



UNIVERSITAT RAMON LLULL

**Escola Tècnica Superior d'Enginyeria La Salle**

Treball Final de Grau

Grau en Enginyeria Informàtica

Procesamiento del lenguaje de signos y  
su conversión a texto

Alumne

Mar Galiana Fernández

Professor Ponent

Elisabet Golobardes Ribé  
i  
Núria Valls Canudas

---

## **ACTA DE L'EXAMEN DEL TREBALL FI DE CARRERA**

---

Reunit el Tribunal qualificador en el dia de la data, l'alumne

Mar Galiana Fenández

va exposar el seu Treball de Fi de Carrera, el qual va tractar sobre el tema següent:

Investigació d'un model predictiu pel reconeixement del llenguatge de signes.

Acabada l'exposició i contestades per part de l'alumne les objeccions formulades pels Srs. membres del tribunal, aquest valorà l'esmentat Treball amb la qualificació de

Barcelona,

VOCAL DEL TRIBUNAL

VOCAL DEL TRIBUNAL

PRESIDENT DEL TRIBUNAL

UNIVERSIDAD: RAMON LLULL

GRUPO DE INVESTIGACIÓN: DATA SCIENCE FOR THE DIGITAL SOCIETY

# PROCESAMIENTO DEL LENGUAJE DE SIGNOS Y SU CONVERSIÓN A TEXTO

---

Trabajo Final de Grado

Tutoras:

Núria Valls Canudas y Elisabet Golobardes Ribé

Autora:

Mar Galiana Fernández

## **En castellano**

En este proyecto se ha querido implementar un modelo predictivo que utilízase imágenes como datos de entrada. Estas imágenes hacen referencia a treinta y ocho gestos estáticos del lenguaje de signos americano (ASL). Los gestos pertenecen al abecedario inglés y a los números del cero al diez.

Los modelos predictivos utilizados han sido árboles de decisiones y redes neuronales. Con el árbol de decisión se utilizó el algoritmo Extreme Gradient Boosting y se obtuvo un rendimiento del 100 %. Finalmente fue descartado debido a que, al aumentar el número de muestras del dataset, su coste temporal se incrementó demasiado para poder ser utilizado bajo las condiciones de ejecución del proyecto.

A lo que hace referencia a las redes neuronales, se implementaron dos tipos: una artificial y otra convolucional. Se utilizaron dos tipos de clasificador: multi-clase y binario, utilizando la estrategia One vs All.

Los mejores resultados se obtuvieron con la red neuronal convolucional multi-clase, consiguiendo un rendimiento del 99,52 % con un tiempo total de ejecución de 1 hora, 55 minutos y 36 segundos, y un coste de memoria de 12,88 GB.

## En inglés

In this project we wanted to implement a predictive model that uses images as an input data. These images refer to thirty-eight static gestures, belonging to the American Sign Language (ASL). The gestures represent the English alphabet and the numbers from zero to ten.

There have been two predictive models used: decision trees and neural networks. The Extreme Gradient Boosting algorithm was used with the decision tree, obtaining an accuracy of 100 %. It was eventually discarded because, as the number of samples in the dataset increased, its time cost increased too much to be used under this project execution conditions.

As far as neural networks are concerned, two types were implemented: one artificial and one convolutional. Two types of classifier were used: multi-class and binary, using the One vs All strategy.

The best results were obtained with the multi-class convolutional neural network, achieving an accuracy of 99.52 % with a total execution time of 1 hour, 55 minutes and 36 seconds, and a memory cost of 12,88 GB.

## **En catalán**

En aquest projecte s'ha volgut implementar un model predictiu que utilitzés imatges com a dades d'entrada. Aquestes imatges fan referència a trenta-vuit gestos estàtics del llenguatge de signes americà (ASL). Els gestos pertanyen a l'abecedari anglès i als valors numèrics del zero al deu.

Els models predictius utilitzats han estat arbres de decisions i xarxes neuronals. Amb l'arbre de decisió s'ha fet servir l'algoritme Extreme Gradient Boosting, obtenint un rendiment del 100 %. Finalment va ser descartat, ja que, en augmentar el nombre de mostres del dataset, el seu cost temporal es va incrementar massa per a poder ser utilitzat sota les condicions d'execució del projecte.

Al que fa referència a les xarxes neuronals, es van implementar dos tipus: una artificial i una altra convolucional. Es van utilitzar dos tipus de classificador: multiclasse i binari, utilitzant l'estrategia One vs All.

Els millors resultats es van obtenir amb la xarxa neuronal convolucional multiclasse, aconseguint un rendiment del 99,52 % amb un temps total d'execució d'1 hora, 55 minuts i 36 segons, i un cost de memòria de 12,88 GB.

## **Palabras claves**

Reconocimiento de lenguaje de signos (SLR), lenguaje de signos americano (ASL), red neuronal artifical, red neuronal convolucionarial, árboles de decisión, XGBoost, clasificador mulit-clase, clasificado One vs All, clasificador One vs Rest.

# Índice

<b>1. Introducción</b>	<b>13</b>
<b>2. Estado del arte</b>	<b>14</b>
2.1. El lenguaje de signos . . . . .	14
2.2. Aproximaciones actuales . . . . .	16
2.2.1. Datos . . . . .	17
2.2.2. Técnicas . . . . .	22
2.2.3. Resultados . . . . .	26
<b>3. Fundamentos</b>	<b>28</b>
3.1. Tipos de tareas de clasificadores . . . . .	28
3.1.1. Clasificación binaria . . . . .	28
3.1.2. Clasificación multi-clase . . . . .	29
3.1.3. Clasificación multi-etiqueta . . . . .	30
3.1.4. Clasificación desequilibrada . . . . .	31
3.2. Métodos de clasificación . . . . .	32
3.2.1. Redes Neuronales . . . . .	32
3.2.2. Árboles de decisión . . . . .	38
<b>4. Condiciones de ejecución del proyecto</b>	<b>42</b>
<b>5. Experimentación</b>	<b>44</b>
5.1. Experimentación 1: Búsqueda del dataset . . . . .	44
5.2. Experimentación 2: Uso de redes neuronales . . . . .	46
5.2.1. Red neuronal artificial . . . . .	46
5.2.2. Red neuronal convolucional . . . . .	48
5.2.3. Resultados de las redes neuronales . . . . .	51
5.3. Experimentación 3: Uso de árboles de decisión . . . . .	57
5.4. Comparativa de las experimentaciones 2 y 3 . . . . .	61
5.5. Experimentación 4: Incremento dataset . . . . .	63
5.6. Experimentación 5: Mejorar rendimiento de la CNN . . . . .	69
5.6.1. Tamaño del <i>Batch</i> y número de <i>Epochs</i> . . . . .	70
5.6.2. Algoritmos de optimización . . . . .	72
5.6.3. Índice de aprendizaje e impulso . . . . .	73
5.6.4. Inicialización del peso de la red . . . . .	75
5.6.5. Función de activación neuronal . . . . .	77
5.6.6. Regulación del <i>Dropout</i> . . . . .	78
5.6.7. Número de neuronas . . . . .	81
5.6.8. Resultados de la optimización . . . . .	82
5.7. Comparativa de las experimentaciones 4 y 5 . . . . .	87
5.8. Experimentación 6: Uso de clasificadores binarios . . . . .	89
<b>6. Propuesta</b>	<b>95</b>

<b>7. Metodología utilizada</b>	<b>98</b>
<b>8. Estructura del proyecto</b>	<b>100</b>
8.0.1. Definición de las estrategias . . . . .	100
8.0.2. Secuencia de ejecución de las estrategias . . . . .	102
8.0.3. Diagrama de clases UML . . . . .	105
<b>9. Costes del proyecto</b>	<b>108</b>
<b>10. Conclusiones y líneas de futuro</b>	<b>109</b>
10.1. Conclusiones . . . . .	109
10.2. Líneas de futuro . . . . .	111
<b>A. Anexo</b>	<b>112</b>
A.1. Proyecto Github . . . . .	112
A.2. Diagrama de clases UML . . . . .	113
A.2.1. Ampliación de las estructuras . . . . .	114
A.2.2. Ampliación de las estrategias . . . . .	115
A.2.3. Ampliación del modelo . . . . .	116
A.2.4. Ampliación de las excepciones, almacenamiento y registrador . . . . .	117

## Índice de figuras

1.	Alfabeto del lenguaje de signos americano (ASL). Todos los signos son estáticos, excepto la 'J' y la 'Z'. . . . .	15
2.	Números del lenguaje de signos americano (ASL), . . . . .	15
3.	Estadísticas del dataset alemán utilizado en el artículo <i>An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences</i> [1]. . . . .	18
4.	Escenario para la recopilación de datos en el artículo <i>Fingerspelling recognition with semi-Markov conditional random fields</i> [2] .	19
5.	Posición de la cámara en dos obtenciones de muestras. La primera con una posición frontal y la segunda con una visión satélite con la cámara situada en el sombrero. Obtenido del estudio <i>Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video</i> [3]. . . . .	20
6.	Ejemplo del uso de un sistema de posicionamiento magnético, obtenido del estudio <i>Gesture Recognition of Sign Language Alphabet Using a Magnetic Positioning System</i> [4] . . . . .	21
7.	Diagrama de los cuatro estados de uno de los modelos oculto de Márkov implementado en el estudio <i>Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video</i> [3]. . . . .	22
8.	Estructura de la implementación del modelo predictivo con RNN en el estudio <i>Phonologically-meaningful Subunits for Deep Learning-based Sign Language Recognition</i> [5]. . . . .	23
9.	Comparación de las estrategias de ajuste de múltiples modelos de clasificación binaria con datos de diversas clases. . . . .	30
10.	Funcionamiento de la obtención del resultado de una red neuronal. .	33
11.	Funcionamiento de la obtención del resultado de una red neuronal. .	33
12.	Estructura de una red neuronal. La imagen ha sido obtenida del artículo <i>Are artificial neural networks black boxes?</i> [6] . . . . .	34
13.	Modelo de una red neuronal artificial . . . . .	36
14.	Modelo de una red neuronal convolucional . . . . .	37
15.	Modelo de una red neuronal recurrente . . . . .	38
16.	Estructura y partes de un árbol de decisión. . . . .	39
17.	Estructura de un árbol de decisión basada en las particiones. Esta imagen ha sido obtenido del artículo <i>Classification and Regression Trees</i> [7]. . . . .	40
18.	Comparativa de diferentes algoritmos de árboles de decisión. Esta tabla ha sido obtenido del artículo <i>A Survey on Decision Tree Algorithms of Classification in Data Mining</i> [8]. . . . .	41
19.	Imágenes obtenidas del dataset <i>Sign Language Gesture Images Dataset</i> [9] en color. . . . .	44

20.	Imágenes obtenidas del dataset <i>Sign Language Gesture Images Dataset</i> [9] preprocesadas con la intención de mejorar el entrenamiento de las redes neuronales. . . . .	45
21.	Estructura de la red neuronal artificial implementada. . . . .	46
22.	Gráfica de la función de activación ReLu ( <i>Rectified Linear Activation Function</i> ). . . . .	47
23.	Comportamiento de la función de activación <i>softmax</i> . . . . .	47
24.	Resultado de la ejecución de la función <i>summary</i> con la red neuronal artificial. . . . .	48
25.	Estructura de la red neuronal convolucional implementada. . . .	49
26.	Explicación del parámetro kernel en la capa convolucional de una red neuronal. . . . .	49
27.	Explicación del parámetro stride en la capa convolucional de una red neuronal. . . . .	50
28.	Resultado de la ejecución de la función <i>summary</i> en la red neuronal convolucional. . . . .	51
29.	Uso de la librería xgboost en la implementación del boosted decision tree. . . . .	57
30.	Imágenes obtenidas del dataset <i>Sign Language Gesture Images Dataset</i> [9] en color. . . . .	64
31.	Imágenes obtenidas del dataset <i>Sign Language Gesture Images Dataset</i> [9] preprocesadas con la intención de mejorar el entrenamiento de las redes neuronales. . . . .	64
32.	Imágenes obtenidas del dataset <i>ASL test and train</i> [10]. . . . .	65
33.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>batch and epoch</i> .	70
34.	Resultado de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>batch and epoch</i> . . . . .	71
35.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>optimization algorithms</i> . . . . .	72
36.	Resultados obtenidos en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>optimization algorithms</i> . . . . .	73
37.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>learn rate and momentum</i> . . . . .	74
38.	Resultados de la ejecución de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>learn rate and momentum</i> . . . . .	75
39.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>network weight initialization</i> . . . . .	76
40.	Resultados de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>network weight initialization</i> . .	77

41.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>neuron activation function</i> . . . . .	78
42.	Resultado de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>neuron activation function</i> . . . . .	78
43.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>dropout regularization</i> . . . . .	79
44.	Resultados de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>dropout regularization</i> . . . . .	80
45.	Parámetros evaluados en la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>number neurons</i> . . . . .	81
46.	Resultados de la ejecución de la estrategia <i>Hyperparameter Optimization Strategy</i> con el parámetro <i>number neurons</i> . . . . .	81
47.	Modelo de la red neuronal convolucional obtenida a partir de los resultados de las ejecuciones de la estrategia <i>Hyperparameter Optimization Strategy</i> . . . . .	82
48.	Parámetros requeridos en el constructor de la clase <i>Conv2D</i> de la librería <i>keras.layers</i> . . . . .	83
49.	Archivo del proyecto que contiene los valores de los atributos optimizados obtenidos a partir de las ejecuciones de la estrategia <i>Hyperparameter Optimization Strategy</i> . . . . .	84
50.	Diagrama de flujo de la estrategia <i>Train Binary Neural Network</i>	90
51.	Diagrama de flujo de la estrategia <i>Accuracy Binary Neural Network</i>	91
52.	Diagrama de flujo del entrenamiento de una red neuronal binaria con una estrategia One vs All. . . . .	92
53.	<i>Output</i> de la ejecución de la estrategia <i>Accuracy Binary Neural Network</i> con las señales: 'A', 'B' y 'C'. . . . .	93
54.	<i>Output</i> de la ejecución de la estrategia <i>Accuracy Binary Neural Network</i> con las señales numéricas. . . . .	94
55.	Estructura de la red neuronal propuesta. . . . .	95
56.	Diagrama de flujo de los pasos a seguir para poder utilizar el clasificador propuesto. . . . .	96
57.	Explicación de la metodología espiral. . . . .	98
58.	Secuencia que el usuario debe seguir para poder ejecutar la estrategia <i>Setup</i> del proyecto. . . . .	102
59.	Secuencia que el usuario debe seguir para poder ejecutar la estrategia <i>Help</i> del proyecto. . . . .	103
60.	Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en un árbol de decisión. . . . .	103
61.	Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en una red neuronal, ya sea convolucional o artificial, con datos multi-clase. . . . .	104
62.	Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en una red neuronal convolucional binaria con una estrategia <i>One vs All</i> . . . . .	104

63.	Secuencia que el usuario debe seguir para poder ejecutar la estrategia <i>ShowOptimizedHyperparameter</i> del proyecto. . . . .	105
64.	Tiempo, en horas, invertido en el proyecto. Teniendo en cuenta: la definición del proyecto, el aprendizaje de los fundamentos, la implementación y la documentación. . . . .	108
65.	Estructura del proyecto del repositorio de <i>Github</i> donde se encuentra el proyecto realizado. . . . .	112
66.	Diagrama de clases UML siguiendo una perspectiva conceptual donde se visualiza la estructura del proyecto. . . . .	113
67.	Ampliación del paquete <i>Structures</i> del diagrama de clases mostrado en la imagen 66. . . . .	114
68.	Ampliación del paquete <i>StrategyFactory</i> del diagrama de clases mostrado en la imagen 66. . . . .	115
69.	Ampliación del paquete <i>Model</i> del diagrama de clases mostrado en la imagen 66. . . . .	116
70.	Ampliación de los paquetes <i>Storage</i> , <i>Exception</i> y <i>Logger</i> del diagrama de clases mostrado en la imagen 66. . . . .	117

## Índice de tablas

1.	Comparativa de los resultados obtenidos en diez proyectos dependiendo de las técnicas y los tipos de datos utilizados. . . . .	26
2.	Comparativa del tamaño del dataset cuando se guardan las imágenes en un directorio, o bien cuando se utiliza un pickle sin comprimir. . . . .	52
3.	Comparativa del tamaño del dataset cuando se guardan las imágenes en un directorio, o bien cuando se utiliza un pickle comprimido. . . . .	52
4.	Comparativa del tiempo de ejecución de la estrategia SaveDatabase utilizando dos modelos con diferente procesado cuando los archivos generados no se comprimen. . . . .	53
5.	Comparativa del tiempo de ejecución de la estrategia SaveDatabase utilizando dos modelos con diferente procesado cuando los archivos generados son comprimidos. . . . .	53
6.	Tamaño del archivo <i>h5</i> por los dos tipos de redes neuronales con los dos datasets que se están utilizando actualmente. . . . .	54
7.	Tiempo de ejecución de la estrategia de <i>TrainingNeuralNetwork</i> dependiendo del dataset y del tipo de red neuronal. . . . .	54
8.	Comparativa del <i>accuracy</i> de los dos tipos de redes neuronales dado dos modelos con diferente procesado. . . . .	54
9.	Tiempo de ejecución de la estrategia de <i>AccuracyNeuralNetwork</i> dependiendo del dataset y del tipo de red neuronal. . . . .	55
10.	Comparativa del coste de memoria en las tres estrategias entre los dos tipos de redes neuronales, dado dos modelos con diferente procesado. . . . .	55
11.	Comparativa del coste de tiempo en las tres estrategias entre los dos tipos de redes neuronales, dado dos modelos con diferente procesado. . . . .	56
12.	Comparativa del coste de memoria en la estrategia <i>TrainDecisionTree</i> , dado dos modelos con diferente procesado. . . . .	58
13.	Comparativa del coste de tiempo en la estrategia <i>TrainDecisionTree</i> , dado dos modelos con diferente procesado. . . . .	58
14.	Comparativa de la <i>accuracy</i> del árbol de decisión, dado dos modelos con diferente procesado. . . . .	59
15.	Comparativa del coste de tiempo en la estrategia <i>AccuracyDecisionTree</i> , dado dos modelos con diferente procesado. . . . .	59
16.	Comparativa del coste de tiempo en el uso del árbol de decisión como modelo de predicción dado dos modelos con diferente procesado. . . . .	60
17.	Comparativa de la <i>accuracy</i> entre las estructuras implementadas dado dos modelos con diferente procesado. . . . .	61
18.	Comparativa de la duración del entrenamiento entre las estructuras implementadas dado dos modelos con diferente procesado. . .	61

19.	Comparativa de la duración de la predicción entre las estructuras implementadas dado dos modelos con diferente procesado. . . . .	62
20.	Comparativa de la <i>accuracy</i> de las tres estructuras implementadas dado dos datasets y una combinación de ambos. . . . .	66
21.	Comparativa del tiempo de entrenamiento de las tres estructuras implementadas dado dos datasets y una combinación de ambos. . . . .	67
22.	Comparativa del tiempo de predicción de las tres estructuras implementadas dado dos datasets y una combinación de ambos. . . .	67
23.	Comparativa de la red neuronal convolucional sin <i>Dropout</i> o con un <i>Dropout</i> con un índice igual a cero. . . . .	84
24.	Comparativa de la red neuronal convolucional con los parámetros optimizados con y sin <i>Dropout</i> . . . . .	85
25.	Comparativa de los resultados de las redes neuronales convolucionales de los apartados 5.5 y 5.6. . . . .	88
26.	Características de la ejecución del modelo predictivo con redes neuronales binarias con las muestras de los signos 'A', 'B' y 'C'. .	93
27.	El coste de tiempo de todas las estrategias utilizadas para obtener las predicciones del modelo propuesto . . . . .	96
28.	El coste de memoria de todas las estrategias utilizadas para obtener las predicciones del modelo propuesto . . . . .	97

## 1. Introducción

Hay un pequeño porcentaje de la población que presenta un trastorno en el habla y/o una discapacidad auditiva, impidiéndoles comunicarse en un lenguaje verbal. Este porcentaje es de alrededor de un 5% [5]. En el 2021 había un total de 7.752.840.550 habitantes en el mundo, por lo tanto, se estima que unas 3.876 millones de personas presentaban un problema en el habla o en la escucha.

Este elevado número de personas se comunican a través del lenguaje de signos, el cual no suele ser conocido por las personas que utilizan el lenguaje verbal. Esto dificulta su incorporación, debido a que no se pueden hacer entender.

Con el grupo de investigación *Data Science for the Digital Society*, se ha de desarrollado un proyecto que tiene como objetivo ayudar a facilitar el entendimiento entre una comunicación gestual y una verbal. El método que se utiliza es implementando un modelo predictivo que, a partir de imágenes de gestos del lenguaje de signos, pueda deducir cual es su texto correspondiente.

Hay diversos lenguajes de signos, el que se ha utilizado en este proyecto es el *American Sign Language* (ASL). El motivo de esta elección es que, de este, se dispone de mucha información y es más fácil encontrar datasets ya preparados para ser utilizados en un modelo predictivo.

Las muestras utilizadas disponen de treinta y ocho gestos, las veintisiete letras del abecedario inglés y los números del cero al once. Se han seleccionado estos signos ya que son gestos estáticos, no se requiere de un movimiento para interpretarlos. Al utilizar las imágenes sin un contexto temporal como los datos del modelo, no se podrían predecir gestos dinámicos.

Para poder encontrar el mejor modelo predictivo se han hecho varias experimentaciones, se han probado con clasificadores basados en redes neuronales y en árboles de decisión. Una vez obtenidos los primeros resultados, se intentó mejorar su rendimiento de diversas formas: variando el dataset, realizando transformaciones a las muestras y modificando las estructuras de los modelos.

Las comparaciones entre los modelos predictivos se han realizando evaluando su rendimiento, el coste de tiempo y el de memoria. A partir de estos tres parámetros se han ido descartando los modelos que tenían un rendimiento demasiado leve o unos costes demasiado elevados. Obteniendo así los modelos que mejor se adaptan a estas características.

## **2. Estado del arte**

El análisis del estado del arte, que se realizará en este proyecto, hace referencia a las tecnologías y ayudas que disponen las personas sordomudas a la hora de comunicarse con una persona que no sepa el lenguaje de signos.

### **2.1. El lenguaje de signos**

La comunicación cuando uno de los hablantes presenta una discapacidad auditiva y/o un trastorno en el habla se dificulta. Esto no es solo causado por el echo de que la mayoría de la población no tiene los conocimientos necesarios para poder comunicarse en un lenguaje no verbal; sino por el echo de que existen unos 140 lenguajes de signos, con sus respectivos alfabetos y gesto, según la organización Ethnologue [11].

Se estima que un 5 % de la población global tienen dificultades en el habla o de escucha [5]. De ese fragmento de la población, entre un 5 y un 10 % nacen con padres que presentan la misma discapacidad o trastorno, dejando entre un 90 y un 95 % de estos recién nacidos rodeado de un lenguaje natural visual-gestual en el cual no acaban de poder adaptarse [12]. A demás, a la mayoría de las personas que nacieron sin la capacidad de oír se les dificulta mucho la lectura [13].

Todos los lenguajes de signos emplean varias articulaciones para poder comunicarse, como son: las manos, la parte superior del cuerpo, la cabeza y la cara para expresar rasgos gramaticales simultáneamente [14].

Hay diferentes tipos de gestos dependiendo del tipo de información que se pretenda transmitir. Basado en las investigaciones de Kendon [15] y Quek [16], una posible diferenciación de estos gestos sería:

- Los gestos deícticos, implican señas para establecer la identidad o ubicación espacial de un objeto a partir de una señalización.
- Los gestos manipuladores, generalmente se realizan con movimientos a mano alzada para imitar manipulaciones físicas, como en interfaces de realidad virtual.
- Los gestos específicos, que definen un conjunto de símbolos para comunicarse con las máquinas.
- La gesticulación, es una de las formas más naturales de gesticular. Estos son a menudo impredecible y difícil de analizar.

- Los gestos de lenguaje, son los gestos que se utilizan para los lenguajes de señas. Se realizan mediante una serie de gestos. que se combinan para formar estructuras gramaticales.

De entre todos los lenguajes de signos, el más común es el americano (ASL). La imagen 1 muestra los signos que hacen referencia al abecedario de este. Todos los gestos son estáticos excepto las letras 'J' y 'Z' [17]

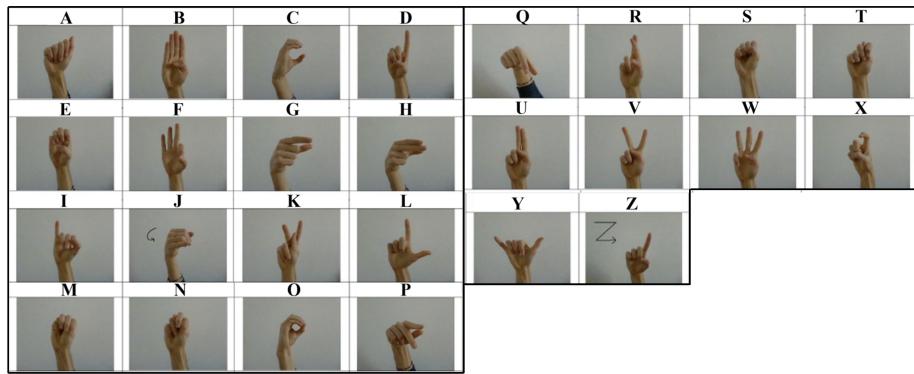


Figura 1: Alfabeto del lenguaje de signos americano (ASL). Todos los signos son estáticos, excepto la 'J' y la 'Z'.

Por último, la imagen 2 muestra la representación gestual de los signos numéricos, del uno al diez, del lenguaje de signos americano (ASL).



Figura 2: Números del lenguaje de signos americano (ASL),

## **2.2. Aproximaciones actuales**

Durante las últimas décadas, se han realizado diversas investigaciones en el campo del reconocimiento del lenguaje de signos (SLR). La mayoría se pueden dividir en dos enfoques:

- Los que hacen un reconocimiento de los signos estáticos, como es el abecedario, los números, algunas expresiones... Estos signos no tienen un movimiento, son capaces de ser representados a partir de un único gesto.
- Los que hacen un reconocimiento donde se incluyen los signos dinámicos, los cuales suceden en un espacio temporal.

Hay otra categorización en las investigaciones realizadas, la cual depende del modo de obtención de los datos. Podemos encontrar dos enfoques:

- Un enfoque basado en la visión, donde se hace uso de cámaras o detectores de imágenes. En estos se encuentran limitaciones a la hora de realizar y analizar las imágenes, se debe tener en cuenta la iluminación, el color y el fondo utilizado. Es importante poder diferenciar la parte del cuerpo que realiza el signo de la parte de la imagen que no aporta información.
- Un enfoque no basado en la visión. En este se hace uso de guantes, sensores de movimiento, rastreadores... Estos también presentan limitaciones, ya que se requiere de un coste adicional de *Hardware* y restringe los movimientos permitidos por los interpretes.

Durante la realización de este apartado se han investigado diversos artículos y libros de proyectos realizados sobre el reconocimiento del lenguaje de signos. Finalmente se ha dado más importancia a diez proyectos realizados con diferentes características, las cuales son: el tipo de datos empleados, las técnicas utilizadas y los resultados obtenidos.

### 2.2.1. Datos

En los artículos investigados se diferencian tres tipos de datos utilizados para el entrenamiento del modelo predictivo implementado. Estos son: imágenes, vídeos o coordenadas.

Las **imágenes** se suelen utilizar en los reconocimientos estáticos, ya que no requieren de ningún espacio temporal.

En el artículo *Sign language recognition using convolutional neural networks* [18], se adquieren las imágenes a partir de una cámara web. Una vez adquiridas son procesadas de tres formas:

- En el primer paso se aplica el algoritmo *Hue, Saturation, Value* (HSV), donde los fondos son detectados y eliminados. Este modelo divide el color de una imagen en 3 partes: la tonalidad, la saturación y el valor.
- El segundo paso de este procesamiento consiste en la segmentación, transformando las imágenes a escala de grises. Citando el artículo: "Por mucho que este proceso resulte en la pérdida de color en la región del gesto de la piel, también mejorará la robustez de nuestro sistema a los cambios de iluminación" [18].

