

# Red neuronal convolucional para la detección de aves exóticas en peligro de extinción

Guillermo Eduardo Narváez  
guillermo.narvaez.21.95@gmail.com / Universidad de Lima, Perú

Recepción: 17-6-2019 / Aceptación: 8-8-2019

**RESUMEN.** En los últimos años, los índices de comercialización ilegal de animales se han incrementado debido a la falta de conciencia sobre el cuidado y la preservación de nuestro ecosistema. Entre los años 2000 y 2015, 67 749 animales silvestres fueron extraídos del Perú de manera ilegal. Entre ellos 29 591 eran aves exóticas (43 %), 26 951 anfibios (40 %), 8600 reptiles (13 %) y 2607 mamíferos (4 %); valorizados en mil dólares, aproximadamente. La propuesta para reducir significativamente estos actos ilícitos se basa en contar con un aplicativo que pueda identificar a las especies que se intentan comercializar ilegalmente. Dicho aplicativo hará uso de las arquitecturas de las redes neuronales convolucionales denominada VGGnet16, la cual permitirá identificar correctamente al animal. Se ha trabajado con una base de datos de animales reales obteniéndose una efectividad del 89 %.

**PALABRAS CLAVE:** reconocimiento de objetos, redes neuronales convolucionales, comercio de vida silvestre

## Convolutional Neural Network for Detecting Endangered Exotic Birds

**ABSTRACT.** In the last years, the rates of illegal wildlife trade have increased due to the lack of awareness for caring and preserving our ecosystem. Between the years 2000 and 2015, sixty-seven thousand seven hundred forty-nine (67,749) wild animals were illegally taken from Peru, including 29,591 exotic birds (43%), 26,951 amphibians (40%), 8,600 reptiles (13%) and 2,607 mammals (4%) valued at approximately USD 1,000. The proposal to significantly reduce these illegal acts is based on having an application that can identify the species attempted to be illegally traded. Such application will use the architectures of convolutional neural networks called VGGNet16, which will allow the correct identification of the animal. The present research used a database of real animals, obtaining an effectiveness of 89%.

**KEYWORDS:** object recognition, convolutional neural networks, wildlife trade

## 1. INTRODUCCIÓN

El tráfico de animales es el tercer comercio ilegal después de las drogas y las armas (Quevans, Falcón y Elías, 2014). Estos actos conllevan a la depredación de la fauna o biodiversidad que para la mayoría de las personas suele pasar desapercibido. Esto tiene como consecuencia la muerte de cientos e incluso miles de animales durante su captura y transporte (Carrión, 2017). Dicha actividad está en crecimiento, no solo a nivel local sino mundial (Quevans *et al.*, 2014).

En la tabla 1 se detalla la finalidad y los países destino donde los animales son comercializados de manera ilegal (Quevans *et al.*, 2014).

Tabla 1  
*Finalidad del tráfico de fauna silvestre en el Perú*

Finalidad	Descripción
Coleccionistas particulares y zoológicos	Europa: Holanda, Bélgica, Austria, Suiza, Francia, Alemania, Italia Reino Unido y España. Asia: Singapur, Hong Kong, Japón y Filipinas. Norteamérica: Estados Unidos y Canadá.
Científicos y la industria biomédica	Los investigadores ilegales están en búsqueda de nuevas especies y sus derivados para la producción de medicamentos.
Comercialización internacional en tiendas de mascotas	Esto incentiva el tráfico ilegal de animales silvestres ya que los precios varían y dependiendo de la especie existe una mayor demanda.
Otras finalidades	Trofeos de caza deportiva, actividades religiosas, etc.

Fuente: Estrategia nacional para reducir el tráfico ilegal de fauna silvestre en el Perú 2017-2027 y su plan de acción 2017-2022. Serfor (2017).

En el Perú existen altos índices de tráfico de animales debido a que en Europa muchas de nuestras especies se encuentran altamente valorizadas (Barletta, Pereira, Robert y Yoguel, 2013). Algunas de dichas especies son el cóndor andino, el gallito de las rocas, loro cabeza roja, entre otros, 8479 aves fueron decomisadas en pésimas condiciones en el período de 2002-2007 en Lima, Perú (Quevans, *et al.*, 2014). Según Carrión (2017), durante el proceso de intervención en las acciones contra el tráfico ilegal y depredación de fauna silvestre se incautaron cacatúas en botellas de plástico (véase la figura 1).



