DESARROLLO DE UNA APLICACIÓN PARA EL RECONOCIMIENTO ALFANUMÉRICO DE LA LENGUA DE SEÑAS COLOMBIANA USANDO ALGORITMOS DE INTELIGENCIA ARTIFICIAL

JADER ALEJANDRO MUÑOZ GALINDEZ

Trabajo de grado para optar por el título de
Ingeniero Físico

Director:

Dr. Rubiel Vargas Cañas

Universidad del Cauca
Facultad de Ciencias Naturales, Exactas y de la Educación
Departamento de Física
Popayán, 2023

DESARROLLO DE UNA APLICACIÓN PARA EL RECONOCIMIENTO ALFANUMÉRICO DE LA LENGUA DE SEÑAS COLOMBIANA USANDO ALGORITMOS DE INTELIGENCIA ARTIFICIAL

JADER ALEJANDRO MUÑOZ GALINDEZ

Trabajo de Grado presentado a la Facultad de Ciencias Naturales, Exactas y de la Educación de la Universidad del Cauca, para la obtención del título de ingeniero físico.

Director:

Dr. Rubiel Vargas Cañas

Universidad del Cauca
Facultad de Ciencias Naturales, Exactas y de la Educación
Departamento de Física
Popayán, 2023

NOTA DE ACEPTACIÓN

Aprobado por el comité de grado en cumplimiento de los requisitos exigidos por la Universidad del Cauca para optar al título de Ingeniero Físico.

PhD. Rubiel Vargas Cañas

DIRECTOR

MSc. Nataly Tobar Muñoz

JURADO

MSc. Carlos Felipe Ordoñez Urbano

JURADO

Fecha de sustentación: 10 de abril de 2023

Agradecimientos

A Dios por darme la vida

A mis padres, Tulia Galindez y Servio Muñoz por su amor, apoyo, durante todos estos años ayudándome a cumplir mis objetivos.

A mi familia por su apoyo incondicional.

A mi hermana Karol y todos mis hermanos por siempre persuadir que cada día puedo dar un poquito más de mí mismo.

A todas las personas, compañeros y amigos que me acompañaron en este camino (Ana, Angie, Johana, Verónica, Darwin, Jhonatan, Juan Diego, Duván, Santiago)

A la Universidad del Cauca.

A mi director Rubiel Vargas Cañas por dirigir, apoyándome y aconsejarme con sus ideas para que este trabajo de grado fuese posible.

A todos los voluntarios quienes ayudaron obtener los datos con los que se desarrolló esta investigación

A los profesores del departamento de física.

Al Grupo de Óptica y laser (GOL) por adoptarme en su laboratorio.

Gracias

Dedicado a mi

Hermano Oscar Iván

“En la memoria nadie muere,

pero no sabes cómo duele”

Contenido

Lista de figuras

Lista de tablas

Resumen

La lengua de señas colombiana (LSC) hace parte del patrimonio cultural de Colombia, es protegido por el estado y es la lengua oficial de las personas con discapacidad de origen auditivo. Aunque existen diversas herramientas implementadas por el gobierno nacional, aún existe una brecha grande para la inclusión social, por lo cual, son frecuentes las discriminaciones en ámbitos sociales y laborales, sin tener la posibilidad de dar a entender sus pensamientos e ideas por falta de intérpretes y desconocimiento de esta lengua. Por esta razón, se desarrolló como objetivo principal, un sistema mediante inteligencia artificial para la interpretación dinámica alfanumérica de la lengua de señas colombiana, mediante interpretación por deletreo de palabras y cantidades. Para cumplir con este objetivo, se empleó la metodología CRISP-DM, iniciando con la creación de un dataset dinámico con la participación de 77 voluntarios no expertos en LSC, realizando tres tipos de señas: alfabeto, números y palabras, siendo los dos primeros necesarios para cumplir con el objetivo. A las grabaciones obtenidas se les extrae seis fotogramas y se realiza un filtrado y procesamiento que los convierten en dos dataset, uno enfocado a cuerpo completo de tamaño 255x255 pixeles y otro enfocado en la mano de tamaño 120x120 pixeles. Esta información se modeló de dos formas empleando métodos del aprendizaje profundo; un primer modelo de combinación de redes neuronales convolucionales preentrenadas (CNN) con una doble red neuronal recurrente de memoria a largo y corto plazo (BILSTM) y un segundo modelo de coordenadas, extraídas por el software mediapipe hands que alimentan una red BILSTM directamente. El ajuste fino se realizó mediante validación cruzada con el uso herramientas de regularización como aumentación de datos, early stopping, regularización L2 y Dropout. Los resultados de evaluación arrojaron que los modelos de combinación obtienen una menor exactitud, siendo esta no convergente a cuerpo completo, pero si en el enfoque en mano, con un 75.9% de exactitud; no obstante, los modelos de coordenadas superan estos resultados con una exactitud del 75.5% y 85.7% a cuerpo completo y enfoque en mano respectivamente, siendo este último, desplegado en una plataforma local usando la librería de streamlit, para realizar la interpretación sobre la imagen captada en cámara directamente. Por tanto, se desarrolló un sistema de interpretación alfanumérico de la lengua de señas colombiana bajo condiciones dinámicas que puede ser desplegado en una página web para tener acceso en cualquier horario, lo cual, aporta a la inclusión social de personas con discapacidad auditiva, al permitir interpretación de sus ideas y pensamientos de manera clara y efectiva, lo que a su vez les permite incluirse plenamente en la sociedad.

1. Introducción

1.1 Descripción del problema

La comunicación es un factor importante en la sociedad, es una forma de expresar sentimientos y emociones, por tanto, tiene implicaciones culturales, sociales y psicológicas que afectan el desarrollo de la persona [1]. El intercambio de información se realiza entre un emisor y un receptor; el medio por el cual se ejecuta se llama canal, teniendo implicaciones en los signos usados para la comunicación, ya sea escrita, verbal y no verbal. Es esta última, la que abarca a la población con discapacidad auditiva, la cual, según estimaciones DANE 2020, es de alrededor de medio millón de personas, quienes han desarrollado su propia lengua de señas y gestos para la comunicación [2]. La lengua de señas Colombiana ha recibido múltiples influencias de diversos países, en especial de España por inmigrantes sordos educado en ese país. Actualmente, es reconocido por el estado como la lengua oficial de las personas sordas en Colombia, logrando protección legal que ha permitido realizar diversas iniciativas de enseñanza, promoción y sensibilización, para la inclusión de las personas sordas en la sociedad [3].

En Colombia, a través de la ley 982 de 2005 se protege y brinda apoyo a personas con discapacidad auditiva, dando herramientas para la inclusión social en ámbitos sociales, académicos, laborales, entre otros. Sin embargo, a pesar de la legislación y conciencia social, aun se tiene muchos desafíos por cubrir; uno de los mayores, es referentes a la educación en su lengua natural, donde la falta de escuelas especializadas e intérpretes en escuelas regulares limita las posibilidades de acceso a la información, como consecuencia, hay una disminución en las oportunidades laborales, así como una discriminación, limitando la capacidad de ganarse la vida plena e independiente. Además, esto pone en riesgo la vida de las personas, ya que hospitales y servicios de emergencia en muy pocos casos poseen intérpretes de LSC que ayuden a realizar una interpretación en caso de emergencia. Ahora bien, el desarrollo de comunicación escrita por personas con discapacidad auditiva es posible, mediante la ejecución de las señas de cada una de las letras del alfabeto y los números, sin embargo, su implementación se realiza con reconocimiento estático y enfocado sobre la mano, cuya metodología a desarrollar, deja de lado muchas letras que poseen movimiento y no brindan la oportunidad de establecer los espacios de orientación conversacional.

Para afrontar este reto, se han desarrollado diferentes tipos de investigaciones haciendo uso de técnicas de inteligencia artificial, como estimadores de pose con redes convolucionales, para la clasificación de lengua de señas Colombiana logrando resultados casi perfectos [4], sin embargo, estos resultados no muestran variabilidad en los componentes de entrenamiento, así como, aplicación de técnicas clásicas del aprendizaje automático para la clasificación de vocales y números con resultados de 70% de exactitud [5], como también desarrollos estáticos con imágenes en control de iluminación [6]–[8] que modelan a partir de la extracción de características, sin embargo, en Colombia no se realiza una interpretación de carácter dinámico, por el contrario, su realización se lleva a través de modelos estáticos implementados en condicionadas dinámicas, cumpliendo la labor pero

dejan de lado muchas características y riquezas de lengua de señas. Internacionalmente se realizan este tipo de investigaciones mediante mapas de profundidad, donde su interpretación entre cuadro y cuadro es de unos pocos milisegundos [9], así como, la creación de dataset de carácter dinámicos y su modelamiento a partir de la combinación de modelos de aprendizaje profundo, para la interpretación de lengua de señas estadounidense, cuyos resultados son del 98.6% de exactitud [10].

Por lo anterior, se evidencia la necesidad de investigar sobre el desarrollo de aplicaciones con el uso de inteligencia artificial, que puedan resolver la interpretación de la lengua de señas Colombiana de forma dinámica de torso o cuerpo completo. Por este motivo, se plantea la siguiente pregunta de investigación: ¿Cómo la inteligencia artificial ayudaría a la inclusión de personas con discapacidad auditiva, con aplicaciones de entrada dinámico asequibles y no invasivas para la interpretación de la lengua de señas Colombiana?

1.2 Objetivos

1.2.1 Objetivo general

Desarrollar un sistema mediante inteligencia artificial para la interpretación dinámica alfanumérica de la lengua de señas colombiana.

1.2.2 Objetivos específicos

- Recopilar y crear un dataset con la información de señas y gestos alfanuméricos de la LSC.
- Implementar una combinación de técnicas de inteligencia artificial para la interpretación alfanumérica de la LSC.
- Evaluar el desempeño, alcance y limitaciones del modelo de interpretación alfanumérico de la LSC a partir de las métricas empleadas en inteligencia artificial.

1.3 Descripción de la metodología

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) es un método orientado a trabajos de minería en el cual se cubre fases del proyecto, tareas y relaciones entre estas. Consta de seis etapas (Figura 1): comprensión del problema, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue, estas fases no son rígidas y permite el movimiento entre estas.

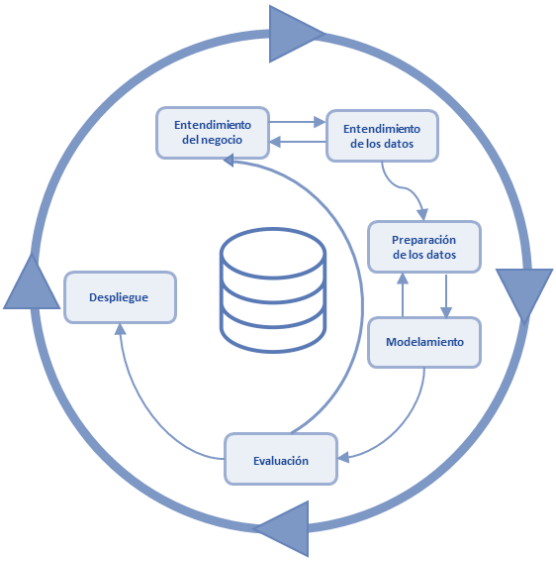


Figura 1: Metodología CRISP-DM
Fuente: Elaboración propia

1.3.1 Comprensión del negocio

Consiste en identificar los requerimientos y problemáticas asociadas a la clasificación de la lengua de señas, al realizar revisión el estado del arte teniendo en cuenta normas y características de este mismo, para formalizarlos como objetivos técnicos y elaborar un plan de trabajo.

1.3.2 Comprensión de los datos

En esta fase se realiza la recopilación de los datos, se revisa tanto la calidad como las particularidades presentes y se da un primer filtrado a la información obtenida, preparando la información para fases posteriores.

1.3.3 Preparación de los datos

Este paso consta de realizar la extracción de cuadros de la base de datos inicial, se selecciona mediante un segundo filtrado la óptima, así como, se construye nueva información si es necesario para obtener nuevas características permitiendo ampliar la base de datos inicial.

1.3.4 Modelado de los datos

Esta fase es una de las más importantes, se construye los primeros modelos de clasificación de la lengua de señas al realizar combinaciones de modelos, teniendo en cuenta el buen manejo de los datos, así como la implementación de técnicas de regularización para obtener modelos óptimos que eviten el sobre entrenamiento.

1.3.5 Evaluación del modelo

Este paso consta del uso de diferentes métricas como validación cruzada, matriz de confusión, exactitud entre otras para la evaluación de los modelos, se tiene en cuenta el estado del arte para realizar las comparaciones en mismas condiciones de evaluación.

1.3.6 Despliegue

Consiste en preparar el mejor modelo obtenido para desarrollar una aplicación, se hace uso de librerías de código abierto donde se crea una aplicación de fácil uso y acceso, cuyo objetivo es lograr interpretación a tiempo real.

1.4 Contribuciones

- Jader Alejandro Muñoz Galindez y Rubiel Vargas Cañas ponencia en “Encuentro Internacional de Ciencia para la Paz y el Desarrollo” en Popayán – Cauca el 30 de noviembre del 2022, premiada como segunda mejor ponencia del encuentro.

- Jader Alejandro Muñoz Galindez y Rubiel Vargas Cañas, artículo titulado “Modelo de interpretación de lengua de señas colombiana usando inteligencia artificial” sometido a la Revista de investigación, desarrollo e innovación RIDI, catalogada categoría B en Publindex.

2. Marco Teórico

2.1 Marco Conceptual

2.1.1 Personas en situación de discapacidad

La organización mundial de la salud (OMS) define a las personas en situación de discapacidad, a aquellas que tengan deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, al interactuar con diversas barreras, puedan impedir su participación plena y efectiva en la sociedad, en igualdad de condiciones con las demás [1]. Se estima que la población con alguna discapacidad a nivel mundial es de alrededor del 15% y según estimaciones de ministerio de salud para agosto del 2020 en Colombia, hay cerca de 1,3 millones de personas presentaba alguna discapacidad.

2.1.2 Personas en situación de discapacidad auditiva

La discapacidad auditiva es la pérdida o anormalidad de la función anatómica y/o fisiológica del sistema auditivo [11], teniendo como consecuencia la perdida de la audición en medida parcial o total por diversas razones tales como genéticas, infecciosas, ocupaciones, traumáticas, envejecimiento, entre otras. Según el grado de disminución de la audición se puede clasificar como; sordo a aquel que no posee audición suficiente y que no puede sostener una comunicación y socialización natural y fluida en lengua oral e hipoacúsica; a aquel que por alguna razón se ve afectado su audición de forma negativa llegando a perdidas leves o profundas de su audición [3].

2.1.3 Lengua de señas colombiana - LSC

Es la lengua natural de la población sorda, la cual forma parte de su patrimonio cultural y es tan rica y compleja en gramática y vocabulario como cualquier lengua oral. Como cualquier otra lengua tiene su propio vocabulario, expresiones idiomáticas y gramáticas diferentes a las del español. Los elementos de esta lengua -las señas individuales-, son la configuración, la posición y la orientación de las manos en relación con el cuerpo y con el individuo, la lengua también utiliza el espacio, dirección y velocidad de movimientos, así como la expresión facial para ayudar a transmitir el significado del mensaje [12]. En Colombia, esta se remonta al año 1920, en un internado católico bogotano. Apareciendo en el año 1957 la primera asociación de sordos en Bogotá y un año después en Cali. Los sistemas de señas recibieron influencia de la lengua de señas española, a través de inmigrantes o sordos colombianos educados en España alrededor de los años 50 [13]. Hoy en día, el lenguaje de señas colombiano hace parte del patrimonio inmaterial, cultural y lingüístico de Colombia, declarado por el ministerio de cultura a través de la Ley 982 de 2005, con lo cual, se garantiza su preservación y divulgación.

Ahora, con la Ley estatutaria 1618 de 2013 se garantizan los derechos a personas con discapacidad, mediante medidas de inclusión, acciones afirmativas y ajustes razonables, encargando su implementación a las entidades públicas de nivel nacional, regional y local. Por su parte, se les exige a las entidades privadas a promover, difundir, respetar y visibilizar el ejercicio efectivo de todos los derechos de las personas en condición de discapacidad bajo el velo del consejo para la inclusión de la discapacidad.

2.1.3.1 Reglas del LSC

En la lengua de señas existen diferentes normas de uso para ejecutar las señas, dactilológicamente, se procura que las señas se realicen con la mano dominante con la palma siempre al frente evitando movimiento bruscos pero que tracen correctamente las señas, pero en general, cuando es requerido el uso de ambas manos para ejecutar señas, se debe realizar con la seña con la mano dominante y la otra mano recibir la seña como soporte, además, se debe considerar un espacio tridimensional desde la cintura a la cabeza donde se realiza la ejecución evitando tapar la boca expresando en este mismo espacio las señas asignadas al discurso (Figura 2), expresiones faciales y corporales acordes al contexto. También, La estructura gramatical del LSC que tiene cuatro condiciones que hacen que se diferencie del lenguaje verbal, estas son: la presencia de verbos en presente infinitivo, no conjugación de verbos, la no utilización de artículos ni conectores y el diferente orden ante una misma oración [14].

Figura 2: Reglas de la lengua de señas
Fuente: Elaboración propia

2.1.3.2 Abecedario LSC

El abecedario LSC (Figura 3) consta de las 27 letras del alfabeto realizadas con una sola mano, de las cuales, seis de ellas son de carácter dinámico y 21 estáticas, pudiendo ser realizadas con cualquier mano si se respeta la forma de cada una de estas, es así, como al no tener seña propia algunas palabras es necesario realizarlas mediante el alfabeto dactilar y dar a entender las palabras respetando el orden.

2.1.3.3 Números del LSC

Los números son una expresión de las cantidades muy necesarias para una conversación en los distintos canales de comunicación, en el LSC, los numero del cero al cinco son estáticos y pueden ser realizados levantando los dedos para expresar la cantidad, sin embargo, cuando se trata de numero mayores al cinco, estos presentan movimiento y pueden ser combinación de estos para expresar cantidades intermedias (Figura 4a). A su vez, los números grandes que pueden expresar cantidades de miles y millones son realizados con ambas manos (Figura 4b).

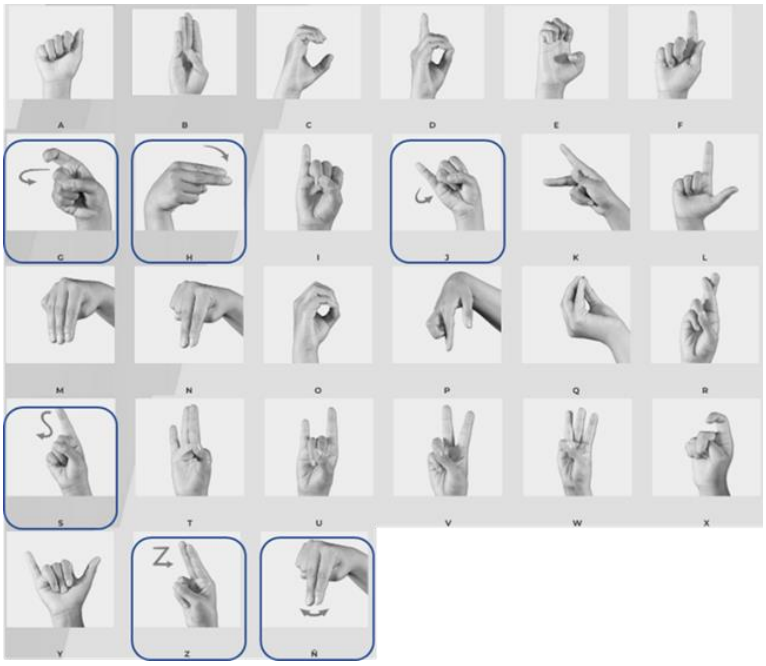


Figura 3: Alfabeto LSC, las señas en recuadro azul responden a un carácter dinámico
Fuente: [13]

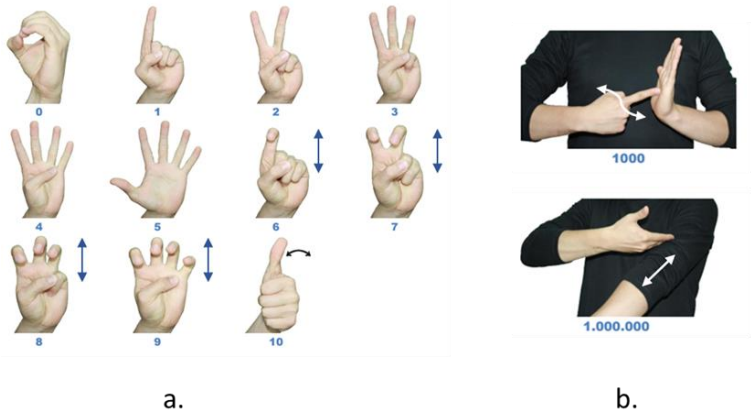


Figura 4: a) Números del cero al diez LSC b) Números mil y millón LSC
Fuente: [14]

2.1.4 Inteligencia artificial

La inteligencia artificial es un campo de la ciencia de la computación, que se dedica al diseño, modelamiento e implementación de sistemas inteligentes que automáticamente dan una respuesta sistemas complejos surgidos en el mundo real [15], sirviendo de apoyo en diversas aplicaciones informáticas, medicas, biológicas, entre otras; sin embargo, a diferencia de las personas proyectan una ventaja, ya que los dispositivos basados en inteligencia artificial no necesitan descansar y pueden analizar grandes volúmenes de información a la vez. Asimismo, la proporción de errores es significativamente menor en las máquinas que realizan las mismas tareas que sus contrapartes humanas [16]. La inteligencia artificial posee subcampos (Figura 5) y estos se vuelven más complejos respecto al tipo de información de entrada que se suministra.

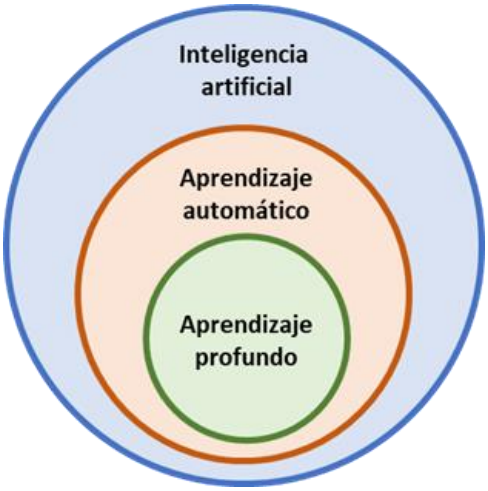


Figura 5: Inteligencia artificial
Fuente: Elaboración propia

2.1.4.1 Aprendizaje automático

El aprendizaje automático o machine learning es parte del grupo de inteligencia artificial, el cual, es una técnica aplicada en variedad de campos para la identificación de procesos donde el conocimiento no es evidente a partir de los datos para la toma de decisiones [17]. Estas técnicas se basan en algoritmos estadísticos y matemáticos para mejorar tareas tales como conceptos de aprendizaje, modelamiento predictivo, entre otras, buscando principalmente la optimización de problemas permitiendo al computador reconocer automáticamente complejo patrones y tomar decisiones acertadas a nueva información [18].

Estos algoritmos de aprendizaje automático dividen en tres tipos: el aprendizaje supervisado; que para sus predicciones es necesaria información etiquetada previamente al entrenamiento, aprendizaje no supervisado; que generaliza la información a partir de

patrones en datos no etiquetados y el aprendizaje por refuerzo; que su objetivo es lograr la mejor optimización a partir de premios y penalizaciones al agente. Realizando estas predicciones se busca tener una ventaja en el gasto computacional para realizar tareas realizadas manualmente donde el conocimiento a un problema debe ser abundante para cubrir todas las posibilidades que se puedan presentar [16].