Una vez realizada la transformación a una escala de grises, se modifica el tamaño a un marco de 64 por 64 píxeles.

- Por último, se realiza la extracción de características. "En nuestro caso, las características que se consideran cruciales son los píxeles binarios de las imágenes. Escalar las imágenes a 64 píxeles nos ha llevado a obtener características suficientes para clasificar de manera efectiva los gestos del lenguaje de señas estadounidense. En total, tenemos 4096 características, obtenidas después de multiplicar 64 por 64 píxeles." [18].

El dataset utilizado dispone de 10 signos del alfabeto del lenguaje de signos americano (ASL). Las letras son: 'A', 'B', 'C', 'D', 'K', 'N', 'O', 'T' e 'Y'. "Nuestro conjunto de datos tiene 2000 imágenes de gestos de signos estadounidenses, de las cuales 1600 son para entrenamiento y el resto 400 son para propósitos de prueba" [18].

En el artículo *Deep Learning-based sign language translation system* [19], se implementa un sistema capaz de detectar la posición de las manos a partir de imágenes obtenidas por cámara. Se utiliza un dataset que contiene 24 gestos

del lenguaje de signos americano (ASL).

Los artículos analizados también hacen uso de los **vídeos** como tipos de muestra, estas son los más comunes entre los SLR. Se suelen utilizar en el reconocimiento de gestos dinámicos, ya que requieren de una secuencia de imágenes para detectar la totalidad del gesto.

Un ejemplo del uso de este tipo de muestra es el artículo *Large-scale Video Classification with Convolutional Neural Networks* [20]. En este se explica el porque es más difícil el procesado de este tipo de datos, “a diferencia de las imágenes que se pueden recortar y volver a escalar a un tamaño fijo, los vídeos varían mucho en extensión temporal y no se pueden procesar fácilmente con una arquitectura de tamaño fijo.” [20]

Los datos son tratados como un conjunto de secciones de un mismo vídeo de tamaño fijo. “Dado que cada sección contiene varios fotogramas contiguos en el tiempo, podemos ampliar la conectividad de la red en la dimensión temporal para conocer las características espacio-temporales.” [20]

En el articulo *An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences* [1] se utiliza un dataset del lenguaje de signos alemán. Este contiene “4.839 terminologías, 8.257 videoclips, 947.756 fotogramas y 113.717 palabras en total.” [1]. Estos valores se muestran en la imagen 3.

	Train	Dev	Test
Vocab.	2,887	951	1,001
Clips	7,096	519	642
Frames	827,354	55,775	64,627
Tot. words	99,081	6,820	7,816

Figura 3: Estadísticas del dataset alemán utilizado en el artículo *An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences* [1].

El *Fingerspelling recognition with semi-Markov conditional random fields* [2] es otro articulo en el que se implementa un SLR a partir de vídeos. En este se estudia el abecedario americano (ASL), con la intención de poder deletrear, analizando las estadísticas de sus secuencias.

“Los experimentos utilizan grabaciones de vídeo de cuatro intérpretes nativos de ASL. Los datos se registraron a 60 fotogramas por segundo en un entorno de estudio” [2]. Con tal de poder guardar las muestras se creó el escenario mostrado en la imagen 4, con la intención de facilitar la obtención de muestras.



Figura 4: Escenario para la recopilación de datos en el artículo *Fingerspelling recognition with semi-Markov conditional random fields* [2]

“Cada intérprete representó una lista de 300 palabras a medida que aparecían en la pantalla de una computadora. Cada palabra se deletreaba dos veces, lo que arrojaba 600 instancias de palabras por cada intérprete. Las listas contenían palabras en inglés y extranjeras, incluidos nombres propios y sustantivos comunes en inglés” [2].

Una vez obtenidas todas las muestras y creado el dataset, se dividió en una proporción 80:10:10. El 80 % de los datos se utiliza como un conjunto de entrenamiento, el 10 % como un conjunto de desarrollo para ajustar los hiperparámetros y el 10 % restante como un conjunto para la prueba final.

Otros estudios, como es el caso de *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video* [3] han querido analizar en qué posición de la cámara se adquirían mejores resultados. La imagen 5 muestra las dos posiciones, la primera desde una visión frontal y la segunda desde una visión satélite, con la cámara situada en el sombrero del intérprete.



Figura 5: Posición de la cámara en dos obtenciones de muestras. La primera con una posición frontal y la segunda con una visión satélite con la cámara situada en el sombrero. Obtenido del estudio *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video* [3].

Las muestras contienen una visión global del interprete, estudios como *Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation* [21] utilizan frameworks como OpenPose para la detección de la posición de las manos de los interpretes.

Los últimos datos utilizados como muestras de un modelo predictivo son las **coordenadas**. Se obtienen a partir de componentes, tales como: sensores de movimiento, guantes, rastreadores... Su utilización es menos frecuente, ya que es incómodo para el interprete, su uso no es recomendable cuando llueve y es difícil de transportar ya que requieren de una computadora [18].

Otro tipo de componente que se puede utilizar para la obtención de este tipo de datos es el enfoques basados en modelos 3D, como se utiliza en el estudio *Exploiting Recurrent Neural Networks and Leap Motion Controller for Sign Language and Semaphoric Gesture Recognition* [22].

“El método propuesto utiliza un modelado basado en el esqueleto, donde una representación virtual de manos esqueléticas (u otras partes del cuerpo) se asigna a segmentos específicos” [22].

Cada muestra se identifica a partir del ángulo de la articulación junto con las longitudes del segmento. Luego, “mide las variaciones en el tiempo de las articulaciones del esqueleto cuyas coordenadas espaciales son adquiridas por un *Leap Motion Controller* (LMC)” [22].

Se comprobó que para adquirir una mayor precisión también se debía considerar “la información de los ángulos intradedo y la información espacial de la palma de la mano” [22].

Por último, otro componente a utilizar es el sistema de posicionamiento magnético, utilizado en el estudio *Gesture Recognition of Sign Language Alpha-*

*bet Using a Magnetic Positioning System [4].*

Este trabajo propone el uso de un sistema de posicionamiento magnético para reconocer los gestos estáticos asociados al alfabeto del lenguaje de signos. “Mide la posición en 3D y la orientación de los dedos dentro de un volumen operativo de aproximadamente 30 x 30 x 30 cm, donde los nodos receptores se colocan en posiciones conocidas” [4]. Esta estructura se muestra en la imagen 6.

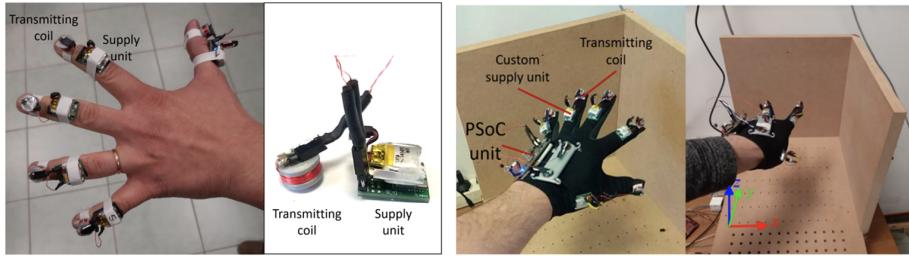


Figura 6: Ejemplo del uso de un sistema de posicionamiento magnético, obtenido del estudio *Gesture Recognition of Sign Language Alphabet Using a Magnetic Positioning System* [4]

Una ventaja de este componente, en comparación con los otros mencionados, es que “mide la posición absoluta de los dedos. Permite una alta precisión y un seguimiento sin desviaciones. Además, funciona también en presencia de obstrucciones causadas por objetos o partes del cuerpo” [4].

Una vez se obtienen todas las muestras, se divide en dataset en una proporción 90:10, el 90 % para el entrenamiento y el 10 % restante para las pruebas finales.

### 2.2.2. Técnicas

Entre los artículos investigados se diferencian diversos tipos de técnicas para el entrenamiento del modelo predictivo.

En el proyecto *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video* [3], se utiliza el **modelo de Márkov**.

A partir de dos sistemas basados en el modelo oculto de Márkov, se puede predecir oraciones del lenguaje de señas americano. “Se consideró una topología inicial de cinco estados para el signo más complejo. Para manejar señales menos complicadas, se especificaron transiciones de salto que permitieron que la topología emulara estrictamente tres o cuatro estados” [3]. El modelo de cuatro estados se puede ver representado en la imagen 7.

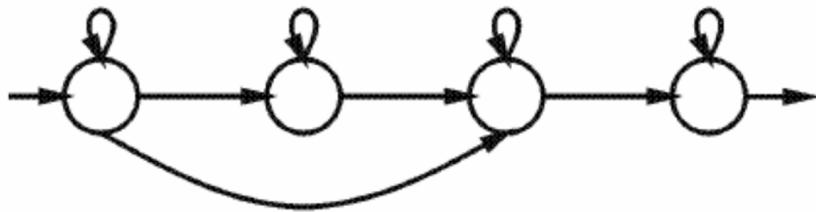


Figura 7: Diagrama de los cuatro estados de uno de los modelos oculto de Márkov implementado en el estudio *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video* [3].

En otros estudios también se ha utilizado el modelo de Márkov, como es el caso de *Fingerspelling recognition with semi-Markov conditional random fields* [2]. Pero a diferencia del anterior, se desarrolló un modelo condicional semi-Márkov, donde las características se definen sobre los segmentos de vídeo. Para cada característica o símbolo, se entrenó un clasificador.

En el mismo estudio, también se entrenó un clasificadores de redes neuronales. Se consideraron varios tipos de funciones en la capa de salida: lineal, softmax y sigmoid.

Otra técnica que se ha utilizado en más de un estudio son las **redes neuronales recursivas**.

En el artículo *Phonologically-meaningful Subunits for Deep Learning-based Sign Language Recognition* [5] “se emplean versiones bidireccionales apiladas de estas RNN, combinando la salida de las capas de avance y retroceso cons-

tituyentes a través de la concatenación de vectores. Las RNN bidireccionales demuestran ser mejores para modelar la temporalidad” [5].

La imagen 8 muestra la estructura del modelo predictivo implementado.

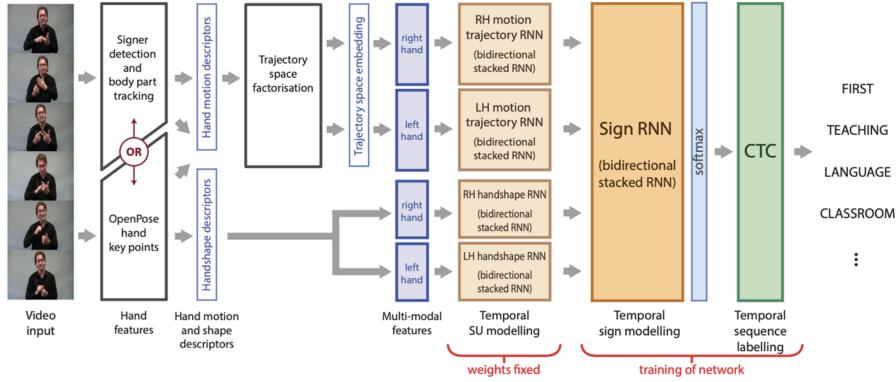


Figura 8: Estructura de la implementación del modelo predictivo con RNN en el estudio *Phonologically-meaningful Subunits for Deep Learning-based Sign Language Recognition* [5].

En el artículo *Exploiting Recurrent Neural Networks and Leap Motion Controller for Sign Language and Semaphoric Gesture Recognition* [22] se hace uso de Deep LSTM (DLSTM). Esta arquitectura permite entrenar la red neuronal en diferentes tiempos sobre una secuencia de entrada.

Este objetivo se ha podido cumplir mediante el uso de una conjunto de redes neuronales recurrentes (RNN) con arquitectura Long Short Term Memory (LSTM), un tipo particular de red neuronal profunda (DNN) donde las conexiones entre sus nodos forman un ciclo dirigido.

“Las RNN, a diferencia de las DNN comunes, pueden modelar información contextual a largo plazo de secuencias temporales obteniendo así excelentes resultados en el análisis de sonido y el reconocimiento del habla. El LSTM es una arquitectura en la que un RNN utiliza unidades especiales en lugar de una función de activación común. Las unidades LSTM ayudan a propagar y preservar el error a través del tiempo y las capas” [22].

En el artículo *An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences* [1] se utiliza el algoritmo heurístico FSDC para comprimir la densidad de los fotogramas de entrada sin eliminar la información clave.

El sistema que se propone está compuesto por un módulo de Neural Machine Translation (SMT) con un codificador optimizado, una red neuronal convolucional temporal y una red híbrida Gated Recurrent Unit (GRU) bidireccional, que comprime la información de los tokens obtenidos en las muestras.

Otra técnica frecuentemente utilizada es el algoritmo **Support Vector Machines** (SVM), un clasificador binario de aprendizaje automático supervisado que tiene como objetivo encontrar los parámetros del hiperplano que mejor separa los datos en dos clases [4].

En el artículo *Gesture Recognition of Sign Language Alphabet Using a Magnetic Positioning System* [4], se considera un conjunto de algoritmos de clasificación, como el k vecinos más cercanos (KNN), regresión logística y otros, pero el que les dio los mejores resultados fue SVM con kernel Gaussiano.

“El SVM no está diseñado para clasificación binaria únicamente. Incluso si la formulación inicial estaba destinada a la clasificación binaria, cada clasificador que realiza la clasificación binaria puede extenderse a la clasificación multiclasa mediante el uso de uno de los enfoques entre Uno contra Uno (OvO) o Uno contra Todos (OvA). En este artículo, utilizamos un enfoque OvO” [4].

De entre todas las técnicas, la más utilizada es la **Red Neuronal Convolucional** (CNN).

Se han encontrados diversos artículos donde se hacía uso de esta red neuronal para implementar el modelo predictivo, entre ellos está el *Sistema de reconocimiento de lenguaje de señas mediante red neuronal convolucional y visión por computadora* [18].

El modelo implementado dispone de tres capas, todas utilizan la función de activación ReLu. La capa de salida está conectada con diez unidades, y la función de activación es Softmax. En la compilación del modelo se utiliza la entropía cruzada de categorías como función de pérdida y Adam como optimizador.

Hay algunos estudios que utilizan este modelo con muestras dinámicas, como es el caso de *Large-scale Video Classification with Convolutional Neural Networks* [20]. Propone dos soluciones para que este modelo tenga en cuenta la temporalidad:

- Modificando los filtros convolucionales de la primera capa para extenderlos en el tiempo.
- Posicionando dos redes independientes de un solo cuadro, a cierta distancia en el tiempo, y fusionando sus salidas en el procesamiento.

En el artículo *Deep Learning-based sign language translation system* [19] se combinan dos técnicas mencionadas. Se propone una estructura de un sistema de traducción de lenguaje de señas basado en redes neuronales convolucionales y el algoritmo de Support Vector Machine para la clasificación de la imagen.

Este entrenamiento del sistema híbrido se realiza utilizando una técnica de validación cruzada y se valida mediante el estimador de Monte Carlo. El sistema propuesto permite incrementar la velocidad y la precisión del sistema diseñado.

Por último, en el proyecto *Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation* [21] se utilizan dos modelos de **redes neuronales convolucionales en 3D**. Se utilizaron Multilayer Perceptron (MLP) y la función SoftMax para la clasificación.

“Se ha utilizado una red neuronal convolucional 3D (3DCNN) para el modelado de vídeo, que es una versión ampliada de las CNN estándar que utiliza filtros espacio-temporales. Esta arquitectura se ha explorado previamente en varios campos de análisis de vídeo para la representación de características espacio temporales” [21].

La característica más importante de 3DCNN es su capacidad para representar jerárquicas de datos espacio-temporales. Pero como inconvenientes, requiere más parámetros que 2DCNN y una dimensión de kernel adicional, lo que dificulta el entrenamiento.

### 2.2.3. Resultados

Una vez se ha analizado toda la información de los destinos artículos y libros que se han consultado, se ha creado la tabla 1. Esta es un resumen de los proyectos más significativos que se han encontrado, donde se esquematizan los resultados obtenidos dependiendo de la técnica y los datos utilizados.

TÉCNICAS / DATOS	Imágenes	Vídeos	Coordenadas
<b>Markov</b>	–	Entre un 92 % y un 97 % de accuracy [3]	–
<b>Semi-Markov con Neural Network</b>	–	Mejora en el LER de un 16,3 % a un 11,6 % [2]	–
<b>Support Vector Machines (SVM)</b>	–	–	Entre 77,19 % y 96,79 % de accuracy [4]
<b>Convolutional Neural Networks (CNN)</b>	Accuracy superior al 90 % [18]	Entre 55,3 % y 63,9 % de accuracy [20]	–
<b>CNN y SVM</b>	Accuracy del 99,9 % RMSE de 0,0126 [19]	–	–
<b>Recurrent Neural Network (RNN)</b>	–	71,9 % de accuracy [5]	Entre 91,43 % y 97,62 % de accuracy [22]
<b>Gated recurrent unit (GRU) y el algoritmo heurístico FSDC</b>	–	Reducción de un 9,28 % de los datos de entrenamiento. Reducción de un 10 % del tiempo de procesamiento. [1]	–
<b>3DCNN</b>	–	87,69 % de accuracy [21]	–

Tabla 1: Comparativa de los resultados obtenidos en diez proyectos dependiendo de las técnicas y los tipos de datos utilizados.

Es difícil hacer una comparación entre ellos, ya que cada autor ha mostrado la precisión del modelo predictivo a partir de diferentes parámetros. Estos han sido: la *accuracy*, el *letter error rate* (LER), el error cuadrático medio (RMSE),

el coste de memoria y el coste de tiempo. Aún así, nos podemos hacer una idea de cuales pueden ser las técnicas mas recomendadas dado los datos utilizados.

Siguiendo la tabla 1, si se utiliza un dataset estático, sería recomendable utilizar un modelo predictivo basado en redes neuronales convolucionales con el algoritmo de aprendizaje supervisado SVM como clasificador.

Si se está utilizando un dataset dinámico, donde se utilizan datos en formato de vídeo, una muy buena opción sería hacer uso de las redes neuronales convolucionales en 3D.

Por último, si los datos se están obtenido en formato de coordenadas, ya sea por el uso de guantes, modelos basados en 3D o un sistema con posicionamiento magnético, una red neuronal recurrente podría dar una precisión muy elevada.

Aún así, se debe tener en cuenta que los resultados no solo dependen del tipo de datos utilizado, pero la tabla 1 permite esquematizar los resultados obtenidos para facilitar su evaluación.

### **3. Fundamentos**

El Machine learning consiste en una serie de técnicas que permiten implementar algoritmos para poder llevar a cabo una clasificación. Permite parsear datos, aprender de ellos y ser capaz de hacer una predicción [23].

La clasificación es una tarea donde se predice la etiqueta correspondiente a los datos de entrada del sistema. Requiere un conjunto de datos de entrenamiento con muchos ejemplos de entradas y salidas de los que aprender [24].

Un modelo utilizará el conjunto de datos de entrenamiento y calculará la mejor manera de relacionar los ejemplos de datos de entrada con las etiquetas. Este conjunto de datos debe ser suficientemente representativo del problema y tener muchos ejemplos de cada etiqueta de clase [24].

Para poder llevar a cabo esta técnica, se debe hacer una extracción y selección de las características de los datos a evaluar, esto requiere un gran conocimiento del dominio. Deep Learning permite automatizar estos procesos [25].

#### **3.1. Tipos de tareas de clasificadores**

Una tarea de clasificación es el tipo de predicción que se realiza en función del problema y los datos disponibles. Los más comunes son: binario, multi-clase, multi-etiqueta y desequilibrado [26].

##### **3.1.1. Clasificación binaria**

Esta tarea de clasificación se diferencia por utilizar únicamente dos etiquetas. Cualquiera de las muestras del conjunto de datos puede tomar un valor de entre dos clases.

Posibles ejemplos a esta tarea sería una predicción de verdadero a falso, como puede ser la detección de correos no deseado, limitando los valores a spam o no spam.

Normalmente, las tareas de clasificación binaria involucran una clase que es el estado normal y otra clase que es el estado anormal. En el ejemplo anterior, la clase spam sería el estado anormal [27].

Los algoritmos más comunes en el uso de la clasificación binaria son [28] :

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Decision Trees
- Support Vector Machine (SVM)
- Naive Bayes

### 3.1.2. Clasificación multi-clase

La clasificación de clases múltiples se refiere a aquellas tareas de clasificación que tienen más de dos etiquetas de clase.

El número de etiquetas de clase puede ser muy grande en algunos problemas. Por ejemplo, un modelo puede predecir que una foto pertenece a uno entre miles o decenas de miles de rostros en un sistema de reconocimiento facial [29].

Los algoritmos más populares para esta clasificación son:

- Random Forest
- K-Nearest Neighbors (KNN)
- Decision Trees
- Gradient Boosting
- Naive Bayes

Los algoritmos que están diseñados para la clasificación binaria se pueden adaptar a problemas que disponen de más de una clase. Para poder llevarlo a cabo se requiere especificar la estrategia, la cual puede diferenciarse entre [30] :

- **One vs Rest o One vs All** : se implementa una clasificación binaria para cada clase, frente a todas las demás clases.
- **One vs One**: se implementa una clasificación binaria para cada par de clases.

La imagen 9 muestra la comparación entre los tipos de estrategias definidas.

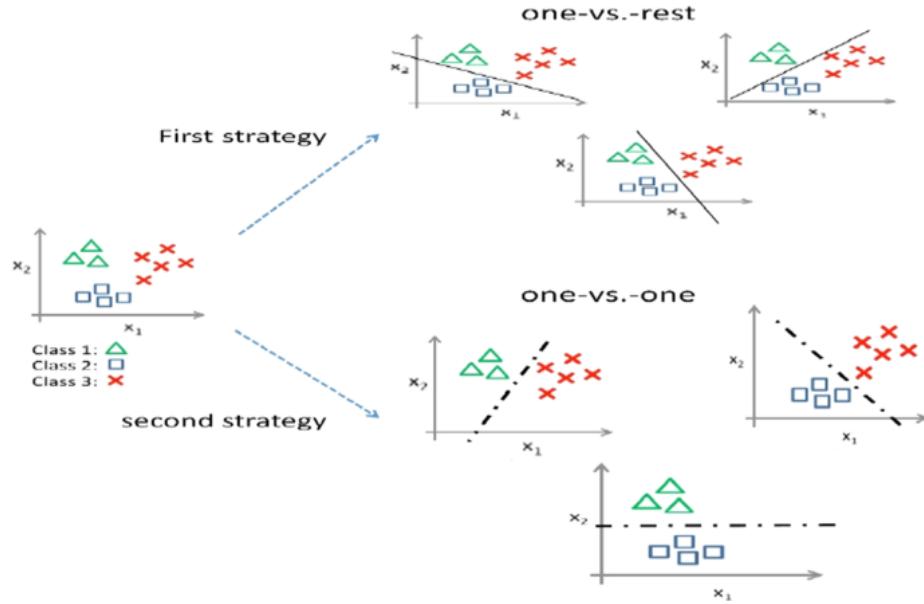


Figura 9: Comparación de las estrategias de ajuste de múltiples modelos de clasificación binaria con datos de diversas clases.

### 3.1.3. Clasificación multi-etiqueta

La mayoría de los problemas de clasificación asocian una sola clase a cada muestra. Sin embargo, hay muchas tareas de clasificación en las que cada instancia se puede asociar con una o más clases [31] [32].

Un ejemplo sería la categorización de una prenda, esta puede ser un pantalón, una camiseta, unos zapatos.. pero también se puede categorizar dependiendo de su color. Con lo que nos podemos encontrar con una muestra que deba ser evaluada como: pantalón - azul. En este caso podemos ver como a una misma instancia se han asignado dos etiquetas.

Se pueden utilizar versiones de algoritmos de clasificación estándar para su implementación, los más comunes son[33]:

- Árboles de decisión de etiquetas múltiples.

- Bosques aleatorios de múltiples etiquetas.
- Aumento de gradiente de etiquetas múltiples.

### 3.1.4. Clasificación desequilibrada

La clasificación desequilibrada se refiere a tareas de clasificación en las que el número de ejemplos de cada clase se distribuye de forma desigual.

Suelen ser tareas de clasificación binaria en las que la mayoría de los ejemplos del conjunto de datos de entrenamiento pertenecen a la clase normal y una minoría de ejemplos pertenecen a la clase anormal.[34].

Posibles ejemplos serían la detección de fraudes o las pruebas de diagnóstico médico. Los algoritmos utilizados deben dar preferencia a la clase minoritaria, los más utilizados son [35]:

- Logistic Regression.
- Decision Trees.
- Support Vector Machines (SVM).

## 3.2. Métodos de clasificación

Hay diversos métodos de clasificación, entre ellos se encuentran las redes neuronales y los árboles de decisión.

### 3.2.1. Redes Neuronales

Las redes neuronales son "máquinas de aprendizaje distribuidas, adaptativas y generalmente no lineales". [36] Están formadas por neuronas denominadas elementos de procesamiento (PE), interconectadas entre ellas donde cada conexión tienen asignadas un peso. [36]

Una red neuronal se compone de tres partes, diferenciadas por capas: la capa d'entrada, la oculta y la de salida. Los nodos de la capa de entrada no disponen de enlace de entrada y los nodos de la capa de salida no disponen de enlaces de salida. Esta definición no se cumple con las redes neuronales recurrentes, donde la capa de salida estás conectada a la capa oculta. [37]

La capa de entrada es la primera capa en una red neuronal profunda y proporciona los parámetros de entrada necesarios para procesar la información. No se realiza ningún calculo.

En las capas ocultas se realizan los cálculos necesarios sobre las entradas recibidas de las capas anteriores y pasan el resultado a la siguiente. Es crucial decidir el número de capas y el número de elementos de procesamiento para aumentar su eficiencia. Cuanto mayor sea el número de capas ocultas, más profunda es la red.

Por ultimó, la capa de salida proporciona el resultado final después de recibir los *outputs* de las capas anteriores. [38]

El resultado obtenido a partir de este modelo depende de tres funciones:

- La función de activación: esta función es opcional. Su objetivo es definir si el valor de la salida es válido o no. En el caso de que no lo sea lo desactivará, no siendo utilizado el valor de esta neurona en la siguiente capa. [39]
- La función de propagación: es el valor obtenido a partir del sumatorio de todas las entradas, multiplicado por el peso asignado a la conexión entre las neuronas. Se le debe añadir el valor de *bias*, el cual en castellano hace referencia al término sesgo y que en la imagen 10 se muestra con el símbolo

$W_0$ . Este dispone de un valor constante que incrementa o decremente el sumatorio de todas las entradas. Permite desplazar la función de activación hacia la izquierda o hacia la derecha, mejorando el aprendizaje del modelo. [40]

- la función de transferencia: esta función traduce las señales de entrada en señales de salida entendibles. Las más comunes son: *threshold*, *sigmoid*, *softmax* y *gaussian*. [41]

Estas funciones se pueden visualizar en la imagen 10:

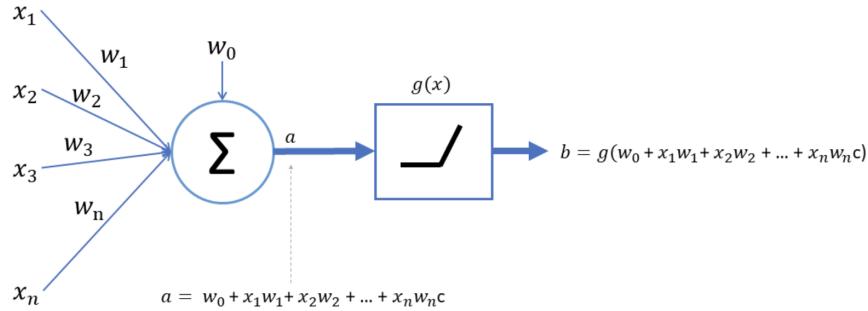


Figura 10: Funcionamiento de la obtención del resultado de una red neuronal.

La imagen 11 muestra el mismo funcionamiento que la imagen 10, pero visualizando la unión entre las diferentes neuronas.