*Figura 1.* Cacatúa dentro de una botella de plástico

Fuente: Carrión (2017)

Lamentablemente, los lugares de control a nivel nacional son deficientes y no existen suficientes especialistas ni agentes ambientales de la policía ecológica; por lo que el proceso de intervención no es efectivo (Carrión, 2017) y no existe ninguna base de datos digital oficial adecuadamente documentada (López Tarabochia, 2016).

Por lo anteriormente expuesto, surge la necesidad de apoyar el proceso de intervención en las acciones contra el tráfico ilegal y depredación de fauna silvestre mediante el desarrollo de un aplicativo que permita identificar correctamente el tipo de ave que se está traficando. Esta identificación está enlazada con un registro de base de datos que brinda la información si es una especie o no en peligro de extinción.

## 2. ESTADO DEL ARTE

### 2.1 Redes neuronales convolucionales para la detección de objetos a través de imágenes

El reconocimiento de objetos consiste en la extracción de parámetros que permiten una identificación e interpretación inconfundible del objeto. El proceso de reconocimiento de objetos

comienza con el preprocesamiento de las imágenes, seguido de un modelo de clasificación previamente entrenado (Rossius, 2013).

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) se han establecido como una poderosa clase de métodos para problemas de reconocimiento de imágenes (Karpathy, Toderici, Shetty, Leung, Sukthankar y Fei-Fei, 2014). Las arquitecturas de las CNN pueden contener hasta sesenta millones de parámetros (miles de imágenes). Esta idea es clave debido a que las capas internas de las CNN pueden actuar como un extractor genérico de representación de imágenes que puede ser entrenado con un conjunto de datos (Oquab, Bottou, Laptev y Sivic, 2014).

### 2.1.1 Aplicación de CNN en animales

La taxonomía es un elemento muy importante para la gestión de la biodiversidad ya que estas informan sobre las características esenciales y los nombres de las especies. Asimismo, permite mejorar los estudios científicos y programas de vigilancia para cuidar el ecosistema.

Una de las aplicaciones fue la identificación de los géneros de las hormigas haciendo uso de las CNN, para lo cual se les utilizó para identificar y clasificar automáticamente a las hormigas con el fin de conocer sobre su taxonomía. Se utilizó un repositorio (<https://antweb.org/>) donde se obtuvieron imágenes de diferentes ángulos de las hormigas (cabeza, perfil dorsal y superior) (véase la figura 2) (Marques *et al.*, 2018).



Figura 2. Ejemplo de las perspectivas de la cabeza, perfil forsal y superior de una especie llamada *Odontomachus chelifer*

Fuente: Marques *et al.* (2018)

La arquitectura utilizada fue AlexNet con 8 capas convolucionales, 5 capas RELU, 3 Max. Pooling y 3 Full Connected (véase la figura 3).

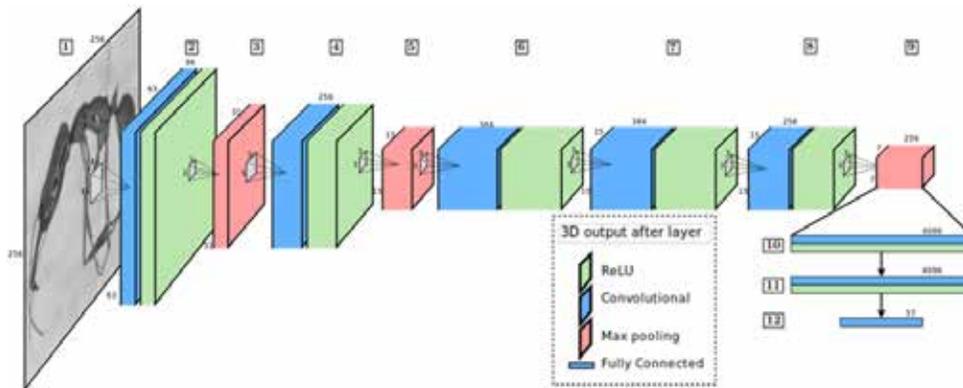


Figura 3. Arquitectura CNN utilizada para realizar la clasificación de imágenes  
Fuente: Marques *et al.* (2018)

Se consideraron diversos clasificadores como exactitud, precisión media y precisión mínima. Los mejores resultados se obtuvieron cuando se trabajaron con imágenes desde el punto de vista de la cabeza. Esto se debe a que la clasificación de vistas de cabeza puede ser más precisa, porque la posición de las piernas de la muestra de perfil y vistas dorsales podría añadir cierto ruido. Por otro lado, los géneros de las hormigas tienen diferencias morfológicas más pronunciado en la cabeza (Marques *et al.*, 2018).