2.1.4.2 Aprendizaje profundo

El aprendizaje profundo es un subcampo del aprendizaje automático de alto nivel usado para resolver problemas de alta complejidad y de gran cantidad de datos no estructurados, realiza abstracciones en la información mediante el uso de redes neuronales artificiales que, según su profundidad, extraen patrones que diferencian la información durante su etapa de entrenamiento. Actualmente, se utiliza en el reconocimiento de voz, el procesamiento del lenguaje natural, la visión artificial, identificación de vehículos en los sistemas de asistencia al conductor, entre otros [8].

2.1.5 Redes neuronales artificiales

Las redes neuronales artificiales son una abstracción matemática que modela la forma de procesamiento de la información en sistemas nerviosos biológicos, especialmente el cerebro humano que corresponde al de un sistema altamente complejo, no-lineal y paralelo[18]; intentan resolver estos problemas mediante el intercambio de experiencias pasadas en problemas ya resueltos, sin embargo, su conducta no es igual al del cerebro humano que actúa como procesador de la información eficiente y de gran plasticidad que las computadoras tradicionales no poseen.

2.1.5.1 Neurona Artificial

La neurona artificial se conoce como perceptrón, cuya finalidad es la solución de problemas binarios, mediante una entrada de un vector de *x* valores *(x1,x2,…xm)* que son características del sistema a clasificar y produce una salida binaria 0 o 1, matemáticamente se define:

$$f(x) = \begin{cases} 1, & wx + b > 0 \\ 0, & en\ caso\ contrio \end{cases} \quad (1)$$

Sin embargo, en la solución de problemas no lineales, es requerido la composición de más neuronas, pasando la suma de pesos por una función de activación que introduce no linealidades, con el objetivo de clasificar la información dividiendo los múltiples hiperplanos. Esta composición es conocida como perceptrón multicapa (MLP) (Figura 6), que en este caso las entradas y salidas son visibles desde afuera, mientras que las capas del medio se encuentran ocultas [19].

2.1.5.2 Funciones de activación

Las funciones de activación permiten introducir no linealidades en los pesos de cada neurona, permitiendo que el aprendizaje se adapte poco a poco reduciendo progresivamente los errores cometidos por la red [19], estas son de diversos tipos (Tabla 1) y su uso depende del tipo de capa en el que son aplicadas.

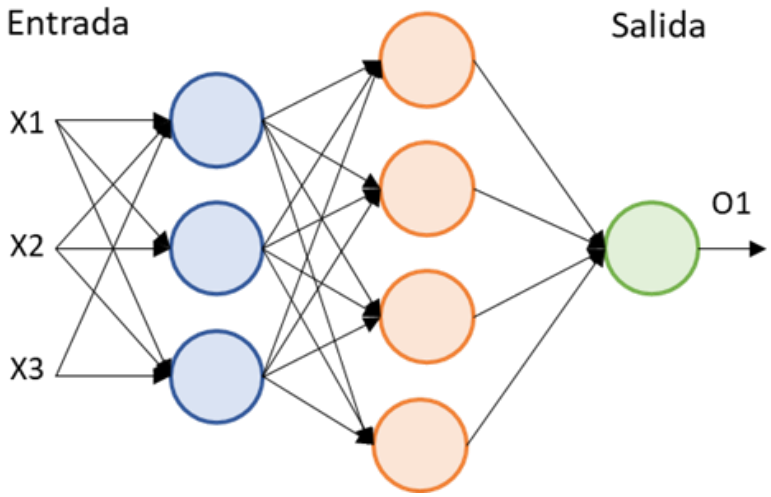


Figura 6: Red neuronal multicapa
Fuente: Elaboración propia

Tabla 1: Funciones de activación

Función de activación	Expresión matemática
Identidad	$f(a) = a$
ReLu	$\max(0, x)$
SoftMax	$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$
Sigmoide	$f(x) = \frac{1}{1 - e^{-x}}$
Tangente hiperbólica	$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$
Gaussiana	$f(u_i) = ce^{-\frac{u_i}{\sigma}}$

2.1.6 Redes neuronales convolucionales CNN

Las CNN es un tipo de modelo del aprendizaje profundo para el procesamiento de información que tiene patrón de cuadricula, tales como imágenes, está inspirado en la organización de la corteza visual animal y diseñado para que aprenda de forma automática y adaptativa a jerarquías espaciales de características de patrones de bajo a alto nivel [20]. Las CNN se componen de tres capas importantes que son: convolución, agrupación y conexión total (Figura 7) cuya combinación da una respuesta ante los valores de entrada de la red.

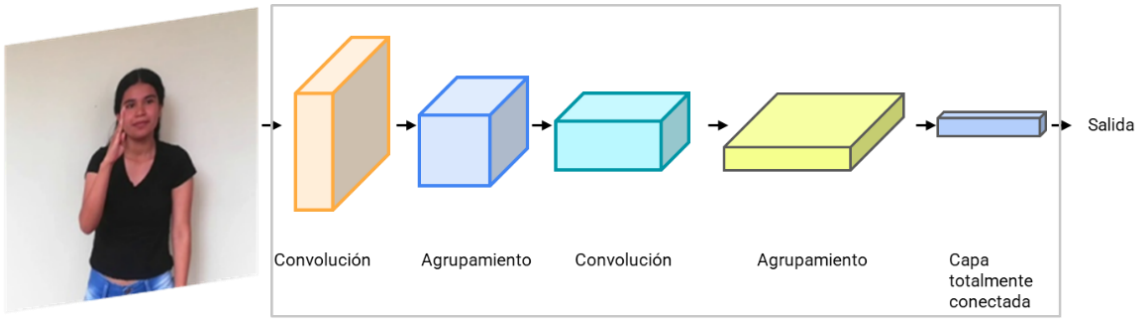


Figura 7: Red neuronal convolucional
Fuente: Elaboración propia

2.1.6.1 Capas de convolución

Consiste en un conjunto de filtros o extractores de características, aplicados mediante el producto escalar (Figura 8) a imágenes para obtener información tales como textura, bordes, entre otras. El producto de la convolución, es conocido como mapa de características y este proceso se repite para formar un numero arbitrario de mapas de características, que ayudan a generalizar la información para la clasificación de esta misma [17].

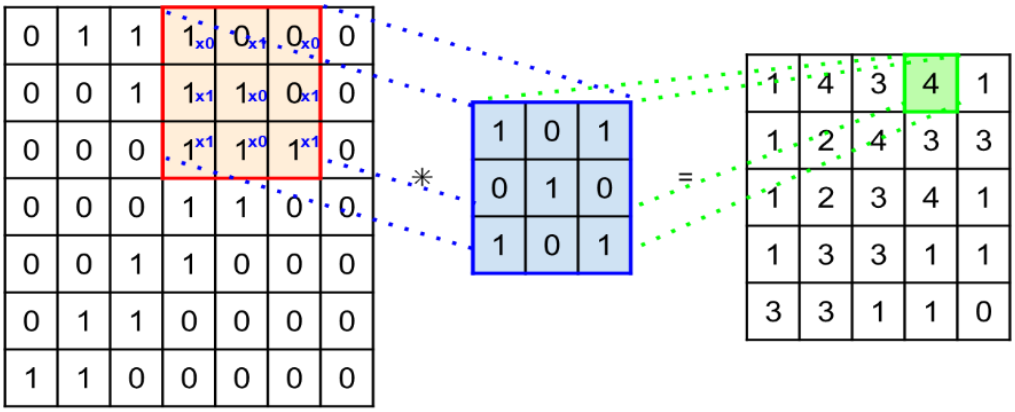


Figura 8: Capa de convolución
Fuente: Elaboración propia

2.1.6.2 Capas de agrupamiento

Las capas agrupamiento, tienen la finalidad de simplificar o reducir la dimensión espacial de la información derivada de los mapas de características, ayudando a la caracterización y ubicando rasgos predominantes en ella. Estas capas, pueden ser de diferente tipo ayudando a las necesidades de optimización del modelo, sin embargo, las más usadas son las de agrupamiento máximo y agrupamiento promedio (Figura 9), que dan un nuevo mapa de características respecto al valor y tamaño del núcleo agregado.

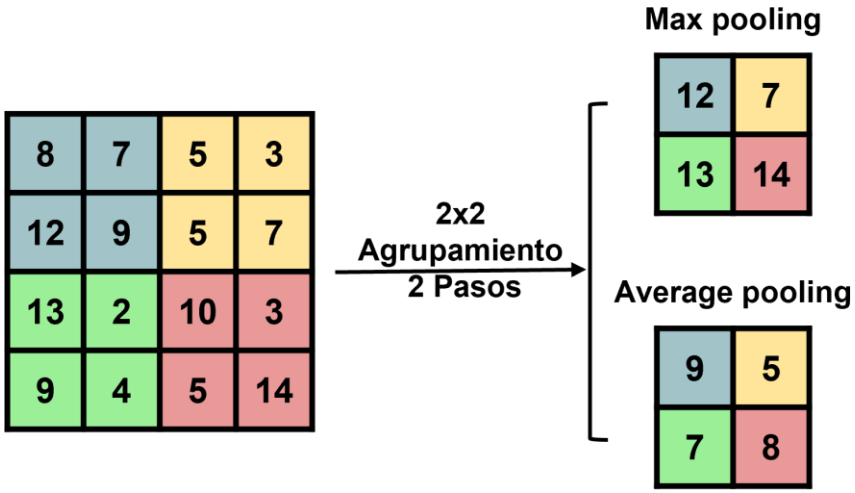


Figura 9: Tipos de capas de agrupamientos
Fuente: Elaboración propia

2.1.6.3 Capas conexión total

La salida de los mapas de características de la convolución final, son típicamente aplanados y transformados en un vector unidimensional que son conectados a una o más capas de conexión total, en otras palabras, a una red neuronal multicapa, cuya función es asignar las probabilidades para cada clase en la tarea de clasificación [21].

2.1.7 Redes Neuronales Recurrentes RNN

Las Redes neuronales recurrentes (RNN), son un tipo de redes neuronales diseñadas para reconocer patrones en secuencias de datos como pueden ser textos, genomas, escritura, palabra hablada o series temporales numéricas [22]. Se suelen indexar con valores continuos o discretos, llamados *timesteps*, con los que se accede a momentos particulares de la secuencia [23]. Son mediante estos *timesteps* que se reconoce patrones sin importar la posición que se encuentre en los datos de entrada, habilitando la posibilidad a recibir secuencias de tamaños diferentes, es así, que las estructuras de datos de entrada y salida presentan diseño uno-uno, uno-muchos, muchos-muchos y muchos-uno (Figura 10), pero estos presentaran dependencia de los datos previos a estos para realizar la predicción.

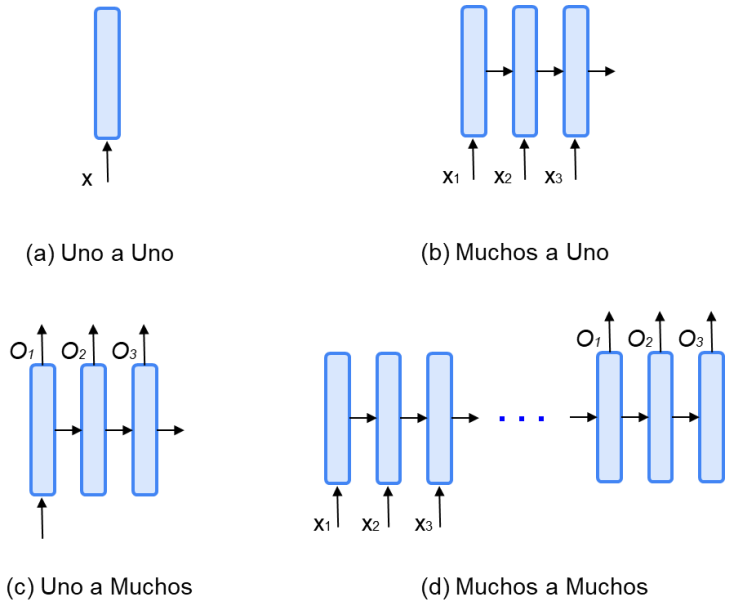


Figura 10: Tipos de RNN según entrada y salida
Fuente: Elaboración propia

La estructura de una RNN recibe este nombre al aplicar una recurrencia entre el estado previo y el actual, matemáticamente podemos definir esta relación de recurrencia como:

$$S_t = f(S_{t-1}, X_t) \tag{2}$$

Donde, *f* es una función derivable, S_t es un vector de valores llamado estado interno de la red de la etapa t y X_t es la entrada de la red en la etapa t, es asi como podemos analizar los estados anteriores con el estado S_{t-1} siendo este el resumen de los estados previos [18]. El analisis de una RNN se realiza mediante el desenvolvimiento de la red como redes de n estapas independientes (Figura 11) con sus respectivas entradas y salidas pero que comparten sus pesos con la red siguiente.

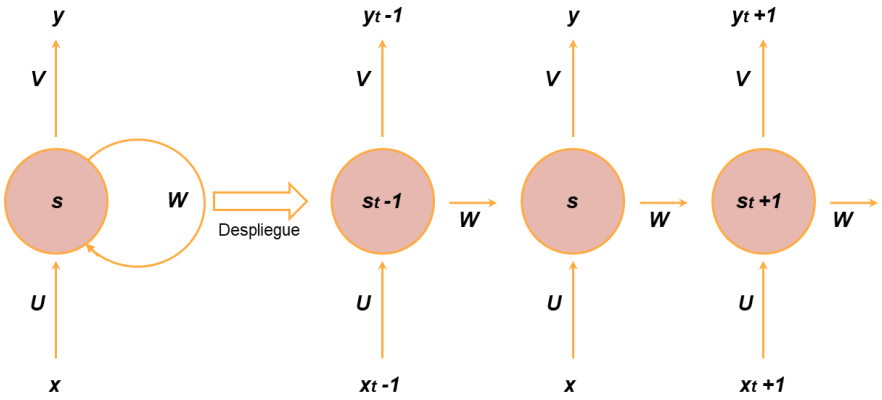


Figura 11: Red neuronal recurrente
Fuente: Elaboración propia

Ahora bien, la composición de la red presenta tres parámetros, siendo: *U,* la transformación de la entrada en la etapa t; *W,* la transformación del estado S_{t-1} al estado S_t; por último, *V,* la asignación del estado S_t actual a la salida Y_t, matemáticamente una red recurrente se puede modelar como:

$$S_t = f(S_{t-1} * W + X_t * \mathrm{U}) \quad (3)$$

$$Y_t = S_t * V \quad (4)$$

Siendo f una función de activación y Y_t la respuesta de la RNN, entonces, es notable que el compartir W del estado previo al estado actual, conlleva a presentar memoria en la red. Sin embargo, estas redes presentan un problema llamado short-term memory, derivado del bien conocido problema asociado al desvanecimiento del gradiente, problema debido al proceso matemático de la propagación hacia atrás o backpropagation [24], por esta razón, se desarrollaron otros modelos de redes neuronales que disminuyen este tipo de problemas como lo son:

2.1.7.1 Long short-term memory (LSTM)

La idea clave de las LSTM es la celda estado (en adición a los estados ocultos de la RNN) donde la información puede ser explícitamente escrita o removida para que el estado permanezca constante sino hay interferencia externa [17]. Estas se componen de tres compuertas llamadas, puerta de olvido, puerta de actualización y puerta de salida, cuyas entradas y salidas son vectores que se conforman por una red neuronal, una función de activación softmax y un multiplicador (Figura 12).

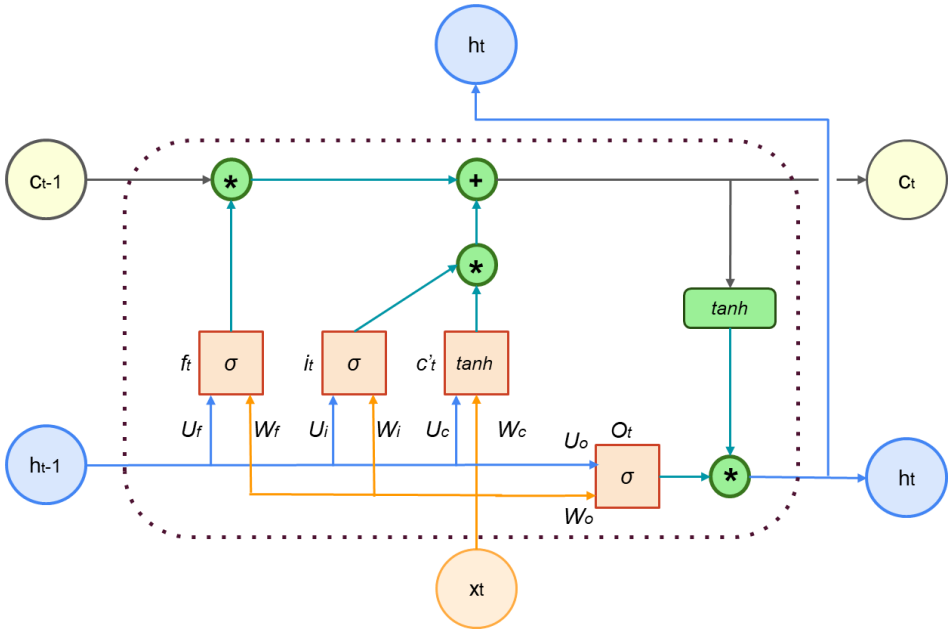


Figura 12: Componentes de la LSTM
Fuente: Elaboración propia

Ahora bien, el diagrama de la LSTM en la parte superior es la denominada celda de estado c, que representa la memoria de la unidad y esta es alterada únicamente por interacciones lineares, por lo que al no sufrir ninguna transformación mayor la información nunca es alterada de gran manera en una sola iteración, permitiéndole recordar información del pasado con más facilidad [15] y la parte inferior es el estado oculto h que es una versión filtrada de la celda de estado.

Matemáticamente, se analiza el comportamiento de una celda LSTM compuerta a compuerta empezando por la puerta de olvido, que basa su decisión de salida de la celda previa h_{t-1} y la entrada común x_t, cuyo resultado depende de la función de activación softmax, determinando la información descartable que no pasara a la celda estado.

$$f_t = \sigma(W_f X_t + U_f\, h_{t-1}) \tag{5}$$

Ahora, la actualización de la información en la celda de estado se realiza en dos partes, considerando la entrada común y el estado oculto anterior. El input gate decide la información que se va a agregar y este se filtra con la probabilidad de los candidatos a hacer agregados a la memoria.

$$i_t = \sigma(W_i X_t + U_i\, h_{t-1}) \tag{6}$$

$$C'_t = tanh(W_c X_t + U_c\, h_{t-1}) \tag{7}$$

Estos dos elementos son agregados a la celda de estado decidiendo que nueva y vieja información serán incluidos en la memoria.

$$C_t = f_t * c_{t-1} \oplus i_t * C'_t \tag{8}$$

El nuevo estado oculto, será una filtración de la información de entrada y el estado oculto anterior con la celda de estado.

$$O_t = \sigma(W_o X_t + U_o\, h_{t-1}) \tag{9}$$

Finalmente, el estado oculto anterior es calculado mediante la multiplicación de O_t con la función de transferencia tangente hiperbólico de la memoria de la red.

$$h_t = o_t * \tanh(C_t) \tag{10}$$

2.1.8 Mediapipe

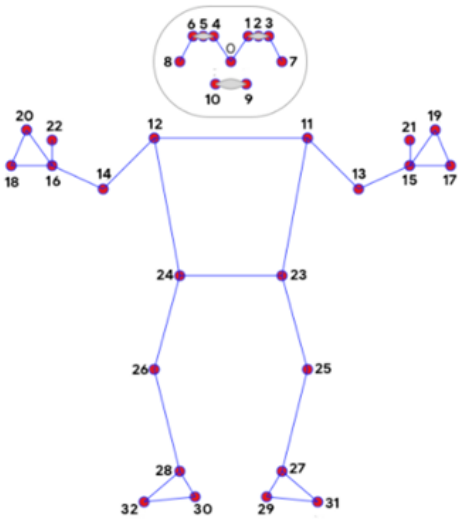
Mediapipe es un framework para construir canalizaciones de interferencias sobre datos sensoriales artificiales,el cual es diseñado con aprendizaje automático [25], para realizar implementaciones rápidas de detección del cuerpo, logrando crear prototipos de rápido despliegue.

2.1.8.1 Mediapipe Pose

Usando un detector, la canalización localiza primero la región de interés (ROI) de la persona/postura dentro del marco. Posteriormente, el rastreador predice los puntos de referencia de la pose y la máscara de segmentación dentro del ROI usando el marco recortado del ROI como entrada. En casos de video, el detector, solo es invocado cuando es necesario tomando como referencia el primer cuadro y cuando no se detecte el cuerpo en el fotograma [26]. Los 32 puntos detectados (Figura 13) sobre el cuerpo pueden ser seleccionados de manera preferente y según las condiciones.

2.1.8.2 Mediapipe Hands

Es una herramienta que realiza seguimiento de manos y dedos de alta fidelidad. Emplea el aprendizaje automático (ML) para inferir 21 puntos (Figura 14) de referencia de una mano, en un espacio tridimensional a partir de un solo cuadro, emplea una canalización de aprendizaje automático que consta de varios modelos que trabajan juntos: un modelo de detección de palma que opera en la imagen completa y devuelve un cuadro delimitador de mano orientado y un modelo de punto de referencia de la mano que opera en la región de la imagen recortada definida por el detector de la palma y devuelve puntos clave de la mano [27]. Los puntos inferidos corresponden a la palma y falanges en los dedos.



0. Nariz
1. Ojo izquierdo interno
2. Ojo izquierdo
3. Ojo izquierdo externo
4. Ojo derecho interno
5. Ojo derecho
6. Ojo derecho externo
7. Oreja izquierda
8. Oreja derecha
9. Boca derecha
10. Boca izquierda
11. Hombro izquierdo
12. Hombro derecho
13. Codo izquierdo
14. Codo derecho
15. Muñeca izquierda
16. Muñeca derecha
17. Meñique izquierdo
18. Meñique derecho
19. Índice izquierdo
20. Índice derecho
21. Pulgar izquierdo
22. Pulgar derecho
23. Cadera izquierda
24. Cadera derecha
25. Rodilla izquierda
26. Rodilla derecha
27. Tobillo izquierdo
28. Tobillo derecho
29. Tacón izquierdo
30. Tacón derecho
31. Índice pie izquierdo
32. Índice pie derecho

Figura 13: Puntos característicos de mediapipe pose
Fuente: [26]

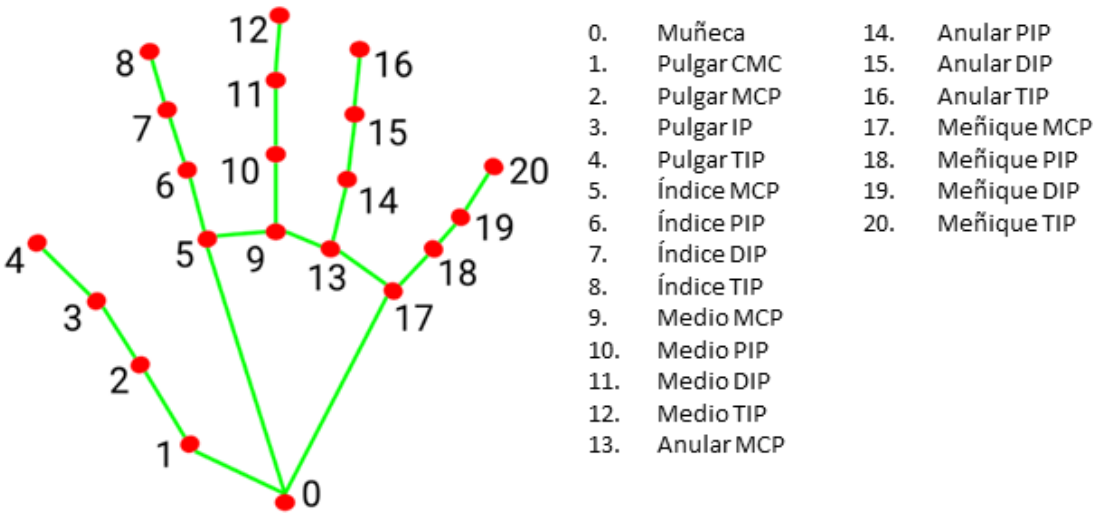


Figura 14: Puntos característicos mediapipe hands
Fuente: [27]

2.2 Marco referencial

En esta sección, se presenta un resumen de los principales trabajos identificados en la revisión de la literatura. Los trabajos fueron seleccionados teniendo en cuenta las siguientes características: Tamaño y variabilidad de la base de datos en condiciones no controladas y desempeño dinámico, extraídos de diferentes bases de datos como la IEEE Xplore, Scielo, entre otras. La organización de los referentes bibliográficos se realiza de nivel externo a interno, teniendo el orden de referentes internacional, nacional y regionales.