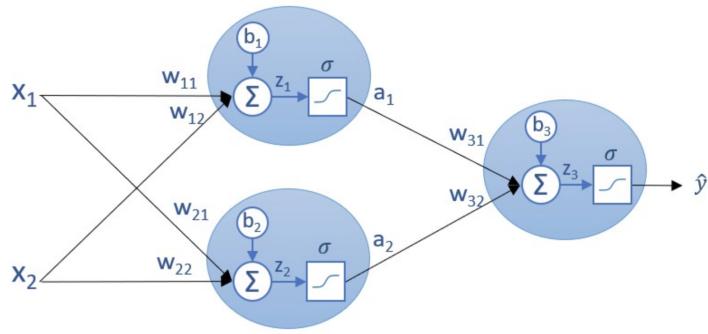


Figura 11: Funcionamiento de la obtención del resultado de una red neuronal.

Matemáticamente, una red neuronal se puede representar de la siguiente forma [6]: supongamos que disponemos de una estructura con  $n$  neuronas de entrada ( $x_1, \dots, x_n$ ),  $h$  neuronas ocultas ( $z_1, \dots, z_h$ ) y  $m$  neuronas de salida ( $y_1, \dots, y_m$ ).

Dada esta estructura, disponemos de las siguientes características:

- $T_j$  es el *bias* de las neuronas  $z_j$ .
- $W_{ij}$  es el peso asignado a la conexión entre la neurona  $x_i$  a la neurona  $z_j$ .
- $B_{jk}$  es el peso asignado a la conexión entre la neurona  $z_j$  a la neurona  $y_k$ .

Esta estructuración se muestra en la imagen 12.

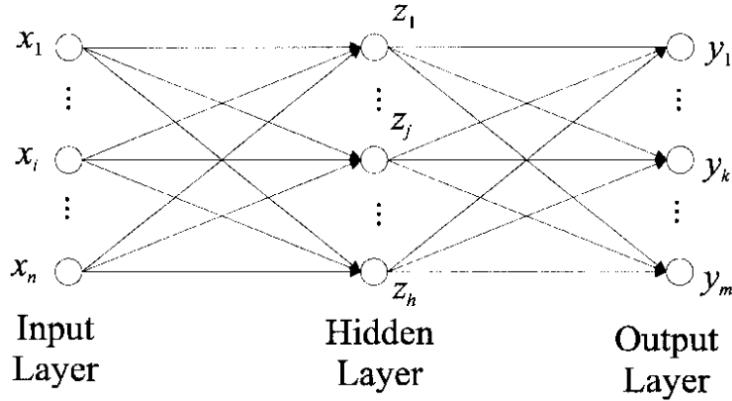


Figura 12: Estructura de una red neuronal. La imagen ha sido obtenida del artículo *Are artificial neural networks black boxes?* [6]

La función que se calcula para poder realizar las predicciones presenta la siguiente expresión:

$$F(x_1, \dots, x_n) = (y_1, \dots, y_m)$$

Donde cada neurona de salida obtiene su valor a partir de la siguiente función:

$$y_k = gA\left(\sum_{j=1}^h z_j B_{jk}\right)$$

Y donde cada neurona oculta obtiene su valor a partir de la siguiente función:

$$z_k = fA\left(\sum_{i=1}^n x_i W_{ij} + T_j\right)$$

Las funciones de activación se representan como  $gA$  y  $fA$ . Suelen ser continuas, acotadas, no decreciente y no lineal. La más común es:

$$fA(x) = \frac{1}{1 + e^{-x}}$$

Este conjunto de clasificadores es uno de los más utilizados para implementar los modelos predictivos, ya que dispone de muchas ventajas, como [36]:

- Puede aprender de un gran número de muestras.
- Una vez entrenada, donde se establecidos las conexiones, su ejecución puede ser muy rápida debido a la gran cantidad de elementos que trabajan en paralelo.
- Dispone de un alto grado de resistencia al error. Las predicciones no se suelen ver afectadas si un solo elemento de procesamiento falla o si algunas conexiones se rompen.

Aún sus numerosas ventajas, también dispone de inconvenientes. El mayor de todos es la falta de comprensión de cómo resuelve las predicciones. Al no desglosar un problema en sus elementos lógicos, sino que lo resuelven mediante un enfoque holístico, es difícil entender su lógica. Esto supone una desventaja cuando se quiere determinar si la solución proporcionada es correcta. [38]

Hay diferentes tipos de redes neuronales en Deep Learning, pero las más comunes son:

- **Redes neuronales artificiales (ANN)**

La red neuronal artificial es un grupo de múltiples neuronas en cada capa. También se conoce como una red neuronal Feed-Forward porque las

entradas se procesan solo en la dirección de avance.

Son capaces de aprender cualquier función no lineal gracias a la función de activación, sin esta, la red solo aprende la función lineal y nunca puede llegar a realizar relaciones complejas. Se suelen utilizar para resolver problemas con datos de imagen o texto.

Una de las principales razones detrás de la aproximación universal es la función de activación. Las funciones de activación introducen propiedades no lineales a la red. Esto ayuda a la red a conocer cualquier relación compleja entre entrada y salida. [42]

En la imagen 13 se muestra la estructura del modelo de una red neuronal artificial.

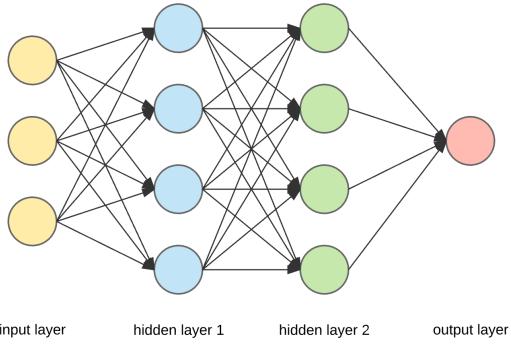


Figura 13: Modelo de una red neuronal artificial

#### ■ Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales son especialmente frecuentes en proyectos de procesamiento de imágenes y vídeos.

Los componentes básicos de este tipo de redes neuronales son filtros, también conocidos como núcleos. Se utilizan para extraer las características relevantes de las muestras de entrada mediante la convolución, creando un mapa de características.

Las características que captura son las referentes a la disposición de los píxeles y la relación entre ellos en una imagen. Permite identificar el objeto

con precisión, la ubicación de este y su relación con otros objetos en una imagen. [43]

En la imagen 14 se muestra un proceso de entrenamiento y predicción que siguen estos tipos de redes neuronales.

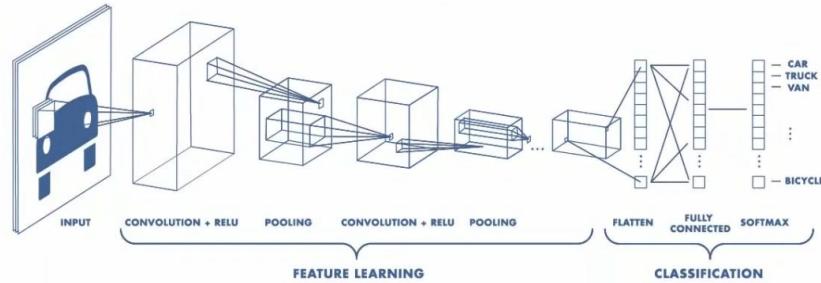


Figura 14: Modelo de una red neuronal convolucional

#### ■ Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes tienen una conexión de retroalimentación de la capa de salida a la primera capa oculta. Esta restricción de bucle asegura que la información secuencial se capture en los datos de entrada.

Se suelen utilizar con datos que tienen una secuencialidad, como son las series de tiempo, textos y audio.

Los RNN comparten los parámetros en diferentes etapas de tiempo. Esto se conoce popularmente como Parameter Sharing. Esto da como resultado menos parámetros para entrenar y disminuye el costo computacional. [44]

La imagen 15 muestra la estructura de una red neuronal recurrente.

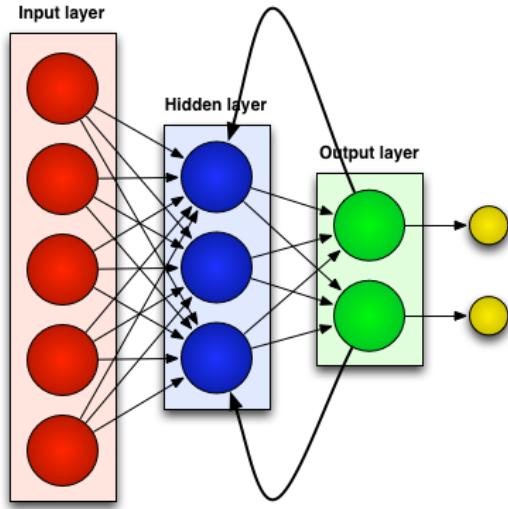


Figura 15: Modelo de una red neuronal recurrente

### 3.2.2. Árboles de decisión

El árbol de decisión es un modelo de predicción que se utiliza en el Machine learning. Su aprendizaje es un proceso para encontrar las reglas óptimas en cada nodo del árbol interno de acuerdo con la métrica seleccionada.

Tiene una estructura similar a un árbol inverso, donde los valores se encuentran en sus hojas. Para llegar a la hoja, la muestra se propaga a través de los nodos, comenzando en el raíz y en cada nodo se toma la decisión del nodo descendiente al que transferir la muestra. La decisión es tomada en función de la característica de la muestra seleccionada. [7]

Su estructura es similar a un diagrama de flujo, donde cada nodo interno realiza una prueba de un atributo en concreto y cada rama representa el resultado de la prueba. la etiqueta de clase está representada por cada nodo hoja (o nodo terminal). También se puede estructurar un árbol donde un nodo analice más de una característica. [8]

El árbol de decisión se puede construir relativamente rápido en comparación con otros métodos de clasificación y suelen obtener una precisión similar o mejor. [8]

En la imagen 16 se muestra la estructura descrita del árbol de decisión.

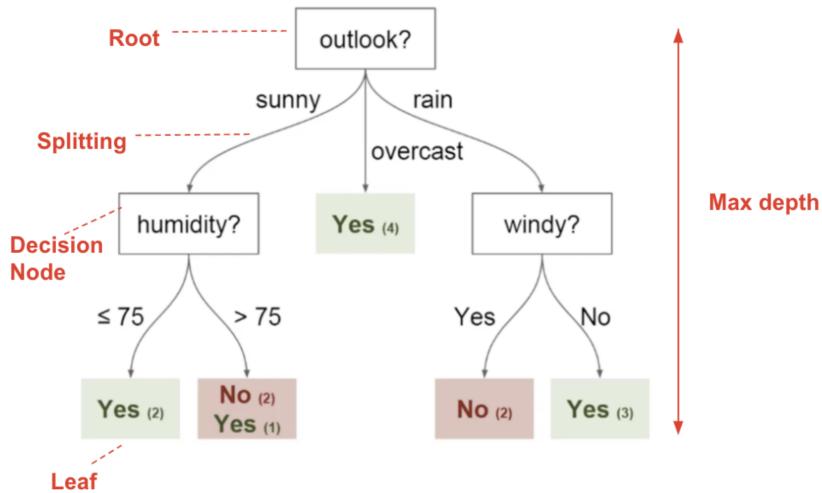


Figura 16: Estructura y partes de un árbol de decisión.

En el entrenamiento se utilizan vectores como datos de entrada, cada uno perteneciente a una clase conocida, y la salida son los valores asignados a las clases. [45]

Se ha demostrado que la eficiencia de este modelo aumenta cuando las dimensiones del árbol son reducidas. Otra forma de aumentar su eficiencia es generando múltiples árboles de los datos, utilizándolos todos en la predicción de nuevas instancias. [7]

En la figura 17 se muestra la estructuración de un árbol de decisión a partir de un conjunto de muestras de entrenamiento. Podemos ver como en la gráfica hay tres clases distinguidas:

- La del color azul con un valor numérico igual a uno.
- La del color rojo con un valor numérico igual a dos.
- La del color verde con un valor numérico igual a tres.

En el entrenamiento se analizan las características que diferencian las clases mencionadas. Se definen las particiones, en este ejemplo se pueden visualizar cuatro:  $[(-6, -4), (-1.4, 0.7)]$ ,  $[(-1.4, -4), (5, 0.7)]$ ,  $[(-6, 0.7), (-0.6, 8)]$  y  $[(-0.6,$

$(0.7), (5, 8)]$ .

Dado estas particiones se construye el árbol de decisión de la imagen 17. Siguiendo los puntos de corte, se definen los límites de cada clase.

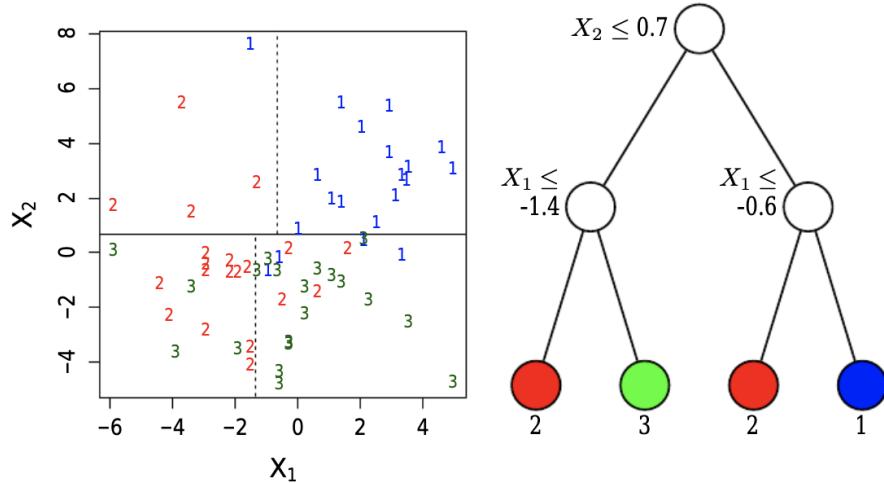


Figura 17: Estructura de un árbol de decisión basada en las particiones. Esta imagen ha sido obtenido del artículo *Classification and Regression Trees* [7].

Las tres técnicas más conocidas, que los árboles de decisión utilizan para realizar la clasificación, son: ID3, C4.5 y CART. [8]

- Algoritmo ID3: construye el árbol de decisiones aplicando una búsqueda exhaustiva, de arriba hacia abajo, a través de todos los conjuntos de atributos en cada nodo del árbol. Seleccionar el atributo con el que se realizarán las pruebas de cada nodo investigando cual es el que aporta una mayor ganancia de información. Gracias a este algoritmo, se reduce la información de clasificación necesaria en cada nodo. [8]
- Algoritmo C4.5: es una extensión del algoritmo ID3 anterior. Este puede aceptar datos con valores categóricos o numéricos. Se genera un umbral para poder trabajar con valores continuos. El algoritmo C4.5 puede manejar fácilmente los valores perdidos, ya que los valores de los atributos perdidos no se utilizan en sus cálculos. [8]
- Algoritmo CART: sus siglas significan *Classification And Regression Trees*. Este algoritmo permite construir tanto clasificaciones como árboles de re-

gresión. La construcción del árbol de clasificación por CART se basa en la división binaria de los atributos. Utiliza el índice de Gini (una medida de desigualdad) para realizar esta división. [8]

La imagen 18 muestra una comparativa de estos tres algoritmos.

<i>Features</i>	<i>ID3</i>	<i>C4.5</i>	<i>CART</i>
Type of data	Categorical	Continuous and Categorical	continuous and nominal attributes data
Speed	Low	Faster than ID3	Average
Boosting	Not supported	Not supported	Supported
Pruning	No	Pre-pruning	Post pruning
Missing Values	Can't deal with	Can't deal with	Can deal with
Formula	Use information entropy and information Gain	Use split info and gain ratio	Use Gini diversity index

Figura 18: Comparativa de diferentes algoritmos de árboles de decisión. Esta tabla ha sido obtenido del artículo *A Survey on Decision Tree Algorithms of Classification in Data Mining* [8].

Hay algoritmos que se utilizan para aumentar el rendimiento de los árboles de decisión. Uno de ellos es el algoritmo XGboost, Extreme Gradient Boosting. Es conocido por su alta velocidad y rendimiento en comparación con todos los demás algoritmos de aprendizaje automático. De los algoritmos mostrados anteriormente, el único que puede soportar el XGboost es el algoritmo CART. [46]

XGboost hace uso del algoritmo Gradient Boosting. Se caracteriza por corregir el error anterior cometido por el modelo y aprender de él. Dispone de diversas ventajas, como son: [47] [48]

- Utiliza una técnica de regularización para eliminar el *overfitting* del modelo.
- Está habilitado con la función de validación cruzada.
- Utiliza funciones objetivas para evaluar el rendimiento del modelo y poder manejar las métricas de validación definidas por el usuario.
- Permite guardar la matriz de datos y poder recargarla posteriormente, con la intención de ahorrar recursos y tiempo.

## 4. Condiciones de ejecución del proyecto

Este proyecto ha sido realizado en diferentes entornos, a medida que se necesitaba una mayor capacidad se ha ido evolucionando para poder ser ejecutado en entornos adecuados.

Las primeras ejecuciones se realizaron en el ordenador personal, el cual es un portátil MacBook Pro con las siguientes características:

- Identificador del modelo: MacBookPro15,1
- Nombre del procesador: Intel Core i7 de 6 núcleos
- Velocidad del procesador: 2,2 GHz
- Cantidad de procesadores: 1
- Cantidad total de núcleos: 6
- Caché de nivel 2 (por núcleo): 256 KB
- Caché de nivel 3: 9 MB
- Memoria: 16 GB
- Versión del firmware del sistema: 1554.140.20.0.0 (iBridge: 18.16.14759.0.1,0)

A medida que se necesitaba una mayor capacidad de memoria se fueron buscando nuevas alternativas. La primera fue Google Colab<sup>1</sup>, una herramienta enfocada a la educación y la investigación del Machine Learning. Es un entorno de Jupyter Notebook<sup>2</sup> que no requiere una configuración para su uso.

Se utilizó la versión gratuita, pero esta estaba muy limitada. Anteriormente se podían llegar a ejecutar proyectos que requiriesen un coste de memoria muy elevado, pero actualmente esto se ha restringido en la versión gratuita.

Por último se utilizó Jupyter notebook, un entorno de desarrollo interactivo basado en web para desarrollar software de código abierto.

Durante toda el proceso de desarrollo del proyecto se ha ido alternando entre Jupyter notebook y el Mac, hay ejecuciones que no se han logrado completar, ya que se requería una capacidad aún mayor. Pero en general estos dos entornos han sido suficientes para el desarrollo del proyecto.

---

<sup>1</sup><https://colab.research.google.com/>

<sup>2</sup><https://jupyter.org/>

El IDE en el que se ha programado la mayor parte del código ha sido PyCharm<sup>3</sup>, una herramienta para el desarrollo de software perteneciente a Jet Brain<sup>4</sup>.

Cuando falta poco para final el proyecto se caducó la licencia, con lo que se tuvo que cambió de IDE. Se pasó a utilizar Visual Studio Code<sup>5</sup>, un editor de código fuente desarrollado por Microsoft.

Por último, el *environment* utilizado en Python ha sido conda<sup>6</sup>, un gestor de paquetes y de entornos de código abierto multiplataforma.

---

<sup>3</sup><https://www.jetbrains.com/es-es/pycharm/>

<sup>4</sup><https://www.jetbrains.com/es-es/>

<sup>5</sup><https://code.visualstudio.com/>

<sup>6</sup><https://docs.conda.io/>

## 5. Experimentación

El objetivo de este proyecto ha sido poder interpretar el lenguaje de signos y mostrar su significado en formato texto, a partir de imágenes. Para poder alcanzar este objetivo se han hecho diversas pruebas y experimentos.

### 5.1. Experimentación 1: Búsqueda del dataset

En la primera versión se ha utilizado un dataset de la comunidad de datos Kaggle.<sup>7</sup> Consiste en 37 gestos diferentes que representan el abecedario inglés, con las letras de la 'A' a la 'Z', los números del 0 al 9 y el carácter espacio. Cada gesto dispone de 1500 imágenes de 50x50 píxeles. Cada imagen está repetida, ya que hay dos versiones de la misma, una ya preprocesada con la intención de mejorar el entrenamiento y las pruebas de las redes neuronales, y otra versión con las imágenes originales en color. Se han utilizado las dos versiones para poder comparar sus resultados.

En la figura 19 se muestran diez imágenes que hacen referencia a diez gestos obtenidos del dataset de la segunda versión mencionada.



Figura 19: Imágenes obtenidas del dataset *Sign Language Gesture Images Dataset* [9] en color.

Como se puede contemplar en la figura 19, todas las imágenes han sido tomadas con un fondo de un mismo color, centrándose únicamente en la mano que realiza el gesto. La tonalidad de las imágenes cambia dependiendo del gesto, pero todas las imágenes que hacen referencia al mismo, contienen una tonalidad

<sup>7</sup><https://www.kaggle.com>

similar.

A continuación, en la figura 20, se muestran otras diez imágenes que hacen referencia a los mismos gestos que en la figura 19, pero estas han sido obtenidas de la primera versión mencionada.



Figura 20: Imágenes obtenidas del dataset *Sign Language Gesture Images Dataset* [9] preprocesadas con la intención de mejorar el entrenamiento de las redes neuronales.

En total, este dataset dispone de unas 55.500 imágenes, englobando las dos versiones.

## 5.2. Experimentación 2: Uso de redes neuronales

Una vez obtenido el dataset, se decidió empezar utilizando redes neuronales, ya que es una de las soluciones más recomendadas y utilizadas para el reconocimiento de imágenes, como se ha podido apreciar en el estado del arte en el apartado 2.2.2.

### 5.2.1. Red neuronal artificial

Se analizaron diferentes tipos de redes neuronales, como se especifica en los fundamentos, en el apartado 3.2.1. Finalmente se decidió empezar por la implementación de una red neuronal artificial, con el objetivo de poder entender con mayor precisión como se implementa este modelo de clasificación.

La imagen 21 muestra la estructura de la red neuronal implementada.

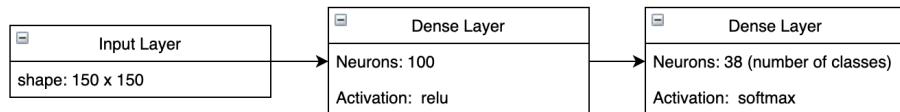


Figura 21: Estructura de la red neuronal artificial implementada.

Esta estructura está formada por tres capas: una de entrada, una oculta y una de salida.

La capa de entrada requiere un conjunto de imágenes donde todas ellas presenten el mismo tamaño. En este caso, se han utilizado imágenes con unas dimensiones de 150 x 150 píxeles.

La capa oculta dispone de 100 neuronas con una función de activación ReLu (*Rectified Linear Activation Function*), una función lineal por partes. En el caso de que la entrada sea positiva devolverá el valor introducido, en el caso de que sea negativa el valor a devolver será cero. Esta función de activación es una de las mas comunes en las redes neuronales, ya que facilita el entrenamiento consiguiendo un rendimiento elevado. La imagen 22 muestra la gráfica de esta función.

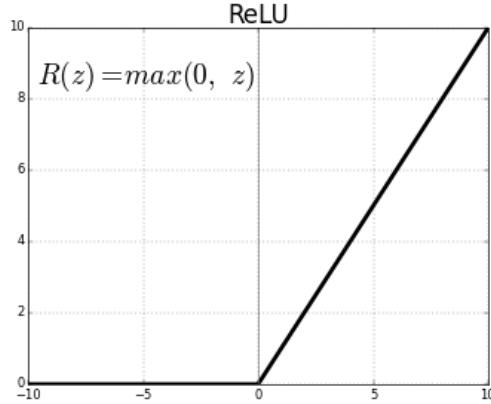


Figura 22: Gráfica de la función de activación ReLU (*Rectified Linear Activation Function*).

Por último, la capa de salida dispone de 38 neuronas, este valor dependerá del dataset utilizado. En este caso es 38, ya que hay 37 clases en el dataset seleccionado y se debe que incrementar en uno para poder disponer del valor *bias*. La función de activación que se utiliza en esta capa es la de *softmax*.

*Softmax* es una función matemática que convierte un vector numérico en un vector de probabilidades. Normaliza los valores de salida convirtiéndolos en probabilidades donde su sumatorio valga uno. La figura 23 muestra el resultado de esta función matemática.

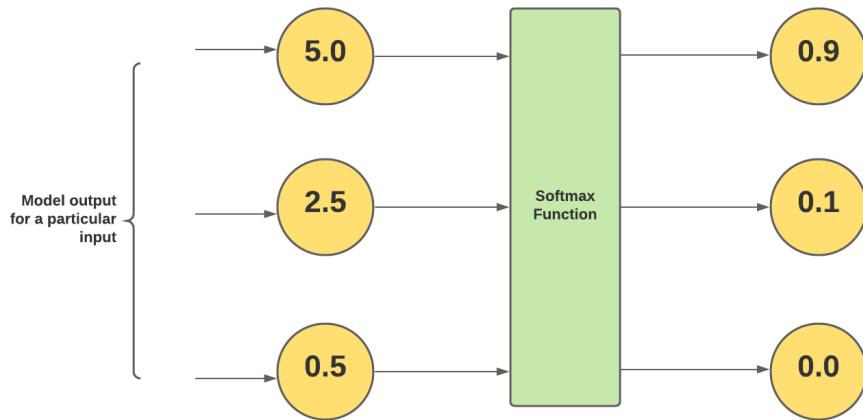


Figura 23: Comportamiento de la función de activación *softmax*.

En la figura 24 se muestran las características del modelo cuando es entrena con el dataset mencionado. Las imágenes a color son transformadas a una escala de gris, con la intención de disminuir el coste de memoria.

Model: "sequential"		
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 100)	2250100
dense_1 (Dense)	(None, 38)	3838
<b>Total params: 2,253,938</b>		
<b>Trainable params: 2,253,938</b>		
<b>Non-trainable params: 0</b>		

Figura 24: Resultado de la ejecución de la función *summary* con la red neuronal artificial.

### 5.2.2. Red neuronal convolucional

Una vez implementada la red neuronal presentada en el apartado 5.2.1, se decidió llevar a cabo la implementación de una red neuronal convolucional, con la intención de poder comparar los resultados obtenidos entre ellas.

Como se muestra en los resultados de los artículo evaluados en el estado del arte del apartado 2.2.3, la convolucional es la que presenta el mejor rendimiento cuando se utilizan datos que no dependen de la temporalidad.

La imagen 25, muestra la estructura de la red neuronal convolucional implementada.

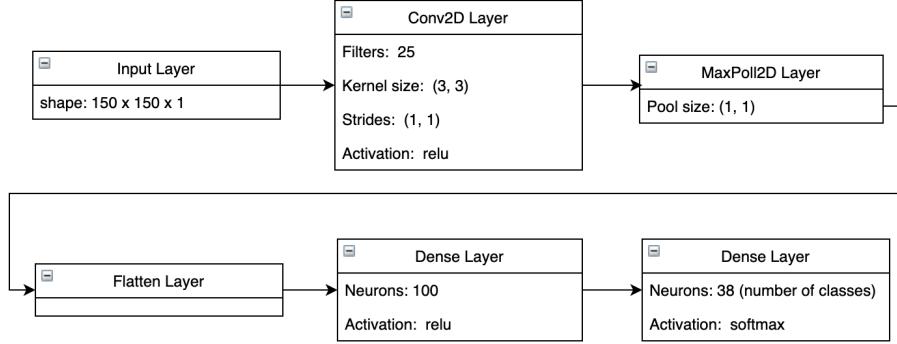


Figura 25: Estructura de la red neuronal convolucional implementada.

Esta estructura dispone de seis capas: una de entrada, cuatro ocultas y una de salida.

La capa de entrada requiere un conjunto de imágenes que dispongan del mismo tamaño, en este caso se han utilizado unas dimensiones de 150x150x1.

La primera capa oculta es la convolucional, la cual dispone de 25 filtros, de un tamaño del kernel de 3x3, un stride de 1x1 y la función de activación ReLu, explicada en el apartado 5.2.1.