Asimismo, varios estudios demuestran que los virus son los responsables de más de cien enfermedades que componen una enorme carga a nivel mundial (Musso, Rodríguez-Morales, Levi, Cao-Lormeau y Gubler, 2018). Alguno de ellos es el dengue, el chikunguya y zika. Motta y otros (2019) proponen una aplicación basada en CNN para la clasificación de los mosquitos en el campo, llevando a cabo una clasificación morfológica automática. Trabajaron con un conjunto de datos que incluía 4056 imágenes de los mosquitos. Dichas imágenes sirvieron para realizar un entrenamiento en tres arquitecturas de CNN: LeNet, AlexNet y GoogleNet (Motta *et al.*, 2019) (véase la figura 4).

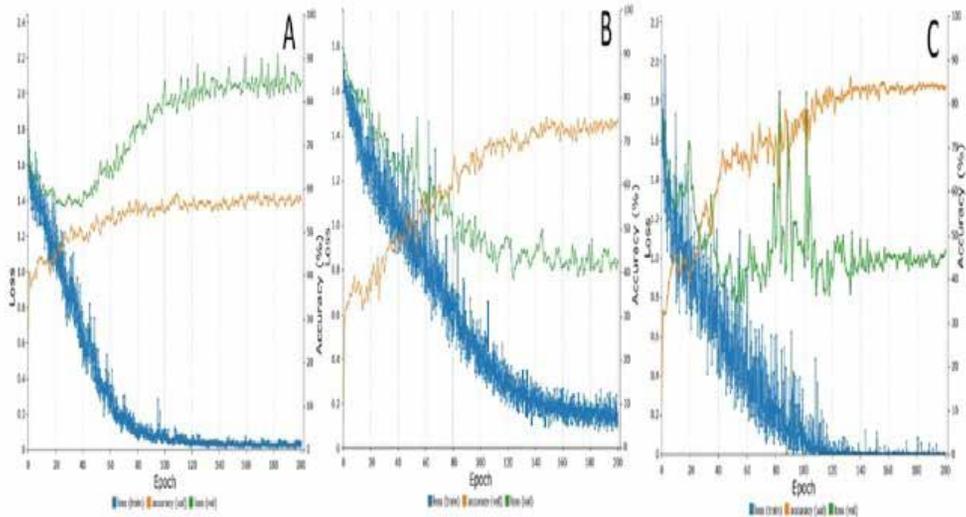


Figura 4. Representación del proceso de capacitación de cada CNN: (1A) LeNet, (1B) AlexNet y (1C) GoogLeNet (línea naranja: validación de precisión; línea verde: validación de pérdida; línea azul: entrenamiento de pérdida). (1A) LeNet (después de 200 épocas): porcentaje de precisión (validación) 57,50 %, pérdida (validación) 2,07, pérdida (entrenamiento) 0,03; (1B) AlexNet (después de 200 épocas): porcentaje de precisión (validación) 74,69 %, pérdida (validación) 0,83, pérdida (entrenamiento) 0,11; (1C) GoogLeNet (después de 200 épocas): porcentaje de precisión (validación) 83,88 %, pérdida (validación) 1,03, pérdida (entrenamiento) 0,00

Fuente: Motta *et al.* (2019)

## 2.2 Aplicaciones móviles con CNN

Durante la vigesimotercera conferencia de reconocimiento de patrones se presentaron estudios sobre el reconocimiento de objetos haciendo uso de celulares centrados en tres arquitecturas: AlexNet, Network In Network y GooLeNet. La aplicación del Museo del Mar en España es un ejemplo de estos propuestos (Tobías, Ducournau, Rousseau, Mercier y Fablet, 2017)

Tobías *et al.* (2017) propuso diferentes cámaras que fueron utilizadas para poder adquirir imágenes con diferentes calidades. Todas las imágenes estuvieron sujetas a un paso de procesamiento previo que incluyó la normalización del tamaño (véase la figura 5).



