansancio aparecía y por creer en mí incluso en los días en los que yo dudaba. Tu presencia fue un impulso importante para seguir adelante y este logro también lleva una parte de ese apoyo que me brindaste.

Finalmente, a mí misma, por no rendirme, por mantenerme firme a pesar del cansancio, las dudas y los desafíos. Por cada noche de esfuerzo, cada lágrima contenida, cada momento en el que pensé en rendirme, pero decidí continuar. Por la valentía de seguir, incluso cuando el camino no era claro, y por creer, aunque fuera en silencio, que podía lograrlo. Porque llegar hasta aquí no fue fácil, pero ha valido completamente la pena, y este logro es prueba de que todo esfuerzo, al final, florece.

Karen R.

AGRADECIMIENTOS

Expresamos nuestro más sincero agradecimiento a Dios, por guiarnos durante este proceso académico, brindarnos fortaleza en los momentos de dificultad y permitirnos culminar esta etapa tan importante de nuestra formación profesional.

A nuestras familias, por su apoyo incondicional, comprensión, paciencia y motivación constante. Gracias por estar presentes en cada paso de este camino, por creer en nuestras capacidades y por impulsarnos a seguir adelante aun en los momentos más exigentes.

A nuestros docentes y asesores, por su orientación, compromiso y acompañamiento durante el desarrollo de este proyecto de grado. Sus conocimientos, recomendaciones y aportes fueron fundamentales para fortalecer tanto la investigación como el desarrollo de la propuesta tecnológica.

A la Universidad CESMAG y a la Corporación Universitaria Minuto de Dios — UNIMINUTO, por brindarnos los espacios académicos, formativos e institucionales necesarios para llevar a cabo este proceso investigativo.

De manera especial, agradecemos a las personas que participaron en las pruebas, validaciones y actividades relacionadas con el proyecto, especialmente a quienes aportaron desde su conocimiento y experiencia en la Lengua de Señas Colombiana. Su colaboración fue fundamental para orientar el desarrollo del aplicativo y valorar su funcionamiento en un contexto real de uso.

Finalmente, agradecemos a cada integrante de este equipo por su compromiso, responsabilidad, esfuerzo y perseverancia. Este proyecto representa el resultado de un trabajo conjunto, construido con dedicación, aprendizaje constante y el deseo de aportar, desde la tecnología, a la inclusión y accesibilidad comunicativa de las personas con discapacidad auditiva.

Yeferson, Juan y Karen

TABLA DE CONTENIDO

Pág.

Contenido

INTRODUCCIÓN 23
I. PROBLEMA DE INVESTIGACIÓN 25
A. OBJETO O TEMA DE ESTUDIO 25
B. LÍNEA DE INVESTIGACIÓN 25
C. PLANTEAMIENTO DEL PROBLEMA 26
D. FORMULACIÓN DEL PROBLEMA 28
E. OBJETIVOS 28
1) Objetivo general 28
2) Objetivos específicos 28
F. JUSTIFICACIÓN 28
G. DELIMITACIÓN 31
II. MARCO TEÓRICO 32
A. ANTECEDENTES 32
1. Antecedentes Internacionales 32
2. Antecedentes Nacionales 33
3. Antecedentes Regionales (Nariño) 33
B. SUPUESTOS TEÓRICOS 34
C. VARIABLES DEL ESTUDIO 40
D. FORMULACIÓN DE HIPÓTESIS 42
1. Hipótesis de investigación 42
2. Hipótesis nula 42
3. Hipótesis alterna 43
III. METODOLOGÍA 44
A. PARADIGMA 44
B. ENFOQUE 44
C. MÉTODO 44
D. TIPO DE INVESTIGACIÓN 45
E. DISEÑO DE LA INVESTIGACIÓN 45
F. POBLACIÓN 46
G. MUESTRA 46

H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN 46
I. VALIDEZ DE LAS TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN 48
J. CONFIABILIDAD DE LAS TÉCNICAS DE RECOLECCIÓN 48
K. INSTRUMENTO DE RECOLECCIÓN DE DATOS 49
IV. RESULTADOS DE LA INVESTIGACIÓN 51
A. ANÁLISIS DE TÉCNICAS Y METODOLOGÍAS PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA 51
B. DESARROLLO DE UN APLICATIVO WEB PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA (LSC), UTILIZANDO LAS TÉCNICAS Y METODOLOGÍAS PREVIAMENTE ANALIZADAS. 67
C. VALIDAR EL SOFTWARE CON UNA PRUEBA DE CONCEPTO APOYADA POR UN INTÉRPRETE DE LENGUA DE SEÑAS. 125
V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS 129
A. ANÁLISIS DE TÉCNICAS Y METODOLOGÍAS PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA 129
B. DESARROLLO DE UN APLICATIVO WEB PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA (LSC), UTILIZANDO LAS TÉCNICAS Y METODOLOGÍAS PREVIAMENTE ANALIZADAS. 135
C. VALIDAR EL SOFTWARE CON UNA PRUEBA DE CONCEPTO APOYADA POR UN INTÉRPRETE DE LENGUA DE SEÑAS. 144
CONCLUSIONES 149
RECOMENDACIONES 152
REFERENCIAS BIBLIOGRÁFICAS 153
ANEXOS 160
Anexo. 2 Carta de validación de intérprete. 160
Anexo. 3 Carta de validación de persona sorda. 160
Anexo. 4 Validación con estudiante sordo 160
Anexo. 5 Pruebas de funcionamiento 160
Anexo. 6 Videos con muestra de pruebas 163
Anexo. 7 Registro de asistencia de estudiantes 163
Anexo. 8 Formularios resumen de las encuestas 163
Anexo. 9 Certificado de registro ante el DNDA 163

LISTA DE FIGURAS

Fig. 1 Lengua de Señas Colombiana (LSC) 47
Fig. 2 Arquitectura general de una red neuronal convolucional (CNN) 53
Fig. 3 Estructura general de una red neuronal recurrente (RNN) 54
Fig. 4 Estructura interna de una red LSTM 56
Fig. 5 Estructura de una red BiLSTM con procesamiento bidireccional 57
Fig. 6 Esquema del mecanismo de atención sobre secuencias 58
Fig. 7 Esquema de arquitecturas Transformer 60
Fig. 8 Representación basada en secuencia de fotogramas de video 63
Fig. 9 Representación basada en extracción de puntos clave 64
Fig. 10 Organización de clases 69
Fig. 11 Videos por clase 70
Fig. 12 Data final 70
Fig. 13 Dataset externo no controlado. 71
Fig. 14 . Arquitectura CNN. 74
Fig. 15 Evidencia gráfica del proceso. 75
Fig. 16 Grafica de precisión. 76
Fig. 17 Grafica de perdida. 77
Fig. 18 Matriz de confusión de CNN. 78
Fig. 19 Resultado de inferencia Fuente: Elaboración propia. 79
Fig. 20 Resultado de inferencia 79
Fig. 21 Arquitectura general del modelo Transformer 80
Fig. 22 Evidencia gráfica del proceso. 81
Fig. 23 Gráficos de precisión. 83
Fig. 24 Graficas de perdida. 84
Fig. 25 Matriz de confusión Transformer. 84
Fig. 26 Evidencia gráfica del proceso. 85
Fig. 27 Evidencia gráfica del proceso. 86
Fig. 28 Evidencia gráfica del proceso. 86
Fig. 29 Arquitectura general red Bi-LSTM. 87
Fig. 30 Gráficas de precisión. 89
Fig. 31 Gráficas de perdida. 90
Fig. 32 Matriz de confusión Bi-LSTM. 91
Fig. 33 Evidencia gráfica del proceso. 92
Fig. 34 Evidencia gráfica del proceso. 92
Fig. 35 Evidencia gráfica del proceso. 93
Fig. 36 Arquitectura general de LSTM + Attention. 94
Fig. 37 Evidencia gráfica del proceso. 95
Fig. 38 Gráfica de precisión. 97
Fig. 39 Gráfica de perdida. 97
Fig. 40 Matriz de inferencia interna. 98
Fig. 41 Matriz de inferencia externa. 99
Fig. 42 Evidencia gráfica del proceso. 100
Fig. 43 Evidencia gráfica del proceso. 101
Fig. 44 Evidencia gráfica del proceso. 101
Fig. 45 Evidencia gráfica del proceso. 102

Fig. 46 Salida de inferencia del modelo para adios.....103
Fig. 47 Salida de inferencia del modelo para carnet 103
Fig. 48 Salida de inferencia del modelo para adios.....104
Fig. 49 Salida de inferencia del modelo para cafetería 104
Fig. 50 Resultados de inferencia en back.....109
Fig. 51 Resultados de inferencia en back.....109
Fig. 52 Resultados de inferencia en back.....109
Fig. 53 Resultados de inferencia en back.....110
Fig. 54 Resultados de inferencia en back.....111
Fig. 55 Resultados de inferencia en back.....111
Fig. 56 Resultados de inferencia en back.....112
Fig. 57 Resultados de inferencia en back.....113
Fig. 58 Resultados de inferencia en back.....113
Fig. 59 Primera interfaz.....116
Fig. 60 Versión Final interfaz 116
Fig. 61 Modo de uso incorporado en la interfaz. 117
Fig. 62 Glosario de palabras disponibles en el aplicativo. 117
Fig. 63 Estructura del backend o fragmento del procesamiento de video implementado.....119
Fig. 64 Visualización de la interpretación generada por el sistema en la interfaz. 120
Fig. 65 Validación con intérprete.....125
Fig. 66 Validación con estudiantes UNIMINUTO 126
Fig. 67 Nivel de usabilidad de los usuarios con el aplicativo- CESMAG 126
Fig. 68 Tiempo de respuesta del aplicativo- UNIMINUTO 127
Fig. 69 Nivel de intuitividad 127
Fig. 70 Reconocimiento correcto de la seña 127
Fig. 71 Validación con estudiante sordo.....160
Fig. 72 Validación con estudiante sordo.....160
Fig. 73 Validación con estudiante sordo.....160
Fig. 74 Pruebas de software 161
Fig. 75 Pruebas de software 161
Fig. 76 Pruebas de software 161
Fig. 77 Pruebas de software 162
Fig. 78 Pruebas de software 162
Fig. 79 Pruebas de software 162
Fig. 80 Pruebas de software 163

LISTA DE TABLAS

TABLA. I 61
TABLA. II 65
TABLA. III 66
TABLA. IV 71
TABLA. V 72
TABLA. VI 73
TABLA. VII 106
TABLA. VIII 114
TABLA. IX 118
TABLA. X 123

LISTA DE ANEXOS

Anexo. 1 Resultados de inferencias externas 160
Anexo. 2 Carta de validación de intérprete. 160
Anexo. 3 Carta de validación de persona sorda. 160
Anexo. 4 Validación con estudiante sordo 160
Anexo. 5 Pruebas de funcionamiento 160
Anexo. 6 Videos con muestra de pruebas 163
Anexo. 7 Registro de asistencia de estudiantes 163
Anexo. 8 Formularios resumen de las encuestas 163
Anexo. 9 Certificado de registro ante el DNDA 163
Anexo. 10 Tablero Kanban 163

GLOSARIO

ACCESIBILIDAD DIGITAL: condición que permite que una herramienta tecnológica pueda ser utilizada por personas con diferentes capacidades, facilitando el acceso a la información, la comunicación y la interacción con sistemas digitales de manera clara, funcional e inclusiva.

ACCURACY: métrica utilizada en aprendizaje automático para medir la proporción de predicciones correctas realizadas por un modelo respecto al total de muestras evaluadas. En este proyecto se empleó para observar el desempeño de los modelos durante el entrenamiento y la validación.

APLICATIVO WEB: sistema informático al que se accede mediante un navegador de internet y que permite al usuario interactuar con diferentes funcionalidades sin necesidad de instalar un programa directamente en el dispositivo.

APRENDIZAJE AUTOMÁTICO: área de la inteligencia artificial que permite a los sistemas aprender patrones a partir de datos y mejorar su desempeño en una tarea específica sin ser programados manualmente para cada caso.

ARQUITECTURA DEL MODELO: estructura interna de una red neuronal, conformada por capas, funciones y parámetros que permiten procesar los datos de entrada y generar una salida o clasificación.

BACKEND: componente interno del aplicativo encargado de recibir los videos enviados desde la interfaz, procesarlos, ejecutar el modelo de reconocimiento y devolver una respuesta al usuario.

BI-LSTM: arquitectura de red neuronal recurrente basada en LSTM bidireccional. Procesa la secuencia en dos direcciones, hacia adelante y hacia atrás, permitiendo analizar información pasada y futura dentro de una misma secuencia.

CAPA DENSE: capa completamente conectada de una red neuronal, en la cual cada neurona recibe información de todas las salidas de la capa anterior. Se utiliza para integrar características y apoyar la clasificación final.

CLASIFICACIÓN: proceso mediante el cual un modelo asigna una entrada a una categoría específica. En este proyecto, la clasificación corresponde a la identificación de una palabra en Lengua de Señas Colombiana a partir de un video.

CNN: red neuronal convolucional utilizada principalmente para procesar datos visuales. En este proyecto se evaluó una variante CNN 1D para analizar secuencias de características extraídas de videos.

CONFIANZA DE PREDICIÓN: valor porcentual que indica qué tan segura es la salida generada por el modelo frente a una clase determinada. Una confianza alta indica mayor seguridad del modelo respecto a la predicción realizada.

DATASET: conjunto de datos organizado y etiquetado que se utiliza para entrenar, validar y probar un modelo de inteligencia artificial. En este proyecto estuvo conformado por videos de palabras en Lengua de Señas Colombiana.

DISCAPACIDAD AUDITIVA: condición relacionada con la pérdida parcial o total de la capacidad auditiva, la cual puede generar barreras de comunicación y acceso a la información cuando no existen apoyos adecuados.

ENTRENAMIENTO DEL MODELO: proceso mediante el cual una red neuronal aprende patrones a partir de datos etiquetados. Durante esta etapa, el modelo ajusta sus parámetros para mejorar su capacidad de clasificación.

ÉPOCA: recorrido completo que realiza el modelo sobre el conjunto de datos de entrenamiento. En cada época, el modelo actualiza sus parámetros con el fin de mejorar su desempeño.

EXTRACCIÓN DE CARACTERÍSTICAS: proceso mediante el cual se obtiene información relevante de una imagen o video para ser utilizada por un modelo de reconocimiento. En este proyecto se realizó mediante puntos clave del cuerpo y las manos.

FRAMES: fotogramas o imágenes individuales que componen un video. En el proyecto, los videos fueron representados como secuencias de frames para analizar el movimiento de las señas.

FRONTEND: parte visible del aplicativo web con la que interactúa el usuario. Incluye elementos como botones, formularios, cámara, carga de video, glosario, instrucciones de uso y visualización de la interpretación generada.

FUNCIÓN SOFTMAX: función utilizada en modelos de clasificación multiclase para convertir las salidas del modelo en probabilidades asociadas a cada clase disponible.

INFERENCIA: proceso mediante el cual un modelo ya entrenado recibe nuevos datos de entrada y genera una predicción. En este proyecto, la inferencia ocurre cuando el aplicativo analiza un video y entrega la palabra interpretada.

INTELIGENCIA ARTIFICIAL: área de la informática que permite desarrollar sistemas capaces de aprender a partir de datos, identificar patrones y generar resultados de manera automatizada.

INTÉRPRETE DE LENGUA DE SEÑAS: persona con formación y conocimiento especializado en lengua de señas, encargada de facilitar la comunicación entre personas sordas y personas oyentes.

LANDMARKS: puntos clave detectados en el cuerpo, las manos o el rostro de una persona dentro de una imagen o video. En este proyecto fueron utilizados para representar numéricamente los movimientos de las señas.

LENGUA DE SEÑAS COLOMBIANA (LSC): lengua natural de la comunidad sorda en Colombia, basada en movimientos de las manos, expresiones faciales, orientación corporal y componentes visuales-gestuales con estructura propia.

LOSS: métrica de pérdida que indica el nivel de error del modelo durante el entrenamiento o la validación. Un valor menor de loss representa un mejor ajuste del modelo frente a los datos evaluados.

LSTM: tipo de red neuronal recurrente diseñada para procesar información secuencial y conservar datos relevantes a lo largo del tiempo. En este proyecto fue utilizada para analizar la secuencia de movimientos presentes en las señas.

MARGIN TOP1-TOP2: indicador que representa la diferencia entre la probabilidad de la clase más probable y la segunda clase más probable. Un margen alto indica que el modelo distinguió con mayor claridad la predicción final.

MATRIZ DE CONFUSIÓN: tabla que permite observar el comportamiento de un modelo de clasificación, mostrando cuántas muestras fueron clasificadas correctamente y cuántas fueron confundidas con otras clases.

MEDIAPIPE: framework utilizado para la detección de puntos clave en imágenes y videos. En este proyecto se empleó para extraer landmarks de manos y parte superior del cuerpo.

MECANISMO DE ATENCIÓN: técnica que permite al modelo asignar mayor importancia a los momentos más relevantes de una secuencia, favoreciendo la interpretación de los fragmentos más significativos del gesto.

MODELO DE RECONOCIMIENTO: estructura de inteligencia artificial entrenada para identificar patrones en los datos de entrada y clasificar una seña dentro de una de las palabras contempladas por el sistema.

MOVEMENT: métrica utilizada en el reporte de inferencia para representar el nivel de variación entre frames consecutivos. Permite identificar si la muestra contiene movimiento gestual suficiente para ser interpretada.

NORMALIZACIÓN: proceso de ajuste de los datos para mantener una escala o estructura común antes de ser procesados por el modelo. En este proyecto permitió reducir variaciones asociadas a la posición o tamaño del usuario en el video.

OVERFITTING: fenómeno que ocurre cuando un modelo aprende demasiado los datos de entrenamiento y pierde capacidad para generalizar correctamente ante datos nuevos.

PIPELINE: conjunto ordenado de etapas que permiten procesar los datos desde su entrada hasta la generación de un resultado. En este proyecto incluye recepción del video, extracción de características, normalización, predicción y respuesta final.

PREDICCIÓN: resultado generado por el modelo después de procesar una entrada. En este proyecto, corresponde a la palabra en LSC que el sistema interpreta a partir del video analizado.

PRUEBA DE CONCEPTO: validación inicial realizada para comprobar que una solución tecnológica funciona dentro de un alcance definido y puede responder a la necesidad planteada.

RECALL: métrica de evaluación que indica la capacidad del modelo para identificar correctamente las muestras pertenecientes a una clase específica.

RECONOCIMIENTO DE SEÑAS: proceso mediante el cual un sistema analiza una seña realizada en video y genera una interpretación textual correspondiente a una palabra específica.

RED NEURONAL: modelo computacional inspirado en el funcionamiento del cerebro humano, compuesto por capas de neuronas artificiales que procesan información para aprender patrones y realizar predicciones.

SECUENCIA: conjunto ordenado de datos que conserva una relación temporal. En este proyecto, una secuencia corresponde a una serie de frames o características extraídas de un video de una seña.

TENSORFLOW/KERAS: herramientas de desarrollo utilizadas para construir, entrenar y ejecutar modelos de aprendizaje profundo.

TRANSFORMER: arquitectura de aprendizaje profundo basada en mecanismos de autoatención. Permite modelar relaciones globales dentro de una secuencia, aunque puede requerir mayor cantidad de datos y recursos computacionales.

USABILIDAD: grado en que un sistema puede ser comprendido, aprendido y utilizado de manera clara, sencilla y satisfactoria por los usuarios.

VALIDACIÓN DEL MODELO: proceso mediante el cual se evalúa el comportamiento del modelo con datos que no se usan directamente para ajustar sus parámetros, con el fin de observar su capacidad de generalización.

RESUMEN

El presente trabajo de grado tuvo como propósito implementar un aplicativo web basado en inteligencia artificial para apoyar la comunicación básica y el acceso a la información de estudiantes con discapacidad auditiva en las universidades CESMAG y UNIMINUTO, mediante el reconocimiento de palabras en Lengua de Señas Colombiana (LSC). La investigación surgió a partir de las barreras comunicativas que pueden presentarse en contextos educativos cuando no se cuenta con apoyos suficientes o herramientas tecnológicas accesibles que faciliten la interacción entre estudiantes sordos, docentes y compañeros oyentes [1], [2] .

Metodológicamente, el proyecto se desarrolló bajo un enfoque mixto con predominio cualitativo, articulando la revisión de técnicas y metodologías de reconocimiento, la construcción de un dataset propio, el entrenamiento de modelos de inteligencia artificial, el desarrollo del aplicativo web y la validación mediante una prueba de concepto[3]. Para el reconocimiento de señas se analizaron diferentes arquitecturas de aprendizaje profundo, entre ellas CNN 1D, Transformer, Bi-LSTM y LSTM con mecanismo de atención [4], [5]. Asimismo, se trabajó con secuencias de video procesadas mediante la extracción de puntos clave o landmarks [6], con el fin de representar de manera estructurada los movimientos de las manos y parte superior del cuerpo.

Como resultado, se obtuvo un aplicativo web funcional compuesto por un frontend orientado a la interacción del usuario y un backend encargado de recibir los videos, procesarlos, extraer características y ejecutar el modelo de reconocimiento. La arquitectura LSTM + Attention fue seleccionada como modelo final debido a su comportamiento más consistente durante el entrenamiento, la validación y las pruebas externas de inferencia documentadas. El sistema permitió reconocer un conjunto definido de palabras en LSC, entre ellas adiós, auditorio, ayuda, baño, biblioteca, cafetería, carnet, cuándo y dónde, dentro de las condiciones establecidas para el proyecto.

La validación del software se realizó mediante una prueba de concepto apoyada por una intérprete de Lengua de Señas Colombiana, con la participación de un estudiante sordo y usuarios de las instituciones involucradas. Los resultados permitieron evidenciar que el aplicativo interpreta las señas evaluadas dentro del alcance definido y presenta una interfaz comprensible para los usuarios. Se concluye que la herramienta desarrollada constituye un apoyo tecnológico viable para fortalecer

la comunicación básica en contextos académicos, sin reemplazar la labor del intérprete, sino complementándola mediante una solución digital accesible e inclusiva.

Palabras clave: Accesibilidad digital, aplicativo web, inteligencia artificial, Lengua de Señas Colombiana, reconocimiento de señas.

ABSTRACT

This undergraduate project aimed to implement a web application based on artificial intelligence to support basic communication and access to information for students with hearing disabilities at CESMAG University and UNIMINUTO, through the recognition of words in Colombian Sign Language (LSC). The research emerged from the communication barriers that may appear in educational contexts when there are not enough support resources or accessible technological tools to facilitate interaction among deaf students, teachers and hearing classmates [1], [2].

Methodologically, the project was developed under a mixed approach with qualitative predominance, integrating the review of recognition techniques and methodologies, the construction of a custom dataset, the training of artificial intelligence models, the development of the web application and its validation through a proof of concept [3]. For sign recognition, different deep learning architectures were analyzed, including CNN 1D, Transformer, Bi-LSTM and LSTM with an attention Mechanism [4], [5]. Video sequences were processed through the extraction of key points or landmarks [6], in order to represent hand movements and upper-body information in a structured way.

As a result, a functional web application was obtained, composed of a frontend designed for user interaction and a backend responsible for receiving videos, processing them, extracting features and executing the recognition model. The LSTM + Attention architecture was selected as the final model due to its more consistent behavior during training, validation and documented external inference tests. The system made it possible to recognize a defined set of words in Colombian Sign Language, including goodbye, auditorium, help, bathroom, library, cafeteria, ID card, when and where, within the conditions established for the project.

The software was validated through a proof of concept supported by a Colombian Sign Language interpreter, with the participation of a deaf student and users from the institutions involved. The results showed that the application interprets several of the evaluated signs within the defined scope and presents an understandable interface for users. It is concluded that the developed tool represents a viable technological support to strengthen basic communication in academic contexts, without replacing the role of the interpreter, but rather complementing it through an accessible and inclusive digital solution.

Keywords: artificial intelligence, Colombian Sign Language, digital accessibility, sign recognition, web application.

INTRODUCCIÓN

La inclusión de las personas con discapacidad auditiva parcial o total continúa siendo un desafío para los sistemas educativos en distintos contextos del mundo. En el marco de la Agenda 2030, las Naciones Unidas establecieron los Objetivos de Desarrollo Sostenible como una hoja de ruta para promover sociedades más equitativas, destacándose el ODS 4, orientado a garantizar una educación inclusiva, equitativa y de calidad, y el ODS 10, enfocado en la reducción de las desigualdades [7]. De manera complementaria, la UNESCO ha señalado que, en América Latina, la inclusión educativa sigue enfrentando retos relacionados con el acceso, la permanencia y la participación efectiva de poblaciones históricamente excluidas, entre ellas las personas con discapacidad [8].

A pesar de estos lineamientos internacionales, las barreras comunicativas continúan limitando la participación plena de las personas con discapacidad auditiva en los entornos académicos. Esta situación se hace más evidente cuando las instituciones no cuentan con apoyos humanos suficientes, como intérpretes de lengua de señas, ni con herramientas tecnológicas accesibles que faciliten la interacción, el acceso a los contenidos y el desarrollo de procesos de aprendizaje significativos [2]. En este sentido, la inclusión no depende únicamente del reconocimiento formal del derecho a la educación, sino también de la existencia de condiciones reales que permitan ejercerlo en igualdad de oportunidades.

En Colombia, esta problemática mantiene una alta relevancia. El DANE dispone de estadísticas y procesos de caracterización relacionados con la población con discapacidad, mientras que el INSOR ha venido desarrollando orientaciones y recursos para fortalecer la inclusión educativa de la población sorda, especialmente en los niveles de educación superior [1], [9]. De forma paralela, el marco normativo colombiano ha avanzado mediante disposiciones como la Ley 982 de 2005, que contempla medidas para favorecer el acceso de las personas sordas y sordociegas al sistema educativo con apoyo de interpretación, y la Ley 1618 de 2013, que establece disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad [10], [11]. No obstante, la existencia de estos avances institucionales y jurídicos no elimina por sí sola las barreras de acceso, comunicación y permanencia que aún enfrenta esta población en el ámbito universitario.

En el departamento de Nariño, la situación adquiere un carácter aún más sensible debido a la dependencia que existe frente a los servicios de mediación comunicativa y a la limitada

disponibilidad de apoyos especializados. En este contexto, la referencia a ASORNAR como organización afiliada a FENASCOL permite reconocer la existencia de una comunidad sorda organizada en el departamento, así como la necesidad de fortalecer acciones y mecanismos de accesibilidad que respondan a sus condiciones particulares [12]. Esto pone de manifiesto que la problemática no solo debe entenderse desde una perspectiva nacional, sino también desde realidades regionales donde los apoyos institucionales y tecnológicos pueden resultar insuficientes.

Dentro de este escenario regional, la problemática se concreta en instituciones de educación superior como la Universidad CESMAG y UNIMINUTO, donde el acceso a recursos tecnológicos orientados a apoyar la comunicación de estudiantes con discapacidad auditiva sigue siendo limitado. En estos contextos, las dificultades no se reducen únicamente a la disponibilidad de intérpretes de Lengua de Señas Colombiana (LSC), sino también a la ausencia de herramientas digitales que faciliten la interacción cotidiana entre estudiantes sordos, docentes y compañeros oyentes. Como consecuencia, persisten barreras que afectan la participación académica, el acceso oportuno a la información y el fortalecimiento de procesos de inclusión educativa, lo que justifica la necesidad de proponer soluciones tecnológicas accesibles que contribuyan a mejorar la comunicación y el acompañamiento dentro de estas instituciones.

I. PROBLEMA DE INVESTIGACIÓN

A. OBJETO O TEMA DE ESTUDIO

El desarrollo de esta investigación se llevó a cabo durante los periodos 2025 y 2026-1, en las instalaciones de las Universidades CESMAG y UNIMINUTO, en la ciudad de Pasto. El objetivo principal fue la investigación y desarrollo de una solución tecnológica mediante un aplicativo web basado en inteligencia artificial, que apoye la comunicación y el acceso a la información para estudiantes con discapacidad auditiva. Esta solución logró el prototipado para el reconocimiento de Lengua de Señas Colombiana (LSC). A través de esta herramienta se buscó fomentar la inclusión educativa, la autonomía de los estudiantes con discapacidad y el fortalecimiento de su participación en el ámbito académico de la educación superior.

B. LÍNEA DE INVESTIGACIÓN

Esta investigación se centró en la línea de Tecnologías de la información y la comunicación, ya que contribuye a mejorar los procesos de enseñanza y aprendizaje mediante el uso de tecnologías emergentes. En particular, el proyecto se ajusta a esta línea al haber desarrollado una solución tecnológica orientada a apoyar la comunicación y el acceso a la información de estudiantes con discapacidad auditiva en la Universidad CESMAG, mediante el uso de la Lengua de Señas Colombiana (LSC).

Se trata de una investigación aplicada, enfocada en resolver una problemática real de inclusión educativa, a través del desarrollo de un sistema que facilite la comprensión y expresión del lenguaje mediante medios tecnológicos accesibles y eficientes, promoviendo así una educación más equitativa y de calidad.

Sublínea de investigación

Esta investigación se basó en la sublínea de Lenguajes de Programación, entendida como el desarrollo y aplicación de sistemas digitales que permitieron crear entornos interactivos y funcionales para mejorar la experiencia del usuario. En el contexto de este proyecto, se propuso el diseño e implementación de una interfaz tecnológica que utilice modelos de reconocimiento visual y procesamiento de señas. Esto facilitará a los estudiantes con discapacidad auditiva una comunicación más efectiva en el entorno académico.

Desde la Universidad CESMAG, esta sublínea representa un espacio fundamental para explorar tecnologías inclusivas que integren componentes de programación avanzada y desarrollo de software accesible, con el objetivo de ampliar las oportunidades de aprendizaje y participación en el ámbito universitario.

C. PLANTEAMIENTO DEL PROBLEMA

La Lengua de Señas Colombiana (LSC) es un sistema de comunicación visual-gestual que permite a las personas sordas expresar y comprender ideas mediante movimientos de las manos, expresiones faciales y posturas corporales. Contrario a la creencia común, no es una mera representación gestual del idioma oral, sino una lengua natural con su propia gramática, sintaxis y léxico, desarrollada y evolucionada dentro de las comunidades sordas. Cada país o región posee su propia lengua de señas, adaptada a su contexto cultural y lingüístico.

Este reconocimiento ha sido respaldado a nivel internacional por la Convención sobre los Derechos de las Personas con Discapacidad, adoptada por la Organización de las Naciones Unidas (ONU) en 2006, en la que se insta a los Estados a garantizar el uso e inclusión de las lenguas de señas en todos los ámbitos sociales, culturales y educativo [13].

En Colombia, el proceso de reconocimiento comenzó con la Ley 324 de 1996, que declara a la Lengua de Señas Colombiana (LSC) como la lengua natural de la población sorda en el país [14]. Este avance se fortaleció con la implementación del Decreto 366 de 2009, la Ley 982 de 2005 y la Ley 1618 de 2013, las cuales establecen normas para garantizar la inclusión, accesibilidad y equiparación de oportunidades para las personas con discapacidad auditiva en el ámbito educativo, social y laboral [15] [16] [17].

A pesar de los avances en el reconocimiento de la lengua de señas y la promulgación de leyes inclusivas, las personas sordas en Colombia continúan enfrentando importantes barreras en distintos ámbitos de la vida diaria. En el ámbito social, la falta de intérpretes en espacios públicos y eventos limita su participación activa en la comunidad. En el ámbito laboral, las oportunidades de empleo son escasas debido a la falta de políticas de inclusión efectiva y de tecnologías accesibles que faciliten la comunicación. Culturalmente, el acceso a actividades recreativas, artísticas y deportivas también se ve restringido por la ausencia de adaptaciones en los servicios. Incluso en sectores como la salud y la justicia, donde la comunicación es fundamental, las personas sordas

encuentran obstáculos para recibir atención adecuada o ejercer plenamente sus derechos. Esta exclusión generalizada vulnera su autonomía y limita su desarrollo integral en la sociedad.

Esta situación es particularmente crítica en el ámbito educativo, donde las barreras de comunicación impactan directamente el acceso, la permanencia y el éxito académico de las personas sordas. No obstante, a pesar del sólido marco legal, persisten múltiples barreras para la población sorda, especialmente en el acceso a la educación superior. Según la Comisión de Regulación de Comunicaciones (CRC), más de 459.000 personas en Colombia se reconocen con algún tipo de discapacidad auditiva [18]. Sin embargo, solamente el 5% de los jóvenes sordos accede a la universidad, lo que revela una profunda brecha de inclusión educativa [19].

Entre los factores que acentúan esta brecha se encuentra la limitada disponibilidad de tecnologías accesibles que integren la Lengua de Señas Colombiana (LSC). Aunque existen iniciativas como la aplicación “SEP Investiga”, una herramienta orientada a facilitar procesos de investigación en (LSC) [20], estas no cubren por completo las necesidades comunicativas en entornos universitarios, donde se requiere interacción constante, acceso a contenidos en tiempo real y autonomía en el aprendizaje. Esta carencia tecnológica limita no solo el rendimiento académico, sino también el desarrollo social, emocional y profesional de los estudiantes sordos.

El Instituto Nacional para Sordos (INSOR). entidad adscrita al Ministerio de Educación Nacional que asesora la formulación de políticas para la población sorda, ha señalado que los niveles de implementación real de los derechos de esta comunidad son bajos, especialmente fuera de las grandes capitales [21].

Un caso emblemático de estas dificultades es el del Centro de Relevo, un servicio creado por el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) en alianza con la Federación Nacional de Sordos de Colombia (FENASCOL). Este programa permite a las personas sordas comunicarse con personas oyentes mediante intérpretes de (LSC) en tiempo real, a través de videollamadas, mensajes de texto y plataformas digitales como WhatsApp, brindando un canal vital de comunicación accesible [22]. Sin embargo, en marzo de 2025, el servicio fue suspendido de manera indefinida tras la finalización del convenio entre FENASCOL y el MinTIC. Esta suspensión dejó incomunicadas a más de 500.000 personas sordas en Colombia, generando una fuerte preocupación social y manifestaciones desde la comunidad sorda [23].

Este panorama resulta aún más preocupante cuando se observa el caso de universidades locales como la Universidad CESMAG y UNIMINUTO, instituciones en las que se han identificado estudiantes sordos con barreras de acceso a contenidos, a plataformas tecnológicas inclusivas y a herramientas de aprendizaje adaptadas. La falta de apoyo en este ámbito impide que estos estudiantes tengan las mismas oportunidades para alcanzar su potencial académico y profesional.

Así, aunque existen avances jurídicos y esfuerzos institucionales, las personas sordas en Colombia siguen enfrentando obstáculos estructurales para acceder a una educación superior equitativa y de calidad. Reconocer la magnitud de esta problemática y su impacto en la autonomía, el bienestar y las oportunidades de vida de la comunidad sorda es un paso fundamental para generar conciencia y acción desde la academia, la tecnología y la política pública.

D. FORMULACIÓN DEL PROBLEMA

¿Cómo puede fortalecerse la comunicación básica y el acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG y UNIMINUTO a partir del reconocimiento de palabras en Lengua de Señas Colombiana en contextos académicos?

E. OBJETIVOS

1) Objetivo general

Implementar un aplicativo web integrando Lengua de Señas Colombiana (LSC), que apoye la comunicación y el acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG Y UNIMINUTO.

2) Objetivos específicos

- Analizar las técnicas y metodologías apropiadas para el diseño de modelos en reconocimiento de la lengua de señas colombiana (LSC).
- Desarrollar un aplicativo web para el reconocimiento de la lengua de señas colombiana (LSC), utilizando las técnicas y metodologías previamente analizadas.
- Validar el software con una prueba de concepto apoyada por un intérprete de lengua de señas.

F. JUSTIFICACIÓN

La educación superior para personas con discapacidad auditiva enfrenta desafíos importantes, especialmente por las barreras comunicativas que se presentan en los entornos académicos y por la escasez de herramientas tecnológicas que faciliten la comunicación mediante Lengua de Señas Colombiana (LSC). En Colombia, las cifras disponibles sobre discapacidad evidencian la necesidad de fortalecer estrategias de inclusión social y educativa. Como antecedente estadístico, se ha reportado que el 6,3 % de la población presentaba alguna discapacidad y que, dentro de esta población, el 17,3 % correspondía a discapacidad auditiva[24]. De igual manera, el DANE dispone de información estadística y procesos de caracterización relacionados con la población con discapacidad, los cuales permiten reconocer la magnitud de esta problemática en el país[25]. En el departamento de Nariño, esta situación adquiere mayor relevancia debido a la limitada disponibilidad de recursos tecnológicos y apoyos especializados que favorezcan el acceso, la permanencia y la participación de esta población en la educación superior.

Desde el componente normativo, el proyecto también se justifica por su relación con los principios de inclusión, accesibilidad, igualdad de oportunidades y eliminación de barreras para las personas con discapacidad. En Colombia, la Ley 1346 de 2009 aprobó la Convención sobre los Derechos de las Personas con Discapacidad, cuyo propósito es promover, proteger y asegurar el goce pleno y en condiciones de igualdad de los derechos humanos y libertades fundamentales de esta población [26]. Asimismo, la Ley 982 de 2005 establece disposiciones orientadas a la equiparación de oportunidades para las personas sordas y sordociegas, incluyendo el reconocimiento de la lengua de señas y la necesidad de apoyos que favorezcan su participación en distintos contextos sociales y educativos [27]. De igual manera, la Ley 1618 de 2013 dispone medidas para garantizar el pleno ejercicio de los derechos de las personas con discapacidad, mientras que el Decreto 1421 de 2017 reglamenta la atención educativa a esta población en el marco de la educación inclusiva [28], [29]. En este sentido, el desarrollo de una herramienta tecnológica de apoyo a la comunicación se articula con estos lineamientos, al proponer una alternativa orientada a reducir barreras comunicativas y fortalecer la participación académica de estudiantes con discapacidad auditiva.

Por otro lado, desde el ámbito económico y operativo, este proyecto plantea una alternativa viable para complementar los apoyos comunicativos existentes y reducir, en ciertas situaciones, la dependencia exclusiva de intérpretes. No se pretende reemplazar la presencia ni la función del intérprete en Lengua de Señas Colombiana, sino complementar su labor y ampliar las posibilidades

de comunicación para estudiantes que no cuentan con acompañamiento constante. De esta manera, la herramienta propuesta puede contribuir a fortalecer la autonomía de los estudiantes con discapacidad auditiva, al mismo tiempo que apoya a las instituciones en la optimización de recursos y en la ampliación de estrategias inclusivas.

