



PROYECTO DE APLICACIÓN RECONOCIMIENTO DE IMÁGENES LENGUAJE DE SEÑAS COLOMBIANO

J. M. Mejía, A. F. Robles, J. A. Thomas.

Proyecto de Aplicación / Universidad Jorge Tadeo Lozano Bogotá, Colombia

*e-mail: jesidm.mejiac@utadeo.edu.co, andresf.roblesg@utadeo.edu.co, juliea.thomasv@utadeo.edu.co.

RESUMEN: Las comunicaciones a lo largo del planeta son un método importante de transmitir una idea, pensamiento u opinión, en ocasiones cuando se hace referencia al lenguaje de señas, a pesar de que es un problema de gran complejidad ya que se debe hacer un análisis, estudio y reconocimiento de dichos gestos y sus diferentes formas que definen una letra o número. En el desarrollo de este artículo se busca identificar los gestos de las manos que realizan hombres y mujeres aplicando una técnica de machine learning que toma los datos iniciales y generar un algoritmo que, dada una imagen, sea capaz de reconocer el número o letra representada.

PALABRAS CLAVES: Entrenamiento, Probabilidad, Algoritmos de clasificación, Gestos, Red Neuronal, Python.

ABSTRACT: *Communications throughout the planet are an important method of transmitting an idea, thought or opinion, sometimes when reference is made to sign language, despite the fact that it is a highly complex problem since analysis, study and recognition of these gestures and their different forms that specify a letter or number. In the development of this article, we seek to identify the hand gestures performed by men and women by applying a machine learning technique that takes the initial data and generate an algorithm that, given an image, capable of recognizing the number or letter represented.*

KEYWORDS: *Training, Probability, Classification algorithms, Gestures, Neural Network, Python*

1. INTRODUCCIÓN

La población en Colombia con pérdida parcial o total de la audición esta alrededor de los 560.000 provocando en esta población un problema y dificultad de comunicación y es por esta razón que el uso de lenguaje mediante señas ha sido una manera sencilla de poder expresarse y se ha tomado como su lengua materna.

Cada país tiene uno o a veces dos o más lenguajes de señas, aunque diferentes lenguajes de señas pueden compartir las mismas raíces lingüísticas. El lenguaje de señas no es un lenguaje internacional, pero existen características universales en los lenguajes de señas. Esto hace posible que usuarios de distintos lenguajes de señas puedan entenderse entre sí de una manera más rápida que los usuarios de lenguajes hablados. Al igual que los lenguajes hablados, los lenguajes de señas tienen una estructura que puede ser dividida en segmentos más pequeños como frases, signos u otras unidades más pequeñas. Los lenguajes de señas son usados de acuerdo con ciertas reglas gramaticales.

Las personas que utilizan un lenguaje de señas enfrentan grandes obstáculos para comunicarse con las personas que utilizan lenguajes hablados ya que la mayoría de estas desconoce el significado de los signos utilizados en los lenguajes de señas [1].

Lo que se busca es generar un algoritmo de machine learning, que sea capaz de interpretar, aprender u obtener representaciones ideales de los datos sin intervención del usuario, mediante el uso de modelos apropiados y bases de datos con grandes cantidades para las diferentes señales del lenguaje de señas Colombiano y que de esta manera, se facilite a cada individuo con discapacidad auditiva comunicarse con las demás personas, para tener una conversación clara y entendible con las personas que no conocen este lenguaje.

2. MARCO TEORICO

LA LENGUA DE SEÑAS EN COLOMBIA

Poco se sabe sobre el origen de la lengua de señas en Colombia, ya que no existen suficientes testimonios escritos para precisar sus comienzos. Se dice que sus orígenes se remontan a 1920, en un internado católico bogotano. En 1957 aparece la primera asociación de sordos en Bogotá y un año después otra en Cali. Parece que estos sistemas de señas recibieron influencia de la lengua de señas española, a través de inmigrantes o de sordos colombianos educados en España, en los años 50. El hecho de que el Gobierno Nacional hubiera decretado y reconocido, según la Ley número 324 de 1996, en su artículo 2°, la lengua de señas colombiana como propia



de la comunidad sorda del país, no es más que una corroboración legal de lo que esta comunidad ya sabía y los investigadores que han trabajado el tema afirmaban hace muchos años: la lengua de señas es una lengua natural, con su propia gramática, sintaxis, vocabulario, usada por una comunidad específica.