Figura 5. Ejemplo de imágenes de la base de datos del Museo del Mar.  
Fuente: Tobías *et al.* (2017)

Una de las principales ventajas de la utilización de CNN es que los modelos pueden ser desplegados en casi cualquier dispositivo y el tiempo de entrenamiento es directamente proporcional a la cantidad de datos a procesar y el número de iteraciones aplicadas (Tobías *et al.*, 2017). Por ejemplo, en este experimento que se realizó se ejecutó, aproximadamente, en dos a tres días utilizando varios dispositivos: CPU i7.6820HQ, CPU Intel AT 2.70GHz, GPU Nvidia Tesla K80 y un Ipad con un procesador de 1,3 Ghz de doble núcleo (Tobías *et al.*, 2017). Estos datos se aprecian en la tabla 2.

Tabla 2

*Tiempo computacional promedio y memoria de requerimientos profundos*

	GPU	CPU	Mobile	Memory
AlexNet	22,6 ms	339,0 ms	382 ms	230 MB
GoogLeNet	17,6 ms	452,1 ms	992 ms	50 MB

Fuente: Tobías *et al.* (2017)

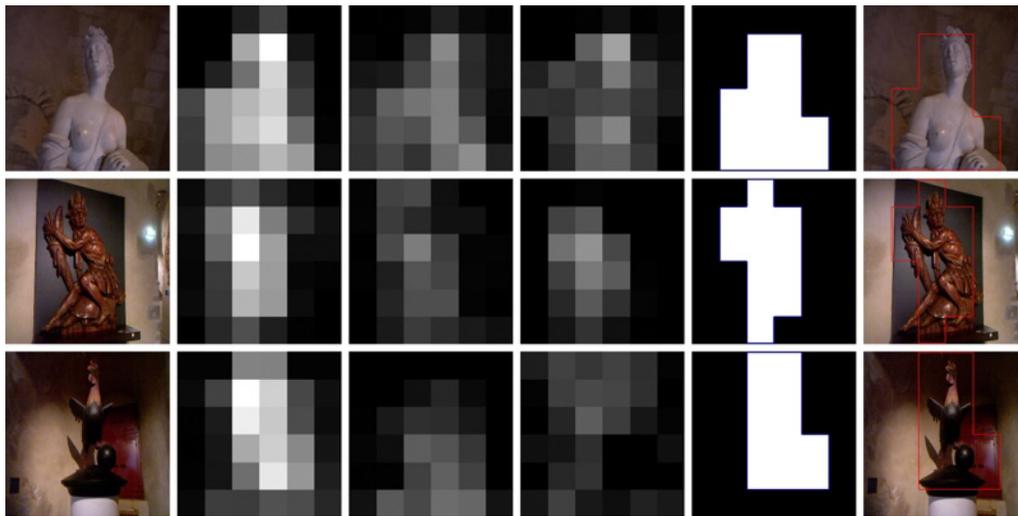


Figura 6. Localización e identificación de objetos, caso de estudio del Museo del Mar  
Fuente: Tobías *et al.* (2017)

### 3. METODOLOGÍA

La metodología de la investigación consiste en que el usuario al abrir el aplicativo activará la cámara del celular de manera automática. Las imágenes capturadas, automáticamente, pasarán por un proceso de identificación de la CNN y si, en el caso, dicha ave está en peligro de extinción se procederá a generar el acta (véase la figura 7).

