



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

**Implementación de un modelo de Deep Learning para la traducción de lenguaje de señas
para personas con discapacidades del habla**

Adrian Gómez Sánchez Bendezú AA
Asesor: Marks Calderón

Lima, 3 de junio de 2024

Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ac odio tempor orci dapibus ultrices in iaculis nunc sed. Vivamus arcu felis bibendum ut tristique et egestas quis ipsum. Odio morbi quis commodo odio aenean sed adipiscing diam donec. Donec ultrices tincidunt arcu non sodales neque sodales ut. Fusce ut placerat orci nulla pellentesque dignissima enim sit amet. Facilisi etiam dignissima diam quis enim lobortis. Sit amet justo donec enim diam vulputate ut pharetra. Gravida in fermentum et sollicitudin ac orci phasellus egestas. Ultricies tristique nulla aliquet enim tortor at auctor. Nullam vehicula ipsum a arcu cursus vitae congue mauris. Convallis posuere morbi leo urna molestie at elementum eu facilisis. Elit at imperdiet dui accumsan sit amet nulla. Amet consectetur adipiscing elit pellentesque habitant morbi tristique senectus et. Mauris in aliquam sem fringilla ut morbi. Ultricies integer quis auctor elit sed vulputate mi sit. Nulla pellentesque dignissima enim sit amet venenatis urna cursus eget. Ac feugiat sed lectus vestibulum mattis ullamcorper. Eu augue ut lectus arcu bibendum. Rhoncus dolor purus non enim praesent elementum.

Nulla facilisi cras fermentum odio eu feugiat pretium. Massa massa ultricies mi quis hendrerit. Id leo in vitae turpis massa sed elementum. Quis vel eros donec ac odio tempor orci. Netus et malesuada fames ac turpis egestas integer eget aliquet. Velit ut tortor pretium viverra suspendisse potenti. Ut enim blandit volutpat maecenas. Nibh tellus molestie nunc non blandit. Mus mauris vitae ultricies leo integer malesuada nunc vel. Vel elit scelerisque mauris pellentesque pulvinar pellentesque habitant. Neque viverra justo nec ultrices dui sapien eget. Vitae aliquet nec ullamcorper sit. Dui id ornare arcu odio ut sem nulla pharetra diam. Et magnis dis parturient montes. Varius morbi enim nunc faucibus.

Palabras claves: uno, dos, tres, cuatro

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ac odio tempor orci dapibus ultrices in iaculis nunc sed. Vivamus arcu felis bibendum ut tristique et egestas quis ipsum. Odio morbi quis commodo odio aenean sed adipiscing diam donec. Donec ultrices tincidunt arcu non sodales neque sodales ut. Fusce ut placerat orci nulla pellentesque dignissima enim sit amet. Faciliti etiam dignissima diam quis enim lobortis. Sit amet justo donec enim diam vulputate ut pharetra. Gravida in fermentum et sollicitudin ac orci phasellus egestas. Ultricies tristique nulla aliquet enim tortor at auctor. Nullam vehicula ipsum a arcu cursus vitae congue mauris. Convallis posuere morbi leo urna molestie at elementum eu facilisis. Elit at imperdiet dui accumsan sit amet nulla. Amet consectetur adipiscing elit pellentesque habitant morbi tristique senectus et. Mauris in aliquam sem fringilla ut morbi. Ultricies integer quis auctor elit sed vulputate mi sit. Nulla pellentesque dignissima enim sit amet venenatis urna cursus eget. Ac feugiat sed lectus vestibulum mattis ullamcorper. Eu augue ut lectus arcu bibendum. Rhoncus dolor purus non enim praesent elementum.

Nulla facilisi cras fermentum odio eu feugiat pretium. Massa massa ultricies mi quis hendrerit. Id leo in vitae turpis massa sed elementum. Quis vel eros donec ac odio tempor orci. Netus et malesuada fames ac turpis egestas integer eget aliquet. Velit ut tortor pretium viverra suspendisse potenti. Ut enim blandit volutpat maecenas. Nibh tellus molestie nunc non blandit. Mus mauris vitae ultricies leo integer malesuada nunc vel. Vel elit scelerisque mauris pellentesque pulvinar pellentesque habitant. Neque viverra justo nec ultrices dui sapien eget. Vitae aliquet nec ullamcorper sit. Dui id ornare arcu odio ut sem nulla pharetra diam. Et magnis dis parturient montes. Varius morbi enim nunc faucibus.