La imagen 26 muestra, de una forma visual, el funcionamiento del parámetro kernel. Una red neuronal convolucional es un conjunto de capas que se definen por la acción de varios filtros en la entrada. A esos filtros se les suele llamar kernels. Su tamaño está definido por el anchura y la altura de la máscara de filtros.

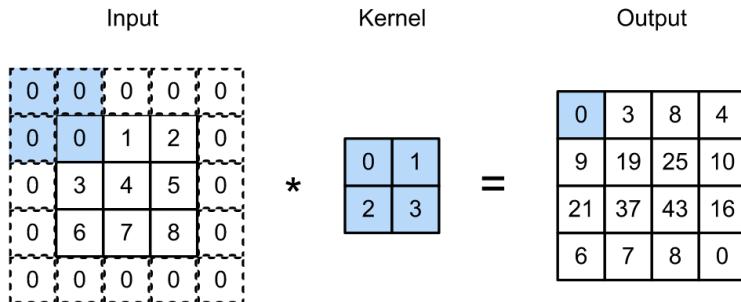


Figura 26: Explicación del parámetro kernel en la capa convolucional de una red neuronal.

La imagen 27 define el significado del valor asignado al stride. En el caso que

se le asigne un valor igual a uno, el filtro se moverá una distancia de un píxel o una unidad a la vez. El tamaño del filtro afecta el volumen de salida, por lo que se suele establecer un valor entero.

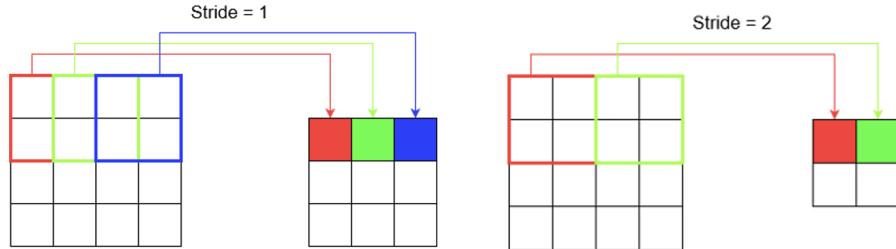


Figura 27: Explicación del parámetro stride en la capa convolucional de una red neuronal.

La segunda capa interna es una capa de *Max Polling*. Esta devuelve el píxel con el máximo valor en el conjunto de píxeles definidos por la máscara del kernel. Esta capa ayuda a prevenir el *overfitting*.

La tercera capa interna recibe como input un mapa de características agrupado. Esta se encargará de transformar la matriz recibida en un columna. Se necesita este capa ya que, a continuación, estos datos serán insertados en una capa densa de una dimensión, como las utilizadas de una red neuronal artificial.

La última capa interna y la capa de salida tienen las mismas características que las dos últimas de la red neuronal artificial definida en el apartado 5.2.1.

La figura que se muestra a continuación contiene el resultado del entrenamiento de la red neuronal convolucional que se acaba de especificar. El dataset utilizado es el que contiene las imágenes a color, pero igual que se ha hecho con la red neuronal artificial, las imágenes han sido procesadas en blanco y negro, para disminuir el coste de memoria.

```

Model: "sequential"

-----  

Layer (type)          Output Shape       Param #  

-----  

conv2d (Conv2D)        (None, 148, 148, 25)    250  

-----  

max_pooling2d (MaxPooling2D) (None, 148, 148, 25)    0  

-----  

flatten (Flatten)      (None, 547600)        0  

-----  

dense (Dense)          (None, 100)           54760100  

-----  

dense_1 (Dense)         (None, 38)            3838  

-----  

Total params: 54,764,188  

Trainable params: 54,764,188  

Non-trainable params: 0

```

Figura 28: Resultado de la ejecución de la función *summary* en la red neuronal convolucional.

### 5.2.3. Resultados de las redes neuronales

Se han implementado tres nuevas estrategias para poder entrenar las redes neuronales especificadas. Estas estrategias permiten guardar los modelos que se utilizan, facilitando la ejecución con diferentes datasets y modelos de redes neuronales, no teniendo que ejecutar de nuevo el procesamiento de las imágenes y el entrenamiento en cada cambio que se realice en el código.

El inconveniente de tener el proyecto separado en estrategias es que se genera un aumento en el coste de memoria, ya que se necesita guardar los cambios realizados en archivos para poder utilizarlos en la siguiente ejecución.

Las estrategias implementadas son:

- **SaveDatabase:** tiene como objetivo procesar el dataset, transformándolo al formato adecuado: cambiando el tamaño y el color de cada imagen. Una vez se han hecho las modificaciones pertinentes, se guarda el modelo en un archivo denominado pickle, con extensión *.pkl*.

De cada dataset se generarán dos pickles, uno con las imágenes de training y otro con las de testing. Con el dataset con el que estamos trabajando actualmente, en la versión de las imágenes a color, el pickle de testing tiene un tamaño de 3 GB y el de training de 6,99 GB, 9,99 GB en total.

En la tabla 2 se puede apreciar la comparación de los tamaños de las dos versiones del dataset cuando se guardan en un directorio, tal y como se descargaron de Kaggle <sup>8</sup>, respecto al tamaño del archivo pickle.

	<b>Tamaño directorio</b>	<b>Tamaño pickle</b>
<b>Dataset en color</b>	106,7 MB	9,99 GB
<b>Dataset preprocesado</b>	81,6 MB	9,99 GB

Tabla 2: Comparativa del tamaño del dataset cuando se guardan las imágenes en un directorio, o bien cuando se utiliza un pickle sin comprimir.

Como se puede apreciar en la tabla 2, el archivo pickle tiene un tamaño mayor que el propio dataset, pero aún así es beneficioso ya que su lectura es mucho más rápida y nos permite no tener que procesar los datos en cada ejecución.

Con tal de poder reducir el coste de memoria, se ha utilizado la librería gzip <sup>9</sup> para poder comprimir los archivos creados. Con este nuevo formato de compresión se ha conseguido reducir el coste de memoria un 22,4 %, con el dataset en color, y un 89,43 %, con el dataset procesado, como se muestra en la tabla 3. Se puede observar como, con este nuevo tipo de archivo creado, el tipo de procesado en las imágenes afecta a su tamaño.

	<b>Tamaño pickle comprimido</b>	<b>Tamaño pickle sin comprimir</b>
<b>Dataset en color</b>	7,75 GB	9,99 GB
<b>Dataset preprocesado</b>	1,06 GB	9,99 GB

Tabla 3: Comparativa del tamaño del dataset cuando se guardan las imágenes en un directorio, o bien cuando se utiliza un pickle comprimido.

En la tabla 4, se muestra el coste temporal de la ejecución de esta estrategia cuando se utilizan los archivos generados en la tabla 2. Se puede apreciar como es un tiempo reducido, no llega a los cinco minutos, y es

---

<sup>8</sup><https://www.kaggle.com/>

<sup>9</sup><https://docs.python.org/3/library/gzip.html>

independiente del tipo de procesado. Su diferencia es del orden de segundos, lo que puede variar por cada ejecución.

Tiempo ejecución	
Dataset en color	00:03:01 horas
Dataset preprocesado	00:03:56 horas

Tabla 4: Comparativa del tiempo de ejecución de la estrategia SaveDatabase utilizando dos modelos con diferente procesado cuando los archivos generados no se comprimen.

En cambio, cuando se utilizan los archivos generados en la tabla 3, el coste temporal incrementa entre un 6,18 % y un 6,84 %.

Tiempo ejecución	
Dataset en color	00:09:35 horas
Dataset preprocesado	00:10:17 horas

Tabla 5: Comparativa del tiempo de ejecución de la estrategia SaveDatabase utilizando dos modelos con diferente procesado cuando los archivos generados son comprimidos.

Dependiendo del recurso más limitado, se escogerá una opción o la otra. En nuestro caso, la memoria es el recurso limitado, por este motivo se comprimirán los archivos asumiendo el coste adicional de tiempo.

- **TrainNeuralNetwork:** esta estrategia tiene como objetivo entrenar la red neuronal y guardar el modelo en un archivo con extensión *h5*. Para poder ejecutarla se necesita especificar el tipo de red neuronal que se quiere utilizar. Actualmente hay dos posibilidades: la red neuronal artificial y la convolucional. Otro parámetro que también se debe especificar es el pickle con el que se entrenará la red neuronal.

El modelo se guardará en el archivo *h5* con la intención de poder utilizarlo en otras estrategias, como la que se explicará a continuación, sin tener que volver a entrenarla para poder trabajar con este modelo.

El tamaño de este archivo depende del tipo de red neuronal utilizada, en este caso el modelo de la artificial tiene un tamaño de 27,1 MB, y la convolucional de 657,2 MB, independientemente del pickle utilizado. Esta información se muestra de una forma más gráfica en la tabla 6:

	Tamaño modelo red neuronal artificial	Tamaño modelo red neuronal convolucional
<b>Dataset en color</b>	27,1 MB	657,2 MB
<b>Dataset preprocesado</b>	27,1 MB	657,2 MB

Tabla 6: Tamaño del archivo *h5* por los dos tipos de redes neuronales con los dos datasets que se están utilizando actualmente.

El tiempo de ejecución de esta estrategia varía mucho dependiendo del tipo de red neuronal que se esté entrenando y del pickle seleccionado. La red neuronal convolucional puede llegar a tardar alrededor de una hora más que la artificial, como se muestra en la tabla 7.

	Tiempo entreno red neuronal artificial	Tiempo entreno red neuronal convolucional
<b>Dataset en color</b>	00:05:54 horas	00:55:45 horas
<b>Dataset preprocesado</b>	00:03:29 horas	01:00:16 hora

Tabla 7: Tiempo de ejecución de la estrategia de *TrainingNeuralNetwork* dependiendo del dataset y del tipo de red neuronal.

- **AccuracyNeuralNetwork:** Una vez entrenada la red neuronal, se habrá creado un archivo *h5* donde se habrá guardado el modelo entrenado. Esta estrategia procesa el modelo guardado e intenta hacer las predicciones con la parte del dataset definida para el testing, una vez tiene el resultado lo analiza para poder comprobar el número de aciertos. El resultado se mostrará por el terminal.

En la tabla 8 se muestra una comparativa del *accuracy* entre las diferentes versiones del dataset: la original y la preprocesada con las dos redes neuronales implementadas.

	<i>Accuracy</i> red neuronal artificial	<i>Accuracy</i> red neuronal convulacional
<b>Dataset en color</b>	72.27 %	34.46 %
<b>Dataset preprocesado</b>	100 %	99.96 %

Tabla 8: Comparativa del *accuracy* de los dos tipos de redes neuronales dado dos modelos con diferente procesado.

Como se aprecia en la tabla 8, la red neuronal artificial tiene una *accuracy* mas elevada en los dos Dataset, con lo que se han deducido dos posibilidades:

- No se están introduciendo suficientes datos en la red neuronal convolucional para poder hacer un buen entrenamiento, ya que este tipo de redes necesitan muchos datos para poder aprovechar todo su potencial.
- La configuración con los que se han implementado la red neuronal convolucional no son las más adecuadas para este dataset.

En cuanto al tiempo de ejecución de esta estrategia, si que se encuentra diferencia entre el tipo de red neuronal, pero su duración no llega al minuto, como se muestra la tabla 9.

	<b>Tiempo accuracy red neuronal artificial</b>	<b>Tiempo accuracy red neuronal convolucional</b>
<b>Dataset en color</b>	11 segundos	48 segundos
<b>Dataset preprocesado</b>	15 segundos	43 segundos

Tabla 9: Tiempo de ejecución de la estrategia de *AccuracyNeuralNetwork* dependiendo del dataset y del tipo de red neuronal.

En futuros experimentos se comprobará si las suposiciones que se han mencionado son acertadas o no, con la intención de mejorar la *accuracy* de las dos redes neuronales.

Una vez analizadas cada una de estas estrategias, podemos comparar el coste global, tanto de memoria como de tiempo. En la tabla 10 se muestra la suma de los costes de memoria que ha habido en las tres estrategias, dependiendo del dataset y del tipo de red neuronal.

	<b>Coste memoria red neuronal artificial</b>	<b>Coste memoria red neuronal convolucional</b>
<b>Dataset en color</b>	7,78 GB	8,41 GB
<b>Dataset preprocesado</b>	1,09 GB	1,72 GB

Tabla 10: Comparativa del coste de memoria en las tres estrategias entre los dos tipos de redes neuronales, dado dos modelos con diferente procesado.

El coste de memoria ha sido calculado a partir de la suma de los archivos que se generan en cada estrategia. En concreto, la suma del tamaño del pickle y del archivo que contiene el modelo entrenado de la red neuronal. Como conclusión, se puede observar como el dataset es el que supone la mayor diferenciación en el coste de memoria.

En la tabla 11 se muestra el coste de tiempo de las tres estrategias, sumando el tiempo de ejecución de cada una de ellas, dependiendo del dataset y de la red neuronal utilizada.

	<b>Coste tiempo red neuronal artificial</b>	<b>Coste tiempo red neuronal convolucional</b>
<b>Dataset en color</b>	00:15:40 horas	01:05:55 horas
<b>Dataset preprocesado</b>	00:14:14 horas	01:11:16 hora

Tabla 11: Comparativa del coste de tiempo en las tres estrategias entre los dos tipos de redes neuronales, dado dos modelos con diferente procesado.

En la tabla 11 se puede observar como hay una diferencia abismal entre el coste del tiempo de ejecución de la red neuronal artificial a la convolucional, prácticamente la diferencia es de una hora. Con el coste de tiempo es indiferente el dataset utilizado, como mucho hay cinco minutos de diferencia.

En cuanto a la conclusión de la implementación de las redes neuronales, se ha observado, en la tabla 8, como la red neuronal artificial ha dado mejores resultados que la convolucional, y como el procesado del dataset ha significado un mejora inimaginable en los dos tipos de redes. Esta información se tendrá en cuenta cuando se vaya a utilizar un dataset diferente al actual.

Durante la explicación de cada estrategia se han contemplado diferentes mejoras que se llevarán a cabo en futuras experimentaciones, con la intención de poder obtener el mejor resultado con el mínimo coste posible de las redes neuronales.

### 5.3. Experimentación 3: Uso de árboles de decisión

Una vez se han obtenido resultados y conclusiones de las redes neuronales, se ha decidido experimentar con otro tipo de estructuras, como son los árboles de decisión.

Se ha implementado un árbol de decisión que utiliza el algoritmo *Extreme Gradient Boosting*. El motivo de la utilización del algoritmo es que este ayuda a aumentar la velocidad y el rendimiento del modelo de clasificación, como se especifica en el apartado de fundamentos 3.2.2.

Se ha utilizado la librería xgboost para poder llevar a cabo su implementación, de donde se ha utilizado la clase XGBClassifier. La imagen 29 muestra como se ha hecho uso de esta librería para entrenar el árbol de decisión:

```
xgboost_model = XGBClassifier()
xgboost_model.fit(x_train, y_train, verbose=True, eval_set=[(x_test, y_test)])
```

Figura 29: Uso de la librería xgboost en la implementación del boosted decision tree.

Se ha entrenado el árbol de decisión con los dos mismos datasets con los que se han entrenado las redes neuronales, con la intención de poder comparar la *accuracy* final entre las dos estructuras.

Para poder llevar a cabo este entrenamiento se han implementado dos nuevas estrategias, con el mismo objetivo que con las redes neuronales: poder separar la ejecución del entrenamiento de la ejecución de la predicción, con tal de no tener que entrenarlo cada vez que se quiera hacer una predicción. Las estrategias creadas son:

- **TrainDecisionTree:** Esta estrategia tiene el objetivo de entrenar el árbol de decisión dado un pickle que contendrá el dataset con el que se quiere trabajar.

Una vez entrenado, se guardará el modelo en un archivo con extensión *.pickle.dat*, el cual nos permitirá recuperar dicho modelo para utilizarlo en otras estrategias.

En la tabla 12 se muestra el coste de memoria que genera el archivo donde se guarda el modelo entrenado del árbol de decisión. Se puede apreciar como el modelo del *boosted decision tree* entrenado con el dataset preprocesado ocupa el doble que el entrenado con el dataset a color transformado

a escalas de grises.

Tamaño del modelo del árbol de decisión	
Dataset en color	1,1 MB
Dataset preprocesado	2,2 MB

Tabla 12: Comparativa del coste de memoria en la estrategia *TrainDecisionTree*, dado dos modelos con diferente procesado.

Para poder evaluar la eficiencia de este modelo predictivo también se debe tener en cuenta el coste de tiempo en su entrenamiento. Como podemos ver en la tabla 13, el tiempo de ejecución mínimo de esta estrategia es de dos horas, una hora más que con la red neuronal, como se comparará más adelante. También podemos apreciar que hay una diferencia de una hora y media entre el tiempo de entreno de ambos datasets analizados.

Tiempo ejecución del entrena- miento del árbol de decisión	
Dataset en color	03:35:45 horas
Dataset preprocesado	02:07:53 horas

Tabla 13: Comparativa del coste de tiempo en la estrategia *TrainDecisionTree*, dado dos modelos con diferente procesado.

- **AccuracyDecisionTree:** La segunda estrategia que se ha implementado para poder utilizar el aprendizaje basado en árboles de decisión tiene como objetivo mostrar el *accuracy* del modelo entrenado. Se especificará el modelo que se quiere utilizar, el cual será el archivo con extensión *.pickle.dat* que esta estrategia procesará. Una vez se haya leído el modelo se hará una predicción con todos los datos de testigo del dataset seleccionado, el cual deberá ser el mismo con el que se hizo el entrenamiento. El tanto por ciento de aciertos se mostrarán por el terminal, informando así del *accuracy* de modelo.

En la tabla 14 se muestran los resultados de esta estrategia, comparando la *accuracy* de los dos modelos entrenados con los dos datasets con los que se han hecho todos los experimentos hasta el momento.

<i>Accuracy</i> del árbol de decisión	
<b>Dataset en color</b>	100 %
<b>Dataset preprocesado</b>	100 %

Tabla 14: Comparativa de la *accuracy* del árbol de decisión, dado dos modelos con diferente procesado.

Como se muestra en la tabla 14, se obtiene una *accuracy* máxima con ambos datasets, lo que muestra que con este tipo de aprendizaje se llegan a obtener resultados muy óptimos para las condiciones del problema actual que se intenta resolver.

En cuanto a los costes de esta estrategia, nos centraremos en el coste de tiempo ya que no se genera ningún archivo adicional que pudiese incrementar el coste de memoria, el único coste significativo que hay es el propio coste de ejecución.

Se ha ejecutado esta estrategia dos veces, una por cada dataset con el que estamos trabajando, la duración de cada ejecución se muestra en la tabla 15:

Tiempo ejecución del <i>accuracy</i> del árbol de decisión	
<b>Dataset en color</b>	00:00:23 horas
<b>Dataset preprocesado</b>	00:00:20 horas

Tabla 15: Comparativa del coste de tiempo en la estrategia *AccuracyDecisionTree*, dado dos modelos con diferente procesado.

Se puede apreciar como el tiempo de ejecución de esta estrategia es mínimo con ambos datasets. Se han obtenido muy buenos resultados con este método de aprendizaje: una *accuracy* del 100 % con un tiempo de predicción de unos 20 segundos.

Una vez analizadas las dos estrategias, podemos comparar el coste global de tiempo, ya que el coste de memoria únicamente se ha mencionado el de la primera estrategia, por lo tanto el coste global de memoria será el mismo que el de la estrategia de *TrainDecisionTree*.

En la tabla 16 se muestra la suma de los costes de tiempo que ha habido en las dos estrategias, dependiendo del dataset utilizado.

	<b>Coste tiempo del árbol de decisión</b>
<b>Dataset en color</b>	03:42:42 horas
<b>Dataset preprocesado</b>	02:14:24 horas

Tabla 16: Comparativa del coste de tiempo en el uso del árbol de decisión como modelo de predicción dado dos modelos con diferente procesado.

El coste de tiempo ha sido calculado a partir de la suma de los tiempos de ejecución de ambas estrategias junto a los valores obtenidos en la tabla 5. El alto valor de este coste es debido a la estrategia de entrenamiento, que es la causante de este incremento, ya que la predicción tenía un coste del orden de segundos. Esto demuestra que el entrenamiento de un árbol de decisión es mucho más costoso que su predicción, deberemos comprobar si, para el problema que queremos resolver en este proyecto, nos puede afectar este coste tan elevado o si es despreciable.

## 5.4. Comparativa de las experimentaciones 2 y 3

Una vez explicada la experimentación del árbol de decisión, se puede hacer la comparativa entre los resultados de este y los de las redes neuronales.

Para poder realizar una comparativa adecuada se deberán utilizar los parámetros que se han ido estudiando durante las experimentaciones: la *accuracy* y los costes, tanto de tiempo como de memoria.

En la tabla 17 se puede apreciar la comparación de la *accuracy* entre las tres estructuras implementadas: la red neuronal artificial, la red neuronal convolucional y el árbol de decisión. Se han utilizado los dos datasets con los que se ha estado trabajando en ambos experimentos.

	<i>Accuracy</i> red neuronal artificial	<i>Accuracy</i> red neuronal convolucional	<i>Accuracy</i> árbol de decisión
<b>Dataset en color</b>	72,27 %	34,46 %	100 %
<b>Dataset preprocesado</b>	100 %	99,96 %	100 %

Tabla 17: Comparativa de la *accuracy* entre las estructuras implementadas dado dos modelos con diferente procesado.

Podemos apreciar cómo al árbol de decisión es el que tiene una mejor *accuracy*, ya que su valor es del 100 % con ambos datasets. Otra conclusión que podemos obtener es que todas las estructuras tienen un precisión muy elevada cuando se utiliza el dataset específicamente procesado para favorecer el entrenamiento.

En cuanto al coste de tiempo, compararemos la duración de la ejecución del entrenamiento y de la predicción por separado. En la tabla 18 se pueden ver la duración del entrenamiento en la red neuronal artificial, la convolucional y el árbol de decisión, estos valores se han obtenido de las tablas: 7 y 13.

	Duración red neuronal artificial	Duración red neuronal convolucional	Duración árbol de decisión
<b>Dataset en color</b>	00:05:54 horas	00:55:45 horas	03:35:45 horas
<b>Dataset preprocesado</b>	00:03:29 horas	01:00:16 hora	02:07:53 horas

Tabla 18: Comparativa de la duración del entrenamiento entre las estructuras implementadas dado dos modelos con diferente procesado.

En la tabla 19, se muestra una comparativa de la duración de las predicciones entre la red neuronal artificial, la convolucional y el árbol de decisión. Estos valores han sido obtenidos de las tablas: 9 y 15.

	Duración red neuronal artificial	Duración red neuronal convolucional	Duración árbol de decisión
<b>Dataset en color</b>	00:00:11 horas	00:00:48 horas	00:00:23 horas
<b>Dataset preprocesado</b>	00:00:15 horas	00:00:43 horas	00:00:20 horas

Tabla 19: Comparativa de la duración de la predicción entre las estructuras implementadas dado dos modelos con diferente procesado.

Comparando los valores mostrados en las dos tablas, podemos observar como el árbol de decisión es el que más se deriva, ya que su duración de entrenamiento llega prácticamente a triplicar el de las redes neuronales. En cuanto a la duración de las predicciones, todos tienen una duración del orden de segundos, siendo la red neuronal convolucional la que mas se acerca al minuto.

Podemos deducir entonces, que el árbol de decisión es el que nos proporciona una mejor *accuracy*, pero nos genera un mayor coste. Puede que el motivo por el cual el coste de tiempo de entrenamiento del árbol de decisión sea mucho más elevado que el de las redes neuronales sea porque este no se ha podido ejecutar en *Jupyter Notebook*. Al intentarlo, la ejecución se detenía en mitad del entrenamiento, por motivos desconocidos, por lo que se ha tenido que realizar el entrenamiento en el ordenador personal. Aún así, en este mismo ordenador también se han entrenado las redes neuronales sin haber muchas diferencia en la duración de la ejecución.

En futuras experimentaciones se intentaran mejorar los valores de la red neuronal convolucional, ya que hay diversos parámetros que pueden incrementar su *accuracy* y que podrían obtenerse con un coste menor al árbol de decisión.

## 5.5. Experimentación 4: Incremento dataset

El dataset que se ha utilizado durante las anteriores experimentaciones disponía de 27.750 muestras, 55.500 en total con los dos datasets con diferente procesado. Con este número de muestras se puede obtener muy buenos resultados, pero aún serían mejores si se utilizaran dos datasets diferentes, ya que el ángulo, el color y la forma de la imagen abarcarían mas posibilidades. Conseguiríamos así una accuracy mejor para cuando se hagan predicciones con imágenes que no pertenezcan al testing del mismo dataset.

Se han evaluado varios datasets, buscando uno que contenga los mismo gestos que el actual pero con un enfoque diferente en los datos. Los enlaces que se muestran a continuación contienen los datasets que se han estado evaluando. Como se puede apreciar, todos pertenecen a la comunidad de datos Kaggle <sup>10</sup>.

- <https://www.kaggle.com/kareemalaa74/dataset-asl-test-and-train>
- <https://www.kaggle.com/grassknotted/asl-alphabet>
- <https://www.kaggle.com/lucasvieirademiranda/aslalphabet>
- <https://www.kaggle.com/ammarnassanalhajali/american-sign-language-letters>

Finalmente se ha optado por escoger el primer enlace mostrado:

<https://www.kaggle.com/kareemalaa74/dataset-asl-test-and-train>

Los motivos de esta elección han sido:

- Dispone de los gestos que hacen referencia a los números. La mayoría de los otros datasets únicamente contenían letras, con lo que se debería de haber utilizado dos dataset, uno para las letras y otro para los números.
- Las imágenes de este dataset han sido tomadas desde diferentes ángulos que las del actual dataset.
- Las imágenes no se han tomado con un fondo de un mismo color, como pasa con el dataset actual, sino que se han realizado con diferentes fondos.

El nuevo dataset tiene como nombre: *Dataset ASL test and train* y se ha obtenido de la comunidad de datos Kaggle. Dispone de los mismos gestos que el dataset *Sign Language Gesture Images Dataset* [9], pero este, adicionalmente, contiene el número 10. Cada gesto dispone de 300 imágenes, 301 con la imagen

---

<sup>10</sup><https://www.kaggle.com>

de test, con unas dimensiones de  $1500 \times 1500$  píxeles.

A continuación, se muestra una comparativa del dataset que se ha estado utilizando, con el nuevo que se va a añadir al actual. Las imágenes 30 y 31 hacen referencia al dataset actual. Son las mismas imágenes que se mostraron y explicaron en la experimentación de la búsqueda del dataset, en la sección 5.1.



Figura 30: Imágenes obtenidas del dataset *Sign Language Gesture Images Dataset* [9] en color.



Figura 31: Imágenes obtenidas del dataset *Sign Language Gesture Images Dataset* [9] preprocesadas con la intención de mejorar el entrenamiento de las redes neuronales.

La imagen 32 muestra diez gestos del dataset *Dataset ASL test and train* [10]. Hace referencia a los mismos gestos que las imágenes 30 y 31, se han situado en el mismo orden para facilitar su comparación. Se puede apreciar como cumple las características explicadas anteriormente del porque se ha seleccionado este dataset.



Figura 32: Imágenes obtenidas del dataset *ASL test and train* [10].

Se ha realizado dos experimentaciones con este dataset: usándolo individualmente, y añadiendo las muestras del dataset *Sign Language Gesture Images* [9]. La intención de esto es poder saber la accuracy y los costes de ambos datasets por separado, y poder compararlo con los valores que nos dan estas características cuando se utilizan a la par.

Para poder realizar la segunda experimentación, se han tenido que implementar unos cambios en el proyecto, ya que hasta ahora no se podía seleccionar más de un dataset. Se ha pensado la mejor solución para este problema, y se han acabado encontrado dos posibles implementaciones:

- Permitir seleccionar más de un dataset en la creación de un pickle, en la estrategia *SaveDatabase*, con lo que en las otras estrategias no sería necesario realizar ningún cambio, ya que seguiría utilizando únicamente un pickle.
- En un pickle únicamente se puede almacenar un dataset, no teniendo que cambiar esta estrategia. El cambio se debería realizar en las otras estrategias, ya que estas deberían permitir seleccionar tantos pickles como el usuario desee.