2.2.1 Referentes internacionales

- **Reconocimiento de la lengua de señas americano mediante extracción de características de estimación de pose de la mano [28]**

Esta investigación realiza clasificación de la lengua de señas estadounidense (ASL) usando un algoritmo de detección de manos mediante la variante mediapipe hand, realizando la estimación en imágenes RGB que se obtienen directamente de una cámara web. La disposición de este algoritmo lleva a detectar 21 puntos sobre las manos, cuyas coordenadas tridimensionales están relacionados con el tamaño de la imagen. Estos puntos en la imagen, pueden ser interpretados como vectores y se obtienen dos descriptores que corresponde a la distancia entre articulaciones y ángulos entre vectores 3D. Con la obtención de 210 vectores se crean 210 descriptores de distancias además de 630 ángulos que alimentan el modelo de inteligencia artificial. Se realiza pruebas en tres repositorios estáticos de ASL que presentan distintas características, el primero denominado ASL Alphabet dataset, que posee un total de 780.000 imágenes que corresponden a 3.000 por cada clase donde las imágenes son capturadas en sombra dificultando su reconocimiento, el segundo, llamado Massey dataset, que posee 1.815 imágenes, 65 muestras para la clase t y 70 para las otras 25 clases, cuyas capturas tienen una buena distinción entre el fondo

oscuro y el color de mano y finalmente Finger Spelling A Dataset, un repositorio de 65.774 imágenes en RGB y profundidad correspondiendo a 2.615 a 3.108 muestras, donde las clases dinámicas J y Z no son tenidas en cuenta, además, son capturadas en un ambiente no controlado. Estas imágenes, alimentan a algoritmos de clasificación multiclase llamado máquina de soporte de vectores, donde su núcleo tiene implícito la distancia euclidiana y una variante de árboles de decisión, denominado máquina de aumento del gradiente de luz. Este modelo es entrenado usando la métrica de accuracy y se utiliza validación cruzada de cinco divisiones, entregando resultados favorables para la clasificación de estas señas, obtienen un accuracy de 99,39% para Massey dataset, 87,60% para ASL Alphabet dataset, y 98,45% para Finger Spelling.

- **Reconocimiento de lengua de señas americano usando aprendizaje profundo** [29]

Este articulo realiza una comparación, entre la estimación de poses captadas con cámaras de profundidad y el reconocimiento a través de imágenes RGB, empleando aprendizaje profundo. Este repositorio cuenta con el alfabeto que ha sido dividido en 41.258 imágenes para el entrenamiento y 2.728 para prueba, donde cada muestra contiene una imagen RGB, un mapa de profundidad y una máscara de segmentación, toda de dimensión de 320x320 pixeles. Esta base de datos, se normaliza y redimensiona a un tamaño de 244x244 pixeles, alimentando una red neuronal convolucional preentrenada llamada SqueezeNet, que ha sido modificada, removiendo las capas superiores y agregando dos capas densas seguidas de una función softmax, que brinda la probabilidad de interpretación, además, como técnica de regularización se usa un proceso de aumento de datos que agrega más muestras de entrenamiento a la red neuronal. El modelo ha sido entrenado usando NVIDIA K80 GPU durante diez épocas y obtiene un accuracy en datos de prueba del 83,29%, mostrando que la red ha aprendido características del ASL, permitiendo la predicción en tiempo real como también el ser almacenado en un dispositivo móvil.

- **Reconocimiento de deletreo dactilar de lengua de señas en tiempo real empleando redes neuronales convolucionales de mapas de profundidad** [9]

En este trabajo se realiza clasificación de señas de 31 clases entre alfabeto y números usando una base de datos que contiene información de datos de profundidad; esta colección de datos cuenta con 31.000 muestras de mapas de profundidad de resolución 320x240, el dataset contiene 1.000 imágenes por cada clase que fueron realizadas por cinco sujetos de prueba, sin embargo se excluyen las señas que requieren información temporal como lo es la J y Z, las imágenes de profundidad son obtenida mediante condiciones controladas para disminuir condiciones de ruido que afecten la buena toma de los datos, además, se redimensiona la imagen a un tamaño de 227x227. La información recolectada alimenta una red neuronal convolucional preentrenada conocidas como CafeNet equivalente a AlexNet, esta arquitectura posee cinco capas de convolución, tres capas de agrupamiento máximo y tres capas totalmente conectadas, que son las encargadas de realizar la clasificación de las características y obtener la probabilidad de interpretación. El modelo es entrenado con épocas fijas de 8.000 y 4.000 para

reentrenamiento y sintonización fina respectivamente, en un equipo con tarjeta gráfica Nvidia GeForce GTX Titan y finalmente obteniendo un accuracy de del 83,58% a 85,49% con un retraso en la predicción de una imagen de tres milisegundos.

- **Una red neuronal convolucional de ocho capas con agrupación estocástica, normalización por lotes y abandono para el reconocimiento de deletreo manual de la lengua de señas china** [30]

Este estudio propone el reconocimiento de la lengua de señas china, mediante el uso de una red neuronal convolucional, combinando distintas técnicas para favorecer la interpretación de las señas. La colección de información posee 30 categorías que incluyen 26 letras básicas monosílabas y cuatro letras doble silabas, que fueron realizadas por 33 voluntarias; obteniendo 1.320 capturas, las cuales, han sido normalizadas y redimensionadas a un tamaño de 256x256, además, a estos datos se les aplico la técnica aumentación de datos, usando el 80% del repositorio para tal función, con características de escalado de 0,7 a 1,3, inyección de ruido, traslaciones aleatorias, correcciones gamma, transformación afín, y PCA aumentación de color, logrando aumentar el dataset por cada imagen original a 180 nuevas imágenes, con lo cual, el dataset ahora se compone de 191.136 imágenes. Esta información alimenta una red neuronal convolucional de ocho capas, divididas en seis capas convolucionales y dos capas totalmente conectadas con técnicas de regularización, además, de las variaciones en la capa de agrupamiento probando diferentes algoritmos para la obtención de la siguiente capa convolucional. El accuracy obtenido en datos de prueba durante el promedio de diez experimentos, arrojan resultados de 89,32 ± 1,07 con agrupamiento estocástico, siendo el mejor modelo para el reconocimiento de las características del alfabeto del lenguaje de señas chino.

- **Herramienta dinámica para la interpretación de deletreo dactilar de la lengua de señas** [10]

Esta investigación, propone el reconocimiento del alfabeto de la lengua de señas estadounidense en un enfoque dinámico, usando una base de datos compuesta por videos en condiciones semicontroladas. La extracción de cuadros importantes que caracterizan la señas, alimentan una red neuronal convolucional cuyo dataset tiene un tamaño de 80Mb, que incluyen 32.400 imágenes en 26 etiquetas, que han sido extraídas a un cuadro por segundo, convertidas a escala de grises y mediante técnicas de preprocesamiento tales como filtrado de piel y detección de bordes Canny, que sirven para extraer primeras características de las señas, además, de obtener descriptores visuales locales mediante el uso el uso Transformación de características invariante de escala (SIFT). La validación de este modelo, se realiza usando la técnica de validación cruzada con unas 5.000 imágenes y obtienen un accuracy del 98,66%, mediante este modelo se construye una transmisión en vivo para realizar la detección y prueba de las señas.

2.2.2 Referentes nacionales

- **Reconocimiento del abecedario de la lengua de señas Colombiana con Redes Neuronales Convolucionales** [8]

Este trabajo, diseña un método de reconocimiento de señas estáticas del abecedario de la lengua de señas Colombiana (LSC), realiza una metodología que combina arquitecturas de aprendizaje profundo y técnicas de procesamiento de imágenes. En primer lugar, se realiza una recolección de información con las manos de doce voluntarios, en ambientes no controlados y cámaras de dispositivos móviles, así como, videos disponible en YouTube, realizando las 21 letras estáticas del alfabeto [a, b, c, d, e, f, i, k, l, m, n, o, p, q, r, t, u, v, w, x, y], en las cuales, se elimina el fondo y realiza una técnica de regularización denominado aumentación de datos. Las imágenes que conforman la base de datos, son redimensionadas a una resolución de 32 x 32 pixeles y trabajadas en el espacio de color YCrCb; el conjunto de datos se conforma finalmente por 2.364 imágenes, que se dividen en una proporción de alrededor de 80% entrenamiento y 20% prueba, que corresponde a 1.875 y 489 respectivamente; alimentando una red neuronal convolucional, que posee tres capas de convolucionales de núcleo 3x3 y función de activación ReLu. Este procedimiento se realiza durante 100 experimentos y se obtiene la media de los resultados de accuracy, usando en cada experimento la estrategia de early stopping, el cual, procede a parar el entrenamiento alrededor de las 25 épocas. Los resultados obtenidos poseen un accuracy de 79,2% para las 21 clases, siendo capaz de reconocer características de forma y figura de la mano, en trabajos futuros se tiene la pretensión de implementar en tiempo real para realizar traducción las señas a texto incluyendo señas que poseen comportamiento dinámico.

- **Modelo computacional para reconocimiento de lengua de señas en un contexto colombiano** [31]

El desarrollo de esta investigación, implementa un software de reconocimiento de la lengua de señas Colombiana, el cual, prioriza los gestos donde parte de las manos se relacione con el torso o incluso partes de la cabeza, bajo este criterio, se construye un repositorio con 22 diferentes gestos realizados por cinco personas, entre las cuales se encuentra un intérprete de LSC. El repositorio cuenta con 3.168 imágenes a blanco y negro de resolución 640x380 pixeles, obtenidas de cuadros cada 3,3 ms en condiciones semicontroladas; las imágenes son redimensionadas a una resolución de 320x240 pixeles y divididas en un 70% de manera estratificada alimentando una red neuronal convolucional, que mediante las técnicas de Grid Search y validación cruzada, busca los mejores parámetros de red y verifica la eficiencia del modelo respectivamente. Obteniendo el mejor modelo, se emplea la técnica de transfer learning para ahorrar tiempo y mejorar resultados de previos entrenamientos. En la evaluación con los datos de prueba, se obtiene un accuracy del 68%, siendo el modelo capaz de reconocer las señas de LSC, sin embargo, los resultados son afectado drásticamente por condiciones de iluminación; con este modelo se construye una aplicación web, que ayuda a facilitar la evaluación con datos de prueba y como trabajo futuro, se pretende agregar un sistema de procesamiento natural del lenguaje que permita mejorar la semántica de la comunicación.

- **Método automático de Reconocimiento de la Lengua de Señas Colombiana para vocales y números del cero al cinco utilizando SVM y KNN** [5]

Esta publicación, presenta el reconocimiento de vocales y números del cero al cinco de la lengua de señas colombiano, basando su metodología en seis pasos, iniciando con la adquisición de imágenes usando una cámara digital Fujifil Finepix camera(S4800), capturando tres fotografías en tres perspectivas diferentes y cuya suma total es de 3.324 imágenes en tamaño de 4.608x2.592 pixeles. En segunda etapa, realizan el procesamiento de los datos, redimensionando las imágenes a una resolución 461x260 pixeles, siendo este el diez por ciento de la imagen original y cambiando el espacio de color RGB por el denominado YCbCr; sobre este espacio de color, se establecen rangos de segmentación de las imágenes según el color de piel y mediante la tercera etapa, denominada extracción de características, extraen información relevante de las imágenes segmentadas mediante métodos de momentos HU, histogramas orientados de gradiente, características morfológicas y descriptor elíptico de Fourier. En la cuarta etapa, realizan un muestreo usando el método de validación cruzada de cinco pliegues, esta información se divide en porcentajes de 70-30%, 75-25% y 80-20%; en la última etapa, se emplea algoritmos de máquina de soporte de vectores (SVM), con núcleo de función de base radial and K-vecinos cercanos (KNN). Los resultados reflejan que los datos tienen una mejor adaptación a la combinación de descriptores HOG-EF, en una proporción de 80-20% para el algoritmo SVM y les permite obtener un accuracy del 70,0%.

- **Un sistema de reconocimiento de gestos para la lengua de señas colombiana basado en redes neuronales convolucionales** [32]

Esta investigación, realiza el reconocimiento de 24 letras estáticas del alfabeto; las imágenes son capturadas tomando en consideración la distancia a la que se realiza una conversación, en rangos 0,2 a 0,7m, generando una base de datos que cuenta con un total de 24.000 imágenes, siendo 1.000 imágenes por cada categoría que son capturadas en diferentes lugares, ángulos, distancias a cámara y condiciones de iluminación. Estas imágenes son normalizadas y redimensionadas sin tener en cuenta el aspecto radial a un tamaño de 256x256 pixeles; siendo la entrada para un modelo NASNet, en una proporción de 75% y 25% para entrenamiento y prueba respectivamente. Los resultados de la evaluación arrojan un porcentaje de accuracy del 88,0% sobre las imágenes de prueba, siendo capaz de reconocer la gran mayoría de las 24 señas.

- **Reconocimiento de la lengua de señas colombiana mediante redes neuronales convolucionales y captura de movimiento** [4]

Este articulo presenta el diseño de un modelo predictivo para el reconocimiento de la lengua de señas caracterizado por números y palabras empleadas en el sector hotelero y turístico, realizando la adquisición de datos con ayuda de un algoritmo de detección de manos sobre las cuales se dibujan 21 puntos característicos; se contó con dos personas con las cuales se capturan 1.000 imágenes por cada clase extraída en un tamaño 200x200 pixeles

obteniendo un total de 39.000 imágenes. Las pruebas del desempeño del modelo se realizaron con 50 imágenes por cada seña con alrededor de 1.100 épocas obteniendo resultados en datos de prueba similares a datos de entrenamiento obteniendo 97,6%, sin embargo, esto depende de las condiciones de las imágenes capturadas y en señas similares el porcentaje tendiendo a bajar.

- **Modelo de identificación de señas del alfabeto del lengua de señas colombiana (LSC) basado en computación inteligente** [7]

Existen diferentes métodos de reconocimiento de la lengua de señas, en base a esto, muchas de estas cuentas con ruido a la salida que hacen difícil el reconocimiento de señas, la propuesta de este articulo realiza la interpretación estática del alfabeto de lengua de señas colombiana LSC empleando cámaras Leap Motion. La base de datos ha sido obtenida mediante 38 sujetos de prueba y el reconocimiento se realizó con diferentes técnicas de aprendizaje automático entre las que se encuentran, máquina de soporte de vectores (SVM), perceptrón multiclase (MLP), bosques aleatorios (RF) y modelo de replanteos; cuyos resultados superan en un máximo del cuatro por ciento a los otros modelos, obteniendo un accuracy de 97,41%, pudiendo interpretar correctamente el alfabeto LSC.

2.2.3 Referentes regionales

- **Sistema de reconocimiento automático de lengua señas colombiano mediante Kinect y Leap Motion** [33]

Esta es una propuesta de estudio, que realiza detección de lengua de señas, como apoyo a la inclusión de personas en situación de discapacidad auditiva, mediante el uso de dispositivos ópticos de captura (Leap motion), además, de emplear plataforma 3D creada en Unity. Se creo una base de datos compuesta de 100 capturas por letras, usando las 22 letras estáticas del alfabeto. El repositorio de información se compone de 2.200 datos que tras ser preprocesados alimentan un modelo de aprendizaje automático que proporciona identificación y clasificación de los datos, utilizando una herramienta conocida como “Machine learning model builder” y cuyo modelo optimo es la base para crear el modelo 3D de reconocimiento.

- **Traductor de símbolos de alfabeto del lengua de signos colombiano al lenguaje escrito** [6]

Este trabajo de investigación consiste en el reconocimiento de lengua de señas colombiana LSC, recurriendo a las 22 letras estáticas del alfabeto con 15 voluntarios que realizaron el ejercicio de las señas. Las imágenes obtenidas, sirven para la creación del repositorio en un ambiente controlado, a las cuales, se les realiza una segmentación necesaria para la extracción de características usando método Surf (Funciones robustas aceleradas), Sift (Transformación de características invariantes de escala), momentos Zernike y momentos invariantes HU. Sin embargo, se elige un descriptor propio realizado por contornos y se redimensiona la imagen que abarca todo su contorno a un valor de 300 pixeles que a su vez están normalizados. En la clasificación se elige un modelo llamado Deformación

dinámica del tiempo (DTW) obteniendo un accuracy del 79% con desviación estándar del 6%, presentando mayor probabilidad de reconocimiento las letras A, C, D, F, H, I, K, O, P, Q, R, T, U y Y en comparación al resto letras disponibles en el alfabeto.

2.3 Vigilancia tecnológica

Las herramientas como modelos de interpretación, aplicaciones de enseñanza de la lengua de señas, ayudan a cerrar brechas de comunicación y ayudan a la inclusión de personas con discapacidad auditiva, este ha sido el objetivo de muchos investigadores que dedican su tiempo a esta tarea que presenta dificultades para realizar una interpretación completa, sin embargo, aunque estos desarrollos no modelen todas las condiciones posibles, son grandes aportes para centrar en el desarrollo de un modelo completo.

2.3.1 Aplicaciones móviles

IncluSeñas: Es una aplicación móvil gratuita presente en Google Play (Figura 15) creada para aprender lengua de señas colombiana, además, de otras lenguas de señas presentes en la región. Esta aplicación realiza interacción entre el usuario dándole lección de diferentes temas como abecedario, colores, números, alimentos, acciones, entre otras; ofreciendo la posibilidad de abrir nuevas lecciones, aprobando juegos y trivias que prueban el conocimiento aprendido.

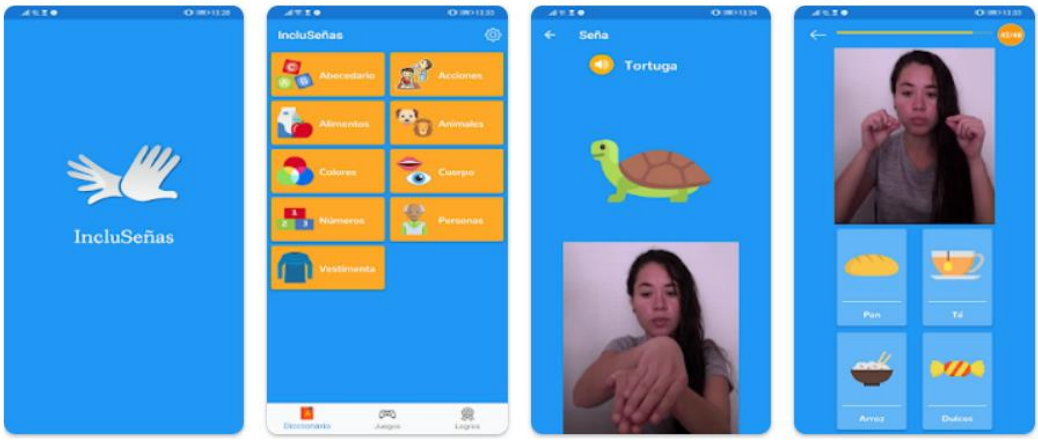


Figura 15:Aplicación IncluSeñas
Fuente: [38]

PROFEenSEÑAS: Es otra aplicación presente en Google Play (Figura 16) de forma gratuita la cual puede ser usada de modo offline, tiene 200 imágenes en 20 categorías, entre las cuales se encuentran animales, deportes, días, educación, colores, entre otras, de la lengua de señas colombiana, esta incluye gráficos de orientación para realizar las señas indicadas de tal manera que se pueda aprender de una forma rápida y ágil.

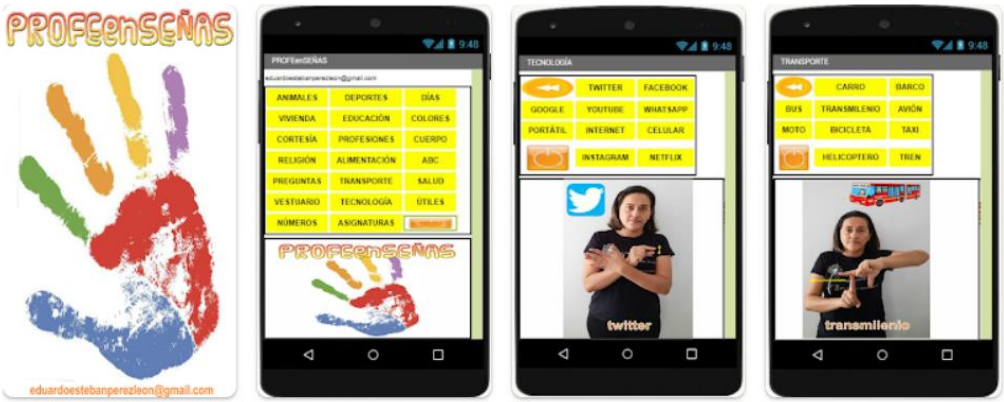


Figura 16: Aplicación PROFEenSEÑAS
Fuente: [39]

2.3.2 Páginas Web

Centro de relevo Colombia: El gobierno nacional mediante el ministerio de tecnologías de la información y las comunicaciones MinTic y apoyada por la Federación Nacional de Sordos de Colombia (Fenascol), ha desarrollado el denominado Centro de relevo Colombia, cuyo objetivo principal es atender las necesidades comunicativas de las personas en situación de discapacidad auditiva, quienes con el uso de herramientas tecnológicas, pueden interactuar con personas oyentes gracias a un servicio de interpretación en línea [34]. Estos servicios son totalmente gratuitos pudiendo ser accedidos las 24 horas, los siete días de la semana, con duración de diez a quince minutos en su aplicación o página web www.centroderelevo.gov.co. Entre los servicios ofrecidos se encuentran:

- **Servicio relevo de llamadas:** donde una persona sorda, puede solicitar una llamada a una persona oyente para dar un mensaje o solicitar un servicio.
- **Video mensajes por WhatsApp:** en el cual un intérprete de LSC grabará mensajes de máximo dos minutos y transmitirá el mensaje a la persona solicitada.
- **Servicio de interpretación en línea SIEL**, cuyo propósito es de servir de interprete en tiempo real, en una intermediación con una persona oyente que se encuentra en el mismo lugar

Diccionario básico de la lengua de señas Colombiana: En el año 2011, atraves del Ministerio de Educación Nacional y el Instituto Nacional para Sordos presentan el Diccionario Básico de la Lengua de Señas Colombiana; este instrumento lexicográfico es el primer paso a la estandarización de esta lengua, en la medida en que recoge la identidad y sentido de pertenencia de la comunidad sorda colombiana. Siendo este empleado para un aprendizaje inicial de la lengua, generando condiciones para la superación de barreras de comunicación que permitan a las personas sordas hacer uso pleno de sus derechos como ciudadanos colombianos y participar en una colectividad que cada día sea más incluyente [13].

