



UNIVERSIDAD ESAN  
FACULTAD DE INGENIERÍA  
INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

**Implementación de modelo de análisis de tendencias en noticieros por procesamiento de  
imágenes: Aplicación de Big Data Analytics**

Oscar Eduardo Salazar García  
Asesor: Junior Fabian Arteaga

Lima, 25 de julio de 2024

# **Índice general**

# **Índice de Figuras**

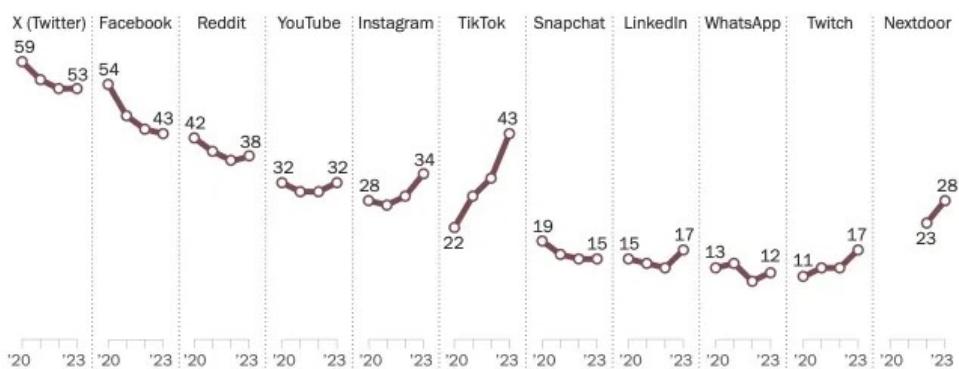
# **Índice de Tablas**

# Capítulo 1

## PLANTEAMIENTO DEL PROBLEMA

### 1.1. Descripción de la Realidad Problemática

En la era digital actual, el consumo de contenido audiovisual en redes sociales ha crecido exponencialmente, convirtiéndose en una fuente invaluable de información y entretenimiento. Una de los principales viene a ser la búsqueda de la actualidad del mundo, de acuerdo a una encuesta realizada por Pew Reasearch Center al sector adulto de EEUU, 3 de cada 10 adultos utilizan regularmente Youtube como plataforma de noticias, siguiendo Instagram y Tiktok. La cantidad de información que se genera día a día es abrumadora y el tiempo escaso.



**Figura 1.1:** Porcentaje de usuarios en base a su consumo de noticias en redes sociales en 2023.

Fuente: [PewReasearchCenter, 2023](#). Encuesta a adultos de EEUU realizada en 2023.

De acuerdo a la ??, se puede observar que plataformas con tendencia a ‘Más información en menos tiempo’ empiezan a ser las preferidas entre el público, observando el sorprendente aumento de plataformas como Instagram y Tiktok. Los medios informativos están tomando en cuenta estos cambios y han iniciado a informar en pocos segundos lo más relevante.

vante del día mediante videos cortos en sus redes sociales, para llegar a este tipo de público. Pero el principal desafío diario es cómo seleccionar el contenido que represente los hechos más importantes ocurrido en la actualidad.

Dentro de esta problemática se encuentra la gran cantidad de información audiovisual que se genera cada minuto y la limitada capacidad de realizar análisis para extraer información significativa. Este proceso implica enfrentarse a una serie de desafíos técnicos y operativos que requieren soluciones innovadoras y eficientes. En este contexto, las herramientas de Big Data se destacan como una solución crucial. Estas herramientas proporcionan la infraestructura necesaria para manejar, procesar y analizar grandes volúmenes de datos de manera efectiva.

Con el uso de herramientas como PySpark, es posible gestionar datos distribuidos a gran escala, lo que resulta fundamental para el análisis de videos. PySpark permite procesar los datos de video de manera eficiente, extrayendo y analizando frames para identificar eventos clave o patrones significativos. Esta capacidad es esencial dado que los métodos tradicionales de análisis no son suficientes para manejar la escala y complejidad del contenido audiovisual disponible.

En particular, YouTube juega un papel crucial en este análisis. Según Data Reportal en la ??, en el contexto peruano, YouTube es la segunda página más visitada, con un promedio de 227 millones de visitas mensuales y 11.4 millones de visitantes únicos al mes en 2023. YouTube se convierte así en una fuente inagotable de contenido relevante para el análisis de noticias y eventos actuales. La plataforma no solo alberga una vasta biblioteca de videos, sino que también proporciona metadatos valiosos, como descripciones, etiquetas y estadísticas de visualización, que enriquecen el análisis del contenido. Aplicar herramientas de big data en YouTube permite automatizar la extracción y el procesamiento de contenido, facilitando la selección de los clips más relevantes que capturan los eventos importantes del día.

Esta automatización optimiza tanto el tiempo como los recursos necesarios para curar contenido, mejorando la precisión en la selección de información significativa.

La generación de clips informativos a partir de videos largos presenta múltiples beneficios para los medios de comunicación. Primero, facilita la creación de resúmenes concisos y precisos que destacan los puntos clave de las noticias, permitiendo a los usuarios acceder a la información relevante de manera rápida. Este enfoque es crucial en un entorno donde la atención del usuario es limitada y la demanda de contenido digerible es alta. Esto no solo aumenta el alcance y el impacto de las noticias, sino que también mejora la participación del público y fomenta una mayor interacción con el contenido.

La automatización en la generación de clips también contribuye a la eficiencia operativa



**Figura 1.2:** Top webs más visitadas en Perú en 2023.

Fuente: [DataReportal, 2024](#). Basado en el tráfico web de Diciembre de 2022 a Noviembre del 2023.

de los medios. Al reducir la necesidad de intervención manual en la selección y edición de contenido, los recursos se liberan para enfocarse en otras áreas críticas, como la investigación y el análisis en profundidad. Esto permite a los medios mantenerse competitivos en un entorno en constante cambio y con una demanda creciente de contenido actualizado y relevante.

### 1.1.1. Problema General

¿De qué manera el uso de herramientas de Big Data y modelos de aprendizaje automático podría facilitar la selección y generación de clips informativos concisos a partir de contenido audiovisual en plataformas de redes sociales, optimizando la eficiencia operativa de los medios de comunicación?

### 1.1.2. Problemas Específicos

- ¿Cómo afecta la sobrecarga de información audiovisual en plataformas como YouTube la identificación de eventos relevantes y actuales para el público?
- ¿Qué técnicas y algoritmos de aprendizaje automático son más efectivos para automatizar la selección de clips de video que capturan los eventos importantes del día?
- ¿Cómo puede la implementación de herramientas de Big Data mejorar la eficiencia operativa de los medios de comunicación, reduciendo la necesidad de intervención manual en la selección y edición de contenido audiovisual?

- ¿Cuáles son las técnicas más adecuadas para el preprocesamiento y normalización de la base de datos de videos acerca de la actualidad?

## 1.2. Objetivos de la Investigación

### 1.2.1. Objetivo General

Desarrollar una metodología automatizada utilizando herramientas de Big Data y modelos de aprendizaje automático para procesar y analizar contenido audiovisual de YouTube, con el fin de generar clips informativos concisos que resuman los eventos más importantes del día.

### 1.2.2. Objetivos Específicos

- Optimizar tanto el tiempo como los recursos necesarios para el proceso completo de la herramienta.
- Mejorar la eficiencia operativa de los medios de comunicación reduciendo la necesidad de intervención manual en la selección y edición de contenido audiovisual, liberando recursos para que los medios se enfoquen en áreas críticas
- Desarrollar un sistema de análisis automatizado utilizando PySpark para manejar y procesar grandes volúmenes de datos de video, extrayendo y analizando frames para identificar eventos clave y patrones significativos.
- Seleccionar los procesos adecuados para mejorar los resultados y la calidad de los clips generados por el programa

## 1.3. Justificación de la Investigación

### 1.3.1. Teórica

Esta investigación se realiza para determinar cómo el uso de herramientas de Big Data y modelos de aprendizaje automático puede optimizar la selección y generación de clips informativos concisos a partir de contenido audiovisual en YouTube. La teorización se centra en la

capacidad de estas herramientas para manejar la sobrecarga de información y mejorar la eficiencia operativa en el proceso de curación de contenido. Este estudio contribuirá al cuerpo de conocimiento en el campo del análisis automatizado de contenido audiovisual, proponiendo un marco teórico para la implementación de tecnologías avanzadas en la industria de los medios de comunicación.

### **1.3.2. Práctica**

Al culminar la investigación, se espera que los medios de comunicación puedan utilizar el sistema automatizado basado en Big Data y aprendizaje automático para seleccionar y generar clips informativos relevantes de manera eficiente. Esto facilitará la creación de resúmenes concisos que destaqueen los puntos clave de las noticias, permitiendo a los usuarios acceder a información relevante rápidamente. La implementación práctica de esta solución tecnológica ayudará a los medios a adaptarse a las nuevas tendencias de consumo de contenido, donde la rapidez y concisión son cruciales. Además, mejorará la participación del público y fomentará una mayor interacción con el contenido.

### **1.3.3. Metodológica**

. El desarrollo del sistema automatizado de análisis de contenido audiovisual utilizando herramientas de Big Data y aprendizaje automático permitirá manejar grandes volúmenes de datos de video de manera eficiente. La metodología incluye el uso de PySpark para la gestión de datos distribuidos y técnicas de procesamiento de imágenes para la extracción de frames significativos. La implementación de algoritmos de clustering ayudará a identificar patrones y eventos relevantes. Esta metodología no solo optimiza el tiempo y los recursos necesarios para curar contenido, sino que también mejora la precisión en la selección de información significativa, ofreciendo una solución tecnológica avanzada para los desafíos actuales en el análisis de contenido audiovisual.

## **1.4. Delimitación del Estudio**

### **1.4.1. Espacial**

El estudio se llevará a cabo en el contexto de plataformas de redes sociales, enfocándose específicamente en YouTube como la principal fuente de contenido audiovisual para el análisis.

Se seleccionarán videos de noticias y eventos actuales relevantes para el público peruano, dado que YouTube es una de las páginas más visitadas en Perú, con un alto volumen de visitas mensuales y visitantes únicos. La investigación se centrará en la capacidad de YouTube para proporcionar tanto contenido audiovisual como metadatos valiosos para el análisis.

### **1.4.2. Temporal**

Los datos que se extraerán serán de la primera semana de Agosto, se realizará la recopilación, procesamiento y análisis. Al ser una herramienta de uso periódico, los resultados se generarán minutos posteriores a la culminación del proceso de recopilación.

### **1.4.3. Conceptual**

La presente investigación se centrará en el desarrollo de un sistema automatizado para la selección y generación de clips informativos a partir de contenido audiovisual utilizando herramientas de Big Data y modelos de aprendizaje automático. El enfoque estará en el análisis de videos de noticias en YouTube, aplicando técnicas de procesamiento de imágenes y algoritmos de clustering para identificar eventos relevantes.