Keywords: uno, dos, tres, cuatro

Para mi X, Y,X

Agradecimientos

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ac odio tempor orci dapibus ultrices in iaculis nunc sed. Vivamus arcu felis bibendum ut tristique et egestas quis ipsum. Odio morbi quis commodo odio aenean sed adipiscing diam donec. Donec ultrices tincidunt arcu non sodales neque sodales ut. Fusce ut placerat orci nulla pellentesque dignissima enim sit amet. Faciliti etiam dignissima diam quis enim lobortis. Sit amet justo donec enim diam vulputate ut pharetra. Gravida in fermentum et sollicitudin ac orci phasellus egestas. Ultricies tristique nulla aliquet enim tortor at auctor. Nullam vehicula ipsum a arcu cursus vitae congue mauris. Convallis posuere morbi leo urna molestie at elementum eu facilisis. Elit at imperdiet dui accumsan sit amet nulla. Amet consectetur adipiscing elit pellentesque habitant morbi tristique senectus et. Mauris in aliquam sem fringilla ut morbi. Ultricies integer quis auctor elit sed vulputate mi sit. Nulla pellentesque dignissima enim sit amet venenatis urna cursus eget. Ac feugiat sed lectus vestibulum mattis ullamcorper. Eu augue ut lectus arcu bibendum. Rhoncus dolor purus non enim praesent elementum.

Nulla facilisi cras fermentum odio eu feugiat pretium. Massa massa ultricies mi quis hendrerit. Id leo in vitae turpis massa sed elementum. Quis vel eros donec ac odio tempor orci. Netus et malesuada fames ac turpis egestas integer eget aliquet. Velit ut tortor pretium viverra suspendisse potenti. Ut enim blandit volutpat maecenas. Nibh tellus molestie nunc non blandit. Mus mauris vitae ultricies leo integer malesuada nunc vel. Vel elit scelerisque mauris pellentesque pulvinar pellentesque habitant. Neque viverra justo nec ultrices dui sapien eget. Vitae aliquet nec ullamcorper sit. Dui id ornare arcu odio ut sem nulla pharetra diam. Et magnis dis parturient montes. Varius morbi enim nunc faucibus.

Índice general

Índice de Figuras	8
Índice de Tablas	9
1. PLANTEAMIENTO DEL PROBLEMA	10
1.1. Descripción de la Realidad Problemática	10
1.2. Formulación del Problema	12
1.2.1. Problema General	12
1.2.2. Problemas Específicos	12
1.3. Objetivos de la Investigación	12
1.3.1. Objetivo General	12
1.3.2. Objetivos Específicos	13
1.4. Justificación de la Investigación	13
1.4.1. Teórica	13
1.4.2. Práctica	13
1.4.3. Metodológica	14
1.5. Delimitación del Estudio	14
1.5.1. Espacial	14
1.5.2. Temporal	14
1.5.3. Conceptual	14

1.6.	Hipótesis	15
1.6.1.	Hipótesis General	15
1.6.2.	Hipótesis Específicas	15
1.6.3.	Matriz de Consistencia	15
2.	MARCO TEÓRICO	16
2.1.	Antecedentes de la investigación	16
2.1.1.	DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation	16
2.1.2.	Sign Language Fingerspelling Recognition Using Depth Information and Deep Belief Networks	17
2.1.3.	Deep learning-based sign language recognition system for static signs .	18
2.1.4.	Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras	19
2.1.5.	Using Deep Learning in Sign Language Translation to Text	20
2.2.	Bases Teóricas	21
2.2.1.	Lenguaje de Señas:	21
2.2.2.	Accesibilidad:	22
2.2.3.	Tecnología Asistiva:	22
2.2.4.	Discapacidad Auditiva:	22
2.2.5.	Inclusión Educativa:	23
2.2.6.	Comunicación Inclusiva:	23
2.2.7.	Discapacidad Auditiva:	23
2.2.8.	Tecnología de Reconocimiento de Señas	24
2.2.9.	Derechos de las Personas con Discapacidad	24
2.2.10.	Cultura Sorda	24
2.3.	Marco Conceptual	25