Finalmente se optó por implementar la segunda opción. A pesar de tener que realizar más cambios en el proyecto actual, ya que prácticamente se debía cambiar la estructura de todas las estrategias, se pensó que, en cuanto a coste de memoria, sería mejor tener un único pickle por dataset, que no tener diferentes pickles con variaciones del mismo dataset. Como se ha comentado anteriormente en la tabla 3, un pickle tiene un tamaño del orden de gigabytes.

Una vez realizados los cambios convenientes para poder utilizar más de un pickle en una estrategia, se ejecutó el proyecto para poder averiguar la *accuracy* de este nuevo dataset.

A continuación, se compararan los resultados y el coste de las ejecuciones de las tres siguientes combinaciones de datasets:

- Con el dataset *Sign Language Gesture Images* [9], que hace referencia a la imagen 30.
- Con el nuevo dataset, *ASL test and train* [10], mostrado en la imagen 32.
- Combinando ambos datasets.

Los tres tablas, que hay a continuación, referenciadas con los números: 20, 21 y 21, muestran la comparativa de las tres combinaciones de datasets que se acaban de explicar. La tabla 20 contiene la información de la accuracy de la red neuronal artificial, la convolucional y el árbol de decisión, estas tres estructuras serán las que se van a ir comparando con las combinaciones de datasets.

	Dataset <i>Sign Language Gesture Images</i>	Dataset <i>ASL test and train</i>	Combinación de ambos datasets
<i>Accuracy</i> red neuronal artificial	72,27 %	33,85 %	2,88 %
<i>Accuracy</i> red neuronal convolucional	34,46 %	98,95 %	99,30 %
<i>Accuracy</i> árbol de decisión	100 %	100 %	—

Tabla 20: Comparativa de la *accuracy* de las tres estructuras implementadas dado dos datasets y una combinación de ambos.

La tabla 21 compara el tiempo de ejecución de los entrenamientos entre las diferentes estructuras. El formato de la tabla es el mismo que el anterior, facilitando así la lectura de la información de ambas tablas y pudiendo compararlas a la vez.

	<b>Dataset Sign Language Gesture Images</b>	<b>Dataset ASL test and train</b>	<b>Combinación de ambos datasets</b>
Tiempo entreno red neuronal artificial	00:05:54 horas	00:01:01 horas	00:12:09 horas
Tiempo entreno red neuronal convolucional	00:55:45 horas	00:27:51 horas	01:35:02 horas
Tiempo entreno árbol de decisión	03:35:45 horas	03:38:52 horas	–

Tabla 21: Comparativa del tiempo de entrenamiento de las tres estructuras implementadas dado dos datasets y una combinación de ambos.

Por último, la tabla 22, muestra la comparativa del tiempo de ejecución de las predicciones.

	<b>Dataset Sign Language Gesture Images</b>	<b>Dataset ASL test and train</b>	<b>Combinación de ambos datasets</b>
Tiempo predicción red neuronal artificial	00:00:11 horas	00:00:04 horas	00:00:41 horas
Tiempo predicción red neuronal convolucional	00:00:48 horas	00:00:24 horas	00:01:43 horas
Tiempo predicción árbol de decisión	00:00:23 horas	00:00:10 horas	–

Tabla 22: Comparativa del tiempo de predicción de las tres estructuras implementadas dado dos datasets y una combinación de ambos.

Como se ha podido contemplar, en la tabla no se muestran los resultados de las combinaciones de ambos datasets utilizando el árbol de decisión como modelo de predicción. El motivo es que no se pudo acabar de realizar la ejecución de su entrenamiento, después de 24h únicamente se había completado el 30 %. Por lo tanto, si no hubiese ningún problema durante la ejecución, como podría ser un coste de memoria superior al que el dispositivo podría soportar, el tiempo total hubiese sido de 80 horas, 3,33 días. Se paró la ejecución en el 30 % ya que se dedujo que este tiempo de entrenamiento era inviable bajo las características en las que se está realizando este proyecto.

Hasta el momento, el árbol de decisión era una solución que nos estaba aportando un *accuracy* muy elevada, a pesar de tener el tiempo de entrenamiento más alto. Pero viendo que no es capaz de proceder con el entrenamiento de ambos datasets, cuando tenemos otras estructuras implementadas que también dan muy buenos resultados con un tiempo de entrenamiento asumible, se ha decidido descartar este modelo de predicción.

En cuanto a la red neuronal artificial, podemos apreciar como esta dispone del tiempo de ejecución más reducido en, prácticamente, todas las ejecuciones. Aún así, en la tabla 20 se muestra como disminuye la *accuracy* hasta el 2,88 % cuando se utilizan ambos datasets. Se realizaron varias ejecuciones para poder asegurar que este valor no variaba, pero la *accuracy* de todas estas ejecuciones no varió, la máxima desviación fue de una unidad.

Con estos resultados podemos verificar que la red neuronal artificial no es un modelo fiable en este proyecto, por lo tanto, no se puede utilizar como solución a este problema.

Por último, la red neuronal convolucional dio una *accuracy* del 99,30 %, un valor muy elevado y totalmente valido para la solución. Su tiempo de ejecución en el entrenamiento es mas elevado que el de la red neuronal artificial, pero es un tiempo asequible. Se concluye que este modelo es el óptimo de los implementado hasta ahora.

## 5.6. Experimentación 5: Mejorar rendimiento de la CNN

Como se ha demostrado en la sección 5.5, la estructura implementada con la que se han obtenido los mejores resultados, tanto de rendimiento como de costes, ha sido la red neuronal convolucional.

Todos los experimentos que se hagan a partir de ahora se realizarán con la unión de los datasets y con la red neuronal convolucional.

Aún haber obtenido muy buenos resultados con esta red neuronal, en esta sección se intentará acotar sus parámetros para poder obtener el mayor rendimiento con las mínimas pérdidas. Se ha utilizado la librería Scikit-learn<sup>11</sup> para poder implementar la estrategia que realizará la comparativa de parámetros. La nueva estrategia implementada se denomina: *Hyperparameter Optimization Strategy*

Se ha seguido el artículo de *Grid search hyperparameters deep learning models python keras* [49] para la implementación de este experimento. Se ha utilizado como guía y modelo, pero se ha modificado y adaptado a las necesidades del actual problema.

Se comenzó haciendo las pruebas con todos los datos de ambos datasets, se estaba haciendo una estimación del valor que deberían asignarse al tamaño del *Batch* y el número de *Epochs*. Ejecutando el training de la red neuronal convolucional con estos parámetros variables, se comprobó que no era posible realizar la ejecución con todos los datos. Después de cuatro días ejecutándose, cuando se calculaba que aún faltaba el doble de tiempo, como mínimo, para finalizar su ejecución, el dispositivo se quedó sin espacio de memoria y tuvo que interrumpirla.

Como solución, se implementó una modificación en el proyecto. En lugar de leer todos los datos de los *Pickles*, archivos donde se almacenan las imágenes procesadas, únicamente se obtienen 20 imágenes de cada señal. Consiguiendo así disminuir el tiempo de entrenamiento de la red neuronal. Los cambios en el código han sido notorios, ya que se ha tenido que implementar una nueva funcionalidad donde se pudiesen obtener el mismo número de imágenes por cada señal, la reducción del dataset debía continuar siendo óptima.

La estrategia *Hyperparameter Optimization Strategy* consta de siete atributos a optimizar, el cual se especifica en los parámetros de la ejecución de esta. A continuación se explicará cada atributo y los resultados obtenidos de este.

---

<sup>11</sup><https://scikit-learn.org/stable/>

### 5.6.1. Tamaño del *Batch* y número de *Epochs*

Para la realización de este entrenamiento, se ha utilizado el mismo modelo que el implementado en la red neuronal convolucional, ya que es el que se busca mejorar. En esta ejecución se optimizan los siguientes atributos:

- El tamaño del *Batch*, el cual define la cantidad de datos con los que trabajar.
- El número de *Epochs* que define el número de veces que el algoritmo de aprendizaje se ejecutará por cada conjunto de datos de entrenamiento.

El parámetro que se debe introducir para poder ejecutar la estrategia optimizando estos atributos es “*batch\_and\_epoch*”.

Los parámetros utilizados para el clasificador se muestran en la imagen 33.

```
@staticmethod
def __get_parameters_for_batch_epochs(n_classes, image_size):
    batch_size = [10, 20, 40, 60, 80, 100]
    epochs = [10, 50, 100]

    param_grid = dict(batch_size=batch_size, epochs=epochs, num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_batch_epochs, verbose=2)

    return param_grid, classifier
```

Figura 33: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *batch and epoch*

El tamaño del *Batch* se ha evaluado con los valores: 10, 20, 40, 60, 80 y 100. En cambio, el parámetro del número de *Epochs* es menos variado, ya que oscila entre los valores: 10, 50 y 100.

La imagen 34 muestra los resultados de esta ejecución. La mejor combinación de parámetros es de un tamaño del *Batch* y un número de *Epochs* igual a 10.

```

Best: 1.000000 using {'batch_size': 10, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 10, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 10, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 10, 'epochs': 100}
1.000000 (0.000000) with: {'batch_size': 20, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 20, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 20, 'epochs': 100}
1.000000 (0.000000) with: {'batch_size': 40, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 40, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 40, 'epochs': 100}
1.000000 (0.000000) with: {'batch_size': 60, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 60, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 60, 'epochs': 100}
1.000000 (0.000000) with: {'batch_size': 80, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 80, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 80, 'epochs': 100}
1.000000 (0.000000) with: {'batch_size': 100, 'epochs': 10}
1.000000 (0.000000) with: {'batch_size': 100, 'epochs': 50}
1.000000 (0.000000) with: {'batch_size': 100, 'epochs': 100}
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 02:55:57 [hh:mm:ss]

```

Figura 34: Resultado de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *batch and epoch*

Podemos apreciar como todos las combinaciones tienen el mismo resultado, un *accuracy* del 100 % sin ninguna pérdida. Se utilizará el resultado que se muestra como el mejor, con un valor de 10 en ambos atributos. Aún así, teniendo como objetivo únicamente la *accuracy*, se podría utilizar cualquiera de las combinaciones probadas y se obtendría el mismo resultado.

Esta ejecución ha tenido un coste de tiempo de 2 horas, 55 minutos y 57 segundos. Es un tiempo muy reducido comparado con el de la primera ejecución, cuando se utilizó todo el dataset. Como se ha comentado anteriormente, este se estuvo ejecutando durante cuatro días cuando se paró su ejecución sin haber llegado a la mitad de esta. Con las modificaciones realizadas en el proyecto y estimando el tiempo total de ejecución con el dataset completo, se ha conseguido reducir un 98,44 % el coste del tiempo de ejecución.

Los valores obtenidos en este experimento se utilizarán en las optimizaciones de los siguientes atributos.

### 5.6.2. Algoritmos de optimización

Una vez obtenidos los mejores valores para los atributos del tamaño del *Batch* y el número de *Epochs*, se ha querido evaluar cual sería el algoritmo de entrenamientos más óptimo.

Para poder implementar esta funcionalidad se ha utilizado la estrategia, comentada anteriormente, *Hyperparameter Optimization Strategy*. Para poder ejecutar esta optimización se debe introducir el parámetro *optimization algorithms*.

En la imagen número 35, se muestran los parámetros introducidos en este clasificador. Los algoritmos con los que se han hecho las comparaciones son: *SGD*, *RMSprop*, *Adagrad*, *Adadelta*, *Adam*, *Adamax* y *Nadam*.

```
@staticmethod
def __get_parameters_for_optimization_algorithm(n_classes, image_size):
    optimizer = ['SGD', 'RMSprop', 'Adagrad', 'Adadelta', 'Adam', 'Adamax', 'Nadam']

    param_grid = dict(optimizer=optimizer, num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_optimizer_algorithm, epochs=EPOCHS, batch_size=BATCH_SIZE,
                                 verbose=2)

    return param_grid, classifier
```

Figura 35: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *optimization algorithms*.

Como se muestra en la imagen anterior, los valores del tamaño del *Batch* y el número de *Epochs* son constantes, ya que adquieren el valor del mejor resultado obtenido en la ejecución del apartado 5.6.1. Estos resultados se guardan en un archivo con extensión *.py*, donde actualmente se encuentran los valores de *EPOCHS* y de *BATCH\_SIZE*.

En la siguiente imagen, se muestran los resultados de la ejecución de esta estrategia con el clasificador y los parámetros mostrados en la figura 35.

```
In [15]: python3 Src/main.py --showOptimizedHyperparameter optimization_algorithms asl_alphabet_gray_150x150px sign_gesture_gray
38/38 - 7s - loss: 2.1799e-04 - accuracy: 1.0000
Epoch 7/10
38/38 - 7s - loss: 1.8106e-04 - accuracy: 1.0000
Epoch 8/10
38/38 - 7s - loss: 1.5470e-04 - accuracy: 1.0000
Epoch 9/10
38/38 - 7s - loss: 1.3471e-04 - accuracy: 1.0000
Epoch 10/10
38/38 - 7s - loss: 1.1921e-04 - accuracy: 1.0000
Best: 1.000000 using {'optimizer': 'SGD'}
1.000000 (0.000000) with: {'optimizer': 'SGD'}
1.000000 (0.000000) with: {'optimizer': 'RMSprop'}
1.000000 (0.000000) with: {'optimizer': 'Adagrad'}
1.000000 (0.000000) with: {'optimizer': 'Adadelta'}
1.000000 (0.000000) with: {'optimizer': 'Adam'}
1.000000 (0.000000) with: {'optimizer': 'Adamax'}
1.000000 (0.000000) with: {'optimizer': 'Nadam'}
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 00:33:55 [hh:mm:ss]
```

Figura 36: Resultados obtenidos en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *optimization algorithms*.

El algoritmo que ha obtenido el mejor resultado es el *SGD*, *Stochastic gradient descent*, con un *accuracy* del 100 %. Aún así, si observamos los resultados obtenidos con los otros algoritmos veremos que su resultado no varía, con obtenemos la máxima *accuracy*.

La duración de esta ejecución ha sido de 33 minutos y 55 segundos, dos horas y media menos que la anterior ejecución. El motivo de la reducción de este coste de tiempos es que únicamente se debían probar siete combinaciones, una por cada algoritmo.

En las siguientes ejecuciones utilizaremos el algoritmo que se muestra como el mejor, pero en caso de querer repetir este experimento sería interesante poder variar el algoritmo para comprobar otros parámetros de su ejecución, como el coste de tiempo, por ejemplo.

### 5.6.3. Índice de aprendizaje e impulso

En el apartado anterior hemos deducido que el algoritmo de entrenamiento que se utilizará será el *Stochastic gradient descent*, el cual tiene dos parámetros a configurar:

- El índice de aprendizaje (*learn rate*): define la velocidad a la que el modelo se adapta al problema. Cuanto menor sea su valor, mayor deberá ser el tamaño del *Batch* de entrenamiento. Su valor oscila entre los rangos de 0.0 y 1.0

- El impulso (*momentum*): permite acelerar al gradiente en la dirección correcta. En lugar de utilizar solo el gradiente del paso actual, el *momentum* acumula el gradiente de los pasos anteriores para determinar la dirección a seguir.

El parámetro que se debe introducir para poder ejecutar la estrategia optimizando estos atributos es *learn\_rate\_and\_momentum*.

En la imagen que hay a continuación se muestran los parámetros introducidos en el clasificador. El índice de aprendizaje es evaluado a partir de los valores: 0.001, 0.01, 0.1, 0.2 y 0.3. En cambio, el *momentum* puede adquirir uno de los siguientes valores: 0.0, 0.2, 0.4, 0.6, 0.8 o 0.9.

```
@staticmethod
def __get_parameters_for_learn_rate_and_momentum(n_classes, image_size):
    learn_rate = [0.001, 0.01, 0.1, 0.2, 0.3]
    momentum = [0.0, 0.2, 0.4, 0.6, 0.8, 0.9]

    param_grid = dict(learn_rate=learn_rate, momentum=momentum, num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_learn_rate_and_momentum, epochs=EPOCHS,
                                 batch_size=BATCH_SIZE, verbose=2)

    return param_grid, classifier
```

Figura 37: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *learn rate and momentum*.

En la figura 38 se muestra el resultado de esta ejecución, donde podemos apreciar como los mejores parámetros son la combinación de un índice de aprendizaje del 0.001 y de un *momentum* de 0.

```
[INFO]: Best: 1.000000 using {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.001, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.01, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.02, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.9, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.0, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.2, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.4, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.6, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.8, 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'learn_rate': 0.03, 'momentum': 0.9, 'num_classes': 39}
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 01:29:37 [hh:mm:ss]
```

Figura 38: Resultados de la ejecución de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *learn rate and momentum*.

Dado los resultados obtenidos, utilizaremos el algoritmo *Stochastic gradient descent* sin *momentum*, ya que es un coste adicional innecesario que no nos aporta un mayor rendimiento en la red neuronal convolucional.

Con todas las otras combinaciones obtenemos el mismo *accuracy*, esto ha estado sucediendo en todas las ejecuciones que se han realizado en la optimización de los hiper-parámetros. Se debe remarcar que, tal y como se muestra en la tabla 20, se había conseguido un rendimiento del 99,30 % en la red neuronal convolucional que se está intentando mejorar con ambos datasets. Su valor inicial es prácticamente perfecto, es de esperar que con una *accuracy* tan elevada se obtengan tales resultados con todas las combinaciones.

Esta ejecución ha tenido un coste temporal de 1 hora, 29 minutos y 37 segundos.

#### 5.6.4. Inicialización del peso de la red

Cada nodo de la red neuronal tiene asignado una ponderación que se utiliza para calcular la suma de los *inputs*. El algoritmo de optimización *Stochastic gradient descent*, el cual se está utilizando, va incrementando los pesos con la intención de minimizar la función de pérdida.

Este algoritmo necesita tener asignados unos valores al inicio de la optimización, y es el inicializador el que definen la forma de establecer los pesos iniciales según la capa de la red neuronal. [50]

El parámetro que se debe introducir para poder ejecutar la estrategia optimizando este atributo es "*network weight initialization*". Con este podremos comparar el rendimiento de las distintas heurísticas y acabar asignado la que nos aporte una mayor precisión.

Como se muestra en la figura 39, los inicializadores utilizados son: *uniform*, *lecun uniform*, *normal*, *zero*, *glorot normal*, *glorot uniform*, *he normal* y *he uniform*.

Estos son los principales, pero en el enlace que se encuentra a continuación, se muestran todos los inicializadores que la dispone la librería Keras.<sup>12</sup>

<https://keras.io/api/layers/initializers/>

```
@staticmethod
def __get_parameters_network_weight_init(n_classes, image_size):
    init_mode = ['uniform', 'lecun_uniform', 'normal', 'zero', 'glorot_normal', 'glorot_uniform', 'he_normal',
                'he_uniform']

    param_grid = dict(init_mode=init_mode, num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_network_weight_init, epochs=EPOCHS, batch_size=BATCH_SIZE,
                                 verbose=2)

    return param_grid, classifier
```

Figura 39: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *network weight initialization*.

En la imagen 40 se muestran los resultados de esta ejecución, donde podemos apreciar que el inicializador *uniform* es el clasificado como el más óptimo. Aún así, podemos ver, como nos ha estados sucediendo en todas las optimizaciones hasta ahora, que los otros inicializadores nos aportan la misma eficiencia.

---

<sup>12</sup><https://keras.io/>

```
In [16]: !python3 Src/main.py --showOptimizedHyperparameter network_weight_initialization asl_alphabet_gray_150x150px sign_gestu
Epoch 7/10
38/38 - 8s - loss: 0.0026 - accuracy: 1.0000
Epoch 8/10
38/38 - 8s - loss: 0.0022 - accuracy: 1.0000
Epoch 9/10
38/38 - 8s - loss: 0.0019 - accuracy: 1.0000
Epoch 10/10
38/38 - 8s - loss: 0.0016 - accuracy: 1.0000
[INFO]: Best: 1.000000 using {'image_size': (150, 150), 'init_mode': 'uniform', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'uniform', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'lecun_uniform', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'normal', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'zero', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'glorot_normal', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'glorot_uniform', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'he_normal', 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'image_size': (150, 150), 'init_mode': 'he_uniform', 'num_classes': 39}
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 00:22:42 [hh:mm:ss]
```

Figura 40: Resultados de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *network weight initialization*.

Al tener únicamente ocho combinaciones a probar, la duración de esta ejecución ha sido menor a todas las realizadas hasta ahora, con un tiempo de 22 minutos y 42 segundos.

### 5.6.5. Función de activación neuronal

Otro parámetro a optimizar en la red neuronal convolucional es la función de activación. Permite utilizar la información importante, descartando los datos irrelevantes.

La función de activación comprobará si una neurona debe activarse o no utilizando operaciones matemáticas. Transformará la suma de ponderaciones del nodo en un valor de salida que se enviará a la siguiente capa, este valor servirá para identificar si la neurona está activa [39].

Las funciones de activación probadas en esta ejecución, tal y como se muestra en la imagen 41, son: *relu*, *softmax*, *softplus*, *softsign*, *tanh*, *sigmoid*, *hard\_sigmoid* y *linear*.

Se han utilizado las funciones más comunes, pero en el siguiente enlace se muestran todas las funciones de activación disponibles en la librería *Keras*<sup>13</sup>.

<https://keras.io/api/layers/activations/>

El parámetro que se debe introducir para poder ejecutar la estrategia optimizando este atributo es *"neuron activation function"*.

---

<sup>13</sup><https://keras.io/>

```

@staticmethod
def __get_parameters_neuron_activation_function(n_classes, image_size):
    activation = ['relu', 'softmax', 'softplus', 'tanh', 'sigmoid', 'hard_sigmoid', 'linear']

    param_grid = dict(activation=activation, num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_neuron_activation_function, epochs=EPOCHS,
                                 batch_size=BATCH_SIZE, verbose=2)

    return param_grid, classifier

```

Figura 41: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *neuron activation function*.

En la imagen 43 se muestran los resultados de esta ejecución. Podemos apreciar como se obtienen diferentes valores dependiendo de la función de activación, tres de los cuales tiene una *accuracy* máxima: *relu*, *softplus* y *linear*.

```

In [18]: a1n.py --showOptimizedHyperparameter neuron_activation_function asl_alphabet_gray_150x150px sign_gesture_gray_150x150px
Epoch 7/10
38/38 - 7s - loss: 0.1879 - accuracy: 0.6027
Epoch 8/10
38/38 - 7s - loss: 0.1398 - accuracy: 0.9120
Epoch 9/10
38/38 - 7s - loss: 0.1044 - accuracy: 0.9840
Epoch 10/10
38/38 - 7s - loss: 0.0792 - accuracy: 0.9973
[INFO]: Best: 1.000000 using {'activation': 'relu', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'activation': 'relu', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 0.400000 (0.000000) with: {'activation': 'softmax', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'activation': 'softplus', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 0.877333 (0.167852) with: {'activation': 'softplus', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 1.333333 (0.471405) with: {'activation': 'tanh', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 0.000000 (0.000000) with: {'activation': 'sigmoid', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 0.000000 (0.000000) with: {'activation': 'hard_sigmoid', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: 1.000000 (0.000000) with: {'activation': 'linear', 'image_size': (150, 150), 'num_classes': 39}
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 00:30:51 [hh:mm:ss]

```

Figura 42: Resultado de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *neuron activation function*.

Se utilizará la función de activación *ReLU*, la cual es la más utilizada en las redes neuronales convolucionales, ya que acelera su entrenamiento al tener un cálculo tan simple. Cualquier elemento negativo se iguala a cero, no utiliza operaciones costosas tales como exponentiales, multiplicaciones o divisiones.

La ejecución de la optimización de este parámetro ha tenido un coste temporal de 30 minutos y 51 segundos.

### 5.6.6. Regulación del *Dropout*

El *Dropout* [51] es una técnica que ayuda a regular el *overfitting* en todo tipo de redes neuronales con un coste computacional leve. Este método descarta o ignora, aleatoriamente, una cierta cantidad de neuronas y es muy común

en redes neuronales convolucionales bidimensionales con una gran cantidad de nodos en la red. Esta técnica se puede añadir después de una o varias capas.

En la configuración de esta técnica, utilizando la librería *Keras*<sup>14</sup>, se debe introducir un atributo denominado ”*dropout rate*”. Este parámetro tiene un rango de valores decimales del cero al uno. Por ejemplo, si se configura con un *dropout rate* del 0,2, el 20 % de los nodos serán descartados aleatoriamente.

Otro atributo a configurar es el *kernel constraint*, el cual restringe los pesos de las capas ocultas garantizando que la norma máxima de los pesos no excede el valor definido.

Estos dos atributos son los que se van a intentar optimizar en la ejecución de esta estrategia introduciendo el parámetro *dropout regularization*.

En la imagen 43 se muestra el clasificador de este modelo, con los posibles rangos de valores para el *dropout rate* y el *kernel constraint*.

```
@staticmethod
def __get_parameters_dropout_regularization(n_classes, image_size):
    weight_constraint = [1, 2, 3, 4, 5]
    dropout_rate = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]

    param_grid = dict(dropout_rate=dropout_rate, weight_constraint=weight_constraint, num_classes=[n_classes],
                      image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_dropout_regularization, epochs=EPOCHS, batch_size=BATCH_SIZE,
                                 verbose=2)

    return param_grid, classifier
```

Figura 43: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *dropout regularization*.

En la imagen 44 se muestran los resultados más significativos de esta ejecución, ya que el *output* ocupaba demasiado como para poder incluirlo en una única imagen.

---

<sup>14</sup><https://keras.io/>

```

[INFO]: Best: 1.000000 using {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 0.805333 (0.242153) with: {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}
[INFO]: 0.640000 (0.376652) with: {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 2}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 0.664000 (0.469530) with: {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 4}
[INFO]: 0.978667 (0.016438) with: {'dropout_rate': 0.0, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 5}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.1, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}
[INFO]: 0.677333 (0.456320) with: {'dropout_rate': 0.1, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 2}
[INFO]: 0.976000 (0.033941) with: {'dropout_rate': 0.1, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 0.920000 (0.107530) with: {'dropout_rate': 0.1, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 4}
[INFO]: 0.997333 (0.003771) with: {'dropout_rate': 0.1, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 5}
[INFO]: 0.666667 (0.471405) with: {'dropout_rate': 0.2, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.2, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 2}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.2, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 0.336000 (0.469530) with: {'dropout_rate': 0.2, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 4}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.2, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 5}
[INFO]: 0.773333 (0.314915) with: {'dropout_rate': 0.3, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}
[INFO]: 0.624000 (0.439102) with: {'dropout_rate': 0.3, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 2}
[INFO]: 0.706667 (0.235272) with: {'dropout_rate': 0.3, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 1.000000 (0.000000) with: {'dropout_rate': 0.3, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 4}
[INFO]: 0.792000 (0.294156) with: {'dropout_rate': 0.3, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 5}
[INFO]: 0.464000 (0.288666) with: {'dropout_rate': 0.4, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}
[INFO]: 0.450667 (0.414167) with: {'dropout_rate': 0.4, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 2}
[INFO]: 0.309333 (0.398522) with: {'dropout_rate': 0.4, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 3}
[INFO]: 0.746667 (0.335894) with: {'dropout_rate': 0.4, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 4}
[INFO]: 0.664000 (0.469530) with: {'dropout_rate': 0.4, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 5}
[INFO]: 0.594667 (0.411065) with: {'dropout_rate': 0.5, 'image_size': (150, 150), 'num_classes': 39, 'weight_constraint': 1}

```

Figura 44: Resultados de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *dropout regularization*.

Podemos ver como la opción que el optimizador muestra como la ideal tiene un *dropout rate* de 0, lo que significa que ningún nodo será descartado. En cuanto al *dropout\_regularization*, que en la imagen anterior se muestra como el *weight\_constraint*, tiene la máxima *accuracy* con un valor de 3, combinado con un *dropout rate* nulo.

Aún así, cuando se vaya a entrenar la red neuronal con los parámetros optimizados se harán dos comprobaciones, uno sin *Dropout* y otro con la combinación de un *dropout rate* con un valor igual a 0,2 y un *kernel constraint* igual a 3. Esta dos combinación se muestran en la imagen 44 con una *accuracy* del 100 %.