La investigación sobre la lengua de señas en Colombia es casi reciente. Se inició aproximadamente hace 10 años cuando investigadores, universidades y asociaciones emprendieron algunos estudios básicos sobre aspectos para la enseñanza y el aprendizaje de la lengua de señas y otros sobre el funcionamiento lingüístico de esta. Poco a poco la descripción y estudio de la lengua de señas colombiana se han venido consolidando gracias a entidades como el Instituto Nacional para Sordos, INSOR [2]. En las figuras 1 y 2 se puede observar el alfabeto manual y números utilizados en Colombia para el lenguaje de señas.

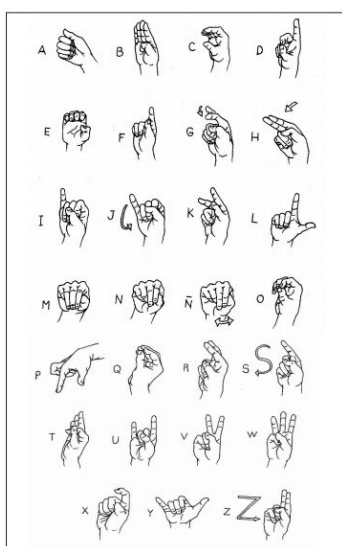


Fig. 1 Alfabeto Manual



Fig. 2 Números

APRENDIZAJE AUTOMÁTICO

El aprendizaje automático, también conocido como machine learning, es un campo de estudio de la inteligencia artificial que se encarga de otorgar a las computadoras la capacidad de aprender sin necesidad de ser programadas explícitamente.

Esta tecnología se centra en el desarrollo e implementación de algoritmos que sean capaces de aprender de los datos a los que son expuestos además de generar predicciones acerca de los mismos. Se utiliza principalmente para resolver tareas en las que la creación explícita de algoritmos específicos resulta difícil o imposible.

Algunas de las áreas en las que se ha utilizado esta tecnología son la clasificación de imágenes, detección de spam en correo electrónico, creación de vehículos autónomos, procesamiento de lenguaje natural, analizar y predecir movimientos de la bolsa de valores, detección de anomalías, etc.

APRENDIZAJE SUPERVISADO

Este tipo de algoritmos aprenden por medio de ejemplos, reciben un grupo de datos y el tipo de resultados que se espera que produzcan a partir de ellos, con esa información son capaces de aprender las reglas que deben ser aplicadas a las entradas para producir los resultados esperados

- **Clasificación.**

Consiste en identificar a que categoría pertenece un dato utilizando un algoritmo previamente entrenado con un conjunto de datos pertenecientes a categorías conocidas. La categoría que le será asignada al nuevo elemento pertenece al conjunto de categorías con las que fue entrenado previamente el algoritmo.

Un algoritmo capaz de realizar la clasificación de datos es conocido como clasificador. Existen diversos modelos matemáticos que permiten realizar la clasificación de datos, uno de los más comunes es conocido como regresión logística. [1].

RED NEURONAL ARTIFICIAL

Las RNA son sistemas de procesamiento de la información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas. Consisten en un gran número de elementos simples de procesamiento llamados nodos o neuronas que están organizados en capas. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. Los pesos representan la información que será usada por la red neuronal para resolver un problema determinado.

Así, las RNA son sistemas adaptativos que aprenden de la experiencia, esto es, aprenden a llevar a cabo ciertas



tareas mediante un entrenamiento con ejemplos ilustrativos.

Mediante este entrenamiento o aprendizaje, las RNA crean su propia representación interna del problema, por tal motivo se dice que son auto organizadas. Posteriormente, pueden responder adecuadamente cuando se les presentan situaciones a las que no habían sido expuestas anteriormente, es decir, las RNA son capaces de generalizar de casos anteriores a casos nuevos. [3].