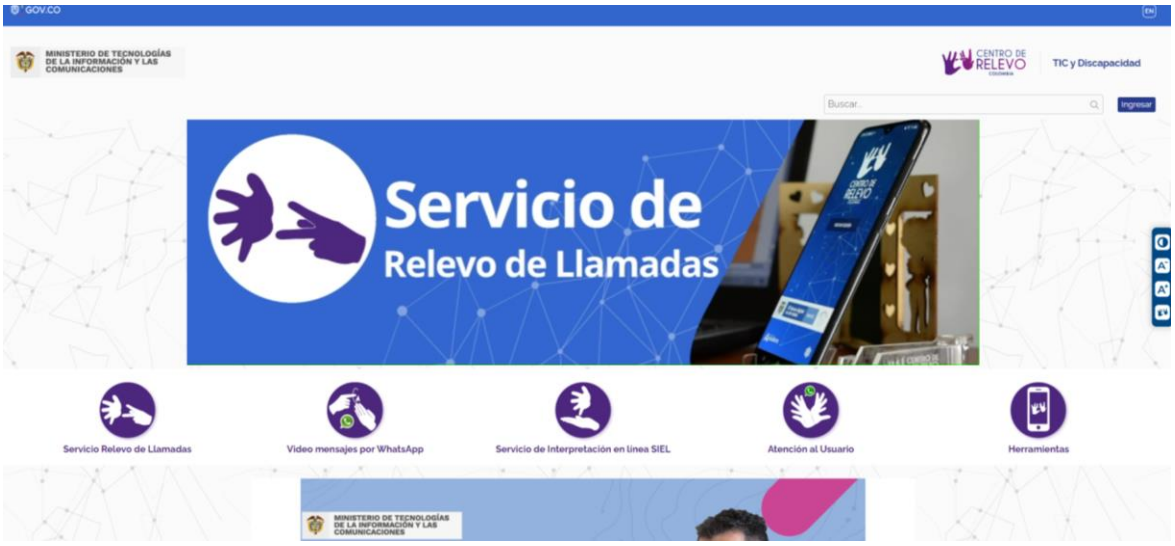


Figura 17: Pagina web Centro de relevo
Fuente: [34]

incluir

El instituto nacional para sordos INSOR: cuyo propósito, busca contribuir al mejoramiento de la calidad de la educación de las personas sordas de Colombia (Figura 18), el cual, produce recursos educativos en lengua de señas colombiana, dirigidos a estudiantes sordos, docentes, padres de familia, modelos lingüísticos, intérpretes y demás actores responsables de la atención educativa de las personas sordas del país [35]. Ha creado una base donde se puede aprender la lengua de señas colombiana, ofreciendo repositorios de contenidos cortos, clases en vivo e imágenes interactivas de las señas que componen la lengua todo esto a través de su página web www.educativo.insor.gov.co.

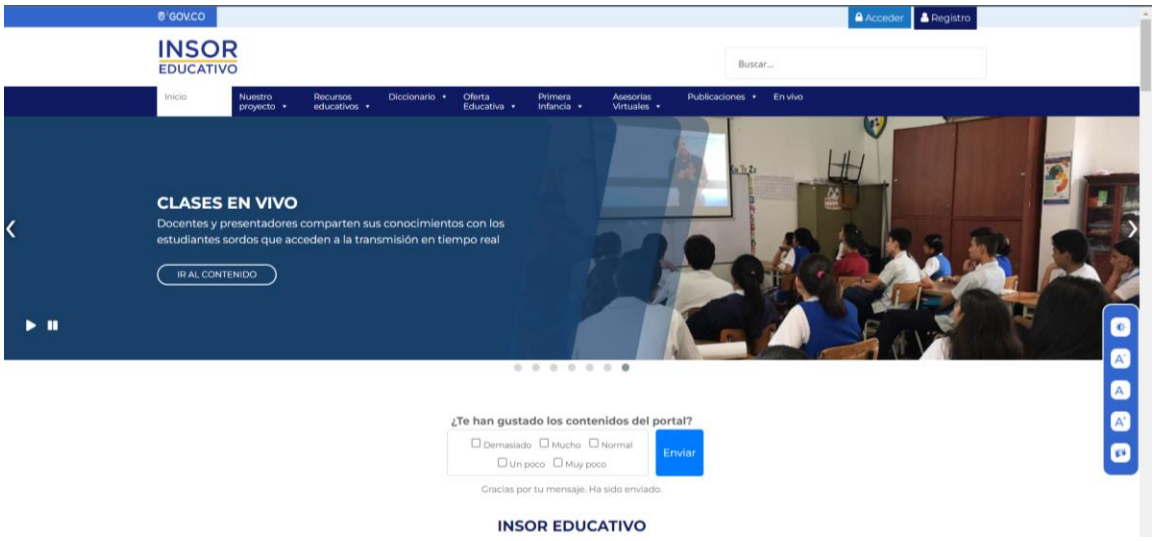


Figura 18: Pagina web INSOR
Fuente: [35]

2.4 Análisis de literatura

El siguiente cuadro comparativo muestra las diferencias entre las investigaciones comparando sus dataset, modelos elegidos y resultado obtenido en la interpretación de la lengua de señas y serán tenido en cuenta como presentes comparativos respecto al trabajo aquí presentado.

Tabla 2: Comparación de la literatura

Ref	Modelo	Contenido	Característica Dataset	País	Enfoque	Voluntarios	Tamaño repositorio	Exactitud %
[8]	CNN	Alfabeto	Estático	COLOMBIA	Mano	12	2364	79.20
[31]	Transfer learning	Palabras	Estático	COLOMBIA	Cuerpo completo	5	3168	68.00
[5]	SVM - HOG-EF	Vocales y números cero-cinco	Estático	COLOMBIA	Mano	NaN	3324	70.00
[32]	NASNet	Alfabeto	Estático	COLOMBIA	Mano	NaN	24000	88.00
[4]	CNN + PointHands	Palabras y números uno - cinco	Dinámico	COLOMBIA	Cuerpo completo	2	39000	97.60
[7]	Staking model	Alfabeto	Estático	COLOMBIA	Mano	38	NaN	97.41
[6]	Dynamic Time Warping (DTW)	Alfabeto	Estático	COLOMBIA	Mano	15	2200	79.00
[28]	SVM - LightGBM	Alfabeto	Estático	ESTADOS UNIDOS	Mano	5	2524	99.39
[29]	SqueezeNet	Alfabeto	Estático	ESTADOS UNIDOS	Mano	NaN	43986	83.29
[9]	CafeNet	Alfanumérico	Estático	ESTADOS UNIDOS	Mano	5	31000	84.56
[30]	CNN 8 capas	Alfabeto	Estático	CHINA	Mano	33	1320	89.32
[10]	CNN	Alfabeto	Dinámico	ESTADOS UNIDOS	Mano	NaN	32400	98.66

Como se puede evidenciar en la revisión de la literatura (Tabla 2), existe un déficit para realizar, interpretación alfanumérica de lengua de señas a cuerpo completo de carácter dinámico, ya que la gran mayoría de estas investigaciones, se han centrado en producir modelos de interpretación estáticos trabajando en condiciones dinámicas, en consecuencia, algunas clases son dejadas de lado o se presenta una pobre aproximación.

Además, como es evidente a nivel nacional y regional, las pocas investigaciones realizadas con intenciones dinámicas se han centrado en palabras, con bases de datos de poca variabilidad y cuya interpretación por la gran apertura de las extremidades, es sencilla de clasificar, dejando brechas al realizar una comunicación dactilar dinámica alfanumérica de la lengua de señas colombiana.

3. Metodología

Para la realización de esta investigación se hizo uso de la metodología CRISP-DM (Figura 19) que es una metodología empleada en la minería de datos; cuenta con seis pasos, los cuales, son pasos no secuenciales, pudiendo retroceder entre cada uno de ellos para realizar una mejor adaptación de los datos.

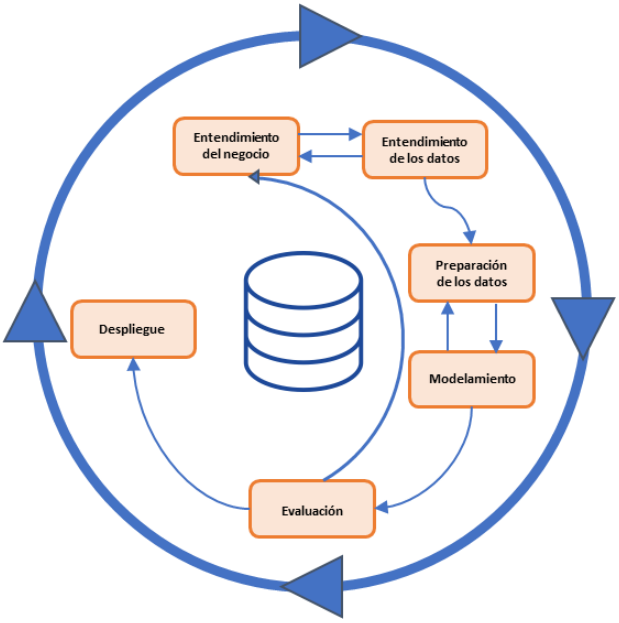


Figura 19: Metodología CRIP-DM
Fuente: Elaboración propia

3.1 Comprensión del negocio

Esta sección hace referencia a la consulta bibliográfica o estado del arte realizada durante el segundo capítulo, donde se referencia las investigaciones realizadas en otras lenguas de señas, así como conceptos necesarios para entender la importancia de la lengua de señas y de desarrollo de intérpretes dinámicos.

3.2 Entendimiento de los datos

3.2.1 Recolección de los datos

En este paso, se realizó la recolección de los datos teniendo en cuenta las normas y condiciones de la lengua de señas; se inició la recolección de los datos mediante grabaciones de participantes realizando las señas del alfabeto, números y algunas palabras del LSC. En consecuencia, se les pidió a voluntarios expertos en LSC, su consentimiento informado para ser grabados y se les instruyo en un periodo de diez minutos las señas y el procedimiento para realizar la captura de video. La grabación se realizó continuamente, empezando con las señas alfanuméricas y terminando con las palabras, con ayuda del instructor detrás de la cámara que indicaba cual debía ser la seña a realizar (Figura 20); los videos fueron capturados a través de teléfono celular Huawei P20 lite, en resolución 640x480 pixeles a una velocidad de 30 cuadros por segundo, en fondo claro a una distancia aproximadamente de dos metros sin tener control de iluminación ni de la vestimenta utilizada por los participantes.

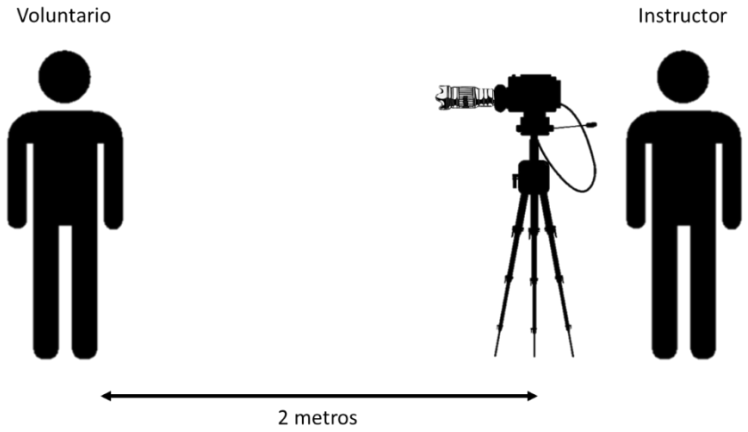


Figura 20: Grabación de voluntarios
Fuente: Elaboración propia

3.2.2 Descripción de los datos

Una vez finalizado el proceso de captura, se revisa la base recolectada y se extrae información respecto al tamaño y tiempo de duración de los videos. Se organizan respecto a su nombre por defecto y se realizan copias de seguridad en una carpeta local como digital para evitar pérdidas de la información base.

3.2.3 Exploración de los datos

En este paso se realiza un proceso de búsqueda detallada en los datos de los participantes, con el propósito de comprender las singularidades y dificultades que pudieron haber surgido durante el proceso de grabación de las señas. Además, se tomó en cuenta

la vestimenta utilizada por los participantes, ya que algunos colores o patrones pueden ser distracciones visuales o afectar la claridad de las señas.

3.2.4 Verificación de calidad

Analizados los datos obtenidos, se realizó un primer filtrado de la información, teniendo en cuenta parámetros que ayuden a mejorar la adaptación de los datos, estos deben tener características como: realización de las señas y condiciones de iluminación mínima, que, aunque se pretende realizar una buena adaptación, en algunas muestras se dificulta observar las señas, por lo tanto, estos datos son omitidos.

3.3 Preparación de los datos

3.3.1 Preparación del dataset

En esta etapa, se realiza la extracción de cuadros a partir de los videos depurados, por esa razón, mediante un algoritmo desarrollado con la librería de código abierto OpenCv; se extraen seis cuadros, considerando la ejecución de las señas dentro del primer segundo y cuyos periodos de extracción, se cuentan en 166 ms. Se dispone de su almacenamiento automáticamente, creando una carpeta para cada voluntario y dentro cuenta con subcarpetas que se disponen respecto a cada clase, del mismo modo, los cuadros extraídos se nombran teniendo en cuenta la carpeta de almacenamiento y el número de cuadro que representa en la secuencia.

3.3.2 Selección de información

Con el procedimiento anterior, se posee la información necesaria para alimentar los modelos de inteligencia artificial, sin embargo, como la información presenta tres tipos de datos, se procede a separar la información alfanumérica de la de palabras, ya que, solo dos de ellas son de interés para esta investigación. Por esta razón, se realiza la separación de estas mismas dejando de juntas las alfanuméricas separadas de la de palabras, que serán usadas en trabajos posteriores.

3.3.3 Filtrado de información

En este paso, se realiza un filtrado más general de señas similares, correspondientes a los números ceros, dos y tres, cuyas señas son equivalentes a las consonantes O, V, y W respectivamente (Figura 21). Este procedimiento resulta crucial para garantizar que el modelo pueda interpretar de manera correcta las señas, reduciendo significativamente la posibilidad de errores de interpretación y aumentando la precisión del resultado final.

3.3.4 Construcción de nueva información

Ahora, se procede a realizar una eliminación de información no relevante en las imágenes correspondientes al fondo, por esta razón, se recorta alrededor de la persona, desde la cabeza hasta debajo de la cadera, dejando la información de importancia y omitiendo la información extra no relevante. De igual manera, se redimensiona la imagen a

un tamaño de 255x255 pixeles (Figura 22) sin tener en cuenta el aspecto radial, para manejar un tamaño único en todo el conjunto de datos.

3.3.5 Integración de nueva información

En esta acción, se procede a la creación de una subbase de datos a partir de la información de cuerpo completo; se recorre todas las secuencias y se aplica el detector de pose (mediapipe pose), teniendo en cuenta la mano dominante para realizar un enfoque sobre esta. La creación de esta subbase toma el mismo nombre y formato que la base de datos original.

3.3.6 Formación de características

A partir de la información adquirida, se hace uso del detector de manos (mediapipe hands), para extraer las coordenadas sobre ambos conjuntos de datos. Esta adquisición, considera que en las imágenes donde el detector no ha encontrado los puntos característicos, deban ser rellenados valores de cero, así como, grabar los puntos anteriores sino encuentra puntos entre cuadros consecutivos de la misma clase, para completar la serie de tiempo de seis pasos.

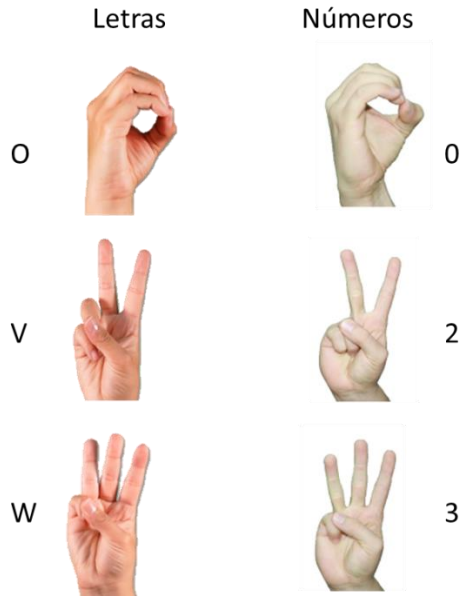


Figura 21: Similitud de letras entre números
Fuente: Elaboración propia

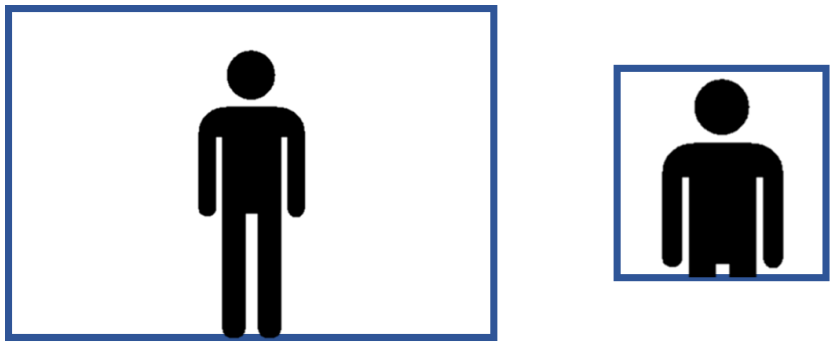


Figura 22: Recorte de imagen
Fuente: Elaboración propia

3.4 Modelamiento de los datos

En esta etapa se realiza los modelos de interpretación teniendo en cuentas las condiciones de información obtenida en procedimientos previos.

3.4.1 Técnicas de modelado

En esta sección, se tuvo en cuenta, las características de dimensión de la información de entrada, por ello, se realizo distintos modelos que pudiesen trabajar con este tipo de datos como serie de tiempo.

3.4.1.1 Modelo de coordenadas

Este modelo trabaja con la secuencia de datos, obtenidos en la extracción de coordenadas mediante el detector de mano; empleando una red neuronal recurrente doble (BILSTM), que analiza la información en doble sentido para reconocer patrones omitidos en un solo paso (Figura 23).

Figura 23: Modelo de coordenadas
Fuente: Elaboración propia

El tamaño de información que se analiza, corresponde a la cantidad de puntos elegidos en la detección de manos, por la cantidad de cuadros que componen la secuencia y cuyo modelamiento debe ser ajustado por técnicas de regularización para la generalización de nueva información.

3.4.2 Modelo de combinación

En este modelo se llevó a cabo la obtención de las características de las imágenes que componen la serie, que son extraídas mediante las CNN en sus capas de convolución, pero deben ser preparadas con el corte de la red en la capa de aplanamiento, la cual une todos los componentes multidimensionales de los tensores que conforman la red, siendo dependientes de los componentes de la red y de la profundidad de esta (Figura 24), siendo estos modelados como serie de tiempo a treves de la red BILSTM que da finalmente la interpretación de la secuencia de señas.

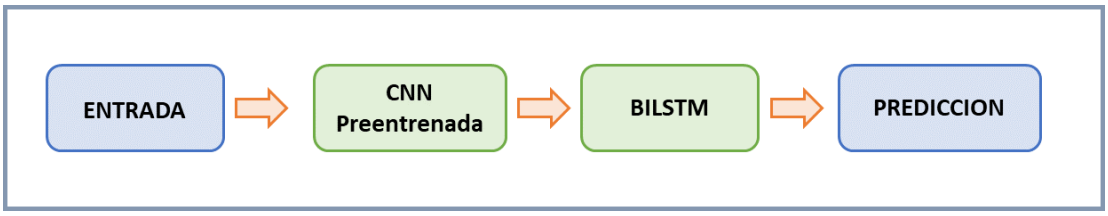


Figura 24: Modelo de combinación
Fuente: Elaboración propia

3.4.3 Normalización de los datos

Se ajustan los valores de las imágenes a un rango entre 0 y 1, dividiendo todos los pixeles entre el mayor valor de la imagen, el cual corresponde al numero entero 255, ya que, el análisis de valores tan grandes en una red convolucional tiende a un mayor gasto de computacional, distorsión en la información e incremento en las redundancias, por ello, se busca tener una escala común en todas las imágenes donde se pueda hacer análisis diferente en cada punto de información.

3.4.4 Construcción del modelo

Se comienza la construcción de modelos, realizando la obtención de características mediante detectores de posición y redes neuronales convolucionales preentrenadas. La construcción del modelo de coordenadas; posee una adquisición de 63 datos en el espacio correspondientes a puntos en los ejes X, Y, y Z, que son agrupados formando un tensor de 6x63 datos, que alimentan una red BILTSM conectada a una red neuronal multicapa, que entrega la probabilidad respecto a cada categoría (Figura 25). Este modelo es de bajo costo computacional logrando converger en pocas épocas siendo entrenado a través de CPU.

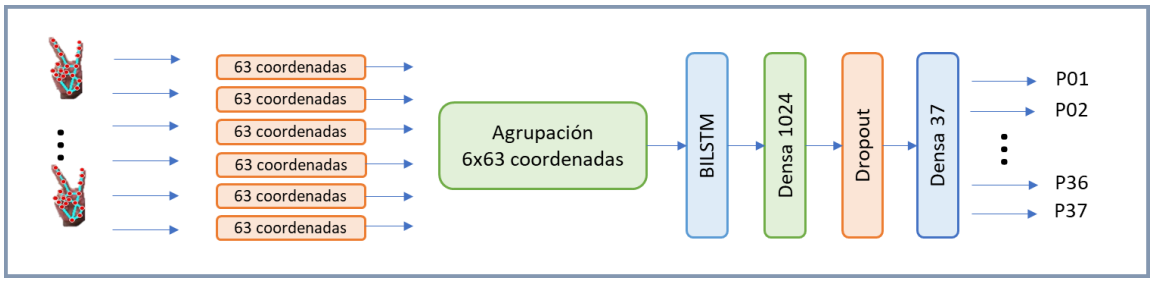


Figura 25: Composición del modelo de coordenadas
Fuente: Elaboración propia

De igual manera, se realiza la extracción de características mediante redes neuronales convolucionales preentrenadas que facilitan el entrenamiento, obtención de patrones y ajuste fino. Los cuales, gracias a previos entrenamientos, han grabado pesos en la red que son útiles, no solo en la disminución del tiempo de entrenamiento, sino también, en la adaptación a nueva información. Es por ello, que existen diferentes estructuras que pueden ser adaptadas como extractores de características con diferente cantidad de parámetros, profundidad, entre otras, que mejoran el comportamiento de redes creadas desde cero. A esta técnica se le conoce como transferencia de aprendizaje y algunos de los modelos más conocidos (Tabla 3) son probados en esta investigación.

Tabla 3: Modelos CNN preentrenados

MODELO	Total de parámetros	Cantidad de características 255x255x3	Cantidad de características 120x120x3
VGG16	14.714.688	25.088	4.608
MOBILENET	3.228.864	65.536	9.216
INCEPTION V3	21.802.784	73.728	8.192
XCEPTION	20.861.480	131.072	32.768

Finalmente esta secuencia de caracteristicas alimenta un modelo de red recurrente denominada LSTM bidireccional,que realiza el tratamiento de informacion en dos direcciones, logrando tener una mejor respuesta y detectar cambios en los datos. En general, son dos redes LSTM que leen la información en diversos sentidos y concatenan sus respuestas logrando una mejor adaptacion a los datos (Figura 26).

