

UNIVERSIDAD NACIONAL DE INGENIERÍA - FACULTAD DE CIENCIAS - ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN

Implementacion de Inteligencia Artificial en el reconomiento de aves migratorias de los Pantanos de Villa

Jose Manuel Ravichagua Marin ¹, Linda Salazar Flores²

Ruiz Arica Milagros Cristina³

Cesar Cruz Martin Salazar

Facultad de Ciencias¹⁻⁴, Universidad Nacional de Ingeniería¹⁻⁴

e-mail¹⁻⁴: linda.salazar.f@uni.pe, jose.ravichagua.m@uni.pe, milagros.ruiz.a@uni.pe

Resumen

En este trabajo de investigación se presenta una aplicación de reconocimiento para la detección automática de aves migratorias de los Pantanos de Villa . Para el desarrollo de nuestra aplicación se hizo mediante deep learning para el reconocimiento , este se llevó a cabo el lenguaje python y para la aplicación se hizo uso de Streamlit.

Abstract

In this research work we present a recognition application for the automatic detection of migratory birds in the Pantanos de Villa. For the development of our application was done using deep learning for recognition, this was carried out in python language and the application was made using Streamlit.

1. INTRODUCCIÓN:

El reconocimiento de imágenes mediante inteligencia artificial es un tema de investigación de larga data en el campo de la visión por ordenador. Aunque a lo largo del tiempo han evolucionado diferentes métodos, el objetivo común del reconocimiento de imágenes es la clasificación de los objetos detectados en diferentes categorías (también se denomina reconocimiento de objetos).

En los últimos años, el aprendizaje automático, en particular la tecnología de aprendizaje profundo,

logró grandes éxitos en muchas tareas de visión por ordenador y comprensión de imágenes. En el presente trabajo se creará una aplicación de reconocimiento de aves migratorias en los pantanos de villa para promover la preservación de las especies.

2. OBJETIVOS

- Generales :
 - Crear una aplicación de reconocimiento de aves migratorias de los Pantanos de Villa mediante técnicas de deep learning.

■ Específicos:

- Hacer uso de redes neuronales convolucionales para la detección de aves ,además de diferentes softwares para la creación de la aplicación.
- Promover la preservación y conocimiento de las especies en peligro de extinción

3. ESTADO DEL ARTE

3.1. Pantanos de Villa

El Refugio de Vida Silvestre los Pantanos de Villa (RVSV) está ubicado en la ciudad de Lima. Este oasis natural alberga una gran cantidad de especies de flora y fauna, pero sobre todo se encuentra protegido por la diversidad de atractivos turísticos que posee. Entre ellos contamos con los 5 espejos de agua, y la presencia de cerca de 210 especies de aves, 13 de peces, 5 de anfibios y reptiles, y más.

El refugio busca principalmente conservar sus espacios naturales y fomentar la investigación de la diversidad biológica.

Los pantanos de Villa son también un lugar preciado para los observadores de aves, en especial en los meses de verano (noviembre y diciembre) donde se puede observar entre 15,000 y 20,000 bandadas de aves, que llegan del hemisferio norte para proseguir su trayectoria hacia el sur del país o bien visitar el refugio.

En sus territorios se registran 210 especies de aves (entre migratorias, residentes y registros eventuales). Las aves migratorias se pueden subdividir incluso de acuerdo a su procedencia: América del Norte, América del Sur, Andes y locales[1].

Aves en peligro crítico de los pantanos de villa :

- El potoyunco peruano (*Pelecanoides garnotii*)
- El pelícano peruano (*Pelecanus thagus*)
- El guanay (*Phalacrocorax bougainvillii*)
- El piquero peruano (*Sula variegata*)

Además, habitan en él aves casi amenazadas como: El zambullidor blanquillo (*Podiceps occidentalis*) y la gallareta gigante (*Fulica gigantea*) Otras especies de aves importante son: el playero manchado (*Actitis macularia*), el pato cabeza verde (*Anas platyrhynchos*), el playerito occidental (*Calidris mauri*), el halcón peregrino (*Falco peregrinus*), el chorlo del ártico (*Pluvialis squatarola*) y la garza tricolor (*Egretta tricolor*).