2.3.1. Aprendizaje Profundo:	25
2.3.2. Redes Neuronales Artificiales:	25
2.3.3. Aprendizaje Supervisado:	25
2.3.4. Redes Neuronales Convolucionales (CNN):	26
2.3.5. Redes Neuronales Recurrentes (RNN):	26
2.3.6. Transferencia de Aprendizaje:	26
2.3.7. Optimización de Modelos:	27
2.3.8. Normalización y Estandarización:	27
2.3.9. Dropout:	27
2.3.10. Modelos Generativos:	28
2.3.11. Redes Neuronales de Procesamiento de Lenguaje Natural (NLP):	28
2.3.12. Aprendizaje No Supervisado y Semi-Supervisado:	28
2.3.13. Data Augmentation:	28
Anexos	30
A. Anexo I: Matriz de Consistencia	31
B. Anexo II: Resumen de Papers investigados	33

Índice de Figuras

- 1.1. % de personas con discapacidad. Fuente: **porcentaje de personas discapacidad** 11
- 1.2. Encuesta de personas con discapacidad . Fuente: **encuesta personas discapacidad** 11

Índice de Tablas

A.1. Matriz de consistencia. Fuente: Elaboración propia	32
B.1. Cuadro Resumen de Papers investigados. Fuente: Elaboración propia	34

Capítulo 1

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la Realidad Problemática

Las discapacidades del habla abarcan una amplia gama de condiciones que afectan la capacidad de una persona para comunicarse verbalmente de manera clara y fluida. Según la American Speech-Language-Hearing Association (ASHA), las discapacidades del habla pueden originarse por una variedad de razones, que van desde dificultades físicas en los órganos responsables de la producción del habla, hasta trastornos neurológicos que impactan la habilidad de hablar de manera clara y fluida.

Para las personas con discapacidades del habla, el lenguaje de señas se convierte en una herramienta invaluable que les permite expresar sus pensamientos, emociones y necesidades de manera efectiva. El lenguaje de señas es un sistema de comunicación visual y gestual utilizado por personas sordas o con discapacidades auditivas para comunicarse entre sí y con personas que pueden escuchar.

La Organización Mundial de la Salud afirmó que aproximadamente 70 millones de personas en el mundo son sordomudas. Un total de 360 millones de personas son sordas, y 32 millones de ellas son niños. Sin embargo, en Perú, de acuerdo con los resultados del Censo de Población y Vivienda 2017, como se puede observar en la Figura 1.1, hay un elevado porcentaje de personas que tienen dificultades para oír y para hablar o comunicarse.

El INEI también afirma que las personas presentan estas capacidades utilizan como apoyo para comunicarse su voz (19,8%), gesto y manos (11,9%) y lenguaje de señas (2,9%). Y debido a estas dificultades, estas personas se ven afectadas en el ámbito social y también laboral, por no poder expresarse debido a sus discapacidades. Según la Organización Mundial de Salud, las personas con estas discapacidades tienen más probabilidades de experimentar

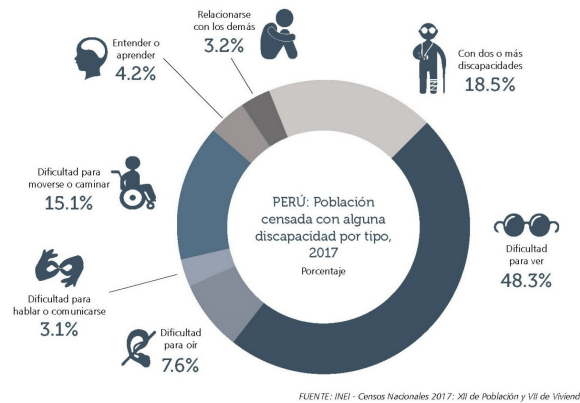


Figura 1.1: % de personas con discapacidad. Fuente: porcentaje de personas discapacidad

pobreza y exclusión social, y tienen menos probabilidades de tener un empleo remunerado que las personas sin discapacidades. En base a encuestas realizadas por la organización Incluyeme, alrededor del 72% de personas con discapacidad se encuentra desempleado, como se puede observar en la Figura 1.2.



Figura 1.2: Encuesta de personas con discapacidad . Fuente: encuesta personas discapacidad

El Deep Learning puede cambiar la forma de comunicación de las personas con discapacidades para comunicarse con personas sin ellas. Al entrenar modelos de Deep Learning con conjuntos de datos que contengan gestos de lenguaje de señas, es posible desarrollar modelos que puedan reconocer y traducir los gestos en tiempo real a texto. Estos modelos pueden capturar la complejidad de los movimientos de manos, dedos y expresiones faciales que son parte integral del lenguaje de señas. Además, al utilizar redes neuronales, estos modelos pueden aprender representaciones complejas de los gestos y expresiones, lo que les permite diferenciar dialectos de lenguaje de señas.

