



CONAHCYT

CONSEJO NACIONAL DE HUMANIDADES
CIENCIAS Y TECNOLOGÍAS



CentroGeo

Centro de Investigación en
Ciencias de Información Geoespacial, A.C.

**CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C.**

CentroGeo

Centro Público de Investigación CONAHCYT

**Clasificación de Imágenes de Vertebrados Marinos y
su Entorno Espacial: una Perspectiva Basada en
Aprendizaje Profundo**

TESIS

Que para obtener el grado de

Maestro en Ciencias de Información Geoespacial

Presenta

Emmanuel Morales García

Directora de Tesis

Dra. Daniela Alejandra Moctezuma Ochoa

Co-directora de Tesis

Dra. Ibiza Martínez Serrano

Ciudad de México

mayo 2024

**CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C.**

CentroGeo

Centro Público de Investigación CONAHCYT

**Clasificación de Imágenes de Vertebrados Marinos y
su Entorno Espacial: una Perspectiva Basada en
Aprendizaje Profundo**

TESIS

Que para obtener el grado de

Maestro en Ciencias de Información Geoespacial

Presenta

Emmanuel Morales García

Directora de Tesis
Dra. Daniela Alejandra Moctezuma Ochoa
Codirectora de Tesis
Dra. Ibiza Martínez Serrano

Sinodales
Dra. Tania Aglaé Ramírez del Real
Dr. Jesús Arturo Monroy Anieva
Examinador Externo
Dra. Cecilia Cruz López

Ciudad de México

mayo 2024

RESUMEN

Esta investigación se centra en el desarrollo de las redes neuronales convolucionales (CNNs) para la clasificación de imágenes de tortugas verdes (*Chelonia mydas*) y delfines nariz de botella (*Turpsions truncatus*) en el Parque Nacional Sistema Arrecifal Veracruzano (PNSAV). El objetivo principal es la propuesta de una arquitectura de CNNs que logrará tener exactitud precisa de los problemas de clasificación como ejemplo identificar una especie de otra, enfermedades en las tortugas o tipos de aletas en delfines. Para el desarrollo metodológico se obtuvieron imágenes mediante el proceso de foto identificación, estas imágenes pasaron el proceso de obtener los datos a nivel píxel y posteriormente aplicación de técnicas de aumento de datos para resolver el problema de desbalance de datos y así trabajar las CNNs. Otra parte fundamental es el proceso de métricas adecuadas para evaluar el modelo una vez procesado los datos. Para la parte espacial se empleó el proceso de obtener imágenes de parámetros como la clorofila y la temperatura del PNSAV en sensores remotos como Sentinel-3, con la finalidad de verificar la distribución de las especies con esos dos parámetros. Los resultados obtenidos fueron que los modelos CNNs tienen una precisión del 98% de clasificación correcta de las imágenes, se confirman estos resultados con la validación cruzada en cinco particiones dando robustez a los resultados, se considera que los modelos fueron óptimos y aprendieron correctamente. Por otra parte, la distribución espacial tiene importantes implicaciones para la conservación y el estudio de la vida marina en el parque nacional, al proporcionar una herramienta precisa y eficiente para monitorear y estudiar la población de delfines y tortugas en la región.

Palabras clave: Aprendizaje automático, especies marinos, clasificación, imágenes PR.

ABSTRACT

This research focuses on the development of convolutional neural networks (CNNs) for image classification of green turtles (*Chelonia mydas*) and bottlenose dolphins (*Turpsions truncatus*) in the Parque Nacional Sistema Arrecifal Veracruzano (PNSAV). The main objective is the proposal of an architecture of CNNs that will achieve precise accuracy of classification problems such as identifying one species from another, diseases in turtles or types of fins in dolphins. For the methodological development, images were obtained through the process of photo identification, these images went through the process of obtaining the data at pixel level and later application of data augmentation techniques to solve the problem of data imbalance and thus work the CNNs. Another fundamental part is the process of appropriate metrics to evaluate the model once the data has been processed. For the spatial part, we used the process of obtaining images of parameters such as chlorophyll and temperature of the PNSAV in remote sensors such as Sentinel-3, in order to verify the distribution of species with these two parameters. The results obtained were that the CNNs models have an accuracy of 98% of correct classification of the images, these results are confirmed with the cross-validation in five partitions giving robustness to the results, it is considered that the models were optimal and learned correctly. Moreover, the special distribution has important implications for the conservation and study of marine life in the national park, by providing an accurate and efficient tool to monitor and study the dolphin and turtle population in the region.

Key words: Machine learning, marine species, classification, PR images.

DEDICATORIA

Este proyecto de investigación y logro académico está dedicado a las personas que más amo en el mundo: mi mamá, mi papá y mi hermana. Su amor incondicional, apoyo inquebrantable y constante inspiración han sido la fuerza motriz detrás de cada paso que he dado en este camino. Agradezco profundamente su presencia en mi vida y el invaluable papel que juegan en mi crecimiento personal y profesional. Sin ustedes, este logro no sería posible. Gracias por ser mi mayor motivación y por creer en mí siempre.