## **1.5. Hipótesis**

### **1.5.1. Hipótesis General**

Mediante el uso de herramientas de Big Data y modelos de aprendizaje automático se logrará optimizar la selección y generación de clips informativos concisos a partir de contenido audiovisual en plataformas de redes sociales, mejorando la eficiencia operativa de los medios de comunicación.

### **1.5.2. Hipótesis Específicas**

- Mediante la implementación de técnicas de procesamiento de imágenes y clustering, se logrará una mejor identificación y selección de frames significativos, mejorando la calidad y relevancia de los clips informativos generados.
- El uso de técnicas y algoritmos avanzados de aprendizaje automático permitirá superar la

sobrecarga de información audiovisual, facilitando la identificación de eventos relevantes y actuales para el público.

- La aplicación de técnicas adecuadas de preprocesamiento y normalización de datos de videos aumentará la calidad de los clips generados, optimizando el rendimiento del sistema automatizado.
- La implementación de un sistema automatizado basado en Big Data y aprendizaje automático mejorará significativamente la precisión y relevancia de los clips de video seleccionados, demostrando la efectividad de estas técnicas.

### 1.5.3. Matriz de Consistencia

A continuación se presenta la matriz de consistencia elaborada para la presente investigación (véase Anexo ??).

# **Capítulo 2**

## **MARCO TEÓRICO**

### **2.1. Antecedentes de la investigación**

En esta sección se presentarán diversos artículos de investigación o tesis las cuales abordarán diversas técnicas y enfoques que se emplearon para afrontar problemas similares al de esta tesis. Asimismo, a continuación se presenta un cuadro resumen (véase Anexo ??) de lo que se presenta en esta sección.

#### **2.1.1. Efficient Video Classification Using Fewer Frames**

##### **2.1.1.1. Planteamiento del problema**

a necesidad de reducir las operaciones de punto flotante (FLOPs) y la huella de memoria. Aunque los modelos compactos actuales son más ligeros en términos de memoria, todavía requieren un número significativo de FLOPs ya que procesan todos los cuadros de un video. La propuesta de esta investigación es desarrollar modelos de clasificación de videos que procesen menos cuadros, disminuyendo así el número de FLOPs.

##### **2.1.1.2. Objetivos**

- Desarrollar modelos de clasificación de video que tengan una huella de memoria pequeña (menor a 1 GB) y que sean eficientes
- Crear modelos que procesen menos cuadros de video para reducir el número de FLOPs, manteniendo una alta eficiencia computacional.

- Utilizar un modelo maestro computacionalmente intensivo que procesa todos los cuadros para entrenar a un modelo estudiante más eficiente que procesa solo una fracción de los cuadros.
- Realizar una evaluación exhaustiva con tres tipos de modelos de clasificación de video: modelos recurrentes, modelos de agrupación y agregación, y modelos de agrupación y agregación eficientes en memoria.

### 2.1.1.3. Metodología

La metodología del documento incluye varias etapas, a partir de la recolección de datos hasta la implementación del modelo. A continuación, se presenta un resumen de la metodología:

1. **Recopilación de datos:** Se utilizó el conjunto de datos YouTube-8M (versión 2017), que contiene 8 millones de videos con múltiples clases asociadas a cada video. Los videos tienen una longitud promedio de 200 segundos.
2. **Entrenamiento del profesor y estudiante:** Se entrena una red neuronal compleja (profesor) que procesa todos los fotogramas de un video para generar una representación detallada del mismo. Se seleccionan solo algunos fotogramas del video (cada j-ésimo fotograma) para ser procesados por la red neuronal del estudiante. Se entrena una red neuronal más simple (estudiante) para que genere representaciones similares a las del profesor, utilizando solo los fotogramas seleccionados.
3. **Minimización de pérdidas:** Se minimizan las diferencias entre las representaciones del profesor y del estudiante mediante funciones de pérdida específicas, como la pérdida de error cuadrático.
4. **Evaluación y Comparación:** Se evalúa el rendimiento del estudiante en términos de precisión y eficiencia computacional, comparándolo con el modelo del profesor y otros métodos base.

### 2.1.1.4. Resultados

Los resultados obtenidos se presentan en un cuadro comparativo entre los diferentes modelos utilizados en el estudio. Los modelos comparativos 1 y 2 representan otras configuraciones de modelos de aprendizaje profundo probadas por los autores para comparar la efectividad de DeepASL.

**Tabla 2.1:** Resultados de la Clasificación de Videos Utilizando Menos Fotogramas

<b>Modelo</b>	<b>k</b>	<b>GAP</b>	<b>mAP</b>	<b>FLOPs (Billones)</b>
<b>Tiempo de Evaluación (hrs.)</b>				
Teacher-Skyline 13.00		0.811	0.414	5.058
Uniform-k 7.61	10	0.759	0.324	0.167
Uniform-k 8.20	20	0.785	0.363	0.268
Uniform-k 9.11	30	0.795	0.378	0.520

## 2.1.2. Sign Language Fingerspelling Recognition Using DepthInformation and Deep Belief Networks

### 2.1.2.1. Planteamiento del Problema

El reconocimiento de la ortografía manual en el lenguaje de señas, que implica traducir signos realizados con los dedos en texto. Este proceso es desafiante debido a la similitud visual de algunos signos, la invisibilidad de ciertos dedos y la variabilidad significativa en la representación de los signos por parte de diferentes usuarios. Este sistema traductor de lenguaje de señas contiene una técnica de profundidad utilizando redes de creencias profundas (DBNs) para la detección de dedos en el deletreo del lenguaje de señas. Se utilizan momentos de Zernike y histogramas de gradientes orientados (HOG) como características discriminativas.

### 2.1.2.2. Objetivos

- Desarrollar un marco de trabajo para el reconocimiento de la ortografía manual del lenguaje de señas.
- Utilizar información de profundidad e intensidad para mejorar la precisión del reconocimiento.
- Comparar el enfoque propuesto con otros métodos existentes en términos de precisión y rendimiento.
- Implementar una Red de Creencias Profundas (DBN) compuesta por tres Máquinas de Boltzmann Restringidas (RBM) como clasificador.

### 2.1.2.3. Fundamento Teórico

El reconocimiento de señas ha evolucionado significativamente desde los años 90, utilizando enfoques basados en sensores y visión. Este estudio se enfoca en un enfoque basado en visión, utilizando cámaras de profundidad como el Kinect de Microsoft. Se emplean descriptores de Histogramas de Gradientes Orientados (HOG) y momentos de Zernike como características discriminativas, conocidos por su simplicidad y buen rendimiento.

### 2.1.2.4. Metodología

- Segmentación de la imagen de profundidad: Utiliza un algoritmo adaptativo basado en k-means para separar el área de la mano del fondo.
- Extracción de características:
  - Se extraen descriptores HOG para capturar la información de la forma local en la imagen de intensidad.
  - Se utilizan momentos de Zernike para capturar la información de la forma global en la imagen de profundidad.
- Clasificación: Se emplea un DBN compuesto por tres RBM para clasificar los signos de la ortografía manual.
- Base de Datos:
  - **Dataset A:** Capturado en condiciones de iluminación constante por cinco usuarios, con un total de 120,000 imágenes (RGB y profundidad) representando 24 letras del alfabeto de ortografía manual.
  - **Dataset B:** Capturado con variaciones en la iluminación por nueve usuarios.

### 2.1.2.5. Resultados

**Tabla 2.2:** Comparación de Tasa de Clasificación y Desviación Estándar (STD) entre enfoques

Enfoque	Tasa de Clasificación (%)	Desviación Estándar (STD)
Pugeault y Bowden (2011)	75	-
Zhu y Wong (2012)	89.23	0.14
Otiniano et al. (2015)	91.26	0.17
Enfoque Propuesto	94.37	0.09

### 2.1.3. Deep learning-based sign language recognition system for static signs

#### 2.1.3.1. Planteamiento del problema

El reconocimiento automático del lenguaje de señas es una tarea desafiante debido a la variabilidad en las posturas de las manos, las condiciones ambientales y la necesidad de modelos robustos que puedan manejar estas variaciones. En particular, la lengua de señas india (ISL) no ha sido extensamente investigada, y no existe un sistema de reconocimiento en tiempo real para esta lengua. Este estudio aborda la creación de un sistema robusto para el reconocimiento de señas estáticas en ISL utilizando redes neuronales convolucionales (CNN).

#### 2.1.3.2. Objetivos

- Desarrollar un sistema robusto para el reconocimiento de señas estáticas en la lengua de señas india (ISL).
- Utilizar redes neuronales convolucionales (CNN) para mejorar la precisión y eficiencia del reconocimiento de señas.
- Evaluar la efectividad del sistema utilizando diferentes modelos de CNN y optimizadores.
- Comparar el rendimiento del sistema propuesto con otros enfoques existentes en términos de precisión, recall y F-score.

#### 2.1.3.3. Fundamento Teórico

Las redes neuronales convolucionales (CNN) han demostrado ser especialmente eficaces para la tarea de reconocimiento de imágenes debido a su capacidad para aprender jerarquías de características a diferentes niveles de abstracción. En este contexto, se han utilizado variaciones de CNN para el reconocimiento de gestos de la mano, como max-pooling CNN, 3D CNN, y sistemas híbridos que combinan CNN con modelos de Markov ocultos (HMM).

Este estudio se enfoca en la creación de un sistema de reconocimiento de señas estáticas utilizando una arquitectura de CNN optimizada para la lengua de señas india. El sistema propuesto emplea técnicas de preprocesamiento de imágenes, como la normalización y el redimensionamiento, y entrena el modelo utilizando una combinación de optimizadores como Adam, RMSprop y SGD.

#### 2.1.3.4. Metodología:

##### 1. Adquisición de datos:

- Se recopilaron un total de 35,000 imágenes de señas estáticas de 100 clases diferentes, incluyendo 23 letras del alfabeto inglés, dígitos del 0 al 10 y 67 palabras de uso común.

##### 2. Preprocesamiento de datos:

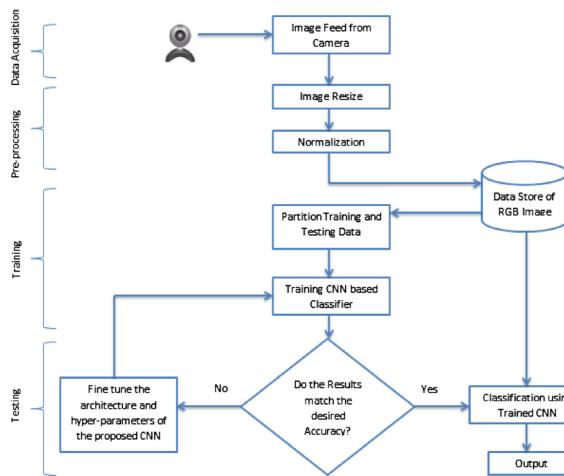
- Redimensionamiento de imágenes a 128x128 píxeles.
- Normalización para cambiar el rango de valores de intensidad de los píxeles.