Desde el ámbito ingenieril, el proyecto aporta al estudio y aplicación de áreas como el reconocimiento gestual, la inteligencia artificial y la accesibilidad digital, promoviendo la investigación de soluciones innovadoras orientadas a la inclusión de personas con discapacidad auditiva[30]. En este sentido, el aplicativo web se fundamenta en técnicas de inteligencia artificial aplicadas al reconocimiento e interpretación de la Lengua de Señas Colombiana [31]. En relación con herramientas existentes, como la desarrollada por la Universidad Nacional de Colombia, orientada al aprendizaje y reconocimiento de señas, esta propuesta busca aportar desde el reconocimiento de señas realizadas por la persona no oyente, con el fin de favorecer la comunicación con personas oyentes en contextos académicos [32]. También existen antecedentes de proyectos tecnológicos orientados a la inclusión social y educativa de personas con discapacidad auditiva, lo que demuestra que el uso de herramientas digitales puede convertirse en un apoyo relevante para los procesos de comunicación, aprendizaje y participación[33].

El uso de tecnologías accesibles representa un apoyo significativo para los procesos educativos, ya que abre nuevas oportunidades para la comunicación, la participación y el aprendizaje de personas con discapacidad auditiva. En este sentido, la implementación del aplicativo no solo contribuye a mejorar la interacción entre estudiantes con discapacidad auditiva, docentes y compañeros, sino que también impulsa la investigación y el avance de tecnologías de accesibilidad orientadas a fortalecer una sociedad más inclusiva y equitativa [2]. De igual forma, experiencias recientes en instituciones de educación superior muestran que la adopción de tecnologías de apoyo puede favorecer una participación más inclusiva de estudiantes con limitaciones auditivas en actividades formativas.

Al ser una herramienta accesible tanto para estudiantes como para docentes, el aplicativo facilita la comunicación y el acceso a la información en entornos sociales y educativos. Su implementación en la Universidad CESMAG y UNIMINUTO permite evaluar el comportamiento del sistema mediante pruebas con usuarios, observando su aporte en la comunicación, la autonomía y la interacción académica de personas con discapacidad auditiva. De esta manera, el proyecto no solo

responde a una necesidad tecnológica, sino también a una necesidad institucional relacionada con el fortalecimiento de entornos educativos más accesibles, participativos e inclusivos.

En cuanto al beneficio aportado, el desarrollo de esta herramienta fomenta la accesibilidad, la inclusión y la autonomía de las personas con discapacidad auditiva, a la vez que contribuye a la digitalización de las instituciones de educación superior mediante la propuesta de entornos más accesibles y equitativos [24]. Se espera que el uso de este tipo de aplicativos inspire a otras instituciones a adoptar soluciones tecnológicas orientadas a la inclusión, fortaleciendo el compromiso universitario con la diversidad, la accesibilidad y la participación de poblaciones con discapacidad [25].

G. DELIMITACIÓN

El proyecto se delimitó en cinco fases: análisis, planificación, diseño, pruebas y validación de un aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana (LSC), con fines educativos. La etapa de desarrollo se llevó a cabo en la Universidad CESMAG, ubicada en la ciudad de Pasto, Nariño.

La investigación se limitó al reconocimiento de un conjunto definido de palabras, sin abordar la interpretación completa de la Lengua de Señas Colombiana. En el desarrollo y validación del proyecto participaron estudiantes, intérpretes de LSC y miembros de la comunidad sorda, así como funcionarios vinculados al área de inclusión educativa de la universidad CESMAG. Asimismo, las pruebas piloto contaron con el apoyo de intérpretes de Lengua de Señas Colombiana, quienes acompañaron el proceso de validación funcional del aplicativo.

II. MARCO TEÓRICO

A. ANTECEDENTES

1. Antecedentes Internacionales

En el ámbito internacional, el desarrollo de soluciones tecnológicas para el reconocimiento de lengua de señas ha avanzado significativamente. Uno de los trabajos más pioneros fue realizado por Váradi et al. (2021) con el sistema SignAll, en Estados Unidos. Esta solución integral combina visión artificial, modelado 3D y procesamiento de lenguaje natural para traducir en tiempo real las señas de la Lengua de Señas Americana (ASL) a texto o voz. Aunque este estudio es previo a los cinco años establecidos, se destaca por su enfoque innovador en la combinación de tecnologías para la traducción de señas, lo que puede inspirar al proyecto en cuanto a la integración de diferentes herramientas tecnológicas en un sistema robusto[34]. Este avance aporta directamente al proyectar un sistema que combine múltiples tecnologías (visión artificial, redes neuronales y procesamiento) para el reconocimiento de la Lengua de Señas Colombiana, adaptando y mejorando estos principios al contexto necesario.

El sistema presentado por Álvarez (2023) para la Lengua de Señas Mexicana (LSM) utiliza una cámara RGB-D y aprendizaje automático para la detección precisa de puntos característicos del cuerpo, especialmente manos, rostro y cuerpo, los cuales son fundamentales en la articulación de las señas. El sistema ha sido probado con 3000 secuencias de datos y 30 señas diferentes. Este modelo y su metodología de captura de movimiento pueden aportar al proyecto al ofrecer una base para trabajar con diferentes tipos de señas, incluyendo las dinámicas, que son fundamentales para ampliar la cantidad de signos reconocidos en el sistema [35]. Además, la implementación de cámaras RGB-D en el proyecto puede ser útil en el contexto colombiano, donde las condiciones de captura pueden ser variadas.

Por último, Gaure et al. (2025) desarrollaron un traductor basado en redes neuronales para la interpretación en tiempo real de diferentes lenguas de señas. Este proyecto se centra en la traducción simultánea de señas a texto, lo cual puede ser muy útil para el desarrollo de un sistema multilingüe que reconozca la Lengua de Señas Colombiana (LSBC). La importancia de este trabajo radica en su potencial para adaptarse a varios tipos de lenguas de señas, lo cual podría ser relevante para futuras aplicaciones que incluyan no solo la (LSC), sino otras lenguas de señas regionales

[36]. Este enfoque ayudará a pensar en la escalabilidad del aplicativo web, permitiendo futuras expansiones.

2. Antecedentes Nacionales

En Colombia, el uso de tecnologías para el reconocimiento de la Lengua de Señas Colombiana (LSC) también ha sido explorado de manera significativa. Gutiérrez Leguizamón et al. (2022) desarrollaron un sistema de reconocimiento basado en redes neuronales convolucionales (CNN) y captura de movimiento para identificar las señas de la (LSC). Este enfoque permite un análisis detallado de los gestos y ha demostrado ser eficaz en entornos reales, lo cual es importante para el proyecto, ya que se necesita un sistema que funcione en situaciones cotidianas y no solo en entornos controlados [37]. Este estudio aportará a la comprensión sobre cómo integrar la captura de movimiento y redes neuronales para lograr un sistema de reconocimiento preciso y adaptado a diferentes contextos sociales.

A su vez, Suat Rojas et al. (2021) diseñaron un sistema que reconoce el abecedario de la (LSC) mediante redes neuronales convolucionales, alcanzando una precisión del 79.2%. Este trabajo ha permitido avanzar en la automatización del reconocimiento de señas estáticas, lo cual es un aporte directo al proyecto, especialmente si se considera que el abecedario es una de las bases para la enseñanza y el aprendizaje de la (LSC) [38]. Este estudio también destaca por su uso de técnicas avanzadas de procesamiento de imágenes, algo que se puede aplicar para mejorar la precisión en la identificación de las señas.

Finalmente, Quintero (2024) implementó una herramienta de aprendizaje automático con redes neuronales, entrenando modelos con 1500 videos en (LSC). Este corpus de datos permite una mayor precisión en el reconocimiento y es un recurso valioso para mejorar el aprendizaje de la (LSC) en entornos digitales [39]. En el contexto del proyecto, el uso de bases de datos similares será fundamental para entrenar el aplicativo web y mejorar su capacidad de reconocer una mayor cantidad de señas de manera eficiente y precisa.

3. Antecedentes Regionales (Nariño)

A nivel regional, en Nariño, el proyecto “Thomasito”, desarrollado por Enriquez Palacios y Ramírez Gracia (2023), utiliza ajustes razonables para facilitar la enseñanza de matemáticas

mediante la Lengua de Señas Colombiana. Esta herramienta digital se enfoca en la enseñanza de operaciones matemáticas básicas como la suma y resta, lo cual tiene implicaciones directas para la integración de la (LSC) en la educación cotidiana [40]. Para el proyecto, este enfoque educativo ayuda a pensar no solo en el reconocimiento, sino también en cómo la tecnología puede facilitar el aprendizaje de la (LSC) en áreas específicas.

Por otra parte, Bravo y Narváez (2022) diseñaron una plataforma de aprendizaje accesible para estudiantes sordos en programas técnicos. Esta plataforma utiliza videolecciones y recursos visuales adaptados a las necesidades de la comunidad sorda. Este trabajo es relevante porque subraya la importancia de crear plataformas que no solo reconozcan las señas, sino que también promuevan su enseñanza de forma interactiva y accesible. Esto es clave para el proyecto, ya que implica un enfoque inclusivo que no solo se limite al reconocimiento de signos, sino también a su enseñanza efectiva [41]. Este antecedente muestra cómo integrar la tecnología en la educación para personas sordas o con cierto grado de discapacidad auditiva en un contexto local.

Además, la Gobernación de Nariño (2022) ha promovido proyectos de inclusión para la comunidad sorda a través de la implementación de sistemas de comunicación accesibles. Estos esfuerzos a nivel departamental son clave para crear una infraestructura inclusiva que facilite la integración social de las personas sordas [42]. Para este proyecto, este tipo de iniciativas gubernamentales destaca la importancia de la inclusión social y la implementación de políticas públicas que respalden la tecnología para personas con discapacidad auditiva.

B. SUPUESTOS TEÓRICOS

- **El lenguaje como base de la comunicación humana**

El lenguaje es la capacidad que permite realizar la expresión de ideas, emociones, pensamientos y conocimientos mediante sistemas de signos organizados en forma orales, escritas y gestuales, aparte de ser un medio de comunicación, el lenguaje constituye la base sobre la cual los seres humanos construyen cultura, identidad, saberes y estructuras sociales. A través de él, se establece la interacción, la interpretación del entorno y la transmisión de valores, convirtiéndose en un factor esencial del desarrollo humano [43].

En el contexto del presente proyecto, el lenguaje cobra una importancia particular, al considerar que la lengua de señas, como manifestación gestual del lenguaje, posibilita a las personas con discapacidad auditiva participar plenamente en entornos educativos. Comprender la naturaleza del lenguaje es, por tanto, fundamental para diseñar soluciones tecnológicas que reconozcan y traduzcan los signos, promoviendo una comunicación inclusiva y efectiva.

- **Qué es la lengua**

La lengua como complemento del lenguaje es un sistema de signos con reglas gramaticales que a través de la historia ha jugado un papel fundamental para transmitir conocimientos, valores y tradiciones que perduran a lo largo del tiempo y preservar la cultura [44].

El presente proyecto reconoce la lengua de señas colombiana (LSC) como un sistema auténtico, cuya preservación es fundamental para garantizar los derechos de comunicación de la comunidad con discapacidad auditiva.

- **Comunicación**

Según la Central Washington University la comunicación es una disciplina de las ciencias sociales que estudia cómo, por qué y con qué efectos las personas utilizan el lenguaje y los medios de comunicación para transmitir información. Analiza el envío, la recepción y la interpretación de mensajes en diferentes contextos y a través de diversos canales, tanto tradicionales como digitales [45].

En el entorno educativo, garantizar una comunicación una para personas con discapacidad auditiva implica desarrollar herramientas inclusivas. El proyecto se sustenta en la premisa de que una comunicación fluida, mediada por tecnologías de reconocimiento de señas, puede facilitar la participación activa de los estudiantes sordos en entornos de aprendizaje, reduciendo barreras lingüísticas y sociales.

- **Información**

Según el Tecnológico Nacional de México [46], la información representa el contenido significativo transmitido en un proceso comunicativo. Constituye datos procesados que adquieren valor para quien los recibe, permitiendo la toma de decisiones, la generación de conocimiento y la

acción informada. En la sociedad contemporánea, caracterizada por el acceso masivo a datos, la correcta gestión de la información es esencial para la innovación y el progreso. Este proyecto aborda la información desde una perspectiva inclusiva, en la que los signos captados a través de lenguas de señas deben ser interpretados, procesados y transformados en datos comprensibles para todos. De este modo, la información gestual se convierte en un puente de comunicación entre comunidades sordas y oyentes, facilitando el aprendizaje colaborativo y la integración educativa.

- **Discapacidad**

Discapacidad se entiende como el resultado de la interacción entre personas con deficiencias físicas, mentales, intelectuales o sensoriales y las barreras que impone el entorno, las cuales limitan su participación plena y efectiva en la sociedad en igualdad de condiciones con los demás. Esta perspectiva, establecida por la Convención de las Naciones Unidas, reconoce que la discapacidad no reside únicamente en las condiciones individuales, sino en las actitudes sociales y estructuras físicas que impiden el ejercicio de los derechos. Comprender la discapacidad desde este enfoque es fundamental para promover la inclusión, la equidad y la justicia social [47].

Dentro del proyecto, asumir esta definición de discapacidad permite orientar el desarrollo de soluciones tecnológicas que eliminen barreras, promuevan la participación y garanticen el acceso igualitario a la información, la educación y la comunicación. Al integrar este marco de derechos humanos, se fortalece la autonomía de las personas con discapacidad y se avanza hacia una sociedad más accesible e incluyente.

- **Discapacidad auditiva**

La discapacidad auditiva implica la pérdida parcial o total de la capacidad para percibir sonidos, afectando de manera significativa la comunicación oral y el acceso a la información. Esta condición puede ser congénita o adquirida, y su impacto varía según la edad de aparición, el grado de pérdida y las oportunidades de acceso a sistemas de apoyo comunicativo. Reconocer la discapacidad auditiva no solo desde una perspectiva médica, sino también desde un enfoque social y de derechos humanos, es clave para construir entornos inclusivos [48].

Dentro del proyecto, comprender la discapacidad auditiva resulta esencial para diseñar soluciones tecnológicas adaptadas a las necesidades reales de los usuarios sordos. Facilitar el acceso a la lengua de señas mediante herramientas de reconocimiento automático contribuye no solo a la accesibilidad educativa, sino también al fortalecimiento de la autonomía, la participación y el ejercicio pleno de los derechos comunicativos de esta población.

- **Lengua de señas**

Según el Banco Interamericano de Desarrollo, la lengua de señas no solo constituye un medio de comunicación para las personas sordas, sino que también impacta profundamente en su desarrollo personal y social. La implementación de tecnologías que permitan interpretar la lengua de señas surge como una solución ante la limitada disponibilidad de intérpretes certificados, promoviendo la accesibilidad y eliminando barreras de comunicación. Estas tecnologías facilitan la comunicación en tiempo real entre personas sordas y oyentes, fortaleciendo la inclusión social a través de sistemas innovadores de video-interpretación y procesamiento de señas [49]. El proyecto propone la implementación de un sistema basado en inteligencia artificial para el reconocimiento automático de señas, lo que permitirá convertir los gestos en texto o voz. Esta funcionalidad resulta crucial para la inclusión educativa, ya que facilita la interacción entre estudiantes sordos y oyentes, promoviendo entornos de aprendizaje colaborativos y accesibles.

- **Solución tecnológica**

Según la Agencia de Servicios Públicos Digitales de Andalucía, una solución tecnológica se entiende como la respuesta corporativa estándar a una necesidad específica, ya sea de carácter global, como el desarrollo de un sitio web, o concreta, como la integración con una plataforma particular para atender una necesidad determinada [50]. Estas soluciones pueden ser autosuficientes o requerir la combinación de varias herramientas para alcanzar su objetivo.

En el presente proyecto, la solución tecnológica se centra en el diseño e implementación de un sistema de reconocimiento de lengua de señas destinado a entornos educativos. Esta iniciativa pretende responder a la necesidad de comunicación inclusiva de los estudiantes sordos, permitiéndoles una participación activa en las actividades académicas y fortaleciendo los procesos de inclusión dentro de las instituciones universitarias.

- **Aplicativo web**

Un aplicativo web es un programa que se ejecuta en un servidor externo y al que se accede a través de un navegador web, sin necesidad de instalarlo en el equipo local. Esto facilita su uso principalmente en computadoras y mejora la accesibilidad para los usuarios. Estas aplicaciones manejan su gestión y actualizaciones de manera centralizada en el servidor, lo que disminuye los costos y las complicaciones para quienes las utilizan. Son comunes en áreas como la educación, el comercio electrónico y las redes sociales, debido a su capacidad para adaptarse a diversas necesidades y contextos, además de fomentar la inclusión digital [51].

En el presente proyecto, se propone el desarrollar un aplicativo web que incorpore inteligencia artificial para el reconocimiento básico de la Lengua de Señas Colombiana (LSC), Esta herramienta busca ser accesible para los usuarios, permitiendo su uso tanto en aulas presenciales como en entornos remotos. La elección de una solución web facilita la escalabilidad, el mantenimiento y la integración con los recursos institucionales, contribuyendo a promover la equidad educativa y a disminuir las barreras tecnológicas que enfrentan los estudiantes con discapacidad auditiva en la educación superior.

- **Inteligencia artificial (IA)**

Según el Parlamento Europeo, la inteligencia artificial se define como la capacidad de una máquina para exhibir habilidades similares a las humanas, como el razonamiento, el aprendizaje, la creatividad y la planificación [52]. Los sistemas de IA perciben su entorno, interactúan con él, procesan información y toman decisiones orientadas a objetivos específicos. Además, son capaces de adaptar su comportamiento analizando los resultados de acciones anteriores, lo que les permite operar de manera autónoma en diversos contextos.

Aplicada al proyecto, la inteligencia artificial constituye el núcleo del sistema de reconocimiento de señas. Gracias a su capacidad para identificar patrones visuales complejos característicos de la lengua de señas, se posibilita una traducción automatizada que elimina barreras de comunicación y promueve una inclusión educativa más efectiva para los estudiantes sordos.

- **Redes neuronales**

Las redes neuronales artificiales son modelos computacionales inspirados en la estructura y funcionamiento del cerebro humano. Están diseñadas para reconocer patrones y aprender a partir de datos, imitando los procesos de aprendizaje y toma de decisiones de los seres humanos [53]. Estas redes se componen de capas de nodos o "neuronas", organizadas en una capa de entrada, una o más capas ocultas y una capa de salida, a través de las cuales la información es procesada y transmitida hasta generar un resultado final, como una clasificación o una predicción.

Dentro del proyecto, las redes neuronales artificiales desempeñan un rol fundamental al ser la base sobre la cual se construye el sistema de reconocimiento de señas. Mediante el entrenamiento con grandes volúmenes de imágenes y gestos, las redes pueden aprender a identificar las diferentes señas, optimizando su precisión y su capacidad de adaptación a variaciones individuales en la forma de signar.

- **Redes neuronales convolucionales (CNN)**

Las redes neuronales convolucionales (CNN) son un tipo especializado de redes diseñadas para procesar datos con una estructura de cuadrícula, como las imágenes. Su arquitectura permite extraer automáticamente características espaciales relevantes a través de filtros de convolución, lo que las hace especialmente útiles en el reconocimiento visual [53].

En este proyecto, las CNN son esenciales para interpretar correctamente los movimientos, configuraciones de manos y expresiones faciales características de la lengua de señas. Su capacidad de generalizar a partir de ejemplos facilita que el sistema identifique gestos en diferentes condiciones de iluminación, ángulos o variaciones individuales, incrementando así su robustez y eficacia en entornos educativos reales.

- **Modelos de reconocimiento de imágenes**

El reconocimiento de imágenes es una aplicación de machine learning que permite a los dispositivos y al software identificar objetos, lugares, personas, textos y acciones dentro de imágenes o videos digitales. Esta tecnología resulta fundamental en múltiples sectores, desde la detección de defectos en procesos industriales hasta la identificación de anomalías médicas y el desarrollo de vehículos autónomos [54].

En el contexto del proyecto, los modelos de reconocimiento de imágenes representan la base tecnológica para interpretar los gestos propios de la lengua de señas. A través del entrenamiento de redes neuronales con grandes volúmenes de datos visuales, se logra que el sistema identifique y traduzca de forma precisa las diferentes señas realizadas por los usuarios, promoviendo así una comunicación más accesible y efectiva en entornos educativos.

C. VARIABLES DEL ESTUDIO

1. Variables dependientes

- Reconocimiento de lengua de señas colombiana (LSC)
- Usabilidad del aplicativo
- Fiabilidad del aplicativo

2. Variables independientes

- Aplicación web

A. DEFINICIÓN NOMINAL DE LAS VARIABLES

1. Variables Dependientes (VD)

- **Reconocimiento de lengua de señas colombiana (LSC)**

Es la capacidad del aplicativo web para poder reconocer y procesar los gestos y expresiones de la lengua de señas colombiana que utilizan las personas sordas o con cierto grado de discapacidad auditiva, dando como resultado la interpretación de la palabra que quiere trasmitir la persona no oyente. El reconocimiento se da a través de tecnologías como IA, redes neuronales, que permite comparar y procesamiento de dichas señas[55].

- **Usabilidad**

Se define como la capacidad del aplicativo web para ser comprendido, utilizado de manera eficiente y satisfactoria por los usuarios. Esta variable contempla aspectos como la facilidad de uso, la claridad de la interfaz, la accesibilidad y la satisfacción con la herramienta tecnológica[56].

- **Fiabilidad**

Se refiere a la capacidad del aplicativo web para tener un rendimiento estable cuando se utiliza en situaciones de estrés operativo durante periodos de tiempo determinados. Esta variable incluye

subcaracterísticas como la disponibilidad, la tolerancia de los fallos y la recuperación ante errores que puedan presentarse [56].

2. Variable Independiente (VI)

- **Aplicación web**

Se define como la herramienta tecnológica desarrollada para capturar o recibir videos de señas, procesarlos mediante un modelo de reconocimiento e interpretar palabras en Lengua de Señas Colombiana dentro del conjunto contemplado por el sistema. Su propósito es apoyar la comunicación básica de estudiantes con discapacidad auditiva en el entorno académico. Es un hecho que el uso de tecnologías expande muchas opciones para acceder a información además de facilitar tareas cotidianas y educativas [57].

B. DEFINICIÓN OPERATIVA DE LAS VARIABLES

Las variables operativas se midieron según la naturaleza de cada indicador. Para aspectos de percepción se utilizó una escala tipo Likert de cinco niveles: 1 = Muy bajo, 2 = Bajo, 3 = Medio, 4 = Alto y 5 = Muy alto [58]. Para identificar fallos del software se empleó una escala nominal con opciones “Sí” y “No”, complementada con preguntas abiertas en caso de presentarse errores o anomalías [59].

1. Variable Dependiente (VD)

- **Reconocimiento de lengua de señas colombiana (LSC)**

Esta variable se refiere a la capacidad del aplicativo web para reconocer e interpretar correctamente las señas ejecutadas por el usuario, dentro del conjunto de palabras contemplado en la versión implementada del sistema. Para su evaluación, se tuvieron en cuenta los resultados generados durante las pruebas funcionales, considerando la correspondencia entre la seña realizada y la interpretación emitida por el aplicativo, así como el tiempo de respuesta requerido para procesar el video y devolver la salida correspondiente.

- **Usabilidad**

Esta variable se refiere al grado en que los usuarios pueden comprender, aprender y utilizar el aplicativo web de manera fácil, clara y satisfactoria. Para su medición, se tuvieron en cuenta

aspectos como la facilidad de uso, la comprensión del funcionamiento general de la herramienta, la claridad de la interfaz, la accesibilidad percibida y la satisfacción del usuario durante la interacción con el sistema [56].

- **Fiabilidad**

Esta variable se refiere a la capacidad del aplicativo web para mantener un funcionamiento estable durante su uso, respondiendo de manera consistente ante las solicitudes realizadas por el usuario. Su evaluación considera aspectos como la estabilidad del sistema durante las pruebas, la continuidad del funcionamiento sin fallos evidentes y la capacidad de ofrecer una respuesta operativa dentro de las condiciones de uso del aplicativo [56].

2. Variables Independiente (VI)

- **Aplicación web**

Esta variable se ejecutó mediante el uso del aplicativo web por parte de los participantes durante las pruebas funcionales, a través de sus dos modalidades de entrada: carga de video previamente grabado y grabación en tiempo real. Su presencia en el estudio se evidenció en la interacción directa de los usuarios con la interfaz, el envío del video al sistema y la recepción de la interpretación generada por el aplicativo.

D. FORMULACIÓN DE HIPÓTESIS

1. Hipótesis de investigación

El diseño de un aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana contribuyó al apoyo de la comunicación básica y al acceso a la información de estudiantes con discapacidad auditiva en las universidades CESMAG y UNIMINUTO, dentro del entorno académico.

2. Hipótesis nula

El diseño de un aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana no contribuyó al apoyo de la comunicación básica y al acceso a la información de estudiantes con

discapacidad auditiva en las universidades CESMAG y UNIMINUTO, dentro del entorno académico.

3. Hipótesis alterna

El diseño de un aplicativo web basado en reconocimiento de palabras en Lengua de Señas Colombiana contribuyó principalmente al apoyo de la comunicación básica de estudiantes con discapacidad auditiva en el entorno académico, más que al acceso amplio a contenidos o a procesos complejos de interpretación.

III. METODOLOGÍA

A. PARADIGMA

La presente investigación se desarrolló bajo el paradigma positivista, debido a que partió de una problemática concreta e identificable en un contexto real: la necesidad de apoyar la comunicación básica y el acceso a la información de estudiantes con discapacidad auditiva en el entorno académico de las universidades CESMAG y UNIMINUTO. Este paradigma se orienta a la obtención de conocimiento objetivo, verificable y sustentado en la observación, el análisis y la validación de hechos observables [60].

En coherencia con este enfoque, el proyecto se orientó al diseño de una solución tecnológica basada en inteligencia artificial, incorporando redes neuronales recurrentes (RNN) para el reconocimiento básico de palabras en Lengua de Señas Colombiana (LSC). Desde esta perspectiva, la problemática fue abordada mediante procedimientos técnicos y metodológicos que permitieron analizar secuencias gestuales, estructurar un modelo de reconocimiento, integrarlo dentro de un aplicativo web y validar su funcionamiento en condiciones de uso definidas. De esta manera, la investigación buscó generar resultados medibles, observables y sustentados en evidencia, en correspondencia con los principios del paradigma positivista.

B. ENFOQUE

El enfoque de la investigación que orientó este proyecto es mixto con predominio cualitativo, ya que se buscó comprender la problemática y las necesidades de la comunidad sorda en relación con la interpretación de la Lengua de Señas Colombiana (LSC). Según Denzin y Lincoln [61], este enfoque permite analizar las experiencias y significados que las personas atribuyen a una situación. De manera complementaria, se emplearon encuestas de carácter exploratorio para obtener una visión general del contexto, sin implicar un análisis cuantitativo profundo, como lo señalan Hernández, Fernández y Baptista [62]. Posteriormente, se desarrolló una prueba de concepto con el fin de validar la viabilidad técnica de la solución propuesta.

C. MÉTODO

Se adoptó un enfoque no experimental de tipo descriptivo, orientado a la validación del aplicativo web. En este estudio no se realizaron mediciones antes y después ni asignación aleatoria de los participantes, sino que se evaluó el sistema en un único momento tras su uso, mediante encuestas aplicadas a los usuarios. Este tipo de diseño es adecuado en contextos educativos y sociales, donde se busca evaluar herramientas en condiciones reales sin limitar el acceso a los participantes [63]. No obstante, al no existir manipulación de variables ni comparación entre grupos, no es posible establecer relaciones causales, y los resultados deben interpretarse con cautela debido a posibles sesgos que afectan la validez interna [64].

D. TIPO DE INVESTIGACIÓN

La presente investigación es de tipo cualitativa, debido a que permitió comprender y describir la problemática de comunicación básica y acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG y UNIMINUTO. Asimismo, se analizaron las percepciones de los usuarios y la validación funcional del aplicativo apoyada por intérprete de Lengua de Señas Colombiana, de acuerdo con los planteamientos metodológicos de Quijano Vodniza sobre investigación cualitativa[65].

E. DISEÑO DE LA INVESTIGACIÓN

El diseño de la investigación fue descriptivo, debido a que permitió caracterizar el funcionamiento del aplicativo web en un contexto académico real, sin manipular las condiciones del entorno ni establecer relaciones causales entre variables [66]. Este diseño se ajustó al propósito del estudio, ya que permitió observar, describir y valorar el comportamiento del sistema durante las pruebas funcionales y la validación con usuarios e intérprete de Lengua de Señas Colombiana.

Desde este diseño, se analizaron aspectos relacionados con el reconocimiento de palabras en Lengua de Señas Colombiana, la correspondencia entre la seña realizada y la interpretación generada por el aplicativo, el tiempo de respuesta, la facilidad de uso, la presencia de fallos y la percepción de los usuarios frente a la herramienta, elementos propios de las pruebas de usabilidad y evaluación de sistemas interactivos [67]. De esta manera, los datos recolectados permitieron describir el desempeño funcional del sistema y su aporte como herramienta de apoyo a la

comunicación básica y al acceso a la información de estudiantes con discapacidad auditiva en UNICESMAG y UNIMINUTO.

F. POBLACIÓN

La población del estudio estuvo conformada por estudiantes sordos o con discapacidad auditiva pertenecientes al contexto universitario del departamento de Nariño, Colombia. Esta delimitación respondió al enfoque de la investigación, centrado en el ámbito de la educación superior, así como a la viabilidad de acceso a los participantes dentro de las instituciones contempladas. De manera complementaria, durante el proceso de validación se contó con el acompañamiento de intérpretes de Lengua de Señas Colombiana (LSC), quienes facilitaron la comunicación y apoyaron el desarrollo de las pruebas.

G. MUESTRA

La muestra de la investigación se seleccionó mediante un muestreo no probabilístico por conveniencia y estuvo conformada por dos estudiantes sordos o con discapacidad auditiva pertenecientes a la Universidad CESMAG. Inicialmente se contempló la participación de estudiantes de UNIMINUTO; sin embargo, estos no fueron incluidos en la muestra final debido a su egreso de la institución al momento de la ejecución de las pruebas.

H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Al principio del proyecto no se contaba con una base de datos que proporcione toda la información estandarizada sobre la lengua de señas colombiana (LSC), dada la situación se recurrió a técnicas mixtas que ayudaron a recolectar la información necesaria, como:

Entrevistas Personales

Esta parte fue fundamental, tener entrevistas directamente con la población sorda o con cierto grado de discapacidad auditiva reforzó la investigación, ya que al observar cómo es la comunicación cotidiana se pudo empezar a detectar cómo son sus gestos o cuáles son sus señas más comunes.

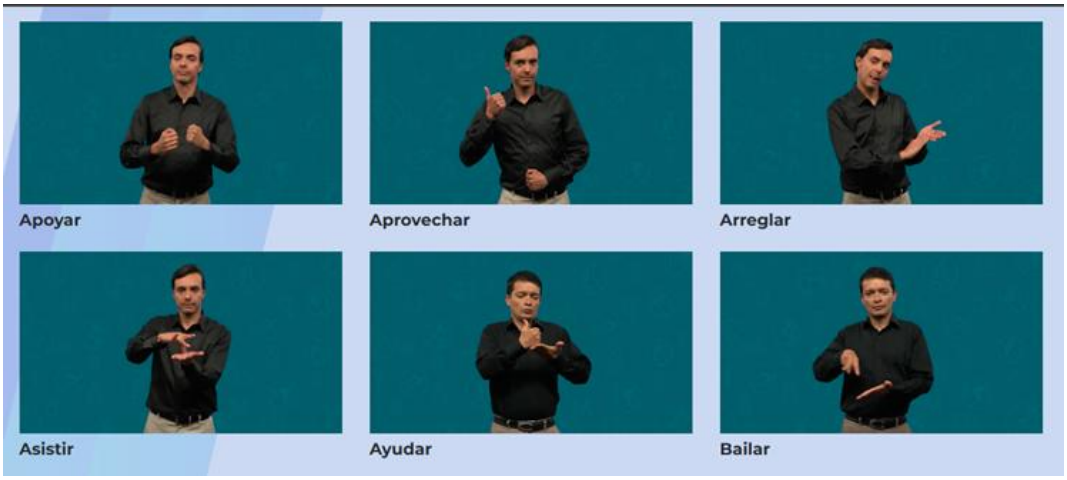


Fig. 1 Lengua de Señas Colombiana (LSC)

Fuente: adaptado de [68]

Prueba funcional no participante

Esta técnica se empleó con el fin de evaluar el comportamiento del sistema en condiciones de uso definidas, mediante las dos modalidades de entrada disponibles en la herramienta: la grabación en tiempo real y la carga de video previamente almacenado. A través de estas pruebas, los participantes ejecutaron las señas frente al sistema y se observó la respuesta emitida por el aplicativo en relación con la interpretación generada. Esta técnica permitió recopilar información sobre el funcionamiento general de la herramienta, el flujo de uso y la correspondencia entre la seña realizada y la salida producida por el sistema.

Encuesta

La encuesta se utilizó como técnica para recolectar información sobre la percepción de los participantes frente al aplicativo web. Su aplicación permitió obtener información relacionada con la usabilidad de la herramienta, la claridad de la interfaz, la facilidad de uso y la apreciación de los usuarios sobre la correcta interpretación de las señas realizadas. Esta técnica se aplicó mediante un formulario estructurado en Google Forms, lo que facilitó la organización y comparación de las respuestas obtenidas. El hacer encuestas identifica las opiniones de los usuarios y detecta patrones fundamentales[69].

Validación por juicio experto

Esta técnica se desarrolló mediante una prueba de concepto apoyada por una intérprete de Lengua

de Señas Colombiana (LSC), quien actuó como referente experto durante la validación del software. Su participación permitió contrastar la interpretación generada por el aplicativo con el criterio profesional de una persona con conocimiento especializado en el área. Esta técnica aportó un sustento cualitativo a la validación del sistema, al permitir examinar la correspondencia entre las señas ejecutadas y la interpretación emitida por la herramienta.

I. VALIDEZ DE LAS TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN

La validez de las técnicas aplicadas en la investigación se orientó a priorizar que los aspectos centrales del estudio fueran evaluados de manera pertinente y coherente con los objetivos planteados. En primer lugar, la validez de contenido se respaldó mediante la participación de una intérprete de Lengua de Señas Colombiana (LSC), quien actuó como referente experto para revisar la pertinencia, claridad y representatividad de las señas, las grabaciones y los materiales empleados durante el proceso de desarrollo y validación.

En segundo lugar, la validez de criterio se estableció a partir del contraste entre la interpretación generada por el aplicativo y la valoración realizada por la intérprete de LSC. Este contraste permitió examinar el comportamiento del sistema frente a un referente externo confiable, sustentado en el criterio profesional de una persona con conocimiento especializado en el área.

Finalmente, la validez de constructo se abordó a partir de variables como la usabilidad del aplicativo y la correcta interpretación de las señas por parte del sistema. Estas dimensiones fueron consideradas mediante pruebas funcionales y a través de la información recolectada con usuarios reales, lo que permitió valorar la correspondencia entre la herramienta desarrollada y las necesidades de apoyo a la comunicación básica en la población objetivo.

J. CONFIABILIDAD DE LAS TÉCNICAS DE RECOLECCIÓN

La confiabilidad de las técnicas aplicadas en la investigación se sustentó en la aplicación uniforme de las pruebas, procurando mantener procedimientos y condiciones similares con los participantes durante las sesiones realizadas. Esto permitió contar con un criterio de comparación más consistente entre los resultados y reducir la influencia de factores externos que pudieran afectar el desarrollo de las pruebas[70].

De manera complementaria, la recolección de información se apoyó en una encuesta estructurada aplicada mediante Google Forms, orientada a recoger la percepción de los participantes sobre aspectos relacionados con la usabilidad del aplicativo y la correcta interpretación de las señas. Para ello, se emplearon preguntas organizadas de forma homogénea, lo que favoreció la obtención de respuestas comparables entre los distintos participantes. Asimismo, se utilizó una escala tipo Likert, ampliamente empleada en investigaciones educativas y sociales para recoger opiniones y percepciones de forma estructurada [58].

Adicionalmente, se realizó una prueba piloto previa a la aplicación principal de los instrumentos, con el fin de identificar posibles ajustes en la redacción de las preguntas, en la dinámica de aplicación y en las condiciones generales de las pruebas. Esta etapa permitió mejorar la claridad del instrumento y favorecer condiciones más estables y equitativas durante la recolección de la información [70].

K. INSTRUMENTO DE RECOLECCIÓN DE DATOS

Para la presente investigación, se aplicaron los siguientes instrumentos para la recolección de información:

1. **Guía de entrevista:** se utilizó para orientar el diálogo con los participantes y recoger información relacionada con sus formas de comunicación, las señas de uso frecuente y las necesidades percibidas dentro del contexto académico. Durante su aplicación se contó con el apoyo de un intérprete de Lengua de Señas Colombiana (LSC), lo que facilitó el desarrollo de la interacción y la comprensión mutua entre los participantes.

2. **Registro fotográfico:** se utilizó como apoyo para documentar la ejecución de señas, los movimientos gestuales y algunos aspectos observados durante la interacción de los participantes. Este instrumento permitió revisar posteriormente detalles asociados a la forma de ejecución y a posibles factores que incidían en el reconocimiento por parte del sistema.

3. **Grabación de video:** se empleó un dispositivo móvil para registrar las señas realizadas por los participantes. Este material fue utilizado tanto para la conformación del conjunto de datos como para las pruebas funcionales del aplicativo, permitiendo disponer de secuencias visuales necesarias para el procesamiento y evaluación del sistema.

4. **Formulario en Google Forms:** se aplicó como instrumento para recolectar la percepción de los participantes sobre la usabilidad del aplicativo, la claridad de la interfaz y la correcta interpretación de las señas por parte del sistema. Este formulario se estructuró con preguntas organizadas de manera homogénea y apoyadas en una escala tipo Likert[58].

5. **Acta de validación:** se utilizó como soporte formal durante la prueba de concepto realizada con la intérprete de Lengua de Señas Colombiana, dejando constancia de la valoración emitida sobre el funcionamiento del software y su capacidad de interpretación dentro del alcance definido en la investigación.

IV. RESULTADOS DE LA INVESTIGACIÓN

A. ANÁLISIS DE TÉCNICAS Y METODOLOGÍAS PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA

Durante el desarrollo del proyecto se analizaron distintas técnicas empleadas en el reconocimiento automático de señas. A partir de esta revisión, se identificaron como enfoques más relevantes las redes neuronales convolucionales (CNN), redes recurrentes (RNN), arquitecturas LSTM, modelos bidireccionales (BiLSTM), mecanismos de atención y arquitecturas basadas en Transformers.

En cuanto al procesamiento de la información visual, se analizaron fotogramas en formato RGB y con la extracción de características a partir de puntos clave (landmarks) mediante MediaPipe. Para la construcción del conjunto de datos, se consideraron variables relacionadas con el movimiento de las manos, la configuración de los dedos, la orientación y la secuencia temporal de los gestos presentes en los videos. Estas variables se utilizaron como entrada para el entrenamiento de los modelos.

