

# Plataforma en la nube para la clasificación de objetos utilizando inteligencia artificial

Cloud platform for object classification using artificial intelligence

 Mauricio Marcelo Peredo Claros<sup>1</sup>,  Edgar Ramos Silvestre<sup>2</sup>

<sup>1</sup> Director del Departamento Académico de Biomédica, Universidad Privada del Valle, Cochabamba, Bolivia. [mperedoc@univalle.edu](mailto:mperedoc@univalle.edu)

<sup>2</sup> Director del Departamento Académico de Electrónica y Telecomunicaciones, Universidad Privada del Valle, Cochabamba, Bolivia. [eramoss@univalle.edu](mailto:eramoss@univalle.edu)

## RESUMEN

La Inteligencia Artificial (IA) es una de las áreas que más interés atrae en el ámbito de las Ciencias de la Computación, tanto desde el punto de vista científico-académico como del de sus múltiples aplicaciones en distintos sectores de actividad. A lo largo de la historia de la Inteligencia Artificial han surgido desarrollos y aplicaciones que la han convertido en un área de conocimiento consolidada e interés económico y social. Sin embargo, durante los últimos años, la IA ha experimentado un desarrollo excepcional, motivado por la aparición de tecnologías que han supuesto un gran avance en la disciplina y por la disponibilidad de recursos en hardware que han hecho viable su aplicación en distintos dominios.

El interés de la sociedad en la Inteligencia Artificial ha crecido de forma proporcional a este desarrollo tecnológico. Prueba de ello es que gobiernos y administraciones en todos los niveles del Estado en diferentes países alrededor del mundo han promovido multitud de programas para financiar actividades de investigación, desarrollo e innovación en IA. Además, son muchas las empresas que han decidido llevar a cabo inversiones en IA para aplicar esta tecnología en distintos puntos de su cadena de valor.

Aunque no existe una definición oficial y única de la Inteligencia Artificial todas ellas concuerdan en señalar que se trata de un campo de las Ciencias de la Computación que persigue desarrollar sistemas informáticos que exhiban características intelectuales similares a los humanos tales como la capacidad de razonar, aprender, generalizar, solucionar problemas, percibir y utilizar lenguaje natural, en resumen “imitar a los humanos”

Este artículo está centrado en explicar el desarrollo de una plataforma en la nube para la clasificación de objetos utilizando Inteligencia Artificial, su principal aplicación será la de incrementar el conocimiento referente a este campo y coadyuvar en diferentes soluciones que se puedan plantear durante el uso de esta.

**Palabras clave:** Inteligencia Artificial. Plataforma en la nube. Ciencias de la computación.

**Citar como:** Peredo Claros, M. M., & Ramos Silvestre, E. Plataforma en la nube para la clasificación de objetos utilizando inteligencia artificial. Journal Boliviano de Ciencias, UNIVALLE, 18(53), 26-47. <https://doi.org/10.52428/20758944.v18i53.253>

**Revisado:** 28/11/2022

**Aceptado:** 01/12/2022

**Publicado:** 30/12/2022

**Declaración:** Los autores declaran no tener ningún conflicto de intereses en la publicación de este documento.

**Fuentes de financiamiento:** No hubo fuentes de financiamiento.

Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de la Creative Commons. Licencia de atribución (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

Derechos de autor 2022  
Mauricio Peredo, Edgar Ramos.



---

**ABSTRACT**

Artificial Intelligence (AI) is one of the areas that attracts the most interest in the field of Computer Science, both from the scientific-academic point of view and from its applications in multiple sectors of activity. Throughout the history of Artificial Intelligence, developments and applications have emerged that have made it a consolidated area of knowledge and of proven economic and social interest. However, in recent years, AI has undergone exceptional development, motivated by the appearance of technologies that have represented a great advance in the discipline and by the availability of hardware resources that have made its application viable in different domains.

Society's interest in Artificial Intelligence has grown proportionally to this technological development. Proof of this is that governments and administrations at all levels of the State in different countries around the world have promoted a multitude of programs to finance research, development and innovation activities in AI. In addition, many companies have decided to invest in AI to apply this technology at different points in their value chain.

Although there is no official and unique definition of Artificial Intelligence, all of them agree that it is a field of Computer Science that seeks to develop computer systems that exhibit intellectual characteristics similar to humans, such as the ability to reason, learn, generalize, solve problems, perceive and use natural language.

This project is focused on proposing a cloud platform for classifying objects using Artificial Intelligence, its main application will be to increase knowledge regarding this field and contribute to different solutions that may arise during its use.

**Keywords:** Artificial Intelligence. Cloud platform. Computer's Science.

**1. INTRODUCCIÓN**

El concepto Inteligencia Artificial fue creado en la década de 1950 por varios investigadores, sin embargo, el término en sí fue acuñado por John McCarthy en 1956 durante el desarrollo de la conferencia de Dartmouth, y lo definió como "la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes". Fueron necesarios 5 años de trabajo previo durante los cuales se habían propuesto definiciones distintas que en ningún caso lograron ser aceptadas en su totalidad por la comunidad investigadora. Finalmente, el término fue propuesto en dicha conferencia por John McCarthy, Marvin Minsky y Claude Shannon. También Alan Turing, considerado uno de los padres de la Inteligencia Artificial, escribió en 1950, su ensayo "Computing Machinery and Intelligence" años antes de esta conferencia y cuando todavía no existía el término "Inteligencia Artificial", este trabajo fue considerado un trabajo seminal en el campo. En él, se puede leer "Propongo que se considere la siguiente pregunta, '¿Pueden pensar las máquinas?'". Turing menciona también que, a finales del siglo XX, una máquina debería ser capaz de pasar una prueba de inteligencia estándar, conocida como el "test de Turing" (Universidad Da Coruña *et al.*, 2021).

De igual manera, (Iglesias *et al.*, 2020) que menciona “La inteligencia artificial es un término que engloba aquellos sistemas informáticos capaces de detectar su entorno, pensar, aprender y tomar decisiones en función de las entradas que reciben y sus objetivos.”

La inteligencia artificial se transformó en un término muy usado en las aplicaciones y tareas muy complejas para el ser humano. Con frecuencia se usa el termino como sus diferentes subdivisiones, de igual manera esto se ve reflejado en una nota del 2019 de la Comisión Económica para América Latina y el Caribe en donde cita:

“Los países latinoamericanos tienen la capacidad para aprovechar todo el potencial de la IA, sin embargo, debido a las limitaciones sociales y económicas, se ha realizado poca inversión en el gobierno, la industria y la investigación para avanzar en IA. Esto es una desventaja, ya que la IA es una tecnología importante y fundamental en la cuarta revolución industrial, y dada su naturaleza multipropósito, poder exponencial y capacidad predictiva podría ser una herramienta importante para abordar diversos desafíos que afectan el desarrollo de la región.” (CEPAL, 2019).