Lo que conocemos es una gota, lo que no conocemos es un océano.

(Isaac Newton)

AGRADECIMIENTOS

Agradezco CONAHCYT por otorgarme el apoyo con el número de beca 997728 y con esto lograr el objetivo de culminar la Maestría en Ciencias de Información Geoespacial, en el Instituto de Ciencias de Información Geoespacial en CDMX.

Quiero expresar mi más sincero agradecimiento a dios y la sn por permitirme culminar un escalón más en el ámbito profesional, también a mis padres por todo el apoyo que me brindan y a mi hermana que siempre está para mí, los amo mucho.

A mi directora de Tesis la Dra. Daniela, por siempre tenerme paciencia y sobre todo el gran apoyo que me dio durante mi estancia en la Maestría, sin duda no me equivoque en elegirla, gracias por sus enseñanzas y asesoramiento en mi trayecto escolar. A mi Codirectora la Dra. Ibiza, gracias también por la paciencia, por las imágenes y por la retroalimentación a mi proyecto, es usted un gran ser humano. A mi comité supervisor la Dra. Tania, que siempre estuvo presente en mis evaluaciones cuatrimestrales y a sus observaciones para la mejora de mi trabajo, al Dr. Rodrigo, gracias por sus comentarios y aportaciones, también por la clase de aprendizaje profundo que fue donde nació mi interés por esta gran área de estudio. Finalmente, a mis sinodales la Dra. Cecilia, sin duda la retroalimentación fue muy provechosa durante mi estancia en la FEI y al Dr. Jesús, por sus aportaciones al trabajo de investigación.

Quiero agradecer a mi Ceci (amiguita), gracias por todo el apoyo que siempre me das, muchos consejos que tuve en ese lapso fueron muy provechosos, también los buenos momentos y los malos cuando decaía y me dabas muchos ánimos, sabe que te quiero mucho y gracias nuevamente por adoptarme, eres la mejor.

Martita, gracias madrina por todo el apoyo que me das y ánimos, también uno que otro regaño, pero sobre todo que siempre estas al pendiente de mi vida profesional y quieres lo mejor para mí. Gracias por esos grandes consejos, te quiero mucho y bueno tu sabes lo mucho que luche por este día.

A mis amigas (Hermanas de vida) de la Maestría que fue mejor que me dejo la CDMX, Brendis (la más seria) sabes lo mucho que te quiero y como olvidar los ánimos que nos dábamos mutuamente cuando decaíamos o nos sentíamos solos, pero también como olvidar los buenos momentos que pasamos juntos esos no los voy a nombrar (recuerda secreto). Mari hay hermana como olvidar cuando entramos fuiste la primera en hablarme y el match fue super rápido agradezco todo lo que vivimos juntos igual esos malos momentos que compartimos y nos apoyamos y también el festejo de lograr cosas buenas, te quiero muchísimo. Lupita e Iselita gracias por todo el apoyo siempre las quiero mucho que buenos momentos pasamos en mi estadía por el GEO espero no se olviden de mí. Bueno esto es poco de lo que vivimos juntos pero cada una sabe lo que siento por ustedes y el agradecimiento que tengo por todo lo que pasamos juntos.

Índice General

1	Introducción	
1.1	Planteamiento del problema	1
1.2	Hipótesis	2
1.3	Objetivos	3
1.3.1	General	3
1.3.2	Específicos	3
1.4	Justificación de la investigación	3
1.5	Alcances y limitaciones	4
2	Estado del Arte	6
2.1	Aprendizaje profundo en clasificación de imágenes	6
2.1.1	Clasificación de imágenes de especies marinas	7
3	Marco Teórico	17
3.1	Ecosistemas marinos	17
3.2	Zona de estudio	18
3.3	Vertebrados marinos	19
3.4	Clasificación de imágenes	21
3.5	Aprendizaje automático	22
3.5.1	Métricas para la evaluación del modelo	24
3.6	Redes Neuronales	27
3.7	Red neuronal artificial	28
3.8	Redes neuronales convolucionales (<i>Convolutional Neural Networks</i>)	29
3.9	Componentes de la red neuronal convolucional	31
3.9.1	Capa de entrada (Input layer)	31
3.9.2	Capa de convolución (Convolution layer)	31

ÍNDICE GENERAL

3.9.3	Función de activación	32
3.9.4	Capa de agrupación (Pooling layer)	33
3.9.5	Capa totalmente conectada (Fully connected layer)	33
3.9.6	Capa de salida (Output layer)	33
3.9.7	Función de pérdida (Loss Function)	34
3.9.8	Optimización	34
3.10	Arquitecturas conocidas de <i>CNNs</i>	35
3.11	Clases desbalanceadas	36
3.11.1	Aumento de datos	37
3.12	Validación cruzada (<i>Cross Validation</i>)	38
3.13	Análisis de componentes principales (<i>PCA</i>)	40
4	Conjunto de datos	41
4.1	Descripción de la base de datos	41
4.2	Recopilación de datos	42
4.2.1	Foto-identificación	42
4.2.2	Satélites	45
4.3	Diseño de la base de datos	48
5	Metodología	49
5.1	Definición de la metodológica	51
5.1.1	Descripción de las imágenes originales	51
5.1.2	Aumento de datos	51
5.1.3	Clasificaciones	52
5.1.4	División de datos	55
5.1.5	Algoritmos de clasificación (Aprendizaje de máquina)	56
5.1.6	Red Neuronal Convolutiva (<i>CNNs</i>)	57
5.1.7	Validación cruzada	59
5.1.8	Distribución espacial de las especies en el PNSAV	59
6	Resultados y Discusión	60
6.1	Etapa experimental	60
6.2	Resultados	61
6.2.1	Resultados dos clases: Delfín o tortuga	61