##### 3. Entrenamiento del modelo:

- Se utilizó una arquitectura de CNN con capas convolucionales, capas ReLU, capas de pooling máximo y capas totalmente conectadas.
- El modelo se entrenó utilizando el conjunto de datos preprocesado, con una división de 80 % para entrenamiento y 20 % para validación.

##### 4. Pruebas:

- Se evaluó el rendimiento del sistema utilizando diferentes optimizadores (Adam, RMSprop, SGD) y se realizaron experimentos para ajustar el número de capas y filtros.



**Figura 2.1:** Metodología

#### 2.1.3.5. Resultados:

Los resultados del artículo se presentan en la siguiente tabla:

**Tabla 2.3:** Resultados experimentales con respecto a optimizadores y conjuntos de imágenes coloreadas

Modelo	Precisión de Entrenamiento (%)	Optimizador
I	99.17	Adam
II	99.59	RMSProp
III	99.72	SGD

## 2.1.4. Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras

### 2.1.4.1. Planteamiento del Problema

El reconocimiento y la traducción del lenguaje de señas en tiempo real presentan varios desafíos debido a la necesidad de captar y procesar movimientos complejos de las manos y el cuerpo en diferentes condiciones ambientales. Además, la mayoría de los sistemas existentes requieren hardware adicional o soporte de infraestructura, lo que limita su aplicabilidad en situaciones cotidianas. Este estudio aborda la creación de un sistema que utilice solo la cámara de profundidad integrada en los smartphones modernos para traducir el lenguaje de señas a texto en tiempo real.

### 2.1.4.2. Objetivos

- Desarrollar un sistema de traducción de lenguaje de señas en tiempo real utilizando únicamente la cámara de profundidad de los smartphones.
- Mejorar la precisión de la traducción mediante la incorporación de técnicas de procesamiento de imágenes de profundidad y aprendizaje profundo.
- Evaluar el rendimiento del sistema en diversas condiciones ambientales y con diferentes usuarios.
- Asegurar que el sistema funcione en tiempo real con baja latencia y sin la necesidad de servidores externos o dispositivos adicionales.

### 2.1.4.3. Fundamento Teórico

El reconocimiento del lenguaje de señas ha sido una tarea desafiante debido a la complejidad de los movimientos y la variabilidad en las condiciones de captura. Tradicionalmente,

los sistemas de reconocimiento de lenguaje de señas han utilizado cámaras RGB, sensores de movimiento y otros dispositivos adicionales para captar los gestos. Sin embargo, estos enfoques presentan limitaciones en cuanto a la privacidad y la necesidad de hardware adicional.

Las cámaras de profundidad ofrecen una solución potencial a estos problemas al proporcionar información sobre la distancia que es menos sensible a las condiciones de iluminación y más respetuosa con la privacidad. Las cámaras de profundidad capturan imágenes en escala de grises donde los valores de los píxeles representan la distancia desde la cámara al objeto, permitiendo una mejor segmentación y seguimiento de los gestos.

El uso de redes neuronales convolucionales tridimensionales (3DCNN) permite la extracción de características tanto espaciales como temporales de las secuencias de vídeo de profundidad. Este enfoque ha demostrado ser efectivo en la clasificación de actividades humanas y movimientos, lo que lo hace adecuado para la tarea de traducción del lenguaje de señas.

#### **2.1.4.4. Metodología**

##### **1. Adquisición de datos:**

- Se recopilaron un total de 5,000 secuencias de palabras de 50 palabras diferentes del lenguaje de señas coreano (KSL) de 20 participantes.
- Adicionalmente, se recopiló un conjunto de datos en diferentes condiciones ambientales para validar la robustez del sistema.

##### **2. Preprocesamiento de datos:**

- Redimensionamiento de las imágenes a una resolución uniforme.
- Aplicación de un módulo de énfasis en los gestos humanos para resaltar las partes relevantes del vídeo.

##### **3. Entrenamiento del modelo:**

- Uso de una arquitectura de 3DCNN basada en ResNet-18, preentrenada con el conjunto de datos Kinetics-400.
- Aplicación de técnicas de poda de filtros y cuantización de pesos para reducir el tamaño del modelo y el costo computacional.

##### **4. Aumento de datos:**

- Generación de un conjunto de datos aumentado mediante la introducción de ruido de movimiento para mejorar la robustez del modelo.

### 5. Segmentación de palabras:

- Implementación de un módulo de segmentación de palabras basado en una ventana deslizante para procesar secuencias de vídeo que contienen múltiples palabras.

#### 2.1.4.5. Resultados:

**Tabla 2.4:** Resultados experimentales con diferentes variaciones del modelo de inferencia de SUGO

Modelo	Precisión de Entrenamiento (%)
Sin Transfer Learning	74.56
Con Transfer Learning	87.34
Con Aumento de Datos	89.45
Con Poda de Modelo	90.12
Con Énfasis en Gestos Humanos	91.00

### 2.1.5. Using Deep Learning in Sign Language Translation to Text

#### 2.1.5.1. Planteamiento del Problema

La traducción del lenguaje de señas a texto es una tecnología crucial para aquellas personas que tienen dificultades para comunicarse, como aquellas con trastornos del habla, discapacidades auditivas o sordera. La complejidad de los gestos y la necesidad de precisión en la interpretación presentan un desafío significativo. Este estudio se enfoca en abordar estos desafíos mediante la implementación de técnicas de aprendizaje profundo para mejorar la precisión y eficiencia en la traducción del lenguaje de señas a texto.

#### 2.1.5.2. Objetivos

- Investigar y analizar diferentes técnicas de aprendizaje profundo aplicadas a la traducción del lenguaje de señas a texto.
- Desarrollar un sistema de traducción de lenguaje de señas a texto utilizando redes neuronales convolucionales (CNN), clasificación temporal conexionista (CTC) y redes de creencias profundas (DBN).

- Evaluar la efectividad del sistema propuesto comparando con estudios existentes.
- Proponer un modelo eficiente y preciso para la traducción de lenguaje de señas a texto.

### 2.1.5.3. Fundamento Teórico

La traducción del lenguaje de señas a texto es una tecnología que facilita la comunicación para personas con discapacidades auditivas o del habla. El reconocimiento del lenguaje de señas implica captar y entender los gestos complejos realizados por las manos y el cuerpo. Las técnicas tradicionales han utilizado métodos de machine learning que requieren la extracción manual de características, limitando así la precisión y la escalabilidad del sistema.

Con el avance del deep learning, se han desarrollado modelos que permiten la extracción automática de características a partir de datos brutos, mejorando significativamente la precisión y la eficiencia. Entre las técnicas más utilizadas se encuentran las redes neuronales convolucionales (CNN) para la extracción de características de imágenes, la clasificación temporal conexiónista (CTC) para la alineación secuencial y las redes de creencias profundas (DBN) para el modelado jerárquico de características.

Las CNN son especialmente efectivas en el reconocimiento de imágenes debido a su capacidad para aprender jerarquías de características a diferentes niveles de abstracción. La CTC es una técnica que permite la alineación de secuencias de entrada con secuencias de salida sin necesidad de segmentación previa, lo que es particularmente útil para la traducción de secuencias de gestos en texto. Las DBN, por otro lado, son modelos generativos que pueden capturar las dependencias complejas entre características y mejorar el rendimiento del sistema en tareas de reconocimiento.

### 2.1.5.4. Metodología

#### 1. Revisión de Literatura:

- Se realizó una revisión sistemática de la literatura utilizando bases de datos como IEEE Xplore, Elsevier, SpringerLink, Taylor and Francis Online y World Scientific.
- Se filtraron los estudios entre 2017 y 2021, seleccionando aquellos relevantes para el tema de la traducción de lenguaje de señas utilizando técnicas de deep learning.

## 2. Filtrado de Estudios:

- Inicialmente se obtuvieron 40 artículos, de los cuales se seleccionaron 20 tras una revisión detallada basada en el título y el resumen.

## 3. Desarrollo del Modelo:

- Se utilizaron técnicas de CNN, CTC y DBN para desarrollar el sistema de traducción de lenguaje de señas a texto.
- El modelo se entrenó y evaluó utilizando conjuntos de datos disponibles y métodos de validación cruzada.

## 4. Evaluación y Comparación:

- Se comparó el rendimiento del modelo propuesto con estudios existentes en términos de precisión, recall y F-score.
- Se realizaron ajustes y optimizaciones para mejorar la eficiencia y precisión del sistema.

### 2.1.5.5. Resultados:

**Tabla 2.5:** Resultados obtenidos con diferentes técnicas aplicadas

Técnica Aplicada	Precisión (%)	Recall (%)	F1-Score (%)
Redes Neuronales Convolucionales (CNN)	95	93	94
Clasificación Temporal Conexiónista (CTC)	92	90	91
Redes de Creencias Profundas (DBN)	89	87	88

## 2.1.6. Towards a Bidirectional Mexican Sign Language–Spanish Translation System: A Deep Learning Approach

### 2.1.6.1. Planteamiento del Problema

Las personas con discapacidades auditivas enfrentan barreras de comunicación significativas al interactuar con personas oyentes. Para abordar este problema, este artículo propone un sistema de traducción bidireccional de la lengua de señas mexicana (MSL) al español y viceversa. El sistema utiliza modelos de aprendizaje profundo como redes neuronales recurrentes

(RNN), RNN bidireccionales (BRNN), LSTM, GRU y Transformers para encontrar el modelo más preciso para el reconocimiento y la traducción del lenguaje de señas. La detección de puntos clave mediante MediaPipe se emplea para rastrear y comprender los gestos de la lengua de señas. El sistema presenta una interfaz gráfica fácil de usar con modos para traducir entre MSL y español en ambas direcciones. Los usuarios pueden ingresar signos o texto y obtener las traducciones correspondientes. La evaluación del rendimiento muestra una alta precisión, con el modelo BRNN logrando una precisión del 98.8%. La investigación enfatiza la importancia de las características de las manos en el reconocimiento de la lengua de señas. Los desarrollos futuros podrían centrarse en mejorar la accesibilidad y expandir el sistema para admitir otros lenguajes de señas. Este sistema de traducción ofrece una solución prometedora para mejorar la accesibilidad en la comunicación e impulsar la inclusión de personas con discapacidades auditivas.

#### **2.1.6.2. Objetivos**

- Desarrollar un sistema de traducción bidireccional que pueda traducir del español a la lengua de señas mexicana (MSL) y viceversa.
- Comparar modelos de aprendizaje profundo como RNN, BRNN, LSTM, GRU y Transformers para encontrar el más preciso para el reconocimiento y traducción del lenguaje de señas.
- Utilizar MediaPipe para la detección de puntos clave y la comprensión de los gestos de la lengua de señas.
- Evaluar el rendimiento del sistema en diversas condiciones y con diferentes usuarios.
- Mejorar la accesibilidad y la inclusión de personas con discapacidades auditivas.