Por lo tanto, el Machine Learning forma parte de la inteligencia artificial, engloba un conjunto de técnicas para el aprendizaje automático, ya sea en organización de datos, reconocimiento de patrones o simplemente en el aprendizaje autónomo de las máquinas. Algunos ejemplos de aplicación pueden ser la clasificación, clustering, regresión y detección de anomalías (Manzanares González, 2018).

El aprendizaje profundo o Deep Learning intentará comprender los conceptos con mayor precisión, analizando los datos a un alto nivel de abstracción. A través de la comprensión no lineal. Su funcionamiento es similar al del cerebro, En una red neuronal, se combinan capas sucesivas de datos para aprender los conceptos (Oracle, 2022).

“Las redes neuronales artificiales son un paradigma de programación inspirado en el funcionamiento de las neuronas y conexiones del cerebro que sirve para procesar información. Con una cierta estructura, estas son capaces de computar cualquier función real” (Nielsen, 2019).

De igual manera, como se menciona en el siguiente párrafo extraído de un proyecto de investigación, se puede establecer que:

"Los modelos matemáticos de las redes neuronales biológicas fueron introducidos por primera vez en 1943 por Warren S. McCulloch y Walter Pitts. Posteriormente, la primera y más básica de las neuronas artificiales fue el perceptrón que conceptualmente se trata de un algoritmo de clasificación binario en el que una función lineal produce una salida binaria (0 o 1) en función del producto de unos pesos por los valores de entrada. Estos fueron los puntos de partida que dieron paso a un posterior desarrollo de

algoritmos relacionados con las redes neuronales, aunque los nuevos retos de la tecnología y la mejora en la capacidad de computación en la última década han provocado un avance más acelerado en esta materia.

En muchos de los problemas complejos de reconocimiento de patrones con redes neuronales, la cantidad de parámetros que se han de ajustar es muy grande, por lo que el tiempo de computación era demasiado elevado en la mayoría de los casos en el pasado. La mejora de las computadoras y el uso de las tarjetas gráficas para hacer cálculos en paralelo han permitido superar esa barrera en tiempo de computación, haciendo que las redes neuronales sirvan de gran utilidad para muchas tareas" (Arriola, 2018).

En el ámbito de las tecnologías de la información y la comunicación (TIC), la inteligencia artificial es una disciplina que permite resolver problemas para los cuales no existe una solución algorítmica factible o la misma es ineficiente o inaplicable. En la actualidad, las soluciones tecnológicas y los proyectos que aprovechan los conocimientos y las técnicas de la inteligencia artificial para resolver problemas de diversa naturaleza son numerosas y arrojan resultados muy alentadores.

Este artículo está centrado en explicar el desarrollo de una plataforma en la nube para la clasificación de objetos utilizando Inteligencia Artificial, su principal aplicación será la de incrementar el conocimiento referente a este campo y coadyuvar en diferentes soluciones que se puedan plantear durante el uso de esta.

## 2. METODOLOGÍA

Durante el proceso de investigación para realizar una clasificación de objetos de forma precisa y obteniendo métricas de rendimiento aceptables para dicho proceso, es necesario establecer la metodología para el diseño de la red neuronal artificial que será la encargada de realizar el proceso de aprendizaje y por ende aprender las características de cada objeto para poder clasificarlo de forma correcta.

Para este cometido es preciso conocer la evolución y generalización del perceptrón, siendo esta una neurona artificial que permite obtener como salida valores no solamente binarios. Las neuronas constan de los siguientes componentes según explica su publicación (Kriesel, 2007):

- Variables de entrada, debe constar al menos de una. Se representan comúnmente en forma de vector de la siguiente manera,

$$x = [x_1, x_2, x_3, \dots, x_N].$$

- Un peso por variable de entrada que determinará la importancia de cada conexión. El vector de pesos se representará utilizando la notación .

$$w = [w_1, w_2, w_3, \dots, w_N].$$

- Un parámetro llamado sesgo "b". El sesgo indicará la facilidad que tiene la neurona de activarse. Este proceso está basado en las neuronas biológicas, las cuales envían un impulso eléctrico solo si a su entrada (constan de un único axón) hay un potencial eléctrico mayor a un potencial umbral. El sesgo sería la inversa a ese potencial umbral. En el ámbito de reconocer patrones, las salidas de las neuronas indican ciertas características de la muestra con la que se trabaja, en este caso los pesos determinan como está relacionada las características con cada entrada, y el sesgo cómo de fácil de detectar es esa característica.

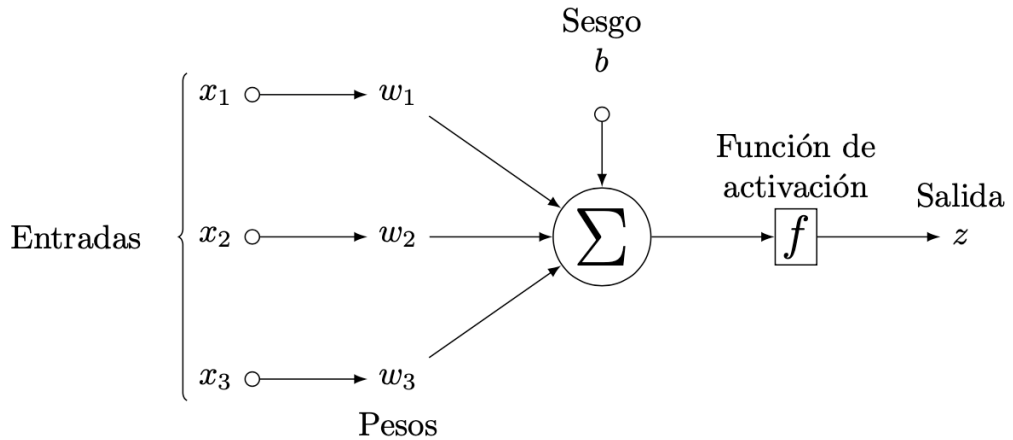
- Una función de propagación que actúa sobre los pesos y el sesgo antes de acceder a la función de activación,  $y = f(x_1, \dots, x_N, w_1, \dots, w_N, b)$ . En el caso del perceptrón, la función de propagación es lineal y tiene la siguiente forma matemática representada en la ecuación (1):

$$y = \sum_{i=1}^N w_i x_i + b \quad (1)$$

- Aunque esta es la función de propagación más comúnmente utilizada, de manera general las funciones implementadas en cada neurona pueden ser no lineales.

- Una función de activación que determinara si se activa la salida de la neurona en función del valor resultante de la función de propagación  $z = f_{act}(y)$ . Entre las funciones de activación más comúnmente utilizadas se encuentran la tangente hiperbólica, la cual ofrece una salida acotada entre -1 y 1, la función sigmoide, una versión escalada de la tangente hiperbólica que ofrece una salida acotada entre 0 y 1 o la unidad lineal rectificadora (ReLU), la cual lleva a cabo la siguiente operación:  $f(y) = \max(0, y)$ .

En la Figura N° 1, se muestra de manera gráfica el esquema de la estructura interna de una neurona artificial.