Técnicas de modelado

En esta sección se presentan los resultados derivados del análisis de distintas arquitecturas de aprendizaje profundo aplicadas al reconocimiento de secuencias gestuales. Para ello, se consideraron diversas técnicas de modelado orientadas al procesamiento de datos provenientes de video.

1. Redes Neuronales Convolucionales (CNN)

Las CNN se consideraron como una arquitectura para el procesamiento de información visual dentro del sistema de reconocimiento de Lengua de Señas Colombiana . Este tipo de modelo está compuesto por capas convolucionales encargadas de extraer características espaciales mediante la aplicación de filtros sobre la entrada, permitiendo identificar patrones locales relevantes [71].

Estas capas se complementan con operaciones de agrupamiento (pooling), orientadas a reducir la dimensionalidad de los datos y conservar las características más representativas [71]. Finalmente,

se emplea una capa completamente conectada (fully connected), basada en perceptrones, encargada de realizar la clasificación de las señales[71].

El perceptrón, como unidad básica de la capa densa, se define matemáticamente como:

$$y = f(\sum_{i=1}^{n} w_i x_i + b) \qquad Ecu.\ 1$$

donde x_i representa las entradas provenientes de las capas anteriores, w_i los pesos asociados a cada entrada, b el sesgo del modelo y f la función de activación utilizada para generar la salida del modelo[71] . Este mecanismo corresponde a la operación básica de decisión en una red neuronal, en la que las entradas ponderadas se combinan para producir una salida de clasificación [71].

El funcionamiento de la CNN se basa en un proceso jerárquico de transformación de la información, en el cual las primeras capas identifican patrones simples como bordes o contornos, mientras que las capas más profundas construyen representaciones más complejas asociadas a estructuras visuales del gesto [71]. En la arquitectura analizada, la red procesa secuencias de video tratadas como conjuntos de fotogramas, lo que permite la extracción progresiva de características hasta obtener una representación final para la clasificación. En esta arquitectura, los perceptrones constituyen la etapa final, integrando la información extraída por las capas convolucionales para generar la predicción de la seña.

Asimismo, se identificó que este enfoque se centra en el procesamiento de información espacial a partir de los fotogramas, sin incorporar mecanismos explícitos para el modelado de la información temporal. Esta característica se consideró como referencia dentro del proceso de selección de alternativas para el desarrollo del sistema.

En la Fig. 2 se presenta la estructura general de una red CNN, en la cual se ilustran las etapas de extracción de características y clasificación.

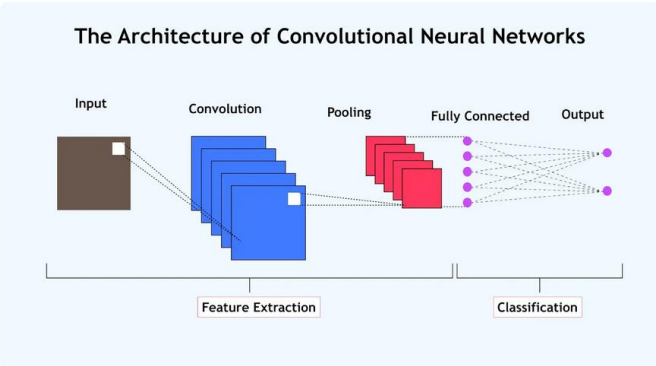


Fig. 2 Arquitectura general de una red neuronal convolucional (CNN)

Fuente: Adaptado de [72]

2. Redes Neuronales Recurrentes (RNN)

Las RNN se consideraron como una arquitectura orientada al procesamiento de información secuencial dentro del sistema de reconocimiento de Lengua de Señas Colombiana. A diferencia de modelos basados en datos estáticos, estas redes incorporan un mecanismo que permite mantener y actualizar información a lo largo de una secuencia, facilitando la representación de dependencias temporales entre los datos [4].

Desde el punto de vista estructural, las RNN están conformadas por unidades recurrentes que reciben información en cada paso de la secuencia y la combinan con un estado interno previo. Este mecanismo se expresa matemáticamente mediante la actualización del estado oculto:

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \qquad Ecu.\ 2$$

donde x_t corresponde a la entrada en el instante $t, ht - 1$ representa el estado previo, W_h y W_x son matrices de pesos, b es el sesgo y f la función de activación que permite generar el nuevo estado h_t[4].

El funcionamiento de esta arquitectura se basa en la propagación de información a través de la secuencia, permitiendo que cada estado capture tanto la entrada actual como el contexto acumulado de los estados anteriores. De esta manera, el modelo construye una representación dinámica de los datos a lo largo del tiempo[4]. En la arquitectura analizada, las RNN procesan secuencias de video interpretadas como una sucesión ordenada de estados que describen la evolución del gesto, lo que

permite modelar la continuidad del movimiento e integrar información de diferentes instantes dentro de una misma representación [4].

La Fig. 3 presenta la estructura general de una RNN, donde se ilustra el flujo de información a través del tiempo y la dependencia entre estados consecutivos. Esta arquitectura se caracteriza por la presencia de conexiones recurrentes que permiten que la información previa influya en el procesamiento actual, manteniendo la continuidad de la secuencia.

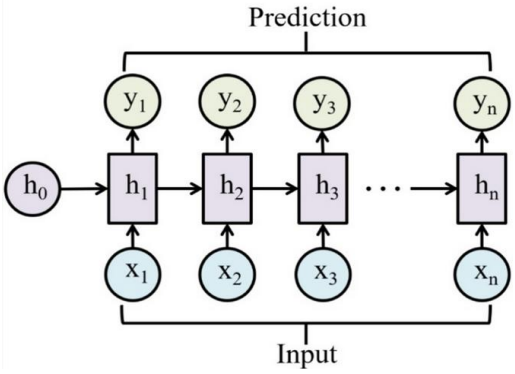


Fig. 3 Estructura general de una red neuronal recurrente (RNN)

Fuente: Adaptado de [4]

3. Redes LSTM (Long Short-Term Memory)

Las LSTM se consideraron como una arquitectura orientada al modelado de información secuencial con capacidad de memoria a largo plazo dentro del sistema de reconocimiento de Lengua de Señas Colombiana. Estas redes corresponden a una variante de las arquitecturas recurrentes que incorporan celdas de memoria y compuertas de control para regular el flujo de información a lo largo de la secuencia[73].

Desde el punto de vista estructural, una unidad LSTM está compuesta por una celda de memoria y tres compuertas principales: entrada, olvido y salida. Estas permiten controlar la información que se incorpora, se descarta y se utiliza para generar la salida en cada instante. Este comportamiento se describe mediante:

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big) \qquad \textit{Ecu. 3}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \qquad \textit{Ecu. 4}$$

$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c) \quad \textit{Ecu. 5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \textit{Ecu. 6}$$

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0) \quad \textit{Ecu. 7}$$

$$h_t = o_t \cdot tanh(C_t) \quad \textit{Ecu. 8}$$

La Ecu. 3 representa la compuerta de olvido f_t, encargada de registrar la información del estado de celda anterior que será conservada o descartada en el instante t; en esta expresión, W_f corresponde a la matriz de pesos, b_f al sesgo, h_{t-1} al estado oculto anterior, x_t a la entrada actual y σ a la función sigmoide. La Ecu. 4 define la compuerta de entrada i_t, la cual registra la información nueva que será incorporada al estado de celda; en este caso, W_i representa la matriz de pesos y b_i el sesgo asociado. La Ecu. 5 presenta el estado candidato de la celda $\tilde{C}_t$, calculado mediante la función tangente hiperbólica $tanh$, donde W_c y b_c corresponden a los pesos y sesgos de esta operación 70.

La Ecu. 6 muestra la actualización del estado de celda C_t, obtenida a partir de la combinación entre la información conservada del estado anterior C_{t-1} y la nueva información propuesta por $\tilde{C}_t$, ponderada por las compuertas f_t e i_t. La Ecu. 7 representa la compuerta de salida o_t, encargada de registrar la información que será enviada al estado oculto; en esta operación, W_0 corresponde a la matriz de pesos y b_0 al sesgo de salida. Finalmente, la Ecu. 8 define el estado ocultoh_t, obtenido a partir de la compuerta de salida o_t y del estado de celda actualizado C_t, mediante la función $tanh$ [73].

El funcionamiento de esta arquitectura se basa en la capacidad de mantener y actualizar información relevante a lo largo de múltiples pasos de la secuencia, lo que permite una representación más estable de los patrones temporales [70]. En la arquitectura analizada, las LSTM procesan secuencias de video interpretadas como estados consecutivos que describen la evolución del gesto, permitiendo modelar la dinámica completa de las señas en el tiempo[73].

Como se muestra en la Fig. 4, esta arquitectura incorpora mecanismos de control que regulan la información que se conserva, se descarta y se transmite en cada paso, manteniendo actualizado el estado de la celda de memoria.

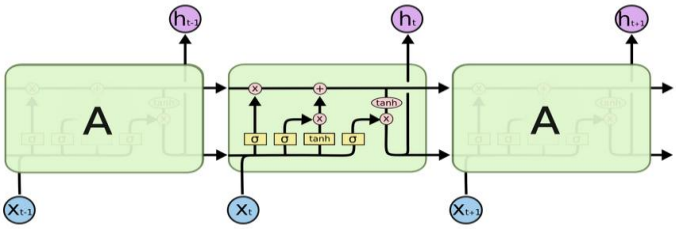


Fig. 4 Estructura interna de una red LSTM

Fuente: Adaptado de [74]

4. Redes BiLSTM (Bidirectional Long Short-Term Memory)

Las BiLSTM se consideraron como una arquitectura de modelado secuencial basada en el procesamiento bidireccional de la información dentro del sistema de reconocimiento de Lengua de Señas Colombiana. Esta arquitectura extiende el funcionamiento de las LSTM mediante la incorporación de dos flujos de procesamiento: uno en orden temporal directo y otro en sentido inverso [72].

Desde la parte estructural, una red BiLSTM está compuesta por dos unidades LSTM que operan sobre la misma secuencia de entrada. La salida bidireccional en el instante t, se obtiene mediante la concatenación de los estados ocultos generados en ambas direcciones, como se muestra en la Ecu. 9**:**

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \qquad \textit{Ecu. 9}$$

Donde h_t corresponde a la representación final de la secuencia en el instante t, integrando información temporal previa y posterior de la secuencia procesada [75]. Esta formulación se incorporó en el modelo para representar el procesamiento de las secuencias de entrada $X \in \mathbb{R}^{32 \times 132}$, conformadas por 32 fotogramas y 132 características extraídas por fotograma. Esta configuración permite integrar información proveniente tanto de estados previos como de estados posteriores dentro de la secuencia [75]:

El funcionamiento de esta arquitectura se basa en el análisis simultáneo de la secuencia en ambas direcciones, lo que permite construir una representación más completa del contexto temporal. En la arquitectura analizada, las BiLSTM procesan secuencias de video interpretadas como estados

consecutivos que describen la evolución del gesto, incorporando información contextual tanto pasada como futura en cada punto de la secuencia[76].

La Fig. 5 presenta la estructura general de una red BiLSTM, donde se combinan los dos recorridos de la secuencia, permitiendo que cada estado integre información de los elementos anteriores y posteriores dentro del proceso de modelado.

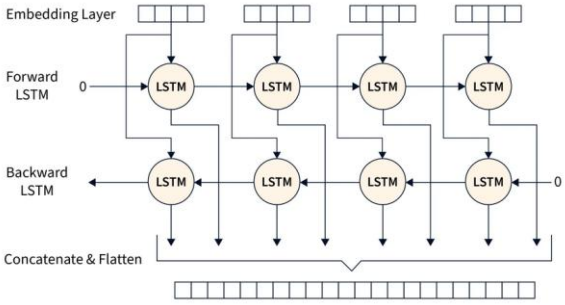


Fig. 5 Estructura de una red BiLSTM con procesamiento bidireccional

Fuente: Adaptado de [76]

5. Mecanismo de Atención (Attention Mechanism)

El mecanismo de atención se incorporó como una técnica complementaria al modelado secuencial dentro del sistema de reconocimiento de Lengua de Señas Colombiana, orientada a ponderar la información a lo largo del tiempo [5]. Este enfoque permite asignar diferentes niveles de importancia a los elementos de una secuencia, priorizando aquellos más relevantes para la representación[5].

el mecanismo de atención se implementa como una capa adicional que opera sobre los estados ocultos generados por una arquitectura secuencial, como las LSTM [5]. A partir de estos estados, se calcula un conjunto de pesos de atención que determina la contribución de cada instante dentro de la secuencia [5]. Este proceso se expresa mediante:

$$e_t = tanh(W_h h_t + b) \quad \textit{Ecu. 10}$$

$$\alpha_t = \frac{exp\,(e_t)}{\sum_k exp\,(e_k)} \qquad \textit{Ecu. 11}$$

$$c = \sum_t \alpha_t h_t \qquad \textit{Ecu. 12}$$

La Ecu. 10 representa el puntaje de atención e_t, calculado a partir del estado oculto h_t mediante la matriz de pesos W_h y el sesgo b, utilizando la función tangente hiperbólica $tanh$; esta expresión registra la relevancia asociada al instante temporal t. La Ecu. 11 define el peso de atención αtαt, obtenido mediante una normalización Softmax sobre los puntajes e_t; en esta ecuación, α_t corresponde al peso asignado al estado oculto h_t, mientras que $\sum_k \exp(e_k)$ representa la suma de los valores exponenciales de todos los puntajes de atención de la secuencia. Finalmente, la Ecu. 12 representa el vector de contexto c, calculado como la suma ponderada de los estados ocultos h_t por sus respectivos pesos de atención α_t; de esta manera, c corresponde a la representación final de la secuencia utilizada posteriormente por las capas densas y la clasificación del modelo.[5].

El funcionamiento de este mecanismo se basa en la generación de una representación ponderada de la secuencia, en la cual cada estado contribuye según su importancia relativa[5]. El mecanismo de atención se centra sobre secuencias de video interpretadas como estados consecutivos del gesto, permitiendo resaltar los instantes más representativos durante la ejecución de cada seña[5].

La Fig. 6 presenta el esquema general del mecanismo de atención aplicado sobre una secuencia, donde se ilustra el proceso de asignación de pesos y la generación de un vector de contexto a partir de los estados del modelo 74. Este mecanismo permite construir una representación en la que se priorizan los estados más relevantes dentro del proceso de modelado [5].

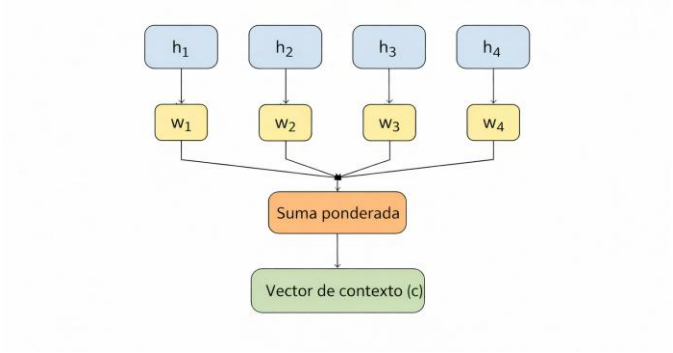


Fig. 6 Esquema del mecanismo de atención sobre secuencias

Fuente: Elaboración propia

6. Arquitecturas basadas en Transformers

Las arquitecturas Transformers se caracterizan por el uso de mecanismos de autoatención que permiten modelar relaciones entre los elementos de una secuencia de manera global [77]. Este enfoque permite que cada elemento interactúe directamente con los demás, sin depender de un procesamiento estrictamente secuencial, lo que facilita la construcción de representaciones más completas de la información [77].

Desde el punto de vista estructural, estos modelos se organizan en bloques que integran capas de atención, transformaciones lineales y funciones de activación [75]. Cada elemento de la secuencia se proyecta en tres representaciones denominadas consulta (Query), clave (Key) y valor (Value), las cuales permiten calcular la atención mediante la siguiente expresión 75:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad \textit{Ecu. 13}$$

La Ecu. 13 representa el mecanismo de atención escalada utilizado en el modelo Transformer, donde la función $\mathrm{Attention}(\mathrm{Q}, \mathrm{K}, \mathrm{V})$ calcula la relación entre las consultas Q, las claves K y los valores V; en esta expresión, el término QK^T corresponde al producto entre consultas y claves, el cual permite medir la similitud entre los elementos de la secuencia, mientras que d_k representa un factor de escala asociado a la dimensión de las claves d_k [77]. Posteriormente, se aplica la función **Softmax** para normalizar estos valores y obtener pesos de atención, los cuales se utilizan para ponderar la matriz de valores V [77]. Como resultado, se obtiene una representación que integra la información relevante de la secuencia de entrada, la cual fue utilizada dentro del modelo Transformer implementado sobre secuencias $\mathrm{X} \in \mathbb{R}^{32 \times 132}$ [77].

El funcionamiento de esta arquitectura se basa en la generación de matrices de atención que cuantifican la relación entre los elementos de la secuencia [77]. A partir de estas relaciones, el modelo produce nuevas representaciones en las que cada elemento incorpora información contextual proveniente de múltiples posiciones, este proceso se repite a través de varias capas, permitiendo una transformación progresiva de la información [77].

a información se procesa a partir de secuencias de video organizadas como estados temporales, lo que permite representar los distintos instantes del gesto y modelar interacciones entre ellos dentro del proceso de reconocimiento.

La Fig. 7 presenta un esquema del mecanismo de atención en arquitecturas Transformer, donde se ilustran las transformaciones de entrada y el cálculo de la atención a partir de las representaciones de consulta, clave y valor [77]. El modelo organiza el procesamiento mediante módulos de codificación y decodificación, en los cuales el mecanismo de atención permite integrar información entre los elementos de la secuencia [77].

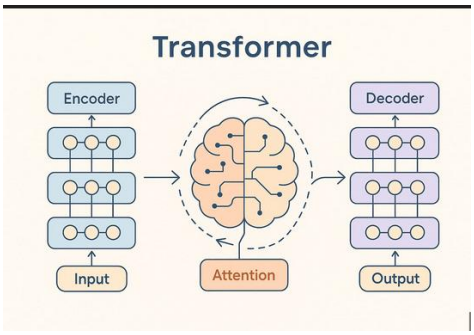


Fig. 7 Esquema de arquitecturas Transformer

Fuente: Adaptado de [78]

La TABLA. I presenta un resumen de las arquitecturas de modelado investigadas, en el cual se describen sus características principales, capacidad de modelado temporal, complejidad computacional, requerimientos de datos, ventajas, limitaciones y su pertinencia dentro del contexto del proyecto

TABLA. I

COMPARACIÓN DE ARQUITECTURAS DE MODELADO EVALUADAS PARA EL RECONOCIMIENTO DE LENGUA DE SEÑAS COLOMBIANA.

Arquitectura	**Tipo de Modelo**	**Capacidad de modelado temporal**	**Complejidad**	**Requerimiento de datos**	**Ventajas**	**Limitaciones**	**Pertinencia para el Proyecto**
CNN	Espacial	Baja (por sí sola)	Media–Alta	Alta	Excelente extracción de patrones espaciales en imágenes	No modela dependencias temporales directamente	Adecuada para señas estáticas, limitada para señas dinámicas
RNN	Secuencial	Media	Media	Media	Modela dependencias temporales	Problema de gradiente en secuencias largas	Parcialmente adecuada, con limitaciones en secuencias complejas
LSTM	Secuencial con memoria	Alta	Media–Alta	Media	Maneja dependencias temporales largas de forma estable	Mayor complejidad que RNN simple	Adecuada para el modelado de señas dinámicas

BiLSTM	Secuencial bidireccional	Muy Alta	Alta	Media	Analiza la secuencia en ambas direcciones	Mayor costo computacional	Pertinente, pero con mayor complejidad
Mecanismos de Attention	Secuencial con ponderación temporal	Muy Alta	Alta	Media	Enfoca el modelo en segmentos relevantes de la secuencia	Implementación más compleja	Pertinente, con mayor complejidad de implementación
Transformers	Secuencial basado en autoatención	Muy Alta	Muy Alta	Alta	Alto rendimiento en tareas secuenciales complejas	Requiere gran volumen de datos y recursos computacionales	Viable teóricamente, pero no óptima para el contexto del proyecto

Fuente: Resumen Adaptado de [4], [5], [72], [73], [77], [78]

Técnicas de captura y representación de datos

En el desarrollo del proyecto se analizaron diferentes métodos para la captura y representación de la información visual asociada a las señas. Se consideraron dos enfoques comunes en sistemas de reconocimiento: la representación basada en video en formato crudo y la representación basada en la extracción de puntos clave (landmarks).

1. Representación basada en video crudo

El enfoque basado en video crudo consiste en el uso de secuencias de fotogramas en formato RGB como entrada al modelo. En este caso, cada fotograma contiene información visual de la escena, incluyendo características como color, iluminación, fondo y elementos del entorno [79].

Las secuencias de video se organizan como conjuntos de imágenes consecutivas que preservan la información temporal asociada al desarrollo de cada seña. Cada fotograma representa un instante dentro de la secuencia, lo que permite registrar la evolución del movimiento a lo largo del tiempo [79].

La Fig. 8 presenta un ejemplo de esta representación, donde se observan distintos fotogramas correspondientes a una secuencia de movimiento.

Fig. 8 Representación basada en secuencia de fotogramas de video

Fuente: Elaboración propia

Este tipo de representación implica el manejo de datos de alta dimensionalidad, debido a la cantidad de información contenida en cada imagen. Asimismo, las secuencias pueden presentar variaciones asociadas a las condiciones de captura, como cambios en la iluminación, el fondo y la calidad del video [79].

2. Representación basada en extracción de landmarks

El enfoque basado en la extracción de puntos clave consiste en la obtención de información estructurada a partir de secuencias de video mediante la detección de puntos anatómicos del cuerpo, especialmente en las manos y la parte superior del torso [6]. Para este propósito, se utilizó el framework MediaPipe, el cual permite identificar puntos clave en manos y en la parte superior del cuerpo a partir de cada fotograma [6].

Cada punto detectado se representa mediante coordenadas en el espacio tridimensional, describiendo la posición de articulaciones en cada instante de la secuencia [78]. A partir de estos datos, se generan estructuras numéricas organizadas como secuencias, en las cuales cada conjunto de coordenadas corresponde a un paso temporal dentro del gesto, es decir, organizadas como secuencias numéricas que describen la posición y el movimiento de las articulaciones a lo largo del tiempo [6].

La Fig. 9 presenta un ejemplo de esta representación, donde se observan los puntos clave detectados durante la ejecución del movimiento.

Fig. 9 Representación basada en extracción de puntos clave

Fuente: Elaboración propia

Estas secuencias se utilizaron como datos de entrada para los modelos de reconocimiento, preservando la información temporal del movimiento mediante la organización ordenada de los puntos clave extraídos de cada fotograma [6].

En la TABLA. II se presenta una síntesis de las técnicas de captura y representación de datos utilizadas, así como sus características principales y la forma en que estructuran la información para su uso en el modelo.

TABLA. II

CARACTERÍSTICAS DE LAS TÉCNICAS DE CAPTURA Y REPRESENTACIÓN DE DATOS UTILIZADAS.

Técnica	Tipo de datos	Forma de representación	Estructura de los datos	Elementos considerados
Video crudo (RGB)	Imágenes	Fotogramas en color (RGB)	Secuencia de imágenes	Color, iluminación, fondo, forma, entorno
Landmarks (puntos clave)	Coordenadas numéricas	Puntos anatómicos en 2D/3D	Secuencia de vectores numéricos	Posición de manos, dedos y parte superior del cuerpo

Fuente: Resumen Adaptado de [6], [79]

Metodologías analizadas

En el desarrollo del proyecto se aplicaron metodologías orientadas a la construcción del sistema de reconocimiento de Lengua de Señas Colombiana. Estas estructuraron el proceso en etapas que incluyen la captura de datos, el procesamiento de la información, el entrenamiento de modelos y la integración en un aplicativo web.

1. Metodología experimental comparativa

La metodología experimental comparativa analiza diferentes arquitecturas de aprendizaje profundo y técnicas de representación de datos en el reconocimiento de secuencias visuales. Este enfoque utiliza métricas de evaluación como la precisión (accuracy) y la pérdida (loss), así como variables asociadas al comportamiento del modelo durante el entrenamiento [80]. Este tipo de metodología es propio del aprendizaje automático, donde los modelos se evalúan mediante métricas de desempeño que permiten comparar su comportamiento bajo distintas configuraciones [80].

2. Metodología de prototipado iterativo

La metodología de prototipado iterativo se basa en ciclos sucesivos de entrenamiento y ajuste del modelo, en los cuales se realizan múltiples iteraciones con modificaciones en la arquitectura, los parámetros de entrenamiento y las estrategias de procesamiento de datos [81].

Cada iteración permite evaluar el comportamiento del modelo mediante métricas como la precisión y la pérdida, facilitando el seguimiento de su evolución a lo largo del proceso [81]. Este enfoque corresponde a prácticas de desarrollo iterativo en sistemas de aprendizaje automático, donde los modelos se optimizan progresivamente mediante ciclos de experimentación [81].

En la TABLA. III se muestra el resumen de las metodologías analizadas para el desarrollo del sistema.

TABLA. III
METODOLOGÍAS ANALIZADAS PARA EL DESARROLLO DEL SISTEMA DE RECONOCIMIENTO DE LSC.

Metodología	Enfoque	Aplicación en el proyecto	Aporte principal
Experimental comparativa	Evaluación técnica	Pruebas con diferentes arquitecturas y representaciones de datos	Permite comparar desempeño y seleccionar la mejor alternativa
Prototipado iterativo	Desarrollo progresivo	Ajuste del modelo mediante ciclos sucesivos de entrenamiento	Facilita la mejora continua del modelo

Fuente: Resumen adaptado de [80], [81], [82]

Métricas de evaluación

Para la evaluación del desempeño de los modelos se emplearon métricas estándar en problemas de clasificación supervisada, específicamente la precisión (accuracy) y la función de pérdida (loss), las cuales permiten analizar tanto la exactitud de las predicciones como el comportamiento del modelo durante el entrenamiento [83].

La precisión (accuracy) cuantifica la proporción de predicciones correctas respecto al total de muestras evaluadas, constituyendo un indicador global del rendimiento del modelo. Se define como:

$$Accuracy = \frac{Número\ de\ predicciones\ correctas}{Número\ total\ de\ muestras} \qquad Ecu.\ 14$$

Esta métrica permite evaluar el grado de acierto del modelo en tareas de clasificación y comparar el desempeño entre diferentes arquitecturas bajo condiciones equivalentes[83].

Por su parte, la función de pérdida (loss) mide el error entre las predicciones generadas por el modelo y las etiquetas reales, proporcionando información sobre la calidad del aprendizaje durante el entrenamiento. En modelos de clasificación, se expresa como lo muestra la Ecu. 15:

$$Loss = -\sum_{i=1}^{N} y_i log(\hat{y}_i) \qquad Ecu.\ 15$$

donde y_i representa la etiqueta real y $\hat{y}_i$ la probabilidad predicha por el modelo para cada clase. La reducción progresiva de la pérdida indica una mejor aproximación del modelo a los datos [84]. En conjunto, estas métricas permiten realizar una evaluación integral del desempeño del modelo, considerando tanto la precisión de las predicciones como la estabilidad del proceso de entrenamiento.

B. DESARROLLO DE UN APLICATIVO WEB PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA (LSC), UTILIZANDO LAS TÉCNICAS Y METODOLOGÍAS PREVIAMENTE ANALIZADAS.

Para el desarrollo del aplicativo web se empleó la metodología Scrumban, entendida como un enfoque ágil y flexible que integra elementos de Scrum y Kanban [85]. Esta metodología permitió organizar el trabajo mediante tareas definidas, priorizadas y ajustadas de acuerdo con las necesidades del proyecto, en lugar de estructurar el desarrollo en fases rígidas. De esta manera, el proceso de construcción del sistema se orientó a la ejecución progresiva de actividades relacionadas con la grabación de los videos para la base de datos, la organización y categorización de las clases, el procesamiento de las muestras, el entrenamiento y evaluación de modelos de inteligencia artificial, la selección del modelo con mejor desempeño, la implementación del backend, la integración con el frontend y la realización de pruebas funcionales del aplicativo. Para la gestión de estas actividades, se tomó la definición de tareas propuesta por Scrum y se organizó el flujo de

trabajo mediante un tablero Kanban, permitiendo visualizar el estado y avance de cada actividad, tal como se evidencia en el **Anexo. 10** .

La elección de Scrumban favoreció un desarrollo adaptable, ya que permitió realizar ajustes durante el avance del sistema a partir de los resultados obtenidos en las pruebas y de las necesidades identificadas durante la validación. En este sentido, las actividades descritas en este apartado corresponden a tareas desarrolladas de forma iterativa e incremental, enfocadas en consolidar una herramienta funcional para el reconocimiento de palabras en Lengua de Señas Colombiana dentro del alcance definido para la investigación.

1. Construcción del dataset y preparación de datos

Después de evaluar las técnicas de reconocimiento aplicables al proyecto, se realizó la construcción de un dataset desde cero, debido a que no se contaba con una base de datos local ajustada al contexto, vocabulario y alcance definidos para el sistema. El conjunto de datos fue conformado por videos de palabras en Lengua de Señas Colombiana, organizados de acuerdo con la seña representada. Esta organización permitió disponer de una base etiquetada para las etapas posteriores de entrenamiento, validación e inferencia.

La construcción del dataset se desarrolló de manera progresiva. En una primera versión, se recopilaron videos realizados por una sola persona, con un conjunto aproximado de cincuenta señas o clases. Los videos fueron organizados en carpetas independientes, de modo que cada carpeta correspondiera a una palabra específica. Esta primera estructura permitió iniciar las pruebas del sistema y establecer una organización base para el manejo del material.

Posteriormente, se realizó una segunda versión del dataset con la participación de dos personas, manteniendo inicialmente un número amplio de clases Fig. 10. Esta nueva versión incorporó muestras adicionales para las señas trabajadas y permitió ampliar el registro de ejecuciones disponibles para cada palabra. Los videos continuaron siendo clasificados por carpetas según la clase correspondiente, conservando la organización necesaria para su uso en los procesos de entrenamiento.

Finalmente, se construyó una versión más controlada del dataset, en la cual se amplió la participación de personas en la grabación de las muestras y se redujo el número de clases trabajadas. Esta versión se orientó al registro de un grupo específico de palabras en Lengua de Señas Colombiana, seleccionadas de acuerdo con el alcance del aplicativo y las condiciones de prueba definidas para el proyecto. Entre las clases incluidas en las versiones finales se trabajaron palabras como **adiós, auditorio, ayuda, baño, biblioteca, cafetería, carnet, cuándo y dónde** como se puede ver en la Fig. 11**.**

El dataset final quedó organizado por clases, conservando una estructura de carpetas que permitió identificar cada palabra y sus respectivos videos Fig. 12. Esta organización facilitó el control del material recolectado, la revisión de las muestras disponibles y su posterior utilización en los procesos de entrenamiento y validación de los modelos. De esta manera, el proyecto contó con una base de datos propia, construida a partir de videos recolectados específicamente para el reconocimiento de palabras aisladas en Lengua de Señas Colombiana.

Durante el proceso de construcción del dataset se registró la participación de un intérprete de Lengua de Señas Colombiana, quien supervisó la correcta ejecución de las señas incluidas en los videos. Asimismo, se contó con la colaboración de una persona sorda en la generación del material, lo cual permitió disponer de muestras representativas para el desarrollo del sistema.

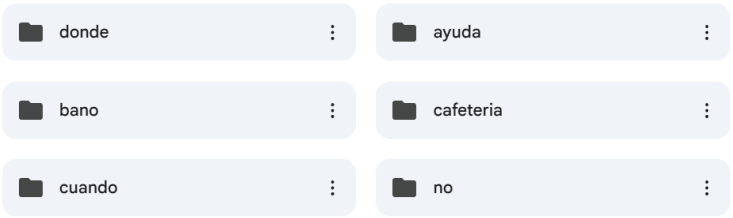


Fig. 10 Organización de clases

Fuente: Elaboración propia

Fig. 11 Videos por clase

Fuente: Elaboración propia

Fig. 12 Data final

Fuente: Elaboración propia

Adicionalmente, se construyó un conjunto externo de pruebas en entorno no controlado, conformado por 15 videos por cada una de las nueve clases reconocidas por el sistema, para un total de 135 videos. Cada video fue grabado por una persona diferente y bajo condiciones variables de captura, como iluminación, distancia, ángulo de cámara, dispositivo utilizado y velocidad de ejecución de la seña, como se muestra en la Fig. 13. Este conjunto no hizo parte del entrenamiento, validación ni prueba interna del modelo, sino que fue utilizado exclusivamente para evaluar el comportamiento del sistema frente a muestras externas no vistas.

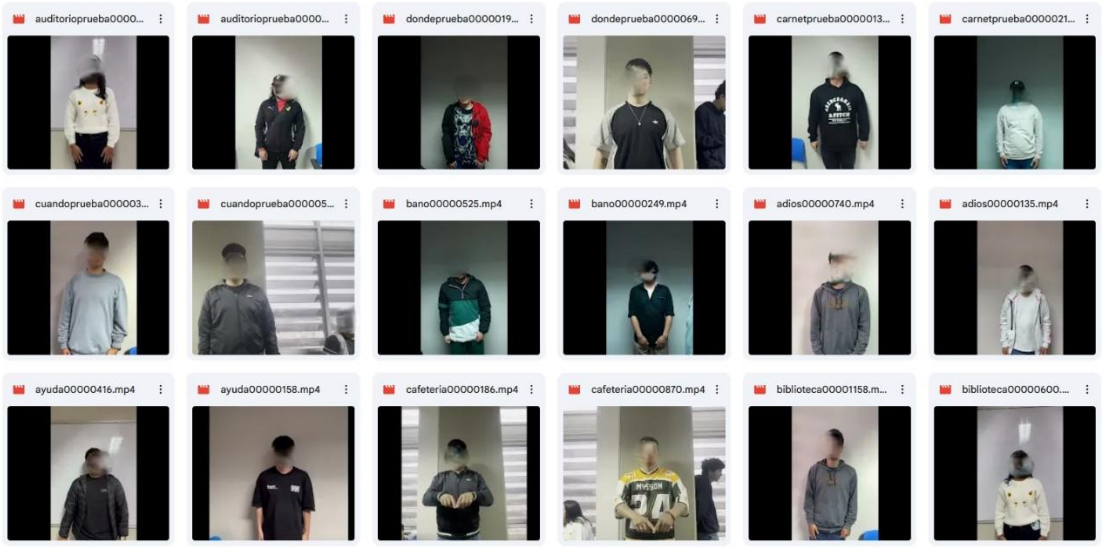


Fig. 13 Dataset externo no controlado.
Fuente: Elaboración propia

En la TABLA. IV se presenta la evolución del dataset en términos de número de participantes, clases y enfoque de construcción, evidenciando los cambios realizados durante el proceso de desarrollo.

TABLA. IV
EVOLUCIÓN DEL DATASET

Versión del dataset	Participantes	Número aproximado de clases	Enfoque de construcción	Resultado registrado
Dataset inicial	1 persona	50 clases	Construcción inicial desde cero con vocabulario amplio	Primera base organizada para pruebas iniciales
Dataset intermedio	2 personas	50 clases	Ampliación de participantes conservando el número de clases	Registro adicional de muestras por clase
Dataset final	25 personas	9 clases	Reducción del vocabulario y ampliación de participantes	Base final utilizada para las pruebas principales
Dataset externo no controlado	135 personas	9 clases	Grabación de 15 videos por clase en condiciones variables	Base externa utilizada para pruebas de inferencia no controladas

Fuente: Elaboración propia

Como se observa en la TABLA. V, el dataset original estuvo conformado por 1644 videos distribuidos en nueve clases. Durante su revisión se identificó una distribución no homogénea entre las categorías, debido a que algunas clases contaban con una mayor cantidad de muestras, como cafetería con 227 videos y adiós con 213 videos, mientras que otras presentaban una cantidad menor, como la clase dónde con 105 videos. Esta diferencia podía generar un desbalance durante el entrenamiento del modelo, favoreciendo las clases con mayor representación y afectando la capacidad de generalización del sistema.

Para reducir este efecto, se realizó un proceso de balanceo del dataset, estableciendo una muestra uniforme de 160 videos por clase. En las clases que superaban esta cantidad, se seleccionaron 160 videos reales. En el caso de la clase dónde, al contar únicamente con 105 videos originales, se generaron 55 muestras artificiales mediante técnicas de aumento de datos, con el fin de alcanzar el mismo número de muestras que las demás clases. No se incrementó esta clase por encima de 160 videos, debido a que un aumento artificial excesivo podía introducir variaciones poco representativas y afectar negativamente el aprendizaje del modelo.

Una vez balanceado, el dataset quedó conformado por 1440 videos. La división de los datos se realizó de forma aleatoria y estratificada por clase, manteniendo una proporción de 60 % para entrenamiento, 30 % para validación y 10 % para prueba. De esta manera, se obtuvieron 864 muestras para entrenamiento, 432 para validación y 144 para prueba. Los conjuntos resultantes fueron mutuamente excluyentes, por lo que los videos utilizados en entrenamiento no hicieron parte de validación ni de prueba. Esto permitió evaluar el modelo con datos no vistos durante su ajuste y observar su capacidad de generalización frente a muestras independientes.

Las técnicas de aumento de datos, como ruido y zoom, se aplicaron sobre el conjunto de entrenamiento, manteniendo los conjuntos de validación y prueba sin modificaciones artificiales, con el propósito de evaluar el comportamiento del modelo sobre muestras reales no alteradas. Todas las muestras fueron recolectadas bajo un ambiente controlado, considerando condiciones similares de iluminación, distancia, encuadre y ejecución de la seña.

TABLA. V
TAMAÑO EXACTO DEL DATASET Y DISTRIBUCIÓN POR CLASE

Clase	Videos originales	Videos reales usados	Videos artificiales generados	Total, balanceado	Entrenamiento 60 %	Validación 30 %	Prueba 10 %
adios	213	160	0	160	96	48	16
auditorio	162	160	0	160	96	48	16
ayuda	172	160	0	160	96	48	16
bano	192	160	0	160	96	48	16

biblioteca	190	160	0	160	96	48	16
cafeteria	227	160	0	160	96	48	16
carnet	188	160	0	160	96	48	16
cuando	195	160	0	160	96	48	16
donde	105	105	55	160	96	48	16
Total	**1644**	**1385**	**55**	**1440**	**864**	**432**	**144**

Fuente: Elaboración propia

Adicionalmente, se realizaron pruebas externas en ambiente no controlado, utilizando 15 videos por cada una de las nueve clases reconocidas por el sistema, para un total de 135 videos, como se visualiza en la TABLA. VI. Estas pruebas fueron realizadas con personas diferentes a las utilizadas en la construcción del dataset balanceado y bajo condiciones variables de captura, como iluminación, distancia, ángulo de cámara, dispositivo utilizado y velocidad de ejecución de la seña. Este conjunto no hizo parte del entrenamiento, validación ni prueba interna del modelo. Su propósito fue evaluar el comportamiento del aplicativo frente a muestras externas no vistas, cercanas a condiciones reales de uso. De esta manera, se complementó la evaluación realizada en ambiente controlado y se obtuvo una referencia adicional sobre la capacidad de generalización del sistema.