#### **2.1.6.3. Fundamento Teórico**

La traducción automática del lenguaje de señas es un campo crucial para mejorar la comunicación y la inclusión de personas con discapacidades auditivas. Tradicionalmente, el reconocimiento del lenguaje de señas se ha basado en técnicas de visión por computadora y aprendizaje automático, utilizando sensores y cámaras para capturar los gestos de las manos y el cuerpo. Sin embargo, estos enfoques a menudo enfrentan desafíos debido a la variabilidad en los gestos y las condiciones ambientales.

El aprendizaje profundo ha revolucionado este campo al proporcionar modelos más robustos y precisos. Las redes neuronales recurrentes (RNN) y sus variantes, como las redes LSTM y GRU, son especialmente efectivas para tareas de secuencias temporales debido a su capacidad para mantener una memoria a largo plazo. Los modelos bidireccionales (BRNN y BLSTM) mejoran aún más esta capacidad al procesar la información en ambas direcciones, capturando mejor el contexto de los gestos. Los Transformers, una arquitectura más reciente, han demostrado un rendimiento sobresaliente en tareas de traducción y reconocimiento de secuencias debido a su capacidad para capturar dependencias a largo plazo de manera eficiente.

La detección de puntos clave mediante MediaPipe es una técnica avanzada que permite el seguimiento en tiempo real de los puntos de interés en las manos, la cara y el cuerpo, mejorando la precisión en la interpretación de los gestos de la lengua de señas. Este estudio combina estas técnicas avanzadas para desarrollar un sistema de traducción bidireccional entre MSL y español, evaluando su rendimiento en diversas condiciones para garantizar su robustez y eficacia.

#### **2.1.6.4. Metodología**

##### **1. Selección de Hardware:**

- Comparación de placas de computación como Raspberry Pi 4 Model B, Up Square y Nvidia Jetson Nano.
- Evaluación de la velocidad de inferencia de cada placa utilizando el modelo Holistic de MediaPipe.

##### **2. Selección de Características:**

- Definición de los puntos clave que son estadísticamente significativos para el proceso de inferencia del modelo.
- Reducción de puntos clave para optimizar los recursos computacionales del Jetson Nano.

##### **3. Recopilación de Datos:**

- Selección de diez frases aplicables en un entorno escolar.
- Recolección de aproximadamente 1000 muestras de cada signo de seis individuos.

##### **4. Definición del Modelo:**

- Implementación y evaluación de diferentes arquitecturas de RNN como Standard RNN, LSTM, BRNN, GRU y Transformers.
- Entrenamiento y evaluación de cada modelo para su efectividad en la clasificación de signos en MSL.

## 5. Diseño de la Interfaz Gráfica:

- Desarrollo de una interfaz gráfica de usuario (GUI) para mejorar la usabilidad y accesibilidad del sistema.

### 2.1.6.5. Resultados

**Tabla 2.6:** Resultados del modelo propuesto

Modelo	Épocas	Precisión (%)
RNN	44	94.8
BRNN	49	97.0
GRU	83	98.8
LSTM	70	98.8
BLSTM	31	98.6
Transformer	46	94.2

## 2.1.7. Computational Model for Sign Language Recognition in a Colombian Context

### 2.1.7.1. Planteamiento del Problema

Las personas sordas enfrentan significativas barreras de comunicación al interactuar con personas oyentes, especialmente en servicios públicos y privados. Actualmente, la plataforma en línea de intérprete de lenguaje de señas en Colombia no ofrece disponibilidad y confidencialidad completa. Este artículo propone un sistema que usa aprendizaje de máquina para reconocer automáticamente las señas colombianas y mejorar la comunicación, proporcionando un repositorio de imágenes de señas y un modelo computacional para predecir el significado de nuevas imágenes.

### 2.1.7.2. Objetivos

- Desarrollar un sistema de reconocimiento automático de señas colombianas.

- Crear un repositorio de imágenes de señas colombianas.
- Utilizar técnicas de aprendizaje de máquina para entrenar y evaluar un modelo de reconocimiento de señas.
- Comparar el rendimiento del modelo propuesto con otros modelos existentes.
- Proveer nuevas herramientas para mejorar la comunicación de las personas sordas con la población oyente.

#### 2.1.7.3. Fundamento Teórico

El reconocimiento del lenguaje de señas es crucial para mejorar la comunicación de personas sordas. En Colombia, la disponibilidad de intérpretes de señas es limitada, lo que dificulta la interacción con servicios públicos y privados. El aprendizaje de máquina y, en particular, el aprendizaje profundo han mostrado gran potencial en el reconocimiento de patrones complejos, como los gestos de las señas.

Las redes neuronales convolucionales (CNN) son efectivas para la tarea de reconocimiento de imágenes, ya que pueden aprender jerarquías de características visuales. La construcción de un repositorio de imágenes es esencial para entrenar estos modelos, permitiendo que el sistema reconozca una variedad de señas en diferentes condiciones.

Este estudio implementa un sistema de reconocimiento de señas basado en CNN, utilizando un repositorio de imágenes recopilado específicamente para este propósito. Las imágenes se procesan para normalizar su tamaño y se utilizan técnicas de aumento de datos para mejorar la robustez del modelo. El sistema propuesto se evalúa comparando su rendimiento con otros modelos existentes y analizando su precisión en diversas condiciones.

#### 2.1.7.4. Metodología

##### 1. Construcción del Repositorio de Imágenes:

- Recopilación de 3168 imágenes de 22 diferentes gestos del lenguaje de señas colombiano.
- Las imágenes se obtuvieron de videos grabados a cinco personas realizando las señas.

## 2. Procesamiento de Imágenes:

- Ajuste de tamaño de las imágenes a 320x240 píxeles.
- Conversión de imágenes a formato numérico para ser tratadas por la CNN.

## 3. Entrenamiento del Modelo:

- Uso de una arquitectura de CNN con capas convolucionales y de MaxPooling.
- Aplicación de técnicas de Grid Search y Cross Validation para optimizar los hiperparámetros del modelo.

## 4. Evaluación del Modelo:

- Comparación de diferentes modelos de CNN en términos de precisión y loss.
- Validación del modelo utilizando una técnica de Cross Validation con 5 pliegues.

### 2.1.7.5. Resultados

**Tabla 2.7:** Resultados del modelo propuesto

Modelo	Épocas	Precisión (%)
CNN sin Dropout	15	96.93
CNN con 25 neuronas y Tanh	15	98.8

## 2.1.8. Research of a Sign Language Translation System Based on Deep Learning

### 2.1.8.1. Planteamiento del Problema

La comunicación para personas con discapacidad auditiva y del habla es una barrera significativa debido a la falta de intérpretes de lenguaje de señas y la baja popularidad del lenguaje de señas entre la población general. El reconocimiento de lenguaje de señas puede mejorar esta situación al proporcionar una forma más conveniente de estudio, trabajo y vida para estas personas. Sin embargo, los métodos de reconocimiento de lenguaje de señas enfrentan desafíos debido a la variabilidad en los gestos, la complejidad del fondo y la iluminación. Este estudio propone un sistema de traducción de lenguaje de señas basado en aprendizaje profundo para abordar estos desafíos y mejorar la precisión del reconocimiento.

### 2.1.8.2. Objetivos

- Desarrollar un sistema de localización de manos basado en Faster R-CNN para reconocer las manos en imágenes de lenguaje de señas.
- Construir una red de extracción de características utilizando 3D CNN y un marco de reconocimiento de lenguaje de señas basado en LSTM.
- Integrar la red de localización de manos, la red de extracción de características 3D CNN y LSTM para construir un algoritmo de reconocimiento de lenguaje de señas.
- Evaluar la precisión del sistema propuesto en un conjunto de datos común y compararlo con otros métodos existentes.

### 2.1.8.3. Fundamento Teórico

El reconocimiento del lenguaje de señas es una tecnología vital para mejorar la comunicación de personas con discapacidad auditiva y del habla. Las técnicas tradicionales para el reconocimiento de señas a menudo enfrentan problemas debido a la variabilidad en los gestos y las condiciones ambientales. El aprendizaje profundo ha demostrado ser una solución prometedora, ofreciendo capacidades avanzadas de extracción de características y modelado.

Las redes neuronales convolucionales (CNN) han demostrado ser efectivas para la extracción de características visuales en imágenes y videos, permitiendo el reconocimiento preciso de patrones complejos. En este estudio, se utiliza una red Faster R-CNN para la localización de manos, lo que permite identificar y aislar las manos del fondo de la imagen, mejorando la precisión del reconocimiento de señas. La Faster R-CNN combina una red de propuestas de región (RPN) y una red de detección en una sola arquitectura, lo que mejora tanto la velocidad como la precisión de la detección.

Además, se emplea una red de CNN 3D para extraer características espaciales y temporales de secuencias de imágenes de señas, capturando información sobre el movimiento y la forma de las manos. La red LSTM se utiliza para el reconocimiento de secuencias de señas, aprovechando su capacidad para modelar dependencias a largo plazo en datos secuenciales. La combinación de estas tecnologías permite un reconocimiento robusto y preciso de señas, superando los desafíos planteados por las condiciones variables y complejas del entorno.

#### 2.1.8.4. Metodología

##### 1. Construcción del Conjunto de Datos:

- Recolección de un conjunto de datos de 40 palabras comunes en lenguaje de señas y 10,000 imágenes de señas.
- División del conjunto de datos en conjuntos de entrenamiento y validación en una proporción de 8:2.

##### 2. Entrenamiento de la Red de Localización de Manos:

- Utilización de Faster R-CNN para la localización de manos en imágenes de señas.
- Entrenamiento de la red utilizando el conjunto de datos recolectado y optimización con Stochastic Batch Gradient Descent (SGD).

##### 3. Extracción de Características con 3D CNN:

- Implementación de una red CNN 3D para la extracción de características de secuencias de imágenes de señas.
- Uso de bloques de convolución y pooling en la red para extraer características relevantes.

##### 4. Reconocimiento de Secuencias con LSTM:

- Integración de una red LSTM para el reconocimiento de secuencias de señas utilizando las características extraídas por la CNN 3D.
- Entrenamiento y ajuste de la red LSTM para mejorar la precisión del reconocimiento.

##### 5. Evaluación del Modelo:

- Evaluación de la precisión del sistema en el conjunto de datos de validación.
- Comparación de los resultados con otros métodos existentes para demostrar la eficacia del enfoque propuesto.