El objetivo principal de esta investigación es lograr un aumento significativo en la comunicación entre personas con discapacidades del habla y aquellas que no las tienen, a través de

la implementación de un modelo de traducción de lenguaje de señas utilizando Deep Learning. Este modelo busca facilitar una interacción más fluida y efectiva, permitiendo a las personas con discapacidades del habla expresar sus pensamientos, emociones y necesidades de manera más accesible y comprensible para quienes no conocen el lenguaje de señas.

1.2. Formulación del Problema

1.2.1. 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?

1.2.2. Problemas Específicos

- ¿De qué manera la falta de conjuntos de datos de lenguaje de señas de cada idioma afectar al modelo Deep Learning?
- ¿De qué manera el modelo Deep Learning pueden diferenciar entre los distintos tipos de lenguajes de señas?
- ¿Qué métricas son las más adecuadas para la precisión y rendimiento de un modelo de traducción de lenguaje de señas?
- ¿Cuáles son las técnicas más adecuadas para el preprocesamiento y normalización de la base de datos de lenguaje de señas?

1.3. Objetivos de la Investigación

1.3.1. Objetivo General

Desarrollar un modelo Deep Learning que se utilizará como medio para la traducción de lenguaje de señas, permitiendo la comunicación entre personas con discapacidades del habla y personas sin conocimiento del lenguaje de señas.

1.3.2. Objetivos Específicos

- Evaluar y comparar diferentes enfoques en los aumentos de datos para mejorar la representación de los conjuntos de datos de lenguaje de señas.
- Utilizar técnicas de aprendizaje automático para mejorar la precisión del modelo en la diferenciación entre los distintos tipos de lenguajes de señas.
- Evaluar diferentes métricas de evaluación de modelos Deep Learning, como Accuracy, Recall, F1-Score para la determinación del modelo más adecuado para la traducción adecuada de lenguaje de señas.
- Realizar comparaciones entre diferentes técnicas de preprocesamiento y normalización de datos de lenguaje de señas, como normalización de iluminación, corrección de gestos ambiguos.

1.4. Justificación de la Investigación

1.4.1. Teórica

Esta investigación se realiza para determinar si se puede utilizar un modelo de Deep Learning para poder mejorar la comunicación de las personas con discapacidades del habla con personas sin conocimiento de lenguaje de señas. Esta solución tiene el potencial de mejorar la calidad de vida de este grupo de personas facilitando su comunicación de manera más efectiva y fluida.

1.4.2. Práctica

Al culminar la investigación, las personas con discapacidades podrán utilizar el modelo de traducción de lenguaje de señas basado en Deep Learning para mejorar de manera significativa la comunicación con las personas sin conocimiento del lenguaje de señas. Este modelo tiene el potencial de ofrecer una solución efectiva y tecnológicamente avanzada para mejorar la accesibilidad y la calidad de vida de este grupo de personas.

1.4.3. Metodológica

. El desarrollo de un modelo de traducción de lenguaje de señas basado en Deep Learning ayudará a que se mejore la calidad de vida de las personas con discapacidades del habla, debido a que permitirá una comunicación más fácil con aquellos que no conocen el lenguaje de señas. Este modelo facilitará la interacción en diversos entornos, como en el trabajo, la educación y la vida diaria. Además, al ser una solución tecnológica, se espera que tenga un impacto positivo en la comunidad de personas con discapacidad del habla en general.

1.5. Delimitación del Estudio

1.5.1. Espacial

El estudio se realizará a nivel de la ciudad de Lima, enfocándose en videos de traductores de lenguaje de señas del Perú, que han brindado acceso a videos para el entrenamiento del modelo.

1.5.2. Temporal

El período de desarrollo de la investigación será de nueve meses comenzando en abril de 2024 con la recopilación de información para la definición de la problemática y temas con el mismo fin.