**Figura N° 1** Representación gráfica de una neurona artificial  
Fuente: Arriola, 2018

Por lo tanto, la salida de una neurona cuya entrada es un vector  $x = [x_1, \dots, x_N]$  está representado por la ecuación (2)

$$z = f_{act} \left( \sum_{i=1}^N w_i x_i + b \right) \quad (2)$$

En este proceso es importante conocer el proceso de convolución en las redes neuronales, o lo que se conoce como redes neuronales convolucionales, se considera a la convolución en el bloque principal para la construcción de una red neuronal convolucional, este término hace referencia a la combinación matemática de dos funciones para poder introducir una tercera, es decir, se encarga de fusionar dos conjuntos de información (Ecuación 3). En el caso particular de estudio (imágenes), se cuenta con las características propias de una imagen como ser el alto y el ancho de esta y de igual manera la cantidad de canales que representan el color (rojo, verde y azul).

$$(I * K)_{i,j} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \sum_{p=0}^C K_{m,n,p} * I_{i+m,j+n,c} + b \quad (3)$$

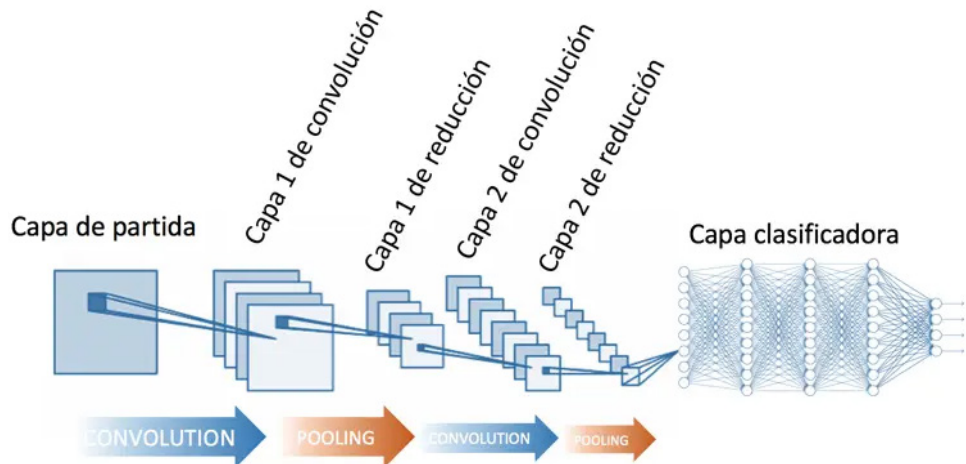
Los modelos de redes neuronales convolucionales tienen ciertas ventajas que sobresalen a las otras arquitecturas de redes neuronales, según menciona (Liu et al., 2020) estas pueden ser:

- Una estructura jerárquica para aprender representaciones de datos con múltiples niveles de abstracción.
- La capacidad de aprender funciones muy complejas.
- Aprender representaciones de características directas en forma automática a partir de datos con un conocimiento mínimo del dominio.

Las redes neuronales convolucionales están compuestas por diferentes filtros o kernels, los cuales se constituyen en un conjunto de parámetros entrenables y que resultan útiles para detectar características como bordes y formas. Un número alto de filtros esencialmente puede llegar a inferir características espaciales de la imagen a analizar en función a los pesos aprendidos en la propagación hacia atrás. Las demás capas convolucionales se pueden usar para la detección de formas aún más complejas a partir de las características aprendidas en otros niveles. Por lo tanto, pueden reducir una imagen en una representación altamente abstracta que es fácil de predecir.

"Las arquitecturas CNNs (Convolutional Neural Network) se pueden encontrar con diversas variantes, sin embargo, en general se componen de capas convolucionales y de submuestreo que se encuentran agrupadas en módulos o bloques convolucionales. Otro elemento de este tipo de modelos son las capas completamente conectadas (densas), tal y como se conocen en las redes neuronales de retroalimentación hacia adelante, colocadas después de los módulos que agrupan varias operaciones de las CNNs. Cuando estos módulos son apilados uno sobre otro da origen a lo que se conoce como un modelo profundo". (Rawat & Wang, 2017)

En la Figura N° 2 se muestra la arquitectura típica de una red neuronal convolucional para la tarea de clasificación de imágenes, en la cual se aprecian los bloques que la conforman.



**Figura N° 2.** Arquitectura típica de una red neuronal en la clasificación de objetos  
Fuente:Clavo, 2017

"La práctica común en la mayoría de los desarrollos recientes de redes profundas es implementar modelos más grandes y profundos para lograr un mejor rendimiento. A medida que el modelo se hace más grande y profundo, los parámetros de la red se incrementan dramáticamente. Como resultado, el modelo se vuelve más complejo de entrenar y en consecuencia es más costoso computacionalmente. Por lo tanto, es muy importante diseñar una arquitectura que proporcione un mejor rendimiento utilizando un número razonablemente menor de parámetros de la red principalmente si este será utilizado en una aplicación que requiere respuestas en tiempo real o se pretende implementar en algún sistema embebido" (García Villanueva & Romero Muñoz, 2020).

### **Recolección de datos**

Para este trabajo se han recopilado diferentes "set de datos" con diferentes imágenes de objetos bajo diferentes características de posición e iluminación, puesto que en la actualidad la adquisición de imágenes resulta una tarea sencilla de realizar, como se resalta en el siguiente párrafo.

"Los conjuntos de datos han jugado un rol clave a través de la historia en la investigación de reconocimiento de objetos, no solamente como una base común para medir y comparar el desempeño de algoritmos, también empujando a esta área de investigación hacia problemas cada vez más complejos y desafiantes. El acceso a una gran cantidad de imágenes en internet permite construir conjuntos de datos completos para capturar una gran riqueza y diversidad de los elementos que componen una escena, lo que permite un rendimiento sin precedentes en el reconocimiento de objetos" (Liu et al., 2020).

Sin embargo, a pesar de que como se menciona anteriormente se tiene gran facilidad para la obtención de imágenes de diferentes objetos mediante el uso de internet, es primordial prestar atención a la construcción del set de datos a utilizar o bien al set de datos que se pueda encontrar, esto principalmente por la diversidad de formas de los objetos, la posición en la que se encuentran en la imagen, el fondo de la imagen, el ruido, etc. Se deben considerar e identificar en los diferentes escenarios para la aplicación en cuestión, para este trabajo de investigación se tomó en cuenta 3 set de datos para los procesos de entrenamiento y validación, el primero obtenido del repositorio del framework a utilizar, el cual contiene imágenes de perros y gatos (Figura N° 3), el segundo resulta en un set de datos de elaboración propia que cuenta con imágenes de plantas medicinales y por último el tercer set de datos que resulta un compendio de imágenes adquiridas de internet que contiene imágenes de personas con barbijo.





**Figura N° 3.** Visualización del dataset de perros y gatos. Fuente: Elaboración propia en base a TensorFlow, 2022.

Estos objetos son visibles para el ojo humano y pueden detectarse automáticamente mediante algoritmos apropiados de procesamiento digital de imágenes, en este proyecto se busca estudiar modelos de redes neuronales para la clasificación automática de objetos.