Figura 1: El potoyunco peruano (*Pelecanoides garnotii*)

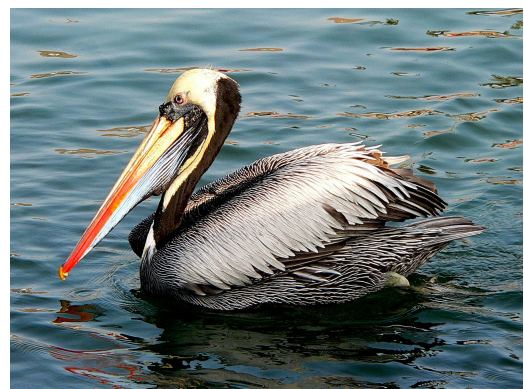


Figura 2: El pelícano peruano (*Pelecanus thagus*)



Figura 3: Gallareta gigante (*Fulica gigantea*)

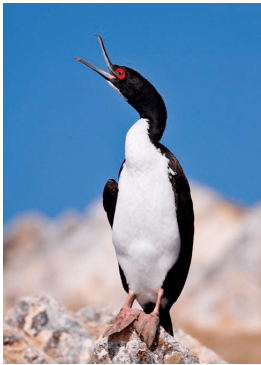


Figura 4: El guanay (*Phalacrocorax bouganvilli*)



Figura 5: El piquero peruano (*Sula variegata*)



Figura 6: El zambullidor blanquillo (*Podiceps occidentalis*)

4. CONCEPTOS PREVIOS

4.1. Inteligencia artificial

De [2] se dice que la Inteligencia Artificial (IA) es una de las ramas de las ciencias de la computación que más interés ha despertado en la actualidad, debido a su enorme campo de aplicación. La búsqueda de mecanismos que nos ayuden a comprender la inteligencia y realizar modelos y simulaciones de estos, es algo que ha motivado a muchos científicos a elegir esta área de investigación. El origen inmediato del concepto y de los criterios de desarrollo de la “IA” se remonta a la intuición del genio matemático inglés Alan Turing y el apelativo “Inteligencia Artificial” se debe a McCarthy quien organizó una conferencia en el Dartmouth College (Estados Unidos) para discutir la posibilidad de construir máquinas “inteligentes”; a esta reunión asistieron científicos investigadores de conocida reputación en el área de las ciencias computacionales como: Marvin Minsky, Nathaniel Rochester, Claude Shannon, Herbert Simon y Allen Newell. Como resultado de esta reunión, se establecieron los primeros lineamientos de la hoy conocida como Inteligencia Artificial; aunque anteriormente ya existían algunos trabajos relacionados. Desde su origen, la IA tuvo que lidiar con el conflicto de que no existía una definición clara y única de inteligencia; así es que no es de sorprender que aún en la actualidad, no exista una definición única de ella. Así como la Psicología ha identificado diferentes tipos de inteligencia humana (emocional, interpersonal, musical, lingüística, kinestésica, espacial, etc.), las distintas definiciones de la inteligencia artificial hacen énfasis en diferentes aspectos; aunque existen similitudes entre ellas. A continuación se presentan algunas de las definiciones iniciales de esta área.

- Estudio de la computación que observa que una máquina sea capaz de percibir, razonar y actuar (Winston, 1992).
- Ciencia de la obtención de máquinas que logren hacer cosas que requerirían inteligencia si las hiciesen los humanos (Minsky, 1968).
- Nuevo esfuerzo excitante que logre que la computadora piense, máquinas con mentes, en el sentido completo y literal (Haugeland, 1985).
- Rama de la ciencia computacional preocupada por la automatización de la conducta inteligente (Luger and Stubblefield, 1993).
- Máquina Inteligente es la que realiza el proceso de analizar, organizar, y convertir los datos en conocimiento, donde el conocimiento del sistema es información estructurada adquirida y aplicada para reducir la ignorancia o la incertidumbre sobre una tarea específica a realizar por esta (Pajares y Santos, 2006)