#### 2.1.8.5. Resultados

Los resultados obtenidos muestran que el método propuesto supera a los métodos existentes en términos de precisión. La tabla a continuación presenta la precisión de diferentes métodos evaluados en el estudio:

**Tabla 2.8:** Resultados de precisión de los métodos evaluados

Método	Precisión (%)
LSTM <sub>fc</sub>	91.6
3D-CNN	91.5
3D ResNet-18+SVM <sub>local</sub>	96.9
3D ResNet-18+SVM <sub>fusion</sub>	98.3
Método propuesto (Our method <sub>fusion</sub> )	99.0

## 2.1.9. A Real-time Portable Sign Language Translation System

### 2.1.9.1. Planteamiento del Problema

Las personas con discapacidades auditivas y del habla enfrentan barreras significativas en la comunicación debido a la falta de intérpretes de lenguaje de señas y la baja familiaridad con el lenguaje de señas entre la población general. Traducir el lenguaje de señas a voz o texto puede mejorar significativamente la comunicación. Sin embargo, los sistemas existentes de reconocimiento de gestos enfrentan desafíos como la variabilidad de los gestos y las condiciones ambientales complejas. Este artículo propone un sistema portátil de traducción en tiempo real del lenguaje de señas taiwanés utilizando guantes de reconocimiento de gestos con sensores flexibles e inerciales, y un teléfono inteligente para la traducción y síntesis de voz.

### 2.1.9.2. Objetivos

- Desarrollar un sistema de reconocimiento de gestos de manos en tiempo real y portátil para traducir el lenguaje de señas taiwanés.
- Utilizar sensores flexibles e inerciales integrados en un guante para capturar la postura de los dedos, la orientación de la palma y el movimiento de la mano.
- Implementar una arquitectura y algoritmo que permitan la traducción precisa de gestos a texto y voz.
- Evaluar la precisión del sistema propuesto en la traducción del lenguaje de señas.

### 2.1.9.3. Fundamento Teórico

El reconocimiento del lenguaje de señas es crucial para mejorar la comunicación de personas con discapacidades auditivas y del habla. Tradicionalmente, los sistemas de reconoci-

miento de gestos se han basado en sensores iniciales o sistemas basados en video. Los sistemas basados en sensores iniciales son portátiles y detectan los gestos mediante guantes equipados con sensores flexibles que capturan la flexión de los dedos y sensores iniciales que capturan la orientación de la palma y el movimiento de la mano.

Los sistemas basados en video utilizan cámaras y técnicas de visión por computadora para rastrear y reconocer los gestos. Aunque estos sistemas pueden proporcionar información útil, su precisión puede verse afectada por condiciones ambientales como la iluminación y el fondo complejo.

El estudio propone un sistema portátil y en tiempo real que combina las ventajas de los sensores flexibles e iniciales con un teléfono inteligente para la traducción del lenguaje de señas. Los sensores flexibles detectan la postura de los dedos, mientras que los sensores iniciales detectan la orientación de la palma y el movimiento de la mano. La señal procesada se transmite al teléfono inteligente mediante Bluetooth, donde se traduce a texto y voz utilizando una base de datos de lenguaje de señas y Google Translator.

#### **2.1.9.4. Metodología**

##### **1. Desarrollo del Guante de Reconocimiento de Gestos:**

- Integración de 10 sensores flexibles en el guante para detectar la flexión de los dedos.
- Integración de sensores iniciales (G-sensor y giroscopio) para detectar la orientación de la palma y el movimiento de la mano.

##### **2. Adquisición y Procesamiento de Señales:**

- Captura periódica y continua de las señales de los sensores flexibles e iniciales.
- Procesamiento de las señales para identificar gestos válidos mediante ciclos de reloj predefinidos.

##### **3. Transmisión de Señales y Traducción:**

- Transmisión de señales codificadas al teléfono inteligente mediante Bluetooth.
- Traducción de los gestos a texto mediante comparación con una base de datos de lenguaje de señas.

- Traducción del texto a voz utilizando Google Translator.

#### 4. Evaluación del Sistema:

- Pruebas del sistema con cinco sujetos y cinco palabras del lenguaje de señas.
- Evaluación de la precisión del sistema utilizando la métrica de Sensibilidad ( $TP / (TP + FN)$ ).

##### 2.1.9.5. Resultados

**Tabla 2.9:** Resultados de sensibilidad del sistema propuesto

Usuario	Gesto	TP	FN	Sensibilidad (%)
A	Lonely	46	4	96
B	Promote	49	1	98
C	Assist	47	3	94
D	Love	49	1	98
E	Protect	48	2	96

#### 2.1.10. Pattern Recognition and Synthesis for a Sign Language Translation System

##### 2.1.10.1. Planteamiento del Problema

Las personas con discapacidad auditiva enfrentan barreras significativas de comunicación debido a la falta de conocimiento generalizado del lenguaje de señas y la escasez de intérpretes. Este problema se agrava en situaciones de emergencia donde la asistencia inmediata de un intérprete no es viable. El estudio aborda el desarrollo de un sistema de traducción del lenguaje de señas japonés (JSL) al texto japonés y viceversa, utilizando técnicas de reconocimiento y síntesis de patrones. El sistema propuesto permitirá a las personas con discapacidad auditiva comunicarse directamente con personas oyentes sin necesidad de un intérprete, mejorando así su acceso a servicios y su integración social.

##### 2.1.10.2. Objetivos

- Desarrollar un subsistema de traducción del lenguaje de señas japonés a texto japonés.
- Crear un subsistema de traducción de texto japonés a animaciones de lenguaje de señas.

- Implementar técnicas de programación dinámica continua (DP) para el reconocimiento de gestos de señas en tiempo real.
- Evaluar la precisión del sistema en la traducción bidireccional entre JSL y texto japonés.
- Mejorar la accesibilidad y la comunicación para las personas con discapacidad auditiva en Japón.

#### 2.1.10.3. Fundamento Teórico

El lenguaje de señas es una forma de comunicación que utiliza gestos manuales, expresiones faciales y movimientos corporales para transmitir información. Las barreras de comunicación para las personas con discapacidad auditiva a menudo resultan en exclusión social y dificultades en el acceso a servicios esenciales. El reconocimiento de patrones y la síntesis de lenguaje de señas pueden resolver este problema mediante la automatización de la traducción entre lenguaje de señas y texto.

El sistema de traducción propuesto se compone de dos subsistemas: uno que traduce el lenguaje de señas japonés (JSL) a texto japonés y otro que traduce texto japonés a JSL mediante animaciones generadas por computadora. La entrada de datos se realiza a través de un guante de datos (DataGlove) que captura los movimientos y las formas de las manos. El reconocimiento de patrones utiliza técnicas de programación dinámica continua (DP) para emparejar los gestos de entrada con patrones predefinidos de señas.

La programación dinámica continua es una técnica utilizada en el reconocimiento de voz que se adapta bien al reconocimiento de señas debido a su capacidad para manejar secuencias de tiempo variables. Este método permite detectar palabras de señas dentro de una secuencia continua de gestos, mejorando la precisión del reconocimiento en tiempo real. Para la síntesis de señas, se utilizan modelos gráficos tridimensionales que generan animaciones de las señas basadas en patrones registrados previamente. Este enfoque facilita la generación de animaciones de señas naturales y comprensibles.

#### 2.1.10.4. Metodología

##### 1. Desarrollo del Subsistema de Traducción de Señal a Texto:

- Utilización de un guante de datos (DataGlove) para capturar los movimientos y formas de las manos.
- Implementación de técnicas de programación dinámica continua (DP) para el reconocimiento de patrones de señas.
- Emparejamiento de patrones de entrada con patrones predefinidos de señas mediante DP.

## 2. Desarrollo del Subsistema de Traducción de Texto a Señal:

- Conversión de texto japonés a etiquetas de palabras de señas utilizando reglas de construcción de oraciones de señas.
- Generación de animaciones de señas utilizando gráficos tridimensionales basados en patrones de movimiento registrados.
- Integración de expresiones faciales y movimientos corporales para mejorar la naturalidad de las animaciones.

## 3. Evaluación del Sistema:

- Prueba del sistema utilizando datos de entrenamiento y validación de gestos de señas.
- Medición de la precisión del reconocimiento y la velocidad de procesamiento del sistema.
- Comparación de los resultados con otros sistemas de reconocimiento de señas existentes.

### 2.1.10.5. Resultados

**Tabla 2.10:** Resultados de precisión y tiempo de reconocimiento

Técnica Utilizada	Precisión del Reconocimiento (%)
Sin Extracción de Características	100
Con Extracción de Características	97.3

## 2.2. Bases Teóricas

### 2.2.1. Inteligencia Artificial:

La inteligencia artificial (IA) es una rama de la informática dedicada al desarrollo de sistemas que pueden realizar tareas que típicamente requieren inteligencia humana. Estas tareas abarcan el reconocimiento de voz, la toma de decisiones, la traducción de idiomas y el reconocimiento de imágenes. La IA incluye diversas subdisciplinas, tales como Machine Learning, Deep Learning, el procesamiento del lenguaje natural (NLP) y Computer Vision.

### 2.2.2. Computer Vision:

Computer Vision una disciplina que permite a las máquinas interpretar y comprender el contenido de imágenes y videos. Las técnicas de visión por computadora se utilizan en una amplia gama de aplicaciones, incluyendo la detección de objetos, el reconocimiento facial, la segmentación de imágenes y la conducción autónoma.

### 2.2.3. Machine Learning:

El aprendizaje automático, conocido como Machine Learning, es un campo de la inteligencia artificial dedicado a crear algoritmos que capacitan a las máquinas para aprender, predecir y tomar decisiones basadas en datos. En vez de ser programadas con instrucciones detalladas para cada tarea, estas máquinas se entrena con datos para reconocer patrones y optimizar su rendimiento mediante la experiencia acumulada. Machine Learning permite que las máquinas mejoren su precisión y eficiencia continuamente a medida que procesan más datos.

### 2.2.4. Deep Learning:

Deep Learning es una subdisciplina del Machine Learning que se centra en el uso de redes neuronales profundas para modelar y aprender representaciones complejas de los datos. A diferencia de los métodos de aprendizaje automático tradicionales, que a menudo requieren características de entrada diseñadas manualmente, las técnicas de aprendizaje profundo permiten que las máquinas aprendan directamente de los datos brutos mediante la utilización de capas de procesamiento no lineales. Estas capas permiten la extracción de características de bajo nivel en las primeras capas y características más abstractas y de alto nivel en las capas superiores.

Este enfoque ha demostrado ser altamente eficaz en una variedad de aplicaciones, incluyendo visión por computadora, procesamiento de lenguaje natural, y reconocimiento de voz.

### **2.2.5. Aprendizaje Supervisado:**

El aprendizaje supervisado es un tipo de aprendizaje automático en el que un modelo se entrena utilizando un conjunto de datos etiquetado, es decir, un conjunto de datos donde cada entrada está asociada a una salida conocida. El objetivo del aprendizaje supervisado es aprender una función que, a partir de las entradas, prediga correctamente las salidas. Existen dos tipos principales de problemas en el aprendizaje supervisado: la clasificación, donde las salidas son categorías discretas, y la regresión, donde las salidas son valores continuos. Algunos algoritmos populares de aprendizaje supervisado incluyen regresión lineal, árboles de decisión, máquinas de vectores de soporte (SVM) y redes neuronales.