La ejecución de esta estrategia ha tenido un coste de tiempo igual a 2 horas, 29 minutos y 33 segundos.

### 5.6.7. Número de neuronas

El último atributo a optimizar es el número de neuronas en las capas de la red neuronal convolucional. El parámetro que se debe introducir para poder ejecutar la estrategia optimizando este atributo es *number\_neurons*.

En la imagen 45 se muestran los rangos de valores que pueden adquirir las dos capas con un número de nodos configurables.

```
@staticmethod
def __get_parameters_number_neurons(n_classes, image_size):
    neurons_conv_layer = [1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140]
    neurons_dense_layer = [10, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140]

    param_grid = dict(neurons_conv_layer=neurons_conv_layer, neurons_dense_layer=neurons_dense_layer,
                      num_classes=[n_classes], image_size=[image_size])
    classifier = KerasClassifier(build_fn=create_model_number_neurons, epochs=EPOCHS, batch_size=BATCH_SIZE,
                                 verbose=2)

    return param_grid, classifier
```

Figura 45: Parámetros evaluados en la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *number neurons*.

En la siguiente imagen aparece el resultado de la ejecución de esta estrategia.

```
[INFO]: Best: 1.000000 using {'image_size': (150, 150), 'neurons_conv_layer': 15, 'neurons_dense_layer': 100, 'num_classes': 39}
[INFO]: 0.666667 (0.471405) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 10, 'num_classes': 39}
[INFO]: 0.773333 (0.209261) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 20, 'num_classes': 39}
[INFO]: 0.666667 (0.471405) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 25, 'num_classes': 39}
[INFO]: 0.666667 (0.471405) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 30, 'num_classes': 39}
[INFO]: 0.666667 (0.471405) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 40, 'num_classes': 39}
[INFO]: 0.800000 (0.282843) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 50, 'num_classes': 39}
[INFO]: 0.960000 (0.051016) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 60, 'num_classes': 39}
[INFO]: 0.624000 (0.303789) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 80, 'num_classes': 39}
[INFO]: 0.466667 (0.410961) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 100, 'num_classes': 39}
[INFO]: 0.994667 (0.007542) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 120, 'num_classes': 39}
[INFO]: 0.952000 (0.067882) with: {'image_size': (150, 150), 'neurons_conv_layer': 1, 'neurons_dense_layer': 140, 'num_classes': 39}
[INFO]: 0.221333 (0.160577) with: {'image_size': (150, 150), 'neurons_conv_layer': 5, 'neurons_dense_layer': 10, 'num_classes': 39}
[INFO]: 0.202667 (0.286614) with: {'image_size': (150, 150), 'neurons_conv_layer': 5, 'neurons_dense_layer': 20, 'num_classes': 39}
[INFO]: 0.666667 (0.471405) with: {'image_size': (150, 150), 'neurons_conv_layer': 5, 'neurons_dense_layer': 25, 'num_classes': 39}
[INFO]: 0.266667 (0.316941) with: {'image_size': (150, 150), 'neurons_conv_layer': 5, 'neurons_dense_layer': 30, 'num_classes': 39}
[INFO]: 0.000000 (0.000000) with: {'image_size': (150, 150), 'neurons_conv_layer': 5, 'neurons_dense_layer': 40, 'num_classes': 39}
```

Figura 46: Resultados de la ejecución de la estrategia *Hyperparameter Optimization Strategy* con el parámetro *number neurons*.

Según el resultado obtenido, la combinación más óptima para el número de neuronas son 15 en la capa convolucional y 100 en la densa oculta.

La ejecución de esta estrategia tiene un coste de tiempo de 10 horas, 56 minutos y 9 segundos. Este parámetro es el que ha tenido la ejecución más larga en la estrategia *Hyperparameter Optimization Strategy*.

#### 5.6.8. Resultados de la optimización

Una vez realizadas todas las optimizaciones se creó un nuevo modelo de red neuronal convolucional mejorado. Este nuevo modelo se muestra en la imagen 47.

```
def __get_improved_sequential_model(n_classes, shape):
    model = Sequential()
    model.add(Conv2D(NEURONS_CONV_LAYER, kernel_size=(3, 3), strides=(1, 1), padding='valid', activation=ACTIVATION,
                    input_shape=(shape[1], shape[2], 1), kernel_initializer=INIT_MODE,
                    kernel_constraint=max_norm(WIGHT_CONSTRAINT)))
    model.add(Dropout(DROPOUT_RATE))
    model.add(MaxPool2D(pool_size=(1, 1)))
    model.add(Flatten())
    model.add(Dense(NEURONS_DENSE_LAYER, kernel_initializer=INIT_MODE, activation=ACTIVATION))
    model.add(Dropout(DROPOUT_RATE))
    model.add(Dense(n_classes, activation='softmax'))

    return model
```

Figura 47: Modelo de la red neuronal convolucional obtenida a partir de los resultados de las ejecuciones de la estrategia *Hyperparameter Optimization Strategy*.

Se han realizado diversos cambios comparado con la red neuronal convolucional presentada en el apartado 5.2.2. Se han modificado el valor de varios atributos, tales como:

- El número de nodos en la capa convolucional y en la capa densa oculta. En la anterior versión de este modelo se utilizaron 25 nodos en la convolucional y 100 en la densa. Actualmente, se han reducido los 25 nodos a 15.
- Anteriormente no se definía ningún inicializador, pero como se muestra en la imagen 48, que hace referencia al constructor de la capa convolucional, el inicializador por defecto es el *glorot\_uniform*. En el nuevo modelo se define con el inicializador *uniform*.
- Por último, se han añadido dos nuevas capas de *Dropout* y un *kernel constraint* en la capa convolucional. Este parámetro tiene un valor nulo cuando

no se define, como se muestra en la imagen 48, por lo tanto, en el modelo anterior no se restringía los pesos de las capas ocultas.

Actualmente se tienen dos conjuntos de valores a probar, uno con un *dropout rate* de 0.2 y un *kernel constraint* de 2, o un 0 de *dropout rate* y un 3 de *kernel constraint*.

```
def __init__(self,
            filters,
            kernel_size,
            strides=(1, 1),
            padding='valid',
            data_format=None,
            dilation_rate=(1, 1),
            groups=1,
            activation=None,
            use_bias=True,
            kernel_initializer='glorot_uniform',
            bias_initializer='zeros',
            kernel_regularizer=None,
            bias_regularizer=None,
            activity_regularizer=None,
            kernel_constraint=None,
            bias_constraint=None,
            **kwargs):
    super(Conv2D, self).__init__(
        rank=2,
        filters=filters,
        kernel_size=kernel_size,
```

Figura 48: Parámetros requeridos en el constructor de la clase *Conv2D* de la librería *keras.layers*.

Se ha entrenado la red neuronal convolucional mejorada con los dos conjuntos de valores del *dropout*, en la siguiente tabla se comparan los resultados obtenidos y los costes temporales y de memoria de cada uno de ellos.

	<i>Sin Dropout rate = 0.0 Kernel Constraint = 3</i>	<i>Dropout rate = 0.2 Kernel Constraint = 2</i>
<b>Accuracy</b>	99,72 %	97,26 %
<b>Coste temporal entrenamiento</b>	01:06:50 hora	01:29:03 hora
<b>Coste temporal predicción</b>	00:01:13 horas	00:01:05 horas
<b>Coste memoria</b>	394,4 MB	394,4 MB

Tabla 23: Comparativa de la red neuronal convolucional sin *Dropout* o con un *Dropout* con un índice igual a cero.

Dados los resultados obtenidos en la tabla 23, podemos concluir que para el problema que estamos intentado solucionar, con los dos datasets combinados y con las características definidas de la red neuronal convolucional, obtenemos un resultado más óptimo sin *dropout*.

En cuanto al coste de ambas ejecuciones, tanto la suma del temporal como el de memoria es menor sin el *dropout*, lo que nos confirma cual es la mejor solución.

Aún así, una *accuracy* del 97,26 % es un resultado excelente, pero al tener un valor más elevado con un coste menor, se utilizará la primera opción.

Todos los resultados obtenidos a partir de las ejecuciones de la estrategia *Hyperparameter Optimization Strategy* se han guardado en un archivo denominado *hyperparameters.py*. La imagen 49 muestra el contenido de este archivo con los valores finales.

```
"""
After executing the --showOptimizedHyperparameter strategy, these are the best results obtained for the convolutional
neural network.
"""

EPOCHS = 10
BATCH_SIZE = 10
OPTIMIZER_ALGORITHM = 'SGD'
LEARN_RATE = 0.001
MOMENTUM = 0.0
INIT_MODE = 'uniform'
ACTIVATION = 'relu'
DROPOUT_RATE = 0.0
WEIGHT_CONSTRAINT = 3
NEURONS_CONV_LAYER = 15
NEURONS_DENSE_LAYER = 100
```

Figura 49: Archivo del proyecto que contiene los valores de los atributos optimizados obtenidos a partir de las ejecuciones de la estrategia *Hyperparameter Optimization Strategy*.

Los valores de la tabla 23 han sido obtenidos a partir de un mismo modelo

modificando los dos atributos diferenciados. Esto significa que en el caso del *dropout rate* igual a cero, se ha entrenado la red neuronal con una capa adicional que no era necesaria. Utilizando la lógica diríamos que una capa de *Dropout* con un índice nulo únicamente estaría incrementando el coste de memoria, ya que en sí no tiene ninguna utilidad.

En la tabla 24 se comparan los mismos atributos que en la tabla 23, pero en este caso las dos redes neuronales convolucionales a comparar se diferencian en si tienen las capas de *Dropout* con un índice nulo, o si no disponen de estas capas.

Esta comparación se lleva a cabo para evaluar si el resultado que se espera utilizando la lógica es el mismo que el resultado real.

	<i>Dropout rate igual a cero</i>	<i>Sin las capas Dropout</i>
<b>Accuracy</b>	99,72 % - 99,75 %	99.46 % - 99.57 %
<b>Coste temporal entrenamiento</b>	01:06:50 hora	01:14:57 hora
<b>Coste temporal predicción</b>	00:01:13 horas	00:01:17 horas
<b>Coste memoria</b>	394,4 MB	394,4 MB

Tabla 24: Comparativa de la red neuronal convolucional con los parámetros optimizados con y sin *Dropout*.

Se han ejecutado las estrategias de entrenamiento y de predicción diversas veces con la intención de poder comparar si realmente tienen características diferentes. En la tabla 24 se muestran los rangos de valores más significativos en esta comparación.

Se aprecia como hay una diferencia significativa en la *accuracy* entre ambas redes. En todos los entrenamientos que se han realizado, en ningún momento el modelo sin *Dropout* ha llegado a igualar el peor valor del modelo con el *dropout rate* nulo.

Al no comprender el porque de esta diferenciación en la *accuracy* entre ambas redes neuronales, se ha hecho una búsqueda exhaustiva por internet. La única información que se ha conseguido obtener ha sido una publicación en *Stack Overflow*<sup>15</sup>, sin ninguna respuesta, y un *bug* abierto en *Git Hub*<sup>16</sup> que hacia referencia exactamente a este error. El *bug* tenía como título:

Keras Dropout layer changes results with dropout=0.0

---

<sup>15</sup><https://es.stackoverflow.com/>

<sup>16</sup><https://github.com/>

Este se puede encontrar en el siguiente enlace:

<https://github.com/tensorflow/tensorflow/issues/10845>

El problema está en que este *bug* se abrió en 2017, y la solución que aparece en los comentarios no hace referencia al estado actual de la librería *tensorflow*. Con lo que no se puede entender exactamente el porque de esta diferenciación.

En este proyecto se utilizará el modelo sin la capa de *Dropout*, ya que no la otra opción no es una solución fiable si contiene un error. Aunque, aparentemente, pueda parecer que nos aporta una *accuracy* más elevada con un coste menor, no entendemos cual es la forma en la que está implementado. Al no poder dar una explicación a este problema, no se puede dar como valido el modelo con un *dropout rate* igual a cero.

En futuras investigación sería recomendable investigar cual es el modelo matemático que hay detrás de esta capa, de la librería *tensorflow*, para poder concluir si este *bug*, no solucionado, supondría un problema en futuros entrenamientos.

## 5.7. Comparativa de las experimentaciones 4 y 5

En la experimentación del apartado 5.6, se han evaluado unas ciertas características del modelo utilizado en la red neuronal convolucional. Se han obtenido unos resultados donde se mostraba cuales eran los valores óptimos para estos atributos, buscando una mayor *accuracy* en la red neuronal.

Después de probar las combinaciones especificadas, se ha obtenido el modelo que mejor se adapta al problema que se intenta solucionar. Se ha deducido que la mejor solución era utilizar una red neuronal con una capa convolucional de 25 nodos y una densa de 100, con el inicializador *glorot\_uniform* y sin ninguna capa de *Dropout*.

Este ha sido el resultado obtenido dada las configuraciones de la optimización, pero lo que no se ha evaluado es la comparación de este resultado con el obtenido en el apartado 5.5. Se podrían dar tres posibilidades en cuanto a la comparación de la *accuracy*:

- No haber ninguna modificación. Aún haber evaluado los valores más óptimos de los atributos a configurar, al tener ya de por si una *accuracy* muy elevada, prácticamente perfecta, estas modificaciones podrían no implicar una mejora.
- Mejorar la eficiencia de la red neuronal. Al haber obtenido los mejores valores centrados exclusivamente en el problema que se intenta solucionar con el dataset con el que se hace el entrenamiento, podríamos haber encontrado una configuración más específica que nos incremente su eficiencia.
- Empeorar la eficiencia de la red neuronal. Las comprobaciones de las optimizaciones no se han podido realizar con todos los datos del dataset por un problema de coste de memoria y tiempo. Al haber hecho las pruebas con un conjunto de datos muy pequeño, comparado con el que se hace el entrenamiento real, se podrían haber obtenido resultados que no acabasen de ser específicos para nuestro problema.

Teniendo presente estas tres posibilidades se realizó el entrenamiento y las predicciones con la red neuronal, supuestamente, optimizada. A continuación, se muestra una tabla con los valores de la comparación.

	<b>CNN no optimizada</b>	<b>CNN optimizada</b>
<b>Accuracy</b>	99,30 %	99,52 %
<b>Coste temporal entrenamiento</b>	01:35:02 hora	01:06:50 hora
<b>Coste temporal predicción</b>	00:01:43 horas	00:01:13 horas
<b>Coste memoria</b>	657,2 MB	394,4 MB

Tabla 25: Comparativa de los resultados de las redes neuronales convolucionales de los apartados 5.5 y 5.6.

Evaluado los valores de la tabla 25, podemos apreciar como la red neuronal optimizada incrementa la *accuracy* 22 décimas. Este incremento puede parecer muy leve, pero al ser un valor de eficiencia tan elevado, el poder incrementar unas décimas significa una gran mejora en el modelo.

Además, podemos apreciar como el coste temporal de entrenamiento ha sido reducido media hora, mejorándose un 30,52 %. y en cuanto al coste de memoria, se han disminuido 262,8 MB, un 39,99 %.

Dados estos resultados, podemos concluir que la experimentación que se ha centrado en la optimización de la red neuronal convolucional ha sido exitosa, y esta es la mejor solución que se tiene hasta el momento.

## 5.8. Experimentación 6: Uso de clasificadores binarios

En el apartado 5.7, se muestra una posible solución a este proyecto, donde se entrena una red neuronal convolucional. Hasta el momento todas las implementaciones que se han realizado han estado centradas en analizar los datos de un modo categórico.

Este método nos ha aportado un rendimiento prácticamente perfecto, si nos centramos en evaluar únicamente su *accuracy*. La solución que se propone tiene un rendimiento del 99,52% con un coste de tiempo de entrenamiento y predicción de una hora y con un coste de memoria de 394,4 MB.

Aún haber obtenido un resultado excelente, se ha querido evaluar otra técnica. El objetivo es ampliar el conocimiento de las diferentes técnicas de clasificación y poder comprobar si esta también daría unos resultados tan satisfactorios.

Como se explica en los fundamentos, en el apartado 3.1, hay diferentes tipos de clasificadores. Hasta ahora, se ha utilizado el clasificador multi-clase, ya que es el que mejor se adapta a las características de los datos con los que se está trabajando. Pero hay otra técnica que se puede utilizar y que no se ha implementado en las experimentaciones anteriores. Esta consiste en adaptar un clasificador binario para que sea capaz de clasificar más de una clase. La explicación de esta técnica se encuentra en el apartado 3.1.2.

Se utilizará el modelo que mejores resultados nos ha proporcionado hasta el momento, en este caso es la red neuronal convolucional presentada en el apartado 5.7. Se modificarán los parámetros necesarios para que se adapte a las características del tipo de clasificador que se implementará.

Para la realización de este experimento se han implementado dos estrategias, siguiendo la misma estructura que con el árbol de decisión y las redes neuronales. Estas estrategias se denominan:

- **Train Binary Neural Network:** Para ejecutarla se deben definir los *Pickles* (archivos donde se almacenan los datos) a utilizar. Una vez definidos, se procesaran los datos con tal de poder crear tantos clasificadores como tipos de clases haya, en este caso, hay una clase por cada letra y número del lenguaje de signos, 39 en total.

Los datos dependerán del clasificador, cada clasificador se centra en identificar si las imágenes son una señal en concreto o no. Es por esto que los datos serán transformados en binarios previamente al entrenamiento de cada modelo.

Cuando se han creado todos los clasificadores, se construyen los modelos de las redes neuronales convolucionales y se entrena basados en los datos de *training*.

Antes de finalizar la ejecución, se almacenan todos los modelos en archivos con extensión *.h5*, comprimidos en un *zip* para minimizar el coste de memoria.

La imagen 50 muestra el diagrama de flujo de esta estrategia.

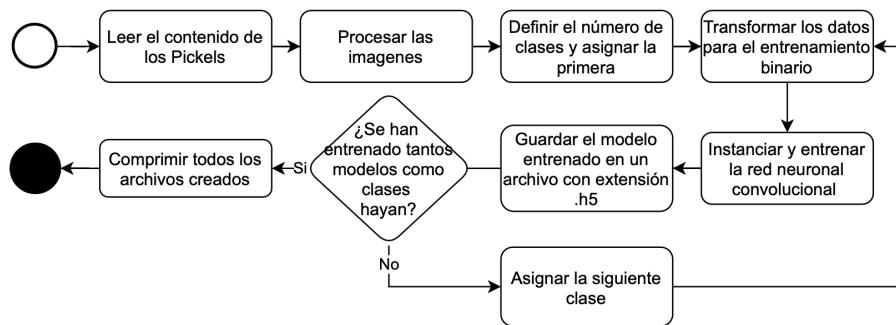


Figura 50: Diagrama de flujo de la estrategia *Train Binary Neural Network*

- **Accuracy Binary Neural Network:** Esta estrategia recuperará los modelos creados en la anterior ejecución. Por cada modelo, introducirá los datos de *testing* y evaluará la predicción de cada una de las redes neuronales binarias, mostrando la *accuracy* de cada una de ellas.

Una vez obtenidas todas las predicciones, se calculará, por cada dato introducido del *testing*, qué modelo ha proporcionado un valor más elevado, pudiendo calcular así cual es la clase resultante. A continuación, se mostrará la *accuracy* global de esta técnica.

La imagen 51 hace referencia al diagrama de flujo de la estrategia *Accuracy Binary Neural Network*.

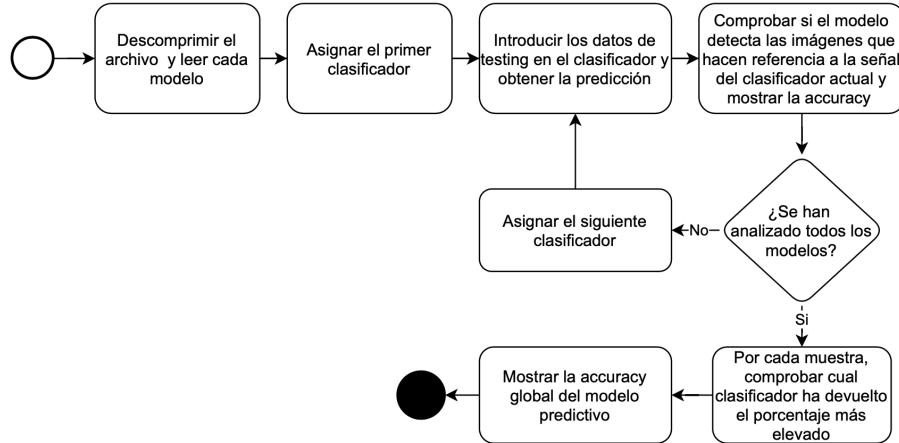


Figura 51: Diagrama de flujo de la estrategia *Accuracy Binary Neural Network*

Estas estrategias poseen un coste muy elevado, tanto de memoria como de tiempo. Se intentó entrenar el modelo predictivo con todas las muestras de los datasets, pero después de ocho horas, no había acabado de entrenar ni 1 red neuronal, y se debían entrenar 39.

Se decidió probar este modelo con menos datos, utilizando únicamente tres clases:

- La señal que hace referencia a la letra 'A'.
- La señal que hace referencia a la letra 'B'.
- La señal que hace referencia a la letra 'C'.

Se han restringido los valores de las muestras, con la intención de únicamente aceptar estos tres valores y reducir el tamaño del dataset. Una vez eliminados las muestras que no nos interesan, se han ejecutado las dos estrategias mencionadas. En la imagen 52 se muestra un diagrama de del entrenamiento con las tres clases mencionadas.

En el primer paso se obtienen las muestras buscadas, y se procesan para obtener las etiquetas (X) y el valor de los píxeles (Y) en dos estructuras diferentes. En el tercer paso, el cual se ejecuta secuencialmente, no en paralelo, se analizan los valores. La estructura, que almacena las etiquetas, modifica el valor de sus elementos, con un valor de 0, si no pertenece a la clase actual, y con uno si pertenece a la clase.

El cuarto paso se entrena todos los modelo, pero, como se muestra el diagrama de flujos de la imagen 50, el tercer y el cuarto paso se llevan a cabo secuencialmente. Una vez se ha finalizado el entrenamiento, se vuelve al paso tres con la siguiente clase.

El paso cinco hace referencia al *output* que proporcionarán estos modelos cuando se realiza la predicción.

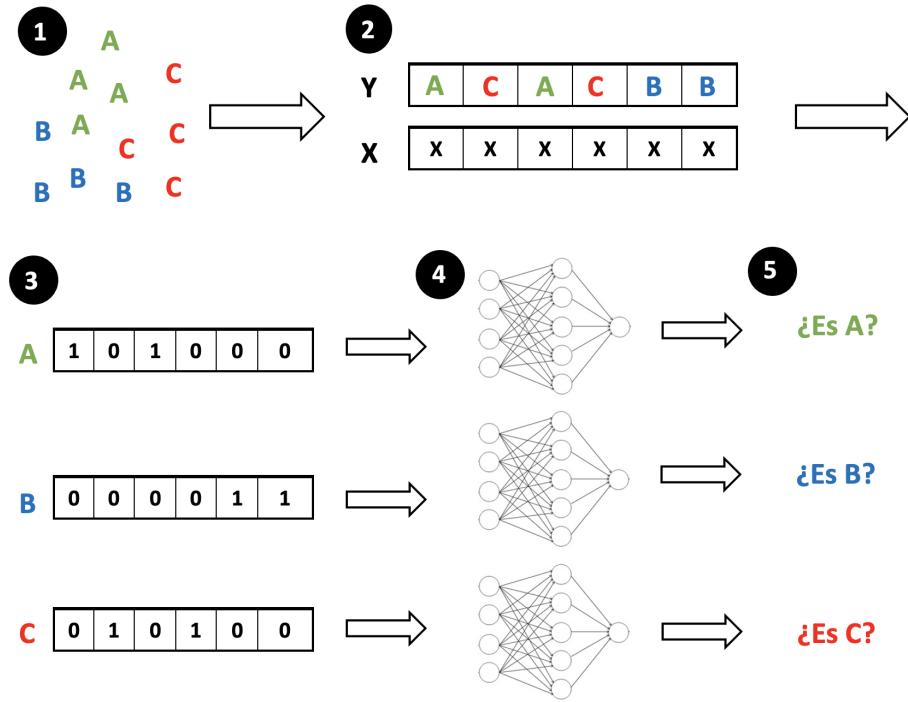


Figura 52: Diagrama de flujo del entrenamiento de una red neuronal binaria con una estrategia One vs All.

En la imagen 53, se muestra la accuracy obtenida con este modelo con las tres clases definidas. Se puede apreciar la accuracy de cada uno de los clasificadores, el cual ha sido:

- Clasificador del signo 'A': 100 %.
- Clasificador del signo 'B': 99,82 %.
- Clasificador del signo 'C': 99,95 %.

La accuracy global se calcula analizando el clasificador que ha devuelto la predicción más elevada, puede pasar que más de uno devuelva el máximo valor, pero como se muestra en la imagen 53 no se ha dado el caso en esta ejecución.

La accuracy global ha sido del 100 %, lo que muestra que esta técnica puede ser muy prometedora.

```
!python3 Src/main.py --accuracyBinaryNeuralNetwork abc abc_asl_alphabet_gray_150x150px-sign_gesture_gray_150x150px_mode
[INFO]: Strategy selected: --accuracyBinaryNeuralNetwork
[INFO]: Arguments entered:
        * Signs to train: abc
        * Neural Network model file: abc, abc_asl_alphabet_gray_150x150px-sign_gesture_gray_150x150px_models.zip
[INFO]: The accuracy of the binary neural network of the sign 'C' is: 100.00%
[INFO]: The accuracy of the binary neural network of the sign 'B' is: 99.82%
[INFO]: The accuracy of the binary neural network of the sign 'A' is: 99.95%
[INFO]: The global accuracy is: 100.00%
[INFO]: 0.00% of the correct samples had more than one solution.
[INFO]: Strategy executed successfully
[INFO]: Execution finished
[INFO]: The duration of the execution has been 00:00:41 [hh:mm:ss]
```

Figura 53: *Output* de la ejecución de la estrategia *Accuracy Binary Neural Network* con las señales: 'A', 'B' y 'C'.

La tabla 26, muestra las características de esta ejecución. Se puede apreciar como comprimir los archivos genera una reducción en el coste de memoria de la ejecución de ambas estrategias.

El tiempo de entrenamiento y de predicción puede parecer leve, comparado con los resultados de otros modelos predictivos implementados en anteriores estrategias. Pero debemos recordar que esta ha sido entrenada únicamente con tres clases, si se hubiese entrenado con 39, como es el dataset real, estos valores se dispararían.

Modelo predictivo con redes neuronales binarias de los signos 'A', 'B' y 'C'	
<b>Accuracy</b>	100 %
<b>Tiempo ejecución del entrenamiento</b>	00:47:38 horas
<b>Tiempo ejecución de la predicción</b>	00:00:41 horas
<b>Coste de memoria de los archivos comprimidos</b>	975,9 MB
<b>Coste de memoria de los archivos sin comprimir</b>	1,18 GB

Tabla 26: Características de la ejecución del modelo predictivo con redes neuronales binarias con las muestras de los signos 'A', 'B' y 'C'.

Una vez obtenidos este rendimiento tan prometedor, se volvieron a ejecutar

las mismas estrategias, pero esta vez admitiendo las clases numéricas, las señales del cero al diez.

En este caso, los resultados no fueron tan favorables. Como se muestra en la imagen 54, la máxima *accuracy* ha sido del 100 %, pero el clasificador de la señal '4' nos proporciona una *accuracy* del 24,80 %, un valor muy pobre.

Al tener clasificadores con un rendimiento tan leve, afectan a la *accuracy* global del modelo predictivo, ya que proporcionan valores, en sus predicciones, más elevados que el clasificador correcto.

Figura 54: *Output* de la ejecución de la estrategia *Accuracy Binary Neural Network* con las señales numéricas.

Estos resultados pueden ser debidos a que el modelo de red neuronal, que se está utilizando, ha sido diseñado para el modelo predictivo multi-clase, no para ser entrenado de forma binaria. Se debería repetir la experimentación del apartado 5.6 con los datos de *training* que se utilizan en este apartado.