TABLA. VI
CANTIDAD DATA DE PRUEBAS NO CONTROLADAS

Clase	Videos externos no controlados
adios	15
auditorio	15
ayuda	15
bano	15
biblioteca	15
cafeteria	15
carnet	15
cuando	15
dónde	15
Total	135

Fuente: Elaboración Propia.

Estas pruebas permitieron contrastar el desempeño del sistema en condiciones menos controladas que las usadas durante la construcción del dataset, sin alterar los conjuntos internos de entrenamiento, validación y prueba.

2. Resultados del modelo CNN 1D Baseline

El modelo CNN fue evaluado con una entrada de 32 × 132, correspondiente a 32 fotogramas por video y 132 características por fotograma. Las clases evaluadas fueron: adios, auditorio, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde. Las etiquetas se mantuvieron sin tildes ni caracteres especiales durante el entrenamiento y la clasificación.

La arquitectura implementada estuvo conformada por capas Conv1D, MaxPooling1D, Dropout, GlobalAveragePooling1D, Dense y Softmax. La estructura general se presenta en la Fig. 14. El modelo se empleó como línea base de comparación**.** En las pruebas de inferencia presentó menor estabilidad frente a las demás arquitecturas evaluadas; por tanto, no fue seleccionado como modelo final**.**

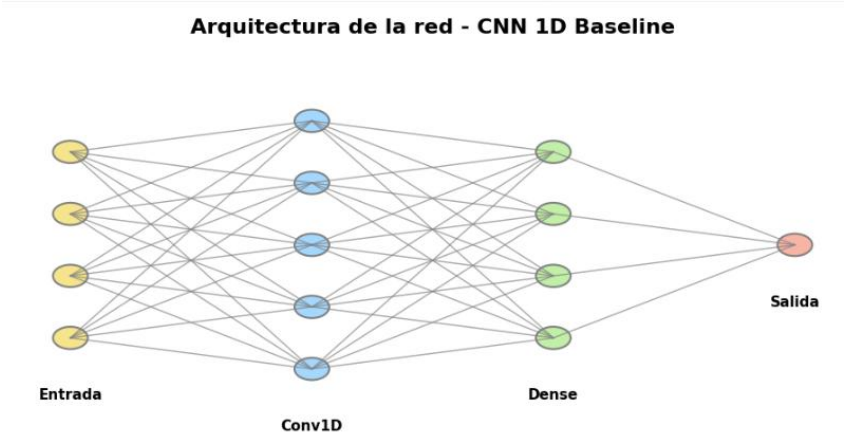


Fig. 14 . Arquitectura CNN.
Fuente: Elaboración propia.

El modelo CNN presentó una entrada (None, 32, 132) y una salida (None, 9), correspondiente a las nueve clases evaluadas. La arquitectura estuvo conformada por dos capas Conv1D. La primera generó una salida (None, 32, 32) con 12.704 parámetros, seguida de MaxPooling1D, que redujo la dimensión temporal a (None, 16, 32). La segunda capa Conv1D produjo una salida (None, 16, 64) con 6.208 parámetros.

La capa GlobalAveragePooling1D transformó la secuencia en un vector (None, 64). Posteriormente, la capa Dense mantuvo una salida (None, 64) con 4.160 parámetros, y la capa final Softmax generó la salida (None, 9) con 585 parámetros.

En total, la arquitectura registró 23.657 parámetros entrenables. La distribución de capas y parámetros se presenta en la Fig. 15.

Model: "CNN 1D Baseline"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 32, 132)	0
conv1d_1 (Conv1D)	(None, 32, 32)	12,704
maxpool1d_1 (MaxPooling1D)	(None, 16, 32)	0
dropout_1 (Dropout)	(None, 16, 32)	0
conv1d_2 (Conv1D)	(None, 16, 64)	6,208
dropout_2 (Dropout)	(None, 16, 64)	0
global_avg_pool (GlobalAveragePooling1D)	(None, 64)	0
dense_1 (Dense)	(None, 64)	4,160
dropout_3 (Dropout)	(None, 64)	0
output_layer (Dense)	(None, 10)	650

Total params: 23,722 (92.66 KB)
Trainable params: 23,722 (92.66 KB)
Non-trainable params: 0 (0.00 B)

Fig. 15 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La formulación matemática presentó el flujo del modelo desde la entrada de datos hasta la predicción final. La *Ecu. 16* representa la entrada del modelo definida como $X \in \mathbb{R}^{32 \times 132}$, correspondiente a 32 fotogramas y 132 características por fotograma. La Ecu. 17 define la operación de convolución unidimensional, donde $z_{t,k}^{(l)}$ corresponde a la salida del filtro k en la capa ll, calculada a partir de los pesos $\omega_{i,c,k}^{(l)}$, la entrada $x_{t+i,c}^{(l-1)}$ y el sesgo $b_k^{(l)}$. La Ecu. 18 representa la función de activación ReLU, mediante la cual se obtiene $a_{t,k}^{(l)}$ como la salida activada. La Ecu. 19 muestra la operación de MaxPooling, donde $p_{t,k}$ corresponde al valor máximo dentro de una ventana de tamaño m. La Ecu. 20 define el promedio global temporal g_k, calculado como la media de las activaciones a lo largo de la dimensión temporal. La Ecu. 21 representa la capa densa, donde z_j corresponde a la combinación lineal de las características de entrada x_i con los pesos w_{ij} y el sesgo b_j. La Ecu. 22 define la función Softmax, utilizada para obtener la probabilidad $\hat{y}_j$ asociada a cada clase. Finalmente, la Ecu. 23 representa la función de pérdida de entropía cruzada categórica, utilizada durante el entrenamiento del modelo para calcular la diferencia entre las etiquetas reales y_j y las predicciones $\hat{y}_j$

$$X \in \mathbb{R}^{32 \times 132} \qquad \textit{Ecu. 16}$$

$$z_{t,k}^{(l)} = \sum_{i=0}^{K-1} \sum_{c=1}^{C_{in}} \omega_{i,c,k}^{(l)} x_{t+i,c}^{(l-1)} + b_k^{(l)} \qquad \textit{Ecu. 17}$$

$$a_{t,k}^{(l)} = ReLU\left(z_{t,k}^{(l)}\right) = max\left(0, z_{t,k}^{(l)}\right) \qquad \textit{Ecu. 18}$$

$$p_{t,k} = max\left(a_{t,k}, a_{t+1,k}, \ldots, a_{t+m-1,k}\right) \qquad \textit{Ecu. 19}$$

$$g_k = \frac{1}{T'} \sum_{t=1}^{T'} a_{t,k} \qquad \textit{Ecu. 20}$$

$$z_j = \sum_{i=1}^{n} w_{ij} x_i + b_j \qquad \textit{Ecu. 21}$$

$$\hat{y}_j = \frac{e^{Zj}}{\sum_{k=1}^{C} e^{Zk}} \qquad \textit{Ecu. 22}$$

$$\mathcal{L} = -\sum_{j=1}^{C} y_j \, log\left(\hat{y}_j\right) \qquad \textit{Ecu. 23}$$

Métricas de entrenamiento del modelo CNN

El modelo CNN fue entrenado durante 50 épocas con una entrada de 32 × 132. La precisión de entrenamiento registró un incremento de 0,20 a 0,90. La precisión de validación inició en 0,37 y alcanzó valores finales entre 0,93 y 0,94.

Las curvas de precisión se presentan en la Fig. 16 y fueron generadas en Python con apoyo de TensorFlow/Keras, NumPy y Matplotlib.

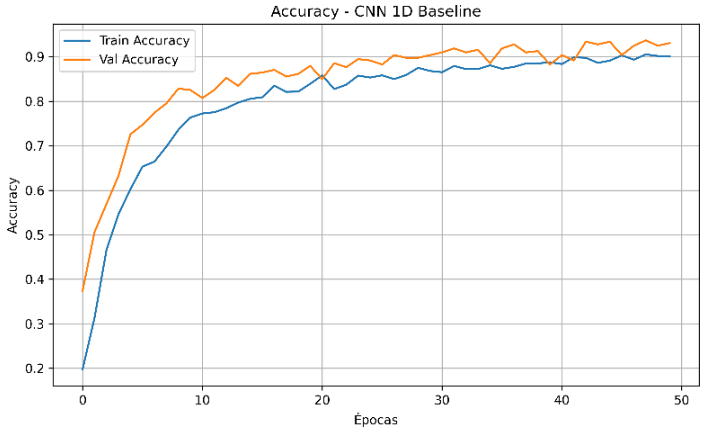


Fig. 16 Grafica de precisión.
Fuente: Elaboración propia.

La pérdida de entrenamiento disminuyó de un valor inicial superior a 2,0 hasta un valor final cercano a 0,24. La pérdida de validación inició aproximadamente en 1,79 y finalizó cerca de 0,17. Las curvas de pérdida se presentan en la Fig. 17.

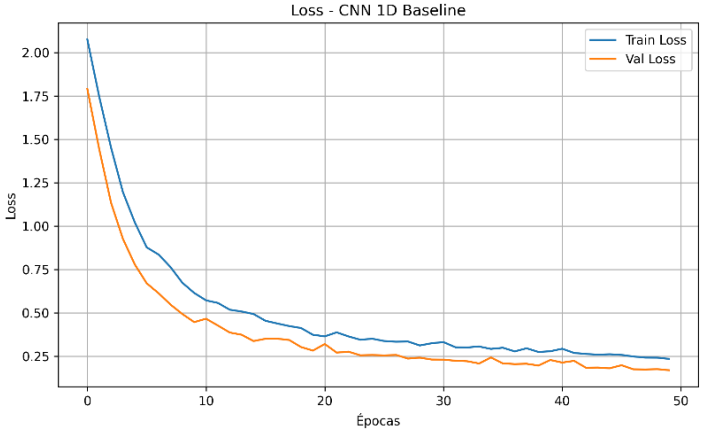


Fig. 17 Grafica de perdida.

Fuente: Elaboración propia.

Matriz de confusión del modelo CNN

La matriz de confusión registró aciertos en las nueve clases evaluadas. Los mayores valores se observaron en cafeteria con 46 aciertos, adios con 41, y carnet y cuando con 38 aciertos cada una. Los menores valores se presentaron en donde con 21 aciertos, y en bano y biblioteca con 29 aciertos cada una. Los errores principales correspondieron a bano - biblioteca con 10 casos y biblioteca - bano con 6 casos. Además, se registraron errores menores en biblioteca - adios con 3 casos, adios – biblioteca con 2 casos, ayuda - donde con 1 caso y cuando - bano con 1 caso. La matriz correspondiente se presenta en la Fig. 18.

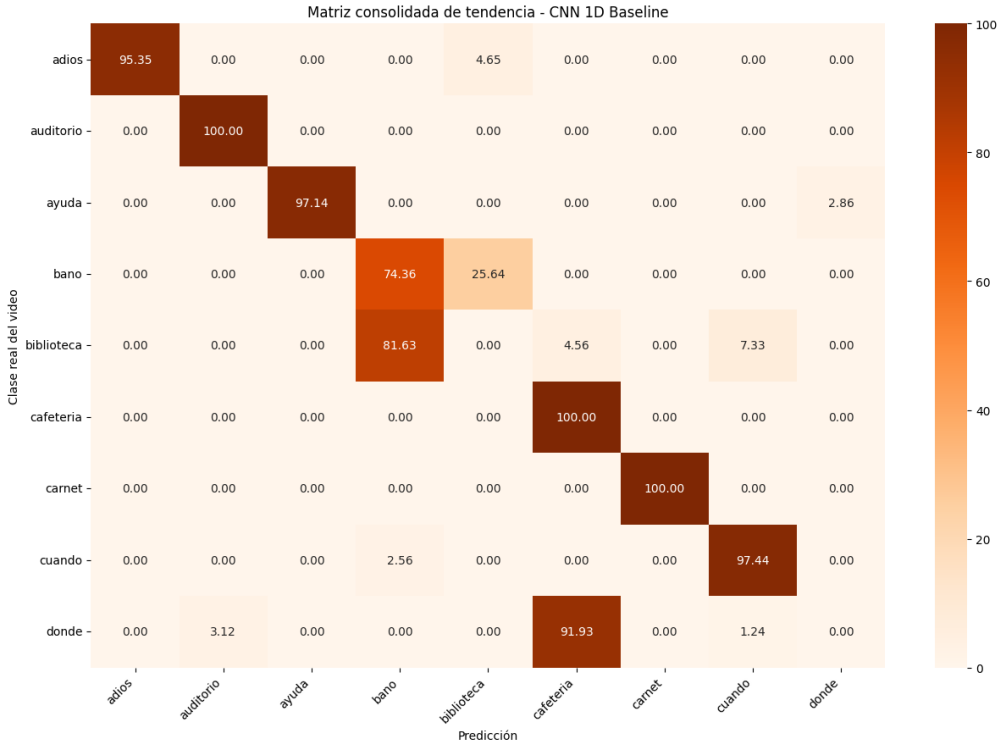


Fig. 18 Matriz de confusión de CNN.

Fuente: Elaboración propia.

Prueba de inferencia externa del modelo CNN

La inferencia externa se realizó con el archivo biblioteca (1).mp4, correspondiente a la clase real biblioteca. El modelo clasificó el video como bano, con una confianza de 81,63 %; por tanto, la predicción no coincidió con la etiqueta real.

El top 3 de predicciones fue: bano con 81,63 %, cuando con 7,33 % y cafeteria con 4,56 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,1780, movimiento medio = 2,2723 y margen = 0,7430. El resultado de la inferencia se presenta en la Fig. 19**.**

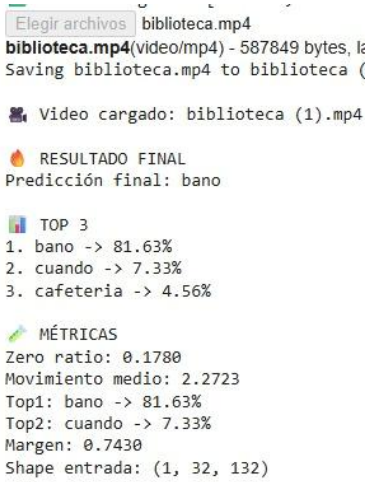


Fig. 19 Resultado de inferencia

Fuente: Elaboración propia.

Otra prueba de inferencia se realizó con el archivo baño (2).mp4, correspondiente a la clase real bano. El modelo clasificó el video como adios, con una confianza de 97,49 %; por tanto, la predicción no coincidió con la etiqueta real.

El top 3 de predicciones fue: adios con 97,49 %, carnet con 1,63 % y biblioteca con 0,29 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,6136, movimiento medio = 2,5040, margen = 0,9586 y tiempo total = 4,3269 s. El resultado de la inferencia se presenta en la Fig. 20**.**

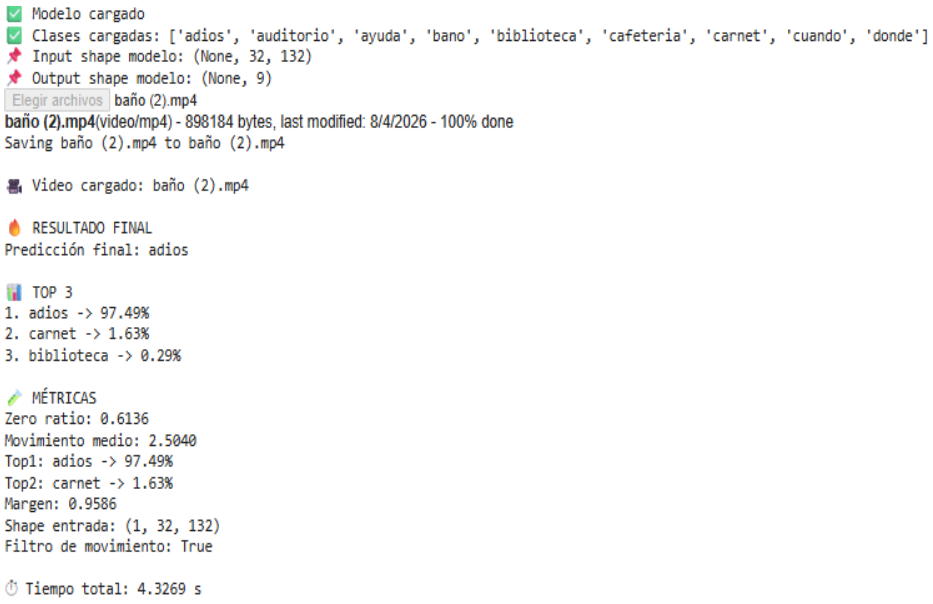


Fig. 20 Resultado de inferencia

Fuente: Elaboración propia.

3. Resultados del modelo Transformer

El modelo Transformer fue evaluado con una entrada de 32 × 132, correspondiente a 32 fotogramas por video y 132 características por fotograma. Las clases evaluadas fueron: adios, auditorio, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde. Las etiquetas se mantuvieron sin tildes ni caracteres especiales durante el entrenamiento y la clasificación.

La arquitectura implementada estuvo conformada por una capa de entrada, proyección de características, codificación posicional, bloque Transformer, capas Dense, GlobalAveragePooling1D y una capa final Softmax para la clasificación multiclase. La estructura general se presenta en la Fig. 21.

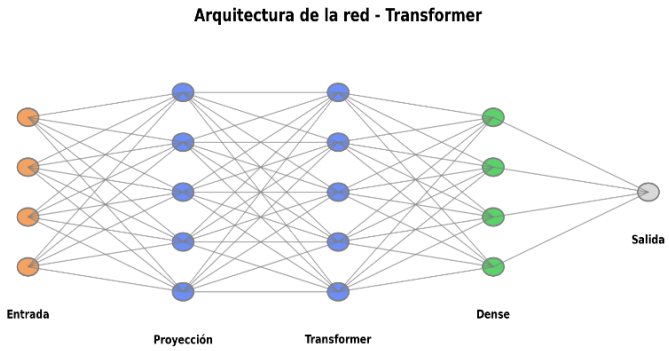


Fig. 21 Arquitectura general del modelo Transformer

Fuente: Elaboración propia.

El modelo Transformer presentó una entrada de (None, 32, 132) y una salida de (None, 9), correspondiente a las nueve clases evaluadas. La capa input_projection registró una salida de (None, 32, 64) con 8.512 parámetros. La capa positional_encoding mantuvo la forma (None, 32, 64). El bloque de atención incluyó una capa MultiHeadAttention con salida de (None, 32, 64) y 8.352 parámetros, seguida de capas Add y LayerNormalization; esta última registró 128 parámetros. El bloque denso interno incorporó dos capas Dense, cada una con salida de (None, 32, 64) y 4.160 parámetros. Posteriormente, GlobalAveragePooling1D transformó la secuencia en un

vector de (None, 64). La capa Dense posterior mantuvo una salida de (None, 64) con 4.160 parámetros. Finalmente, la capa Softmax generó la salida (None, 9) con 585 parámetros.

En total, el modelo Transformer registró 30.185 parámetros entrenables. La distribución de capas y parámetros se presenta en la Fig. 22**.**

RESUMEN DEL MODELO

Model: "transformer"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 32, 132)	0	-
input_projection (Dense)	(None, 32, 64)	8,512	input_layer[0][0]
positional_encoding (PositionalEncodin…	(None, 32, 64)	0	input_projection…
multi_head_attenti… (MultiHeadAttentio…	(None, 32, 64)	8,352	positional_encod… positional_encod…
add_10 (Add)	(None, 32, 64)	0	positional_encod… multi_head_atten…
layer_normalizatio… (LayerNormalizatio…	(None, 32, 64)	128	add_10[0][0]
dense_10 (Dense)	(None, 32, 64)	4,160	layer_normalizat…
dropout_16 (Dropout)	(None, 32, 64)	0	dense_10[0][0]
dense_11 (Dense)	(None, 32, 64)	4,160	dropout_16[0][0]
add_11 (Add)	(None, 32, 64)	0	layer_normalizat… dense_11[0][0]
layer_normalizatio… (LayerNormalizatio…	(None, 32, 64)	128	add_11[0][0]
dropout_17 (Dropout)	(None, 32, 64)	0	layer_normalizat…
global_avg_pool (GlobalAveragePool…	(None, 64)	0	dropout_17[0][0]
dense_1 (Dense)	(None, 64)	4,160	global_avg_pool[…
dropout_final (Dropout)	(None, 64)	0	dense_1[0][0]
output_layer (Dense)	(None, 9)	585	dropout_final[0]…

Total params: 30,185 (117.91 KB)
Trainable params: 30,185 (117.91 KB)
Non-trainable params: 0 (0.00 B)

Fig. 22 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La formulación matemática presentó el flujo del modelo desde la entrada de datos hasta la predicción final. La *Ecu. 24* representa la entrada del modelo definida como $X \in \mathbb{R}^{32 \times 132}$, correspondiente a 32 fotogramas y 132 características por fotograma. La Ecu. 25 define la proyección lineal de la entrada, donde $H = XW_e + b_e$, siendo W_e la matriz de pesos y bebe el sesgo de la proyección. La Ecu. 26 y la Ecu. 27 representan la codificación posicional, en la cual se incorpora información del orden temporal mediante funciones seno y coseno, donde pos corresponde a la posición del fotograma, ij al índice de dimensión y d a la dimensión interna del

modelo. La Ecu. 28 define la suma entre la proyección y la codificación posicional, obteniendo la representación Z. La Ecu. 29 representa el mecanismo de atención escalada, donde se calcula la relación entre consultas Q, claves Ky valores V, utilizando el producto QK^Ty un factor de escala d_k, seguido de una normalización Softmax. La Ecu. 30 representa la normalización con conexión residual, obteniendo Z' a partir de la suma entre Z y la salida de la atención. La Ecu. 31 define la red feed-forward, donde se aplica una transformación con activación $ReLU$. La Ecu. 32 muestra la segunda normalización con conexión residual, generando Z''. La Ecu. 33 representa el promedio global temporal g, calculado sobre la secuencia. La Ecu. 34 define la capa densa, donde Zj corresponde a la combinación lineal de las características. La Ecu. 35 representa la función Softmax, utilizada para obtener la probabilidad $\hat{y}_j$ de cada clase. Finalmente, la Ecu. 36 define la función de pérdida de entropía cruzada categórica, utilizada para medir la diferencia entre la distribución real de las clases y_j y la distribución predicha por el modelo $\hat{y}_j$ durante el entrenamiento.

$$X \in \mathbb{R}^{32 \times 132} \quad \textit{Ecu. 24}$$

$$H = XW_e + b_e \quad \textit{Ecu. 25}$$

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000_{\frac{2i}{d}}}\right), PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000_{\frac{2i}{d}}}\right) \quad \textit{Ecu. 26}$$

$$Z = H + PE \quad \textit{Ecu. 27}$$

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \textit{Ecu. 28}$$

$$MultiHead(Q,K,V) = Concat(head_1,\dots,head_n)W^O \quad \textit{Ecu. 29}$$

$$Z' = LayerNorm\left(Z + MultiHead(Z,Z,Z)\right) \quad \textit{Ecu. 30}$$

$$FNN(x) = ReLU(xW_1 + b_1) + W_2 + b_2 \quad \textit{Ecu. 31}$$

$$Z'' = LayerNorm\left(Z' + FFN(Z')\right) \quad \textit{Ecu. 32}$$

$$g = \frac{1}{T}\sum_{t=1}^{T} Z''t \quad \textit{Ecu. 33}$$

$$Zj = \sum_{i=1} W_{ij} X_i + b_j \qquad \textit{Ecu. 34}$$

$$\hat{y}_j = \frac{e^{Zj}}{\sum_{k=1}^{C} e^{Zk}} \qquad \textit{Ecu. 35}$$

$$\mathcal{L} = -\sum_{j=1}^{C} y_j log\,(\hat{y}_j) \qquad \textit{Ecu. 36}$$

Métricas de precisión del modelo Transformer

El modelo Transformer fue entrenado durante 50 épocas con secuencias de entrada 32 × 132. La precisión de entrenamiento aumentó de 0,29 a 0,95. La precisión de validación inició en 0,53 y alcanzó un rango final entre 0,93 y 0,94.

Los gráficos de precisión se presentan en la Fig. 23.

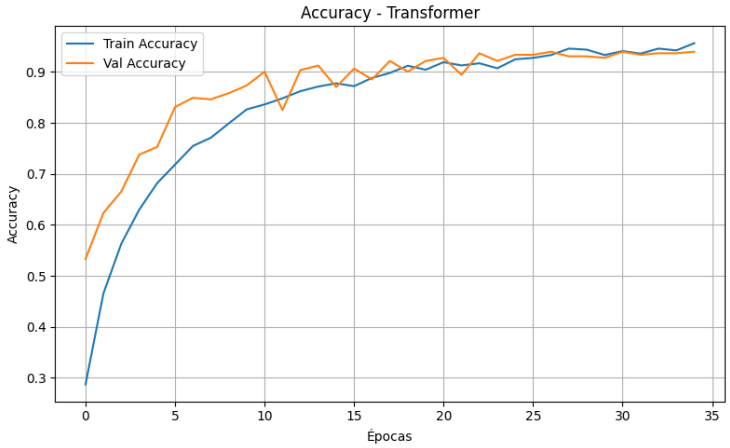


Fig. 23 Gráficos de precisión.

Fuente: Elaboración propia.

Métricas de pérdida del modelo Transformer

El modelo Transformer fue evaluado durante 50 épocas. La pérdida de entrenamiento inició en 1,98 y finalizó en 0,15. La pérdida de validación inició en 1,50 y alcanzó valores finales entre 0,14 y 0,15.

La curva de pérdida se presenta en la Fig. 24.

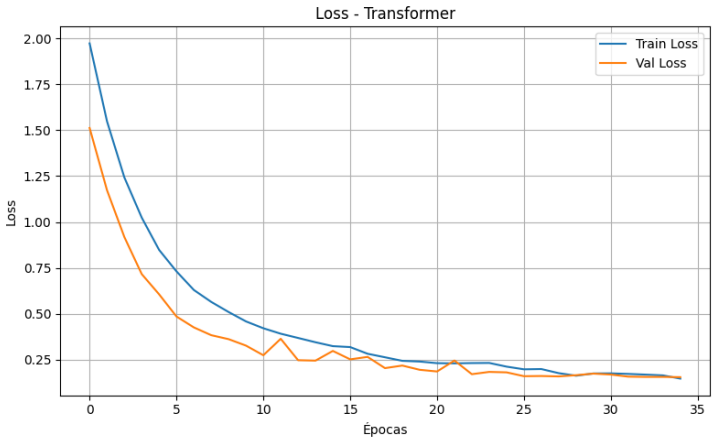


Fig. 24 Graficas de perdida.

Fuente: Elaboración propia .

Matriz de confusión del modelo Transformer

La matriz de confusión registró aciertos en las nueve clases evaluadas. Los mayores valores se observaron en cafeteria con 45 aciertos, adios con 40, carnet con 38 y cuando con 37 aciertos.

Los menores valores se presentaron en donde con 21 aciertos, biblioteca con 31, y auditorio y bano con 33 aciertos cada una.

Los errores principales correspondieron a bano - biblioteca con 5 casos y biblioteca - bano con 4 casos. Además, se registraron errores menores de 1 a 2 casos entre las clases adios, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde. La matriz correspondiente se presenta en la Fig. 25.

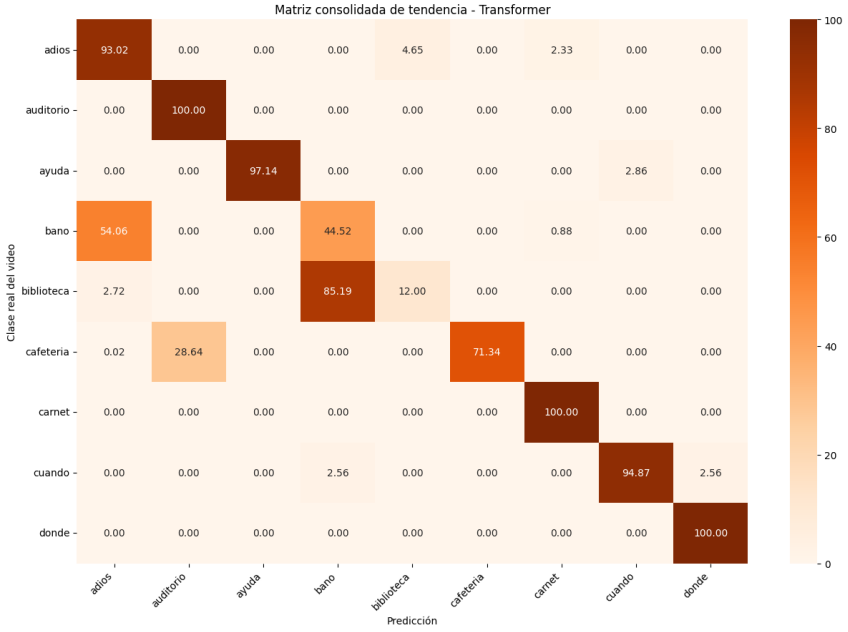


Fig. 25 Matriz de confusión Transformer.

Fuente: Elaboración propia.

Prueba de inferencia externa del modelo Transformer

La inferencia externa se realizó para la clase real cafeteria. El modelo clasificó correctamente el video como cafeteria, con una confianza de 71,34 %. El top 3 de predicciones fue: cafeteria con 71,34 %, auditorio con 28,64 % y adios con 0,02 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,2917, movimiento medio = 3,0615, margen = 0,4270 y tiempo total = 6,6495 s.

El resultado de la inferencia se presenta en la Fig. 26.

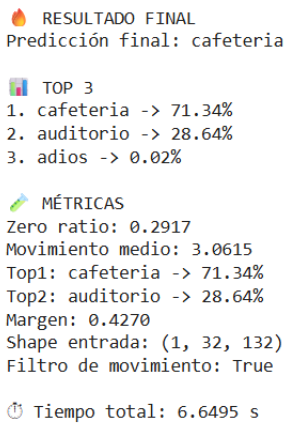


Fig. 26 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La siguiente inferencia externa se realizó con el archivo baño.mp4, correspondiente a la clase real bano. El modelo clasificó el video como adios, con una confianza de 54,06 %; por tanto, la predicción no coincidió con la etiqueta real. El top 3 de predicciones fue: adios con 54,06 %, bano con 44,52 % y carnet con 0,83 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,2064, movimiento medio = 2,0892, margen = 0,0954 y tiempo total = 4,9235 s.

El resultado de la inferencia se presenta en la Fig. 27.

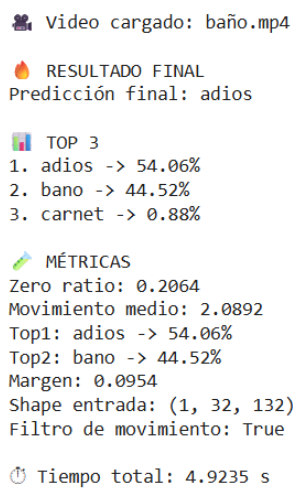


Fig. 27 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia externa se realizó con el archivo biblioteca.mp4, correspondiente a la clase real biblioteca. El modelo clasificó el video como bano, con una confianza de 85,19 %; por tanto, la predicción no coincidió con la etiqueta real. El top 3 de predicciones fue: bano con 85,19 %, biblioteca con 12,00 % y adios con 2,72 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,1780, movimiento medio = 2,2723, margen = 0,7319 y tiempo total = 6,2283 s.

El resultado de la inferencia se presenta en la Fig. 28.

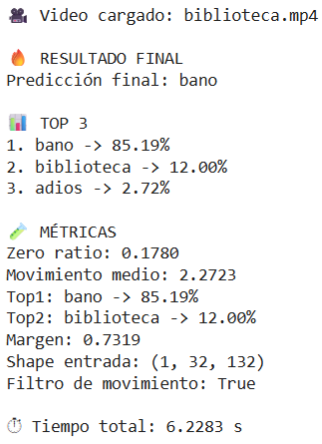


Fig. 28 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

Síntesis de resultados del modelo Transformer

El modelo Transformer registró valores altos en entrenamiento y validación, con una precisión de validación final entre 0,93 y 0,94. En inferencia externa, el archivo baño.mp4 fue clasificado como

adios, aunque su clase real era bano. De igual forma, el archivo biblioteca.mp4 fue clasificado como bano, aunque su clase real era biblioteca. Estos resultados muestran diferencias entre el desempeño observado en entrenamiento-validación y el comportamiento del modelo frente a videos externos.

4. Resultados del modelo Bi-LSTM

El modelo Bi-LSTM fue evaluado con una entrada 32 × 132, correspondiente a 32 fotogramas por video y 132 características por fotograma. Las clases evaluadas fueron: adios, auditorio, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde. Las etiquetas se mantuvieron sin tildes ni caracteres especiales durante el entrenamiento y la clasificación. La arquitectura implementada estuvo conformada por una capa de entrada, dos bloques Bi-LSTM, capas Dense y una capa final Softmax para la clasificación multiclase. La estructura general se presenta en la Fig. 29.

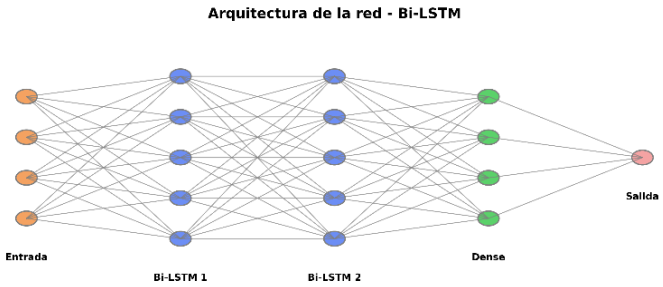


Fig. 29 Arquitectura general red Bi-LSTM.

Fuente: Elaboración propia.

La formulación matemática presentó el flujo del modelo desde la entrada de datos hasta la predicción final. En esta se definió como: *Ecu. 37* representa la entrada del modelo definida como $x \in \mathbb{R}^{32 \times 132}$, correspondiente a 32 fotogramas y 132 características por fotograma. La Ecu. 38 define la compuerta de entrada i_t, calculada a partir de la entrada x_t, el estado oculto anterior h_{t-1} , los pesos W_i, U_i y el sesgo b_i, utilizando la función sigmoide σ. La Ecu. 39 representa la compuerta de olvido ft , encargada de registrar la información del estado de celda anterior que se conserva. La Ecu. 40 define el estado candidato de la celda $\tilde{C}_t$, calculado mediante la función tangente hiperbólica $tanh$. La Ecu. 41 muestra la actualización del estado de celda C_t, combinando la información previa y la nueva mediante las compuertas f_t e i_t. La Ecu. 42 representa la compuerta de salida O_t, que regula la información enviada al estado oculto. La Ecu. 43 define el estado oculto h_t, obtenido a partir del estado de celda actualizado. La Ecu. 44 y la Ecu. 45

representan el procesamiento bidireccional de la secuencia mediante LSTM en dirección hacia adelante y hacia atrás. La Ecu. 46 define la salida bidireccional $h_t^{b_i}$, obtenida mediante la concatenación de los estados ocultos en ambas direcciones. La Ecu. 47 representa el promedio global temporal g, calculado sobre la secuencia completa. La Ecu. 48 define la capa densa, donde Zj corresponde a la combinación lineal de las características con los pesos W_{ij} y el sesgo b_j. La Ecu. 49 representa la función Softmax, utilizada para obtener la probabilidad $\hat{y}j_{\square}$ de cada clase. Finalmente, la Ecu. 50 define la función de pérdida de entropía cruzada categórica, utilizada durante el entrenamiento para calcular la diferencia entre las etiquetas reales y_j y las predicciones del modelo.

$$x \in \mathbb{R}^{32 \times 132} \qquad \textit{Ecu. 37}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i \qquad \textit{Ecu. 38}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_t \qquad \textit{Ecu. 39}$$

$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c \qquad \textit{Ecu. 40}$$

$$C_t = f_{t\odot} C_{t-1} + i_{t\odot} \tilde{C}_t \qquad \textit{Ecu. 41}$$

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o \qquad \textit{Ecu. 42}$$

$$h_t = O_{t\odot} tanh\,(C_{t)} \qquad \textit{Ecu. 43}$$

$$\vec{h}_t = LSTM_{toward}(x_{t)} \qquad \textit{Ecu. 44}$$

$$\overleftarrow{h}_t = LSTM_{toward}(x_{t)} \qquad \textit{Ecu. 45}$$

$$h_t^{b_i} = [\vec{h}_t : \overleftarrow{h}_t] \qquad \textit{Ecu. 46}$$

$$g = \frac{1}{T}\sum_{t=1}^{T} h_t^{b_i} \qquad \textit{Ecu. 47}$$

$$Zj = \sum_{i=1} W_{ij} X_i + b_j \qquad \textit{Ecu. 48}$$

$$\hat{y}j_{=} \frac{e^{z_j}}{\sum_{k-1}^{C} e^{z_k}} \qquad \textit{Ecu. 49}$$

$$\mathcal{L} = -\sum_{j-1}^{C} y_j log\ (\hat{y}j) \qquad \textit{Ecu. 50}$$

Métricas de precisión del modelo Bi-LSTM

El modelo Bi-LSTM fue entrenado durante 50 épocas con secuencias de entrada 32 × 132. La precisión de entrenamiento aumentó de 0,48 a 1,00. La precisión de validación inició en 0,64 y alcanzó un valor final de 0,98.

La curva de precisión se presenta en la Fig. 30.

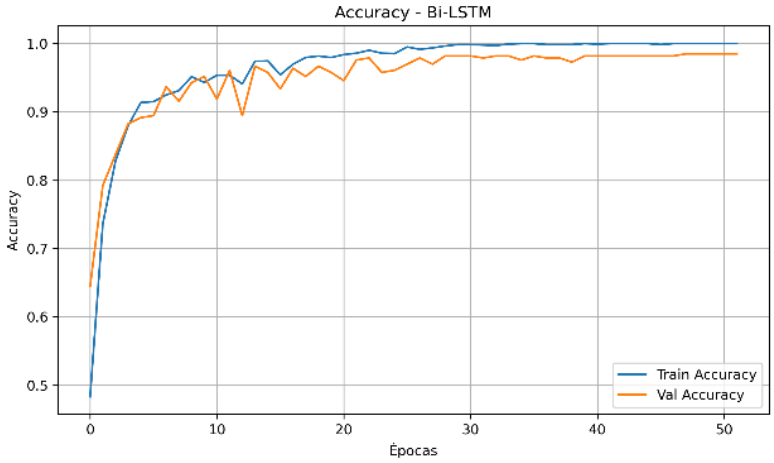


Fig. 30 Gráficas de precisión.