### **2.2.6. Aprendizaje No Supervisado y Semi-Supervisado:**

El aprendizaje no supervisado es una técnica de aprendizaje automático en la que un modelo se entrena utilizando un conjunto de datos que no tiene etiquetas. El objetivo es descubrir patrones y estructuras ocultas en los datos. Los métodos de aprendizaje no supervisado incluyen el clustering, la reducción de dimensionalidad, y los modelos generativos. El aprendizaje semi-supervisado, por otro lado, utiliza una combinación de datos etiquetados y no etiquetados para entrenar un modelo, aprovechando la información limitada de las etiquetas para mejorar el rendimiento en datos no etiquetados.

### **2.2.7. Redes Neuronales:**

Las redes neuronales son un tipo de algoritmos de aprendizaje automático que se inspiran en el funcionamiento del cerebro humano. Estas redes están configuradas para identificar patrones en los datos, lo que les permite aprender y tomar decisiones sin necesidad de programación explícita para cada tarea. Las redes neuronales se aplican en diversas áreas, como el reconocimiento de imágenes, el procesamiento del lenguaje natural, la traducción automática, entre otras muchas aplicaciones.

### **2.2.8. Redes Neuronales Convolucionales (CNN):**

Las redes neuronales convolucionales son una clase de redes neuronales particularmente efectivas para procesar datos con estructura de cuadrícula, como las imágenes. Las CNN utilizan capas convolucionales que aplican filtros a las entradas para extraer características espaciales importantes, seguidas de capas de pooling que reducen la dimensionalidad y la sensibilidad a la traslación. Estas características hacen que las CNN sean ideales para tareas de visión por computadora como la clasificación de imágenes, la detección de objetos y el reconocimiento de patrones. Las CNN han demostrado un rendimiento superior en competiciones y aplicaciones prácticas en comparación con métodos tradicionales de procesamiento de imágenes.

### **2.2.9. Redes Neuronales Recurrentes (RNN):**

Las redes neuronales recurrentes (RNN) son una clase de redes neuronales diseñadas para procesar datos secuenciales. A diferencia de las redes feedforward, las RNN tienen conexiones recurrentes que permiten que la información se mantenga y se utilice a lo largo de secuencias temporales. Esto las hace especialmente adecuadas para tareas como el procesamiento de lenguaje natural, la traducción automática, y el análisis de series temporales. Sin embargo, las RNN tradicionales pueden tener dificultades con secuencias largas debido a problemas de gradiente desvanecido. Para abordar esto, se han desarrollado variantes como las LSTM (Long Short-Term Memory) y las GRU (Gated Recurrent Units), que permiten a las redes capturar dependencias a largo plazo de manera más efectiva.

### **2.2.10. Optimización de Modelos:**

La optimización de modelos se refiere al proceso de ajustar los parámetros de un modelo de aprendizaje automático para mejorar su rendimiento. Este proceso incluye la selección del algoritmo de optimización adecuado, como el descenso de gradiente estocástico (SGD), Adam, o RMSprop, y la afinación de hiperparámetros, como la tasa de aprendizaje y el tamaño del lote. La optimización también puede incluir la implementación de técnicas de regularización, como la regularización L1 y L2, para prevenir el sobreajuste y mejorar la generalización del modelo. La optimización de modelos es una parte crucial del ciclo de desarrollo del aprendizaje automático y puede requerir múltiples iteraciones y experimentos para encontrar la configuración óptima.

### **2.2.11. Redes Neuronales de Procesamiento de Lenguaje Natural (NLP):**

Las redes neuronales de procesamiento de lenguaje natural son modelos diseñados para comprender y generar lenguaje humano. Estas redes incluyen arquitecturas como las redes neuronales recurrentes (RNN), las LSTM, las GRU, y los transformadores. Los modelos NLP se utilizan en una variedad de aplicaciones, incluyendo la traducción automática, el análisis de sentimientos, la generación de texto, y la respuesta a preguntas. Los transformadores, en particular, han revolucionado el campo del NLP con arquitecturas como BERT y GPT, que permiten el procesamiento eficiente de secuencias largas y la captura de contextos complejos.

### **2.2.12. Normalización y Estandarización:**

La normalización y la estandarización son técnicas de preprocesamiento de datos utilizadas para escalar las características de entrada a rangos específicos, mejorando así el rendimiento y la estabilidad del entrenamiento de los modelos de aprendizaje automático. La normalización se refiere al escalado de las características para que caigan dentro de un rango específico, típicamente  $[0, 1]$ . La estandarización implica reescalar las características para que tengan una media de 0 y una desviación estándar de 1. Ambas técnicas ayudan a que los algoritmos de optimización converjan más rápidamente y a que los modelos sean más robustos a diferentes escalas de características.

### **2.2.13. Data Augmentation:**

El data augmentation es una técnica utilizada para aumentar el tamaño del conjunto de datos de entrenamiento mediante la creación de versiones modificadas de las muestras existentes. Esta técnica es especialmente útil en el aprendizaje profundo, donde los grandes volúmenes de datos son cruciales para el rendimiento del modelo. En el contexto de imágenes, el data augmentation puede incluir transformaciones como rotación, escalado, traslación, y cambio de brillo o contraste. Estas transformaciones ayudan a que el modelo generalice mejor a nuevas muestras y previenen el sobreajuste al proporcionar una mayor variedad de ejemplos de entrenamiento.

## 2.3. Marco Conceptual

### 2.3.1. Lenguaje de Señas:

El lenguaje de señas es un sistema completo y natural de comunicación que utiliza gestos de las manos, expresiones faciales y movimientos del cuerpo para transmitir significados. Cada país o comunidad tiene su propio lenguaje de señas, como el American Sign Language (ASL) en los Estados Unidos y la Lengua de Señas Peruana (LSP) en Perú. A diferencia del lenguaje hablado, el lenguaje de señas es visual y espacial, lo que permite a las personas sordas o con problemas auditivos comunicarse de manera efectiva. Cada lenguaje de señas tiene su propia gramática y sintaxis únicas, reflejando la cultura y las particularidades de la comunidad sorda local.

### 2.3.2. Accesibilidad:

La accesibilidad se refiere a la práctica de hacer que los entornos, productos y servicios sean utilizables por todas las personas, independientemente de sus capacidades o discapacidades. En el contexto de la tecnología y la comunicación, la accesibilidad incluye la creación de sitios web, aplicaciones y dispositivos que puedan ser usados por personas con discapacidades visuales, auditivas, físicas o cognitivas. Esto puede incluir características como subtítulos en videos, texto alternativo para imágenes, interfaces de usuario accesibles y compatibilidad con tecnologías asistivas como lectores de pantalla y dispositivos de entrada alternativos. La accesibilidad es fundamental para garantizar la igualdad de oportunidades y la inclusión social.

### 2.3.3. Tecnología Asistiva:

La tecnología asistiva abarca una amplia gama de dispositivos, software y productos que ayudan a las personas con discapacidades a realizar tareas que podrían ser difíciles o imposibles de realizar sin asistencia. Estos dispositivos pueden incluir audífonos, lectores de pantalla, teclados alternativos, sillas de ruedas motorizadas y dispositivos de comunicación aumentativa y alternativa (CAA). La tecnología asistiva no solo mejora la calidad de vida de las personas con discapacidades, sino que también promueve su independencia y participación en la sociedad. La investigación y el desarrollo en este campo buscan continuamente nuevas soluciones para mejorar la accesibilidad y la inclusión.

### **2.3.4. Discapacidad Auditiva:**

La discapacidad auditiva incluye una variedad de condiciones que afectan la capacidad de una persona para oír. Esto puede variar desde una pérdida auditiva leve hasta la sordera profunda. Las personas con discapacidad auditiva pueden enfrentar desafíos significativos en la comunicación, la educación y la participación social. Sin embargo, con el uso de ayudas auditivas, implantes cocleares y tecnologías de comunicación como el lenguaje de señas y los subtítulos, muchas de estas barreras pueden ser superadas. La concienciación y la accesibilidad son esenciales para asegurar que las personas con discapacidad auditiva puedan participar plenamente en todas las áreas de la vida.

### **2.3.5. Inclusión Educativa:**

La inclusión educativa se refiere a la práctica de educar a todos los estudiantes, independientemente de sus habilidades o discapacidades, en entornos de aprendizaje comunes. Este enfoque promueve la igualdad de oportunidades y busca eliminar las barreras que impiden la participación plena de todos los estudiantes. Las estrategias para la inclusión educativa pueden incluir el diseño de currículos accesibles, el uso de tecnologías asistivas, la capacitación de maestros en educación especial y la implementación de apoyos y servicios individualizados. La inclusión educativa no solo beneficia a los estudiantes con discapacidades, sino que también enriquece la experiencia de aprendizaje para todos los estudiantes al promover la diversidad y la empatía.

### **2.3.6. Comunicación Inclusiva:**

La comunicación inclusiva es el enfoque de diseñar y transmitir mensajes de manera que sean accesibles y comprensibles para todas las personas, independientemente de sus habilidades o discapacidades. Esto puede incluir el uso de lenguaje claro y sencillo, formatos accesibles como braille o texto grande, y la provisión de servicios de interpretación de lenguaje de señas. En los medios digitales, la comunicación inclusiva también implica asegurar que los sitios web y las aplicaciones sean compatibles con tecnologías asistivas y que los contenidos multimedia incluyan subtítulos y descripciones de audio. La comunicación inclusiva es fundamental para garantizar que todas las personas puedan acceder a la información y participar plenamente en la sociedad.

### **2.3.7. Discapacidad Auditiva:**

La discapacidad auditiva incluye una variedad de condiciones que afectan la capacidad de una persona para oír. Esto puede variar desde una pérdida auditiva leve hasta la sordera profunda. Las personas con discapacidad auditiva pueden enfrentar desafíos significativos en la comunicación, la educación y la participación social. Sin embargo, con el uso de ayudas auditivas, implantes cocleares y tecnologías de comunicación como el lenguaje de señas y los subtítulos, muchas de estas barreras pueden ser superadas. La concienciación y la accesibilidad son esenciales para asegurar que las personas con discapacidad auditiva puedan participar plenamente en todas las áreas de la vida.

### **2.3.8. Tecnología de Reconocimiento de Señas**

La tecnología de reconocimiento de señas es un campo emergente que utiliza técnicas de visión por computadora e inteligencia artificial para interpretar el lenguaje de señas en tiempo real. Estas tecnologías pueden incluir cámaras y sensores que capturan los movimientos de las manos y el cuerpo, así como algoritmos que procesan y traducen estos movimientos en texto o voz. El reconocimiento de señas tiene el potencial de mejorar significativamente la comunicación entre personas sordas y oyentes, y de aumentar la accesibilidad en una variedad de contextos, desde la educación hasta el comercio y los servicios públicos.