Como ya se ha mencionado anteriormente, dichas redes han logrado un rendimiento de detección impresionante en varias áreas como la medicina y la industria, y en particular aplicados a la clasificación de objetos como se estudia a lo largo de este trabajo.

Los sets de datos para el desarrollo y pruebas de la plataforma están estructurados de diferentes maneras, sin embargo, se tomó en cuenta el 70% del total de imágenes como set de datos de entrenamiento, el 30% restante se constituye en el set de datos de validación y de igual manera se eligen 30 imágenes obtenidas mediante fotografías o imágenes de internet para un proceso de pruebas. Cada clase según el set de datos tiene diferentes cantidades de imágenes, sin embargo, el porcentaje

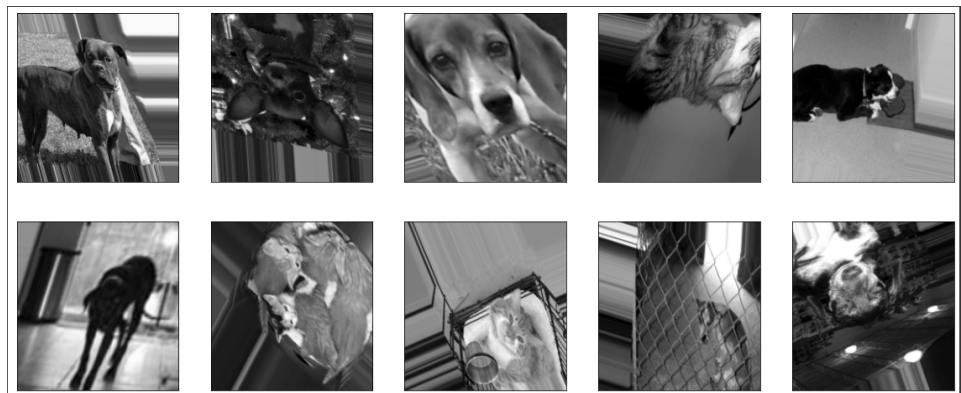
utilizado en el entrenamiento y validación es el mismo, de igual manera con las imágenes de prueba. El detalle de la cantidad de imágenes y clases con las que cuenta cada dataset se describen a continuación en la Tabla 1.

**Tabla N° 1** Cantidad de imágenes y clases de los dataset

DATASET	CLASES	CANTIDAD DE IMÁGENES			TOTAL
		ENTRENAMIENTO	VALIDACION	PRUEBA	
Animales	Perro	8750	3750	100	12600
	Gato	8750	3750	100	12600
Plantas	Manzanilla	351	150	30	531
	Romero	356	152	30	538
	Valeriana	354	152	30	536
Personas	Barbijo	194	83	30	307
	Sin Barbijo	186	80	30	296
	Barbijo incorrecto	160	69	30	259

Fuente: Elaboración propia.

Según la cantidad de imágenes recolectadas para el proceso de entrenamiento se considera necesario un proceso de aumento de datos, este concepto radica en incrementar la cantidad de imágenes ya que al tener mayor cantidad de imágenes el resultado se ve afectado de forma positiva, mejorando considerablemente el resultado de la predicción, con este cometido se utilizan herramientas y librerías del lenguaje de programación Python, con las cuales de forma aleatoria y automática se pueden realizar transformación sobre las imágenes ya presentes en el set de datos y aumentar esas imágenes modificadas como nuevas para incrementar la cantidad total de imágenes, como ejemplo en la Figura N° 4 se puede visualizar los diferentes cambios propuestos.

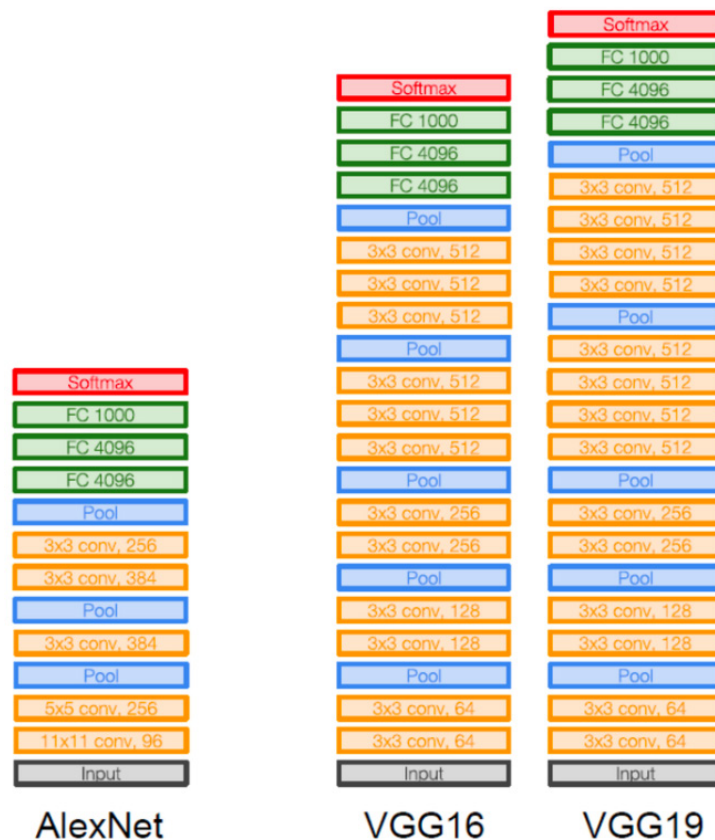


**Figura N° 4.** Imágenes con aumento de datos. Fuente: Elaboración propia en base a TensorFlow, 2022

Como se aprecia en la imagen se cuenta con 7 transformaciones diferentes para complementar los sets de datos con pocas imágenes, este procedimiento se realiza con frecuencia cuando la cantidad de imágenes para el entrenamiento no son las suficientes para contemplar un resultado óptimo esperado, por tal motivo en el presente proyecto se incluye esta opción para el desarrollo y uso de futuros set de datos.

Con el conocimiento referente a los sets de datos, en el siguiente apartado se realizará un análisis para poder visualizar las diferentes clases y tomar en cuenta los parámetros y características propias que tienen las imágenes a utilizar.

Para el desarrollo de una red neuronal para la clasificación se realiza un proceso de verificación de diferentes estructuras posibles y toma en cuenta como punto inicial las redes neuronales pre entrenadas como AlexNet, VGG16 y VGG19, con base a la arquitectura de estas redes, la cual se representa de forma gráfica en la Figura N° 5, se busca establecer una arquitectura de características similares considerando los componentes generales de las redes neuronales de clasificación.



**Figura N° 5.** Redes Neuronales pre entrenadas. Fuente: (Enciclopedia de visión artificial de aprendizaje profundo, 2021)

Con base a los mejores resultados obtenidos para las etapas generales de la red neuronal se determinará las mejores características para la red neuronal propuesta para la plataforma de clasificación.