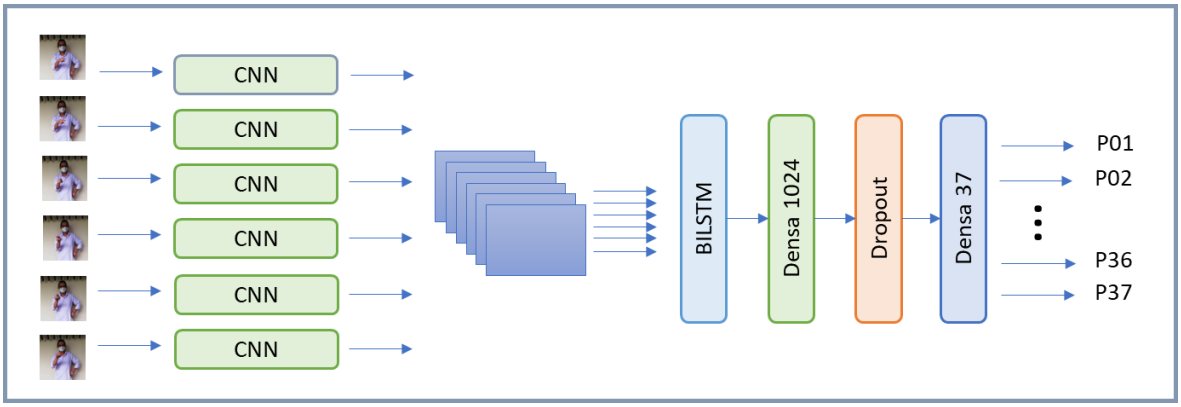


Figura 26: Composición del modelo de combinación
Fuente: Elaboración propia

Este modelo es entrenado con el uso de GPU de gran memoria RAM por la cantidad y costo computacional que conlleva la extracción de características que toman los diferentes modelos, sin embargo, algunos pueden converger en pocas épocas dependiendo de la calidad de la información suministrada.

3.4.5 Técnicas de regularización

Para disminuir la posibilidad de sobre entrenamiento y optimización de los modelos se empleó de técnicas de regularización para mejorar los modelos y convertirlos en robustos ante posibles cambios de escena, algunas de las técnicas conocidas son:

Dropout: Esta técnica en la etapa de entrenamiento desactivara aleatoriamente un porcentaje de las neuronas en capas ocultas, consiguiendo que las neuronas no memoricen los datos de entrada que es lo que precisamente sucede en el sobreajuste de la red (Figura 27).

Early stopping: La idea de esta técnica es la vigilancia continua de los parámetros de rendimiento, guardando sus parámetros en cada época, hasta que se aprecie una validación sostenida.

Aumentación de datos: Es diversificar los datos que se poseen de entrada, obteniendo nueva información creada a partir de transformaciones, a través, de la adición de ruidos, defectos, rotaciones, entre otras, logrando que la información sea ligeramente diferente pero que posea la misma esencia.

Regularización L2: También llamada regularización Ridge, modifica los pesos de la red introduciendo una penalidad en la función de coste original, cuyo objetivo, es lograr que los parámetros sean pequeños, minimizando el efecto de correlación entre los atributos de entrada, logrando que el modelo logre generalizar mejor.

$$L_2(X, w) = L(X, w) + \lambda \sum w_i \text{^} 2 \tag{11}$$

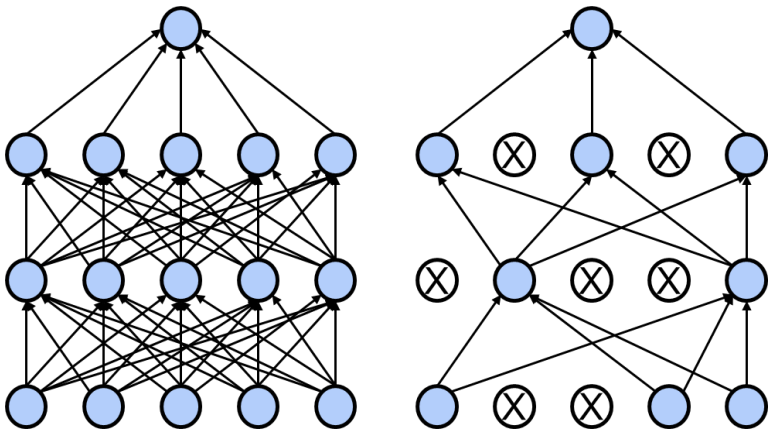


Figura 27: Regularización por Dropout
Fuente: Elaboración propia

3.5 Evaluación

En esta etapa se realiza la evaluación de los modelos a partir de los datos de prueba, probando el rendimiento de los modelos, a partir de voluntarios no visualizados previamente. Algunas de las métricas con las cuales se prueba desempeño de nuestros modelos son:

- **Exactitud:** Es una métrica que hace una descripción general del rendimiento en comparación a todas las clases

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{12}$$

- **Precisión**: Es la relación entre en número de muestras positivas clasificadas correctamente, sobre el número total de muestras clasificadas, como positivas ya sean correctas o incorrecta.

$$Pr = \frac{TP}{TP + FP} \tag{13}$$

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

$$Se = \frac{TP}{TP + FN} \tag{14}$$

- **Puntuación F1:** Es la media armónica entre la precisión y sensibilidad, donde el máximo valor es 1 (estableciendo una precisión y sensibilidad perfecta) y 0 en caso contrario.

$$F1 = \frac{2 \times Pr \times Se}{Pr + Se} \tag{15}$$

- **Matriz de confusión:** Contiene información sobre la predicción de los datos de prueba y establece que tan desajustado, se encuentra el modelo al momento de realizar clasificaciones, mostrando los aciertos y desaciertos cometidos en cada una de las categorías, donde en las columnas representa la predicción y las filas la instancia de verdadera (Figura 28).

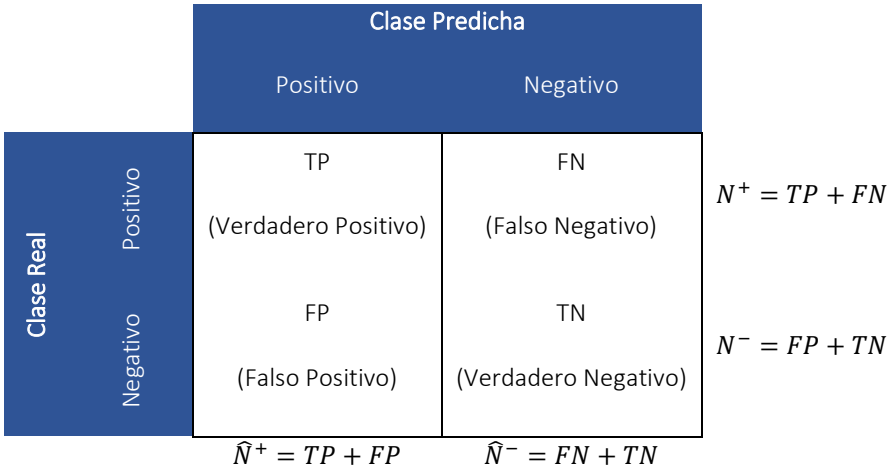


Figura 28: Matriz de confusión
Fuente: Elaboración propia

- **Validación cruzada:** La validación cruzada es un método estadístico, para evaluar y comparar algoritmos de aprendizaje, dividiendo los datos en dos segmentos: uno utilizado para aprender o entrenar un modelo y el otro utilizado para validar el modelo [36]. Esta división se realiza en k segmentos o pliegues de igual tamaño, que ayudan a validar el algoritmo mientras los otros se utilizan para entrenar el modelo, pasando al siguiente pliegue al acabar la iteración con los datos (Figura 29).

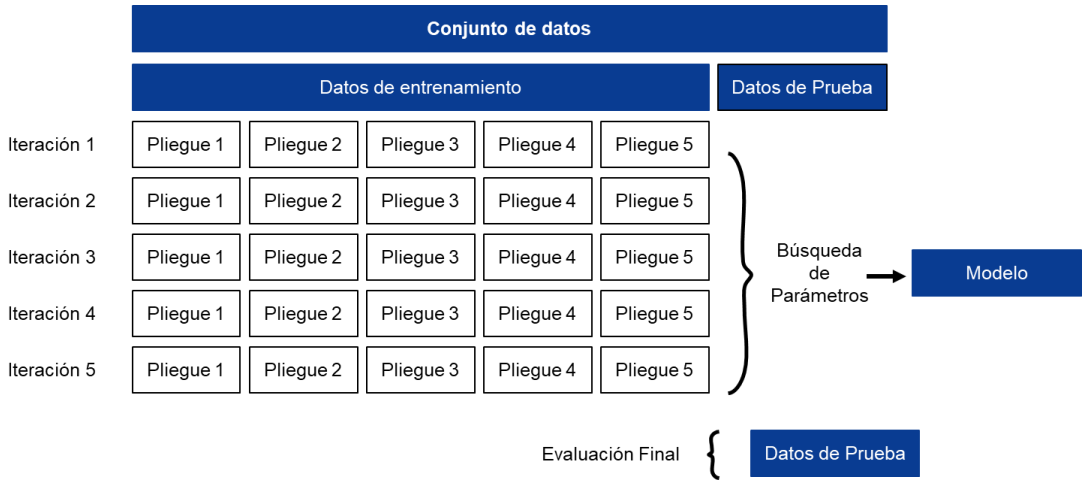


Figura 29: Validación cruzada de cinco pliegues
Fuente: Elaboración propia

3.6 Despliegue

El propósito de la fase de Despliegue es presentar el modelo final de una forma apropiada para su uso por parte de los usuarios finales.

3.6.1 Creación el entorno

Para la presentación de la plataforma se usó la librería Streamlit, librería de código abierto para Python, que facilita la creación e intercambio de aplicaciones web personalizadas, empleado aprendizaje automático y la ciencia de datos [37], que cuenta con múltiples herramientas de ediciones de interfaz , así como, la compatibilidad con diferentes módulos de Python, ayudando a integrar todo en una sola aplicación (Figura 30). Con el uso de las herramientas disponibles en la librería, se dispuso a usar las diferentes cámaras conectadas al computador, localizando esta sección en mitad de la aplicación, al igual que una sección de instrucciones que explican el funcionamiento y el tipo de clases interpretadas por la herramienta.

Figura 30: Herramientas usadas para despliegue
Fuente: Elaboración propia

3.6.2 Interpretación

Mediante la variante streamlit_webrtc, se realiza transmisión en directo en cámara y con ello realizar la interpretación a tiempo real, esta toma los cuadros específicos que son adquiridos a través de la cámara, procesados y extraída la información necesaria para para realizar la interpretación de estos mismos. El resultado del análisis de los datos, es presentado directamente sobre la misma sección de visualización, completando toda una fila de respuestas antes de limpiar la pantalla y empezar nuevamente.

3.6.3 Aplicativo Web

El aplicativo Web, debido al alto consumo de recursos, no ha sido posible alojarlo en servidores básicos de uso libre, por esta razón, se ha optó por trabajar en una máquina local con la finalidad de asegurar un buen rendimiento del aplicativo. Aunque esta solución puede parecer limitante en cuanto a accesibilidad y escalabilidad, garantiza una mayor estabilidad y rapidez en el procesamiento de datos; permitiendo realizar pruebas y mejoras de manera más eficiente, sin la necesidad de estar sujeto a las restricciones del alojamiento web.

4. Resultados y discusión

En esta sección se muestra los resultados obtenidos mediante la metodología desarrollada en el capítulo anterior y se realiza la discusión teniendo en cuenta las condiciones presentadas en el estado del arte.

4.1 Recopilación de información

En esta parte, se presentan los resultados del proceso de adquisición de datos, selección y filtrado de la información que conforman la base datos inicial.

4.1.1 Grabación

Se recolecto 77 videos sobre un fondo de color claro a una distancia aproximada de dos metros. Los voluntarios (Tabla 4), fueron instruidos durante un periodo de diez minutos sobre las señas a ejecutar, grabando continuamente en un único video por participante, empezando por las señas de alfabeto, números y por último palabras. Las condiciones de iluminación y vestimenta son variables a las cuales no se les realizo un control.

Tabla 4: Información de voluntarios

Género	Cantidad	Rango de edad	Diestros	Zurdos
Hombre	45	18-23	44	1
Mujer	32	19-29	30	2

La cantidad de voluntarios se encuentran en la edad de adulto joven, de los cuales, un porcentaje aproximado del 4.0%, tiene como mano dominante la mano zurda y el 96% la mano diestra, siendo este un dato relevante respecto a la forma y visualización de la seña.

La totalidad de clases grabadas (Tabla 5), fueron realizadas por el instructor detrás de cámara, quien daba el orden y explicaba la forma de las señas.

Tabla 5: Contenido de señas LSC grabadas

Señas LSC			
Alfabeto		Números	Palabras
A	Ñ	Cero	Hola
B	O	Uno	Buenos
C	P	Dos	Días
D	Q	Tres	Tardes
E	R	Cuatro	Noches
F	S	Cinco	Yo
G	T	Seis	Nombre
H	U	Siete	Años
I	V	Ocho	Gustar
J	W	Nueve	Licor
K	X	Diez	
L	Y	Mil	
M	Z	Millón	
N			

Todas las señas fueron elegidas para desarrollar una conversación básica en lengua de señas, en la cual, se realice una presentación personal y preguntas a partir de estas mismas; este grupo se conformado por señas estáticas y dinámicas, siendo el mayor número las estáticas (Figura 31 y Figura 32). Las señas dinámicas se componen por las palabras (Figura 33), las letras G, H, J, Ñ, S, Z, los números del seis al diez y por último las expresiones numéricas grandes de mil y millón.

Figura 31: Señas de letras estáticas de A hasta L estático LSC
Fuente: Elaboración propia

Figura 32: Números estáticos de cero al cinco LSC
Fuente: Elaboración propia

Las señas estáticas en su variación respecto al tiempo en video, varían con una pequeña traslación en el espacio y cuya transición de cuadro a cuadro es poco perceptible, en comparación a las palabras (Figura 33), que, por su gran extensión en los brazos, es sencillo de identificar, sin embargo, son dependientes de la dirección del movimiento, permitiendo a una misma seña tener diferentes interpretaciones.

Figura 33: Palabras dinámicas LSC
Fuente: Elaboración propia

Los voluntarios realizan y se adaptan a este tipo de señas de modo más rápido, ya que, esto no implica acomodar los dedos de maneras desconocidas para representar algunas letras, donde algunos participantes no poseen la elasticidad suficiente.

4.1.2 Almacenamiento

La información recolectada tiene un peso digital de 5,24Gb, donde cada video en formato mp4 posee un tamaño de 75Mb que corresponde aproximadamente a tres minutos de grabación. Se realiza una copia de seguridad local y otra en el servicio de la nube OneDrive con los nombres por defecto (VID_año_mes_dia_referencia.mp4) de cada video, pero renombrando la carpeta contenedora como DataLSC (Figura 34).

Figura 34: Carpeta de almacenamiento
Fuente: Elaboración propia

Esta carpeta corresponde a la base de todo el desarrollo de la investigación; las copias digitales se mantienen privadas, para evitar fugas de información, que no estén autorizadas en el consentimiento informado.

4.1.3 Oclusiones y posibles errores

Se realizó una exploración de los datos, mediante la visualización de todos los videos, identificando las dificultades presentadas por los participantes, siendo más común, la fatiga muscular en la realización de las señas, lo cual, afecta directamente la forma de la seña y la velocidad de ejecución, extendiendo el tiempo de grabación debido a los descansos y la realización nuevamente de la seña, también, existió una confusión en algunas personas sobre su mano dominante y la mano de soporte, por lo cual se cambió de mano que ejecuta la seña un tiempo después de empezar la grabación (Figura 35)

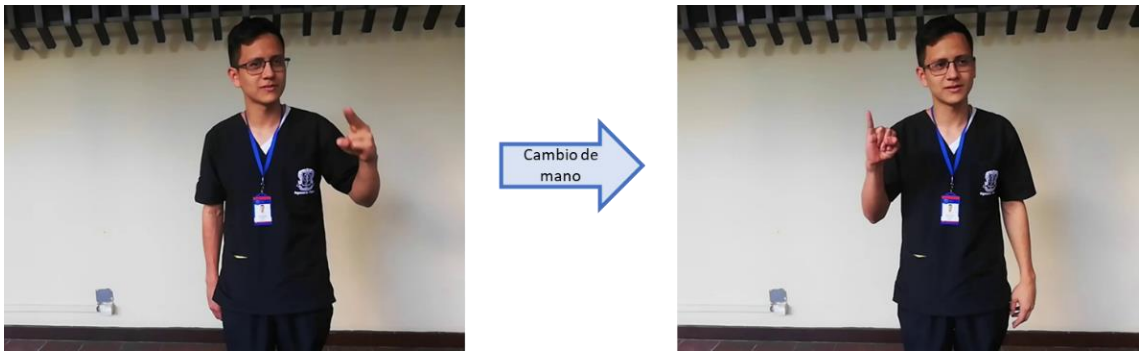


Figura 35: Cambio de mano en la ejecución de señas
Fuente: Elaboración propia

En la imagen (Figura 35), se muestra como el voluntario pasa de ejecutar un cuadro de letras dinámica H a una letra estática. Este cambio de mano es debido a la fatiga muscular o la incorrecta imitación al instructor frente a ellos; siendo para la gran mayoría la primera vez en realizar señas de LSC.

El control sobre la vestimenta no se tomó en cuenta, posibilitando que algunos colores o patrones pueden ser distracciones visuales o afectar la claridad de las señas, pero se cuantificó las prendas o accesorios más comunes usados por los voluntarios, con el propósito de identificar esta posible obstrucción (Tabla 6).

Tabla 6: Accesorios vestidos por voluntarios

Accesorios			
Gorras	Tapabocas	Abrigos	Gafas
4	14	23	18

Estos accesorios son usados respecto a las condiciones de ambiente del día de grabación y salud de los participantes, siendo más probable el uso de abrigos y tapabocas por la temporada de lluvias en la cual se realizó la grabación.

4.1.4 Información útil

Se realizo un primer filtrado de datos, omitiendo siete videos donde algunos participantes presentaron gran dificultad al realizar las señas, ejecutándola más de una vez sin mejora alguna. Con ello en mente y aplicando este primer filtrado, se obtienen 70 videos ideales, los cuales cumplen con la necesidad de esta investigación.

A partir de estos videos, se realizan las bases de datos de series de imágenes, que alimentan los modelos de inteligencia artificial. Los parámetros permitidos (Tabla 7) para dar a entender como video útil, son tomados respecto a la media de todas las grabaciones.

Tabla 7: Parámetros a considerar video útil

Parámetro			
Cantidad de repetición	Visualización respecto al fondo	Cambio de mano	Accesorios
3	Si	Si	Si

Los videos pueden variar su tamaño respecto a estos parámetros, sin embargo, la media de tiempo se establece dentro de los tres minutos de grabación.

4.2 Conjunto de datos

4.2.1 Secuencia de cuadros

Los videos obtenidos son el pilar fundamentos de los modelos de aprendizaje, sin embargo, el entrenamiento debe realizarse mediante imágenes y no por un formato de video, por esta razón, se procede a realizar la extracción de seis cuadros presentes en un segundo, mediante la librería OpenCv; almacenándolos automáticamente en carpetas con nombre: “Per + Numero” (Figura 36) y los archivos en formato JPG son nombrados como: “Per + Numero + Etiqueta + Numero del cuadro.jpg” (Figura 37).

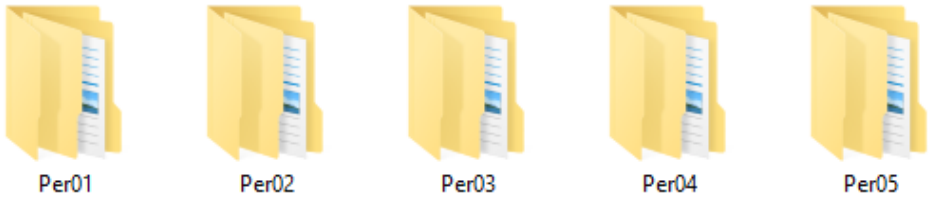


Figura 36: Etiqueta de carpetas
Fuente: Elaboración propia

Figura 37: Etiqueta de carpetas
Fuente: Elaboración propia

La conformación de este nombre, se toma respecto al procesamiento de los videos más antiguos al más actual. La cantidad de imágenes extraídas son 20.964 imágenes en tamaño 640x480 pixeles, que es el tamaño original de los videos.

4.2.2 Separación de categorías

Se procede a realizar una división de clases que conforman DataBase, en dos carpetas diferentes (Tabla 8):

Tabla 8: Separación de información

Bases de Datos			
División	**Carpeta 1**		**Carpeta 2**
Nombre	Alfabeto	Números	Palabras
Cantidad	11.340	5.424	4.200
Total	16.764		4.200
Nombre Base	**LSC70AN640**		**LSC_W70640**

A partir de la información base, se crearon dos repositorios nuevos, que incluyen la carpeta de interés con información alfanumérica y otra con imágenes de palabras que tendrá uso en un modelo de interpretación diferente.

4.2.3 Categorías similares

La conformación del dataset, está representado por las 40 clases alfanuméricas que se presentan en Tabla 5. Sin embargo, como tres categorías numéricas (0,2,3), son similares a las de alfabeto (O, V, W) respectivamente, se procede a eliminar las categorías numéricas y usar doble interpretación, que se resolverá respecto al contexto de la

conversación. En tanto, esto disminuye la cantidad de información de LCS70AN640 (Tabla 9).

Tabla 9: Actualización de información LSC70AN

LSC70AN640	
Cantidad anterior	**Cantidad actual**
16.764	15.504
40 categorías	37 categorías

La eliminación de estos datos disminuye la base a 37 clases de interés, siendo suprimidas 1.260 imágenes, que, aunque es un número considerable de informacion, es necesario sustraerlas para evitar confusiones y aumentar el rendimiento de los modelos de interpretación. Este proceso en el estado del arte no es considerado y usan señas similares en la conformación de las bases de datos.

4.2.4 Redimensionamiento

La base de datos LSC70AN640, posee un tamaño demasiado amplio para alimentar modelos de inteligencia artificial directamente, por tal razón, se extraen los pixeles alrededor de la persona y se redimensiona la imagen a un tamaño de 255x255 pixeles (Figura 38).

Figura 38: Seña dinámica número millón
Fuente: Elaboración propia

La redimensión de la base de datos, enfoca la información sobre la región de interés (manos), con lo cual, disminuye el peso del repositorio en un 67%. Finalmente, este dataset será nombrado como LSC70AN (

Tabla 10) y será la base de entrada los modelos de interpretación.

Tabla 10: Resumen LSCAN70

LSC70AN			
Cantidad	**Categorías**	**Peso [Mbytes]**	**Característica**
15.450	37	251	Dinámico

En resumen, LSC70AN es una base de datos dinámica que corresponde a 37 categorías alfanuméricas, donde cada categoría presenta una secuencia de seis cuadros y el total de imágenes corresponde 15.450 en tamaño 255x255 pixeles, siendo la primer base de datos en Colombia de características dinámicas, comparable con [10] y [30] , cuya base LSA64 emplea guantes de colores para ayudar al reconocimiento de las señas.

4.2.5 Subconjuntos de datos

A partir de las LSC70AN, se obtiene nuevo conjunto de datos igualmente dinámico; mediante el uso algoritmos de detección de pose (Mediapipe Pose), que se enfoca según su documentación en los puntos 20 y 19, correspondiente a la mano diestra y zurda respectivamente. Se realiza un recorte a partir de la imagen base y redimensiona a un tamaño 120x120 pixeles.