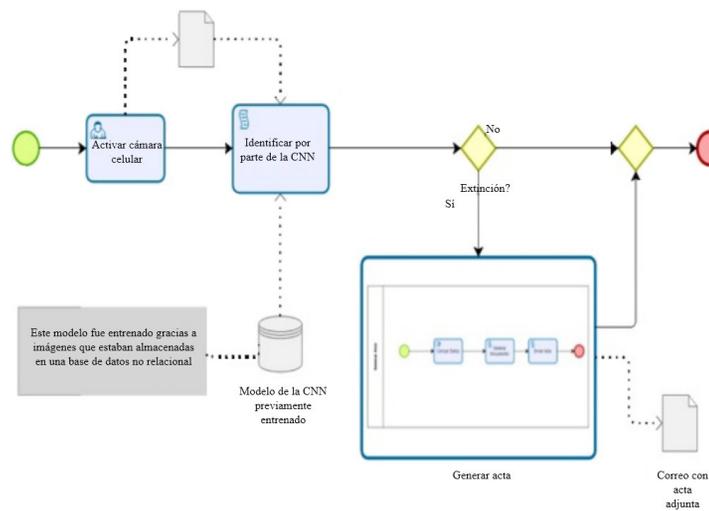


Figura 7. Esquema de la propuesta de solución  
Elaboración propia

### 3.1 Datos

El *dataset* trabajado consta de 600 imágenes donde son 3 clases y cada una de ellas posee 200 imágenes. Estas imágenes se almacenan en una base de datos no relacional conocida como Firebase.

Llevar un registro de las aves será la principal herramienta para que el aplicativo pueda consultarla y saber si realmente es un animal en peligro de extinción (peligro crítico CR, en peligro EN, vulnerable VU) (véase la figura 8).

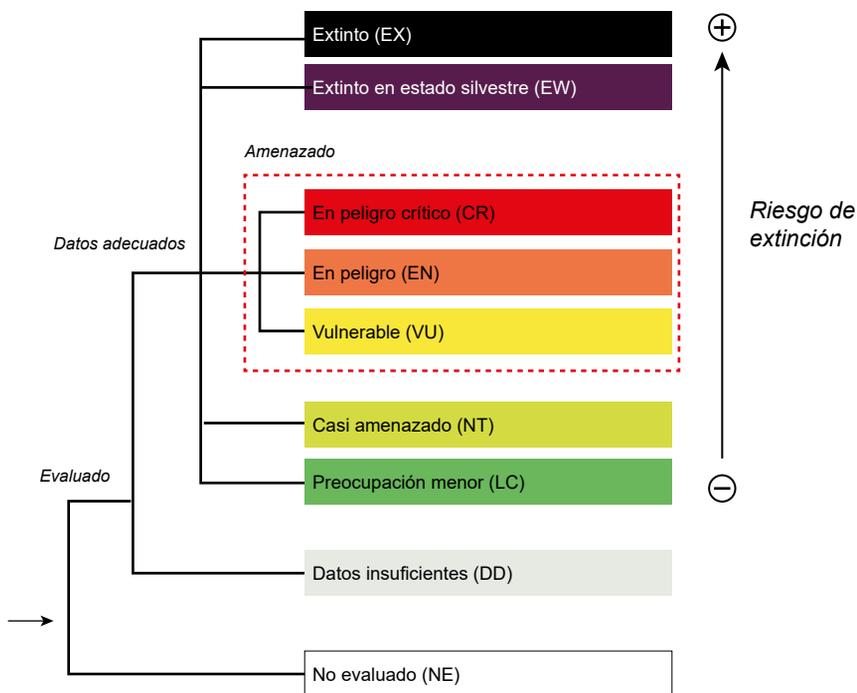


Figura 8. Categorías de amenaza según UICN

Fuente: Guía de identificación y cuidados iniciales de animales silvestres decomisados o hallados en abandono, Serfor (2016)

Luego de haber registrado a los animales haremos uso de las redes neuronales convolucionales para lograr una correcta identificación mediante el uso de imágenes.

## 4. ARQUITECTURA DE LA CNN (VGGNET16)

La arquitectura de la VGG16 consiste en un *InputLayer*, el cual acepta imágenes de 96 x 96 píxeles con una profundidad de 3, las cuales son procesadas por una capa de convolución en 2D que posee 32 filtros con núcleo de 3 x 3. Cada *pool* es de 3 x 3 que permite redimensionar las imágenes ingresadas de 96 x 96 a 32 x 32. Luego se realiza la activación y se va aumentando el tamaño de los filtros de 32 a 64. Se debe considerar que mientras más profunda es la red, menor serán las dimensiones del volumen y más filtros para aprender. Por consiguiente, se dimensiona el tamaño máximo de la agrupación de 3 x 3 a 2 x 2 para que las dimensiones no se reduzcan rápidamente. Estos pasos se realizan hasta llegar a un tamaño del filtro de 128 píxeles. Finalmente, en la capa totalmente conectada se realiza una activación lineal corregida y normalizada (Rosebrock, 2018).