Fuente: Elaboración propia.

Métricas de pérdida del modelo Bi-LSTM

El modelo Bi-LSTM fue evaluado durante 50 épocas. La pérdida de entrenamiento inició en 1,57 y finalizó en 0,32. La pérdida de validación inició en 1,65 y finalizó en 0,35.

La curva de pérdida se presenta en la Fig. 31.

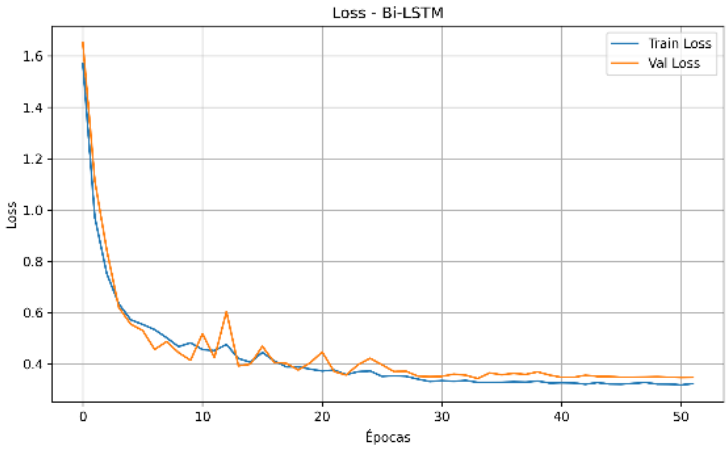


Fig. 31 Gráficas de perdida.

Fuente: Elaboración propia.

Matriz de confusión del modelo Bi-LSTM

La matriz de confusión registró aciertos en las nueve clases evaluadas. Los mayores valores se observaron en cafeteria con 46 aciertos, adios con 43, cuando con 39 y carnet con 38 aciertos. Los menores valores se presentaron en donde con 21 aciertos, auditorio con 33 y ayuda con 34 aciertos. Los errores registrados correspondieron a bano - biblioteca con 2 casos, bano - cuando con 1 caso, ayuda - donde con 1 caso, biblioteca - bano con 1 caso y biblioteca - cuando con 1 caso.

La matriz correspondiente se presenta en la Fig. 32.

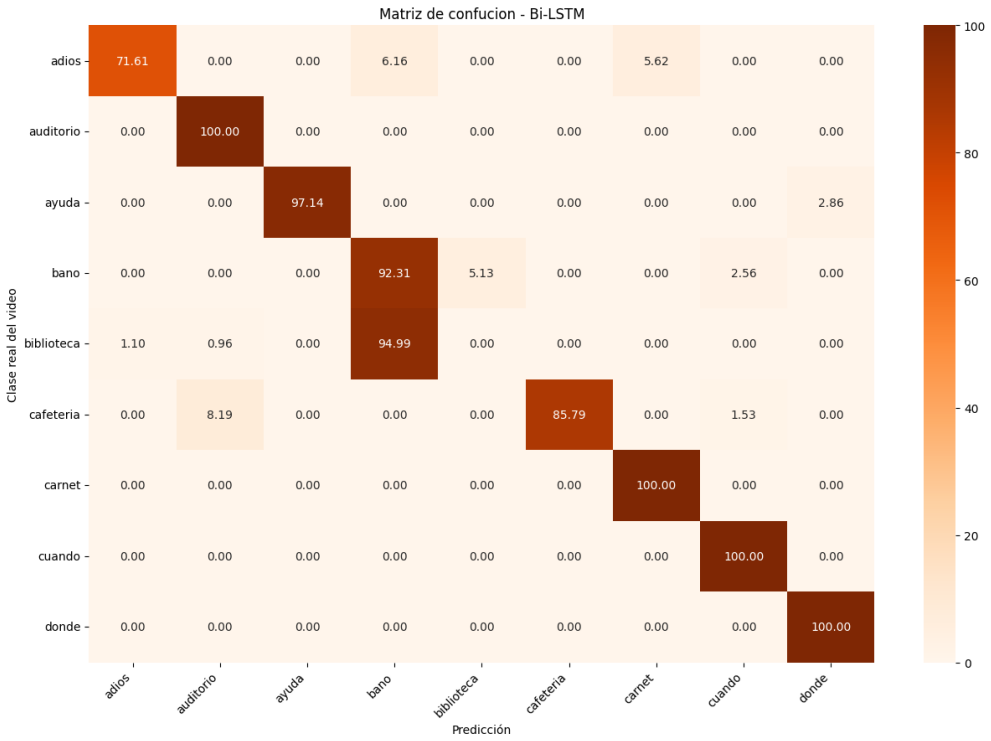


Fig. 32 Matriz de confusión Bi-LSTM.

Fuente: Elaboración propia.

Prueba de inferencia externa del modelo Bi-LSTM

La inferencia externa se realizó con el archivo adios.mp4, correspondiente a la clase real adios. El modelo clasificó correctamente el video como adios, con una confianza de 71,60 %. El top 3 de predicciones fue: adios con 71,60 %, bano con 6,16 % y carnet con 5,62 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,6136, movimiento medio = 2,5040, margen = 0,6545 y tiempo total = 5,9881 s.

El resultado de la inferencia se presenta en la Fig. 33.

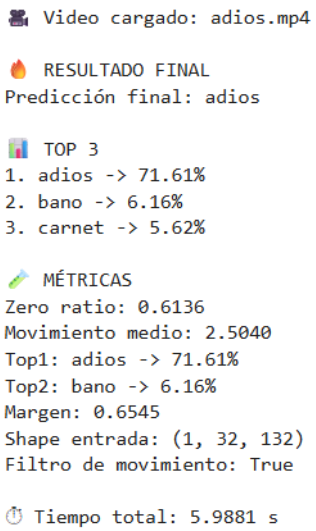


Fig. 33 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia externa se realizó con el archivo biblioteca.mp4, correspondiente a la clase real biblioteca. El modelo clasificó el video como bano, con una confianza de 85,19 %; por tanto, la predicción no coincidió con la etiqueta real. El top 3 de predicciones fue: bano con 85,19 %, biblioteca con 12,00 % y adios con 2,72 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,1780, movimiento medio = 2,2723, margen = 0,7319 y tiempo total = 6,2283 s.

El resultado de la inferencia se presenta en la Fig. 34**.**

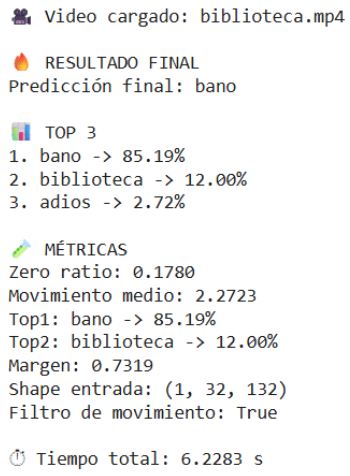


Fig. 34 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia externa se realizó con el archivo cafeteria.mp4, correspondiente a la clase real cafeteria. El modelo clasificó correctamente el video como cafeteria, con una confianza de 85,79

%. El top 3 de predicciones fue: cafeteria con 85,79 %, auditorio con 8,19 % y cuando con 1,53 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,2917, movimiento medio = 3,0615, margen = 0,7761 y tiempo total = 6,5989 s.

El resultado de la inferencia se presenta en la Fig. 35.

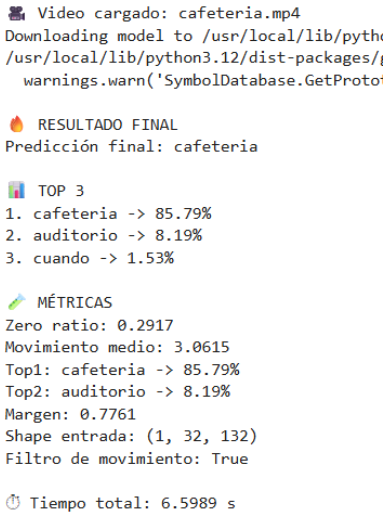


Fig. 35 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

Síntesis de resultados del modelo Bi-LSTM

El modelo Bi-LSTM registró valores altos durante el entrenamiento y la validación, con una precisión de validación final de 0,98. En las pruebas de inferencia externa, los archivos adios.mp4 y cafeteria.mp4 fueron clasificados correctamente. El archivo biblioteca.mp4, correspondiente a la clase real biblioteca, fue clasificado como bano. Estos resultados registraron un comportamiento favorable en dos de las tres pruebas externas documentadas; sin embargo, se mantuvo una confusión entre las clases biblioteca y bano.

5. Resultados del modelo LSTM + Attention estable

El modelo LSTM + Attention estable fue evaluado con una entrada 32 × 132, correspondiente a 32 fotogramas por video y 132 características por fotograma. Las clases evaluadas fueron: adios, auditorio, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde. Las etiquetas se mantuvieron sin tildes ni caracteres especiales durante el entrenamiento y la clasificación.

La arquitectura implementada estuvo conformada por una capa de entrada, dos capas LSTM bidireccionales, capas de normalización, una capa Attention, capas Dense y una capa final Softmax para la clasificación multiclase. La estructura general se presenta en la Fig. 36.

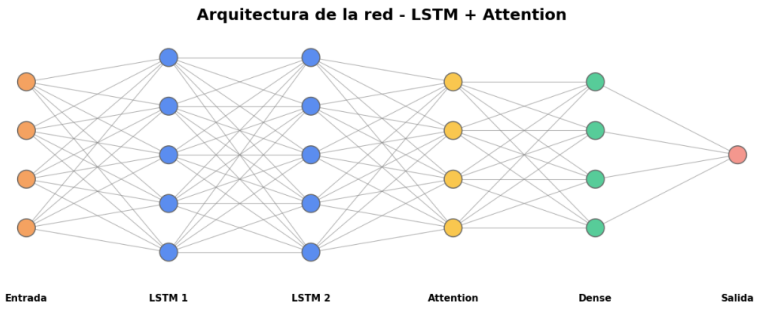


Fig. 36 Arquitectura general de LSTM + Attention.

Fuente: Elaboración propia.

El modelo LSTM + Attention estable fue seleccionado para la integración final del aplicativo web, con base en los resultados obtenidos durante el entrenamiento, la validación y las pruebas externas de inferencia documentadas.

Resumen de arquitectura del modelo LSTM + Attention

El modelo LSTM + Attention estable presentó una entrada (None, 32, 132) y una salida (None, 9), correspondiente a las nueve clases evaluadas. La primera capa BiLSTM generó una salida (None, 32, 256) con 267.264 parámetros, seguida de una capa LayerNormalization con 512 parámetros. La segunda capa BiLSTM presentó una salida (None, 32, 128) con 164.352 parámetros. Posteriormente, la segunda capa LayerNormalization registró 256 parámetros.

La capa AttentionLayer generó una salida (None, 128) con 129 parámetros. Luego, las capas Dense registraron salidas (None, 128) y (None, 64), con 16.512 y 8.256 parámetros, respectivamente. La capa final Softmax generó una salida (None, 9) con 585 parámetros.

En total, el modelo registró 457.866 parámetros entrenables. La distribución de capas y parámetros se presenta en la Fig. 37.

Model: "lstm_attention_estable"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 32, 132)	0
bilstm_1 (Bidirectional)	(None, 32, 256)	267,264
layer_norm_1 (LayerNormalization)	(None, 32, 256)	512
bilstm_2 (Bidirectional)	(None, 32, 128)	164,352
layer_norm_2 (LayerNormalization)	(None, 32, 128)	256
attention_layer (AttentionLayer)	(None, 128)	129
dense_1 (Dense)	(None, 128)	16,512
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
output_layer (Dense)	(None, 9)	585

Total params: 457,866 (1.75 MB)
Trainable params: 457,866 (1.75 MB)
Non-trainable params: 0 (0.00 B)

Fig. 37 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La formulación matemática presentó el flujo del modelo desde la entrada de datos hasta la predicción final. En esta se definió la entrada como Ecu. 51 representa la entrada del modelo definida como $x \in \mathbb{R}^{32 \times 132}$, correspondiente a 32 fotogramas y 132 características por fotograma. La Ecu. 52 define la compuerta de entrada i_t, calculada a partir de la entrada x_t, el estado oculto anterior h_{t-1}, los pesos W_i ,e U_i y el sesgo b_i, mediante la función sigmoide. La Ecu. 53 representa la compuerta de olvido f_t, encargada de registrar la información del estado de celda anterior que se conserva. La Ecu. 54 define el estado candidato de la celda $\tilde{C}_t$, calculado mediante la función tangente hiperbólica. La Ecu. 55 muestra la actualización del estado de celda C_t, combinando la información previa y la nueva mediante las compuertas ft e i_t. La Ecu. 56 representa la compuerta de salidaO_t, que regula la información enviada al estado oculto. La Ecu. 57 define el estado oculto h_t, obtenido a partir del estado de celda actualizado. La Ecu. 58 representa el cálculo del puntaje de atención e_t, obtenido a partir del estado oculto h_t mediante una transformación lineal con la función tanh. La Ecu. 59 define el peso de atención α_t, calculado mediante una normalización Softmax sobre los puntajes e_t. La Ecu. 60 representa la aplicación del peso de atención sobre el estado oculto, obteniendo la representación ponderada r_t. La Ecu. 61 define el promedio global temporal g, calculado a partir de la suma de las representaciones ponderadas a lo largo de la secuencia. La Ecu. 62 representa la capa densa, donde Z_j corresponde a la combinación lineal de las características con los pesos W_{ij} y el sesgo b_j. La Ecu. 63 define la función Softmax, utilizada para obtener la probabilidad $\hat{y}_j$ asociada a cada clase. Finalmente, la

Ecu. 64 representa la función de pérdida de entropía cruzada categórica, utilizada durante el entrenamiento para calcular la diferencia entre las etiquetas reales y_j y las predicciones del modelo.

$$x \in \mathbb{R}^{32 \times 132} \quad \textit{Ecu. 51}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i \quad \textit{Ecu. 52}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_t \quad \textit{Ecu. 53}$$

$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c \quad \textit{Ecu. 54}$$

$$C_t = f_{t\odot} C_{t-1} + i_{t\odot} \tilde{C}_t \quad \textit{Ecu. 55}$$

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad \textit{Ecu. 56}$$

$$h_t = O_{t\odot} tanh\,(C_t) \quad \textit{Ecu. 57}$$

$$e_t = tanh\,(W_a h_t + b_a \quad \textit{Ecu. 58}$$

$$\alpha_t = \frac{exp\,(e_t)}{\sum_{k-1}^{T} exp\,(e_k)} \quad \textit{Ecu. 59}$$

$$r_t = \alpha_t h_t \quad \textit{Ecu. 60}$$

$$g = \frac{1}{T}\sum_{t=1}^{T} r_t \quad \textit{Ecu. 61}$$

$$Z_j = \sum_i W_{ij} g_i + b_j \quad \textit{Ecu. 62}$$

$$\hat{y}_j = \frac{exp\,(Z_j)}{\sum_{k-1}^{C} exp\,(Z_k)} \quad \textit{Ecu. 63}$$

$$\mathcal{L} = -\sum_{j=1}^{C} y_j log\,(\hat{y}_j) \quad \textit{Ecu. 64}$$

Métricas de precisión del modelo LSTM + Attention

El modelo LSTM + Attention estable fue entrenado durante 50 épocas con secuencias de entrada 32 × 132. La precisión de entrenamiento aumentó de 0,20 a 1,00. La precisión de validación inició en 0,56 y alcanzó valores finales entre 0,98 y 0,99.

La curva de precisión se presenta en la Fig. 38.

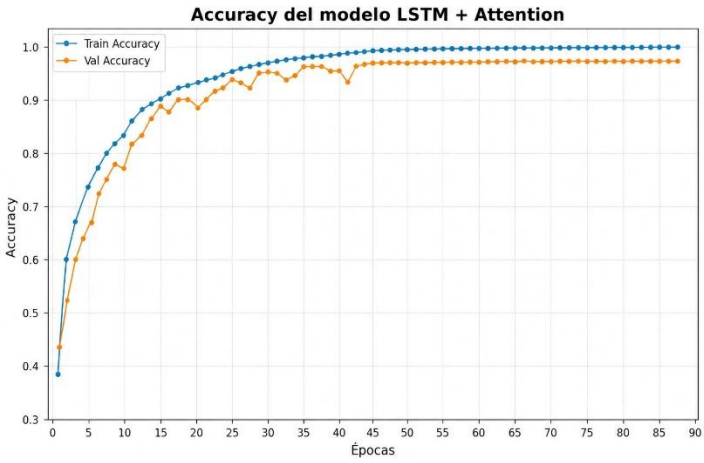


Fig. 38 Gráfica de precisión.

Fuente: Elaboración propia.

Métricas de pérdida del modelo LSTM + Attention

El modelo LSTM + Attention estable fue evaluado durante 50 épocas. La pérdida de entrenamiento inició en 2,22 y finalizó en 0,31. La pérdida de validación inició en 1,65 y finalizó en 0,34.

La curva de pérdida se presenta en la Fig. 39.

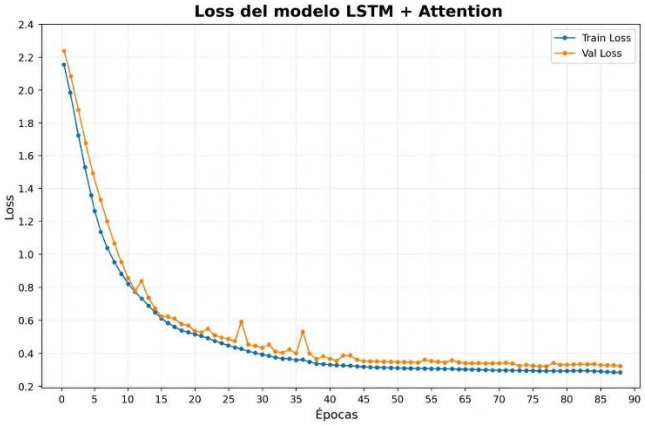


Fig. 39 Gráfica de perdida.

Fuente: Elaboración propia.

Matriz de confusión filtrada por confianza del modelo LSTM + Attention

La matriz de confusión filtrada por confianza del modelo LSTM + Attention evidencio un alto nivel de precision en las nueve clases evaluadas. Los mayores porcentajes de clasificación correcta

se observaron en adios, bano y cafeteria, con valores cercanos al 98%, seguidos de cuando 97.44%, biblioteca 97.37% y donde 97.29%.

Por su parte, ayuda y auditorio presentaron valores de 97.14% y 96.53% respectivamente, manteniendo un desempeño alto y consistente. La clase con menor porcentaje de acierto fue carnet con 90.45%, aunque sigue representando un nivel adecuado de clasificación.

En cuanto a los errores, estos fueron mínimos y se concentraron en confusiones puntuales entre clases semántica y visualmente similares, como biblioteca y cuando ~2.56%, así como ayuda y donde ~2.86%. En general, la baja dispersión de errores confirma la robustez del modelo en condiciones de inferencia con umbral de confianza.

La matriz correspondiente se presenta en la Fig. 40.

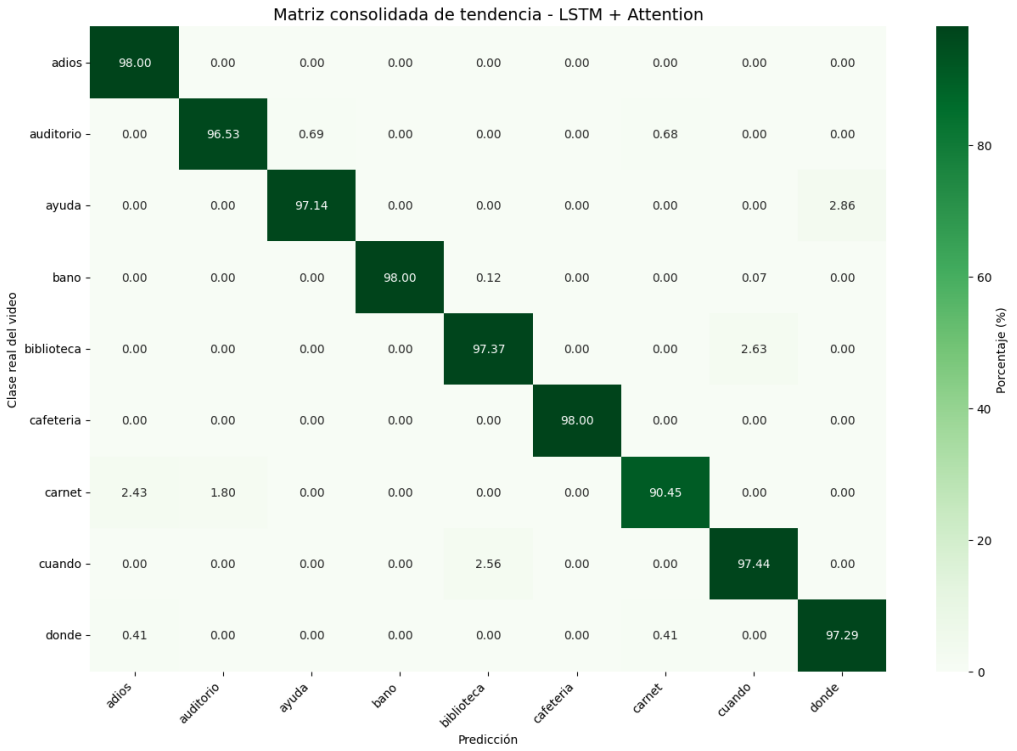


Fig. 40 Matriz de inferencia interna.

Fuente: Elaboración propia.

La Fig. 41 presenta la matriz de confusión probabilística obtenida mediante pruebas de inferencia externa. Esta matriz considera las tres clases con mayor probabilidad por muestra, lo que permite observar la predicción principal y la distribución de probabilidad hacia clases secundarias.

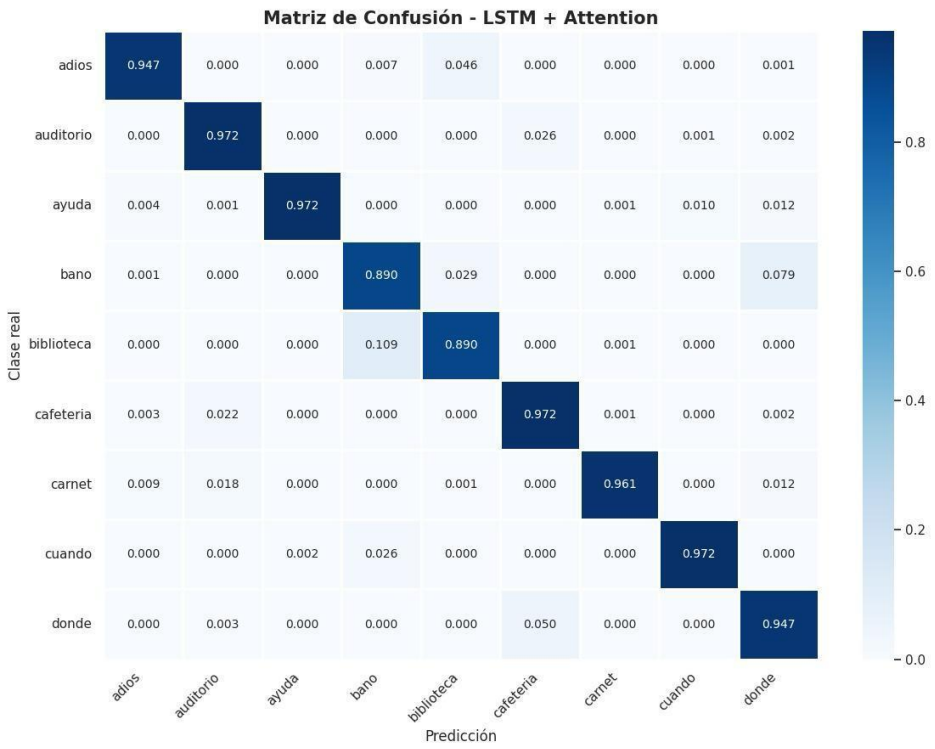


Fig. 41 Matriz de inferencia externa.
Fuente: Elaboración propia.

En la inferencia externa según la Fig. 41, la mayoría de clases presentó porcentajes altos, con valores entre 89 % y 97.2 %. Las principales desviaciones se observaron en clases con similitud gestual, especialmente en relaciones como biblioteca - bano, bano - donde, ayuda - donde y carnet - adios/auditorio. Estos resultados indican que el modelo mantuvo un desempeño favorable en pruebas externas de campo. No obstante, algunas clases registraron probabilidades secundarias asociadas a variaciones en la ejecución de la seña y similitud entre patrones gestuales.

La comparación entre la matriz de entrenamiento Fig. 40 y la matriz de inferencia externa Fig. 41 se evidenció un comportamiento estable del modelo LSTM + Attention. En ambos casos se registraron valores predominantes en la diagonal principal, indicando correspondencia entre la clase real y la predicción generada.

Prueba de inferencia interna del modelo LSTM + Attention

La inferencia interna se realizó con el archivo baño.mp4, correspondiente a la clase real bano. El modelo clasificó correctamente el video como bano, con una confianza de 99,50 %. El top 3 de predicciones fue: bano con 99,50 %, biblioteca con 0,12 % y cuando con 0,07 %. La entrada

procesada presentó una forma (1, 32, 132), con zero ratio = 0,2064, movimiento medio = 2,0892, margen = 0,997, filtro de movimiento = True, normalización de entrenamiento = True y tiempo total = 5,6176 s.

El resultado de la inferencia se presenta en la Fig. 42.

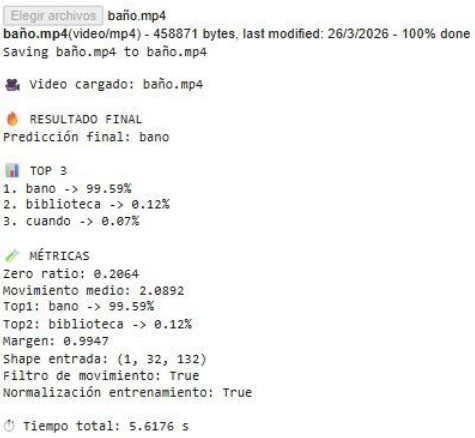


Fig. 42 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia interna se realizó con el archivo carnet.mp4, correspondiente a la clase real carnet. El modelo clasificó correctamente el video como carnet, con una confianza de 90,45 %. El top 3 de predicciones fue: carnet con 90,45 %, adios con 2,43 % y auditorio con 1,80 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,6989, movimiento medio = 1,8561, margen = 0,8801, filtro de movimiento = True, normalización de entrenamiento = True y tiempo total = 4,6256 s.

El resultado de la inferencia se presenta en la Fig. 43.

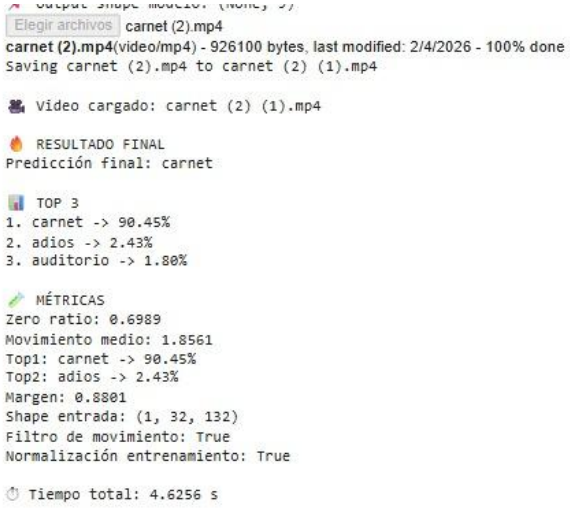


Fig. 43 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia interna se realizó con el archivo auditorio.mp4, correspondiente a la clase real auditorio. El modelo clasificó correctamente el video como auditorio, con una confianza de 96,53 %. El top 3 de predicciones fue: auditorio con 96,53 %, ayuda con 0,69 % y carnet con 0,68 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,1591, movimiento medio = 2,7000, margen = 0,9584, filtro de movimiento = True, normalización de entrenamiento = True y tiempo total = 6,1928 s.

El resultado de la inferencia se presenta en la Fig. 44**.**

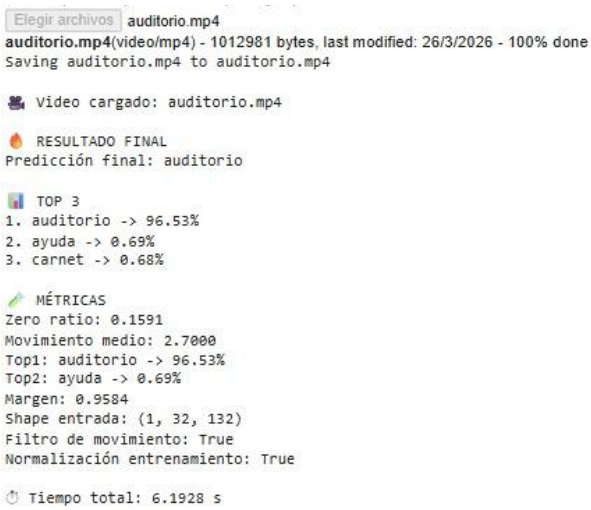


Fig. 44 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia interna se realizó con el archivo donde.mp4, correspondiente a la clase real donde. El modelo clasificó correctamente el video como donde, con una confianza de 97,29 %. El top 3 de predicciones fue: donde con 97,29 %, adios con 0,41 % y carnet con 0,41 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,3390, movimiento medio = 3,7046, margen = 0,9688, filtro de movimiento = True, normalización de entrenamiento = True y tiempo total = 5,0422 s.

El resultado de la inferencia se presenta en la Fig. 45.

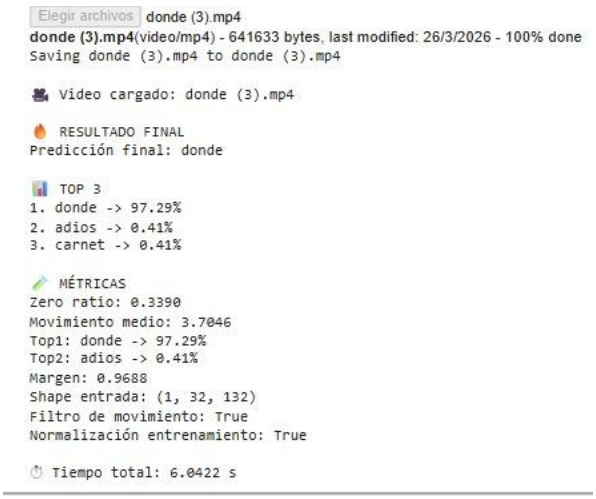


Fig. 45 Evidencia gráfica del proceso.

Fuente: Elaboración propia.

La inferencia interna se realizó con el archivo adiós (14).MP4, correspondiente a la clase real adios. El modelo clasificó correctamente el video como adios, con una confianza de 96,03 %. El top 3 de predicciones fue: adios con 96,03 %, cafeteria con 1,02 % y ayuda con 0,58 %. La entrada procesada presentó una forma (1, 32, 132), con zero ratio = 0,0938, movimiento medio = 3,1270, margen = 0,9500, filtro de movimiento = True y normalización de entrenamiento = True.

El resultado de la inferencia se presenta en la Fig. 46.

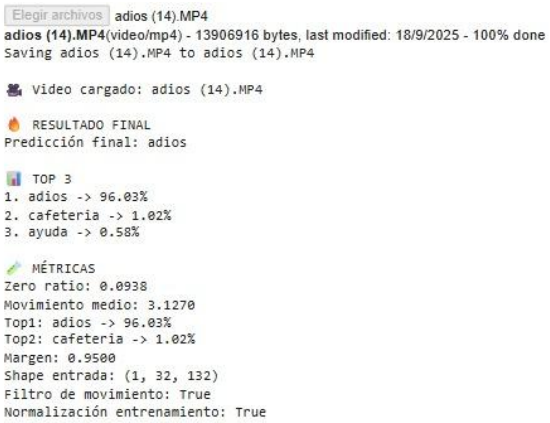


Fig. 46 Salida de inferencia del modelo para adios.

Fuente: Elaboración propia.

Prueba de inferencia externa del modelo LSTM + Attention

La inferencia externa se realizó sobre el conjunto completo de clases consideradas en el modelo, con el fin de ilustrar el comportamiento, se presentan a continuación tres casos representativos correspondientes a las clases carnet, adios y cafeteria. El resto de inferencias realizadas se encuentran consignadas en el **Anexo. 1** .

En el primer caso, se utilizó un video correspondiente a la clase real carnet. El modelo clasificó correctamente la muestra como carnet, con una probabilidad de 94,93 %. El top 3 de predicciones fue: carnet con 94,93 %, auditorio con 2,63 % y donde con 0,90 %. La entrada procesada presentó una forma (1, 32, 132), con 73 fotogramas leídos, zero ratio = 0,0597 y desviación estándar = 0,7428.

El resultado de esta inferencia se presenta en la Fig. 47.

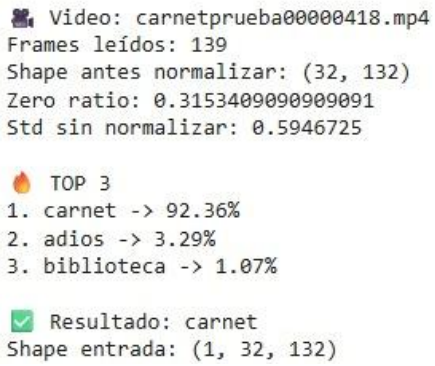


Fig. 47 Salida de inferencia del modelo para carnet

Fuente: Elaboración propia.

En el segundo caso, se utilizó un video correspondiente a la clase real adios. El modelo clasificó correctamente la muestra como adios, con una probabilidad de 92,94 %. El top 3 de predicciones fue: adios con 92,94 %, biblioteca con 5,36 % y bano con 0,52 %. La entrada procesada presentó una forma (1, 32, 132), con 215 fotogramas leídos, zero ratio = 0,0597 y desviación estándar = 0,7460.

El resultado de esta inferencia se presenta en la Fig. 48.

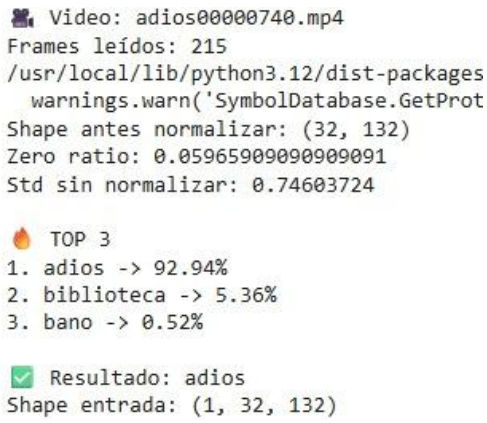


Fig. 48 Salida de inferencia del modelo para adios

Fuente: Elaboración propia.

En el tercer caso, se utilizó un video correspondiente a la clase real cafeteria. El modelo clasificó correctamente la muestra como cafeteria, con una probabilidad de 98,83 %. El top 3 de predicciones fue: cafeteria con 98,83 %, auditorio con 0,56 % y bano con 0,15 %. La entrada procesada presentó una forma (1, 32, 132), con 213 fotogramas leídos, zero ratio = 0,0881 y desviación estándar = 0,9015.

El resultado de esta inferencia se presenta en la Fig. 49.

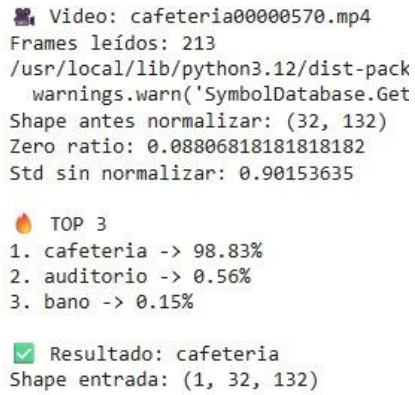


Fig. 49 Salida de inferencia del modelo para cafetería

Fuente: Elaboración propia.

El modelo LSTM + Attention estable registró valores altos en entrenamiento, validación, matriz de confusión filtrada por confianza, reporte de clasificación e inferencia externa. Las pruebas con los archivos baño.mp4, carnet.mp4, auditorio.mp4, donde.mp4 y adiós (14).MP4 generaron predicciones correspondientes con la clase real de cada video. Con base en estos resultados, el modelo LSTM + Attention estable fue seleccionado para la integración final del aplicativo web.

En la TABLA. VII se presenta la comparación de los modelos evaluados, incluyendo sus métricas de desempeño, confusiones entre clases y comportamiento en pruebas externas.

TABLA. VII
TABLA COMPARATIVA PARA SELECCION DE MODELO

Modelo	Entrada	Clases	Acc. val.	Loss val.	Clases con confusión	Errores en inferencia externa	Observaciones	Decisión
CNN 1D Baseline	32×132	9	0.92–0.94	~0.18–0.22	Confusión entre baño–biblioteca; biblioteca–adiós; biblioteca–baño; ayuda–donde; cuando–baño	Confusión entre baño–adiós	No modela bien dependencias temporales	No seleccionado
Transformer	32×132	9	0.95–0.96	~0.12–0.18	Confusión entre adiós–biblioteca; adiós–carnet; ayuda–cuando; baño–biblioteca; baño–cuando; biblioteca–adiós; biblioteca–baño; biblioteca–cuando; biblioteca–donde; cafetería–adiós; cuando–baño; cuando–donde	Confusión entre baño–adiós; biblioteca–baño	Mayor capacidad, pero sensible a regularización	No seleccionado
Bi-LSTM	32×132	9	0.96–0.97	~0.23–0.25	Confusión entre ayuda–donde; baño–biblioteca; baño–cuando; biblioteca–baño; biblioteca–cuando	Confusión entre biblioteca–baño	Captura secuencias, pero sin enfoque selectivo	No seleccionado
LSTM + Attention	32×132	9	0.96–0.97	~0.23–0.25	No se registraron confusiones en las pruebas realizadas	No se observaron errores en las inferencias presentadas	Mejor equilibrio entre precisión y generalización	**Seleccionado**

Fuente: Elaboración propia

Confianza de la predicción

La confianza en las predicciones corresponde a la probabilidad asignada por el modelo a la clase seleccionada, obtenida a partir de la función softmax en la capa de salida. Esta se calcula como:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \quad Ecu.\ 65$$

donde z_i representa la salida del modelo para la clase i y $P(y_i)$ la probabilidad asociada a dicha clase. Este valor indica el grado de certeza del modelo en cada predicción individual.

En las inferencias realizadas, se registraron valores de confianza superiores al 90 % en la clase predicha, lo que evidencia una alta seguridad del modelo en la asignación de etiquetas. No obstante, estos valores no garantizan la correcta clasificación en todos los casos, debido a la existencia de señas con similitud gestual cuyos movimientos pueden resultar parcialmente indistinguibles para el modelo.