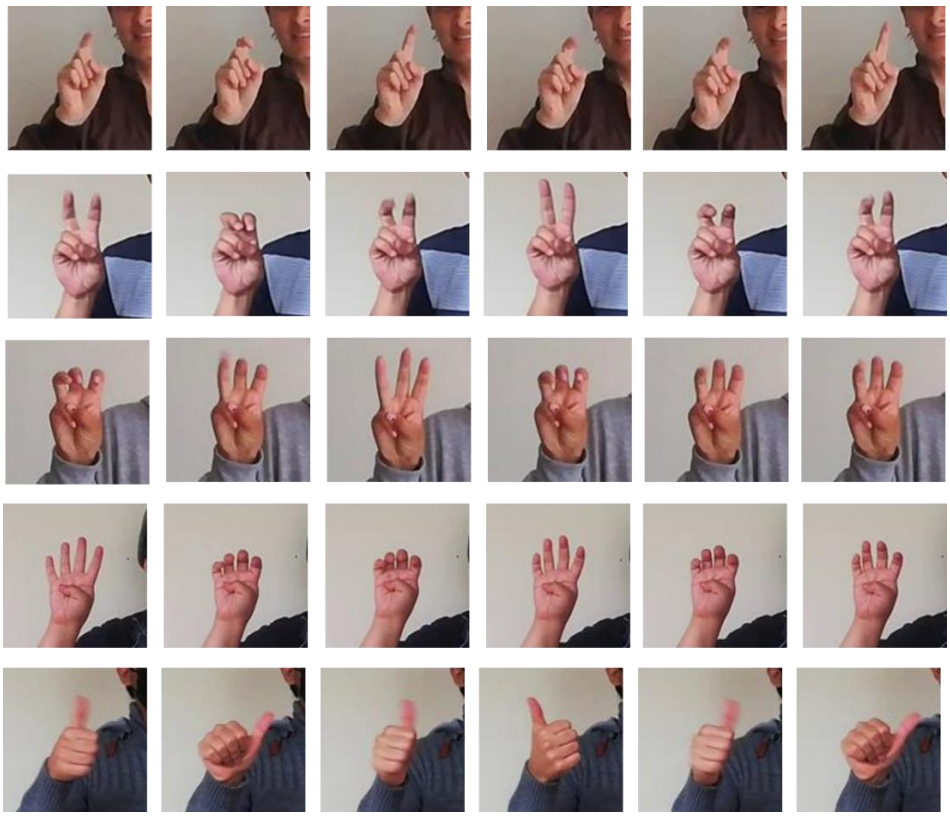


Figura 39: Señas dinámicas números 6-10
Fuente: Elaboración propia

Este subconjunto es un repositorio similar a LSC70AN, que maneja las mismas 37 clases de seis imágenes cada una, teniendo mayor número de clases comparado con [5], que intenta realizar un mismo desarrollo de vocales y números estáticos; también posee más variabilidad en condiciones semicrontroladas, que los dataset creados con el mismo enfoque [8], [32], [7], pero de carácter estático. Este es un desarrollo adicional y requerido para realizar una comparación similar con respecto al estado del arte, donde se enfoca directamente sobre la mano omitiendo el resto de información. Este subconjunto de datos se ha nombrado como LSC70AN_HANDS (Tabla 11)

Tabla 11: Resumen LSC70AN_HANDS

LSC70AN_HANDS					
Cantidad	**Categorías**	**Peso [Mbytes]**	**Característica**	**Tamaño [px]**	**Enfoque**
15.450	37	60,7	Dinámico	120x120	Manos

Este nuevo dataset, ofrece una nueva dimensión de información para la creación de modelos de interpretación dinámicos, diferenciando de la investigación nacional que se ha caracterizado, por tener enfoque sobre mano de carácter estático como [4]. A partir de este subconjunto de datos, se puede obtener nuevas características igualmente dinámicas.

4.2.5.1 Extracción de características

Se extrajeron coordenadas en puntos de la mano, mediante el uso de los 21 puntos de mediapipe hand, a partir de los repositorios LSC70AN y LSC70AN_HANDS. Encontrando coordenadas normalizadas que son independientes del tamaño de imagen, las cuales, permiten utilizar la cámara a diferentes posiciones de grabación.

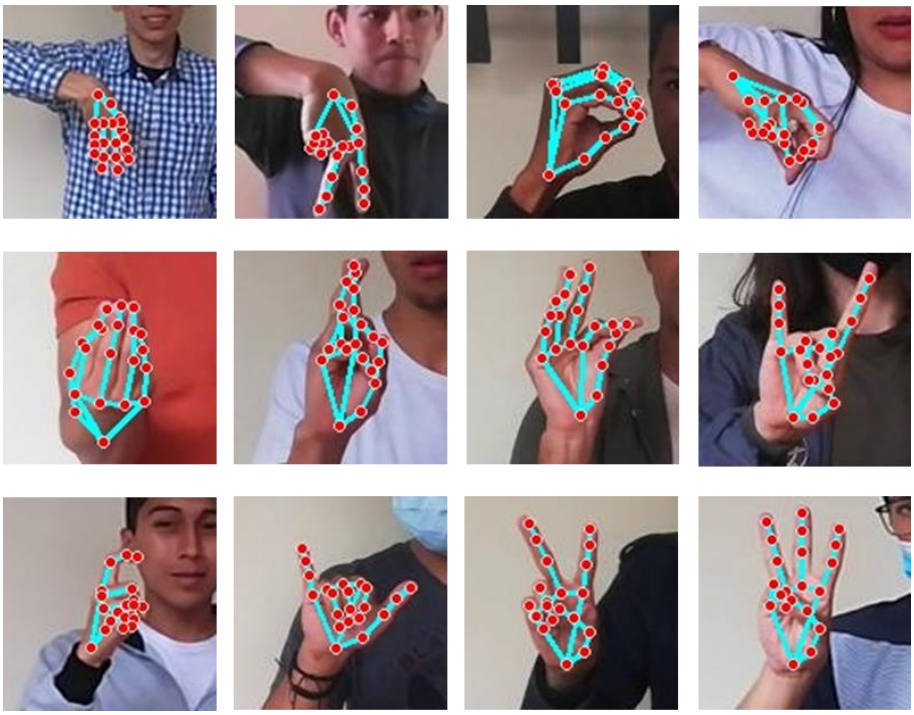


Figura 40: Recolección de puntos característicos letras estáticas M-Y

Fuente: Elaboración propia

Los puntos de características obtenidos, son únicos respecto a cada imagen y se adaptan a las diferentes señas, donde, entre mayor enfoque, realiza una mejor predicción de estos puntos. Con esto, se logra tener por cada puntos tres características que responden a un x,y,z en el espacio cartesiano; en total, las 63 características se graban en un archivo CSV, y se agregan tres columnas finales que contienen información correspondiente al fotograma (número de persona, categoría ,nombre de imagen).

0,492931	-0,014118	Per01	1	Per01_1_0.jpg
0,480766	-0,023479	Per01	1	Per01_1_1.jpg
0,550958	-0,039129	Per01	1	Per01_1_2.jpg
0,574992	-0,038809	Per01	1	Per01_1_3.jpg
0,578369	-0,037223	Per01	1	Per01_1_4.jpg
0,575794	-0,037502	Per01	1	Per01_1_5.jpg

Figura 41: Grabación de los puntos con identificación
Fuente: Elaboración propia

En la Figura 41, se muestran algunas las coordenadas y la identificación de clase uno. En [4] se realiza un proceso similar de extracción de características, sin embargo, cuenta con poca variabilidad de los datos, al ser dos voluntarios quienes realizan todo el proceso de adquisición de datos.

4.3 Modelos de interpretación

En esta sección se muestran los distintos modelos de interpretación, las diferentes técnicas que emplearon y el tipo de enfoque de información.

4.3.1 Consideraciones

Para el modelamiento de los datos, se utilizó técnicas de regularización, aplicadas sobre los datos principales tanto, para aumentar la cantidad y el rendimiento del modelo finalmente obtenido.

4.3.1.1 Aumentación de datos

Se realizó una aumentación de datos, para mejorar la cantidad de posibilidades de interpretación. Este método fue aplicado sobre todos los conjuntos de datos (Figura 42), considerando la misma condición para la secuencia de imágenes de la misma categoría. Esta aplicación de aumento posee tres posibilidades, rotación horizontal, rotación ±25 grados y una combinación de las dos primeras, se aplicó aleatoriamente sobre ambos conjuntos de imágenes, con lo cual, se logra tener el doble de información inicial, permitiendo a los sistemas de inteligencia artificial, realizar interpretaciones más precisas e imparciales, independientemente de la mano dominante. Al contar con esta nueva cantidad datos diversos, los modelos pueden evitar sesgos y obtener una comprensión más completa de los patrones y características de la información, lo que a su vez permite una toma de decisiones más objetiva; logrando algoritmos más robustos que se puedan

implementar en diferentes condiciones escena. De igual manera, sobre estas nuevas imágenes se aplica el detector de puntos de manos, para extraer más cantidad de datos y tener un aumento de datos de puntos característicos.

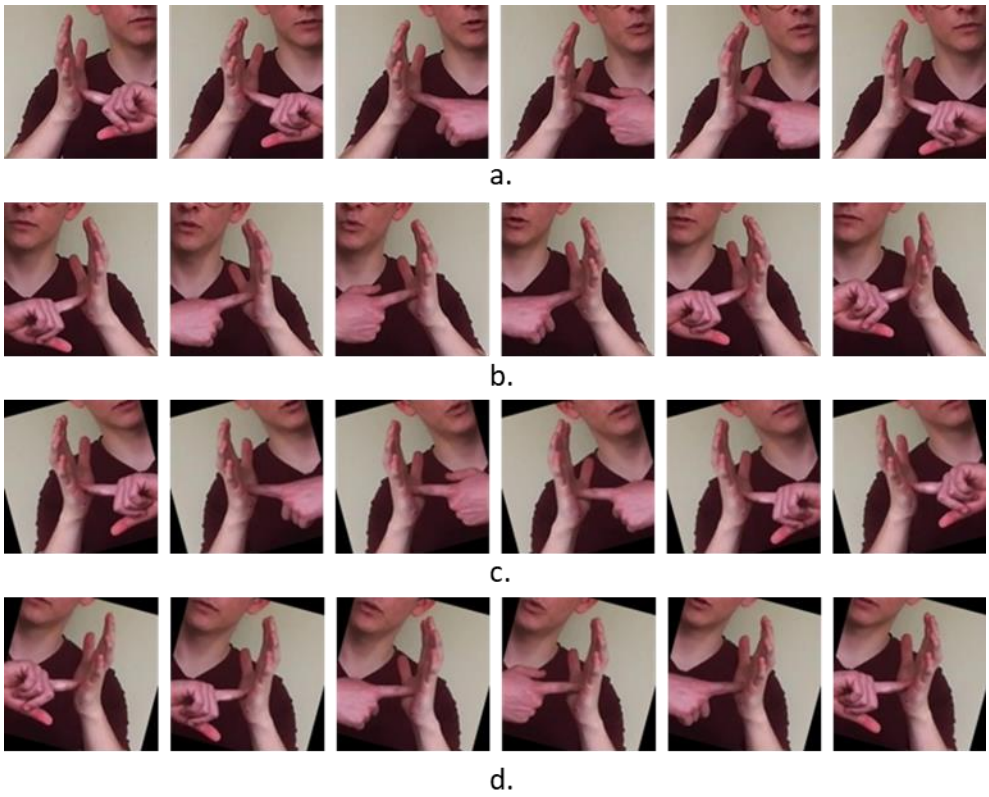


Figura 42: Aumento de datos en dataset de enfoque a mano
a) original, b) espejo c) espejo rotación negativo, d) rotación positiva
Fuente: Elaboración propia

4.3.1.2 Validación cruzada

En todos los modelos, se realiza un proceso de validación cruzada, dividiendo la información a través de cinco pliegues, considerando un porcentaje 75-25% sobre la aumentación de datos; la cantidad de imágenes se divide en 24.808 imágenes de entrenamiento y 6.200 para prueba del modelo, que corresponde 60 y 10 voluntarios para entrenamiento y prueba respectivamente. El desarrollo de esta técnica toma alrededor de 4.962 datos para validación en cada pliegue y se vigila el desarrollo del aprendizaje, empleando la técnica de early stopping configurada a una paciencia de diez épocas.

4.3.2 Modelos de Combinación

Estos modelos son una combinación de redes neuronales convolucionales y redes neuronales recurrentes, aplicando como extracción de características la CNN y manejando los datos como serie de tiempo con la red recurrente BILSTM.

4.3.2.1 Cuerpo completo

Se realiza pruebas con los modelos de transferencia de aprendizaje, abordando la tarea de interpretación de señas al dataset LSC70AN, que posee imágenes de tamaño 255x255 pixeles de cuerpo completo. Las pruebas respecto a estos modelos en combinación de diferentes algoritmos ofrecen diferentes resultados en su rendimiento (Tabla 12).

Tabla 12: Resultados a cuerpo completo modelo de combinación

Modelo CNN + BILSTM				
Métrica	**VGG16**	**Inception**	**Xception**	**MobileNet**
Exactitud	21,50%	19,50%	11,00%	15,50%

Los resultados en este procedimiento son poco alentadores, ya que tienden a la divergencia, a pesar de emplear técnicas de aprendizaje profundo con ajuste de parámetros por regularización, no logrando converger a una solución satisfactoria para la interpretación de la lengua de señas, cuyo máximo rendimiento, en el mejor de los casos es del 20,0%, siendo, un nivel claramente insuficiente para el desarrollo de un intérprete de lengua de señas, que pueda ser usado en situaciones reales, donde el número de aciertos debe ser alto para una comunicación fluida y efectiva entre personas con discapacidad auditiva y oyentes. Estos resultados son estimados, ya que la cantidad de información en la mano respecto a la escena, no es suficiente y se pierde la relevancia sobre todo el contexto de la imagen, omitiendo los pequeños detalles que surgen en la realización de señas, es por ello, por lo que esta técnica de interpretación de lengua de señas, es usada comúnmente enfocado en la mano [8] o en señas de palabras [31], donde la extensión de los brazos ofrece bastante información entre serie de cuadros.

4.3.2.2 Manos

Se probó la transición de las señas con los mismos modelos de redes neuronales preentrenados, en combinación con una red BILSTM, obteniendo sus rendimientos en distintas métricas de evaluación (Tabla 13), a través de los datos de prueba.

Tabla 13: Enfoque en mano CNN+BILSTM

CNN + BILSTM				
Modelo	**Exactitud**	**Precisión**	**Sensibilidad**	**Puntaje F1**
VGG16	61,5%	65,2%	61,5%	59,9%
Inception	**75,9%**	**78,5%**	**75,9%**	**76,0%**
Xception	67.6%	73,9%	67,6%	66.5%
MobileNet	68,4%	71,2%	68,4%	68,0%

Los resultados de exactitud y perdida fueron evaluados respecto a esta técnica. Estos rendimientos mejoran con modelos bastante simples como VGG16, y disminuyen con modelos más complejos como Xception, donde se brinda un número mayor de características por cada secuencia de imágenes, a su vez, los resultados son mejorados respecto a modelos de complejidad intermedia como MobileNet.

En el caso inception, mejora estas características debido a la reestructuración y concatenación de la información en cada paso, logrando extraer patrones diferentes en cada punto de la red y siendo estas lo bastante relevantes, para funcionar como una serie de tiempo de lengua de señas. En adición, los resultados son acordes respecto a modelos CNN de pocas características, ya que [9], realiza prueba con CaffeNet y obtiene resultados altos, en comparación a la extracción de características como transición de tiempo, sin embargo, los resultados tienden a ser menores comparado a [6], debido a la poca variabilidad de los datos realizando su prueba solo con dos personas.

Los rendimientos respecto a cada categoría (Tabla 14), según las métricas de evaluación ofrecen una nueva perspectiva de los resultados de manera local.

Tabla 14: Resultados respecto a categorías modelo combinacion manos

Categoría	Precisión	Sensibilidad	F1-score	Cantidad	Categoría	Precisión	Sensibilidad	F1-score	Cantidad
1	0.389	0.700	0.500	20	L	0.895	0.850	0.872	20
10	**0.760**	**0.950**	**0.844**	**20**	M	0.700	0.700	0.700	20
4	0.731	0.950	0.826	20	**MIL**	**0.895**	**0.850**	**0.872**	**20**
5	1.000	1.000	1.000	20	**MILLÓN**	**0.500**	**0.500**	**0.500**	**20**
6	**0.700**	**0.700**	**0.700**	**20**	N	0.647	0.550	0.595	20
7	**0.833**	**0.750**	**0.789**	**20**	**Ñ**	**0.714**	**0.750**	**0.732**	**20**
8	**0.762**	**0.800**	**0.780**	**20**	O	0.882	0.750	0.811	20
9	**0.941**	**0.800**	**0.865**	**20**	P	0.750	0.750	0.750	20
A	0.800	1.000	0.889	20	Q	0.864	0.950	0.905	20
B	1.000	0.550	0.710	20	R	0.714	0.500	0.588	20
C	1.000	0.800	0.889	20	**S**	**0.667**	**0.600**	**0.632**	**20**
D	1.000	0.550	0.710	20	T	0.941	0.800	0.865	20
E	0.773	0.850	0.810	20	U	1.000	0.700	0.824	20
F	0.692	0.450	0.545	20	V	1.000	0.850	0.919	20
G	**0.812**	**0.650**	**0.722**	**20**	W	0.731	0.950	0.826	20
H	**0.524**	**0.550**	**0.537**	**20**	X	1.000	0.900	0.947	20
I	0.607	0.850	0.708	20	Y	0.704	0.950	0.809	20
J	**0.559**	**0.950**	**0.704**	**20**	**Z**	**0.643**	**0.450**	**0.529**	**20**
K	0.900	0.900	0.900	20					

Los datos resaltados hacen referencia a las categorías dinámicas, cuya evaluación teniendo en cuenta la métrica de precisión, logra un rendimiento mayor al 50% y un máximo del 83,3%, logrando una mejor adaptación de las señas numéricas con respecto a las del alfabeto, sin embargo, las condiciones del número millón, pueden verse afectadas porque se realiza extracción de cuadros sobre mano, y el movimiento de esta seña se ejecuta a través del antebrazo hasta llegar a la palma, con lo cual, esta seña puede ser confundida por cualquier movimiento interpretándolo incorrectamente. Con respecto a las categorías estáticas (

Tabla 15) los resultados tienden a ser satisfactorios con un buen reconocimiento de estas.

Tabla 15: Resultados según su característica

Señas	Precisión	Sensibilidad	F1-score
Estáticas	82,16	78,33	78,74
Dinámicas	71,61	71,53	70,81

Estos resultados se separan respecto a sus categorías estáticas y dinámicas, teniendo el mejor desempeño en las señas estáticas, debido al enfoque sobre la mano que varía muy poco entre transición de cuadros, los cual, brinda información directa sobre el área de interés, limitando la cantidad de errores respecto a las transiciones dinámicas, debidas al desenfoque de la cámara. Este desempeño es predecible, ya que muchas investigaciones estáticas que trabajan con datos parecidos, logran buenos porcentajes de exactitud como [30], que obtiene un 90.9%.

La matriz de confusión (Figura 43), muestra una buena optimización, donde la gran mayoría de las categorías se encuentran sobre la diagonal de la matriz, que es un resultado esperado al realizar buenos modelos de clasificación, sin embargo, hay peculiaridades en estos resultados, principalmente son errores en la interpretación en las señas D,F,R y responden a la clase uno, siendo este un porcentaje de confusión de alrededor del 30%, debido a la similitud en la extensión del dedo índice, que, aunque se diferencian por otras formas que se realizan con los dedos sobre la palma, en algunas imágenes no es completamente diferenciable, clasificándolos incorrectamente. Adicionalmente, hay un error bastante notable sobre las clases S y Z, que poseen una extensión de movimiento amplia entre cuadros, debido a la baja captura de movimiento de la secuencia, en la cual, las señas quedan a mitad de recorrido y no completan su forma. Respecto al primer modelo enfocado en cuerpo, mejora muchos de los parámetros que no pudo asimilar la red, como consecuencia de la poca información que era valiosa respecto al fondo, debido a esto, se obtienen mejores resultados que pueden ser implementados como sistema de interpretación, cumpliendo las normas de la lengua de señas.

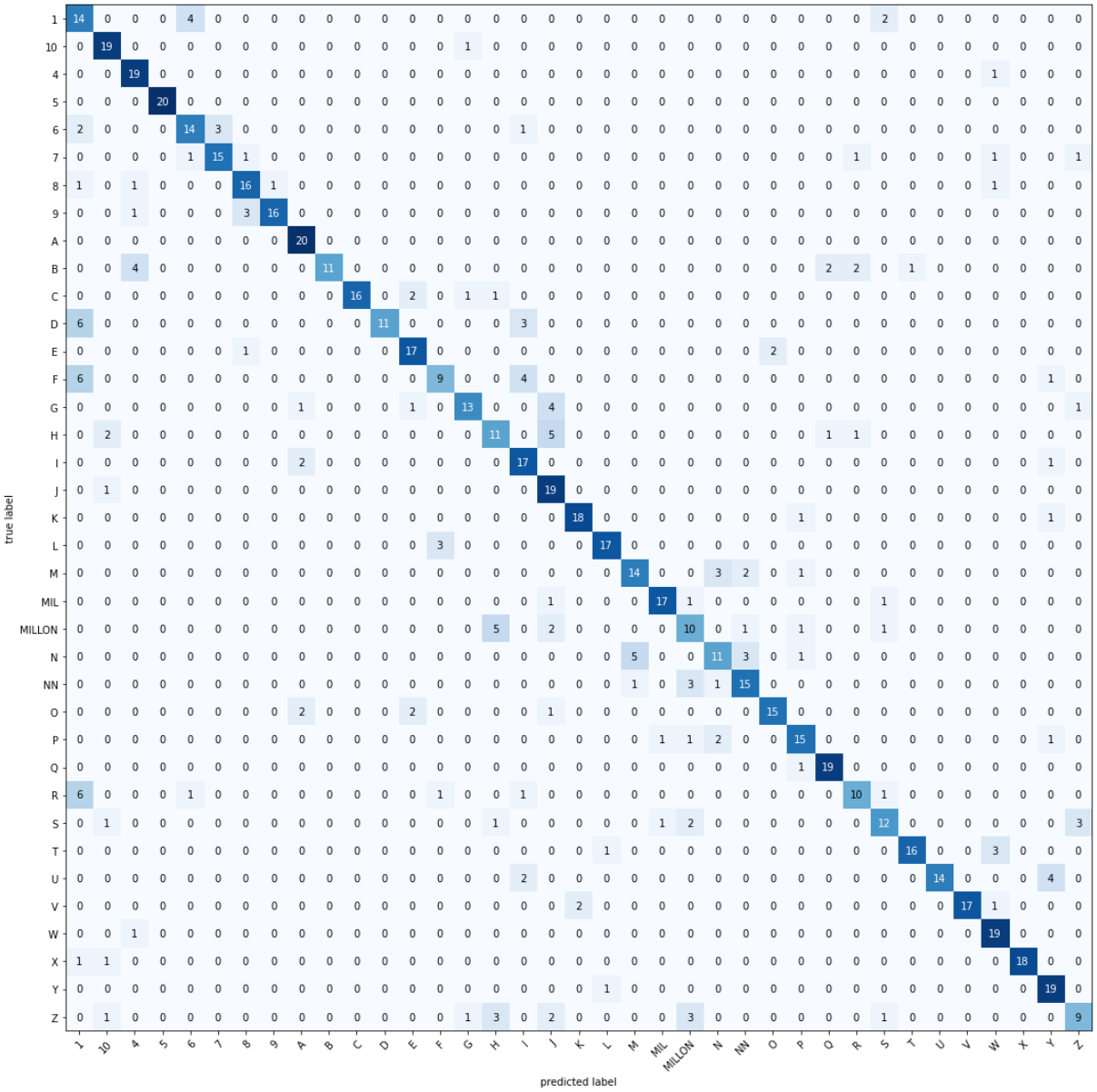


Figura 43: Matriz de confusión modelo combinación manos
Fuente: Elaboración propia

4.3.3 Modelos de Coordenadas

En esta sección, se presentan los resultados obtenidos a través de la extracción de coordenadas, con algoritmos de detección de pose y mano. El resumen del modelo (Tabla 16), muestra los parámetros en cada etapa como solución de la transición de cuadros.

Tabla 16: Modelo de red LSTM

Capa	Tamaño de salida	Número de parámetros
Entrada	[6, 63]	0
BILSTM	128	65,536
Densa	1024	132,096
Dropout	0,3	0
Densa (salida)	37	37,925

Este modelo se sometió ambos dataset de coordenadas, permitiendo validar su rendimiento y capacidad de generalización, por lo cual, se le aplicó métricas de regularización L2 y dropout mediante ensayo y error, para prevenir el sobreajuste y obtener una red con ajuste fino.