4.1.1. Clasificación de la Inteligencia Artificial

La escuela clásica dentro de la IA, utiliza representaciones simbólicas basadas en un número finito de primitivas y de reglas para la manipulación de símbolos (por ejemplo, redes semánticas, lógica de predicados, etc.), los cuales fueron y siguen siendo parte central de dichos sistemas. Otro tipo de representación es el llamado sub-simbólico, el cual utiliza representaciones numéricas (o sub-simbólicas) del conocimiento. El enfoque sub-simbólico de la IA se caracteriza por crear sistemas con capacidad de aprendizaje. Éste se puede obtener a nivel de individuo imitando el

cerebro (Redes Neuronales), a nivel de especie, imitando la evolución. Hasta hace poco era común hablar de Algoritmos Genéticos (AG) en general, en vez de identificar diferentes tipos de AE, ya que el resto de los algoritmos se pueden interpretar como variaciones o mejoras de los AG.[3]

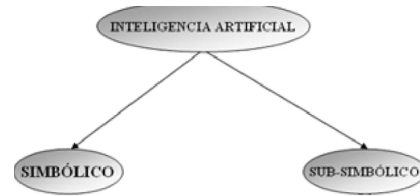


Figura 7: Clasificación de la Inteligencia Artificial

4.2. Machine learning

De [4] Machine learning es un campo de las ciencias de la computación que, de acuerdo a Arthur Samuel en 1959, le da a las computadoras la habilidad de aprender sin ser explícitamente programadas. Sin embargo, machine learning no es un proceso sencillo. Conforme el algoritmo ingiere datos de entrenamiento, es posible producir modelos más precisos basados en datos. Un modelo de machine learning es la salida de información que se genera cuando se entrena su algoritmo de machine learning con datos. Después del entrenamiento, al proporcionar a un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creará un modelo predictivo. A continuación, cuando proporcione al modelo predictivo con datos, recibirá un pronóstico basado en los datos que entrenaron al modelo.[5]

4.2.1. Modelos de Machine Learning

- Modelos lineales

Estos tratan de encontrar una línea que se “ajuste” bien a la nube de puntos que se dis-

ponen. Aquí destacan desde modelos muy conocidos y usados como la regresión lineal (también conocida como la regresión de mínimos cuadrados), la logística (adaptación de la lineal a problemas de clasificación cuando son variables discretas o categóricas-). Estos dos modelos tienen el problema del “overfit”, esto significa que se ajustan “demasiado” a los datos disponibles, con el riesgo que esto tiene para nuevos datos que pudieran llegar. Al ser modelos relativamente simples, no ofrecen resultados muy buenos para comportamientos más complicados

■ Modelos de árbol

Son modelos precisos, estables y más sencillos de interpretar básicamente porque construyen unas reglas de decisión que se pueden representar como un árbol. A diferencia de los modelos lineales, pueden representar relaciones no lineales para resolver problemas. En estos modelos, destacan los árboles de decisión y los random forest (una media de árboles de decisión). Al ser más precisos y elaborados, obviamente ganamos en capacidad predictiva, pero perdemos en rendimiento.

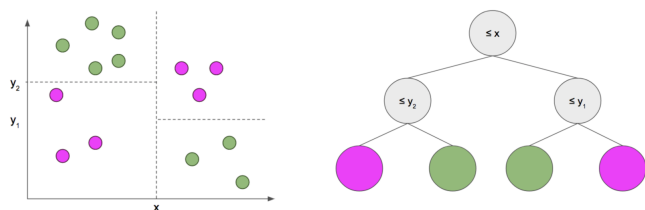


Figura 8: Modelos de árbol

■ Redes neuronales

Las redes artificiales de neuronas tratan, en cierto modo, de replicar el comportamiento

del cerebro, donde tenemos millones de neuronas que se interconectan en red para enviarse mensajes unas a otras. Esta réplica del funcionamiento del cerebro humano es uno de los “modelos de moda” por las habilidades cognitivas de razonamiento que adquieren. El reconocimiento de imágenes o vídeos, por ejemplo, es un mecanismo complejo y una red neuronal es lo mejor para realizarlo. El problema, como ocurre con el cerebro humano, es que son lentas de entrenar y necesitan mucha capacidad de cómputo. Quizás sea uno de los modelos que más ha ganado con la “revolución de los datos”[6]