Análisis comparativo de modelos

A partir de los resultados presentados en la TABLA. VII, se observa que los modelos evaluados presentan diferencias en términos de precisión, pérdida, capacidad de discriminación y comportamiento en inferencias externas.

El modelo CNN 1D Baseline presenta el menor desempeño, evidenciando limitaciones en el modelado de dependencias temporales y registrando múltiples confusiones entre clases.

El modelo Transformer alcanza valores altos de precisión; sin embargo, presenta mayor pérdida y un número considerable de confusiones, lo que afecta su capacidad de generalización en este contexto.

El modelo Bi-LSTM muestra un mejor equilibrio entre precisión y generalización, reduciendo el número de confusiones, aunque aún presenta errores en clases con similitud gestual.

En contraste, el modelo LSTM + Attention presenta el mejor desempeño global, con altos valores de precisión y un comportamiento estable en términos de pérdida dentro del rango observado en las arquitecturas evaluadas. Adicionalmente, no se registraron confusiones en las pruebas ni errores

en las inferencias externas, lo que evidencia una adecuada capacidad de discriminación y generalización. En consecuencia, el modelo LSTM + Attention es seleccionado como la mejor alternativa, al cumplir de manera más consistente con los criterios evaluados, incluyendo precisión, estabilidad, reducción de errores y comportamiento en inferencias externas.

6. Resultados de las extracciones evaluadas en backend

Para seleccionar la configuración de extracción más adecuada para el sistema de reconocimiento de Lengua de Señas Colombiana, se evaluaron tres versiones del pipeline experimental: PIXEL, RED-NOVA y LSC-V1. Cada configuración presentó diferencias en la estructura de entrada, el número de características procesadas y el comportamiento obtenido durante la inferencia. Esta sección registra los resultados de cada configuración, la comparación general entre ellas y los indicadores reportados durante la inferencia. El análisis e interpretación de estos resultados se desarrolla en el capítulo siguiente.

6.1 Resultados obtenidos con PIXEL

La configuración PIXEL fue una de las primeras versiones funcionales del pipeline de extracción. Esta versión utilizó una entrada de 225 características por frame, construida a partir de la pose corporal completa y ambas manos. El tamaño del vector se ajustó mediante recorte o *padding* para cumplir con la forma de entrada requerida por el modelo. Además, incorporó una capa de atención personalizada y una función de velocidad temporal (*temporal_velocity*), por lo que correspondió a una arquitectura experimental inicial.

En las pruebas de inferencia, PIXEL permitió verificar la viabilidad técnica del procesamiento; sin embargo, presentó predicciones incorrectas en las muestras evaluadas. Para la seña auditorio, el sistema generó la predicción donde en 9,464 segundos, como se muestra en la Fig. 50**.** Para la seña bano, la predicción fue adios en 5,820 segundos, como se observa en la Fig. 51**.** Finalmente, para la seña cuando, el sistema generó la predicción si en 4,717 segundos, como se presenta en la Fig. 52**.**

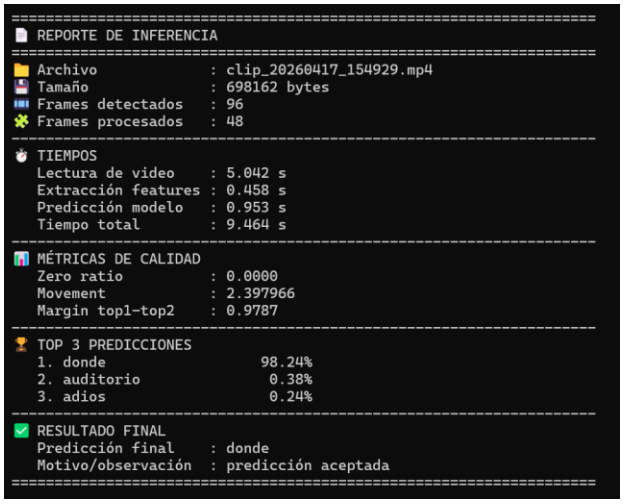


Fig. 50 Resultados de inferencia en back.

Fuente: Elaboración propia

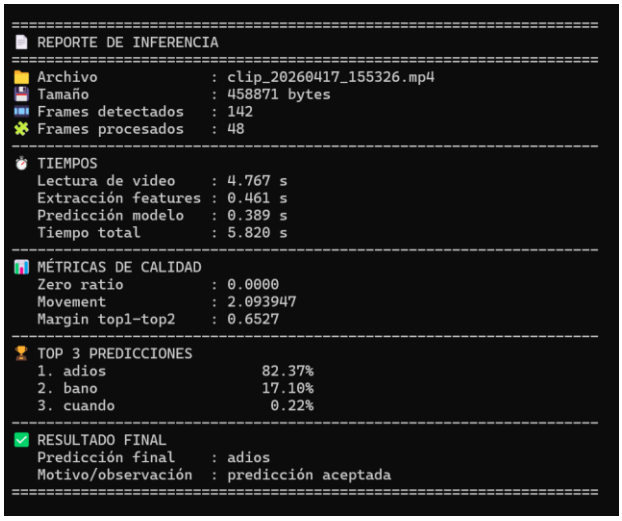


Fig. 51 Resultados de inferencia en back.

Fuente: Elaboración propia

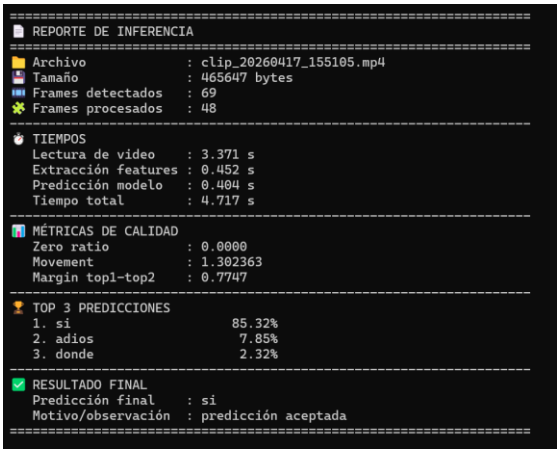


Fig. 52 Resultados de inferencia en back.

Fuente: Elaboración propia

Estos resultados evidenciaron que PIXEL logró ejecutar el flujo completo de inferencia, pero presentó baja estabilidad en la clasificación. Por esta razón, se tomó como una versión base para las mejoras posteriores del sistema.

6.2 Resultados obtenidos con RED-NOVA

La configuración RED-NOVA correspondió a una etapa intermedia del pipeline de extracción. Esta versión trabajó con secuencias de 48 frames y una entrada de 375 características por frame, construida a partir de la pose corporal completa, ambas manos y un bloque facial simulado en ceros, utilizado para conservar la estructura de entrada requerida por el modelo.

En las pruebas de inferencia, RED-NOVA presentó una mejora frente a PIXEL, aunque todavía registró errores y tiempos de procesamiento variables. Para la seña **amigo**, el sistema generó la predicción **hola** en 20,739 segundos, como se muestra en la Fig. 53**.** Para la seña **ayuda**, la predicción fue **gracias** en 2,610 segundos, como se observa en la Fig. 54**.** En el caso de la seña **cafeteria**, el sistema generó una predicción correcta, aunque con baja confianza, en 14,833 segundos, como se presenta en la Fig. 55.

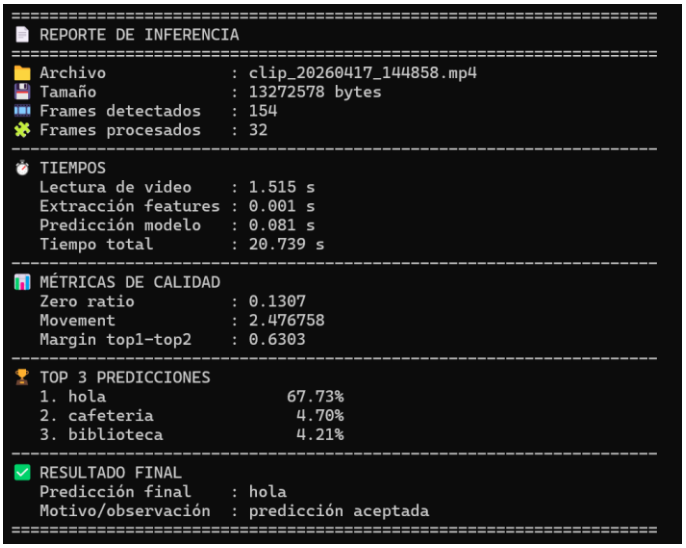


Fig. 53 Resultados de inferencia en back.

Fuente: Elaboración propia

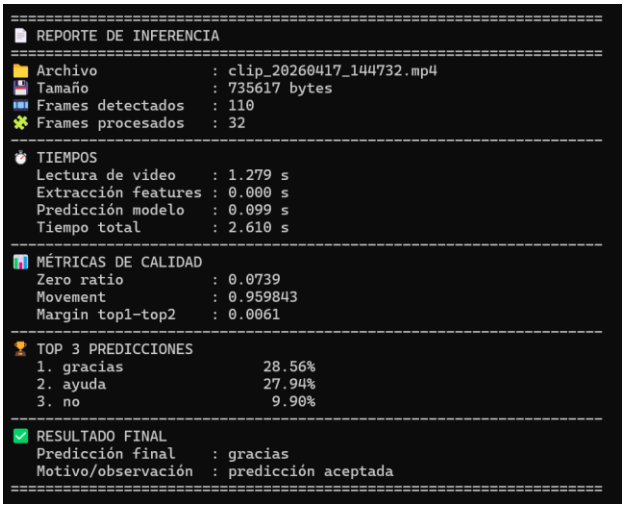


Fig. 54 Resultados de inferencia en back.

Fuente: Elaboración propia

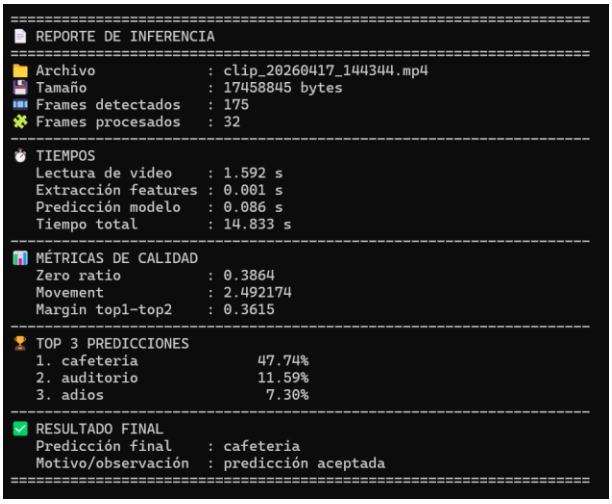


Fig. 55 Resultados de inferencia en back.

Fuente: Elaboración propia

. Estos resultados permitieron identificar una mejora parcial en la estabilidad del reconocimiento respecto a la configuración inicial. Sin embargo, la presencia de predicciones incorrectas y tiempos de inferencia elevados indicó la necesidad de continuar con el ajuste del pipeline hasta consolidar la versión final.

6.3 Resultados obtenidos con LSC-V1

La configuración LSC-V1 presentó el mejor desempeño general en las pruebas de inferencia. Esta versión trabajó con secuencias de 32 frames y una entrada de 132 características por frame,

construida a partir de los hombros como referencia corporal mínima y ambas manos como fuente principal de información gestual.

Esta configuración redujo la dimensionalidad de entrada e incorporó criterios de validación de la muestra, entre ellos *Zero ratio*, nivel de movimiento, umbral de confianza y margen entre las dos clases con mayor probabilidad. Estos indicadores permitieron evaluar la calidad de la secuencia procesada antes de aceptar la predicción final.

En las pruebas realizadas, LSC-V1 obtuvo predicciones correctas con tiempos de respuesta menores frente a las configuraciones anteriores. Para la seña adios, el sistema generó una interpretación correcta en 2,481 segundos, como se muestra en la Fig. 56**.** Para la seña donde, la predicción también fue correcta en 2,495 segundos, como se observa en la Fig. 57**.** Finalmente, para la seña ayuda, el sistema obtuvo una interpretación correcta en 2,380 segundos, como se presenta en la Fig. 58.

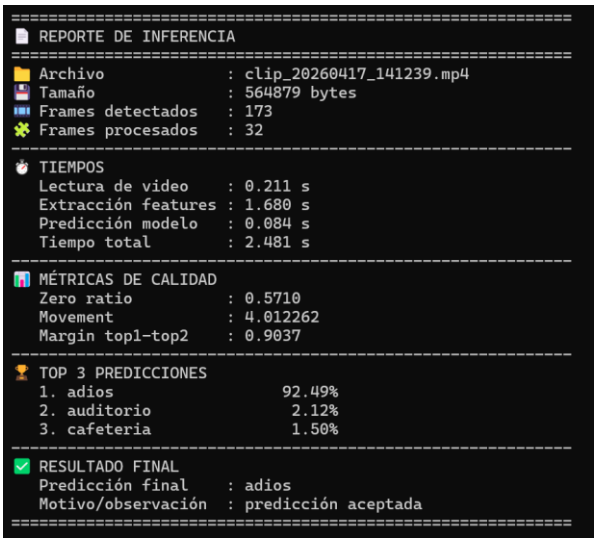


Fig. 56 Resultados de inferencia en back.

Fuente: Elaboración propia

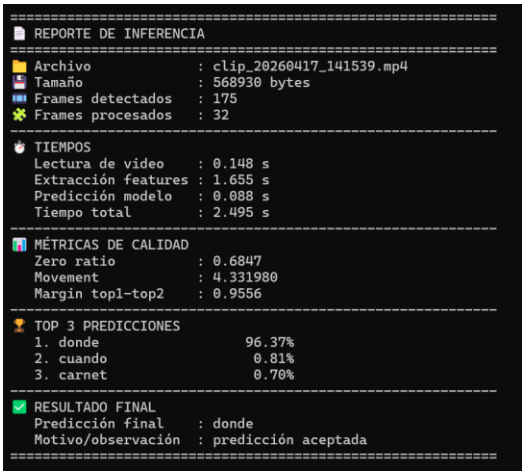


Fig. 57 Resultados de inferencia en back.

Fuente: Elaboración propia

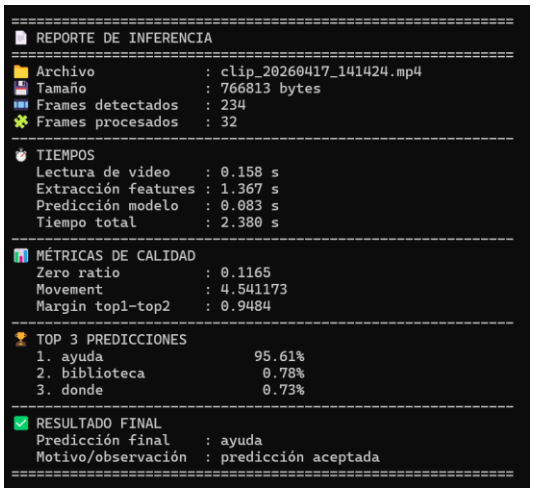


Fig. 58 Resultados de inferencia en back.

Fuente: Elaboración propia

Estos resultados permitieron seleccionar LSC-V1 como la configuración final del pipeline de extracción y como base para la validación del sistema.

6.4 Comparación general de resultados entre las tres configuraciones

La comparación entre las tres configuraciones permitió identificar la evolución del pipeline de extracción. PIXEL correspondió a la primera versión funcional, con una estructura de entrada amplia y resultados iniciales de inferencia. RED-NOVA representó una etapa intermedia, con mayor cantidad de características por frame y una mejora parcial en el reconocimiento. Finalmente, LSC-V1 presentó la estructura más compacta y el mejor comportamiento general durante las pruebas.

Como se observa en la TABLA. VIII, la evolución del sistema pasó de configuraciones con mayor dimensionalidad de entrada hacia una versión más controlada, basada en 32 frames y 132 características por frame. Esta reducción permitió consolidar una configuración más adecuada para la integración y validación final del aplicativo.

TABLA. VIII
COMPARACIÓN DE CONFIGURACIONES DE EXTRACCIÓN EVALUADAS.

Modelo	Secuencia	Características por frame	Resultado general observado
PIXEL	32/48 según pruebas	225	Primera etapa funcional
RED-NOVA	48	375	Segunda mejor configuración
LSC-V1	32	132	Mejor desempeño general

Fuente: Elaboración propia.

La TABLA. VIII resume las diferencias principales entre las configuraciones evaluadas, considerando la secuencia de entrada, el número de características por frame y el resultado general observado durante las pruebas.

6.5 Descripción de los indicadores del reporte de inferencia

Para complementar los resultados, se implementó un reporte de inferencia con los principales indicadores generados durante el procesamiento de cada video. Este reporte permitió registrar información técnica de cada prueba y comparar el comportamiento de las configuraciones evaluadas. El campo Archivo identifica el video procesado. Tamaño registra el peso del archivo. Frames detectados indica la cantidad total de fotogramas identificados en el video original, mientras que Frames procesados señala los fotogramas utilizados para construir la secuencia de entrada del modelo.

El apartado Tiempos registra la duración de las etapas de lectura del video, extracción de características, predicción del modelo y tiempo total de inferencia. Esta información permitió identificar la distribución del procesamiento en cada prueba. En las métricas de calidad, Zero ratio representa la proporción de valores en cero dentro de la secuencia de características. Un valor elevado indica menor cantidad de información útil capturada, asociada principalmente a fallos en la detección de landmarks o a baja calidad de la muestra.

La métrica movement registra la variación entre frames consecutivos y permite verificar si la muestra contiene movimiento gestual suficiente. Por su parte, margin top1-top2 indica la diferencia entre la probabilidad de la clase principal y la segunda clase más probable, aportando información sobre la separación entre predicciones.

Finalmente, el reporte incluye el Top 3 de predicciones, el Resultado final y el campo Motivo/observación, los cuales permiten documentar la salida generada por el sistema y las condiciones asociadas a cada inferencia.

6.6 Cierre de resultados de extracción

Los resultados registraron tres etapas en la evolución del sistema: PIXEL como primera versión funcional, RED-NOVA como configuración intermedia y LSC-V1 como versión final. Esta progresión evidenció la reducción de dimensionalidad, el ajuste de las secuencias de entrada y la optimización del proceso de inferencia. Por su desempeño general, LSC-V1 fue seleccionada como la configuración más adecuada para el sistema propuesto.

7. Resultados de la integración backend y frontend

Como resultado del desarrollo del segundo objetivo específico, se obtuvo un aplicativo web orientado al reconocimiento de palabras en Lengua de Señas Colombiana, implementado bajo una arquitectura cliente-servidor. En esta estructura, el frontend se encargó de la interacción con el usuario y el backend del procesamiento del video, la preparación de la entrada y la ejecución del modelo de reconocimiento.

Esta integración permitió establecer un flujo funcional en el que el usuario puede grabar una seña en tiempo real o cargar un video previamente almacenado. Posteriormente, el archivo es enviado al servidor para su procesamiento y el sistema retorna una interpretación textual visible en la interfaz. En el frontend se desarrolló inicialmente una interfaz básica, centrada en las funciones de carga de video y grabación en tiempo real, como se muestra en la Fig. 59.

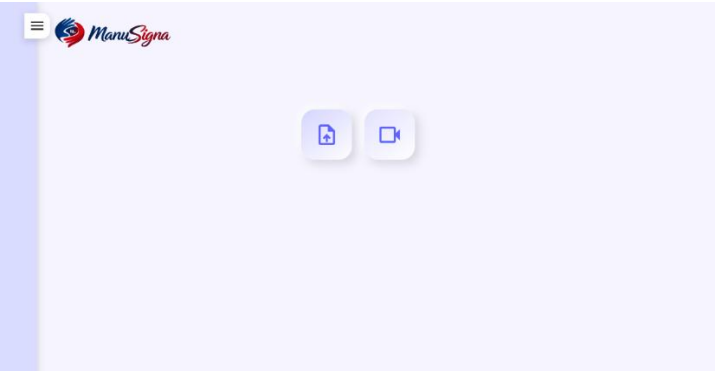


Fig. 59 Primera interfaz

Fuente: Elaboración propia

Esta versión permitió validar la interacción inicial con el sistema, aunque posteriormente fue ajustada a partir de observaciones identificadas durante la revisión de la herramienta. A partir de esta versión inicial, se realizaron ajustes en la organización visual de la página, la distribución de componentes, el encabezado, la presentación de instrucciones, el acceso al glosario y la visualización del resultado de interpretación, como se observa en la Fig. 60**.**

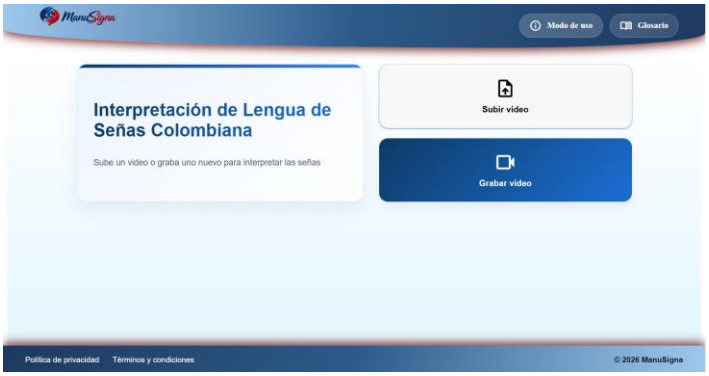


Fig. 60 Versión Final interfaz

Fuente: Elaboración propia

Estos cambios dieron lugar a una versión final con una estructura más organizada y una distribución funcional más clara. Como parte de las mejoras incorporadas, se incluyó un apartado de modo de uso para orientar la captura adecuada de las señas, como se presenta en la Fig. 61.

Fig. 61 Modo de uso incorporado en la interfaz.

Fuente: Elaboración propia

En este apartado se especificaron condiciones de uso como la posición frente a la cámara, la distancia recomendada, la visibilidad desde la cintura hacia arriba y los criterios generales para grabar o cargar un video válido. Asimismo, se integró un glosario con las palabras disponibles en la versión implementada del aplicativo, como se muestra en la Fig. 62.

Fig. 62 Glosario de palabras disponibles en el aplicativo.

Fuente: Elaboración propia

Este recurso permitió informar al usuario sobre el vocabulario reconocido por el sistema.

Adicionalmente, la versión final incorporó un recurso introductorio orientado a facilitar la comprensión inicial del aplicativo. Este componente consistió en una explicación de uso mediante un monólogo hablado acompañado de su interpretación en Lengua de Señas Colombiana, con el fin de complementar la orientación ofrecida por la interfaz.

Como se resume en la TABLA. IX, la evolución de la interfaz incluyó mejoras en la organización visual, las instrucciones de uso, la incorporación del glosario, la presentación del resultado de interpretación y la integración de un recurso introductorio en LSC.

TABLA. IX
CAMBIOS REGISTRADOS ENTRE LA INTERFAZ INICIAL Y LA INTERFAZ FINAL.

Elemento	Interfaz inicial	Interfaz final
Organización visual	Estructura básica centrada en carga y grabación	Distribución más organizada por bloques funcionales
Encabezado	Presentación inicial simple	Encabezado más definido
Instrucciones de uso	Información limitada	Apartado de modo de uso incorporado
Glosario	No visible	Glosario visible con palabras disponibles
Resultado de interpretación	Visualización básica	Presentación más clara del resultado
Recurso introductorio	No incorporado	Monólogo con interpretación en LSC

Fuente: Elaboración propia

Desde la perspectiva del backend, se implementó una estructura encargada de recibir los videos enviados desde el frontend, procesarlos y ejecutar el modelo de reconocimiento. El servidor admitió las dos modalidades de entrada del aplicativo: carga de video previamente grabado y grabación en tiempo real. Como salida, retornó una respuesta estructurada con la interpretación generada por el modelo.

Durante su desarrollo, el backend integró diferentes versiones del pipeline de extracción y preparación de características, ajustadas a las configuraciones evaluadas en el proyecto. Estas versiones incluyeron variaciones en el número de frames, la dimensión de entrada y la selección de puntos clave, como se muestra en la Fig. 63**.**

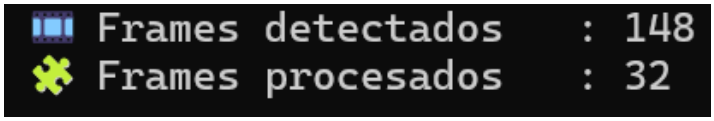


Fig. 63 Estructura del backend o fragmento del procesamiento de video implementado.

Fuente: Elaboración propia

La versión final del backend quedó definida para trabajar con una entrada de 32 frames y 132 características por frame, correspondiente al modelo seleccionado para la integración final del aplicativo.

Como parte de la integración backend-modelo, se implementó un flujo de inferencia encargado de recibir el video, almacenarlo temporalmente, extraer la secuencia de entrada, enviarla al modelo cargado en el servidor y retornar la predicción al frontend. Este flujo permitió articular el procesamiento del video con la respuesta visible en la interfaz web.

En la versión final se incorporaron procesos de lectura del video, selección de fotogramas, extracción de características, normalización de la secuencia y ejecución de la predicción. Asimismo, se incluyó la eliminación del archivo temporal una vez finalizada la inferencia, con el fin de mantener un procesamiento organizado en el servidor.

Finalmente, se implementó una respuesta en formato JSON para la comunicación entre backend y frontend. Esta respuesta incluyó la predicción principal generada por el modelo y, en la versión final, información complementaria del proceso de inferencia, como la probabilidad de la clase reconocida y los tiempos generales de procesamiento. Esta estructura permitió al frontend presentar la interpretación obtenida al usuario.

En el frontend se consolidaron dos modalidades de interacción con el sistema. La primera correspondió a la carga de un video previamente grabado, mediante la selección de un archivo desde el dispositivo del usuario para su envío al backend. La segunda correspondió a la grabación en tiempo real, en la cual el usuario activa la cámara desde la interfaz, realiza la seña y envía el video generado para su análisis.

Ambas modalidades quedaron integradas en un mismo flujo de procesamiento. En ambos casos, el sistema envió el archivo al backend, ejecutó el proceso de inferencia y presentó la interpretación textual en la interfaz, como se muestra en la Fig. 64**.**

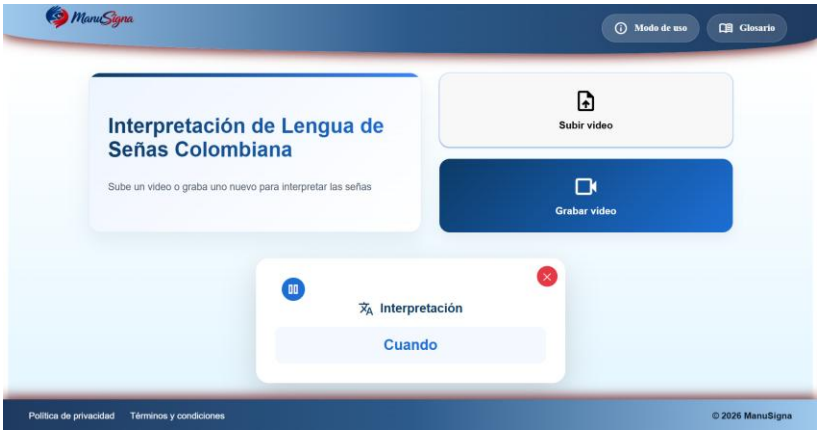


Fig. 64 Visualización de la interpretación generada por el sistema en la interfaz.

Fuente: Elaboración propia

Como resultado de esta integración, se obtuvo una versión funcional del aplicativo web con comunicación entre frontend y backend, procesamiento de video en el servidor, ejecución del modelo de reconocimiento y visualización de la respuesta en pantalla. Esta versión incluyó, además, elementos de apoyo al usuario, como el glosario de palabras disponibles, el apartado de modo de uso y el recurso introductorio con interpretación en Lengua de Señas Colombiana.

Finalmente, la integración entre frontend y backend permitió consolidar una herramienta operativa para el reconocimiento de palabras aisladas en Lengua de Señas Colombiana, dentro del alcance definido para el proyecto. El aplicativo resultante quedó en capacidad de recibir videos, procesarlos mediante el backend y presentar la interpretación correspondiente en la interfaz de usuario.

Para contrastar los resultados de MANUSIGNA con soluciones tecnológicas relacionadas con la Lengua de Señas Colombiana y otros sistemas de apoyo a la comunicación de personas sordas, se compararon los referentes Hablemos, SignAll, SingChat y Cerrando la brecha comunicativa. La comparación consideró el enfoque tecnológico, el tipo de solución, el alcance funcional, las técnicas reportadas y los aportes diferenciales del aplicativo desarrollado.

Hablemos corresponde a una aplicación móvil desarrollada por estudiantes de la Universidad Nacional de Colombia, sede Medellín, orientada a traducir Lengua de Señas Colombiana a texto y audio en tiempo real, así como a convertir texto en lengua de señas mediante un avatar [86]. La información disponible reporta el uso de aprendizaje de máquinas, redes neuronales e inteligencia artificial; sin embargo, no detalla las técnicas, librerías ni metodología implementada [86]. En contraste, MANUSIGNA documentó herramientas específicas como MediaPipe para la extracción de puntos clave, OpenCV para el procesamiento de video, TensorFlow/Keras para el entrenamiento de modelos, y NumPy y Pandas para la manipulación de datos. Además, se enfocó en un aplicativo web para el reconocimiento de palabras específicas en LSC mediante videos cargados o grabados por el usuario.

SignAll corresponde a una solución basada en inteligencia artificial y visión por computador para el reconocimiento y traducción de lengua de señas [87], [88]. También se reporta como una herramienta orientada a facilitar la comunicación entre personas sordas y oyentes, con aplicación inicial en lengua de señas americana [88]. Las fuentes consultadas describen su enfoque general, pero no presentan de forma completa las técnicas, librerías o metodología interna de desarrollo. En comparación, MANUSIGNA se delimitó a la Lengua de Señas Colombiana y al reconocimiento de un conjunto definido de palabras en un entorno web. Además, no se orientó a la traducción completa de conversaciones ni al uso de hardware especializado, sino al procesamiento de videos mediante extracción de características, clasificación con modelos de aprendizaje profundo e integración web.

SingChat se presenta como una aplicación móvil orientada a facilitar la comunicación entre personas oyentes y personas con discapacidad auditiva. Su enfoque principal se relaciona con funciones de mensajería, teclado con alfabeto de lengua de señas, ajustes de tamaño de texto, contraste e interacción entre dispositivos móviles y sistemas de mensajería web. No obstante, las fuentes revisadas no describen una metodología técnica de reconocimiento automático mediante video, extracción de características o arquitecturas de aprendizaje profundo. En contraste, MANUSIGNA se centró en el reconocimiento automático de señas ejecutadas en video, utilizando secuencias de 32 fotogramas y 132 características por fotograma para generar la predicción de una clase entrenada.

El trabajo Cerrando la brecha comunicativa mediante el uso de aprendizaje automático propone el uso de técnicas de aprendizaje automático para reducir la brecha comunicacional entre personas oyentes y personas sordas usuarias de lengua de señas [89]. Este referente se relaciona con el propósito general de MANUSIGNA; sin embargo, el aplicativo desarrollado documentó un proceso práctico de implementación que incluyó construcción del dataset, procesamiento de videos, extracción de puntos clave, entrenamiento de diferentes arquitecturas y selección del modelo LSTM + Attention para su integración final.

A nivel regional, la revisión de fuentes relacionadas con Nariño permitió identificar principalmente iniciativas de formación, cursos de Lengua de Señas Colombiana y servicios de interpretación virtual. No obstante, no se identificó una aplicación regional con características técnicas y funcionales equivalentes a MANUSIGNA, es decir, un aplicativo web orientado al reconocimiento automático de palabras en LSC mediante video, extracción de características y modelos de aprendizaje profundo.

La TABLA. X presenta la comparación entre las aplicaciones relacionadas y MANUSIGNA.

TABLA. X
COMPARACIÓN ENTRE APLICACIONES RELACIONADAS Y MANUSIGNA

Referente	Enfoque del proyecto o aplicación	Técnicas, herramientas o metodología reportada	Diferencia frente a MANUSIGNA	Plus de MANUSIGNA
Hablemos	Aplicación móvil orientada a traducir LSC a texto/audio en tiempo real y convertir texto a lengua de señas mediante un avatar.	La fuente consultada menciona aprendizaje de máquinas, redes neuronales e inteligencia artificial; sin embargo, no expone de manera detallada las técnicas, librerías o metodología interna utilizada para su desarrollo .	Hablemos se orienta principalmente a traducción móvil y representación mediante avatar. MANUSIGNA se enfocó en el reconocimiento automático de palabras en LSC a partir de videos cargados o grabados por el usuario.	MANUSIGNA documenta el flujo técnico completo, incluyendo construcción del conjunto de datos, recorte y etiquetado manual de videos, extracción de características, entrenamiento de modelos, inferencia e integración web.
SignAll	Sistema basado en inteligencia artificial y visión por computador para reconocer y traducir lengua de señas, con aplicación inicial en lengua de señas americana .	Las fuentes consultadas describen el enfoque general de inteligencia artificial y visión por computador, pero no presentan de forma completa las técnicas, librerías o metodología interna del sistema.	SignAll se orienta a una solución de traducción más amplia y enfocada principalmente en ASL. MANUSIGNA se delimitó a Lengua de Señas Colombiana y a un conjunto específico de palabras de uso académico.	MANUSIGNA trabaja directamente con LSC, usando videos grabados en el contexto universitario local y un modelo entrenado para reconocer palabras específicas como adios, auditorio, ayuda, bano, biblioteca, cafeteria, carnet, cuando y donde.
SingChat	Aplicación móvil de comunicación accesible con funciones de mensajería, teclado de señas, ajustes de texto, contraste e interacción entre usuario.	Las fuentes consultadas no exponen de manera detallada una metodología de reconocimiento automático mediante video, extracción de características o modelos de aprendizaje profundo.	SingChat se enfoca en mensajería accesible. MANUSIGNA no se centra en chat, sino en procesar videos de señas y generar una predicción automática de la palabra reconocida.	MANUSIGNA integra reconocimiento automático, permitiendo que el usuario cargue o grabe un video, el sistema procese la seña y entregue una salida asociada a la clase reconocida.

Cerrando la brecha comunicativa	Trabajo orientado a reducir barreras comunicativas mediante el uso de aprendizaje automático aplicado a herramientas lingüísticas para personas sordas.	Se reporta el uso de aprendizaje automático; sin embargo, la comparación se centra en el enfoque general del trabajo frente al desarrollo práctico implementado en MANUSIGNA.	Este referente aborda la brecha comunicativa desde el aprendizaje automático. MANUSIGNA llevó el proceso a un prototipo web funcional con entrenamiento, evaluación de modelos e integración del modelo final.	MANUSIGNA evaluó varias arquitecturas, entre ellas CNN 1D Baseline, Transformer, Bi-LSTM y LSTM + Attention, seleccionando esta última para la integración final del aplicativo.
Iniciativas regionales en Nariño	Se identificaron principalmente cursos, procesos de formación en LSC y servicios de interpretación virtual.	Las iniciativas regionales consultadas no corresponden a sistemas de reconocimiento automático de señas mediante video y aprendizaje profundo.	En la revisión realizada no se identificó una aplicación regional con las mismas características técnicas y funcionales de MANUSIGNA.	MANUSIGNA representa una propuesta tecnológica regional diferenciada, orientada al reconocimiento automático de palabras en LSC mediante video, extracción de características y modelos de aprendizaje profundo.
MANUSIGNA	Aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana mediante videos cargados o grabados por el usuario.	Se implementó con MediaPipe para extracción de puntos clave, OpenCV para procesamiento de video, TensorFlow/Keras para entrenamiento de modelos, NumPy y Pandas para manipulación de datos, y Google Colab para entrenamiento y pruebas.	A diferencia de los referentes revisados, MANUSIGNA documenta de forma explícita las técnicas, herramientas, modelos evaluados y resultados obtenidos durante el desarrollo.	El proyecto integró construcción del conjunto de datos, procesamiento de video, extracción de características en secuencias 32 × 132, evaluación de diferentes redes neuronales, selección del modelo LSTM + Attention e integración en un aplicativo web funcional.

Fuente: Resumen adaptado de [86], [87], [88], [89], [90]

C. VALIDAR EL SOFTWARE CON UNA PRUEBA DE CONCEPTO APOYADA POR UN INTÉRPRETE DE LENGUA DE SEÑAS.

En atención al tercer objetivo específico, se realizó la validación del software mediante una prueba de concepto apoyada por una intérprete institucional de Lengua de Señas Colombiana vinculada a la Universidad CESMAG. Esta actividad permitió verificar el funcionamiento del aplicativo en un escenario de uso real, a partir de la ejecución de señas incluidas en la versión implementada y la revisión de la interpretación generada por el sistema.

Como resultado de esta fase, se obtuvo un acta de validación con participación de la intérprete institucional y un estudiante sordo, como se evidencia en el **Anexo. 2 y Anexo. 3** . Este documento constituyó el soporte formal de la revisión funcional del aplicativo y permitió registrar la comprobación realizada sobre el comportamiento del software frente a las señas evaluadas.

Durante la prueba de concepto, se presentó el aplicativo a la intérprete, indicando la estructura de la interfaz, las modalidades de interacción y el flujo general de reconocimiento. Posteriormente, la participante interactuó con el sistema y ejecutó diferentes señas para contrastar la salida generada por el aplicativo con la seña realizada, como se muestra en la Fig. 65.

Fig. 65 Validación con intérprete
Fuente: Elaboración propia

Adicionalmente, participó un estudiante sordo, lo que permitió incorporar la perspectiva de un usuario perteneciente a la población objetivo. Esta participación aportó evidencia sobre el uso del aplicativo en un contexto cercano al escenario real de aplicación, como se registra en el **Anexo. 4.**

De manera complementaria, se realizaron pruebas funcionales con estudiantes de la Universidad CESMAG y UNIMINUTO, como se observa en la Fig. 66. Estas pruebas permitieron verificar el comportamiento del aplicativo con usuarios de diferentes instituciones y distintos niveles de familiaridad con la Lengua de Señas Colombiana. En los casos en que los participantes no dominaban la LSC, se indicó previamente la seña que debían ejecutar y posteriormente se revisó la interpretación generada por el sistema.

Fig. 66 Validación con estudiantes UNIMINUTO

Fuente: Elaboración propia

Los resultados de las pruebas fueron favorables, debido a que el aplicativo reconoció correctamente varias de las señas ejecutadas dentro de las condiciones establecidas. Además, el flujo general del sistema se mantuvo operativo durante los ejercicios realizados: captura o carga del video, procesamiento de la seña, generación de la interpretación y visualización del resultado. Las evidencias de este proceso se presentan en los **Anexo. 5, Anexo. 6 y Anexo. 7.**

Para complementar la validación, se analizaron las respuestas obtenidas mediante los instrumentos aplicados a los participantes, cuyos resultados se presentan en el **Anexo. 8.** Como se observa en la Fig. 67, la mayoría de los usuarios reportó una valoración favorable frente al uso del aplicativo.