Aún haber obtenido una *accuracy* muy reducida en algunos clasificadores, más de la mayoría tienen un valor superior al 90 %. Con los modelos de red neuronal convolucional optimizados puede llegar a ser una solución muy buena para el problema que se está intentando resolver, siempre y cuando se sea consciente del elevado coste computacional que presenta.

## 6. Propuesta

Al largo del desarrollo de este proyecto se han realizado diversas experimentaciones, donde se han analizado diferentes modelos de predicción. Se han ido comparando los resultados obtenidos con la intención de poder proponer el modelo que aporta un mayor rendimiento, minimizando el coste de memoria y de tiempo.

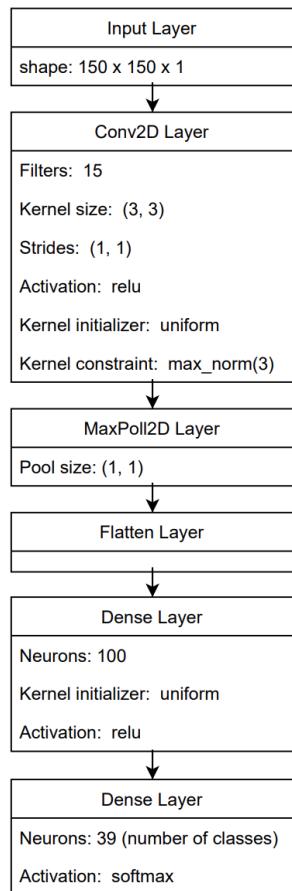


Figura 55: Estructura de la red neuronal propuesta.

Cuando se obtiene este modelo de clasificación debe ser entrenado, en esta implementación se ha utilizado una proporción 70:30, un 70 % de los datos para el entrenamiento y un 30 % para las pruebas.

Se debe tener en cuenta bajo las condiciones que se han realizado estas experimentaciones. Si esta propuesta se llega a implementar en una aplicación real, será ejecutada en tiempo de explotación, lo que reducirá una infinidad el coste de tiempo mostrado.

La propuesta mostrada es con la que se ha obtenido los mejores resultados bajo las condiciones establecidas en el apartado 4. En el caso de querer hacer uso de esta propuesta, se recomienda comprobar si, bajo las circunstancias de las nuevas ejecuciones, alguno de los modelos mostrados en el apartado 5 no supera el rendimiento del modelo propuesto.

El modelo con el que se ha obtenido un mejor rendimiento es la red neuronal convolucional obtenida en la experimentación del apartado 5.6.8. La imagen 55 muestra la estructura de esta red neuronal con las características de cada capa. Esta dispone de seis capas: una de entrada, una de salida y cuatro ocultas.

Utiliza una clasificación multi-clase, con un dataset de treinta y nueve clases diferentes. Estas hacen referencia a todas las letras del abecedario inglés y los valores numéricos del cero al diez.

Tanto en el entrenamiento como en las pruebas es importante seguir los pasos que se muestran en el diagrama de flujo de la imagen 56.

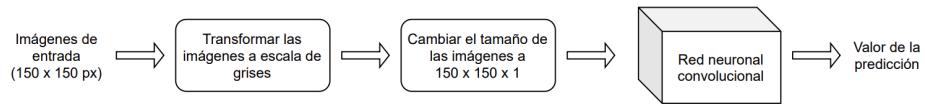


Figura 56: Diagrama de flujo de los pasos a seguir para poder utilizar el clasificador propuesto.

El tamaño de las imágenes es independiente, en la implementación realizada en este proyecto se ha utilizado un tamaño de 1050 x 150 píxeles, pero se puede utilizar el tamaño que se encuentre más conveniente, siempre y cuando al hacer el *resize* sea el correspondiente al de las imágenes introducidas.

Con el modelo de clasificación propuesto se ha obtenido una *accuracy* del 99,52 %, con un tiempo total de ejecución de 1 hora, 55 minutos y 36 segundos, y un coste de memoria de 12,88 GB. Estos valores han sido obtenidos a partir de las tablas 27 y 28

La tabla 27 muestra la duración de la ejecución de todas las estrategias necesarias para poder obtener la predicción final del modelo propuesto.

	Duración
Guardar el dataset <i>Sign Language Gesture Images Dataset</i> [9] en un pickle	00:09:35 horas
Guardar el dataset <i>ASL test and train</i> [10] en un pickle	00:23:09 horas
Entrenamiento del modelo predictivo	01:06:50 hora
Obtener las predicciones del modelo predictivo	00:01:13 horas
<b>Total</b>	<b>01:55:36 horas</b>

Tabla 27: El coste de tiempo de todas las estrategias utilizadas para obtener las predicciones del modelo propuesto

La tabla 28 muestra el coste de memoria que generar los archivos obtenidos a partir de la ejecuciones de todas las estrategias necesarias para poder obtener la predicción final del modelo propuesto.

	Tamaño del archivo
Pickle generado a partir del dataset <i>Sign Language Gesture Images Dataset</i> [9]	7,75 GB
Pickle generado a partir del dataset <i>ASL test and train</i> [10]	4,74 GB
Archivo con extensión <i>.h5</i> que contiene el modelo	394,4 MB
<b>Total</b>	12,88 GB

Tabla 28: El coste de memoria de todas las estrategias utilizadas para obtener las predicciones del modelo propuesto

## 7. Metodología utilizada

El proyecto se ha realizado siguiendo la metodología espiral. Esta consiste en seguir una serie de fases, el cual su conjunto se denomina como ciclo, realizándose en forma de espiral. Las fases son las siguientes:

- Planificación: se analizan los objetivos del nuevo ciclo, la intención es ir incrementando estos objetivos a cada ciclo.
- Análisis de Riesgo: se investigan los factores de riesgo que se pueden llevar a cabo en el nuevo ciclo. En el caso de este proyecto, algunos de los factores de riesgo que se analizaron fueron los costes de tiempo y de memoria de las ejecuciones, los días de vacaciones...
- Implementación: se desarrollan los objetivos definidos en la planificación teniendo en cuenta los factores de riesgo.
- Evaluación: al finalizar el ciclo se debe hacer una evaluación para comprobar si se han logrado los objetivos o si han aparecido nuevos factores de riesgo que no habían sido contemplados. En el caso de este proyecto, se realizaban reuniones al finalizar el ciclo para poder mostrar los resultados obtenidos y poder evaluar si estos resultados eran suficientemente prometedores como para poder continuar desarrollándolos.

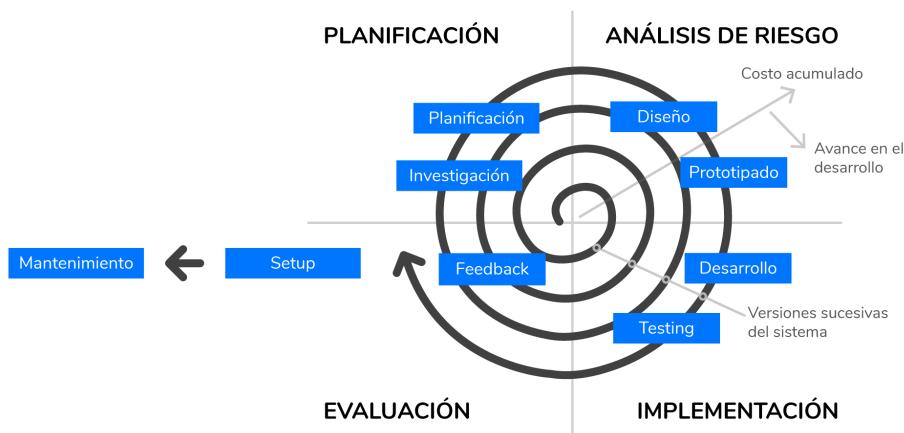


Figura 57: Explicación de la metodología espiral.

Lo que hace referencia a la implementación del proyecto, se han utilizado dos repositorios de Github, uno en el que se almacenan los datasets utilizados y otro donde se encuentra el código y la memoria del proyecto.

Se ha hecho uso de Gitflow<sup>17</sup> para poder utilizar una estructura bien definida en la implementación del proyecto. Para simplificar el control de versiones y poder hacer un uso más preciso del flujo de trabajo de Git, se ha utilizado la herramienta Sourcetree<sup>18</sup>.

El repositorio que contiene el código dispone de un proyecto, en el que se han creado *issues* y *milestones* utilizando una versión del *template* kanban que mejor se adaptaba a las necesidades de este proyecto. Este *template* se muestra en la imagen 65 del anexo A.1.

La versión final del proyecto se encuentra en la rama *master*, pero este dispone de una rama de *development* y varias *features* que siguen activas por si se quiere comprobar los cambios que se realizaron en la implementación de sus funcionalidades.

A continuación se muestran los enlaces de los dos repositorios de Github:

- Repositorio que dispone del código, la memoria y el proyecto de Github:

*<https://github.com/marGaliana/SignLanguageProcessing.git>*

- Repositorio que dispone de los datasets utilizados:

*<https://github.com/marGaliana/SignLanguageProcessingDataset.git>*

---

<sup>17</sup><https://www.atlassian.com/es/git/tutorials/comparing-workflows/gitflow-workflow>

<sup>18</sup><https://www.sourcetreeapp.com/>

## 8. Estructura del proyecto

Durante la realización de los diferentes experimentos se ha creado un proyecto en Python formado por diversas estrategias.

Las estrategias permiten dividir las ejecuciones, guardando los resultados en archivos. El entrenamiento y la predicción de un clasificador baso en árbol de decisión o una red neuronal puede llegar a tener una duración muy elevada, dividir las estrategias permite no tener que realizar el entrenamiento cada vez que se quiera realizar una predicción.

### 8.0.1. Definición de las estrategias

En total se dispone de once estrategias:

- **Setup:** Esta estrategia es la primera que se debe ejecutar. Crea los directorios y los archivos necesarios para poder hacer uso de las otras ejecuciones.
- **Help:** Esta estrategia muestra la descripción y los parámetros requeridos, con sus posibles valores y su orden, de todas las estrategias que se pueden ejecutar.
- **SaveDatabase:** Esta estrategia procesa las imágenes del dataset seleccionado y las dimensiona al tamaño esperado en los modelos predictivos, además de transformarlas a escala de grises. Una vez finalizado el procesamiento del dataset, se guardan las nuevas imágenes en archivos denominados *Pickles*.
- **TrainDecisionTree:** Dado un o un conjunto de *pickles*, se crea el clasificador basado en un árbol de decisión definido en el apartado 5.3 y se realiza su entrenamiento. Una vez finalizado su entrenamiento, se guarda el modelo en un archivo con extensión *.pickle.dat*.
- **AccuracyDecisionTree:** A partir del modelo del árbol de decisión seleccionado, el cual se ha obtenido en la ejecución de la estrategia *TrainDecisionTree*, se realizan las predicciones con el conjunto de imágenes de *testing*. Al finalizar se evalúa el porcentaje de aciertos que se han obtenido

de la predicción.

- **TrainCategoricalNeuralNetwork:** Dado un o un conjunto de *pickles*, se crea el clasificador basado en una red neuronal multi-clase. Esta red neuronal puede ser artificial, con la estructura mostrada en el apartado 5.2.1, o convolucional. Esta última puede contener la estructura definida en el apartado 5.2.2 o el de la solución propuesta final, del apartado 6. Una vez obtenido el clasificador, se realiza su entrenamiento, y, al finalizarlo, se guarda el modelo en un archivo con extensión *.h5*.
- **AccuracyCategoricalNeuralNetwork:** A partir del modelo de la red neuronal multi-clase seleccionada, el cual se ha obtenido en la ejecución de la estrategia *TrainCategoricalNeuralNetwork*, se realizan las predicciones con el conjunto de imágenes de *testing*. Al finalizar se evalúa el porcentaje de aciertos que se han obtenido de la predicción.
- **TrainBinaryNeuralNetwork:** Dado un o un conjunto de *pickles*, se crea el clasificador basado en una red neuronal convolucional binaria con una estrategia *One vs All*. El modelo de esta red está definido en el apartado 5.8. Una vez obtenido el clasificador, se realiza su entrenamiento, y, al finalizarlo, se guarda el modelo en un archivo con extensión *.h5*.
- **AccuracyBinaryNeuralNetwork:** A partir del modelo de la red neuronal convolucional binaria con una estrategia *One vs All* seleccionada, el cual se ha obtenido en la ejecución de la estrategia *TrainBinaryNeuralNetwork*, se realizan las predicciones con el conjunto de imágenes de *testing*. Al finalizar se evalúa el porcentaje de aciertos que se han obtenido de la predicción.
- **ShowOptimizedHyperparameter:** Esta estrategia realiza diversos entrenamientos del modelo de la red neuronal convolucional multi-clase definido en el apartado 5.2.2. Dependiendo de los parámetros introducidos, se optimizan unos atributos o otros. Al finalizar la ejecución, se muestran los resultados obtenidos con las diferentes combinaciones. Estos resultados se han guardado en una archivo Python como variables constantes para poder ser utilizados en el modelo de la red neuronal convolucional del apartado 6.
- **Predict:** Esta estrategia realiza la predicción de la imagen seleccionada. Puede se ejecutado en cualquiera de los modelos generados en las estrategias anteriores.

### 8.0.2. Secuencia de ejecución de las estrategias

El orden de ejecución de estas estrategias es importante para poder entender con exactitud el funcionamiento de este proyecto. Se han creado unos diagramas con la intención de facilitar el entendimiento de este orden.

En primer lugar se debe ejecutar la estrategia de *Setup*, para poder preparar el entorno de ejecución y no obtener ningún problema en la ejecución de las demás estrategias. Esto se puede visualizar en el diagrama de la imagen 58.

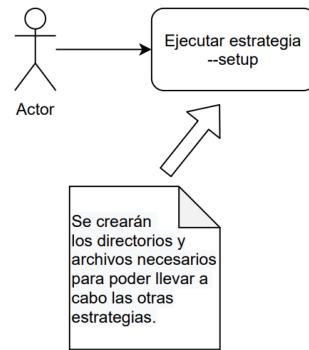


Figura 58: Secuencia que el usuario debe seguir para poder ejecutar la estrategia *Setup* del proyecto.

En el caso de que el usuario no sepa que estrategia se debe utilizar o los parámetros que se deben introducir se puede ejecutar la estrategia *Help*. Esta es independiente de todas las demás, se puede ejecutar en cualquier momento, como se muestra en la imagen 59.

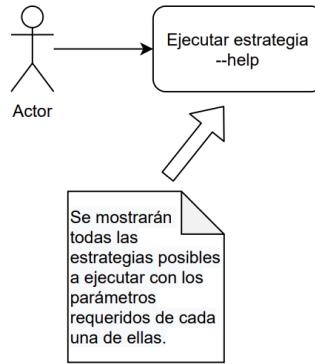


Figura 59: Secuencia que el usuario debe seguir para poder ejecutar la estrategia *Help* del proyecto.

Cuando se quiera realizar una predicción con un clasificador basado en un árbol de decisión se debe seguir el diagrama mostrado en la imagen 60. No hace falta que las ejecuciones de las estrategias mostradas en esta imagen se realicen en el preciso momento en el que se finaliza la estrategia anterior, lo que importa es el orden de esta secuencia. Se puede ejecutar más de una vez la misma estrategia, siempre y cuando se tenga en cuenta que no se solapen los archivos generados al finalizarlas, ya que se perderían los resultados obtenidos de las anteriores ejecuciones.

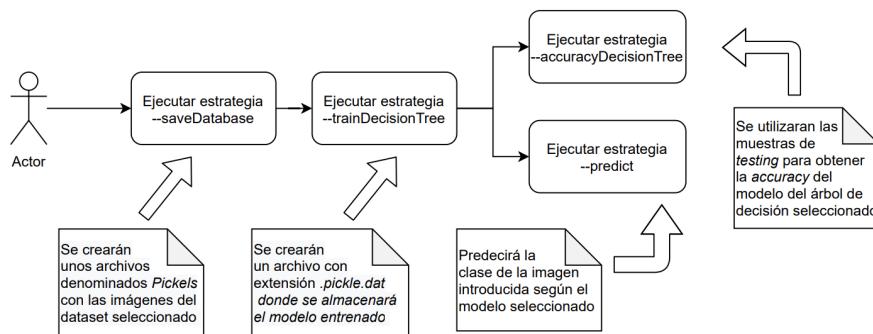


Figura 60: Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en un árbol de decisión.

La imagen 61 muestra el orden de la secuencia requerida para poder entrenar una red neuronal, artificial o convolucional, multi-clase. Se puede percibir como sigue una secuencia similar a la de la imagen 60, el motivo es que se ha intentado seguir un mismo patrón con la intención de facilitarle su comprensión

al usuario.

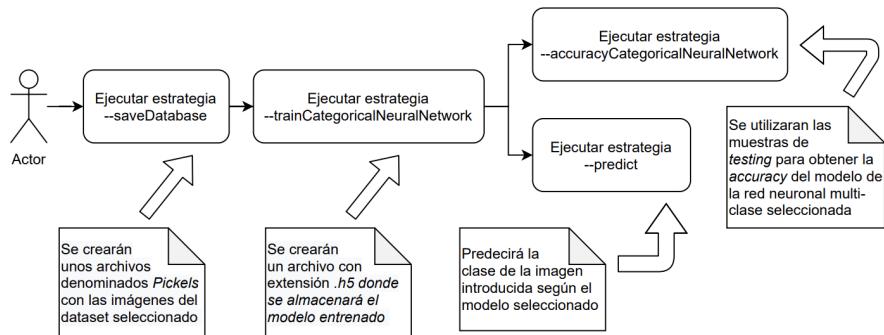


Figura 61: Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en una red neuronal, ya sea convolucional o artificial, con datos multi-clase.

Si se busca hacer una predicción de la red neuronal convolucional binaria con una estrategia *One vs All*, se debe seguir la secuencia mostrada en la imagen 62.

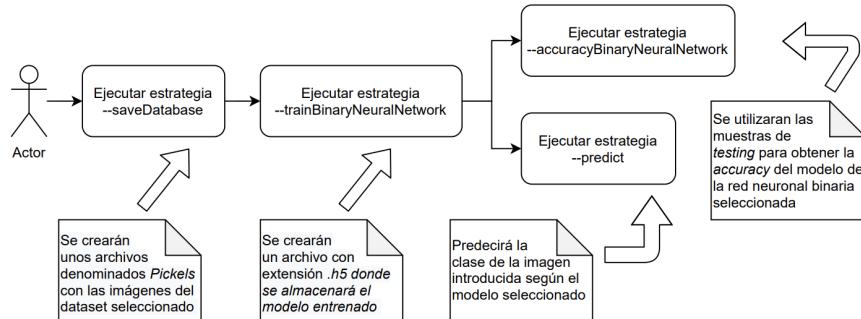


Figura 62: Secuencia que el usuario debe seguir para poder entrenar y predecir un clasificador basado en una red neuronal convolucional binaria con una estrategia *One vs All*.

Por último, si se busca optimizar la red neuronal convolucional se debe seguir la secuencia mostrada en la imagen 63.

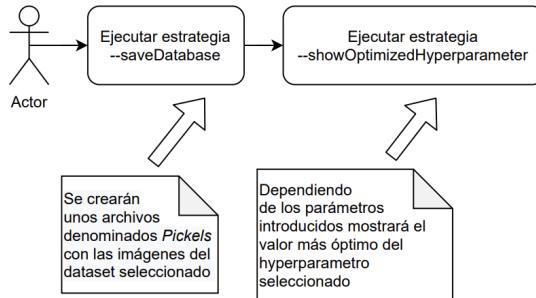


Figura 63: Secuencia que el usuario debe seguir para poder ejecutar la estrategia *ShowOptimizedHyperparameter* del proyecto.

Se pueden combinar las secuencias mostradas, como es en el caso de la ejecución de la estrategia *SaveDatabase*. Esta es requerida en la mayoría de secuencias, y el archivo que se genera al ejecutarla se puede utilizar en todas las otras estrategias que la proceden.

#### 8.0.3. Diagrama de clases UML

Se ha diseñado un diagrama de clases UML siguiendo una perspectiva conceptual. Este se puede encontrar en el anexo A.2.

Uno de los patrones de diseño utilizados que más destaca en este proyecto es el *Factory Method*. Este es un patrón de diseño creacional que permite disponer de una interfaz que será heredada por diversas clases. Una clase padre será la encargada de elegir cual subclase implementará la función buscada. Se ha utilizado para el diseño de las implementaciones de las estrategias, se puede visualizar mejor en la imagen 68 situada en los anexos.

El código se divide en seis paquetes:

- **Las estructuras:** en las estructuras se encuentran todos los modelos predictivos implementados. Este contiene dos subpaquetes: el del árbol de decisión y el de la red neuronal.

A partir de interfaces y enumeradores, se ha conseguido que ambos paquetes muestren una estructura similar y sin tener implementaciones de

las diversas funcionalidades repetidas.

La imagen 67 del anexo A.2.1 muestra una imagen ampliada de este paquete.

- **Las estrategias:** está implementado siguiendo el patrón de diseño creacional *Factory Method*. Una clase padre, denominada: *StrategyFactory* se encarga de instanciar todas las otras estrategias. Hay una clase por cada estrategia implementada en este proyecto.

La imagen 68 del anexo A.2.2 muestra una imagen ampliada de este paquete.

- **El modelo:** contiene las muestras que se utilizan en los entrenamientos y predicciones de los diferentes clasificadores. Es el que se encarga de leer el dataset, transfórmalo en el tamaño y escala de colores correcta y guardarla, posteriormente en los archivos *Pickles*.

Es capaz de leer las imágenes de los directorios que contienen los diferentes datasets, o de los archivos *Pickles*.

La imagen 69 del anexo A.2.3 muestra una imagen ampliada de este paquete.

- **El controlador de almacenamiento:** dispone de un controlador que tiene como objetivo crear y eliminar los directorios utilizados durante las diferentes ejecuciones. Mayoritariamente es utilizado en la estrategia de *Setup*.

La imagen 70 del anexo A.2.4 muestra una imagen ampliada que contiene este paquete.

- **Las excepciones:** hay diversas excepciones implementadas en este proyecto. Este paquete contiene las clases creadas que heredan de la interfaz *Exception* que tienen como objetivo definir con más exactitud el error que se ha generado.

La imagen 70 del anexo A.2.4 muestra una imagen ampliada que contiene este paquete.

- **El registrador:** este paquete contiene la clase que muestra por el terminal toda la información del proceso de la ejecución. Todas las estrategias

tienen la misma instancia de esta clase, con el objetivo que esta sea la única que se comunica con el terminal.

La imagen 70 del anexo A.2.4 muestra una imagen ampliada que contiene este paquete.

Como se ha mencionado, el diagrama tiene una perspectiva conceptual, su intención es poder entender la estructura del proyecto, pero hay muchas relaciones, atributos, métodos... que no han sido incluidos ya que supondrían una mayor densidad sin aportar la información más relevante.

En el caso de que no sea posible apreciar con exactitud el diagrama de clases mostrado en la imagen A.2, el archivo que contiene este diagrama se puede encontrar en el repositorio de *Github*.

## 9. Costes del proyecto

Durante la realización de este proyecto no ha habido ningún coste económico significativo, pero si que se han invertido una gran cantidad de horas.

Han habido muchas ejecuciones que han tardado días en finalizar, si es que llegaban a finalizar. Durante este periodo no se podía continuar avanzando en el proyecto, debido a que el dispositivo, donde se estaba realizando la ejecución, no podía llevar a cabo ninguna otra tarea.

Se intentaba que estos periodos de tiempo coincidiesen en momentos del día en el que no se estaba trabajando, como, por ejemplo, durante la noche. Aún así, las ejecuciones que duraban más de siete horas, se interponían en el proceso de realización del proyecto.

Las ejecuciones de una larga duración necesitaban un control constante, ya que podían ser interrumpidas, ya sea por un error o por falta de espacio en el dispositivo.

En total, teniendo en cuenta el tiempo invertido: en la búsqueda del proyecto a realizar, en el aprendizaje de los fundamentos, en la implementación de los diferentes experimentos y en la realización de la documentación, se calcula que se han invertido alrededor de una 410 horas. Estos tiempos se muestran en la imagen 64.

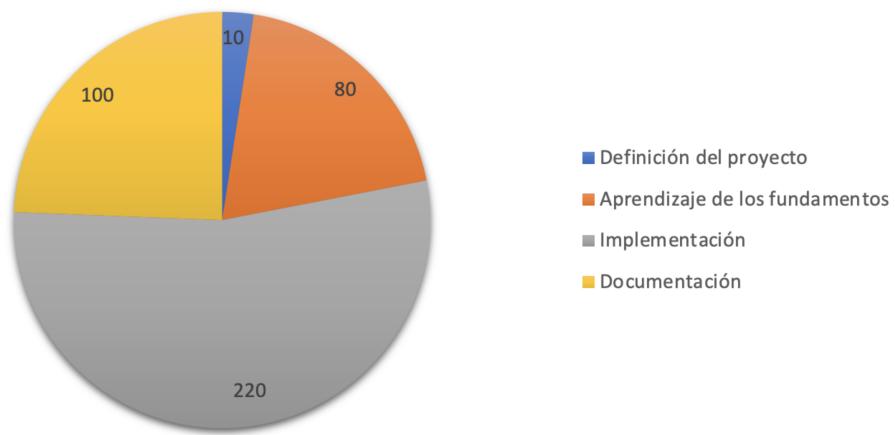


Figura 64: Tiempo, en horas, invertido en el proyecto. Teniendo en cuenta: la definición del proyecto, el aprendizaje de los fundamentos, la implementación y la documentación.

## 10. Conclusiones y líneas de futuro

### 10.1. Conclusiones

En este proyecto se ha realizado un reconocimiento del lenguaje de signos (SLR), implementando diversos tipos de modelos predictivos y comparando sus resultados.

Las muestras, con las que se han procesado estas predicciones, han sido imágenes de gestos del lenguaje de signos americano (ASL). En total se ha dispuesto de treinta y ocho gestos estáticos, los cuales representan el abecedario inglés y los valores numéricos del cero al diez.

Se han utilizado dos datasets, ambos obtenidos de la comunidad de datos *Kaggle*. Se han buscado datasets que contuviesen imágenes realizadas con diferentes perspectivas, con la intención de poder expandir el dominio de cada signo. En total, se ha trabajado con 66.938 muestras, 1.801 imágenes por signo.

Se han implementado dos tipos de estructuras en el modelo predictivo, el árbol de decisión y las redes neuronales. Se han evaluado los resultados obtenidos con cada uno de ellos, contemplando tres parámetros: su rendimiento, su coste temporal y su coste de memoria. Se ha querido buscar el modelo predictivo que aportase la mejor *accuracy* con unos costes viables para las condiciones en las que se ha realizado el proyecto.

El árbol de decisión utiliza el algoritmo *Extreme Gradient Boosting* y se obtuvo la máxima *accuracy* cuando únicamente se entrenaba con las imágenes de un dataset, una proporción de 1.050 imágenes por signo. Al querer utilizar dos datasets en un mismo entrenamiento, su coste temporal se incrementó hasta tal punto, que no se pudo finalizar su ejecución.

Por otro lado, se implementaron dos tipos de redes neuronales: una artificial y otra convolucional. Ambas se entrenaron con todas las combinaciones de datasets posibles. La red neuronal artificial obtuvo su máxima *accuracy*, de un 72,27 %, siendo entrenada con un dataset. Al querer incrementar el número de muestras, su *accuracy* se redujo a un 2,88 %. Aún teniendo unos costes de tiempo y de memoria leves, no disponía del mínimo rendimiento viable como para continuar viéndola como una posible propuesta final.

La red neuronal convolucional empezó dando un rendimiento muy leve, de alrededor de un 34,46 %, cuando únicamente se entrenaba con las muestras de un dataset. Al aumentar el número de muestras, este valor se incrementó hasta un 99,30 %. La razón de este incremento es que la red neuronal convolucional

tiene un mayor rendimiento cuando se dispone de un alto número de muestras. A pesar de haber obtenido una *accuracy* tan elevada, se intentó mejorar este valor. Se realizaron una serie de experimentaciones con la intención de modificar su estructura para poder adaptarla a los requisitos de los datasets utilizados. Estas optimizaciones consiguieron incrementar su rendimiento a un 99,52 %, presentando unos costes de memoria y de tiempo viables.