1.5.3. Conceptual

La presente investigación se centrará en el desarrollo de un modelo de traducción de lenguaje de señas basado en Deep Learning para el idioma de señas peruano, con el objetivo de mejorar la comunicación y la accesibilidad de las personas con discapacidades del habla. Se limitará a la recopilación y análisis de datos relacionados con el lenguaje de señas peruano y las técnicas de Deep Learning, excluyendo otros lenguajes de señas y enfoques de traducción de lenguaje de señas

1.6. Hipótesis

1.6.1. 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 del habla con personas que no conocen el lenguaje de señas, mejorando así su accesibilidad y calidad de vida.

1.6.2. Hipótesis Específicas

- Mediante el uso diferentes enfoques de técnicas de aumento de datos, sea posible mejorar la representación de los conjuntos de datos disponibles y compensar en parte la falta de datos específicos para el español peruano, lo que resultará en un mejor rendimiento del modelo de traducción de lenguaje de señas.
- El modelo Deep Learning aumentará su precisión significativa con lo que respecta de lenguaje de señas, lo que demuestra la eficacia de las técnicas de aprendizaje profundo en este contexto.
- La implementación de métricas de evaluación en los modelos Deep Learning aumentará las diferencias significativas entre los diferentes modelos evaluados, lo que permitirá la identificación del modelo más adecuado para la traducción de lenguaje de señas.
- La implementación de técnicas de preprocesamiento, mejoren la calidad de los datos de lenguaje de señas y con ello aumentar el rendimiento del modelo Deep Learning para la traducción de lenguaje de señas.

1.6.3. Matriz de Consistencia

A continuación se presenta la matriz de consistencia elaborada para la presente investigación (véase Anexo [A.1](#)).

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 [B.1](#)) de lo que se presenta en esta sección.

2.1.1. DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation

DeepASL es un modelo de traducción de lenguaje de señas basada en Deep Learning que permite la traducción de ASL (American Sign Language) tanto a nivel de palabra como de oración de manera no intrusiva, es decir este método no requiere que el usuario use algún equipo que cambie su comportamiento natural. Utiliza una luz infrarroja y un innovador sistema jerárquico bidireccional de redes neuronales recurrentes (HB-RNN) junto con un marco probabilístico basado en la Clasificación Temporal Conexionista (CTC).

2.1.1.1. 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 junto con gráficos relevantes:

1. **Recolección de datos:** Se recolectaron 7,306 muestras de 11 participantes, de las cuales 56 palabras y 100 oraciones comúnmente usadas en lenguaje de señas americano. También se recolectaron 1,178 muestras bajo diferentes condiciones de iluminación, posturas corporales e interferencia en la escena para evaluar la robustez del sistema.
2. **Captura de Señales:** Se utilizó el dispositivo Leap Motion, que emplea luz infrarroja para capturar de manera no intrusiva las señas de ASL. Leap Motion extrae la información de las articulaciones del esqueleto de los dedos, palmas y antebrazos.
3. **Extracción de Características** Se aprovecharon conocimientos del dominio de ASL para extraer las características clave de las señas, incluyendo la forma de la mano, el movimiento de la mano y la ubicación relativa de las dos manos.
4. **Modelado y Traducción:** Se empleó un sistema jerárquico bidireccional de redes neuronales recurrentes (HB-RNN) para modelar la estructura espacial y dinámica temporal de las características extraídas para la traducción a nivel de palabra. Para la traducción a nivel de oración, se adoptó un marco probabilístico basado en la Clasificación Temporal Conexionista (CTC).
5. **Evaluación del Rendimiento:** Se evaluó el rendimiento del sistema en términos de precisión de traducción, robustez bajo diferentes condiciones del mundo real y rendimiento del sistema (tiempo de ejecución, uso de memoria y consumo de energía). DeepASL se implementó en tres plataformas con diferente poder de cómputo: una computadora de escritorio, una plataforma móvil Nvidia Jetson TX1 y una tableta Microsoft Surface Pro 4.

2.1.1.2. Conclusiones

El modelo DeepASL demostró ser robusto en diversas condiciones de iluminación, posturas corporales y con diferentes fuentes de interferencia. El sistema mostró un rendimiento de tiempo de ejecución de 282ms en el peor de los casos y fue capaz de soportar un número suficiente de inferencias para su uso diario en plataformas móviles y de tabletas.

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

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.