Se tiene múltiples opciones para poder realizar una red neuronal de clasificación que logre un desempeño óptimo para el objetivo que se quiere alcanzar, como punto de partida se tomara en cuenta el modelo VGG, principalmente por su versatilidad y la facilidad de implementación, esta arquitectura como se muestra en la Figura N° 5 consta de una serie de capas convolucionales y culmina con una serie de capas densas.

Para comenzar con el proceso de análisis de las imágenes se inicia realizado pruebas con redes neuronales con capas convolucionales, para posteriormente evaluar los modelos generados en el proceso de entrenamiento, para este cometido se realizará tres tipos de análisis para completar una arquitectura. Para llegar a la arquitectura adecuada, se estimará el rendimiento de los modelos empleados con base a métricas de precisión (E) del inglés “Accuracy”, definida por la ecuación (4):

$$E = \frac{P_c}{N_T} = \frac{Vp + Vn}{Vp + Vn + Fp + Fn} \quad (4)$$

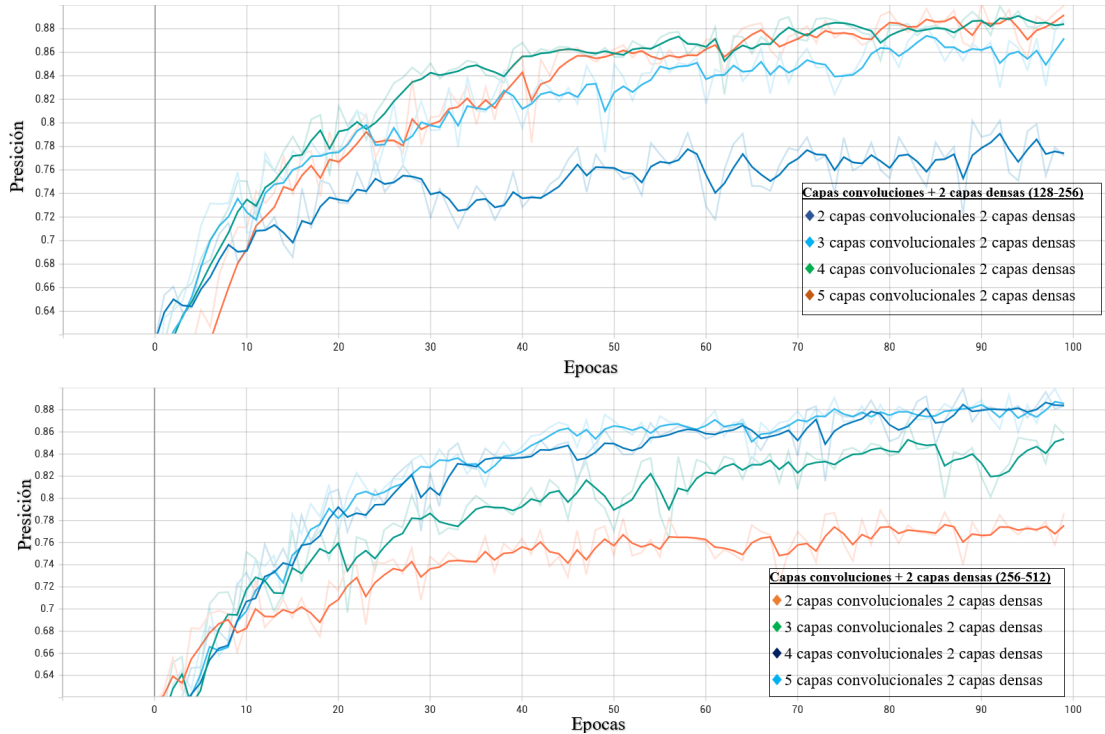
Donde  $P_c$  es la cantidad de predicciones correctas del modelo y el número total de predicciones,  $Vp$  es la cantidad de verdaderos positivos,  $Vn$  es la cantidad de verdaderos negativos,  $Fp$  es la cantidad de falsos positivos y  $Fn$  la cantidad de falsos negativos.

### 3. RESULTADOS

#### Análisis de capas convolucionales

En el proceso de este análisis se considera la capa de entrada de 150 x 150 (imágenes de 150 x 150 píxeles), mediante el empleo de una serie de capas convolucionales se procedió a realizar el entrenamiento y verificar el desempeño de los mejores modelos obtenidos para todas las arquitecturas propuestas. Según la información revisada se considera que, al realizar un incremento en la cantidad de capas de un modelo, se espera también un incremento de la precisión del mismo.

Durante todo este proceso de evaluación se procedió al entrenamiento de los modelos con los tres set de datos planteados para la validación de la plataforma, para tal cometido se obtuvieron modelos para arquitecturas con dos, tres, cuatro y cinco capas convoluciones y dos capas densas al final con diferentes cantidades de neuronas.



**Figura N° 6.** Resultados de clasificar los datos de prueba con los mejores modelos obtenidos durante el proceso de entrenamiento y al incrementar la cantidad de capas convolucionales en la arquitectura. Fuente: Elaboración propia.

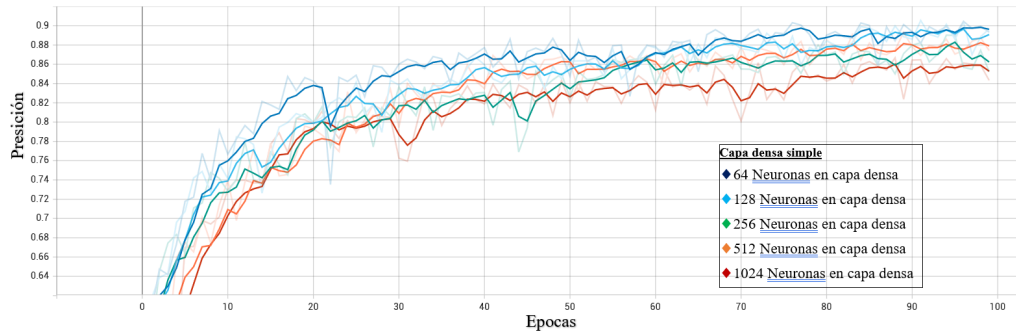
Como se observa en las gráficas anteriores (Figura N° 6), luego de realizar pruebas con los mejores modelos obtenidos en el proceso de entrenamiento, se aprecia que se obtienen los mejores resultados con la arquitectura de 4 capas convolucionales tanto en los modelos que poseen capas densas de 128-256 y 256-512, con base a los resultados obtenidos se tomó en cuenta el modelo con 4 capas convolucionales como parte del modelado final.