### 4.1 Fases de la implementación de la CNN

a) *Creación y configuración de la CNN*

En esta primera fase, se realiza la creación de la clase *SmallerVGGnet* y se agregan las capas respectivas al modelo (Rosebrock, 2018).

b) *Entrenamiento de la CNN*

En esta segunda fase, se implementa el *script* para el entrenamiento de la red. Una vez implementada, se puede entrenar la CNN con Keras en base al entrenamiento deseado (Rosebrock, 2018).

c) *Clasificando imágenes con nuestra CNN y Keras*

En esta tercera fase, la red neuronal está totalmente entrenada y permite clasificar imágenes basándose de una secuencia de comandos (Rosebrock, 2018).

d) *Visualización de resultados*

En esta última fase, se analiza si la CNN ha clasificado correctamente la imagen del animal.

## 5. IMPLEMENTACIÓN

### 5.1 Descripción del *dataset*

Se ha construido un *dataset* de 600 imágenes repartidas en 3 clases de aves. Estas clases de aves fueron *Brotogeris\_versicolorurus*, *Carduelis\_magellanica* y *Sicalis\_flaveola*.

## 5.2 Data de validación y data para testing

Se utilizaron el 70 % de las imágenes para la validación y el 30 % para el *testing*. Esto se logró hacer con la función `sklearn.model_selection.train_test_split` que permite dividir un *dataset* en dos bloques, típicamente bloques destinados al entrenamiento y validación del modelo (llamemos a estos bloques “bloque de entrenamiento” y “bloque de pruebas”) (InteractiveChaos making things simple, s. f.).

## 6. RESULTADOS

Para el entrenamiento se utilizó una MacBookPro Intel Core i7 de séptima generación, 16 Gb de RAM y 4 Gb de Radeon Pro 560.

### 6.1 Métrica utilizada para el análisis

#### 6.1.1 Gráfico de precisión y pérdida

En la siguiente figura 9 se detalla el proceso de entrenamiento de la red durante las 1000 épocas. En cada una de las épocas existe un valor de entrenamiento perdido (*train\_loss*), valor perdido (*val\_loss*), valor ganado (*val\_acc*) y entrenamiento ganado (*train\_acc*). Estos valores van cambiando y generando una gráfica de la cual se concluye que el entrenamiento de la CNN es correcto ya que el valor ganado de las características obtenidas de las imágenes de animales superó al valor perdido durante el entrenamiento en la época 200 en adelante.

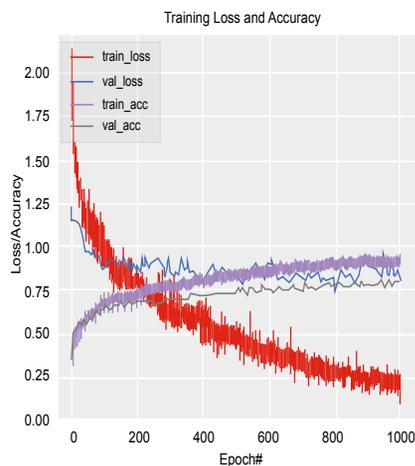


Figura 9. Gráfico del valor ganado y entrenamiento perdido, escenario B

### 6.1.2 Matriz

Una matriz de confusión resume el rendimiento de clasificación de un clasificador con respecto a algunos datos de prueba. Es una matriz bidimensional, indexado en una dimensión por la verdadera clase de un objeto y en el otro por la clase que el clasificador asigna. La figura 10 presenta un ejemplo de matriz de confusión para una clasificación de tres clases de aves (*Brotegeris\_versicolurus*, *Carduelis\_magellanica*, *Sicalis\_flaveola*). La primera fila de la matriz indica que 58 objetos pertenecen a *Brotegeris\_versicolurus* y que 39 fueron clasificados correctamente como pertenecientes a *Brotegeris\_versicolurus*, 9 clasificados erróneamente como pertenecientes a *Carduelis\_magellanica*; y 10 como pertenecientes a *Sicalis\_flaveola*.