2.1.2.1. Metodología

1. **Preprocesamiento de Datos:** Se capturan secuencias de imágenes que representan la dinámica de las señas. Las imágenes son normalizadas y redimensionadas para mantener la consistencia en los datos de entrada.
2. **Modelo Híbrido CNN-RNN:** La CNN se utiliza para extraer características espaciales de cada imagen en la secuencia. Una RNN, específicamente una LSTM, se emplea para capturar dependencias temporales entre las secuencias de imágenes. El modelo híbrido es entrenado con un conjunto de datos etiquetado, ajustando hiperparámetros para optimizar el rendimiento.
3. **Evaluación del Modelo:** En este modelo de traducción, para medir los resultados se enfocan en las métricas de precisión, sensibilidad, especificidad y F1-score. Se realizan pruebas para evaluar la robustez del modelo ante variaciones en las condiciones de iluminación y ángulos de captura.

2.1.2.2. Conclusiones:

La combinación de CNN y RNN mejora significativamente el reconocimiento de señas al capturar tanto características espaciales como temporales. Se identifican desafíos como la necesidad de más datos para entrenar adecuadamente el modelo y la importancia de la diversidad en el conjunto de datos. Futuras investigaciones pueden enfocarse en mejorar la eficiencia computacional del modelo y explorar otras arquitecturas híbridas.

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

El documento describe un estudio sobre el reconocimiento de letras del alfabeto manual utilizando información de profundidad y redes de creencias profundas (DBN). El objetivo principal es mejorar la precisión y eficiencia del reconocimiento de lenguaje de señas, específicamente el dactilológico.

2.1.3.1. Metodología:

1. **Adquisición de Datos:** Se utilizaron sensores de profundidad para capturar imágenes de las manos formando las letras del alfabeto manual. Los datos capturados incluyen tanto imágenes de profundidad como información adicional relevante para el reconocimiento de señas.
2. **Preprocesamiento de Datos:** Las imágenes de profundidad se preprocesaron para resaltar las características importantes y reducir el ruido. Se aplicaron técnicas de normalización y segmentación para preparar los datos para su uso en el modelo de aprendizaje profundo.
3. **Modelo de Aprendizaje:** Se utilizó una Red de Creencias Profundas (DBN) para el reconocimiento de letras. El modelo se entrenó utilizando los datos preprocesados, ajustando los parámetros para optimizar el rendimiento.
4. **Validación y Evaluación:** Se dividieron los datos en conjuntos de entrenamiento y prueba. Se utilizaron métricas de precisión, recall y F1-score para evaluar el rendimiento del modelo. Se realizaron pruebas cruzadas para asegurar la robustez del modelo.

2.1.3.2. Conclusiones:

El uso de información de profundidad en combinación con DBN es una técnica efectiva para el reconocimiento de lenguaje de señas. La metodología propuesta puede ser utilizada para desarrollar sistemas de traducción de lenguaje de señas más eficientes y precisos. Se identificaron áreas para futuras investigaciones, como la inclusión de más datos y la optimización de los modelos para diferentes lenguas de señas.

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

Este trabajo presenta un sistema para la traducción de lenguaje de señas en tiempo real utilizando cámaras de profundidad en dispositivos móviles. El sistema, denominado SUGO, utiliza un modelo de red neuronal convolucional 3D (3DCNN) basado en ResNet-18, adaptado para la extracción de características de secuencias de imágenes en video.

2.1.4.1. Metodología:

1. Arquitectura del modelo: Se utiliza el modelo ResNet-18 adaptado a 3DCNN, preentrenado con el dataset Kinetics-400 y luego reentrenado con un dataset propio.
2. Limpieza y cuantización: Para reducir el tamaño del modelo y su complejidad computacional, se aplican técnicas de poda de filtros y cuantificación de pesos, convirtiendo todos los parámetros de peso a unidades de punto flotante de menor precisión.
3. Aumento de base de datos: Se genera un dataset aumentado para hacer el modelo más robusto frente a ruidos de movimiento. Este dataset incluye datos que imitan los ruidos de movimiento que pueden introducirse en escenarios de uso real.
4. Segmentación de palabras: Se implementa un módulo de segmentación de palabras que utiliza una ventana deslizante para dividir secuencias de video en subconjuntos manejables para la clasificación de palabras individuales.

2.1.4.2. Conclusiones:

La poda de filtros y la cuantificación de pesos reducen el tamaño y la complejidad del modelo, haciéndolo adecuado para su operación en dispositivos móviles con recursos limitados. La adición de datos aumentados durante la fase de entrenamiento mejora la resiliencia del sistema frente a ruidos de movimiento, aunque introduce una ligera degradación (0.3%). El sistema demuestra una capacidad efectiva para clasificar gestos de lenguaje de señas en tiempo real en diversas condiciones de iluminación y movimiento.