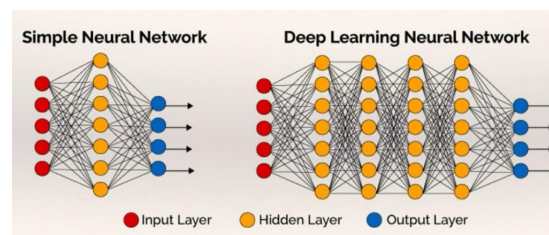


Figura 9: Modelos de árbol

4.2.2. Enfoques hacia el machine learning

Las técnicas de machine learning son necesarias para mejorar la precisión de los modelos predictivos. Dependiendo de la naturaleza del problema que se está atendiendo, existen diferentes enfoques basados en el tipo y volumen de los datos. En esta sección, discutimos las categorías del machine learning.

■ Aprendizaje supervisado

El aprendizaje supervisado comienza típicamente con un conjunto establecido de datos y una cierta comprensión de cómo se clasifican estos datos. El aprendizaje supervisado tiene la intención de encontrar patrones en datos que

se pueden aplicar a un proceso de analítica. Estos datos tienen características etiquetadas que definen el significado de los datos. Por ejemplo, se puede crear una aplicación de machine learning con base en imágenes y descripciones escritas que distinga entre millones de animales.

- Algoritmo de clasificación
 - Regresión logística
- Aprendizaje no supervisado

El aprendizaje no supervisado se utiliza cuando el problema requiere una cantidad masiva de datos sin etiquetar. Por ejemplo, las aplicaciones de redes sociales, tales como Twitter, Instagram y Snapchat, tienen grandes cantidades de datos sin etiquetar. La comprensión del significado detrás de estos datos requiere algoritmos que clasifican los datos con base en los patrones o clústeres que encuentra. El aprendizaje no supervisado lleva a cabo un proceso iterativo, analizando los datos sin intervención humana

- Algoritmo K-means
- Aprendizaje de refuerzo
- El aprendizaje de refuerzo es un modelo de aprendizaje conductual. El algoritmo recibe retroalimentación del análisis de datos, conduciendo el usuario hacia el mejor resultado. El aprendizaje de refuerzo difiere de otros tipos de aprendizaje supervisado, porque el sistema no está entrenado con el conjunto de datos de ejemplo. Más bien, el sistema aprende a través de la prueba y el error. Por lo tanto, una se-

cuencia de decisiones exitosas conduce al fortalecimiento del proceso, porque es el que resuelve el problema de manera más efectiva.[7]

4.3. Deep learning

El aprendizaje profundo, deep learning en inglés, es un campo del aprendizaje automático, machine learning en inglés, que se engloba, a su vez, dentro de la inteligencia artificial. Cuando se trabaja con las técnicas de machine learning se utiliza un conjunto de datos para ajustar un modelo predefinido que resuelve un determinado problema. Los datos con los que trabaja el modelo no suelen ser los datos originales, los datos en bruto del problema. En su lugar, se suelen definir una serie de características, features en inglés, que se obtienen a partir de los datos originales y son las que utiliza el modelo para resolver el problema.[8]

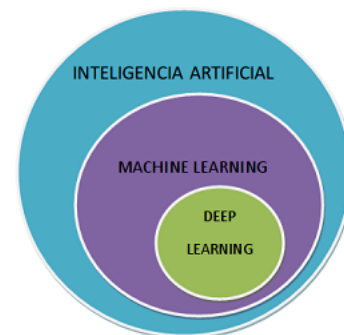


Figura 10: Relación entre inteligencia artificial, machine learning y deep learning