### Análisis de capas densas

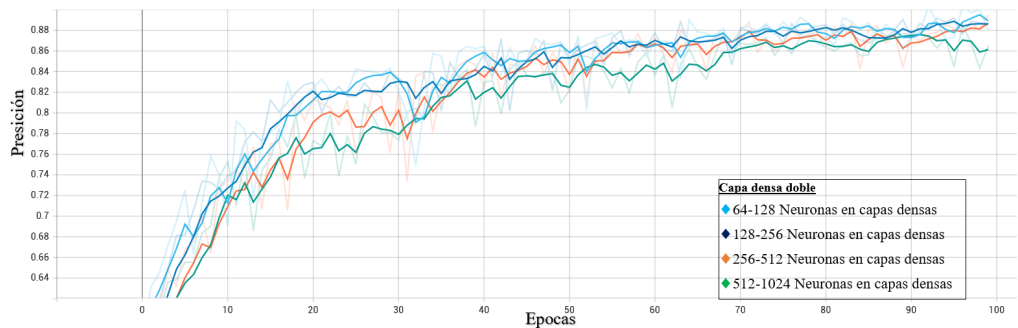
Para este análisis correspondiente a las capas densas a emplear se consideran una y dos capas densas con diferentes cantidades de neuronas, para el proceso de pruebas se utilizó cantidades de 64, 128, 256, 512 y 1024 neuronas para los modelos con una sola capa densa, mientras que para los modelos con dos capas densas se establecieron cantidades de 64 -128, 128 - 256, 256 - 512, 512 - 1024 correspondientes a la primera y segunda capa densa respectivamente.

Las cantidades de neuronas en las capas densas pueden tener cualquier valor, incluso es posible realizar el incremento de uno en uno, esto repercute en el proceso de desarrollo e implementación y se convierte en un proceso poco práctico, debido a la familiaridad de los valores aplicables a potencias de 2, en la presente investigación se establece dicho proceso como patrón para poder determinar la cantidad y la relación de las neuronas de capas densas para el proceso de entrenamiento.

En la Figuras 7 y 8 se presentan los resultados obtenidos para modelos con una capa densa y para modelos con dos capas densas.



**Figura N° 7.** Desempeño utilizando los mejores modelos obtenidos a diferentes cantidades de neuronas en la arquitectura con una capa densa. Fuente: Elaboración propia.



**Figura N° 8.** Desempeño utilizando los mejores modelos obtenidos a diferentes cantidades de neuronas en la arquitectura con dos capas densas. Fuente: Elaboración propia.

En función a las métricas obtenidas en la figuras 6 y 7, se concluye que los mejores modelos obtenidos en primera instancia como capa densa inicial son los modelos de 128 neuronas en la primera capa densa, al realizar pruebas con dos capas densas de salida se toma en cuenta como parte del modelo final el par de 128-256 neuronas en las capas densas respectivamente, ya que demostraron ser el modelo más estable durante las pruebas realizadas.

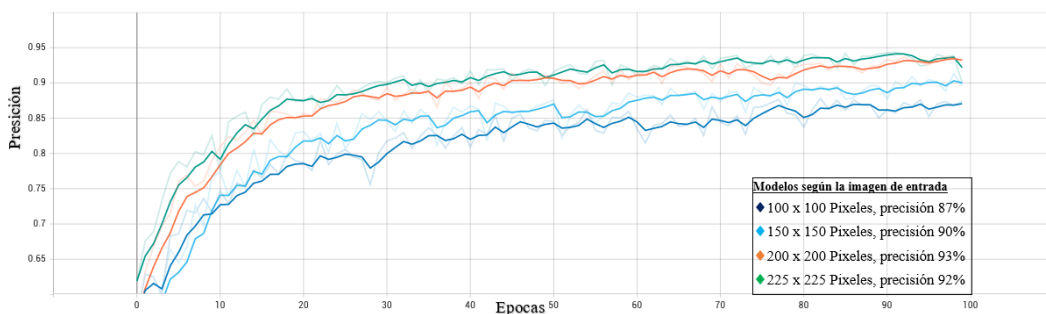
## Análisis de la capa de entrada

Para determinar el tamaño de la capa de entrada del modelo es necesario realizar un análisis del tamaño en píxeles de las imágenes sujetas al entrenamiento, esto ya que se debe considerar la multiplicación de la altura por el ancho de la imagen para tener como cantidad de neuronas de entrada el resultado de dicha multiplicación, dicha cantidad de neuronas de entrada tiene repercusión sobre el rendimiento de la red neuronal.

En esta investigación se tomó en cuenta tamaños de imagen de 100 x 100 a 225 x 225 píxeles, para verificar el modelo con mejor rendimiento según el tamaño de la capa de entrada, se consideran estas dimensiones de imágenes puesto que para tamaños superiores a los propuestos se debe tomar en cuenta que se pueden presentar los siguientes inconvenientes:

- Mayor tiempo de entrenamiento del modelo.
- Mayor espacio de memoria ocupado por el modelo entrenado.
- Mayor tiempo de procesamiento al momento de aplicar el modelo en la plataforma.

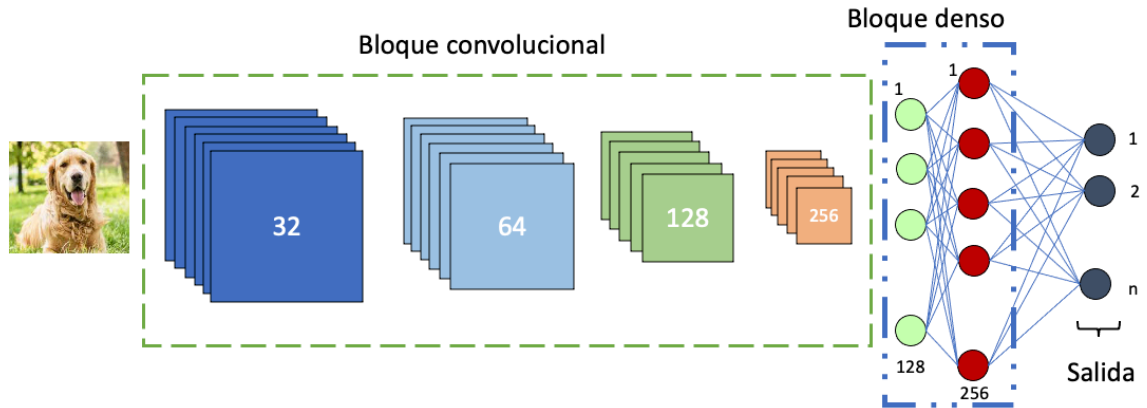
Por otra parte, tamaños menores a los propuestos pueden producir una pérdida de información en las imágenes y de igual manera no permite incrementar el tamaño de las capas convolucionales debido a la reducción producida por las operaciones inherentes a las redes neuronales convolucionales.