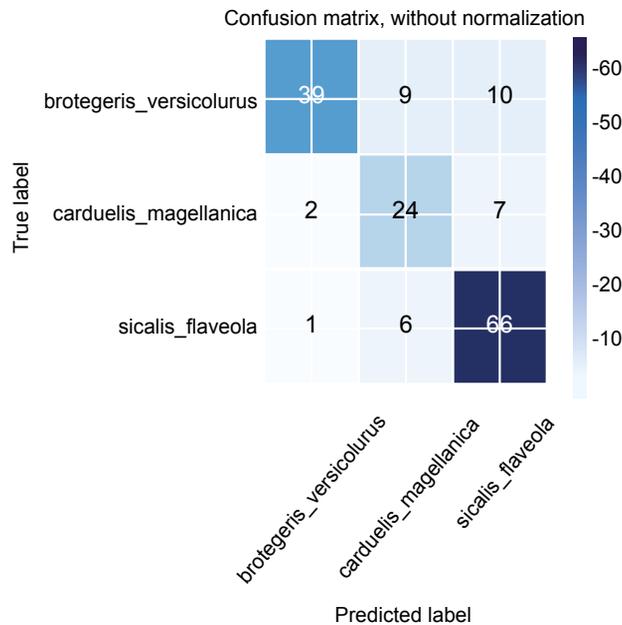


Figura 10. Matriz de confusión  
Elaboración propia

#### 6.1.2.1 Interpretación de las métricas de la matriz de confusión

Para la clase *Brotegeris\_versicolurus* se obtuvo una precisión de un 93 %, un *recall* de 67 % y un *f1-score* de 78 %; en la clase *Carduelis\_maggellanica*, una precisión de un 62 % , un *recall* de 73 % y un *f1-score* de 67 %; en la clase 2, una precisión de un 80 % , un *recall* de 90 % y un *f1-score* de 85 %. Se concluye que el modelo y las predicciones serán correctas (véase la figura 11).

	precision	recall	f1-score	support
0	0.93	0.67	0.78	58
1	0.62	0.73	0.67	33
2	0.80	0.80	0.85	73
avg/total	0.81	0.79	0.79	164

Figura 11. Resumen de los indicadores por cada clase escenario A  
Elaboración propia

## 7. DISCUSIÓN

La precisión promedio de 81 % demuestra que la identificación de imágenes es una característica propia de las redes neuronales convolucionales. Los hallazgos sobre la identificación de objetos abrieron las puertas a dicha investigación. El hecho que desarrollemos una aplicación que tenga la capacidad de identificar a un ave es beneficioso porque agilizará el proceso de intervención policial de animales. Asimismo, se construyó la base de datos no relacional que llevará los registros de las aves en tiempo real. La razón de la elección de una base de datos no relacional se debe a que no es posible llegar a la normalización de la *data* por la diversidad de información. Esta investigación brindará las bases para la posible investigación de clasificación de animales de todas las especies agrupadas por subfamilias y familias.

A medida que el campo de la investigación sea aplicado a la fauna silvestre permitirá lograr nuevos avances y experimentos basados en hipótesis. A nivel mundial existe una creciente necesidad por brindar aportes a la fauna silvestre debido a que el tráfico ilegal de animales ocupa el tercer comercio ilegal.

Aunque identificar correctamente un ave es un primer paso crítico, se debe considerar que las especies trabajadas poseen una taxonomía completa. Por ende, como trabajo futuro se ha proyectado el uso de redes generativas antagónicas (GAN) con el fin de poder reconstruir alguna extremidad o parte de la especie que no se encuentre completa.

Finalmente, nuestro enfoque tiene como objetivo aplicar técnicas de inteligencia artificial a la fauna y de ese modo mitigar el tráfico ilegal de animales.

## 8. CONCLUSIÓN

Se construyó un *dataset* de imágenes de aves y se evaluaron diversas métricas como la precisión, *recall*, *f1-score*, *support* con el fin de identificar automáticamente tipos de ave. Al utilizar nuestro conjunto de datos se proporcionaron resultados prácticos y análisis de desempeño del reconocimiento a lo largo del entrenamiento. Los resultados experimentales demostraron la posibilidad de utilizar el reconocimiento de imágenes para la clasificación de un solo dominio de especie (aves).