Una de las características de las técnicas de deep learning es que la extracción de características que realizan de los datos originales suele generar una representación jerárquica distribuida: se extraen características de menor a mayor complejidad, de forma independiente, definiéndose las características de los niveles superiores en función de las características de los niveles inferiores. El objetivo de las técnicas de

deep learning consiste en encontrar unas características de alto nivel, de forma que el modelo pueda transformar estas características en la salida esperada, con más facilidad que empleando la representación original de los datos. Por ejemplo, en el caso de la clasificación de imágenes, las técnicas de deep learning procesan la imagen de entrada para, en un primer nivel, extraer características de bajo nivel (como colores y bordes), mientras que en niveles superiores se detectan características más complejas, de más alto nivel (como pueden ser las ruedas de un coche o la cara de una persona). Estas características de más alto nivel serán las utilizadas para generar la salida del modelo.[9] En un modelo de deep learning se distinguen dos etapas: una etapa de extracción de características y una etapa de transformación de características. La primera de las etapas consiste en crear la jerarquía de características a partir de los datos en bruto del problema. La segunda etapa toma las características de más alto nivel de la jerarquía y las utiliza para aplicarles una transformación que permita dar la salida esperada por el sistema. Por ejemplo, en el caso de una red neuronal aplicada a la clasificación de imágenes, las primeras capas podrían realizar una extracción de características de bajo nivel y, al ir aumentando la profundidad de la red, las características extraídas serían más complejas, definidas por las características extraídas en las capas de más bajo nivel. La última capa de la red puede utilizarse para aprender a realizar una transformación de las características de más alto nivel a la salida esperada. En la Figura 4.4 se muestra el esquema completo de un modelo de deep learning con sus diferentes etapas.[10]

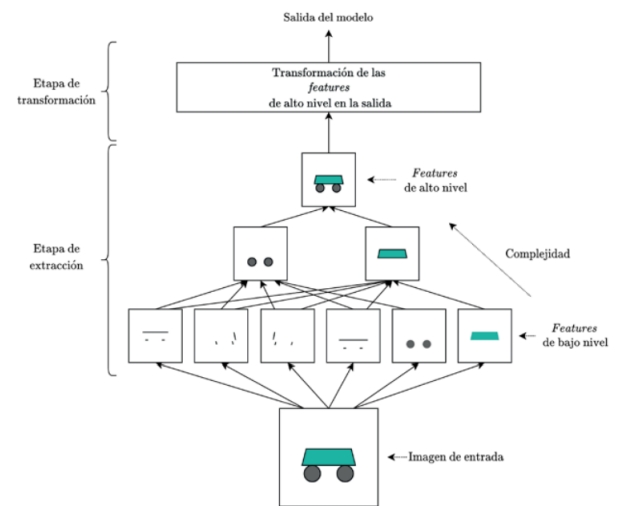


Figura 11: Esquema completo de un modelo de deep learning con sus diferentes etapas

4.4. Redes neuronales

Siendo un subgrupo de la Inteligencia Artificial, una Red Neuronal Artificial (RNA), se puede definir como un sistema de aprendizaje y procesamiento diseñado a imitación de los sistemas nerviosos de los humanos, consistente en una interconexión de unidades, denominadas neuronas artificiales o elementos de proceso, cuyo funcionamiento se inspira en el de las neuronas biológicas que trata de procesar la información.

Las Redes Neuronales se diferencian por topología, dependiendo de las características con las que cuentan:

- Perceptrón monocapa, la más sencilla.
- Perceptrón multicapa, donde se desarrollan capas ocultas y es capaz de eliminar información irrelevante.
- Red Neuronal recurrente, donde cobra un papel importante la temporalidad y, con ella, se dota

a la red neuronal de memoria.

- Redes Neuronales Convolucionales; estas últimas son las más completas.