2.1.5. Using Deep Learning in Sign Language Translation to Text

El documento revisa múltiples estudios sobre el reconocimiento de lenguaje de señas utilizando técnicas de aprendizaje profundo y captura de movimiento. Se enfoca en diferentes lenguajes de señas y diversas metodologías aplicadas, como redes neuronales convolucionales (CNN), redes neuronales de creencias profundas (DBN), y otras técnicas avanzadas de aprendizaje automático. El objetivo principal es analizar y comparar estas metodologías para entender sus fortalezas y limitaciones en el contexto del reconocimiento de lenguaje de señas.

2.1.5.1. Metodología:

1. Redes Neuronales Convolucionales (CNN): Estas redes son eficaces en la extracción de características espaciales de las imágenes, lo que las hace adecuadas para el reconocimiento de gestos y señas.
2. Redes de Creencias Profundas (DBN): Utilizadas para la extracción de características de datos de profundidad, lo que ayuda en la identificación precisa de las formas y movimientos de las manos.
3. Clasificación Temporal Conexista (CTC): Una técnica que permite el reconocimiento de secuencias sin necesidad de segmentación explícita de los datos de entrada.
4. Cámaras de Profundidad a Bordo: Utilizadas para capturar datos en 3D, proporcionando información detallada sobre la posición y el movimiento de las manos.

2.1.5.2. Conclusiones:

Los estudios revisados muestran una alta variabilidad en la precisión y efectividad de las técnicas utilizadas, dependiendo del lenguaje de señas y la complejidad de los gestos. Este estudio utilizó CNN y cámaras de profundidad y lograron una precisión superior al 90% en el reconocimiento de ciertos lenguajes de señas, como el American Sign Language (ASL). Las técnicas enfrentan dificultades cuando se trata de señas complejas o cuando el modelo debe generalizar a diferentes personas y contextos. La disponibilidad y calidad de los datos de entrenamiento son cruciales para el rendimiento de los modelos de aprendizaje profundo. Estudios con grandes conjuntos de datos de alta calidad obtuvieron mejores resultados.

2.2. Bases Teóricas

2.2.1. 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.

2.2.2. 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.

2.2.3. 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.

2.2.4. 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.

2.2.5. 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.

2.2.6. 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.

2.2.7. 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.

2.2.8. 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.

2.2.9. 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.

2.2.10. 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.

2.3. Marco Conceptual

2.3.1. Aprendizaje Profundo:

El aprendizaje profundo es una subdisciplina del aprendizaje automático (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.

2.3.2. Redes Neuronales Artificiales:

Las redes neuronales artificiales son modelos computacionales inspirados en la estructura y función del cerebro humano. Una ANN está formada por unidades básicas llamadas neuronas artificiales, organizadas en tres tipos de capas: la capa de entrada, las capas ocultas y la capa de salida. Cada neurona recibe una señal de entrada, la procesa mediante una función de activación y produce una señal de salida. Las conexiones entre neuronas tienen pesos asociados que se ajustan durante el proceso de entrenamiento para minimizar el error de predicción. Las ANN se utilizan en una amplia variedad de tareas, incluyendo clasificación, regresión y series temporales.

2.3.3. 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.

2.3.4. 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.

2.3.5. Redes Neuronales Recurrentes (RNN):

Las redes neuronales recurrentes (Recurrent Neural Networks, RNN) son una clase de redes neuronales diseñadas para procesar datos secuenciales. A diferencia de las redes feed-forward, 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.

2.3.6. Transferencia de Aprendizaje:

La transferencia de aprendizaje (Transfer Learning) es una técnica que aprovecha un modelo preentrenado en una tarea de origen para mejorar el rendimiento en una tarea de destino relacionada. Esta técnica es especialmente útil cuando se dispone de datos limitados para la tarea de destino, ya que permite reutilizar el conocimiento adquirido en la tarea de origen. En el contexto de redes neuronales, esto a menudo implica utilizar una red preentrenada en un gran conjunto de datos, como ImageNet, y ajustar sus capas finales o agregar nuevas capas para adaptarla a la nueva tarea. La transferencia de aprendizaje ha demostrado ser efectiva en una variedad de dominios, incluyendo visión por computadora y procesamiento de lenguaje natural.