4.3.3.1 Cuerpo completo

El rendimiento del modelo (Tabla 17), muestra la interacción con los datos de prueba, donde se obtienen unas métricas similares, a las alcanzadas con el modelo de combinación enfocado en manos.

Tabla 17: Resultado modelo coordenadas cuerpo completo

Modelo de coordenadas a cuerpo completo				
Modelo	Exactitud	Precisión	Sensibilidad	Puntaje F1
BILSTM	75,5%	79,1%	75,5%	75,1%

El entrenamiento se realizó mediante validación cruzada de cinco pliegues, obteniendo el mejor desempeño en el pliegue número cuatro, a alrededor de época 50 (Figura 44), lo que significa, que después de entrenar varias veces el modelo con diferentes subconjuntos de datos, logra la mejor generalización de la información. Esta técnica es beneficiosa para evitar el sobre ajuste, garantizando el rendimiento generalizando de la información, siendo capaz de aprender patrones de manera efectiva para realizar la interpretación de le lengua de señas.

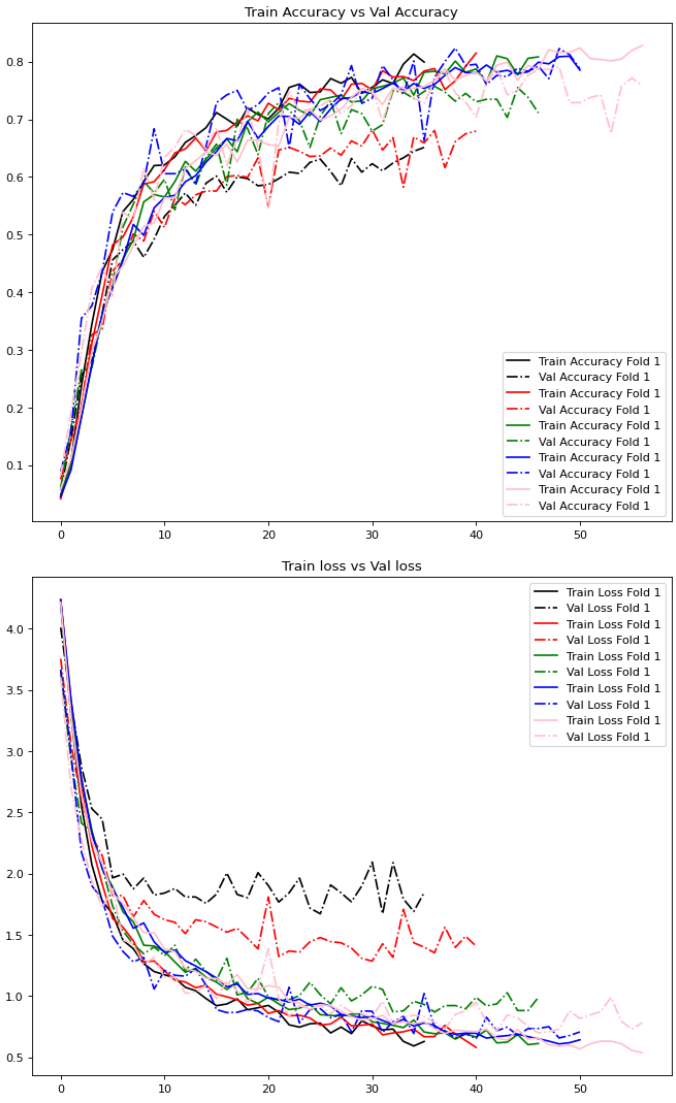


Figura 44: Exactitud y perdida del modelo coordenadas cuerpo completo
Fuente: Elaboración propia

Los resultados obtenidos son bastante alentadores, ya que, mejora resultados previos de mismas condiciones, logrando un mejor rendimiento no solo en términos de exactitud, sino también, en términos computacionales, al realizar menor cantidad de operaciones matemáticas, lo que lo hace más eficiente y rápido en la interpretación de las señas. Adicionalmente, el tener un modelo que posea menor costo computacional implementado a cuerpo completo, beneficia en la reducción de los tiempos de retraso entre secuencia de interpretación, permitiendo procesar y analizar datos con mayor eficiencia, lo cual, es importante al realizar una comunicación más fluida y eficiente. La evaluación en general respecto a cada categoría se realiza mediante la matriz de confusión (Figura 45) mostrando la cantidad de aciertos y desaciertos respecto a otras categorías.

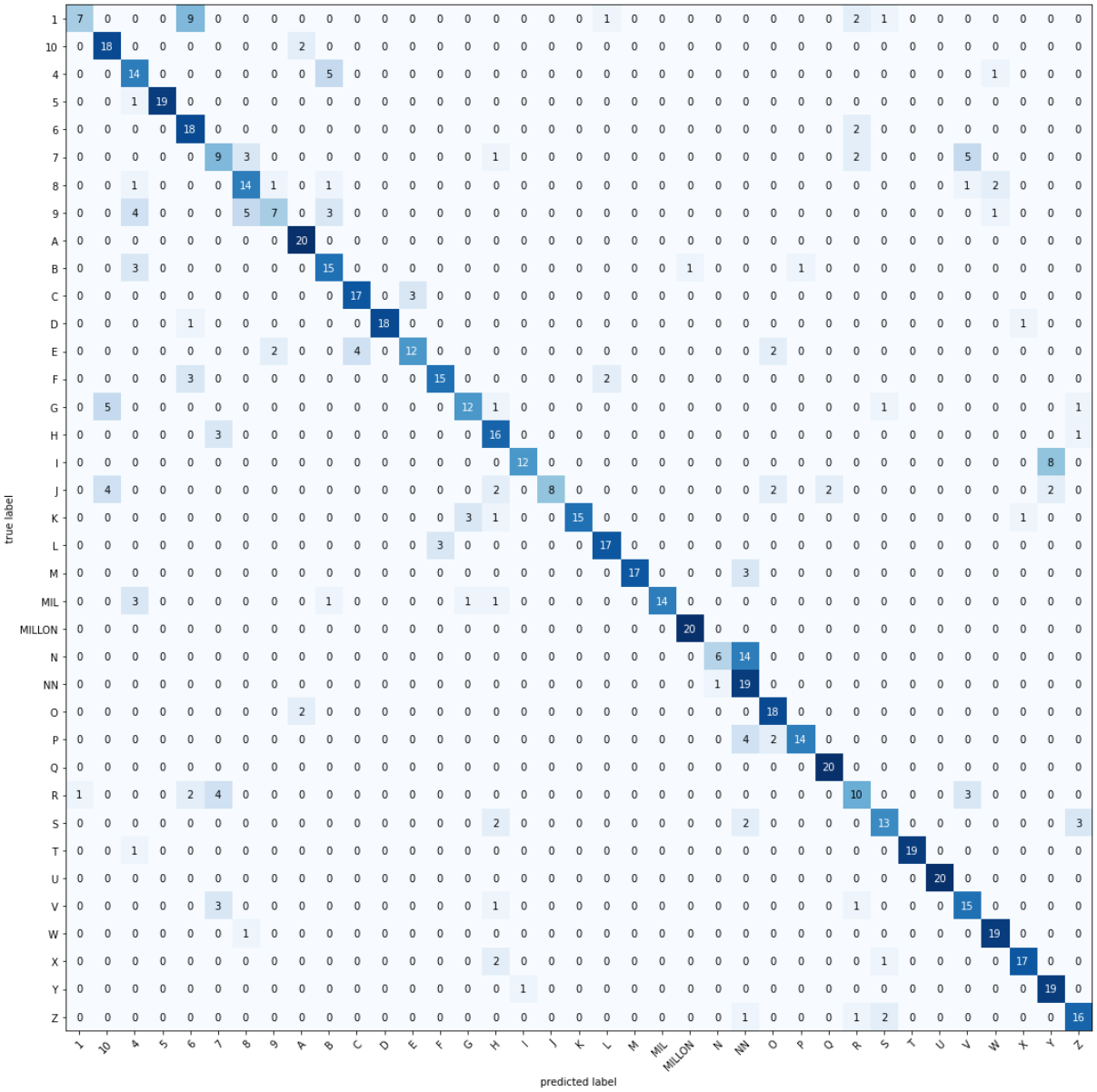


Figura 45: Matriz de confusión modelo coordenadas cuerpo completo
Fuente: Elaboración propia

Estos resultados tienden a ser más aleatorios, pero más significativos por la cantidad, presentando un rendimiento similar al modelo de combinación enfocado en mano, sin embargo, muestra gran cantidad de errores sobre los números, confundiendo la transición estática con la dinámica, debido a que las señas de los números del seis al nueve, son las mismas de los números del uno al cuatro, cuya diferencia radica por el movimiento sobre la falange media. Adicionalmente, bajo las mismas condiciones presenta predicciones erradas en las letras N y Ñ, que contienen la misma forma, pero se diferencian por el movimiento. Como es necesario un algoritmo para obtener los puntos de coordenadas, entonces, sus parámetros deben ser ajustados para que los datos no sean grabados como cero, lo cual, ocurre al no encontrar puntos dentro de la imagen por condiciones de captura. Estos

parámetros se establecen en valores de defecto, pero la adquisición de datos mejora con el ajuste del valor del modelo de complejidad a un valor de uno, logrando el mejor seguimiento y captura de datos sobre la mano. Los resultados respecto a cada categoría se hacen, a través de la tabla de métricas (Tabla 18).

Tabla 18: Resultados del modelo coordenadas cuerpo completo

Categoría	Precisión	Sensibilidad	f1-score	Cantidad	Categoría	Precisión	Sensibilidad	f1-score	Cantidad
1	0.875	0.350	0.500	20	L	0.850	0.850	0.850	20
10	**0.667**	**0.900**	**0.766**	**20**	M	1.000	0.850	0.919	20
4	0.519	0.700	0.596	20	**MIL**	**1.000**	**0.700**	**0.824**	**20**
5	1.000	0.950	0.974	20	**MILLÓN**	**0.952**	**1.000**	**0.976**	**20**
6	**0.545**	**0.900**	**0.679**	**20**	N	0.857	0.300	0.444	20
7	**0.474**	**0.450**	**0.462**	**20**	Ñ	0.442	0.950	0.603	20
8	**0.609**	**0.700**	**0.651**	**20**	O	0.750	0.900	0.818	20
9	**0.700**	**0.350**	**0.467**	**20**	P	0.933	0.700	0.800	20
A	0.833	1.000	0.909	20	Q	0.909	1.000	0.952	20
B	0.600	0.750	0.667	20	R	0.556	0.500	0.526	20
C	0.810	0.850	0.829	20	**S**	**0.722**	**0.650**	**0.684**	**20**
D	1.000	0.900	0.947	20	T	1.000	0.950	0.974	20
E	0.800	0.600	0.686	20	U	1.000	1.000	1.000	20
F	0.833	0.750	0.789	20	V	0.625	0.750	0.682	20
G	**0.750**	**0.600**	**0.667**	**20**	W	0.826	0.950	0.884	20
H	**0.593**	**0.800**	**0.681**	**20**	X	0.895	0.850	0.872	20
I	0.923	0.600	0.727	20	Y	0.655	0.950	0.776	20
J	**1.000**	**0.400**	**0.571**	**20**	**Z**	**0.762**	**0.800**	**0.780**	**20**
K	1.000	0.750	0.857	20					

Los resultados de cada categoría sobre el carácter estático se mantienen en rangos aceptables, sin embargo, hay una disminución bastante notoria sobre las categorías dinámicas, con porcentajes de precisión menores en comparación al modelo de combinación de enfoque a mano, debido, a que los datos no han podido ser enfocados en la secuencia, limitando la cantidad de información relevante de las extremidades superiores, con lo cual, aumenta el porcentaje de identificación estática y disminuye la dinámica.

Tabla 19: Comparación resultados dinámicos - estáticos

Señas	Precisión	Sensibilidad	F1-score
Estáticas	83,53	78,12	79,07
Dinámicas	70,89	70,76	67,76

Al realizar una comparación, entre el modelo de combinación y coordenadas a cuerpo completo, se llega a resultados bastante similares, con una diferencia menor al 1.0 %. Sin embargo, esta implementación a cuerpo completo posee mayor ventaja en costo computacional, ya que no requiere un algoritmo de detección para recortar los cuadros, evitando posibles problemas de alineación al realizar este proceso.

4.3.3.2 Enfoque en manos

La importancia de los enfoques sobre mano en el análisis de señas, radica en su capacidad para proporcionar una mayor distinción entre ellas. Estos enfoques permiten mejorar los resultados de rendimiento en comparación a los otros modelos, ya que se analiza información no vista en diferentes posiciones para probar su desempeño. En la (Tabla 20) se muestran los resultados obtenidos al emplear este tipo de datos.

Tabla 20: Resultado modelo coordenadas mano

Coordenadas + manos				
Modelo	**Exactitud**	**Precisión**	**Sensibilidad**	**Puntaje F1**
BILSTM	85,7%	86,6%	85,7%	85,4%

Los resultados obtenidos muestran un aumento significativo en todas las medidas, lo que se traduce en una exactitud del 85,7%; este valor supera el mejor modelo obtenido en un 10%, lo que indica que los rendimientos son altamente satisfactorios. Este desempeño es esperado, ya que en el estudio de [28], se utiliza un enfoque directo sobre la mano para mejorar la clasificación, permitiendo una evaluación más precisa de las señas, al considerar factores como la variación en la posición de la mano, lo que permite, mejorar la capacidad para identificar y clasificar diferentes señas. De igual modo, el análisis en validación cruzada confirmó las métricas obtenidas en la Tabla 20. En este análisis, se realiza una prueba al finalizar cada iteración de entrenamiento, lo que permite confirmar la estabilidad y consistencia de los resultados obtenidos. La Tabla 21 muestra los resultados de las pruebas realizadas en cada iteración y, como se puede observar, los resultados son consistentes y estables a lo largo de todo el proceso de entrenamiento.

Tabla 21: Exactitud respecto a sus pliegues en validación cruzada

Exactitud					
Modelo	**Pliegue 1**	**Pliegue 2**	**Pliegue 3**	**Pliegue 4**	**Pliegue 5**
BILSTM	86,6%	84,9%	84.7%	85.3%	87.1%

Los resultados de los pliegues de validación cruzada son muy prometedores, ya que se observa que el rendimiento obtenido difiere en un máximo de 1.0%. Esto indica, que el modelo es capaz de proporcionar una respuesta consistente y confiable, ante datos desconocidos, lo cual, es muy importante en cualquier aplicación práctica, teniendo una exactitud media de los pliegues del 85,54%, mostrando coherencia con los resultados obtenidos en la Tabla 20. Sumado a ello, las curvas de exactitud y pérdida (Figura 46), proporcionan información valiosa sobre el comportamiento del modelo. Estas curvas muestran cómo el modelo se desempeña a medida que se realiza el entrenamiento, logrando el mejor resultado en el pliegue número cuatro, sugiriendo que el modelo es más efectivo para identificar patrones y realizar clasificaciones precisas, en ese punto del proceso de entrenamiento.

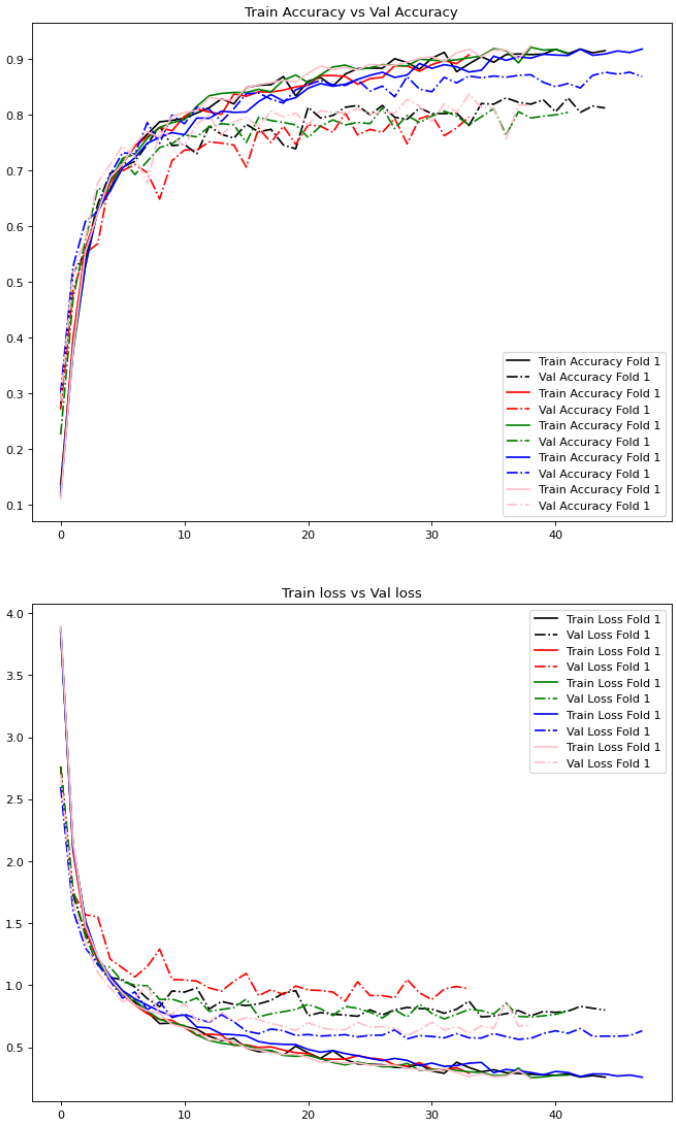


Figura 46: Exactitud y perdida modelo enfoque en mano
Fuente: Elaboración propia

El análisis de las curvas, proporciona evidencia clara que el modelo es capaz de adaptarse de manera efectiva a los datos, logrando esto alrededor de las 50 épocas, donde las curvas de entrenamiento y validación se mantienen juntas en las primeras épocas, lo que sugiere que el modelo está aprendiendo rápidamente de los datos y, a medida que avanza el proceso de entrenamiento, las curvas se desprenden gradualmente y mantiene el valor de pérdida cercano a cero, lo cual, es un indicador de que el modelo está logrando generalizar, antes de entrar en una etapa de sobre-entrenamiento,

La prueba del modelo se realizó nuevamente con la matriz de confusión (Figura 47), para ver el mejor desempeño en todas las categorías, esperando tener el mayor número de datos sobre la diagonal.

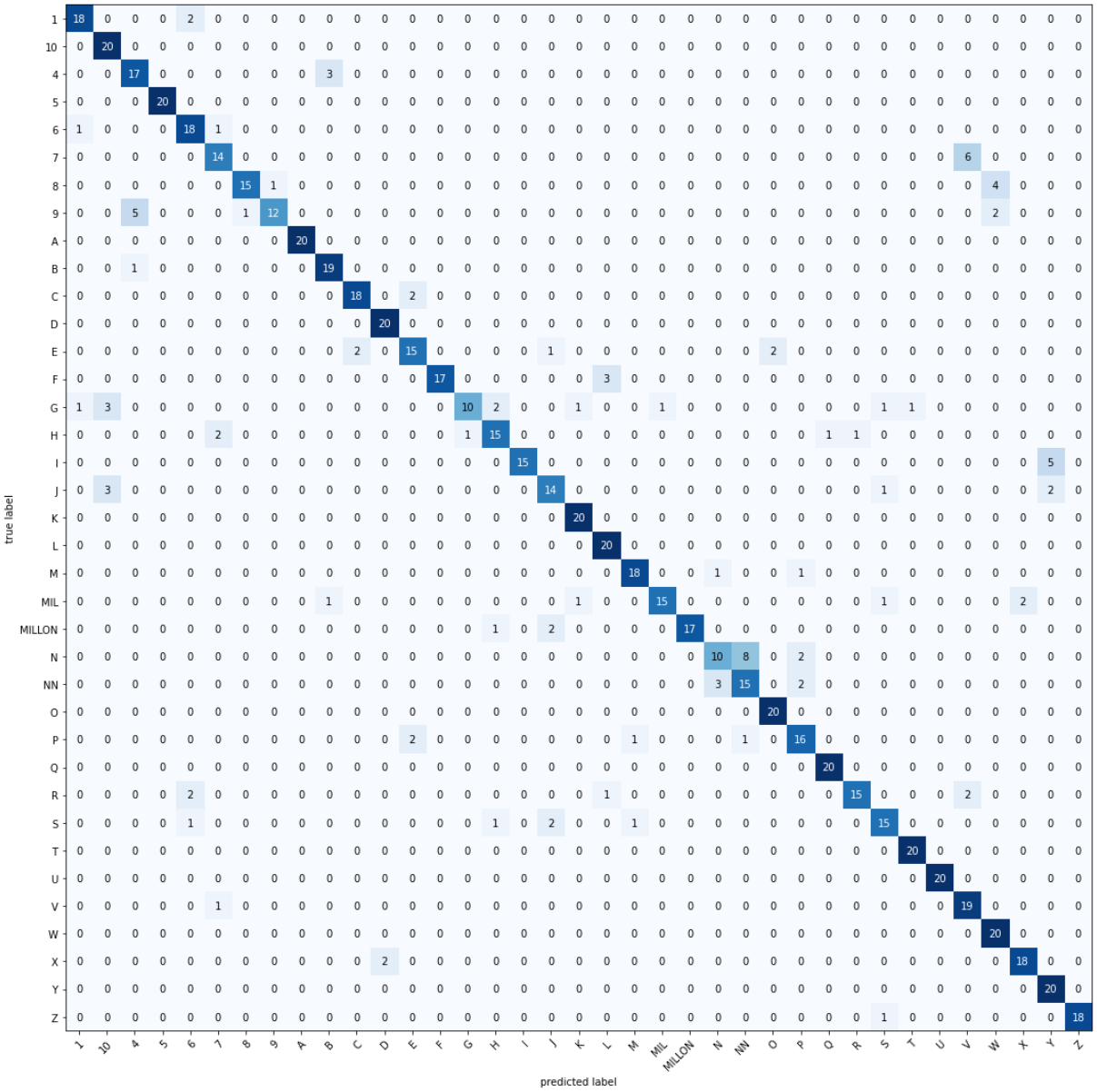


Figura 47:Matriz confusión de enfoque en mano coordenadas
Fuente: Elaboración propia

La matriz de resultados es una herramienta fundamental para evaluar la calidad de los modelos de interpretación. En este caso, los resultados muestran que la matriz está acorde a los resultados previstos, con un número mayor de datos sobre la diagonal y un porcentaje del 14.32 % de los datos erróneos fuera de esta. Es importante señalar que, aunque se han obtenido resultados perfectos en 13 de las 37 categorías, aún hay ciertas categorías que presentan problemas en la clasificación de datos, esto ocurre nuevamente, con las letras N y Ñ, los números del seis al nueve, como también con la clase G. Estos hallazgos sugieren que aún hay margen de mejora en la exactitud de los modelos de interpretación, sin embargo, estos resultados representan un avance significativo en el desarrollo de los

modelos de interpretación, logrando que el modelo actual, sea capaz de clasificar e interpretar con mayor precisión, algunos datos que modelos anteriores no pudieron, debido a las falencias en la asignación de algunos puntos de coordenadas.

Se revisa el rendimiento respecto a cada categoría, para comparar el desempeño de las señas dinámicas como estáticas, como se observa en la Tabla 22.