Una Red Neuronal Convolutiva se utiliza en especial para el análisis de imágenes, transforma las características aprendidas con los datos de entrada y emplea capas convolucionales 2D, lo cual hace que esta arquitectura resulte adecuada para procesar datos 2D como, por ejemplo, imágenes. Este tipo de redes no necesitan una extracción manual de características, ya que realizan dicha extracción de manera automática a partir de las imágenes. Uno de los aspectos más determinantes es que las características más relevantes no se entrenan previamente, sino que se aprenden mientras la red se entrena con un conjunto de imágenes.[11]

Cada parte de una Red Neuronal Convolutiva está entrenada para realizar una tarea, por lo que el entrenamiento de cada una de las partes se desarrolla de manera individual y se efectúa más rápido.

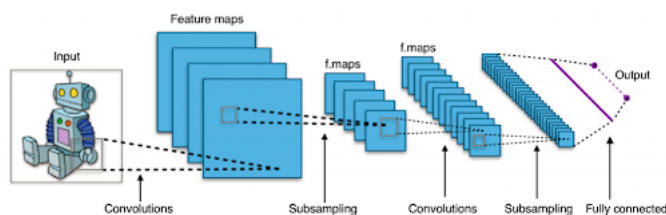


Figura 12: Funcionamiento de una red neuronal

Estas redes trabajan con capas siendo las más típicas la capa de entrada, que generalmente sirve para distribuir entradas en la red; y la capa de salida. Entre estas capas existen capas que se denominan ocultas y pueden existir varios tipos de conexiones entre neuronas ya sea de una misma capa o de diferentes capas. Cada neurona se une con sus capas hasta formar una

red donde se especializan en el reconocimiento en un determinado punto de una imagen, para así reducir el tiempo de trabajo. Una vez reconocido el apartado correspondiente de cada una de las neuronas, los datos se unen en una capa clasificadora para reconocer la imagen en su totalidad. De esta manera, si la imagen que se quiere reconocer es alguna especie de ave, una capa se encargará de reconocer el pico, mientras que otra reconocerá las patas y la siguiente el plumaje o el color. Uniendo todos estos datos, la capa clasificadora deducirá que se trata de, por ejemplo, un canario.

5. Vision por computador

De [11] tal como su nombre indica, el principal objetivo de la visión por computador es intentar imitar la funcionalidad del sistema de visión. Aquellas acciones como diferenciar un coche o diferenciar objetos visibles en un fondo, son acciones que las personas realizan a diario. Esto es una tarea compleja la cual lleva años de investigación para que un ordenador sea capaz de detectarlo con una precisión razonable. Esta nueva funcionalidad de sistema está obteniendo unos avances importantes en los últimos años. La puesta en escena de las redes neuronales profundas y más concretamente de las redes neuronales convolucionales han dado paso a este salto de nivel. El reconocimiento facial, de señales de tráfico o los coches autónomos son algunos de los ejemplos de las aplicaciones que se están consiguiendo. Sin embargo, antes de llegar a este punto hubo un largo camino que es importante mencionar para ver la evolución que ha sufrido la visión por computador. En el año 2001 apareció el algoritmo Viola-Jones capaz de detectar rostros de forma eficaz. El algoritmo detecta rostros

en tiempo real a partir de una cámara web, en aquel momento fue considerada una de las mejores aportaciones al mundo de la visión por computador. Pocos años después, en 2005, apareció Histogram of oriented gradients (HOG). Este nuevo método superó a cretas los algoritmos existentes para el reconocimiento de personas.

5.1. Extracción de Características

Una imagen de entrada contiene una gran cantidad de características. Sin embargo una gran multitud de estas pueden ser irrelevantes y negativas para el paso de la clasificación. Por ello, el primer paso a realizar es la simplificación de la imagen, extrayendo de estas la información más relevante y eliminando aquella que proporciona información significativa para la clasificación. A este proceso se le llama extracción de características. Los siguientes métodos son algunos de los que se utilizaban o utilizan para realizar este proceso.