**Figura N° 9.** Evaluación de los mejores modelos en función a la imagen de entrada. Fuente: Elaboración propia.

La Figura N° 9 se muestra el resultado obtenido realizando las pruebas de funcionamiento de la arquitectura propuesta con diferentes tamaños de imágenes comenzando en 100 píxeles llegando a 225 píxeles, se puede apreciar que para

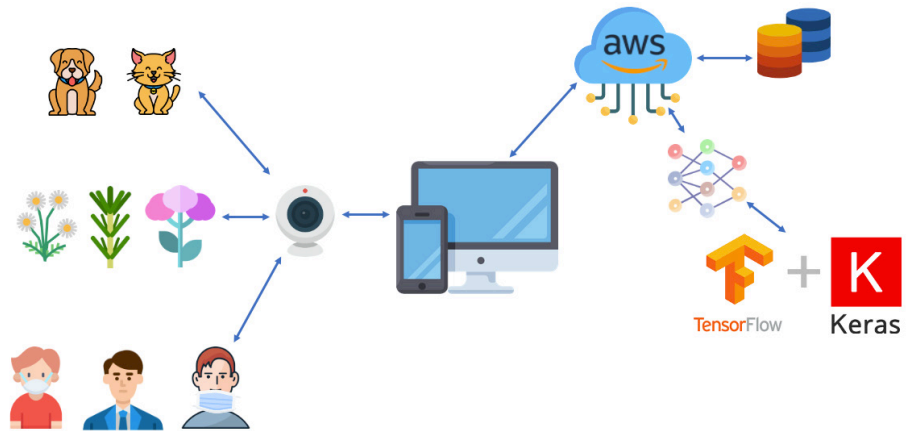
imágenes de 200 pixeles se tienen mejores resultados, sin embargo durante el proceso de entrenamiento existen diferencias considerables en cuanto al tiempo de entrenamiento que requiere cada modelo, este aspecto será considerado en apartados posteriores, sin embargo como arquitectura propuesta se tomará en cuenta una capa de entrada de 200 x 200, considerando los resultados obtenidos.



**Figura N° 10.** Arquitectura propuesta de la red neuronal de clasificación.  
Fuente: Elaboración propia.

En la Figura N° 10 se muestra en forma gráfica el detalle de la arquitectura propuesta que obtuvo los mejores resultados y se observan 4 bloques convolucionales que consisten en una capa convolucional y una capa de máximo muestreo (Max-Pool) e indicando la cantidad de filtros en cada bloque, una capa flatten que servirá para el paso de las capas convolucionales a las capas densas, a continuación se toma en cuenta una capa densa de 128 y otra de 256 neuronas, para concluir con la capa de salida con una activación softmax y completar la clasificación esperada, es importante denotar que la cantidad de neuronas de salida será definida según la cantidad de clases a clasificar.





**Figura N° 11.** Diagrama de Bloques de la Plataforma. Fuente: Elaboración propia.

En la figura 11 se observa el diagrama de bloques de la plataforma desarrollada, la cual cuenta con un inicio de sesión donde se podrá realizar la vinculación de usuarios nuevos y poder iniciar sesión para poder realizar pruebas, todo el diseño de la plataforma basado en el modelo MVC (modelo, vista, controlador), que es el modelo que se usa con frecuencia para este tipo de proyectos, de igual manera la plataforma se encuentra alojada y puesta en producción en AWS (Amazon Web Services), como el objetivo principal es la sencillez de la plataforma solo contara con botones para poder redireccionar a las páginas de prueba ya establecidas según a la pruebas de validación realizadas, de igual manera tendrá un enlace a Google Colaboratory para tener acceso a un código de ejemplo para poder realizar entrenamientos propios en base a set de datos que se deseen analizar. Por último, se tendrá la opción de poder realizar pruebas de funcionamiento de los algoritmos pre entrenados y poder ser visualizados de igual manera como los ejemplos preestablecidos.

#### 4. DISCUSIÓN, DESARROLLO Y ANÁLISIS

En este apartado se evaluará el modelo propuesto para la clasificación de objetos, comparando los resultados obtenidos durante todo el proceso de entrenamiento y validación de los modelos propuestos, a partir de los cuales se determinó una arquitectura básica para la clasificación, para todo este proceso se tomaron en cuenta el 70% de los datos contenidos en los tres dataset planteados para la validación del modelo y para una evaluación de resultados se usó el 30%, para realizar pruebas con los diferentes modelos se determinó 100 épocas para el entrenamiento y se realizaron las validaciones de los modelos propuestos, al igual que se mencionó en el apartado anterior, estos coadyuvaron a la elección y determinación de las características del modelo final.

Se debe tomar en cuenta, que las redes evaluadas en este proceso fueron entrenadas en Google Colaboratory partiendo desde cero en primera instancia sin aumento de datos y posteriormente con aumento de datos, debido a que el entrenamiento sin aumento de datos género resultados ineficientes para la plataforma este proceso fue desestimado y únicamente se cuenta con los resultados del proceso con aumento de datos, de igual manera es importante aclarar que previo al entrenamiento se realizar un procesamiento digital de imágenes para poder estandarizar el tamaño de las imágenes (ancho por alto) según la capa de entrada propuesta.

Durante el proceso de investigación y pruebas realizadas a los modelos se determinó realizar un "Upgrade" a la suscripción de Google Colaboratory, esto con la finalidad de tener mejores prestaciones al momento de evaluar y entrenar los modelos planteados. Luego de realizar el "Upgrade", en la Tabla N° 2 se realiza una comparación de las mejoras obtenidas en cuanto a tiempo de entrenamiento se refiere.

**Tabla N° 2.** Comparación del tiempo de entrenamiento con Colab y Colab pro

ARQUITECTURA	TIEMPO DE ENTRENAMIENTO (MIN)	
	COLAB	COLAB PRO
2 CCN -2 CD	63	29
3 CCN -2 CD	69	30
4 CCN -2 CD	70	30
5 CCN -2 CD	77	35

Fuente: Elaboración propia.

Como se observa en la Tabla 2 el tiempo se reduce aproximadamente en un 50%, esto se debe principalmente a las mejores prestaciones de hardware brindadas en Colab pro, sin embargo esto puede ser variable según la disponibilidad que exista durante la ejecución del código, tal cual como se explica en los términos de uso proporcionados por Google, sin embargo durante todo el desarrollo de la investigación se puede apreciar que los tiempos antes mencionados se mantuvieron estables y sin variaciones considerables al momento del proceso de entrenamiento.

De igual manera, como se mencionó anteriormente durante el proceso de pruebas realizadas a los modelos en función al tamaño de imágenes de entrada se logró evidenciar un incremento en el tiempo de entrenamiento a medida que la imagen aumenta, teniendo una relación directamente proporcional. En la Tabla N° 3 siguiente se muestran los datos obtenidos durante el proceso de entrenamiento evaluando los diferentes modelos y el tamaño de imágenes.

**Tabla N° 3.** Proceso de entrenamiento de los modelos propuestos

ARQUITECTURA PROPUESTA	DIMENSIONES			
	100 X 100	150 X150	200 X 200	225 X 225
(4CCN - 1 CD 256)	28	55	83	123
(3CCN - 1 CD 256)	28	55	84	122
(4CCN - 1 CD 256)	29	56	85	124
(5CCN - 1 CD 256)	30	57	88	125
(2CCN - 2 CD 512)	28	59	89	125
(3CCN - 2 CD 512)	29	60	90	127
(4CCN - 2 CD 512)	32	65	90	128
(5CCN - 2 CD 512)	35	66	92	130

Fuente: Elaboración propia.