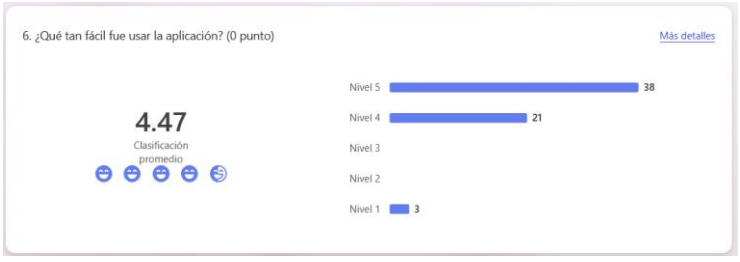


Fig. 67 Nivel de usabilidad de los usuarios con el aplicativo- CESMAG

Fuente: Elaboración propia

En relación con el tiempo de respuesta, los resultados registrados en la Fig. 68 permitieron evidenciar la percepción de los participantes frente al procesamiento del sistema durante las pruebas.

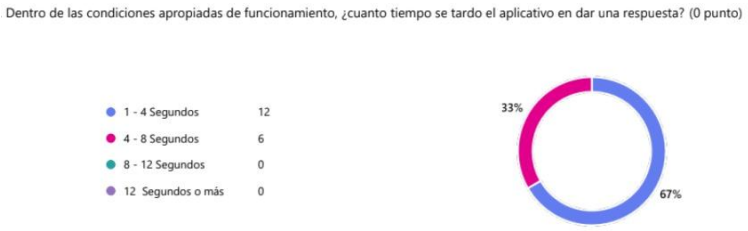


Fig. 68 Tiempo de respuesta del aplicativo- UNIMINUTO

Fuente: Elaboración propia

Asimismo, la Fig. 69 muestra la valoración de los participantes frente al nivel de intuitividad de la interfaz.

Fig. 69 Nivel de intuitividad

Fuente: Elaboración propia

Finalmente, la Fig. 70 presenta la percepción relacionada con el reconocimiento correcto de las señas ejecutadas durante las pruebas.

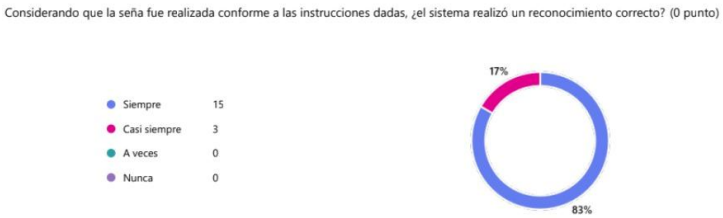


Fig. 70 Reconocimiento correcto de la seña

Fuente: Elaboración propia

En conjunto, los resultados respaldan el cumplimiento del tercer objetivo específico. La validación se realizó mediante una prueba de concepto apoyada por una intérprete de Lengua de Señas

Colombiana, complementada con la participación de un estudiante sordo y pruebas funcionales con usuarios de CESMAG y UNIMINUTO. El acta de validación y las evidencias recolectadas permitieron sustentar que el aplicativo interpreta las señas contempladas en la versión implementada, dentro de las condiciones y límites definidos en la investigación.

V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

A. ANÁLISIS DE TÉCNICAS Y METODOLOGÍAS PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA

1. Análisis de técnicas de modelado

A partir de los resultados obtenidos y su contraste con la literatura reciente, se evidencian diferencias en la forma en que cada arquitectura procesa la información espacial y temporal en el reconocimiento de Lengua de Señas Colombiana (LSC). Estas diferencias impactan tanto el comportamiento durante el entrenamiento como la capacidad de modelar la dinámica del gesto.

1.1 Redes CNN

Las redes neuronales convolucionales (CNN) presentan una alta capacidad para extraer características espaciales en datos visuales, mediante el uso de filtros convolucionales que permiten identificar patrones locales como bordes y contornos[91], [92].

No obstante, este tipo de arquitectura no incorpora mecanismos explícitos para modelar la relación temporal entre fotogramas, ya que procesa cada imagen de manera independiente. En consecuencia, su capacidad se limita a representar el estado del gesto en un instante específico, sin capturar la transición entre estados, lo cual es fundamental en el reconocimiento de señas dinámicas[93].

En el contexto de la Lengua de Señas Colombiana (LSC), esta limitación resulta crítica, dado que el significado de una seña depende no solo de la configuración de la mano, sino de su evolución temporal. Por ello, configuraciones espaciales similares pueden corresponder a señas distintas cuando se consideran sus variaciones en el tiempo.

En este sentido, se infiere que las CNN son adecuadas para la extracción de características espaciales, pero requieren ser complementadas con modelos que integren información temporal, como arquitecturas recurrentes, para un reconocimiento más completo del gesto [93].

1.2 Redes RNN

Las redes neuronales recurrentes (RNN) permiten modelar información secuencial al incorporar un estado interno que se actualiza a lo largo del tiempo, facilitando la captura de relaciones temporales entre los distintos instantes del gesto .

Sin embargo, estas arquitecturas presentan limitaciones en la representación de dependencias de largo alcance, debido al problema del desvanecimiento del gradiente, lo que dificulta la conservación de información relevante en secuencias extensas 90.

En el contexto de la LSC, esta limitación resulta significativa, ya que muchas señas requieren integrar múltiples fases del movimiento. Como consecuencia, las RNN pueden presentar dificultades para modelar adecuadamente la dinámica completa del gesto en escenarios con alta complejidad temporal. La literatura reporta además una menor estabilidad en el entrenamiento de estas arquitecturas en comparación con modelos más avanzados, lo que puede afectar su capacidad de generalización [94].

En este sentido, se infiere que, aunque las RNN representan un avance frente a las CNN en el tratamiento temporal, su desempeño es limitado en tareas que requieren modelar dependencias complejas, siendo superadas por arquitecturas con mecanismos de memoria más robustos.

1.3 Redes LSTM

Las arquitecturas LSTM incorporan celdas de memoria y compuertas de control que permiten regular el flujo de información a lo largo de la secuencia, facilitando la conservación de dependencias de largo alcance y reduciendo el efecto del desvanecimiento del gradiente [95], [96]. Estas compuertas permiten eliminar información irrelevante, incorporar nuevos datos y controlar la salida del estado interno, lo que favorece una representación más estable y completa de la secuencia.

Desde el punto de vista del reconocimiento de la LSC, esta capacidad resulta especialmente relevante, ya que muchas señas se componen de múltiples fases con variaciones en el tiempo. En este sentido, las LSTM permiten integrar información distribuida a lo largo de la secuencia, mejorando la interpretación del gesto completo. Asimismo, la literatura destaca una mayor estabilidad en el entrenamiento y una menor pérdida de contexto en comparación con las RNN tradicionales, lo que se traduce en una representación más robusta de los datos secuenciales[97].

Las LSTM presentan un desempeño superior en el reconocimiento de señas dinámicas, especialmente en escenarios con dependencias temporales prolongadas, cambios de ritmo y coarticulación. Esta capacidad se asocia a sus mecanismos de memoria, que permiten conservar e integrar información a lo largo de la secuencia.

Asimismo, su rendimiento puede potenciarse mediante la integración con mecanismos adicionales, como atención, que optimizan la selección de información relevante. Este comportamiento es consistente con lo reportado en la literatura reciente [95].

1.4 Redes BiLSTM

Las BiLSTM presentan una mayor capacidad para modelar el contexto temporal al procesar la secuencia en ambas direcciones, integrando información pasada y futura en cada instante. Este enfoque permite construir representaciones más completas del gesto en comparación con las LSTM unidireccionales.

En el reconocimiento de la Lengua de Señas Colombiana (LSC), esta característica resulta especialmente relevante, ya que el significado de una seña depende de su evolución completa y no únicamente de estados aislados. En este sentido, la incorporación de información bidireccional favorece la diferenciación de gestos con configuraciones iniciales similares, pero con desarrollos distintos. La literatura reporta que las BiLSTM mejoran la capacidad de discriminación en secuencias complejas al integrar información contextual global [98].

En consecuencia, se infiere que las BiLSTM proporcionan una representación más robusta del gesto frente a modelos unidireccionales. No obstante, este incremento en la capacidad de modelado implica un mayor costo computacional, lo cual debe considerarse en la implementación del sistema.

1.5 Mecanismos de atención

El mecanismo de atención mejora la capacidad del modelo para enfocarse en los segmentos más relevantes de la secuencia, mediante la asignación de pesos de importancia a cada instante. De esta forma, se construye una representación ponderada en la que los elementos más informativos tienen mayor influencia en el proceso de reconocimiento.

En el contexto de la LSC, esta característica resulta especialmente relevante, ya que ciertos momentos del gesto, como cambios en la configuración de la mano o transiciones clave, determinan el significado de la seña. La literatura reporta que la integración de mecanismos de atención en modelos secuenciales mejora la captura de información significativa y el rendimiento en tareas de reconocimiento de señas [99].

En consecuencia, se infiere que la atención actúa como un mecanismo complementario que potencia el modelado secuencial, especialmente cuando la información relevante se distribuye de manera no uniforme en la secuencia.

1.6 Arquitecturas basadas en Transformers

Las arquitecturas basadas en Transformers presentan una alta capacidad para modelar relaciones globales mediante mecanismos de autoatención, permitiendo establecer dependencias directas entre los elementos de la secuencia sin procesamiento estrictamente secuencial[100], [101].

Esta característica facilita la captura de dependencias de largo alcance y la construcción de representaciones contextuales completas, lo cual resulta relevante en el reconocimiento de gestos complejos. No obstante, la literatura señala una alta demanda computacional asociada al mecanismo de autoatención, debido al cálculo de relaciones entre todos los elementos de la secuencia, lo que implica un crecimiento cuadrático en términos de memoria y tiempo de procesamiento[102], [103].

En el contexto de la LSC, esto limita su aplicabilidad en escenarios con recursos restringidos o conjuntos de datos reducidos. En consecuencia, aunque los Transformers ofrecen alto desempeño en condiciones ideales, arquitecturas como LSTM o BiLSTM presentan un mejor equilibrio entre rendimiento y costo computacional en entornos prácticos.

2. Análisis de técnicas de representación de datos

2.1 Representación basada en video crudo

La representación basada en video crudo permite conservar la totalidad de la información visual de la escena, incluyendo color, textura, iluminación y contexto espacial. Esto proporciona una

descripción completa del gesto, capturando tanto información espacial como temporal a partir de secuencias de fotogramas.

Sin embargo, este enfoque implica el manejo de datos de alta dimensionalidad, lo que incrementa la complejidad del entrenamiento en términos de memoria y tiempo de procesamiento. Además, la variabilidad asociada a condiciones de captura, como iluminación o fondo, introduce ruido que dificulta el aprendizaje del modelo .

En el contexto de la lsc, esta limitación resulta crítica, ya que las señas se desarrollan en entornos no controlados, lo que exige una alta capacidad de generalización. La literatura señala que los enfoques basados en video requieren grandes volúmenes de datos y recursos computacionales para lograr un desempeño adecuado [104], [105].

En consecuencia, se infiere que, aunque el video crudo ofrece una representación rica del gesto, su uso implica un incremento significativo en la complejidad del sistema, por lo que debe evaluarse en función de los recursos disponibles.

2.2 Representación basada en landmarks

La representación basada en landmarks reduce la información visual a un conjunto estructurado de coordenadas que describen la posición de puntos clave del cuerpo. Esta transformación disminuye la dimensionalidad de los datos y elimina información irrelevante del entorno, como fondo o iluminación.

Como resultado, se facilita el proceso de aprendizaje, al concentrarse en los elementos directamente asociados al gesto. Además, al organizarse como secuencias de vectores, esta representación favorece la captura de la dinámica temporal del movimiento.

En el contexto de la lsc, este enfoque resulta adecuado, ya que el significado de las señas depende principalmente de la configuración y el movimiento de las manos. En consecuencia, la representación basada en landmarks mejora la robustez del sistema frente a variaciones del entorno.

Asimismo, la reducción de la complejidad de los datos contribuye a una mayor eficiencia computacional en entrenamiento e inferencia. Este comportamiento coincide con lo reportado en

la literatura, donde se destaca que el uso de puntos clave mediante herramientas como MediaPipe permite obtener representaciones más estables y eficientes[106].

En síntesis, se infiere que los landmarks constituyen una alternativa más eficiente y robusta frente al video crudo, especialmente en escenarios con recursos limitados.

3. Análisis metodológico

3.1 Metodología experimental comparativa

El enfoque experimental comparativo permite evaluar arquitecturas y técnicas de representación mediante métricas de desempeño, evidenciando diferencias en precisión, pérdida y comportamiento durante el entrenamiento.

Este análisis muestra que el desempeño de los modelos no depende únicamente de su arquitectura, sino de su interacción con la representación de los datos. En consecuencia, se descartan comparaciones aisladas y se priorizan evaluaciones bajo configuraciones controladas.

Asimismo, se identifican trade-offs entre métricas, como incrementos en precisión asociados a mayor variabilidad o mejoras en pérdida con mayor costo computacional. Esto evidencia la necesidad de un análisis integral del desempeño, considerando múltiples indicadores.

La literatura respalda este enfoque, señalando que métricas como accuracy, precision, recall y F1-score permiten comparar modelos de manera objetiva bajo condiciones equivalentes [107], [108].

3.2 Metodología de prototipado iterativo

El prototipado iterativo permite el desarrollo progresivo del sistema mediante ciclos sucesivos de entrenamiento y ajuste del modelo. Este enfoque facilita la identificación de variables críticas, como hiperparámetros, longitud de secuencias y representación de datos.

Cada iteración aporta información que orienta la optimización del modelo, permitiendo mejorar su precisión, estabilidad y capacidad de generalización. Además, establece un proceso de retroalimentación continua que guía las decisiones de ajuste.

Este enfoque es ampliamente utilizado en aprendizaje automático, donde los modelos se refinan progresivamente mediante ciclos de prueba, evaluación y ajuste hasta alcanzar un desempeño adecuado [109].

En consecuencia, se infiere que el prototipado iterativo resulta clave para adaptar el modelo a las características del problema, especialmente en escenarios con alta variabilidad como el reconocimiento de señas.

B. DESARROLLO DE UN APLICATIVO WEB PARA EL RECONOCIMIENTO DE LA LENGUA DE SEÑAS COLOMBIANA (LSC), UTILIZANDO LAS TÉCNICAS Y METODOLOGÍAS PREVIAMENTE ANALIZADAS.

1. Análisis del desarrollo de modelos y selección de arquitectura

La comparación entre CNN 1D Baseline, Transformer, Bi-LSTM y LSTM + Attention estable evidenció que las métricas internas de entrenamiento y validación no eran suficientes para seleccionar el modelo final. Aunque varias arquitecturas registraron valores altos de precisión y pérdida, las inferencias externas con videos no utilizados durante el ajuste mostraron errores de clasificación en algunos modelos. Por tanto, fue necesario complementar las métricas internas con pruebas externas de inferencia.

CNN 1D funcionó como primera referencia del proceso. Su estructura simple y baja cantidad de parámetros facilitaron el entrenamiento, pero presentó errores en inferencias externas. El modelo Transformer también obtuvo métricas internas altas, aunque registró confusiones entre clases con trayectorias o configuraciones gestuales similares. Por su parte, Bi-LSTM mejoró la captura temporal de las secuencias, pero aún presentó confusiones puntuales. LSTM + Attention estable fue seleccionado para la integración final porque combinó modelado temporal y ponderación de información relevante dentro de la secuencia. Además, sus inferencias externas documentadas fueron más consistentes con las clases reales de los videos evaluados. Esta condición fue relevante porque el aplicativo requería un modelo con buen comportamiento no solo en entrenamiento y validación, sino también en pruebas cercanas al uso real.

En consecuencia, la selección del modelo final no se basó únicamente en el accuracy. La decisión consideró entrenamiento, validación, matriz de confusión, reporte de clasificación e inferencias

externas. Esta evaluación integral respaldó la elección de LSTM + Attention estable como base para la integración final del sistema.

2. Análisis de la evolución de las extracciones en backend

El desarrollo del sistema de reconocimiento de Lengua de Señas Colombiana fue progresivo. Las configuraciones PIXEL, RED-NOVA y LSC-V1 permitieron identificar mejoras en la arquitectura, la extracción de características y la selección de información relevante para el reconocimiento.

Las pruebas evidenciaron que el desempeño del sistema no dependió únicamente del modelo utilizado, sino también de la calidad de la extracción, la selección de características, la longitud de la secuencia y los criterios de validación de la salida. Este comportamiento es coherente con estudios que resaltan la utilidad de los keypoints corporales y manuales en el reconocimiento de señas, así como la necesidad de modelar su dimensión espacial y temporal [110], [111], [112].

La evolución entre configuraciones respondió a un proceso de depuración técnica. Las primeras versiones conservaron mayor cantidad de información, mientras que la configuración final priorizó una representación más compacta y controlada. Este resultado coincide con enfoques recientes basados en *skeletons* y *keypoints*, donde la efectividad no depende solo de acumular *landmarks*, sino de seleccionar los puntos con mayor aporte semántico para el gesto [111], [112].

3. Análisis de PIXEL: funcionalidad inicial y limitaciones estructurales

La configuración PIXEL correspondió a una etapa inicial de experimentación, orientada a lograr un pipeline funcional. Esta versión utilizó pose corporal completa y ambas manos, con una entrada de 225 características por frame. Además, incorporó una función de velocidad temporal, una capa de atención personalizada y la política mixed_float16. La capa de atención permitió ponderar segmentos relevantes de la secuencia [113], mientras que mixed_float16 se empleó como estrategia de precisión mixta en TensorFlow, combinando operaciones de 16 y 32 bits según el hardware disponible [114].

La principal limitación de PIXEL fue la acumulación de complejidad en la entrada y en la arquitectura. La representación mantenía una dimensión elevada y dependía de componentes adicionales, lo que aumentaba la fragilidad del pipeline durante la carga del modelo, la

transformación temporal de la secuencia y la compatibilidad entre bloques internos. Por tanto, aunque la precisión mixta puede aportar mejoras de rendimiento, su efecto depende del entorno de ejecución y no garantiza una mejora operativa directa [114].

Además, esta configuración no contaba con una selección suficientemente depurada de características. El sistema procesaba información corporal y manual amplia, pero parte de estos datos podía ser redundante para la discriminación entre señas. En este sentido, PIXEL permitió comprobar la viabilidad inicial del enfoque, pero evidenció que una mayor complejidad no implicaba necesariamente un mejor desempeño. Este resultado coincide con estudios sobre representaciones basadas en skeletons, donde el rendimiento depende de la organización y modelado de los puntos, más que de su acumulación [111], [112].

En consecuencia, PIXEL fue una etapa funcional, pero limitada. Sus resultados indicaron la necesidad de simplificar la entrada, seleccionar mejor las características y controlar la calidad de la secuencia para obtener inferencias más estables.

4. *Análisis de RED-NOVA: mejora parcial y sobrecarga informativa*

La configuración RED-NOVA representó una etapa intermedia de mejora frente a PIXEL. Esta versión trabajó con secuencias de 48 frames y una entrada de 375 características por frame, conformada por pose corporal completa, ambas manos y un bloque facial simulado en ceros. Esta estructura mantuvo una representación amplia del gesto, coherente con el enfoque de MediaPipe Holistic, el cual permite integrar landmarks de pose, manos y rostro en una representación corporal completa [110].

Los resultados indicaron una mayor estabilidad operativa respecto a la configuración inicial. Sin embargo, la entrada continuó siendo extensa y con información parcialmente redundante. En particular, el bloque facial simulado no correspondía a una señal real del gesto, sino a un recurso para conservar la compatibilidad dimensional del modelo. Por tanto, parte del procesamiento se destinaba a información con bajo aporte semántico para la clasificación. Esta observación coincide con estudios que señalan que la utilidad de los keypoints depende de su relevancia para representar el gesto y no solo de su cantidad [111], [112].

Desde el análisis técnico, RED-NOVA mejoró la continuidad del pipeline, pero no resolvió completamente el problema de representación. El uso de más características por frame y una secuencia más larga incrementó el costo computacional y afectó el equilibrio entre precisión y tiempo de respuesta. En este sentido, el desempeño del sistema seguía limitado por una entrada amplia y menos eficiente de lo necesario. Este comportamiento es coherente con estudios basados en skeletons, donde el reto principal consiste en construir una representación espacial y temporal adecuada [111], [112].

En síntesis, RED-NOVA fue una etapa de transición. Permitió mejorar frente a la primera versión, pero evidenció que aumentar la cantidad de landmarks no garantizaba un mejor desempeño. Este resultado orientó la evolución hacia una representación más compacta y mejor seleccionada.

5. Análisis *de* LSC-V1*:* simplificación útil y control de calidad

La mejora más clara se observó en LSC-V1. Esta configuración redujo la secuencia a 32 frames y la entrada a 132 características por frame, utilizando los hombros como contexto corporal mínimo y ambas manos como fuente principal de información gestual. Esta depuración permitió conservar los elementos más relevantes para el reconocimiento. MediaPipe Holistic permite obtener landmarks de pose y manos en tiempo real [110], y estudios sobre reconocimiento de señas basados en skeletons y transformers señalan que la dinámica espacial y temporal de estos keypoints puede representar información semántica del gesto [111], [112].

El primer factor de mejora fue la reducción funcional de la dimensionalidad. A diferencia de las versiones anteriores, LSC-V1 eliminó información corporal amplia y bloques adicionales con bajo aporte discriminativo. El modelo final conservó el contexto corporal mínimo y priorizó las manos como elemento central del reconocimiento. Este enfoque coincide con trabajos que resaltan la importancia de modelar los landmarks relevantes, en lugar de aumentar indiscriminadamente la cantidad de puntos procesados [111], [112].

El segundo factor fue el control de la secuencia temporal. El pipeline seleccionó frames a partir del conteo reportado por OpenCV y, cuando el video presentaba metadatos insuficientes o inconsistentes, realizó una lectura secuencial alternativa. Posteriormente, aplicó filtrado por movimiento y remuestreo uniforme hasta completar la longitud requerida. Este procedimiento

permitió construir una entrada más representativa del gesto. La importancia del modelado temporal también ha sido reportada en trabajos que combinan MediaPipe Holistic con LSTM y mecanismos de atención para secuencias de señas [112], [115]

El tercer factor fue la incorporación de métricas de calidad en la inferencia. LSC-V1 no aceptó la predicción únicamente por la clase con mayor probabilidad, sino que evaluó Zero ratio, nivel de movimiento, umbral mínimo de confianza y margen entre las dos clases más probables. Con ello, el sistema verificó si la muestra procesada tenía condiciones suficientes para respaldar la salida generada.

En consecuencia, LSC-V1 no solo presentó mayor consistencia en la clasificación, sino también un criterio de aceptación más controlado. Esta configuración redujo la aceptación de salidas ambiguas o con baja calidad de entrada, por lo que fue metodológicamente más adecuada para la integración final del aplicativo

6. Razones principales de la mejora entre extracciones

La comparación entre PIXEL, RED-NOVA y LSC-V1 permitió establecer que la mejora del sistema no dependió de un único cambio, sino de la combinación de varias decisiones técnicas.

La primera decisión fue la depuración de la entrada. Las versiones iniciales conservaron mayor cantidad de landmarks; sin embargo, esto podía incluir información redundante, artificial o poco discriminativa. La versión final adoptó una representación más enfocada en los elementos esenciales del gesto, en coherencia con estudios que resaltan la importancia de seleccionar y modelar keypoints relevantes [111], [112]

La segunda decisión fue la optimización del procesamiento temporal. Al pasar de secuencias más largas a secuencias compactas y mejor seleccionadas, el sistema redujo ruido temporal y concentró la inferencia en fragmentos más representativos. Este enfoque favoreció la estabilidad de la predicción y la eficiencia computacional, aspectos relacionados con propuestas que combinan keypoints con modelos secuenciales o mecanismos de atención [112], [115].

La tercera decisión fue la incorporación de criterios de validación de calidad. A diferencia de las primeras configuraciones, LSC-V1 no se limitó a producir una salida, sino que evaluó si la muestra

tenía condiciones suficientes para respaldar la predicción. Esto permitió diferenciar entre una clasificación posible y una respuesta confiable para el usuario.

La cuarta decisión fue el equilibrio entre precisión y tiempo de respuesta. Las configuraciones iniciales evidenciaron que una entrada más extensa y una arquitectura más compleja no garantizaban mejor desempeño. En cambio, LSC-V1 mostró que una representación más simple y mejor seleccionada podía generar resultados más estables con menor costo computacional. Este criterio fue relevante porque el sistema debía operar dentro de un aplicativo web. Además, TensorFlow señala que las mejoras asociadas a mixed precision dependen del hardware y del entorno de ejecución, por lo que una mayor complejidad no asegura una mejora automática [114].

7. Zero ratio, movement y margin como indicadores de madurez del sistema

En el reporte de inferencia, Zero ratio permitió cuantificar la proporción de valores en cero dentro de la secuencia procesada. Un valor alto indicó baja disponibilidad de información útil, asociada a fallos en la detección de puntos clave o a una muestra insuficiente para el reconocimiento. En la configuración final, este indicador se incorporó como criterio para respaldar o rechazar la predicción generada.

La métrica movement permitió evaluar la variación gestual entre frames consecutivos. Este criterio fue necesario porque un video podía ser leído correctamente, pero no contener movimiento suficiente para representar una seña válida. Su incorporación permitió diferenciar secuencias estáticas de secuencias con actividad gestual real, en coherencia con la importancia de la dinámica temporal en el reconocimiento de señas basado en keypoints [111], [112], [115].

El indicador margin top1-top2 midió la diferencia entre la probabilidad de la clase principal y la segunda clase más probable. Esta métrica permitió identificar si la predicción era suficientemente clara o si existía ambigüedad entre dos clases posibles.

En conjunto, estos indicadores fortalecieron la etapa de inferencia, ya que permitieron evaluar no solo la clase predicha, sino también la calidad de la muestra y la solidez de la decisión. Esto representó una mejora frente a las configuraciones iniciales, en las que la salida dependía principalmente de la clase con mayor probabilidad.

8. Síntesis del análisis de extracciones

En síntesis, la evolución entre PIXEL, RED-NOVA y LSC-V1 evidenció que la mejora del sistema dependió de la selección de información relevante, el control de la secuencia temporal y la incorporación de criterios de calidad en la inferencia. PIXEL permitió comprobar la viabilidad inicial del sistema, pero presentó limitaciones estructurales. RED-NOVA logró una mejora parcial, aunque mantuvo una representación de entrada extensa. Finalmente, LSC-V1 obtuvo el mejor desempeño al integrar una entrada compacta, una secuencia mejor controlada y una lógica de inferencia más confiable. Este comportamiento es consistente con estudios sobre reconocimiento de señas basado en skeletons, atención y secuencias temporales [111], [112], [113], [115].

Este resultado confirma que el avance del sistema estuvo relacionado con la depuración del pipeline de extracción y con la validación de la calidad de la predicción. Por tanto, el modelo final fue seleccionado no solo por su desempeño observado, sino también por la solidez metodológica del proceso de evaluación.

9. Análisis de la integración backend y frontend

En atención al segundo objetivo específico, se diseñó e implementó un aplicativo web para el reconocimiento de palabras en Lengua de Señas Colombiana, integrando visión por computador, aprendizaje profundo y desarrollo web bajo una arquitectura cliente-servidor. El sistema quedó estructurado en cuatro componentes principales: dataset, entrenamiento del modelo, backend de procesamiento e inferencia, y frontend de interacción con el usuario.

El dataset se construyó a partir de videos de palabras en LSC, recolectados bajo orientación de una persona con conocimiento profesional en Lengua de Señas Colombiana. Los videos fueron organizados por clase, lo que permitió consolidar una base etiquetada para las etapas de preprocesamiento, extracción de características, entrenamiento y validación. El almacenamiento en Google Drive facilitó la centralización y disponibilidad del material durante el desarrollo experimental. El entrenamiento se realizó en Google Colab, utilizando rutinas de procesamiento orientadas a representar cada video como una secuencia temporal de fotogramas. Para ello, se aplicaron procedimientos de extracción de características que transformaron la información visual en arreglos numéricos compatibles con las arquitecturas de aprendizaje profundo evaluadas. Esta

representación permitió conservar información asociada a postura, movimiento y configuración gestual.

La etapa de modelado se desarrolló de forma iterativa. Las arquitecturas evaluadas se ajustaron de acuerdo con la forma de entrada generada por cada pipeline de extracción, considerando la longitud de las secuencias, la cantidad de características por frame y la organización de los puntos clave. De esta manera, el entrenamiento integró tanto la configuración del modelo como la consistencia entre la representación del gesto y la estructura computacional encargada de clasificarlo. Una vez entrenado el modelo, se exportó la arquitectura y sus pesos para integrarlos en el backend desarrollado en Python con Flask. Este componente se encargó de recibir los videos enviados desde el frontend, ejecutar la preparación de la entrada, cargar el modelo seleccionado y retornar la interpretación generada. Flask permitió estructurar rutas de procesamiento, gestionar solicitudes HTTP y encapsular la lógica de inferencia dentro de un servicio web ligero.

El backend incorporó rutinas de extracción adaptadas a las configuraciones evaluadas durante el proyecto. Esta decisión fue necesaria porque los modelos no compartían una misma estructura de entrada. Por tanto, cada flujo de inferencia debía conservar correspondencia con las condiciones utilizadas durante el entrenamiento, incluyendo número de frames, dimensionalidad, selección de puntos clave y organización de las características. La inferencia exigió consistencia entre entrenamiento e implementación. Para ello, el backend ejecutó lectura del video, selección de fotogramas, extracción de características, normalización, adecuación dimensional y predicción. Así, el servidor cumplió una doble función: comunicar la interfaz con el modelo y transformar el video bruto en una entrada compatible con la arquitectura seleccionada.

La comunicación entre frontend y backend se implementó mediante una API REST en Flask. A través de solicitudes HTTP, el frontend envió el video al servidor y recibió una respuesta estructurada con la interpretación textual generada por el modelo. Esta separación permitió desacoplar la lógica de presentación de la lógica de reconocimiento, favoreciendo la modularidad y mantenibilidad del sistema. El frontend fue desarrollado en Angular, framework utilizado para estructurar la interfaz del aplicativo y gestionar la interacción del usuario. Esta capa permitió implementar la carga de videos previamente grabados, la grabación en tiempo real mediante la cámara del dispositivo y la comunicación con el backend mediante servicios HTTP.

El flujo operativo del aplicativo contempló dos modalidades de entrada: carga de video y grabación en tiempo real. En ambos casos, el frontend encapsuló el archivo en una solicitud HTTP, el backend procesó el video, ejecutó la inferencia y retornó la predicción. Finalmente, la interfaz presentó la interpretación textual al usuario. Esta estructura permitió mantener un flujo homogéneo, independientemente del origen del video.

La interfaz fue ajustada de forma iterativa a partir de observaciones realizadas durante la revisión del aplicativo. La versión inicial permitió validar las funciones básicas de carga y grabación; posteriormente, se reorganizaron componentes, instrucciones, acciones principales, glosario y visualización del resultado. Estos ajustes permitieron mejorar la claridad funcional y la experiencia de uso. Desde el diseño del sistema, la arquitectura cliente-servidor permitió separar responsabilidades. El frontend gestionó la interacción, captura y presentación de resultados, mientras que el backend concentró el procesamiento, la transformación de los videos y la ejecución del modelo. Esta organización favoreció la modularidad, el mantenimiento y la posibilidad de realizar ajustes independientes en cada capa.

El alcance funcional del aplicativo se delimitó al reconocimiento de palabras individuales previamente definidas en el conjunto de clases entrenadas. Por tanto, el sistema no abordó interpretación continua, análisis sintáctico ni reconocimiento de frases completas en LSC. Esta delimitación permitió mantener coherencia entre el dataset construido, los modelos evaluados, la interfaz desarrollada y los recursos disponibles.

En síntesis, el desarrollo del aplicativo integró la construcción del dataset, la preparación secuencial de los datos, el entrenamiento de modelos, la implementación del backend, la comunicación mediante API REST, el desarrollo del frontend en Angular y el ajuste de la interfaz según observaciones de uso. Como resultado, se obtuvo una herramienta web funcional para el reconocimiento de palabras aisladas en Lengua de Señas Colombiana.

10. Síntesis integradora del desarrollo

La integración de modelos, configuraciones de extracción y aplicativo web evidenció que el avance del proyecto dependió de varias decisiones técnicas. La construcción del dataset permitió trabajar con señas ajustadas al alcance de la investigación. La comparación de arquitecturas permitió

seleccionar un modelo con buen desempeño en validación e inferencias externas. El refinamiento del backend permitió alinear la inferencia con una extracción más estable y compacta. Finalmente, el rediseño del frontend mejoró la interacción del usuario y la presentación del resultado.

En conjunto, el sistema evolucionó desde pruebas experimentales hacia una versión integrada capaz de recibir videos, procesarlos en el backend, ejecutar la inferencia y presentar una interpretación textual en la interfaz web. Esta progresión respalda el cumplimiento del objetivo relacionado con el desarrollo del aplicativo y establece la base para la validación funcional con intérprete, estudiante sordo y participantes de las instituciones involucradas.

C. VALIDAR EL SOFTWARE CON UNA PRUEBA DE CONCEPTO APOYADA POR UN INTÉRPRETE DE LENGUA DE SEÑAS.

La validación permitió analizar el aplicativo en un contexto aplicado, conforme al tercer objetivo específico. La prueba de concepto se desarrolló con apoyo de una intérprete institucional de Lengua de Señas Colombiana vinculada a la Universidad CESMAG. Su participación permitió contrastar la seña ejecutada con la interpretación generada por el sistema desde un criterio experto.

Como soporte del proceso se obtuvieron dos actas de validación. La primera correspondió a la revisión realizada con apoyo de la intérprete institucional de LSC, la cual constituye la evidencia principal del cumplimiento del tercer objetivo específico. Esta acta formalizó la prueba de concepto y registró la valoración del aplicativo frente a las señas evaluadas.

La segunda acta correspondió a la participación de un estudiante sordo, perteneciente a la población objetivo del proyecto. Esta evidencia tuvo carácter complementario y permitió incorporar la perspectiva de un usuario directamente relacionado con el contexto de aplicación del sistema.

En conjunto, ambas actas respaldaron la validación desde dos perspectivas: el criterio experto de la intérprete de LSC y la experiencia de uso de una persona sorda. Aunque el acta de la intérprete representa el soporte principal de la validación, el acta del estudiante sordo fortaleció el proceso desde la población beneficiaria.

De manera complementaria, se realizaron pruebas funcionales con estudiantes de la Universidad CESMAG y UNIMINUTO. Estas pruebas permitieron observar el comportamiento del aplicativo con usuarios externos al equipo desarrollador y con distintos niveles de familiaridad con la LSC.

En los casos en que los participantes no dominaban la lengua de señas, se indicó previamente la seña que debían ejecutar y luego se verificó la interpretación generada por el sistema.

Posteriormente, se aplicaron encuestas a los participantes de CESMAG y UNIMINUTO. En total se recopilaron 80 respuestas: 62 de CESMAG y 18 de UNIMINUTO. Los instrumentos permitieron conocer la percepción de los usuarios sobre facilidad de uso, intuitividad, claridad del resultado, reconocimiento de señas, utilidad comunicativa, estabilidad del sistema, satisfacción general, recomendación y oportunidades de mejora.

El aplicativo fue diseñado principalmente como apoyo para personas sordas o con discapacidad auditiva, por lo que incluyó recursos de orientación como el instructivo en Lengua de Señas Colombiana. La participación de usuarios oyentes tuvo una finalidad funcional y complementaria, orientada a verificar el flujo de carga o grabación de videos, procesamiento de la seña y visualización de la interpretación.

En facilidad de uso, el 95 % de los participantes calificó el aplicativo con valores altos, correspondientes a 4 o 5 en la escala utilizada. Este resultado indica una percepción favorable frente a las acciones principales del sistema: cargar video, grabar una seña y consultar la interpretación generada.

En intuitividad de la interfaz, el 86,2 % de los participantes otorgó calificaciones entre 4 y 5. Además, el 90 % indicó que no se sintió confundido durante el uso de la herramienta. Estos resultados respaldan la claridad del flujo de interacción implementado.

En reconocimiento de señas, el 97,5 % de los participantes indicó que el sistema reconoció correctamente la seña siempre o casi siempre, considerando que esta fue ejecutada según las instrucciones. Este resultado evidencia una percepción favorable del modelo dentro del conjunto de palabras implementadas y bajo las condiciones definidas para la prueba. No obstante, debe interpretarse dentro del alcance de la prueba de concepto, ya que el reconocimiento depende de factores como posición frente a la cámara, iluminación, distancia, ejecución de la seña y calidad del video.

Sobre la necesidad de repetir la seña, el 78,8 % indicó que no fue necesario repetirla, mientras que el 21,2 % señaló que sí tuvo que hacerlo en algún momento. Este resultado muestra un comportamiento favorable en la mayoría de casos, aunque también evidencia que algunas

interpretaciones pueden verse afectadas por condiciones de captura, encuadre corporal, similitud entre gestos o forma de ejecución. Por ello, el apartado de modo de uso resulta necesario para orientar la distancia, visibilidad corporal, iluminación y ejecución adecuada de la seña.

En claridad del resultado, el 100 % de los participantes indicó que la interpretación presentada por el aplicativo fue clara. Este dato respalda la forma en que la interfaz comunica la palabra reconocida después del procesamiento del video.

En utilidad comunicativa, el 93,8 % consideró que la herramienta ayuda a la comunicación entre personas sordas y oyentes, mientras que el porcentaje restante respondió “tal vez”. Este resultado se relaciona con el propósito del proyecto, dado que el aplicativo no busca reemplazar al intérprete de LSC, sino funcionar como apoyo tecnológico para la comunicación básica en contextos académicos.

En cuanto al posible uso académico, el 96,2 % manifestó que sí usaría el aplicativo en este tipo de escenario. Esta percepción evidencia aceptación del sistema como recurso de apoyo en espacios educativos.

Durante la interacción, el 100 % de los participantes indicó que el software no presentó fallos o errores técnicos. Este resultado muestra estabilidad durante las condiciones de prueba, sin que ello implique ausencia absoluta de errores en otros contextos de uso.

La satisfacción general fue positiva: el 93,8 % calificó este aspecto con valores entre 4 y 5, con un promedio general de 4,45 sobre 5. Asimismo, el 98,8 % manifestó que recomendaría la aplicación, lo que evidencia una aceptación favorable después de la interacción.

Respecto al tiempo de respuesta, 61 participantes respondieron esta pregunta. De ellos, el 93,4 % indicó que el aplicativo tardó entre 1 y 8 segundos en generar una respuesta bajo condiciones apropiadas de funcionamiento. Los tiempos superiores reportados pueden relacionarse con conexión, carga del video, dispositivo utilizado o condiciones del procesamiento.