Actualmente, se está trabajando en la expansión de reconocimiento de familias o subfamilias de diversas especies. Finalmente, nos gustaría enfatizar que nuestra aplicación pueda mitigar el tráfico ilegal de animales.

## REFERENCIAS

- Barletta, F., Pereira, M., Robert, V., y Yoguel, G. (2013). Argentina: dinámica reciente del sector de *software* y servicios informáticos. *Revista de la CEPAL*, (110), 137-155. Recuperado de <http://www.cepal.org/publicaciones/xml/1/50511/RVE110Yoqueletal.pdf>
- Carrión, A. (2017). Tráfico de fauna silvestre en estado de emergencia. *El Comercio*. Recuperado de <https://elcomercio.pe/wuf/noticias/trafico-fauna-silvestre-emergencia-402805-noticia/>
- InteractiveChaos making things simple. (s. f.). InteractiveChaos making things simple. Recuperado de <https://www.interactivechaos.com/python/function/traintestsplit>
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., y Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1725-1732.
- López Tarabochia, M. (2016). Perú: Rescatan 25 aves silvestres que iban a ser comercializadas ilegalmente en Europa. *Revista Mongabay*. Recuperado de <https://es.mongabay.com/2016/12/peru-aves-trafico/>
- Marques, A. C., Raimundo, M. M., Cavalheiro, E. M. B., Salles, L. F. P., Lyra, C., y von Zuben, F. J. (2018). Ant genera identification using an ensemble of convolutional neural networks. *PLoS ONE*, 13(1).
- Motta, D., Santos, A. Á. B., Winkler, I., Machado, B. A. S., Pereira, D. A. D. I., ... Badaró, R. B. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLOS ONE*, 14 (1), 1-18.
- Musso, D., Rodriguez-Morales, A. J., Levi, J. E., Cao-Lormeau, V.-M., y Gubler, D. J. (2018). Unexpected outbreaks of arbovirus infections: lessons learned from the Pacific and tropical America. *The Lancet Infectious Diseases*, 18(11), 355-361. doi:10.1016/S1473-3099(18)30269-X

- Oquab, M., Bottou, L., Laptev, I., y Sivic, J. (2014). Learning and transferring midlevel image representations using convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1717-1724.
- Quevans, N., Falcón, N. y Elías, R. (2014). Fauna silvestre y productos derivados decomisados durante el período 2000-2007, Lima-Perú. *Salud y Tecnología Veterinaria*, 1, 14-18. doi:10.20453/stv.2013.105
- Rosebrock, A. (2018). Keras y redes neuronales convolucionales (CNN). *Pyimagesearch*. Recuperado de <https://www.pyimagesearch.com/2018/04/16/keras-and-convolutional-neural-networks-cnns/>
- Rossius, S. (2013). Reconocimiento de objetos mediante WebCam en tiempo real (Doctoral dissertation).
- Serfor. (2017) Estrategia Nacional para reducir el tráfico ilegal de fauna silvestre en el Perú 2017-2027 y su plan de acción 2017-2022. Recuperado de <https://www.serfor.gob.pe/wp-content/uploads/2017/09/PRESENTACION-DE-LA-ESTRATEGIA-NACIONAL-PARA-REDUCIR-EL-TRAFICO-ILEGAL-DE-FAUNA-SILVESTRE.pdf>
- Serfor. (2016). *Guía Identificación y cuidados iniciales de animales silvestres decomisados o hallados en abandono*. Recuperado de <https://www.serfor.gob.pe/wp-content/uploads/2016/08/Guia-de-identificacion-final12ago.pdf>
- S. R. (2013). *Reconocimiento de objetos mediante WebCam en tiempo real*. Universidad Politécnica de Valencia.
- Tobías, L., Ducournau, A., Rousseau, F., Mercier, y G., Fablet, R. (2017). Convolutional Neural Networks for object recognition on mobile devices: A case study. *2016 23rd International Conference on Pattern Recognition (ICPR)*, 3519-3524.
- Zeiler, M. D., y Fergus, R. (2014). Visualizing and understanding convolutional networks. *European conference on computer vision*, 8689, 818-833.