3. TRABAJOS SIMILARES

3.1. Clasificación de Gestos de la Lengua de Señas Colombiana a partir del Análisis de Señales Electromiográficas utilizando Redes Neuronales Artificiales.

La metodología implementada se desarrolló mediante cuatro fases: 1) Adquisición de las señales EMG provenientes de los ocho sensores, registradas por la manilla Myo Armband, 2) extracción de características de las señales empleando la transformada Wavelet, 3) entrenamiento de la red neuronal y 4) validación del método de clasificación utilizando la técnica de validación cruzada; con el uso de la herramienta Matlab.

Las conclusiones del estudio fueron: 1) Se cuenta con un clasificador de redes neuronales artificiales a partir de señales electromiográficas que permite clasificar los 27 gestos del alfabeto de señas colombiano con una precisión del 88.39%. Este clasificador, con pruebas iniciales en ambiente real para la población con discapacidad oyente o hablante en Colombia, es la fase experimental para el desarrollo de una herramienta tecnológica a futuro. 2) El porcentaje de precisión del clasificador obtenido por rangos de edades o de manera global presenta un porcentaje muy similar a otras aproximaciones tecnológicas presentadas en la literatura, lo que permite contrastar la validez de la propuesta desarrollada; 3) La metodología propuesta puede ser escalada para clasificar adicional a los 27 gestos, el vocabulario del lenguaje de señas colombiano. 4) De acuerdo con el análisis de resultados, el clasificador presentó un porcentaje de precisión promedio cercano a los alcanzados por otros métodos mostrados en la literatura. [4].

3.1. Clasificación automática de las vocales en el lenguaje de señas colombiano

La base de datos para el reconocimiento de vocales en el lenguaje colombiano de señas fue obtenida con estudiantes del Instituto Tecnológico Metropolitano (ITM), donde a 151 personas adultas, entre los 18 y 50 años, se les indicó cómo realizar las cinco vocales en dicho lenguaje. Con un fondo uniforme se capturó el

objeto de interés con una cámara Web, de resolución 1024x768 píxeles.

Al conjunto de imágenes obtenidas en la base de datos se les realizó un preprocesamiento que constó de cuatro pasos. El primero de ellos fue aplicar un filtro gaussiano a la imagen original con el fin de suavizar y eliminar ruido en la imagen. El segundo fue una segmentación por color en el espacio de color Hue-Saturation-Value (HSV) que extrajo la totalidad de la mano al tener en cuenta los canales H y S, ya que son los que representan una variación importante para separar el color de la mano del fondo de la escena. El tercer paso realizado fue una operación de cierre para filtrar algunas regiones no pertenecientes al gesto, concluyendo el preprocesamiento con el último paso correspondiente a una umbralización por el método Otsu. Como resultado a esta etapa se obtuvo un único objeto dentro de la escena, correspondiente a la mano. El estudio fue desarrollado con el uso de Toolbox Balu de MATLAB

Las conclusiones del estudio fueron: Se realizó una comparación de clasificadores, orientado a la identificación de las vocales del lenguaje colombiano de señas, junto a una clase extra correspondiente a ninguna vocal. Se creó una base de datos de 151 objetos de prueba para cada clase; se probaron cuatro conjuntos de características; el primero correspondiente a la totalidad de características propuestas en la metodología, el segundo por medio de la transformación de características usando PCA, llegando a un total de 28 características. El tercero y cuarto por SFS-FISHER y SFS-KNN, respectivamente; con un total de 28 características para ambos conjuntos.

Los resultados experimentales mostraron un mayor desempeño en la clasificación de las vocales del lenguaje colombiano de señas, con las características del Conjunto 2: 28 Características por transformación de características usando PCA. El mejor clasificador para este conjunto en particular fue el QDA a 22 características, obteniendo una tasa de desempeño del 96,92%, lo que indica un excelente resultado para el problema planteado. [5].

4. METODOLOGIA.

La base de datos para el reconocimiento de vocales y números en el lenguaje colombiano de señas para hombres y mujeres fue obtenida con Kaggle. Al conjunto de imágenes obtenidas en la base de datos se les aplicó un modelo que consiste en una red neuronal convolucional.

Red Neuronal Convolucional.