Tabla 22: Resultados según las categorías

Categoría	Precisión	Sensibilidad	F1-score	Cantidad	Categoría	Precisión	Sensibilidad	F1-score	Cantidad
1	0.900	0.900	0.900	20	L	0.833	1.000	0.909	20
10	**0.769**	**1.000**	**0.870**	**20**	M	0.900	0.900	0.900	20
4	0.739	0.850	0.791	20	**MIL**	**0.938**	**0.750**	**0.833**	**20**
5	1.000	1.000	1.000	20	**MILLÓN**	**1.000**	**0.850**	**0.919**	**20**
6	**0.783**	**0.900**	**0.837**	**20**	N	0.714	0.500	0.588	20
7	**0.778**	**0.700**	**0.737**	**20**	**Ñ**	**0.625**	**0.750**	**0.682**	**20**
8	**0.938**	**0.750**	**0.833**	**20**	O	0.909	1.000	0.952	20
9	**0.923**	**0.600**	**0.727**	**20**	P	0.762	0.800	0.780	20
A	1.000	1.000	1.000	20	Q	0.952	1.000	0.976	20
B	0.826	0.950	0.884	20	R	0.938	0.750	0.833	20
C	0.900	0.900	0.900	20	**S**	**0.789**	**0.750**	**0.769**	**20**
D	0.909	1.000	0.952	20	T	0.952	1.000	0.976	20
E	0.789	0.750	0.769	20	U	1.000	1.000	1.000	20
F	1.000	0.850	0.919	20	V	0.704	0.950	0.809	20
G	**0.909**	**0.500**	**0.645**	**20**	W	0.769	1.000	0.870	20
H	**0.789**	**0.750**	**0.769**	**20**	X	0.900	0.900	0.900	20
I	1.000	0.750	0.857	20	Y	0.741	1.000	0.851	20
J	**0.737**	**0.700**	**0.718**	**20**	**Z**	**1.000**	**0.947**	**0.973**	**20**
K	0.909	1.000	0.952	20					

La tabla 22, presenta los rendimientos de prueba del modelo de reconocimiento de señas, donde se puede observar, que cada una de las clases mantiene un porcentaje mayor al 70%. De igual modo, los resultados muestran que el modelo ha mejorado significativamente cualquier resultado previo obtenido en este campo, lo que expone que el modelo ha aprendido con un alto nivel de precisión la clasificación de señas dinámicas, específicamente, una alta identificación en señas tales como ocho, diez, mil, millón y la letra Z. Estos resultados demuestran la coherencia general del modelo en datos de prueba, y en la Tabla 23 se establece una comparación entre los resultados dinámicos y estáticos.

Tabla 23: comparación de señas estáticas y dinámicas

Señas	Precisión	Sensibilidad	F1-score
Estáticas	87,69	90,63	88,62
Dinámicas	84,45	76,52	79,32

El hecho que modelo haya logrado métricas más altas, que en cualquier otro punto del proyecto, sugiere que el modelo es confiable y capaz de generalizar bien a datos no vistos

previamente. Cabe destacar que la diferencia entre la precisión de clasificación de señas estáticas y dinámicas es mínima, siendo menor al 4.0%, esto es un logro importante, ya que la clasificación de señas dinámicas ha sido más compleja, debido a la variabilidad que presentan en términos de la velocidad de la secuencia, y la forma en que se realizan las señas.

4.3.4 Modelo de interpretación de palabras LSC

Como un anexo al desarrollo del interprete alfanumérico y con el excedente de datos que se posee, se creó un modelo de interpretación dinámico de palabras LSC, empleando un modelo CNN + BILSTM. El modelo preentrenado elegido para cumplir esta labor fue VGG16, obteniendo una exactitud del 76,0%, que fue ajustado usando técnicas de regularización para la optimización del modelo. El desarrollo en datos de prueba da a obtener la matriz de confusión (Figura 48).

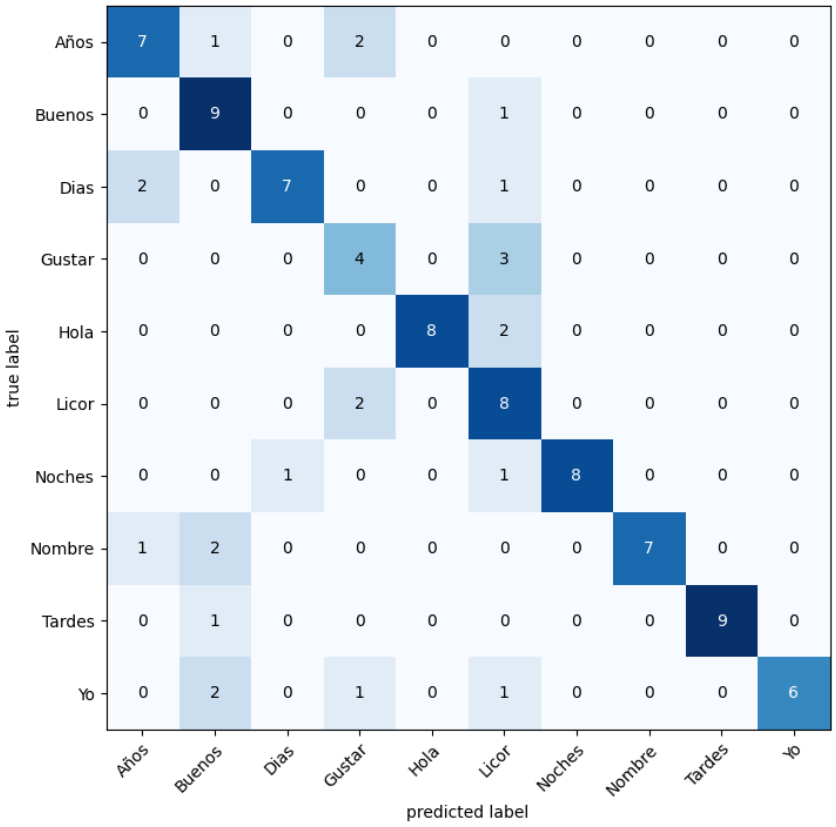


Figura 48: Matriz de confusión modelo de palabras
Fuente: Elaboración propia

El desarrollo de este sistema, demuestra la posibilidad de intérpretes dinámicos a cuerpo completo de lengua de señas, utilizando modelos de combinación; sin embargo, estos resultados pueden ser mejorados, usando un modelo diferente de CNN o empleando coordenadas, con lo cual, se lograría una transición de información más relevante y distintivas de los puntos adquiridos.

4.4 Despliegue

Se desarrollo una plataforma simple de despliegue empleando la librería Streamlit, logrando una aplicación rápido de tipo local (Figura 49), que cuenta con instrucciones y explicación de las categorías posibles de interpretar.

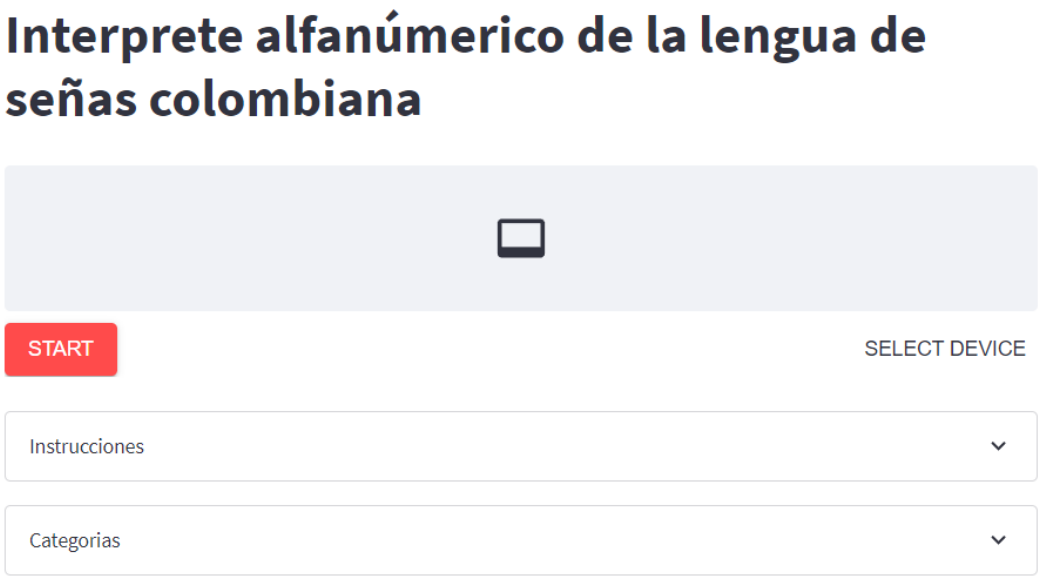


Figura 49: Despliegue del modelo local
Fuente: Elaboración propia

Este desarrollo, habilita la opción de elegir entre los diversos dispositivos de captura de imagen del computador. Su desempeño, es dependiente de la maquina donde este ejecutando el software, por tanto, presenta leves retrasos en la adquisición de la imagen, por ello, es recomendable ejecutar el software en una máquina de buenos recursos con tarjeta gráfica. Adicionalmente, las interpretaciones se realizaron con el modelo de coordenadas en enfoque a mano, adquiriendo las imágenes continuamente y almacenando seis cuadros a cuerpo completo.

A la adquisición de datos, se les realiza un previo procesamiento, donde se igualan condiciones con las que el modelo fue entrenado, pasando la serie de información al modelo para que realice la interpretación (Figura 50).

Figura 50: Voluntarios realizando pruebas de la aplicación
Fuente: Elaboración propia

Las interpretaciones funcionan para todas las categorías, estáticas y dinámicas, a partir, de la escritura por deletreo, que son desplegadas directamente sobre la imagen de la cámara. Sin embargo, como la adquisición es continua, se presentan diversos errores al interpretar algunas letras como se aprecia en la *Figura 51*.

Figura 51: Errores en interpretación
Fuente: Elaboración propia

Para corregir cierta parte de estos errores, y teniendo en cuenta que la mayor confusión en el despliegue es la seña de millón, dado su extensión sobre el brazo de soporte puede ser confundido con cualquier seña de mano abierta, se procede a omitir del despliegue. Adicionalmente, se estableció un intervalo de confianza en el 80%, como mínimo para una interpretación como correcta.

5. Conclusiones y trabajos futuros

5.1 Conclusiones

- Se desarrollo a través de la metodología CRISP-DM, tres modelos de interpretación de lengua de señas colombiana de características dinámicas, empleando distintas técnicas y modelos de inteligencia artificial. Logrando el mejor desempeño en el modelo basado en coordenadas enfocado en mano, con una exactitud del 85.7% y cuyos resultados, se adaptan a condiciones de iluminación y mano dominante empleada por el usuario.
- Se construyó un conjunto de imágenes alfanuméricas de la lengua de señas colombiana, con 15.504 imágenes, obtenidas a través setenta voluntarios en variadas condiciones de iluminación, de los cuales, se extraen dos sub-dataset de imágenes cuadradas de tamaños 255 y 120 pixeles, correspondientes a cuerpo completo y enfoque en mano respectivamente, que encuentran al mismo nivel de conjuntos de datos creados en Argentina y Estados unidos bajo las mismas condiciones

- Los modelos de interpretación por coordenadas, presentan problemas de adquisición de información por desenfoque en el movimiento, como consecuencia, hay una perdida de datos y falla en la interpretación de clases similares que son dinámicamente diferenciables, tales como los números y algunas letras.

- Se creo una herramienta, que facilita la comunicación con personas que presentan discapacidad auditiva, mediante interpretación alfanumérica de la lengua de señas colombiana, permitiendo expresar así sus pensamientos e ideas de manera clara y efectiva, lo que a su vez les permite integrarse plenamente en la sociedad. Adicionalmente, la creación de este sistema es un importante apoyo a los programas implementados por el gobierno nacional en pro de la inclusión social, mejorando el acceso a los mismos servicios y oportunidades que el resto de la población.

5.2 Trabajos futuros

El trabajo realizado deja una serie de líneas abiertas para su extensión futura:

- Desplegar la plataforma de interpretación en un servidor web, con librerías compatibles que faciliten la interpretación de lengua de señas satisfaciendo la necesidad de comunicación a cualquier hora del día,

- Ampliar la cantidad de categorías de información perteneciente al dataset actual, con nuevas etiquetas de carácter dinámico, con palabras para lograr una comunicación más extendida y rica en vocabulario, enfocada a conocer gustos y preferencias de los participantes en la conversación.

- Concatenar modelos de interpretación alfanumérica y palabras, con el propósito de realizar una presentación personal completa, sin interrupciones ante cambio de modelo, aprovechando las fortalezas de cada modelo y superar las limitaciones individuales que se presenta en la comunicación.

- Implementar los modelos de interpretación en plataformas móviles, que sirvan como método rápido de traducción de lengua de señas colombiana, con el uso de librería tensorflow lite, reduciendo el tamaño del desarrollo y siendo accesible a todas las personas que deseen realizar una combinación en LSC.

- Implementar una combinación de modelos con el uso de Transformers, que analicen la información como una aplicación del procesamiento natural del lenguaje, para no limitar distinciones la cantidad de clases y responda ante cambios significativos del contexto de una conversación.

Bibliografía

[1] *OHCHR*, "Convención sobre los derechos de las personas con discapacidad", https://www.ohchr.org/es/instruments-mechanisms/instruments/convention-rights-persons-disabilities (accedido 8 de febrero de 2023).

[2] INSOR, "Informe técnico Estado Goce en Derechos de la Población sorda 2019" https://www.insor.gov.co/insorlab/wp-content/uploads/2021/12/Informe-Tecnico-Estado-Goce-de-Derechos-de- la-poblacion-Sorda-2019.pdf (accedido: 8 de febrero de 2023)

[3] El Congreso de Colombia, "Ley 982 de 2005 - Gestor Normativo - Función Pública". https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=17283 (accedido 8 de febrero de 2023).

[4] J. J. Gutiérrez Leguizamón, J. A. Plazas López, M. J. Suárez Barón y J. S. González Sanabria, "Reconocimiento de lengua de señas colombiana mediante redes neuronales convolucionales y captura de movimiento", *Tecnura*, vol. 26, n.º 74, pp. 70–86, septiembre de 2022. Accedido el 28 de marzo de 2023. [En línea]. Disponible: https://doi.org/10.14483/22487638.19213

[5] G. Jiménez, E. Moreno, R. Guzman, y J. Barrero, "Automatic method for Recognition of Colombian Sign Language for vowels and numbers from zero to five by using SVM and KNN", en *2019 Congreso Internacional de Innovación y Tendencias en Ingenieria (CONIITI)*, oct. 2019, pp. 1-6. doi: 10.1109/CONIITI48476.2019.8960695.

[6] K. V. Monsalve Pineda y J. E. Polo Álvarez, "Traductor de símbolos de alfabeto del lenguaje de signos colombiano al lenguaje escrito", 2016, Accedido: 7 de diciembre de 2022. [En línea]. Disponible en: https://bibliotecadigital.univalle.edu.co/handle/10893/17348

[7] A. H. Nyky Joel, J. P. Navarro Cabiativa, y A. E. Gaona Barrera, "Sign identification model of the Colombian Sign Language (CSL) alphabet based on Computational Intelligence", en *2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)*, jul. 2022, pp. 1-6. doi: 10.1109/ColCACI56938.2022.9905253.

[8] N. E. Suat Rojas, B. S. Montoya Serna, E. M. Pinzón Velásquez, y O. S. Rodríguez Galeano, "Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales", *Orinoquia*, vol. 25, n.º 1, pp. 25-30, jun. 2021, doi: 10.22579/20112629.680.

[9] B. Kang, S. Tripathi, y T. Nguyen, "Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map", sep. 2015.

[10] S.G. Prateek, J. Jagadeesh, R. Siddarth, Y. Smitha, Hiremath, P. G. Sunitha, Pendari and Neha Tarannum, “Dynamic Tool for American Sign Language Finger Spelling Interpreter”, en *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, oct. 2018, pp. 596-600. doi: 10.1109/ICACCCN.2018.8748859.

[11] L. B. R. de Guzmán, “La Deficiencia Auditiva. Identificación De Las Necesidades Educativas Especiales.”, *Rev. Int. Apoyo Inclusión Logop. Soc. Multicult.*, vol. 1, n.º 1, pp. 95-109, 2015.

[12] El Congreso de Colombia, "Ley 2049 de 2020 - Gestor Normativo - Función Pública ", https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=138150#:~:text=Objeto.,derechos%20ling%C3%BC%C3%ADsticos%20que%20le%20corresponden. (accedido 30 de marzo de 2023)

[13] INSOR, “Diccionario básico de la lengua de señas colombiana”, http://www.insor.gov.co/descargar/diccionario_basico_completo.pdf (accedido: 8 de febrero de 2023).

[14] J. Diaz y J. Pulgar, “MODULO 1 – LECCIÓN 2, CURSO BASICO-CONVERSACIONAL LENGUA DE SEÑAS COLOMBIANA L.S.C”. Grupo enseña S.A.S. [En línea]. Disponible en: www.eduacion-incluyente.com

[15] E. Villar *et al.*, *Artificial Intelligence*. IntechOpen, 2021. doi: 10.5772/intechopen.99289.

[16] PlanetadeLibros, *Inteligencia artificial - Lasse Rouhiainen | PlanetadeLibros*. Accedido: 8 de febrero de 2023. [En línea]. Disponible en: https://www.planetadelibros.com/libro-inteligencia-artificial/280581

[17] “Python Deep Learning - Second Edition”, *Packt*. https://www.packtpub.com/product/python-deep-learning-second-edition/9781789348460 (accedido 8 de febrero de 2023).

[18] “Deep Learning with TensorFlow”, *Packt*. https://www.packtpub.com/product/deep-learning-with-tensorflow/9781786469786 (accedido 8 de febrero de 2023).

[19] “Deep Learning with TensorFlow 2 and Keras - Second Edition”, *Packt*. https://www.packtpub.com/product/deep-learning-with-tensorflow-2-and-keras-second-edition/9781838823412 (accedido 8 de febrero de 2023).

[20] R. Yamashita, M. Nishio, R. K. G. Do, y K. Togashi, “Convolutional neural networks: an overview and application in radiology”, *Insights Imaging*, vol. 9, n.º 4, Art. n.º 4, ago. 2018, doi: 10.1007/s13244-018-0639-9.

[21] B. B. Traore, B. Kamsu-Foguem, y F. Tangara, “Deep convolution neural network for image recognition”, *Ecol. Inform.*, vol. 48, pp. 257-268, nov. 2018, doi: 10.1016/j.ecoinf.2018.10.002.

[22] P. Agustín Granell, “Redes Neuronales Recurrentes: Una aplicación para los mercados bursátiles”, jul. 2018, Accedido: 8 de febrero de 2023. [En línea]. Disponible en: http://diposit.ub.edu/dspace/handle/2445/124249

[23] I. Mindlin, “Reconocimiento de Lengua de Señas con redes neuronales recurrentes”, Tesis, Universidad Nacional de La Plata, 2021. Accedido: 7 de diciembre de 2022. [En línea]. Disponible en: http://sedici.unlp.edu.ar/handle/10915/129853

[24] C. Arana, “Redes neuronales recurrentes: Análisis de los modelos especializados en datos secuenciales”, Serie Documentos de Trabajo, Working Paper 797, 2021. Accedido: 8 de febrero de 2023. [En línea]. Disponible en: https://www.econstor.eu/handle/10419/238422

[25] C. Lugaresi *et al.*, “MediaPipe: A Framework for Building Perception Pipelines”. arXiv, 14 de junio de 2019. Accedido: 7 de febrero de 2023. [En línea]. Disponible en: http://arxiv.org/abs/1906.08172

[26] “Pose”, *mediapipe*. https://google.github.io/mediapipe/solutions/pose.html (accedido 1 de marzo de 2023).

[27] “Hands”, *mediapipe*. https://google.github.io/mediapipe/solutions/hands.html (accedido 7 de febrero de 2023).

[28] J. Shin, A. Matsuoka, M. A. M. Hasan, y A. Y. Srizon, “American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation”, *Sensors*, vol. 21, n.º 17, Art. n.º 17, ene. 2021, doi: 10.3390/s21175856.

[29] N. Kasukurthi, B. Rokad, S. Bidani, y D. A. Dennisan, “American Sign Language Alphabet Recognition using Deep Learning”. arXiv, 14 de mayo de 2019. doi: 10.48550/arXiv.1905.05487.

[30] X. Jiang, M. Lu, y S.-H. Wang, “An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language”, *Multimed. Tools Appl.*, vol. 79, n.º 21, pp. 15697-15715, jun. 2020, doi: 10.1007/s11042-019-08345-y.

[31] N. Ortiz-Farfán y J. E. Camargo-Mendoza, “Modelo computacional para reconocimiento de lenguaje de señas en un contexto colombiano”, *TecnoLógicas*, vol. 23, n.º 48, Art. n.º 48, may 2020, doi: 10.22430/22565337.1585.

[32] F. Martínez, F. Betancourt, y M. Arbulú, “A gesture recognition system for the colombian sign language based on convolutional neural networks”, *Bull. Electr. Eng. Inform.*, vol. 9, n.º 5, pp. 2082-2089, 2020, doi: 10.11591/eei.v9i5.2440.

[33] M. A. L. Barrera, D. A. L. Albán, y D. M. Torres, “Sistema de reconocimiento automático de lenguaje señas colombiano mediante Kinect y Leap Motion”, *Boletí-n Inf. CEI*, vol. 7, n.º 3, Art. n.º 3, nov. 2020.

[34] “Centro de relevo”, http://centroderelevo.gov.co/632/w3-channel.html (accedido 15 de diciembre de 2022).

[35] “INSOR | Instituto Nacional para Sordos – Trabajando por la Población Sorda Colombiana”. https://www.insor.gov.co/home/ (accedido 15 de diciembre de 2022).

[36] “ency-cross-validation.pdf”. http://leitang.net/papers/ency-cross-validation.pdf (accedido: 4 de febrero de 2023).

[37] “Streamlit Docs”. https://docs.streamlit.io/ (accedido 4 de febrero de 2023).

[38] “IncluSeñas - Aplicaciones en Google Play”. https://play.google.com/store/apps/details?id=app.aresan.miguel.inclusenas&hl=es&gl=CO (accedido 13 de diciembre de 2022).

[39] “PROFEenSEÑAS - Apps en Google Play”. https://play.google.com/store/apps/details?id=appinventor.ai_eduardoestebanperezleon.aprueba1_copy&hl=es_CO (accedido 2 de abril de 2023).

6. Anexos

6.1 Repositorio de códigos de desarrollo

El seguimiento de los algoritmos y herramientas desarrolladas están disponibles en *https://github.com/dromu/Alphanumeric_LSC*

6.2 Formato de consentimiento informado

CONSENTIMIENTO INFORMADO

Yo _____ mayor de edad, identificado con documento de identidad (o pasaporte) No. _____, doy mi consentimiento a JADER ALEJANDRO MUÑOZ GALINDEZ identificado con cedula de ciudadanía No. 1.007.638.734 de Popayán-Cauca, estudiante del programa de ingeniería física de la universidad del Cauca, para el uso o la reproducción de las secuencias filmadas en vídeo, fotografías o grabaciones de la voz de mi persona. Entiendo que el uso de la imagen o de la voz del participante, será principalmente para fines de investigación académica. Las secuencias filmadas pueden usarse para los siguientes fines:

- Creación base datos de señas de LSC
- Modelamiento computacional.
- Presentaciones de resultados en conferencias.

Se me informará acerca del uso de la grabación en vídeo o fotografías para cualquier otro fin, diferente a los anteriormente citados.

Entiendo que este material puede ser utilizado en diversos medios, incluyendo impresos y electrónicos. Esta autorización es continua y sólo podrá ser revocada por mi rescisión específica de esta autorización.

No existe ningún límite de tiempo en cuanto a la vigencia de esta autorización; ni tampoco existe ninguna especificación geográfica en cuanto a dónde se puede distribuir este material. Esta autorización se aplica a las secuencias filmadas en vídeo o fotografías que se puedan recopilar como parte del desarrollo del trabajo de grado en modalidad de investigación: **DESARROLLO DE UNA APLICACIÓN PARA EL RECONOCIMIENTO ALFANUMÉRICO DEL LENGUAJE DE SEÑAS COLOMBIANO (LSC) USANDO ALGORITMOS DE INTELIGENCIA ARTIFICIAL** y para los fines que se indican en este documento.

Como prueba de mi aceptación, se firma en Popayán-Cauca, a los _____ del mes _____ del año _____

Firma: _____

Nombre y apellidos: _____

Identificación: _____

Celular: _____