- Filtros de cascada

Paul Viola y Michael Jones propusieron un método para la detección de objetos utilizando clasificadores en cascada basados en características Haar.[12]

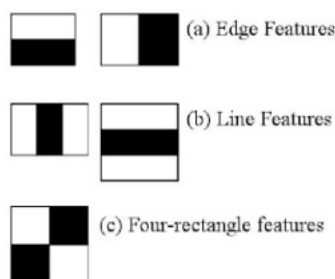


Figura 13: Filtros de Haar

- LBP (Local Binary Patterns):

LBP es un descriptor de texturas introducidos

por Ojala et. al. que calcula una representación local de la textura que se construye comparando cada píxel con los píxeles vecinos. Si la intensidad de un píxel vecino es superior al píxel central, entonces se aplica el valor 0 para el píxel vecino, en caso contrario será 1.[13]

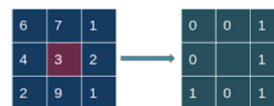


Figura 14: Descriptor LBP que opera en un vecindario 3x3

- HOG (Histogram of Oriented Gradients) [antes mencionado]:

El método HOG fue propuesto en 2005 por Dalal y Triggs y demostró que el método HOG en combinación con máquinas de vectores soporte podía utilizarse para entrenar clasificadores altamente precisos. HOG divide una imagen en ventanas realizando una configuración en rejilla. De cada ventana se obtiene un histograma donde las entradas son las orientaciones de los gradientes.[14]

5.2. Arquitectura de la aplicación

5.2.1. Software :

- Python : Python es un lenguaje de programación de alto nivel, orientado a objetos, con una semántica dinámica integrada, principalmente para el desarrollo web y de aplicaciones informáticas. Es muy atractivo en el campo del Desarrollo Rápido de Aplicaciones (RAD) porque ofrece tipificación dinámica y opciones de encuadración dinámicas.
- Streamlit : Streamlit es un “framework” de

Python de código abierto que permite de manera sencilla e integrada desarrollar aplicaciones gracias a la interacción con otras librerías para su empleo en campos de la teledetección, ciencia de datos, etc.

5.2.2. Hardware:

- Dispositivos móviles.

6. Definición de términos nuevos

- Perceptron : Un perceptrón es una neurona artificial, y, por tanto, una unidad de red neuronal. Efectúa cálculos para detectar características o tendencias en los datos de entrada
- Framework : Es un esquema o marco de trabajo que ofrece una estructura base para elaborar un proyecto con objetivos específicos, una especie de plantilla que sirve como punto de partida para la organización y desarrollo de software.

Referencias

- [1] Onza Perú. Las aves migratorias de los pantanos de villa. 2021.
- [2] J. C. Ponce, A. S. Sprock, A. Torres, A. Casali. Inteligencia artificial. 2014.
- [3] Hebb, Minsky and Papert. Enfoques de clasificación de la Inteligencia Artificial. 1969.
- [4] Adext (2018). INTELIGENCIA ARTIFICIAL. 2 July 2021.
- [5] IBM (s.f). ¿Qué es Machine Learning?. 2021.
- [6] L. J. Sandoval. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS. 2018
- [7] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets, Neural computation, vol. 18 . 2006.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks, in Advances in neural information processing systems. 2012.
- [9] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning requires rethinking generalization. 2016.
- [10] Izaurieta, Fernando. Saavedra, Carlos. (2000) Redes neuronales artificiales. Departamento de Física, Universidad de Concepción Chile.
- [11] Oliva Rodríguez, A. (2018) Desarrollo de una aplicación de reconocimiento en imágenes utilizando Deep Learning con OpenCV. [Tesis de Pregrado] Universidad Politécnica de Valencia.
- [12] Viola, P. y Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. 10.1109/CVPR.2001.990517
- [13] Ojala, T., Pietikainen, M. y Maenpää, T. (2001). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. 10.1109/TPAMI.2002.1017623
- [14] Dalal, N. y Triggs, B. (2005). Histograms of oriented gradients for human detection. 10.1109/CVPR.2005.177
- [15] Mallick, Satya. (2016). A Brief History of Image Recognition and Object Detection. 2018.