2.3.7. 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.

2.3.8. 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.

2.3.9. Dropout:

El dropout es una técnica de regularización utilizada para prevenir el sobreajuste en las redes neuronales. Durante el entrenamiento, el dropout implica desactivar aleatoriamente un porcentaje de neuronas en cada capa en cada paso de entrenamiento, lo que impide que las neuronas desarrollen dependencias específicas entre sí. Esta técnica fuerza a la red a aprender representaciones más robustas y a generalizar mejor en datos no vistos. El dropout se suele aplicar en capas ocultas y se controla mediante una tasa de dropout, que define el porcentaje de neuronas que se desactivan en cada iteración.

2.3.10. Modelos Generativos:

Los modelos generativos son una clase de modelos de aprendizaje automático que intentan aprender la distribución conjunta de los datos de entrada y salida. A diferencia de los modelos discriminativos, que se centran en aprender la frontera de decisión entre diferentes clases, los modelos generativos pueden generar nuevas muestras que se asemejan a las muestras del conjunto de datos original. Ejemplos de modelos generativos incluyen las Redes Generativas Antagónicas (GAN), los Modelos de Boltzmann Restringidos (RBM) y las Redes de Creencias Profundas (DBN). Estos modelos tienen aplicaciones en generación de imágenes, síntesis de voz y modelado de lenguaje.

2.3.11. 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.

2.3.12. 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.

2.3.13. 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.

Anexos

Anexos A

Anexo I: Matriz de Consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS
Problema General	Objetivo General	Hipótesis 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?	Desarrollar un modelo Deep Learning que se utilizará como medio para la traducción de lenguaje de señas, permitiendo la comunicación entre personas con discapacidades del habla y personas sin conocimiento del lenguaje de señas.	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 del habla con personas que no conocen el lenguaje de señas, mejorando así su accesibilidad y calidad de vida.
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas
¿De qué manera la falta de conjuntos de datos de lenguaje de señas de cada idioma afectar al modelo Deep Learning?	Evaluar y comparar diferentes enfoques en los aumentos de datos para mejorar la representación de los conjuntos de datos de lenguaje de señas.	Mediante el uso diferentes enfoques de técnicas de aumento de datos, sea posible mejorar la representación de los conjuntos de datos disponibles y compensar en parte la falta de datos específicos para el español peruano, lo que resultará en un mejor rendimiento del modelo de traducción de lenguaje de señas.
¿De qué manera el modelo Deep Learning pueden diferenciar entre los distintos tipos de lenguajes de señas?	Utilizar técnicas de aprendizaje automático para mejorar la precisión del modelo en la diferenciación entre los distintos tipos de lenguajes de señas.	El modelo Deep Learning aumentará su precisión significativa con lo que respecta de lenguaje de señas, lo que demuestra la eficacia de las técnicas de aprendizaje profundo en este contexto.
¿Qué métricas son las más adecuadas para la precisión y rendimiento de un modelo de traducción de lenguaje de señas?	Evaluar diferentes métricas de evaluación de modelos Deep Learning, como Accuracy, Recall, F1-Score para la determinación del modelo más adecuado para la traducción adecuada de lenguaje de señas.	La implementación de métricas de evaluación en los modelos Deep Learning aumentará las diferencias significativas entre los diferentes modelos evaluados, lo que permitirá la identificación del modelo más adecuado para la traducción de lenguaje de señas.

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

Anexos B

Anexo II: Resumen de Papers investigados

Tipo	N°	Título	Autor	Año	País	Fuente
Problema	1	Copper price estimation using bat algorithm	Dehghani Bogdanovic	2018	United Kingdom	Resources Policy
	2	Alternative techniques for forecasting mineral commodity prices	Cortez, Saydam, Coulton, Sammut	2018	Netherlands	International Journal of Mining Science and Technology
Propuesta	3	Prediction of the crude oil price thanks to natural language processing applied to newspapers	Trastour, Genin, Morlot	2016	USA	Standfort University ML repository
	4	Stock Price Prediction Using Deep Learning	Tipirisetty	2018	USA	Master's Theses San Jose State University
	5	Deep Learning for Stock Prediction Using Numerical and Textual Information	Akita, R., Yoshihara, A., Matsubara, T., Uehara, K.	2016	USA	2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)
Técnica	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 (ICAIIIC)
	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