Las respuestas abiertas permitieron identificar oportunidades de mejora. Las sugerencias más frecuentes fueron ampliar el vocabulario, incorporar más señas, fortalecer el glosario e incluir frases u oraciones completas. También se propuso agregar ejemplos visuales o videos de apoyo para orientar la ejecución correcta de cada seña. Estas observaciones son coherentes con el alcance

actual del sistema, el cual reconoce un conjunto definido de palabras y no realiza interpretación completa de la LSC.

También se reportaron posibles mejoras relacionadas con tiempo de respuesta, cámara, presentación visual de la interfaz y precisión en algunas señas. Estas observaciones no contradicen la valoración favorable del sistema, sino que señalan aspectos relevantes para futuras versiones.

En las respuestas abiertas se identificaron algunos registros ambiguos o no relacionados directamente con la pregunta. Por esta razón, el análisis se apoyó principalmente en las preguntas cerradas, ya que permitieron una lectura más uniforme y comparable. Las respuestas abiertas se utilizaron como complemento cualitativo para identificar tendencias, sugerencias y percepciones adicionales.

Desde el punto de vista metodológico, la validación combinó tres elementos: criterio experto, participación de la población objetivo y pruebas funcionales con usuarios de instituciones educativas. En coherencia con el tercer objetivo específico, el criterio experto de la intérprete de LSC tuvo mayor peso dentro del proceso, mientras que la participación del estudiante sordo y de los estudiantes de ambas universidades aportó evidencia complementaria desde la experiencia de uso.

La prueba de concepto fue pertinente para la etapa de desarrollo alcanzada. Dado que el sistema reconoce señas específicas y no traduce conversaciones completas en LSC, la validación se enfocó en comprobar si el aplicativo interpretaba las señas incluidas en la versión implementada. Esta delimitación evita sobredimensionar el alcance del software y mantiene coherencia con los objetivos de la investigación.

Asimismo, la validación evidenció que el aplicativo puede funcionar como apoyo complementario en contextos educativos. La herramienta no reemplaza la función del intérprete ni aborda toda la complejidad comunicativa de la LSC, pero puede contribuir a interacciones básicas, procesos de sensibilización y apoyo comunicativo entre personas sordas y oyentes.

En consecuencia, el cumplimiento del tercer objetivo específico se sustentó en la prueba de concepto apoyada por una intérprete institucional de LSC y formalizada mediante el acta correspondiente. De forma complementaria, el proceso se fortaleció con el acta del estudiante sordo, pruebas funcionales y encuestas posteriores al uso del aplicativo. Los resultados obtenidos

permitieron validar el sistema dentro del alcance previsto y evidenciaron una percepción favorable frente a facilidad de uso, claridad del resultado, utilidad comunicativa, estabilidad, satisfacción y posible uso académico.

Síntesis de resultados y validación de la hipótesis

A partir de los resultados obtenidos en el desarrollo e implementación del aplicativo web, así como de las pruebas realizadas en escenarios de validación, se evidencia el cumplimiento de los objetivos específicos planteados en la investigación. En este sentido, el sistema desarrollado permitió el reconocimiento de un conjunto definido de palabras en Lengua de Señas Colombiana, aportando al apoyo en la comunicación de estudiantes con discapacidad auditiva.

Asimismo, los resultados permiten afirmar que se da cumplimiento a la hipótesis alterna, dado que la solución propuesta contribuye principalmente a la comunicación básica de los usuarios, más que al acceso amplio a contenidos o a procesos complejos de interpretación. Este comportamiento es coherente con el alcance y las delimitaciones establecidas en el proyecto, evidenciando la pertinencia de la solución desarrollada.

CONCLUSIONES

En el marco del presente trabajo de grado, se desarrolló el artículo titulado *"Comparative Analysis of Deep Learning Models for Colombian Sign Language Interpretation"*, el cual fue estructurado conforme a estándares académicos y científicos con el propósito de ser sometido a evaluación en una revista indexada. Este producto investigativo constituye un resultado relevante del proyecto, ya que sistematiza el proceso metodológico, el análisis comparativo de modelos basados en aprendizaje profundo —incluyendo arquitecturas como Red Neuronal Recurrente (RNN), Long Short-Term Memory (LSTM), Bi-LSTM y Transformer— y los resultados obtenidos en el reconocimiento de la Lengua de Señas Colombiana (LSC). En este sentido, el artículo no solo respalda la rigurosidad del desarrollo realizado, sino que también proyecta el trabajo hacia la comunidad científica, aportando evidencia empírica y metodológica al campo del reconocimiento automático de señas mediante técnicas de inteligencia artificial.

El desarrollo del proyecto permitió implementar un aplicativo web orientado al reconocimiento de palabras en Lengua de Señas Colombiana (LSC), con el propósito de apoyar la comunicación básica y el acceso a la información de estudiantes con discapacidad auditiva en el contexto académico de UNICESMAG y UNIMINUTO. A partir del proceso realizado, se evidenció que la integración de técnicas de inteligencia artificial, visión por computador y desarrollo web contribuye a la creación de herramientas tecnológicas orientadas a la inclusión comunicativa, siempre que se delimiten claramente su alcance, vocabulario disponible y condiciones de uso.

En relación con el primer objetivo específico, se analizaron diferentes técnicas y metodologías aplicadas al reconocimiento de señas, entre ellas CNN, Red Neuronal Recurrente (RNN), Long Short-Term Memory (LSTM), Bi-LSTM, mecanismos de atención y arquitecturas Transformer. Este análisis permitió establecer que las señas requieren un tratamiento secuencial, debido a que su significado depende no solo de configuraciones estáticas, sino también del movimiento, la orientación y la evolución temporal del gesto. En consecuencia, las arquitecturas secuenciales, especialmente LSTM y sus variantes con mecanismos de atención, resultan más adecuadas para este tipo de problema en comparación con modelos centrados exclusivamente en características espaciales.

Asimismo, el análisis de las técnicas de representación de datos permitió determinar que el uso de landmarks o puntos clave constituye una estrategia adecuada frente al procesamiento de video

crudo, debido a que reduce la dimensionalidad de la información, disminuye la influencia de factores externos como el fondo o la iluminación y concentra el procesamiento en las regiones más relevantes del gesto, particularmente manos y parte superior del cuerpo. Esta decisión favoreció la estabilidad del entrenamiento y la integración eficiente del modelo dentro del aplicativo web.

Respecto al segundo objetivo específico, se logró desarrollar un aplicativo web funcional compuesto por un frontend orientado a la interacción del usuario y un backend encargado del procesamiento de la información, incluyendo la recepción de videos, extracción de características, ejecución del modelo de reconocimiento y generación de resultados. El sistema soporta dos modalidades de entrada: carga de video previamente grabado y captura en tiempo real, lo cual amplía las posibilidades de uso dentro del alcance definido como prueba de concepto.

Durante el desarrollo del modelo se evaluaron diferentes arquitecturas, entre ellas CNN 1D Baseline, Transformer, Bi-LSTM y LSTM con mecanismos de atención. Aunque varios modelos presentaron métricas favorables en entrenamiento y validación, las pruebas externas de inferencia evidenciaron diferencias significativas en su capacidad de generalización. A partir de este análisis, el modelo LSTM con atención fue seleccionado como la arquitectura final, debido a su comportamiento más consistente al considerar métricas de entrenamiento, validación, matriz de confusión, reporte de clasificación e inferencias externas documentadas.

La evolución del backend permitió evidenciar que la mejora del sistema no dependió exclusivamente del modelo, sino también de la optimización del proceso de extracción y preparación de características. En este sentido, se consolidó una representación basada en 32 fotogramas y 132 características por fotograma, lo cual permitió equilibrar precisión, estabilidad y tiempo de respuesta, concentrando el procesamiento en información relevante para el reconocimiento de las señas.

En relación con el tercer objetivo específico, se llevó a cabo una prueba de concepto con el apoyo de una intérprete institucional de Lengua de Señas Colombiana, cuya participación permitió validar la correspondencia entre las señas ejecutadas y las interpretaciones generadas por el sistema. Esta validación fue formalizada mediante acta, constituyéndose como evidencia del cumplimiento del objetivo planteado. De manera complementaria, la participación de un estudiante sordo aportó una perspectiva relevante desde la población objetivo, fortaleciendo la evaluación en términos de experiencia de uso, accesibilidad e interacción.

Las pruebas funcionales realizadas con participantes de UNICESMAG y UNIMINUTO evidenciaron una percepción favorable en aspectos como facilidad de uso, intuitividad de la interfaz, claridad de los resultados, utilidad comunicativa y satisfacción general. No obstante, también se identificaron oportunidades de mejora, especialmente relacionadas con la ampliación del vocabulario, incorporación de más señas, fortalecimiento del glosario y evolución hacia estructuras más complejas del lenguaje, lo cual es coherente con las delimitaciones establecidas para el proyecto.

Como resultado adicional, una primera versión del aplicativo se encuentra en proceso de registro ante la Dirección Nacional de Derecho de Autor (DNDA) mostrada en el **Anexo. 9**, lo cual representa un avance en la formalización del producto tecnológico desarrollado. Asimismo, la participación en eventos académicos, como encuentros de investigación y espacios de divulgación científica, permitió socializar los resultados, recibir retroalimentación externa y proyectar el trabajo como una iniciativa de investigación aplicada orientada a la inclusión y accesibilidad.

Finalmente, se concluye que el aplicativo web desarrollado constituye una prueba de concepto viable como herramienta de apoyo para la comunicación básica en contextos académicos, dando cumplimiento al objetivo general de la investigación. De igual manera, se valida la hipótesis alterna, evidenciando que la herramienta contribuye principalmente a la interpretación de un conjunto definido de palabras, más que a la comprensión completa del lenguaje de señas.

En este contexto, la solución propuesta no reemplaza la labor del intérprete de Lengua de Señas Colombiana ni aborda la complejidad total de la comunicación en LSC; sin embargo, se posiciona como un recurso complementario que puede favorecer interacciones básicas, apoyar procesos de inclusión y promover el uso de tecnologías accesibles. En consecuencia, el proyecto demuestra que la aplicación de técnicas de inteligencia artificial al reconocimiento de señas puede contribuir al desarrollo de soluciones inclusivas, siempre que se acompañe de validación especializada, participación de la población objetivo y una delimitación clara de su alcance.

RECOMENDACIONES

Se recomienda ampliar el conjunto de datos incorporando un mayor número de participantes y variaciones en las condiciones de captura, tales como iluminación, distancia, ángulo de cámara y velocidad de ejecución, con el fin de mejorar la capacidad de generalización del modelo en escenarios reales. Asimismo, se sugiere incrementar el vocabulario del sistema, priorizando palabras y expresiones de uso frecuente en contextos académicos, lo cual permitiría fortalecer su utilidad en entornos educativos inclusivos. Como trabajo futuro, se sugiere realizar validaciones con una muestra más amplia de usuarios dentro del contexto regional, incluyendo personas sordas, intérpretes y usuarios oyentes, con el objetivo de evaluar el comportamiento del sistema en escenarios reales de uso. Finalmente, se sugiere optimizar la arquitectura del modelo mediante el ajuste de hiperparámetros y del umbral de confianza, así como explorar técnicas basadas en mecanismos de atención, con el fin de mejorar la precision en clases visualmente similares y fortalecer la robustez del sistema.

REFERENCIAS BIBLIOGRÁFICAS

[1] "DANE - Discapacidad". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/discapacidad

[2] L. D. R. Peñafiel, "Importancia de las TIC en el proceso de aprendizaje de personas con discapacidad auditiva", *Rev. Univ. Informática RUNIN*, núm. 16, Art. núm. 16, dic. 2023.

[3] Asana, "Prueba de concepto (POC): qué es y cómo implementarla [2026] • Asana", Asana. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://asana.com/es/resources/proof-of-concept

[4] I. D. Mienye, T. G. Swart, y G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications", *Information*, vol. 15, núm. 9, p. 517, sep. 2024, doi: 10.3390/info15090517.

[5] Z. Niu, G. Zhong, y H. Yu, "A review on the attention mechanism of deep learning", *Neurocomputing*, vol. 452, pp. 48–62, sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[6] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking", el 18 de junio de 2020, *arXiv*: arXiv:2006.10214. doi: 10.48550/arXiv.2006.10214.

[7] "THE 17 GOALS | Sustainable Development". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://sdgs.un.org/goals

[8] "2025 Latin America Report | Global Education Monitoring Report". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.unesco.org/gem-report/en/publication/2025-latin-america-report

[9] "Educación Superior archivos", Insor Educativo. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://educativo.insor.gov.co/tipoasesorias/educacion-superior/

[10] "Ley 982 de 2005 - Gestor Normativo". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=17283

[11] "Ley 1618 de 2013 - Gestor Normativo". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=52081

[12] "ASORNAR - FENASCOL". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://fenascol.org.co/asociaciones-afiliadas/asornar/

[13] "Microsoft Word - 0722666S.doc". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.un.org/esa/socdev/enable/documents/tccconvs.pdf

[14] "Compilación Jurídica del MINTIC - Ley 324 de 1996". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://normograma.mintic.gov.co/mintic/compilacion/docs/ley_0324_1996.htm?utm_source=chatgpt.com

[15] "articles-182816_archivo_pdf_decreto_366_febrero_9_2009". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.mineducacion.gov.co/1621/articles-182816_archivo_pdf_decreto_366_febrero_9_2009.pdf

[16] "Ley 982 de 2005 - Gestor Normativo". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=17283

[17] "Ley 1618 de 2013 - Gestor Normativo". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=52081

[18] "¿Cuál es la percepción de la comunidad sorda en Colombia sobre el acceso a la TV abierta?", CRC. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.crcom.gov.co/es/noticias/estudios/cual-es-percepcion-comunidad-sorda-en-colombia-sobre-acceso-tv-abierta

[19] Redacción, “Solo el 5% de los jóvenes sordos en Colombia tienen oportunidades de acceder a la educación superior”, DiariOriente. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://diarioriente.com/altiplano/solo-el-5-d.html
[20] L. Ibero, “Lanzan herramienta digital para la investigación en lengua de señas en Colombia”, La Ibero. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.ibero.edu.co/blog/articulos/lanzan-herramienta-digital-para-la-investigacion-en-lengua-de-senas-en-colombia
[21] P. L. L. Munar, “‘Colombia tiene porcentaje bajo en la implementación de los derechos de las personas sordas’, Insor”, infobae. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.infobae.com/colombia/2023/09/24/colombia-tiene-porcentaje-bajo-en-la-implementacion-de-los-derechos-de-las-personas-sordas-insor/
[22] “Inicio”, Centro de relevo. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: http://centroderelevo.gov.co/632/w3-channel.html
[23] “Petición · MinTIC deja incomunicadas a más de 500.000 personas sordas tras suspender Centro Relevo - Colombia · Change.org”. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.change.org/p/mintic-deja-incomunicadas-a-m%C3%A1s-de-500-000-personas-sordas-tras-suspender-centro-relevo
[24] “Situación de discapacidad auditiva en Colombia | OIR AUDIOLOGIA”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://oiraudiologia.com.co/situacion-de-discapacidad-auditiva-en-colombia/
[25] N. Casen, “Nota Estadística No. 1 de 2023”.
[26] “Ley 1346 de 2009 - Gestor Normativo”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=37150
[27] “Ley 982 de 2005 - Gestor Normativo”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=17283
[28] “Ley 1618 de 2013 - Gestor Normativo”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=52081
[29] “Decreto 1421 de agosto 29 de 2017 - Decreto 1421 de agosto 29 de 2017”, Portal MEN - Presentación. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.mineducacion.gov.co/1780/w3-article-381928.html
[30] “La Inteligencia Artificial en la elaboración de un Diccionario de Lengua de Señas”. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://es.linkedin.com/pulse/la-inteligencia-artificial-en-elaboraci%C3%B3n-de-un-se%C3%B1as-granda-ter%C3%A1n-xxhee
[31] S. R. Montoya, “Incluedu, la plataforma de aprendizaje de lengua de señas basada en IA que pone la tecnología al servicio de la inclusión”, Contxto. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.contxto.com/es/tecnologia/incluedu-la-plataforma-de-aprendizaje-de-lengua-de-senas-basada-en-ia-que-pone-la-tecnologia-al-servicio-de-la-inclusion/
[32] A. Montoya, “Desarrollan prototipo virtual de aprendizaje para personas con discapacidad auditiva”, Facultad de Minas | Universidad Nacional de Colombia. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://minas.medellin.unal.edu.co/noticias/facultad/5654-desarrollan-prototipo-virtual-de-aprendizaje-para-personas-con-discapacidad-auditiva
[33] “Un proyecto tecnológico para la inclusión social y educativa”. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.ucundinamarca.edu.co/index.php/noticias-

ucundinamarca/93-noticias-extension-soacha/3600-un-proyecto-tecnologico-para-la-inclusion-social-y-educativa
[34] "SignAll Sign Language Translation App". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.signall.live
[35] K. M. Pérez, "Sistema de reconocimiento del Lenguaje de Señas Mexicano basado en una cámara RGB-D y aprendizaje automático ", sep. 2022, Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://ri-ng.uaq.mx/handle/123456789/3799
[36] A. Gaure, "REAL-TIME SIGN LANGUAGE TRANSLATOR: A MACHINE LEARNING-BASED APPROACH".
[37] J. J. G. Leguizamón, J. A. P. López, M. J. S. Barón, y J. S. G. Sanabria, "Revista Tecnura", *Tecnura*, vol. 26, núm. 74, Art. núm. 74, oct. 2022, doi: 10.14483/22487638.19213.
[38] "Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales | Orinoquia". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://orinoquia.unillanos.edu.co/index.php/orinoquia/article/view/680
[39] A. Quintero Bedoya, "Herramientas de aprendizaje automático con redes neuronales para el reconocimiento de Lengua de Señas Colombiana", 2024, Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://hdl.handle.net/10495/38597
[40] "'Thomasito' herramienta digital basada en ajustes razonables para las matemáticas mediante uso de lengua de señas colombianas -LSC-". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://repositorio.umariana.edu.co/handle/20.500.14112/28068
[41] danielabarco, "Un Nariño más incluyente: Plataforma SERVIR, para mejorar la comunicación con la comunidad sorda", TuBarco Noticias. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://tubarco.news/un-narino-mas-incluyente-plataforma-servir-para-mejorar-la-comunicacion-con-la-comunidad-sorda/
[42] noreply G. Nariño, "Avanzamos en la construcción de un departamento incluyente para las personas sordas de Mi Nariño | Gobernación de Nariño". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://2020-2023.narino.gov.co/noticias/1379-2/
[43] E. rdu, "Lenguaje: instrumento del desarrollo humano", RDU UNAM. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.revista.unam.mx/2021v22n5/lenguaje_instrumento_del_desarrollo_humano/
[44] "Blog - Idioma, lengua, lenguaje y dialecto". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.unaminternacional.unam.mx/es/blog/idioma-lengua-lenguaje-y-dialecto
[45] "Central Washington University | ¿Qué es la comunicación?" Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.cwu.edu/academics/communication/es/que-es-la-comunicacion.php
[46] [En línea]. Disponible en: https://www.merida.tecnm.mx/wp-content/uploads/2022/12/prueba3.pdf
[47] "tccconvs". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.un.org/esa/socdev/enable/documents/tccconvs.pdf
[48] "Sordera y pérdida de la audición". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss
[49] "Tecnología y lengua de señas: transformando la inclusión a través de la comunicación - ¿Y si hablamos de igualdad?" Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://blogs.iadb.org/igualdad/es/tecnologia-lengua-de-senas-inclusion/

[50] "Qué son las Soluciones tecnológicas | Desarrollo de servicios digitales". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://desarrollo.juntadeandalucia.es/soluciones-tecnologicas/que-son-soluciones-tecnologicas
[51] "What is Web Application (Web Apps) and its Benefits? | Definition from TechTarget", Search Software Quality. Consultado: el 20 de mayo de 2025. [En línea]. Disponible en: https://www.techtarget.com/searchsoftwarequality/definition/Web-application-Web-app
[52] "¿Qué es la inteligencia artificial y cómo se usa?", Temas | Parlamento Europeo. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.europarl.europa.eu/topics/es/article/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa
[53] "¿Qué son las redes neuronales y cómo se aplican a la Inteligencia Artificial?", UNIE. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.universidadunie.com/blog/que-son-redes-neuronales
[54] "¿Qué es el reconocimiento de imágenes? | IBM". Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.ibm.com/es-es/think/topics/image-recognition
[55] "(PDF) Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales", *ResearchGate*, Consultado: el 29 de mayo de 2025. [En línea]. Disponible en: https://www.researchgate.net/publication/359262472_Reconocimiento_del_abecedario_de_la_lengua_de_senas_colombiana_con_Redes_Neuronales_Convolucionales
[56] "ISO/IEC 25010:2023", ISO. Consultado: el 19 de mayo de 2025. [En línea]. Disponible en: https://www.iso.org/standard/78176.html
[57] F. Valenzuela, G. Beguerí, y C. Collazos, "Centro de Investigación de la Universidad Distrital Francisco José de Caldas", *Rev. Vínculos*, vol. 11, núm. 2, Art. núm. 2, dic. 2014, doi: 10.14483/2322939X.9666.
[58] "(PDF) Manual Práctico Para El Diseño De La Escala Likert", *ResearchGate*, doi: 10.37646/xihmai.v2i4.101.
[59] C. Ortega, "Escala nominal: Características y ejemplos", QuestionPro. Consultado: el 19 de mayo de 2026. [En línea]. Disponible en: https://www.questionpro.com/blog/es/escala-nominal/
[60] F. B. Ríos, "PARADIGMAS Y PERSPECTIVAS TEÓRICO-METODOLÓGICAS EN EL ESTUDIO DE LA ADMINISTRACIÓN".
[61] N. K. Denzin y Y. S. Lincoln, *The SAGE Handbook of Qualitative Research*. SAGE Publications, 2011.
[62] R. Hernández Sampieri y C. F. Fernandez-Collado, *Metodología de la investigación*, Sexta edición. México D.F.: McGraw-Hill Education, 2014.
[63] A. Velázquez, "Investigación no experimental: Qué es, características y ejemplos", QuestionPro. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.questionpro.com/blog/es/investigacion-no-experimental/
[64] "Quasi-Experimentation: A Guide to Design and Analysis", Guilford Press. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://www.guilford.com/books/Quasi-Experimentation/Charles-Reichardt/9781462540204
[65] A. J. Q. Vodniza, "Guía de investigación cuantitativa".
[66] G. P. Guevara Alban, A. E. Verdesoto Arguello, y N. E. Castro Molina, "Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción)", *RECIMUNDO*, vol. 4, núm. 3, pp. 163–173, jul. 2020, doi: 10.26820/recimundo/4.(3).julio.2020.163-173.

[67] “ISO 9241-11:2018”, ISO. Consultado: el 19 de mayo de 2026. [En línea]. Disponible en: https://www.iso.org/standard/63500.html
[68] “Acciones y distracciones – Insor Educativo”. Consultado: el 6 de mayo de 2025. [En línea]. Disponible en: https://educativo.insor.gov.co/catdiccionario/acciones-y-distracciones/
[69] E. Parks y J. Parks, “A Sociolinguistic Profile of the Peruvian Deaf Community”, *Sign Lang. Stud.*, vol. 10, núm. 4, pp. 409–441, 2010.
[70] “Los distintos tipos de pruebas en software | Atlassian”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.atlassian.com/es/continuous-delivery/software-testing/types-of-software-testing
[71] K.-L. Du, C.-S. Leung, W. H. Mow, y M. N. S. Swamy, “Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era”, *Mathematics*, vol. 10, núm. 24, p. 4730, ene. 2022, doi: 10.3390/math10244730.
[72] “Basic CNN Architecture Explained: Key Layers and Workflow”, upGrad blog. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.upgrad.com/blog/basic-cnn-architecture/
[73] N. Chen, “Exploring the development and application of LSTM variants”, *Appl. Comput. Eng.*, vol. 53, pp. 103–107, mar. 2024, doi: 10.54254/2755-2721/53/20241288.
[74] “Understanding LSTM Networks -- colah’s blog”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[75] F. Landi, L. Baraldi, M. Cornia, y R. Cucchiara, “Working Memory Connections for LSTM”, *Neural Netw.*, vol. 144, pp. 334–341, dic. 2021, doi: 10.1016/j.neunet.2021.08.030.
[76] S. Gureja, “LSTMs and Bi-LSTM in PyTorch”, Scaler Topics. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.scaler.com/topics/pytorch/lstm-pytorch/
[77] Y. Tay, M. Dehghani, D. Bahri, y D. Metzler, “Efficient Transformers: A Survey”, el 14 de marzo de 2022, *arXiv*: arXiv:2009.06732. doi: 10.48550/arXiv.2009.06732.
[78] C. M. Rojas, “Qué son los transformers IA y por qué revolucionaron el lenguaje”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.automatizapro.com.ar/blog/transformers-ia-revolucion-lenguaje/
[79] J. Huang y V. Chouvatut, “Video-Based Sign Language Recognition via ResNet and LSTM Network”, *J. Imaging*, vol. 10, núm. 6, p. 149, jun. 2024, doi: 10.3390/jimaging10060149.
[80] H. Silva y J. Bernardino, “Machine Learning Algorithms: An Experimental Evaluation for Decision Support Systems”, *Algorithms*, vol. 15, núm. 4, p. 130, abr. 2022, doi: 10.3390/a15040130.
[81] “The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://research.google/pubs/the-ml-test-score-a-rubric-for-ml-production-readiness-and-technical-debt-reduction/
[82] C. Schröer, F. Kruse, y J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model”, *Procedia Comput. Sci.*, vol. 181, pp. 526–534, ene. 2021, doi: 10.1016/j.procs.2021.01.199.
[83] J. Terven, D. M. Cordova-Esparza, A. Ramirez-Pedraza, E. A. Chavez-Urbiola, y J. A. Romero-Gonzalez, “Loss Functions and Metrics in Deep Learning”, arXiv.org. Consultado: el 30 de abril de 2026. [En línea]. Disponible en: https://arxiv.org/abs/2307.02694v5
[84] L. Ferrer, “Analysis and Comparison of Classification Metrics”, arXiv.org. Consultado: el 30 de abril de 2026. [En línea]. Disponible en: https://arxiv.org/abs/2209.05355v4

[85] “Scrumban | Atlassian”. Consultado: el 19 de mayo de 2026. [En línea]. Disponible en: https://www.atlassian.com/es/agile/project-management/scrumban?utm_source=chatgpt.com
[86] “Universidad Nacional de Colombia : Sede Medellin - Hablemos, app que traduce de lengua de señas a texto y audio en tiempo real”. Consultado: el 20 de abril de 2026. [En línea]. Disponible en: https://medellin.unal.edu.co/noticias/5361-hablemos-lengua-senas.html
[87] “First system to automatically translate sign language”, CORDIS | European Commission. Consultado: el 20 de abril de 2026. [En línea]. Disponible en: https://cordis.europa.eu/article/id/411590-first-system-to-automatically-translate-sign-language
[88] “The only technology to successfully translate between signed and spoken languages | SIGNALL | Project | Fact Sheet | H2020”, CORDIS | European Commission. Consultado: el 20 de abril de 2026. [En línea]. Disponible en: https://cordis.europa.eu/project/id/854984
[89] M.-F. Silva-Joaqui, K. Márceles-Villalba, y S. Amador-Donado, “Cerrando La Brecha Comunicativa Mediante El Aprendizaje Automático Con Una Herramienta Lingüística Para Personas Sordas”, *Rev. Fac. Ing.*, vol. 33, núm. 69, sep. 2024, Consultado: el 20 de abril de 2026. [En línea]. Disponible en: https://www.redalyc.org/journal/4139/413982387009/html/
[90] E. J. A. Dorado, “Estudiantes colombianos crean el primer ‘WhatsApp’ con lenguaje de señas: SingChat”, El Tiempo. Consultado: el 20 de abril de 2026. [En línea]. Disponible en: https://www.eltiempo.com/tecnosfera/apps/estudiantes-colombianos-crean-el-primer-whatsapp-con-lenguaje-de-senas-singchat-3428916
[91] J. Gu *et al.*, “Recent Advances in Convolutional Neural Networks”, el 19 de octubre de 2017, *arXiv*: arXiv:1512.07108. doi: 10.48550/arXiv.1512.07108.
[92] A. Khan, A. Sohail, U. Zahoora, y A. S. Qureshi, “A Survey of the Recent Architectures of Deep Convolutional Neural Networks”, *Artif. Intell. Rev.*, vol. 53, núm. 8, pp. 5455–5516, dic. 2020, doi: 10.1007/s10462-020-09825-6.
[93] J. G. Pauloski *et al.*, “KAISA: An Adaptive Second-Order Optimizer Framework for Deep Neural Networks”, en *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, nov. 2021, pp. 1–14. doi: 10.1145/3458817.3476152.
[94] “Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications[v1] | Preprints.org”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.preprints.org/manuscript/202408.0748
[95] M. Krichen y A. Mihoub, “Long Short-Term Memory Networks: A Comprehensive Survey”, *AI*, vol. 6, núm. 9, p. 215, sep. 2025, doi: 10.3390/ai6090215.
[96] “Unlocking the Power of LSTM for Long Term Time Series Forecasting”. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://arxiv.org/html/2408.10006v1
[97] B. Ghojogh y A. Ghodsi, “Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey”, el 22 de abril de 2023, *arXiv*: arXiv:2304.11461. doi: 10.48550/arXiv.2304.11461.
[98] Z. Buribayev, M. Aouani, Z. Zhangabay, A. Yerkos, Z. Abdirazak, y M. Zhassuzak, “Enhancing Kazakh Sign Language Recognition with BiLSTM Using YOLO Keypoints and Optical Flow”, *Appl. Sci.*, vol. 15, núm. 10, p. 5685, ene. 2025, doi: 10.3390/app15105685.
[99] D. Kumari y R. S. Anand, “Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM Framework Based on Attention Mechanism”, *Electronics*, vol. 13, núm. 7, p. 1229, ene. 2024, doi: 10.3390/electronics13071229.

[100] T. Lin, Y. Wang, X. Liu, y X. Qiu, "A Survey of Transformers", el 15 de junio de 2021, *arXiv*: arXiv:2106.04554. doi: 10.48550/arXiv.2106.04554.

[101] S. Islam *et al.*, "A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks", el 11 de junio de 2023, *arXiv*: arXiv:2306.07303. doi: 10.48550/arXiv.2306.07303.

[102] K. T. Chitty-Venkata, S. Mittal, M. Emani, V. Vishwanath, y A. K. Somani, "A survey of techniques for optimizing transformer inference", *J. Syst. Archit.*, vol. 144, p. 102990, nov. 2023, doi: 10.1016/j.sysarc.2023.102990.

[103] Y. Tay, M. Dehghani, D. Bahri, y D. Metzler, "Efficient Transformers: A Survey", *ACM Comput Surv*, vol. 55, núm. 6, p. 109:1-109:28, dic. 2022, doi: 10.1145/3530811.

[104] "(PDF) Video-Based Sign Language Recognition via ResNet and LSTM Network", ResearchGate. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.researchgate.net/publication/381769656_Video-Based_Sign_Language_Recognition_via_ResNet_and_LSTM_Network

[105] "Sign language recognition based on deep learning and low-cost handcrafted descriptors". Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://arxiv.org/html/2408.07244v1?utm_source=chatgpt.com

[106] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking", el 18 de junio de 2020, *arXiv*: arXiv:2006.10214. doi: 10.48550/arXiv.2006.10214.

[107] O. Rainio, J. Teuho, y R. Klén, "Evaluation metrics and statistical tests for machine learning", *Sci. Rep.*, vol. 14, núm. 1, p. 6086, mar. 2024, doi: 10.1038/s41598-024-56706-x.

[108] G. Olaoye, "Comparative Study of Machine Learning Models", el 9 de febrero de 2025, *Social Science Research Network, Rochester, NY*: 5130265. doi: 10.2139/ssrn.5130265.

[109] S. Gupta, K. Saluja, A. Goyal, A. Vajpayee, y V. Tiwari, "Comparing the performance of machine learning algorithms using estimated accuracy", *Meas. Sens.*, vol. 24, p. 100432, dic. 2022, doi: 10.1016/j.measen.2022.100432.

[110] "Guía de tareas de detección de puntos de referencia holísticos | Google AI Edge", Google AI for Developers. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://ai.google.dev/edge/mediapipe/solutions/vision/holistic_landmarker?hl=es-419

[111] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, y Y. Fu, "Sign Language Recognition via Skeleton-Aware Multi-Model Ensemble", el 12 de octubre de 2021, *arXiv*: arXiv:2110.06161. doi: 10.48550/arXiv.2110.06161.

[112] N. Song y Y. Xiang, "SLGTformer: An Attention-Based Approach to Sign Language Recognition", el 23 de diciembre de 2022, *arXiv*: arXiv:2212.10746. doi: 10.48550/arXiv.2212.10746.

[113] D. Bahdanau, K. Cho, y Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", el 19 de mayo de 2016, *arXiv*: arXiv:1409.0473. doi: 10.48550/arXiv.1409.0473.

[114] "Precisión mixta | TensorFlow Core", TensorFlow. Consultado: el 18 de abril de 2026. [En línea]. Disponible en: https://www.tensorflow.org/guide/mixed_precision?hl=es-419

[115] S. Srivastava, S. Singh, Pooja, y S. Prakash, "Continuous Sign Language Recognition System using Deep Learning with MediaPipe Holistic", *Wirel. Pers. Commun.*, vol. 137, núm. 3, pp. 1455–1468, ago. 2024, doi: 10.1007/s11277-024-11356-0.

ANEXOS

Anexo. 1 Resultados de inferencias externas

https://drive.google.com/drive/folders/1hMtgPuz3ZLjBaoedQZflgoTbYfbI9_na?usp=sharing

Anexo. 2 Carta de validación de intérprete.

https://drive.google.com/file/d/1VfYfec-ShfjdUqFEJ6cFHEQ8yj-pIs37/view?usp=sharing

Anexo. 3 Carta de validación de persona sorda.

https://drive.google.com/file/d/1OytDWTEaGJZSWnVzQURlSwLA-zfXHrBJ/view?usp=sharing

Anexo. 4 Validación con estudiante sordo

Fig. 71 Validación con estudiante sordo
Fuente: Elaboración propia

Fig. 72 Validación con estudiante sordo
Fuente: Elaboración propia

Fig. 73 Validación con estudiante sordo
Fuente: Elaboración propia

Anexo. 5 Pruebas de funcionamiento

Fig. 74 Pruebas de software

Fuente: Elaboración propia

Fig. 75 Pruebas de software

Fuente: Elaboración propia

Fig. 76 Pruebas de software

Fuente: Elaboración propia

Fig. 77 Pruebas de software

Fuente: Elaboración propia

Fig. 78 Pruebas de software

Fuente: Elaboración propia

Fig. 79 Pruebas de software

Fuente: Elaboración propia

Fig. 80 Pruebas de software

Fuente: Elaboración propia

Anexo. 6 Videos con muestra de pruebas

https://drive.google.com/drive/folders/12E-KYFyYPpxcOnAi0j1kuiN27w8oTvEs?usp=sharing

Anexo. 7 Registro de asistencia de estudiantes

https://drive.google.com/drive/folders/1HNocGlNXx1RNOMaQRX9_BUS6-InYLxCW?usp=sharing

Anexo. 8 Formularios resumen de las encuestas

https://drive.google.com/drive/folders/1B1TgrI_zBt8fAAPe6ZclqbOu3YPPrW8p?usp=sharing

Anexo. 9 Certificado de registro ante el DNDA

https://drive.google.com/drive/folders/1QNnUBRSJL03fy7vLDIbboB935xlUj0sV?usp=sharing

Anexo. 10 Tablero Kanban

https://drive.google.com/drive/folders/16vwEwLm6fLxf00UOooh77_2Qgcpfhud6?usp=sharing

UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN	**CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)**	**CÓDIGO:** AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

San Juan de Pasto, 10 de agosto de 2026

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado **APOYO A LA COMUNICACIÓN Y ACCESO A LA INFORMACIÓN DE ESTUDIANTES CON DISCAPACIDAD AUDITIVA EN UNICESMAG Y UNIMINUTO MEDIANTE UN APLICATIVO WEB.** presentado por el (los) autor(es) **Yeferson Audelo Córdoba Gómez**, **Karen Yisenia Rodríguez Rosero** y **Juan Carlos Cordoba Rodriguez** del Programa Académico Ingeniería de sistemas al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,

DIEGO JAVIER VILLARREAL MORENO
C.c. 1085291480 Pasto
Ingeniería de sistemas
3187885536
djvillarreal@unicesmag.edu.co

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Yeferson Audelo Cordoba Gomez	**Documento de identidad:** 1087048925
Correo electrónico: yefersoncorgomez@gmail.com	**Número de contacto:** 3118584751
Nombres y apellidos del autor: Juan Carlos Cordoba Rodriguez	**Documento de identidad:** 1004598675
Correo electrónico: 2.juanca73@gmail.com	**Número de contacto:** 3145494538
Nombres y apellidos del autor: Karen Yisenia Rodriguez Rosero	**Documento de identidad:** 1004634307
Correo electrónico: Karenrodros17@gmail.com	**Número de contacto:** 3206072974
Nombres y apellidos del autor:	**Documento de identidad:**
Correo electrónico:	**Número de contacto:**
Nombres y apellidos del asesor: Diego Javier Villarreal Moreno	**Documento de identidad:** 1085291480
Correo electrónico: djvillarreal@unicesmag.edu.co	**Número de contacto:** 3187885536
Título del trabajo de grado: **APOYO A LA COMUNICACIÓN Y ACCESO A LA INFORMACIÓN DE ESTUDIANTES CON DISCAPACIDAD AUDITIVA EN UNICESMAG Y UNIMINUTO MEDIANTE UN APLICATIVO WEB.**	
Facultad y Programa Académico: Facultad de Ingeniería, Ingeniería de sistemas.	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.

b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.
d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

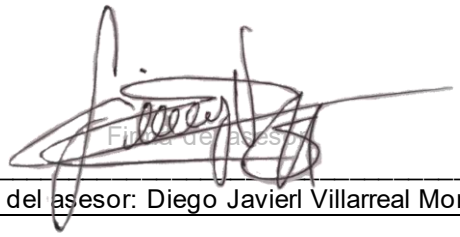
NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permito(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 10 días del mes de agosto del año 2026

Yeferson Córdoba Firma del autor	Juan Carlos C. Firma del autor
Nombre del autor:Yeferson Audelo Cordoba Gomez	Nombre del autor:Juan Carlos Cordoba Rodriguez
Firma del autor	Firma del autor
Nombre del autor: Karen Yisenia Rodriguez Rosero	Nombre del autor:



_____

Nombre del asesor: Diego Javierl Villarreal Moreno