Las redes neuronales convolucionales (CNN) son herramientas ampliamente utilizadas para el aprendizaje profundo. Son específicamente adecuados para imágenes como entradas, aunque también se usan para otras aplicaciones como texto, señales y otras respuestas

continuas. Se puede entender como un conjunto de capas de procesamiento, de modo que puede verse como un diagrama secuencial de bloques. [6].

Cada capa de la CNN es un bloque con tres principales variables: entrada, pesos y salida. en estas capas se presenta una característica fundamental como donde la salida de una capa se convierte en la entrada de la próxima. Este proceso es secuencial y puede ser no lineal, en cada capa se realiza una función específica. Las CNN tiene como principal tarea extraer patrones, en este caso sobre imágenes, debido a ello utilizan un entrenamiento no supervisado [3].

5. ARQUITECTURA

Para la creación de la red neuronal convolucional, se utilizaron las librerías Keras, Numpi, Pandas, Matplotlib.pyplot, Tensorflow los cuales permite trabajar bajo el lenguaje de Python

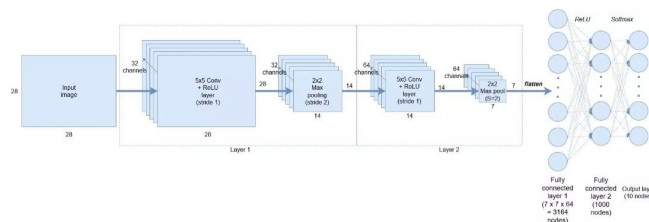


Fig. 3 Red neuronal convolucional que se utilizó como guía para la elaboración de la arquitectura [7]. Imagen tomada de <https://adventuresinmachinelearning.com/keras-tutorial-cnn-11-lines/>

Como se observa en la figura 3, se presenta un modelo guía de la arquitectura implementada en el desarrollo del presente trabajo, con la diferencia que se utilizaron valores en la primera capa en vez de 35 canales, se utilizaron 100 y en la segunda capa, en vez de 64 se utilizaron 500.

Para la etapa de preprocesamiento se implementó un código el cual se encarga de realizar el procesamiento de imágenes y la adaptación de los datos para la red neuronal. La rutina *preprocess_image* es la encargada del escalamiento y normalización de las imágenes. Adicionalmente, con cada imagen preprocesada, se actualizará la lista *img_sign* de etiquetas que almacenará el signo que representa cada señal.

A continuación, en la figura 4 se observa un resumen de la arquitectura implementada.

Model: "sequential_1"		
Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 96, 96, 255)	6630
max_pooling2d_1 (MaxPooling2D)	(None, 48, 48, 255)	0
conv2d_2 (Conv2D)	(None, 44, 44, 255)	1625880
max_pooling2d_2 (MaxPooling2D)	(None, 22, 22, 255)	0
flatten_1 (Flatten)	(None, 123420)	0
dense_1 (Dense)	(None, 1000)	123421000
dense_2 (Dense)	(None, 11)	11011
Total params: 125,064,521		
Trainable params: 125,064,521		
Non-trainable params: 0		

Fig. 4 Resumen de la arquitectura

6. RESULTADOS

Una vez implementada la arquitectura y desarrollado el código Python, se procede con la evaluación del modelo para determinar su precisión por lo que se obtuvo un valor del 91.39% figura 5.