Para finalizar, se decidió probar un modelo predictivo que dispusiese de un clasificador basado en una red neuronal convolucional binaria, con una estrategia One vs All, ya que hasta el momento se había utilizado una estrategia multi-clase. El coste de tiempo fue demasiado elevado como para poder probarlo con todo el conjunto de muestras, por lo que se entrenó con un máximo de once clases. Los resultados no fueron muy satisfactorios, ya que su *accuracy* general fue de un 21.87 %. Pero se debe tener en cuenta que la mayoría de los clasificadores obtuvo un *accuracy* individual superior al 90 %.

Finalmente, se concluyó que el modelo predictivo que mejor resultados aporataba era el implementado a base de una red neuronal convolucional multi-clase. Con este se obtenía una *accuracy* del 99,52 % con unos costes viables para las condiciones de ejecución del proyecto.

## 10.2. Líneas de futuro

En futuras investigaciones sería interesante poder realizar las ejecuciones en un entorno diferente, a ser posible en un entorno de ejecución explosiva. Muchos de los modelos predictivos implementados son muy prometedores, pero las condiciones de ejecución actuales no permiten poder utilizarlos. Puede que en un entorno diferente, la propuesta final varíe.

También sería interesante poder reducir el coste de memoria, hasta el momento no ha sido un impedimento para la realización del proyecto. Es debido a esto por lo que se ha preferido invertir el tiempo en probar nuevos modelos predictivos, en lugar de mejorar el coste de memoria que presentaban los ya implementados.

Se ha utilizado un dataset con 38 gestos diferentes, pero todos ellos son estáticos. Si se quiere hacer uso del modelo predictivo utilizado, se debería deletrear cada palabra, siendo incómodo para el usuario. Se podría ampliar el número de clases para poder incluir los gestos dinámicos. Se tendrían que hacer varias modificaciones al modelo actual para poder tener en cuenta la temporalidad de los gestos.

Por último, en la experimentación del apartado 5.6, se demostraba como una red neuronal convolucional presentaba un mayor rendimiento si disponía de una capa de *Dropout* con un índice igual a cero. No se ha seleccionado esta estructura para la propuesta final ya que es incoherente, y según la información que se ha encontrado, esto puede ser debido a un *bug* de Tensorflow. Sería conveniente continuar con esta investigación, con la intención de averiguar si realmente se puede utilizar esta estructura en la propuesta final.

Este proyecto me ha permitido profundizar en las técnicas de Machine Learning y Deep Learning. Mis conocimientos de este sector eran muy básicos al empezarlo, pero a medida que he ido implementando diferentes estructuras y estrategias de modelos predictivos, he podido ampliar mis conocimientos. Aún tengo mucho por aprender, pero cada una de las realización de estas experimentaciones me ha enseñado a entender mejor los conceptos que antes desconocía.

## A. Anexo

### A.1. Proyecto Github

Tablero de Kanban situado en el repositorio <sup>19</sup> de Github del proyecto actual. Dispone de cuatro columnas: *To do*, *In progress*, *Documentation* y *Blocked in documentation*.

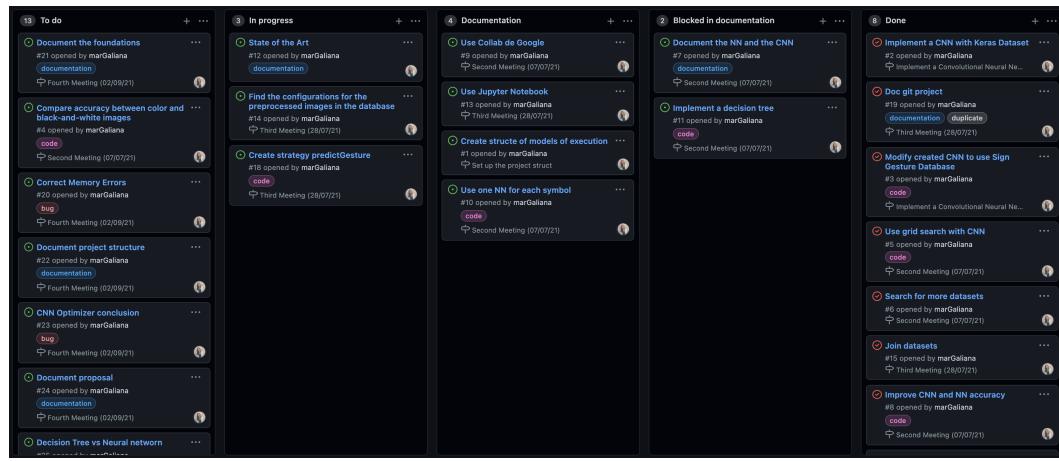


Figura 65: Estructura del proyecto del repositorio de *Github* donde se encuentra el proyecto realizado.

<sup>19</sup><https://github.com/marGaliana/SignLanguageProcessing.git>

## A.2. Diagrama de clases UML

Diagrama de clases UML conceptual del proyecto realizado en Python.

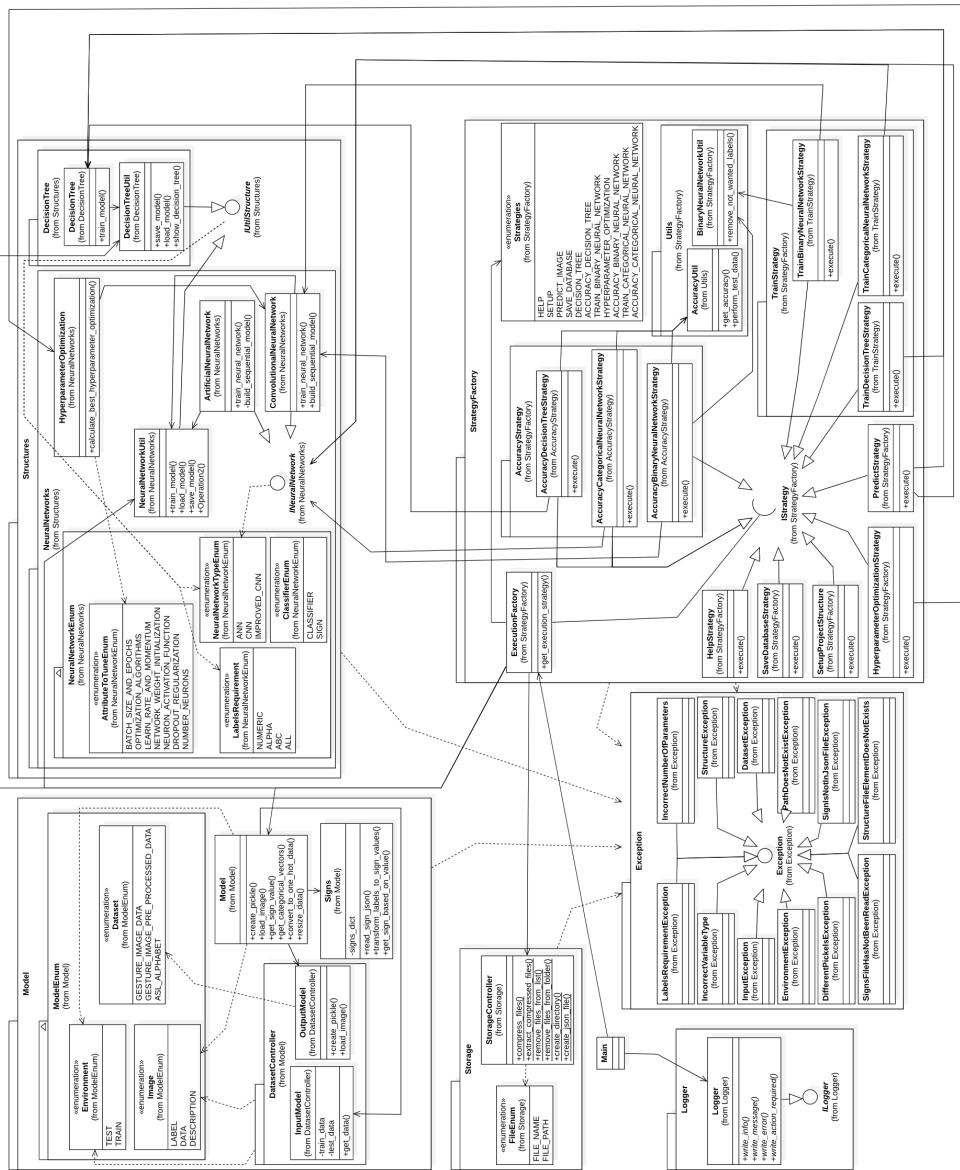


Figura 66: Diagrama de clases UML siguiendo una perspectiva conceptual donde se visualiza la estructura del proyecto.

### A.2.1. Ampliación de las estructuras

Ampliación del paquete de estructuras del diagrama de clases conceptual UML.

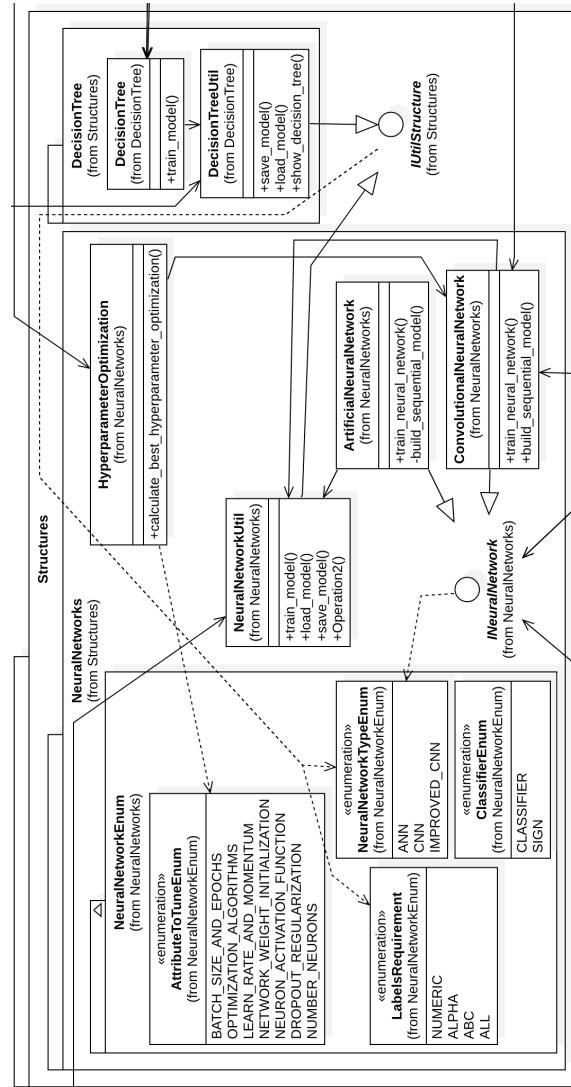


Figura 67: Ampliación del paquete *Structures* del diagrama de clases mostrado en la imagen 66.

### A.2.2. Ampliación de las estrategias

Ampliación del paquete de *Factory Strategy* del diagrama de clases conceptual UML.

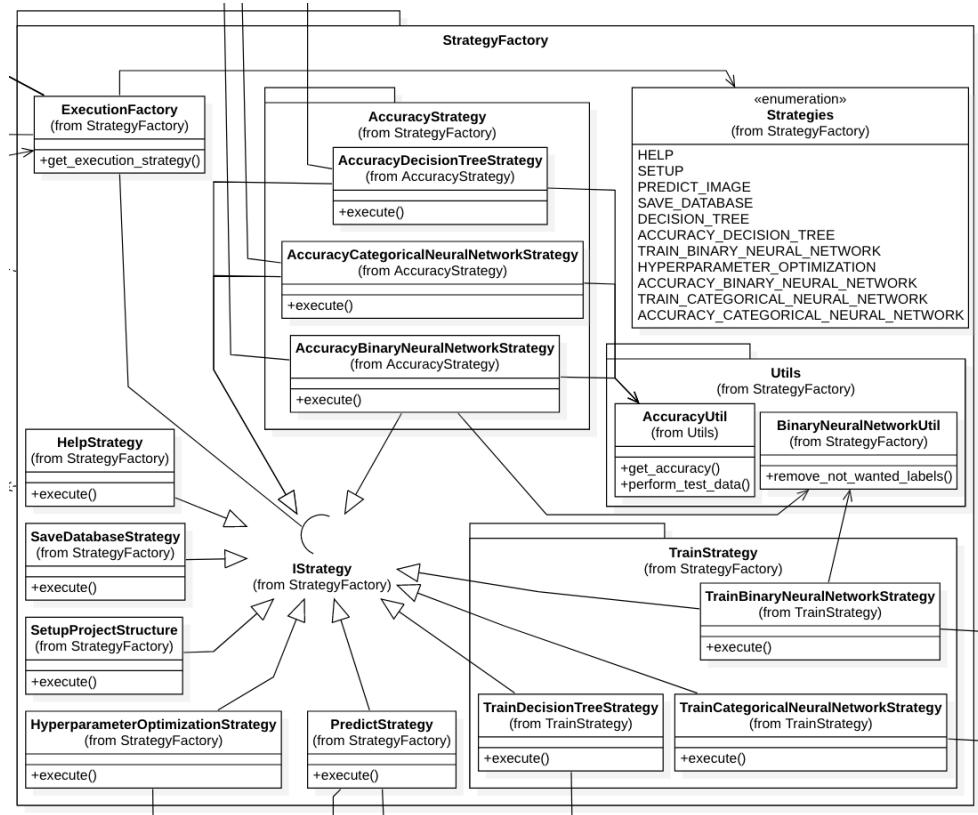


Figura 68: Ampliación del paquete *StrategyFactory* del diagrama de clases mostrado en la imagen 66.

### A.2.3. Ampliación del modelo

Ampliación del paquete de *Model* del diagrama de clases conceptual UML.

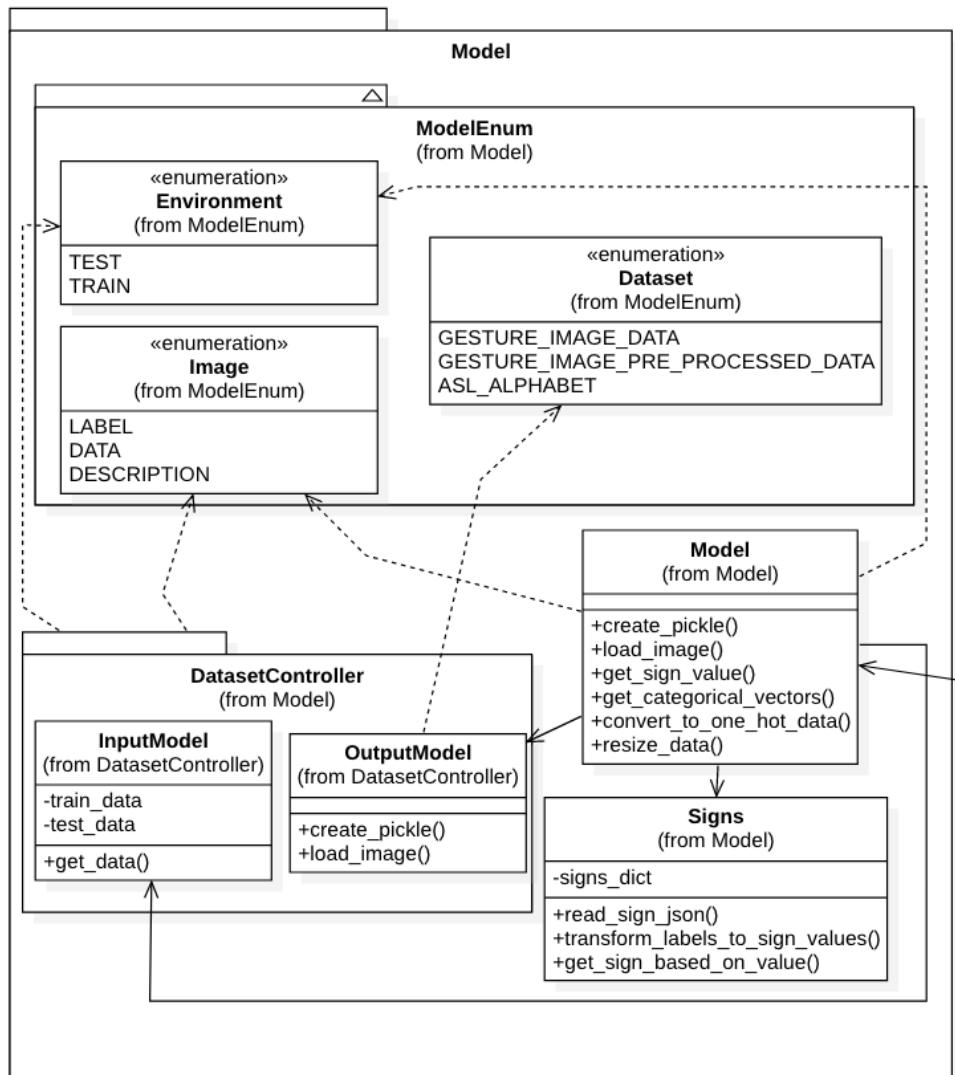


Figura 69: Ampliación del paquete *Model* del diagrama de clases mostrado en la imagen 66.

#### A.2.4. Ampliación de las excepciones, almacenamiento y registrador

Ampliación de los paquetes de *Storage*, *Exception* y *Logger* del diagrama de clases conceptual UML.

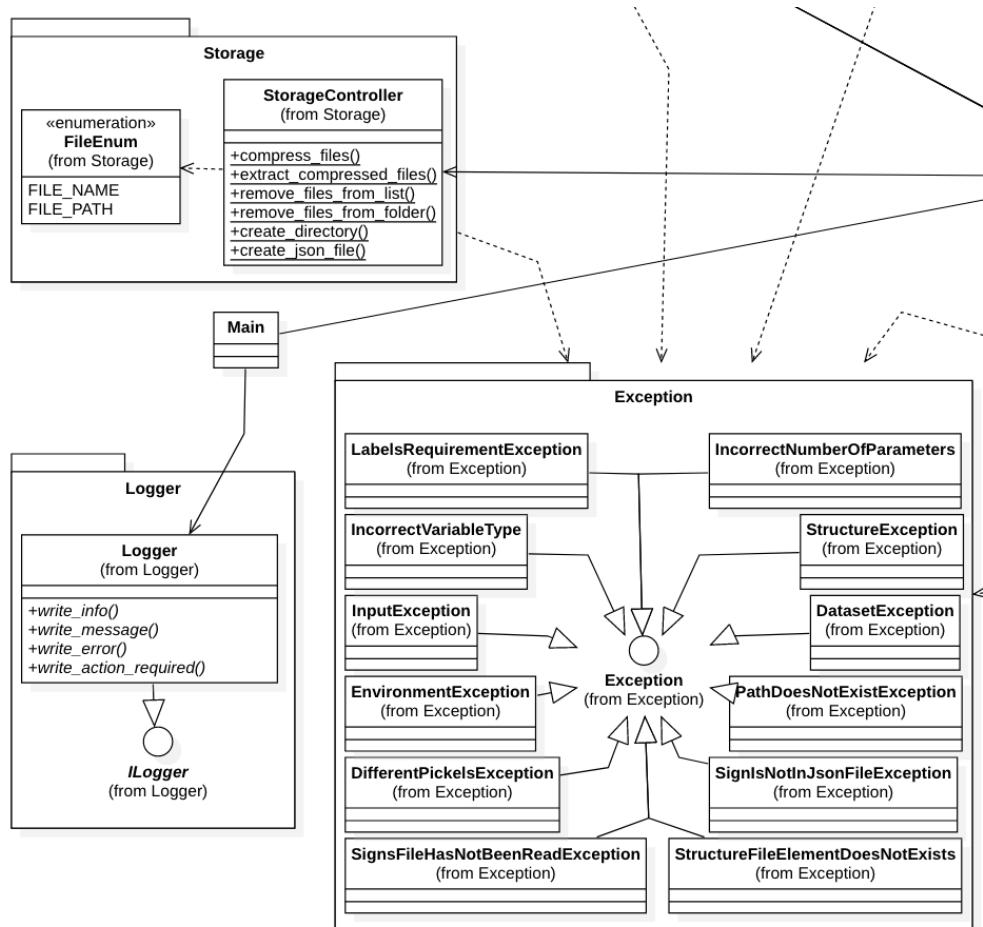


Figura 70: Ampliación de los paquetes *Storage*, *Exception* y *Logger* del diagrama de clases mostrado en la imagen 66.

## Referencias

- [1] Jiangbin Zheng, Zheng Zhao, Min Chen, Jing Chen, Chong Wu, Yidong Chen, Xiaodong Shi, and Yiqi Tong. An improved sign language translation model with explainable adaptations for processing long sign sentences. In *Computational Intelligence and Neuroscience*, volume 2020, page 11, 2020.
- [2] Taehwan Kim, Gregory Shakhnarovich, and Karen Livescu. Fingerspelling recognition with semi-markov conditional random fields. *2013 IEEE International Conference on Computer Vision*, pages 1521–1528, 2013.
- [3] Thad Starner, Joshua Weaver, and Alex Pentland. Real-time american sign language recognition using desk and wearable computer based video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20:1371 – 1375, 01 1999.
- [4] Matteo Rinalduzzi, Alessio Angelis, Francesco Santoni, Emanuele Buchicchio, A. Moschitta, Paolo Carbone, Paolo Bellitti, and Mauro Serpelloni. Gesture recognition of sign language alphabet using a magnetic positioning system. *Applied Sciences*, 11:5594, 06 2021.
- [5] Mark Borg and Kenneth Camilleri. Phonologically-meaningful subunits for deep learning-based sign language recognition. In *ECCV Workshops*, pages 199–217, 01 2020.
- [6] J.M. Benitez, J.L. Castro, and I. Requena. Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks*, 8(5):1156–1164, 1997.
- [7] Wei-Yin Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1:14 – 23, 01 2011.
- [8] Himani Sharma and Sunil Kumar. A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*, 5, 04 2016.
- [9] Ahmed Khan. Sign language gesture images dataset. <https://www.kaggle.com/ahmedkhanak1995/sign-language-gesture-images-dataset>, 2019.
- [10] Kareem. Dataset asl test and train. <https://www.kaggle.com/kareemalaa74/dataset-asl-test-and-train>, 2021.
- [11] Alexey Karpov, Irina Kipyatkova, and Milos Zelezny. Automatic technologies for processing spoken sign languages. *Procedia Computer Science*, 81:201–207, 2016.
- [12] Gerald Berent. The syntax of american sign language: Functional categories and hierarchical structure (review). *Language*, 77:839–842, 01 2001.

- [13] Josep Quer and Markus Steinbach. Handling sign language data: The impact of modality. *Frontiers in Psychology*, 10:483, 2019.
- [14] Brita Bergam. Elisabeth engberg-pedersen, space in danish sign language. the semantics and morphosyntax of the use of space in a visual language. international studies on sign language research and communication of the deaf, vol. 19. hamburg: Signum verlag, 1993. 406 pp. *Nordic Journal of Linguistics*, 18(1):67–74, 1995.
- [15] Adam Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004.
- [16] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. Multimodal human discourse: Gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9(3):171–193, September 2002.
- [17] Scott K. Liddell. *American Sign Language Syntax*. De Gruyter Mouton, 2021.
- [18] Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. Sign language recognition using convolutional neural networks. In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, editors, *Computer Vision - ECCV 2014 Workshops*, pages 572–578, Cham, 2015. Springer International Publishing.
- [19] John Bush Idoko. Deep learning based sign language translation system. *NICOSIA*, 1:1–109, 2016.
- [20] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.
- [21] Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Al-sulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, Hassan Mathkour, and Mohamed Amine Mekhtiche. Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. *IEEE Access*, 8:192527–192542, 2020.
- [22] Danilo Avola, Marco Bernardi, Luigi Cinque, Gian Luca Foresti, and Cristiano Massaroni. Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. *IEEE Transactions on Multimedia*, 21(1):234–245, Jan 2019.
- [23] Giuseppe Bonacorso. *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Packt Publishing, 2017.

- [24] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fmri: A tutorial overview. *NeuroImage*, 45(1, Supplement 1):S199–S209, 2009. Mathematics in Brain Imaging.
- [25] Francis Quintal Lauzon. An introduction to deep learning. In *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pages 1438–1439, 2012.
- [26] Yagang Zhang. *New Advances in Machine Learning*. ItechOpen, 2017.
- [27] King-Shy Goh, Edward Chang, and Kwang-Ting Cheng. Svm binary classifier ensembles for image classification. In *Proceedings of the Tenth International Conference on Information and Knowledge Management, CIKM '01*, page 395–402, New York, NY, USA, 2001. Association for Computing Machinery.
- [28] Nicolás García-Pedrajas and Domingo Ortiz-Boyer. An empirical study of binary classifier fusion methods for multiclass classification. *Information Fusion*, 12(2):111–130, 2011.
- [29] Zhao Zhang, Weiming Jiang, Jie Qin, Li Zhang, Fanzhang Li, Min Zhang, and Shuicheng Yan. Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3798–3814, 2018.
- [30] Seokho Kang, Sungzoon Cho, and Pilsung Kang. Constructing a multi-class classifier using one-against-one approach with different binary classifiers. *Neurocomputing*, 149:677–682, 2015.
- [31] Eduardo Gonçalves, Alexandre Plastino, and Alex Freitas. A genetic algorithm for optimizing the label ordering in multi-label classifier chains. In *A Genetic Algorithm for Optimizing the Label Ordering in Multi-label Classifier Chains*, pages 469–476, 11 2013.
- [32] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. *Mining Multi-label Data*, pages 667–685. Springer US, Boston, MA, 2010.
- [33] André C. P. L. F. de Carvalho and Alex A. Freitas. *A Tutorial on Multi-label Classification Techniques*, pages 177–195. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [34] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5375–5384, 2016.
- [35] Yuchun Tang, Yan-Qing Zhang, Nitesh V. Chawla, and Sven Krasser. Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288, 2009.

- [36] Kevin Gurney. *An Introduction to Neural Networks*. Taylor and Francis, Inc., USA, 1997.
- [37] Erkam Guresen and Gulgun Kayakutlu. Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, 3:426–433, 2011. World Conference on Information Technology.
- [38] B. Muller, J. Reinhardt, and M.T. Strickland. *Neural Networks an Introduction*. Springer-Verlag, Berlin, 1995.
- [39] Forest Agostinelli, Matthew Hoffman, Peter Sadowski, and Pierre Baldi. Learning activation functions to improve deep neural networks, 2015.
- [40] Stuart Geman, Elie Bienenstock, and René Doursat. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1):1–58, 01 1992.
- [41] Anil Kumar Lohani, N. Goel, and Karan Bhatia. Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modelling under different input domains. *Hydrological Processes*, 25:175 – 193, 01 2011.
- [42] R. Aggarwal and Yonghua Song. Artificial neural networks in power systems. iii. examples of applications in power systems. *Power Engineering Journal*, 12(6):279–287, 1998.
- [43] Phil Kim. *Convolutional Neural Network*, pages 121–147. Apress, Berkeley, CA, 2017.
- [44] Thomas Laurent and James von Brecht. A recurrent neural network without chaos, 2016.
- [45] LearningDecisionTreeClassifiers. Learning decision tree classifiers. *ACM Comput. Surv.*, 28(1):71–72, March 1996.
- [46] Omer Sagi and Lior Rokach. Approximating xgboost with an interpretable decision tree. *Information Sciences*, 572:522–542, 2021.
- [47] Rory Mitchell, Andrey Adinets, Thejaswi Rao, and Eibe Frank. Xgboost: Scalable gpu accelerated learning, 2018.
- [48] Rajesh Kumar. Malware classification using xgboost-gradient boosted decision tree. *Advances in Science Technology and Engineering Systems Journal*, 5:536–549, 09 2020.
- [49] Jason Brownlee. Grid search hyperparameters deep learning model python keras. <https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/>, August 2016.
- [50] Jason Brownlee. Weight initialization for deep learning neural networks. <https://machinelearningmastery.com/weight-initialization-for-deep-learning-neural-networks/>, February 2021.

- [51] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