### **2.3.9. Derechos de las Personas con Discapacidad**

Los derechos de las personas con discapacidad están protegidos por una serie de leyes y convenios internacionales que buscan garantizar la igualdad de oportunidades, la no discriminación y la plena inclusión en la sociedad. Uno de los marcos más importantes es la Convención sobre los Derechos de las Personas con Discapacidad (CRPD) de las Naciones Unidas, que establece estándares para la accesibilidad, la educación, el empleo y la participación social. A nivel nacional, muchos países tienen leyes que prohíben la discriminación por motivos de discapacidad y promueven la accesibilidad en el transporte, la vivienda y los servicios públicos.

### **2.3.10. Cultura Sorda**

La cultura sorda es un conjunto de prácticas, creencias y valores compartidos por las personas sordas que utilizan el lenguaje de señas como su principal medio de comunicación. Esta cultura celebra la identidad sorda y promueve la aceptación de la sordera no como una

discapacidad, sino como una característica única y valiosa. La comunidad sorda tiene su propio patrimonio, arte, literatura y tradiciones, que a menudo se transmiten a través de la narración de historias y la actuación en lenguaje de señas. La cultura sorda también aboga por los derechos de las personas sordas y trabaja para promover la accesibilidad y la inclusión en todos los aspectos de la sociedad.

# **Capítulo 3**

## **METODOLOGÍA DE LA INVESTIGACIÓN**

### **3.1. Diseño de la investigación**

En este segmento del documento se explica cual fue el tipo y enfoque del trabajo de investigación, al igual que la población y la muestra.

#### **3.1.1. Tipo de Investigación**

El diseño del presente trabajo de investigación es de tipo no experimental, ya que no se manipulan las variables independientes y se observa su relación en el contexto de la implementación de un modelo de traducción de lenguaje de señas utilizando técnicas de Deep Learning. La etapa inicial del trabajo involucra el análisis de vídeos de lenguaje de señas, los cuales serán procesados mediante técnicas de Deep Learning.

El nivel del presente trabajo de investigación es explicativo, dado que se enfoca en la implementación de un modelo de traducción de lenguaje de señas utilizando Deep Learning para personas con discapacidad del habla. El objetivo es determinar cómo se puede lograr la traducción efectiva del lenguaje de señas a texto y viceversa, estableciendo una relación entre los gestos y movimientos específicos del lenguaje de señas y su correspondencia con el texto escrito.

### 3.1.2. Enfoque de Investigación

El enfoque del presente trabajo de investigación es cuantitativo, dado que se basa en la utilización de instrumentos para la identificación y medición del lenguaje de señas a través de técnicas de visión computacional. Estas técnicas proporcionarán resultados numéricos que, tras un análisis estadístico, servirán como entrada para implementar modelos de Deep Learning.

## 3.2. Población y muestra

**Tabla 3.1:** Descripción del Estudio

Categoría	Descripción
<b>Población</b>	Personas con discapacidades del habla y del escucha.
<b>Muestra</b>	<ul style="list-style-type: none"> <li>■ 260 vídeos MP4 de diálogos en lenguaje de señas extraídas de Universidad Pontificia Universidad Católica del Perú.</li> <li>■ 355 vídeos MP4 de narración de historias en lenguaje de señas extraídas de Universidad Pontificia Universidad Católica del Perú.</li> <li>■ 103 vídeos MP4 de nombres, estados y acciones en lenguaje de señas extraídas de Universidad Pontificia Universidad Católica del Perú.</li> </ul>
<b>Unidad de análisis</b>	Un vídeo de una persona comunicándose con lenguaje de señas

## 3.3. Operacionalización de Variables

## 3.4. Técnicas de recolección

La base de datos se compone de vídeos en formato MP4 proporcionados por la Universidad Pontificia Universidad Católica del Perú. Estos vídeos fueron seleccionados por su

**Tabla 3.2:** Operacionalización de Variables

Variable	Dimensión	Indicador	Cálculo
<b>Independiente:</b> <b>Modelo Deep Learning</b>	Base de datos	Volumen de datos	Cantidad de videos MP4
	Indicadores de rendimiento de modelo	Precisión, recall, F1-score	Fórmulas

relevancia y calidad para el estudio del lenguaje de señas. Los videos MP4 contienen diálogos, narraciones de historias, y nombres, estados y acciones en lenguaje de señas.

## 3.5. Técnicas para el procesamiento y análisis de la información

### 3.5.1. Metodología de la implementación de la solución

#### 3.5.1.1. Recolección de Datos

La recolección de datos es una fase crítica en cualquier proyecto de Deep Learning. En esta tesis, se han recopilado un total de 718 vídeos MP4 de lenguaje de señas, extraídos de la Pontificia Universidad Católica del Perú. Los datos se dividen en tres categorías principales:

- **260 vídeos MP4 de diálogos en lenguaje de señas:** Estos vídeos contienen conversaciones entre dos o más personas utilizando lenguaje de señas.
- **355 vídeos MP4 de narración de historias en lenguaje de señas:** En estos vídeos, las personas narran historias completas usando lenguaje de señas.
- **103 vídeos MP4 de nombres, estados y acciones en lenguaje de señas:** Estos vídeos muestran la señalización de nombres, estados emocionales y diversas acciones.

### 3.5.1.2. Preprocesamiento de Datos

El preprocesamiento de datos es una etapa esencial para preparar los datos brutos para ser utilizados por los modelos de Deep Learning. Las tareas de preprocesamiento incluyen:

- **Extracción de Frames:** Se extraen frames de los vídeos a intervalos regulares para capturar las diferentes poses y movimientos de las manos.
- **Aumento de Datos:** Se aplican técnicas de aumento de datos, como rotación, escalado y traslación, para aumentar la variabilidad y robustez del conjunto de datos.
- **Etiquetado:** Cada frame o secuencia de frames se etiqueta con la correspondiente señal del lenguaje de señas que representa.
- **Normalización:** Los frames se normalizan para asegurarse de que todos los datos tengan el mismo formato y escala.

### 3.5.1.3. Uso de Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales (CNN) son efectivas para la tarea de reconocimiento de imágenes y, en este caso, para la identificación de señales en los frames extraídos de los vídeos. Las etapas clave incluyen:

- **Diseño de la Arquitectura CNN:** Se diseña una arquitectura de CNN adecuada para la tarea, con capas convolucionales, de pooling y totalmente conectadas.
- **Entrenamiento:** La CNN se entrena utilizando los datos preprocesados, ajustando los pesos de la red para minimizar el error en la predicción de las señales.
- **Validación:** Se valida el rendimiento de la CNN utilizando un conjunto de datos de validación separado, ajustando los hiperparámetros según sea necesario.

### 3.5.1.4. Integración de LSTM

Las redes neuronales de memoria a largo plazo (LSTM) son ideales para manejar secuencias de datos y capturar dependencias temporales. En esta fase:

- **Extracción de Secuencias:** Se crean secuencias de frames para capturar el movimiento y la continuidad de las señales.

- **Diseño de la Arquitectura LSTM:** Se diseña una arquitectura LSTM que pueda procesar las secuencias de frames y capturar las dependencias temporales entre ellos.
- **Entrenamiento del Modelo:** El modelo LSTM se entrena para aprender las transiciones y la continuidad de las señales en las secuencias de frames.
- **Combinación con CNN:** La salida de la CNN se combina con la LSTM para mejorar la precisión en la predicción de señales.

### 3.5.1.5. Evaluación del Modelo

La evaluación del modelo es crucial para determinar su eficacia y áreas de mejora. Las actividades incluyen:

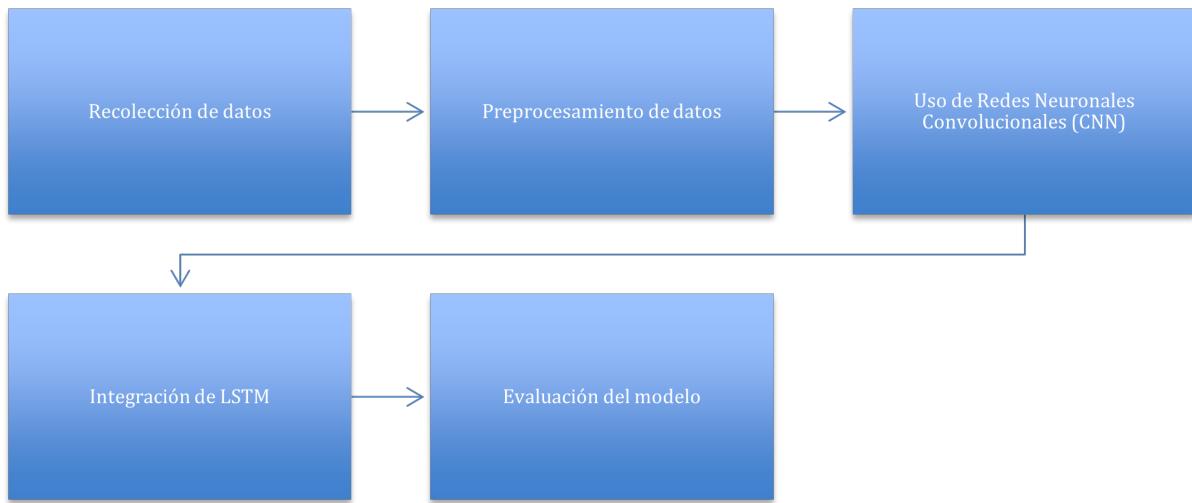
- **Métricas de Evaluación:** Se utilizarán métricas como Recall, F1-Score y Accuracy para evaluar el rendimiento del modelo.
- **Pruebas con Datos de Prueba:** Se prueba el modelo con un conjunto de datos de prueba que no se ha utilizado durante el entrenamiento para evaluar su rendimiento en datos no vistos.
- **Análisis de Errores:** Se realiza un análisis de errores para identificar patrones comunes en las predicciones incorrectas y mejorar el modelo.
- **Ajustes y Mejoras:** Basado en los resultados de la evaluación, se ajustan los hiperparámetros y se realizan mejoras en la arquitectura del modelo para aumentar su rendimiento.