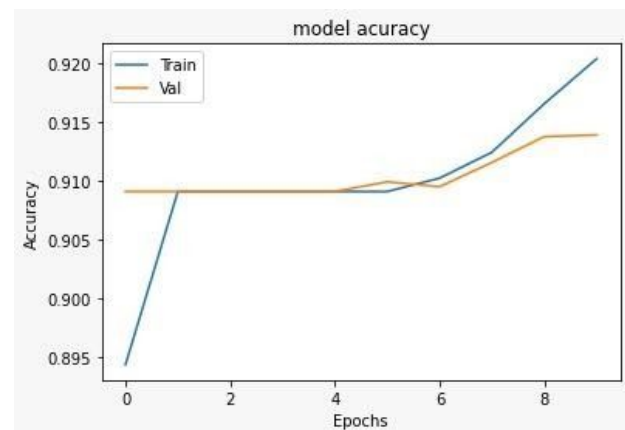


Fig. 5 Precisión del modelo

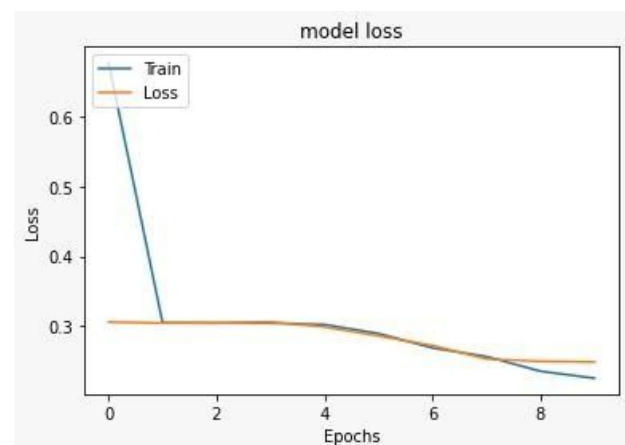


Fig. 6 Grafica perdida de datos

7. CONCLUSIONES



Siendo las redes neuronales una herramienta de gran ayuda en especial en este tipo de casos de reconocimiento de patrones, se pudo presentar un modelo con una precisión considerable para la identificación de imágenes del lenguaje de señas.

El proyecto modelado se creo en principio para el reconocimiento de vocales y números, pero tiene gran forma de expansión en el reconocimiento de los diferentes gestos del lenguaje de señas y a futuro brindar una gran ayuda a personas con discapacidad auditiva para poder comunicarse con los demás.

Entrenar este tipo de modelos tiene un alto de grado de complejidad, recursos de hardware y tarda un tiempo

considerable, pero el reto que se puede concluir de este proyecto se encuentra en el poder computacional para la identificación y reconocimiento de imágenes.

8. BIBLIOGRAFÍA

- [1] Carlos Alberto Monares Zabaleta, “Interpretación del lenguaje de señas utilizando redes neuronales”, UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, Facultad de Ingeniería, México, 2017.
- [2] INSTITUTO NACIONAL PARA SORDOS, INSOR, “DICCIONARIO BÁSICO DE LA LENGUA DE SEÑAS COLOMBIANA”, 2006.
- [3] A. Cifuentes, E. Mendoza, M. Lizcano, A. Santrich, S. Moreno, “Desarrollo de una red neuronal convolucional para reconocer patrones en imágenes”, Universidad Simón Bolívar, Barranquilla-Colombia, 2019. [Online]. Available: <http://revistas.unisimon.edu.co/index.php/identific/article/view/4007>
- [4] E. Galvis-Serrano, I. Sánchez-Galvis, N. Flórez, S. Zabala-Vargas, “Clasificación de Gestos de la Lengua de Señas Colombiana a partir del Análisis de Señales Electromiográficas utilizando Redes Neuronales Artificiales”, Facultad de Ingeniería de Telecomunicaciones, Universidad Santo Tomás, Bucaramanga-Colombia, 2019, [Online]. Available: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-07642019000200171
- [5] D. Botina-Monsalve, M. Domínguez-Vásquez, C. Madrigal-González, A. Castro-Ospina, “Clasificación automática de las vocales en el lenguaje de señas colombiano”, Instituto Tecnológico Metropolitano, Colombia, 2017, [Online]. Available: <https://www.redalyc.org/jatsRepo/3442/344255038006/html/index.html>
- [6] Learn About Convolutional Neural Networks, [Online]. Available: <https://www.mathworks.com/help/deeplearning/ug/introduction-to-convolutional-neural-networks.html>
- [7] Classify Images Using Convolutional Neural Networks & Python, [Online]. Available: <https://medium.com/@randerson112358/classify-images-using-convolutional-neural-networks-python-a89cecc8c679>
- [8] Lic. Franco Ronchetti, “Reconocimiento de gestos dinámicos y su aplicación al lenguaje de señas”, UNIVERSIDAD NACIONAL DE LA PLATA, Facultad de Informática, Argentina, diciembre 2016.