Como se observa según se incrementa el tamaño de imagen se incrementa también el tiempo necesario para el entrenamiento del modelo, para poder determinar el modelo y el tamaño de imagen adecuado, se pone en consideración la precisión del modelo obtenido, para este cometido se puede apreciar la Tabla N° 4, en la cual se muestra el mejor porcentaje de precisión según el tiempo planteado, todo esto para la arquitectura escogida en función a estos resultados.

**Tabla N° 4.** Relación de tiempo y precisión según la imagen de entrada

ARQUITECTURA	TIEMPO (MIN)	% ENTRENAMIENTO	%VALIDACION
(4CCN - 2 CD -100)	35	84,3	87,06
(4CCN - 2 CD - 150)	65	87,5	90,02
(4CCN - 2 CD - 200)	91	90,6	93,27
(4CCN - 2 CD - 225)	126	91,93	92,18

Fuente: Elaboración propia.

Por último, para tener en cuentas las ventajas obtenidas gracias a la suscripción de Google Colabratory Pro se tiene en cuenta la Tabla N° 5 comparativa del tiempo requerido por cada época para el entrenamiento en función al tamaño de imagen.

**Tabla N° 5.** Relación de tiempo por época con Colab y Colab Pro y precisión según la imagen de entrada

ARQUITECTURA	TIEMPO x EPOCA (SEG)		%VALIDACION	
	COLAB	COLAB PRO	COLAB	COLAB PRO
(4CCN - 2 CD - 100)	67	18	86,9	87,06
(4CCN - 2 CD - 150)	150	39	90,05	90,02
(4CCN - 2 CD - 200)	215	56	93,24	93,27
(4CCN - 2 CD - 225)	296	77	92,19	92,18

Fuente: Elaboración propia.

Como se aprecia en los datos obtenidos se tiene una reducción de aproximadamente un 70% del tiempo utilizado para el entrenamiento por época teniendo una suscripción a Colab Pro, esto se debe principalmente a que con la suscripción se tiene acceso a más memoria RAM y un GPU con mejores prestaciones que en la versión regular, sin embargo, los resultados obtenidos en cuando a precisión son muy similares.

## 5. CONCLUSIONES

En primer lugar, se estudió las redes neuronales desde los aspectos más básicos hasta la manera de clasificar objetos en imágenes. Se explicó tanto su estructura, como los algoritmos utilizados para el entrenamiento y los parámetros que interactúan en el mismo.

En segundo lugar, se ha generado un set de datos de imágenes con diferencias entre clases, con el objetivo de aprender diferentes modelos de clasificación que sean generales y robustos al momento de realizar una predicción y pueda ayudar al desarrollo de nuevas aplicaciones e interfaces que sean de ayuda a futuros proyectos de investigación.

En tercer lugar, se pudo evaluar tres redes neuronales diferentes para la clasificación de imágenes, para poder obtener un modelo básico y de comprensión sencilla para la implementación de nuevos procesos de clasificación.

En cuanto a la clasificación de objetos, como trabajo futuro, se podría realizar una actualización de la plataforma actual para la clasificación y detección de objetos, de igual manera se podría plantear como opción el contar con un proceso

de entrenamiento completo y un proceso de transfer learning con base a redes neuronales pre entrenadas, ya sea congelando los parámetros ya entrenados y únicamente realizando únicamente el entrenamiento de capas específicas para el proyecto en específico

En cuando al clasificador utilizado, aumentar el set de datos podría derivar y ayudar considerablemente a mejorar las ratios de clasificación y la capacidad de generalización de modelos. Cuanto mayor variedad de escenarios se tenga en el conjunto de datos con el cual se entrene el modelo, más probable será que el sistema adquiera el conocimiento necesario para poder clasificar y diferenciar entre un objeto y otro.

Se pudo documentar todo el proceso tanto de entrenamiento como de validación de las redes neuronales propuestas, esto con la finalidad de que en futuros trabajos sean punto de partida para mejorar los mismos para el desarrollo de nuevos proyectos con base a los objetivos planteados en el actual proyecto de investigación.

## REFERENCIAS

Arriola, I. (2018). *Detección de objetos basada en Deep Learning y aplicada a vehículos autónomos*. <https://addi.chu.es/handle/10810/28983?show=full>

CEPAL. (2019, September 3). *Cumbre de Inteligencia Artificial en América Latina*. Comisión Económica Para América Latina y El Caribe. <https://www.cepal.org/es/notas/cumbre-inteligencia-artificial-america-latina>

Clavo, D. (2017, July 13). *Clasificación de redes neuronales artificiales*. <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>

Enciclopedia de visión artificial de aprendizaje profundo. (2021, October 10). *VGG, GoogLeNet*. [https://wikidocs-net.translate.google/137251?\\_x\\_tr\\_sl=auto&\\_x\\_tr\\_tl=es&\\_x\\_tr\\_hl=es&\\_x\\_tr\\_pto=wapp](https://wikidocs-net.translate.google/137251?_x_tr_sl=auto&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=wapp)

García Villanueva, M., & Romero Muñoz, L. (2020). Diseño de una arquitectura de Red Neuronal Convolutiva para la clasificación de objetos. *Ciencia Nicolaita*.

Iglesias, E., Garcia, A., Puig, P., & Benzaqué, I. (2020). *Inteligencia artificial: Gran oportunidad del siglo XXI: Documento de reflexión y propuesta de actuación*. <https://publications.iadb.org/publications/spanish/document/Inteligencia-artificial-Gran-oportunidad-del-siglo-XXI-Documento-de-reflexion-y-propuesta-de-actuacion.pdf>

Kriesel, D. (2007). *A brief Introduction on Neural Networks* (Citeseer, Ed.). <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.6774&rep=rep1&type=pdf>

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, 128(2), 261–318. <https://doi.org/10.1007/S11263-019-01247-4/FIGURES/21>

---

Manzanares González, A. (2018). *Detector de baches con Deep Learning*. Universidad Pompeu Fabra.

Nielsen, M. (2019, December). *Neural networks and deep learning*. <http://neuralnetworksanddeeplearning.com/about.html>

Oracle. (2022). *Inteligencia Artificial, Machine Learning, Deep Learning: una historia de muñecas rusas*. <https://www.oracle.com/es/database/cloud/algoritmos-machine-learning.html>

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, 29(9), 2352–2449. [https://doi.org/10.1162/NECO\\_A\\_00990](https://doi.org/10.1162/NECO_A_00990)

TensorFlow. (2022, October 18). *TensorFlow Datasets*. 2022. [https://www.tensorflow.org/datasets/catalog/cats\\_vs\\_dogs?hl=en](https://www.tensorflow.org/datasets/catalog/cats_vs_dogs?hl=en)

Universidade da Coruña, Universidade de Santiago de Compostela, & Universidade de Vigo. (2021). máster universitario en Inteligencia Artificial / Master in Artificial Intelligence. [https://www.uvigo.gal/sites/uvigo.gal/files/contents/paragraph-file/2021-09/MemoriaVerificacionMIA\\_SUG.pdf](https://www.uvigo.gal/sites/uvigo.gal/files/contents/paragraph-file/2021-09/MemoriaVerificacionMIA_SUG.pdf)