## 3.5.2. Metodología para la medición de resultados

### Accuracy

El accuracy medirá la proporción de predicciones correctas sobre el total de predicciones realizadas por el modelo Deep Learning. Se utilizará para obtener una visión general del rendimiento del modelo.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{Ecuación 3.1})$$



**Figura 3.1:** Metodología Iterativa

## Recall

El Recall medirá la capacidad del modelo para identificar correctamente todas las instancias positivas.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (\text{Ecuación 3.2})$$

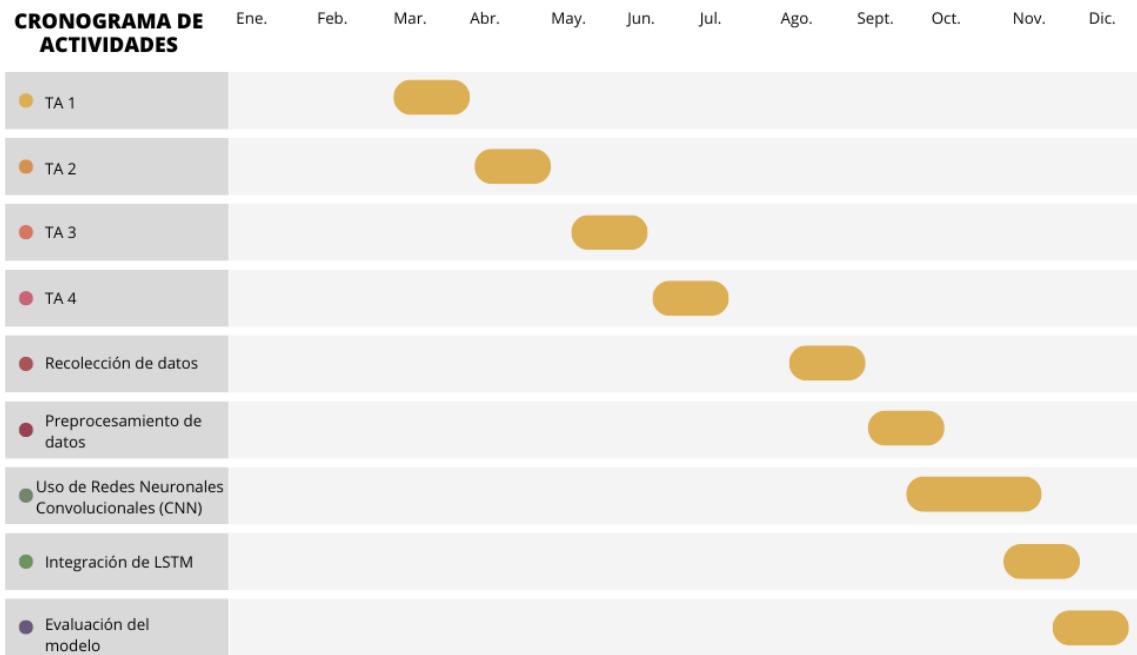
## F1-Score

Se utilizará en caso el modelo presente un desequilibrio entre las clases y se necesita un equilibrio tanto en el Accuracy como en el recall.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{Ecuación 3.3})$$

## 3.6. Cronograma de actividades

## 3.7. Presupuesto

**Figura 3.2:** Metodología Iterativa**Tabla 3.3:** Presupuesto del Proyecto

Item	Costo (Soles)
Laptop	3500
Cámara	250
<b>Total</b>	<b>3750</b>

## **Anexos**

## **Anexos A**

### **Anexo I: Matriz de Consistencia**

PROBLEMAS	OBJETIVOS	HIPÓTESIS
Problema General	Objetivo General	Hipótesis General
¿De qué manera el uso de herramientas de Big Data y modelos de aprendizaje automático podría facilitar la selección y generación de clips informativos concisos a partir de contenido audiovisual en plataformas de redes sociales, optimizando la eficiencia operativa de los medios de comunicación?	Desarrollar una metodología automatizada utilizando herramientas de Big Data y modelos de aprendizaje automático para procesar y analizar contenido audiovisual de Youtube, con el fin de generar clips informativos concisos que resuman los eventos más importantes del día.	Mediante el uso de herramientas de Big Data y modelos de aprendizaje automático se logrará optimizar la selección y generación de clips informativos concisos a partir de contenido audiovisual en plataformas de redes sociales, mejorando la eficiencia operativa de los medios de comunicación.
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas
¿Cómo afecta la sobrecarga de información audiovisual en plataformas como YouTube la identificación de eventos relevantes y actuales para el público?	Optimizar tanto el tiempo como los recursos necesarios para el proceso completo de la herramienta.	Mediante la implementación de técnicas de procesamiento de imágenes y clustering, se logrará una mejor identificación y selección de frames significativos, mejorando la calidad y relevancia de los clips informativos generados.
¿Qué técnicas y algoritmos de aprendizaje automático son más efectivos para automatizar la selección de clips de video que capturan los eventos importantes del día?	Mejorar la eficiencia operativa de los medios de comunicación reduciendo la necesidad de intervención manual en la selección y edición de contenido audiovisual, liberando recursos para que los medios se enfoquen en áreas críticas	El uso de técnicas y algoritmos avanzados de aprendizaje automático permitirá superar la sobrecarga de información audiovisual, facilitando la identificación de eventos relevantes y actuales para el público.
¿Cómo puede la implementación de herramientas de Big Data mejorar la eficiencia operativa de los medios de comunicación, reduciendo la necesidad de intervención manual en la selección y edición de contenido audiovisual?	Desarrollar un sistema de análisis automatizado utilizando PySpark para manejar y procesar grandes volúmenes de datos de video, extrayendo y analizando frames para identificar eventos clave y patrones significativos.	La aplicación de técnicas adecuadas de preprocesamiento y normalización de datos de videos aumentará la calidad de los clips generados, optimizando el rendimiento del sistema automatizado.

**Tabla A.1:** Matriz de consistencia. Fuente: Elaboración propia

Problemas	Objetivos	Hipótesis	Variables	Indicadores		Índice	
				Gestos y signos correctamente interpretados	Total de gestos y signos		
Problema General ¿De qué manera el uso de un modelo Deep Learning podría facilitar la comunicación para personas con discapacidades del habla para interactuar con personas que no conocen el lenguaje de señas peruanas?	Objetivo General Desarrollar un modelo Deep Learning que se utilizará como medio para la traducción de lenguaje de señas, de esta manera mejorar la comunicación entre personas con discapacidades del habla y personas sin conocimiento del lenguaje de señas.	Hipótesis General Mediante el desarrollo de un modelo de traducción de lenguaje de señas basado en Deep Learning se logrará mejorar la comunicación para personas con discapacidades y calidad de vida de las personas con discapacidades de habla al utilizar el modelo de traducción de lenguaje de señas, mejorando así su accesibilidad y calidad de vida.	X Uso de un modelo Deep Learning	Proporción de gestos y signos correctamente interpretados por el modelo de Deep Learning	Gestos y signos correctamente interpretados Total de gestos y signos		
Problema Específico ¿De qué manera la falta de conjuntos de datos de lenguaje de señas peruanos podría afectar al modelo Deep Learning?	Objetivo Específico Evaluar diferentes estrategias para la carencia de falta de datos para mejorar la representación de los conjuntos de datos de lenguaje de señas.	Hipótesis Específica Mediante el uso diferentes estrategias para afrontar la ausencia de datos, mejorar la representación de los conjuntos de datos disponibles y compensar en parte la falta de datos específicos para el español peruano.	X1 Impacto de la falta de conjuntos de datos de lenguaje de señas	Porcentaje de gestos y signos correctamente interpretados por el modelo	Nivel de satisfacción percibido Máximo nivel de satisfacción posible		
Problema Específico ¿De qué manera el modelo Deep Learning puede diferenciar entre los distintos tipos de lenguajes de señas?	Objetivo Específico Utilizar técnicas Deep learning para mejorar la precisión del modelo en la diferenciación entre los distintos tipos de lenguajes de señas.	Hipótesis Específica El modelo Deep Learning aumentará su precisión significativa con lo que respecta de lenguaje de señas peruanas, lo que demuestra la eficacia de las técnicas.	X2 Capacidad del modelo Deep Learning para diferenciar entre distintos tipos de lenguajes de señas	Porcentaje de correcta diferenciación entre distintos tipos de lenguajes de señas por el modelo Deep Learning	Precisión después de aplicar técnicas de aumento de datos Precisión antes de aplicar técnicas de aumento de datos		
Problema Específico ¿Qué métricas son las más adecuadas para la precisión y rendimiento de un modelo de traducción de lenguaje de señas?	Objetivo Específico Evaluar diferentes métricas de evaluación de modelos Deep Learning para la determinación del modelo más adecuado para la traducción de lenguaje de señas.	Hipótesis Específica La implementación de métricas de evaluación en los modelos Deep Learning aumentará las diferencias significativas entre los diferentes modelos evaluados.	X3 Métricas de evaluación de modelos Deep Learning	Precisión del modelo en la diferenciación entre los distintos tipos de lenguajes de señas	Correcta diferenciación entre lenguajes de señas Total de diferenciaciones intentadas		
Problema Específico ¿Cuáles son las técnicas más adecuadas para la precisión y rendimiento de un modelo de traducción de lenguaje de señas?	Objetivo Específico Realizar comparaciones entre diferentes técnicas de preprocesamiento y normalización de datos de lenguaje de señas.	Hipótesis Específica Mediante la implementación de técnicas de preprocesamiento, mejorar la calidad de los datos de lenguaje de señas peruanos y con ello aumenta el rendimiento del modelo Deep Learning para la traducción de lenguaje de señas y en menor.	X4 Técnicas de preprocesamiento	Valor numérico de la precisión obtenida por el modelo	Número de predicciones correctas Número total de predicciones		
			Y4 Calidad de los datos	Valor numérico de la precisión	No aplica	Número total de técnicas de preprocesamiento seleccionadas	Número total de técnicas de preprocesamiento seleccionadas
					Mejoría en la calidad de los datos - Alumentado en el rendimiento del modelo		

**Figura A.1:** Matriz de consistencia Completo. Fuente: Elaboración propia

## **Anexos B**

### **Anexo II: Resumen de Papers investigados**

Tipo	Nº	Título	Autor	Año	País	Fuente
Problema	1	DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation	Biyi Fang, Jililian Co, Mi Zhang	2018	USA	Michigan State University
	2	Sign Language Fingerspelling Recognition Using Depth Information and Deep Belief Networks	Hai-Feng Zhao	2017	China	Huaiyin Institute of Technology
Propuesta	3	Deep learning-based sign language recognition system for static signs	Ankita Wadhawan	2019	USA	Neural Computing and Applications
	4	Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras	HYEONJUNG PARK	2021	Sout Korea	School of Integrated Technology
Técnica	5	Computational Model for Sign Language Recognition in a Colombian Context	Nelson Ortiz-Farfán, Jorge E. Camargo-Mendoza	2020	Colombia	2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)
	6	Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing	Yun, Sim, Seok	2019	Japan	2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)
	7	A Method of Optimizing LDA Result Purity Based on Semantic Similarity	Jingrui, Z., Qinglin, W., Yu, L., Yuan, L.	2017	China	2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)
	8	Qualitative Stock Market Predicting with Common Knowledge Based Nature Language Processing: A Unified View and Procedure	Rao, D., Deng, F., Jiang, Z., Zhao, G.	2015	USA	2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics
	9	Fuzzy Bag-of-Words Model for Document Representation	Zhao, R., Mao, K.	2018	USA	IEEE Transactions on Fuzzy Systems ( Volume: 26 , Issue: 2 , April 2018 )

**Tabla B.1:** Cuadro Resumen de Papers investigados. Fuente: Elaboración propia

# BIBLIOGRAFÍA

DataReportal. (2024). Top Websites: Similarweb ranking Peru. <https://datareportal.com/reports/digital-2024-peru?rq=Peru>

PewResearchCenter. (2023). Social media sites by portion of users who regularly get news there. <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>