ÍNDICE GENERAL

6.2.2	Validación cruzada: Delfín o tortuga.	64
6.2.3	Resultados dos clases: Tortugas enfermas o sanas	67
6.2.4	Validación cruzada: Tortugas enfermas o sanas.	71
6.2.5	Resultados tres clases: Aletas enteras, otras y típicas	73
6.2.6	Validación cruzada: Aletas enteras, otras y típicas.	74
6.2.7	Resultados de las imágenes satelitales y distribución de los verte- brados marinos	79
6.3	Discusión	84
7	Conclusiones	86
7.1	Reflexión de la investigación	87
7.2	Contribución de la investigación	88
7.3	Trabajos a futuro	89
	Bibliografía	89

Índice de Figuras

Figura 1	Mapa: Polígono del Parque Nacional Sistema Arrecifal Veracruzano (PNSAV). Fuente: (Elaboración propia).	19
Figura 2	Visualización de las características de la especie. Fuente: (Ruíz,2014, pág. 17).	20
Figura 3	Visualización de las características de la especie. Fuente: Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV.	20
Figura 4	Matriz de confusión. Fuente: Elaboración propia.	25
Figura 5	Curva ROC. Fuente: Elaboración propia.	27
Figura 6	Ejemplo de red neuronal biológica. (Fuente: Makinde <i>et al.</i> , 2012 .)	28
Figura 7	Ejemplo de una red neuronal artificial. Fuente: (Elaboración propia).	29
Figura 8	Ejemplo del proceso de una <i>CNNs</i> . Fuente (Elaboración propia).	30
Figura 9	Ejemplo de la convolución para bordes. Fuente (Elaboración propia).	32
Figura 10	Gráfico de las funciones de activación . Fuente (Elaboración propia).	32
Figura 11	Ejemplo de la capa de salidas. Fuente (Elaboración propia).	33
Figura 12	Arquitectura LeNet 5. Fuente: (Golovko <i>et al.</i> , 2017).	36
Figura 13	Ejemplo de clases desbalanceadas. Fuente (Elaboración propia).	37
Figura 14	Ejemplo del proceso de validación cruzada. Fuente (Elaboración propia).	39
Figura 15	Ejemplo gráfico <i>PCA</i> . Fuente (Elaboración propia).	40
Figura 16	Frecuencia de imágenes originales. Fuente (Elaboración propia).	42
Figura 17	Foto-identificación. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	43

ÍNDICE DE FIGURAS

Figura 18	Álbum de imágenes de los vertebrados marinos. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	44
Figura 19	Coordenadas geográficas de las imágenes. Fuente (Elaboración propia).	44
Figura 20	Misión Sentinel-2. Fuente (“Copernicus Data Space Ecosystem”, 2023).	46
Figura 21	PNSAV: Color verdadero (RGB). Fuente (Sentinel-2).	46
Figura 22	Misión Sentinel-3 . Fuente(“Copernicus Data Space Ecosystem”, 2023).	47
Figura 23	Imágenes de temperatura y clorofila PNSAV. Fuente(Sentinel-3).	48
Figura 24	Proceso metodológico	49
Figura 25	Imágenes con deformaciones geométricas para el aumento de datos. Fuente (Elaboración propia).	52
Figura 26	Imágenes de tortuga y delfín. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	53
Figura 27	Imágenes de tortugas. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	53
Figura 28	Clasificaciones de aletas dorsales. Fuente (Sarasota Dolphin Research Program, 2006).	54
Figura 29	Aletas dorsales completas. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	54
Figura 30	Aletas dorsales típica. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	55
Figura 31	Otras aletas dorsales. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).	55
Figura 32	Esquema de arquitectura de <i>CNNs</i> propuesta. Fuente: Elaboración propia.	58
Figura 33	Curva de aprendizaje 1.	62
Figura 34	Clasificaciones basada en las predicciones del modelo	63
Figura 35	Área bajo la curva	64

ÍNDICE DE FIGURAS

Figura 36	Componentes principales para las clases tortugas o delfines	65
Figura 37	Curva de aprendizaje 2.	68
Figura 38	Matriz de confusión del conjunto de entrenamiento con datos aumentados.	69
Figura 39	Predicciones.	70
Figura 40	Curva de ROC con datos aumentados.	70
Figura 41	Componentes principales para las clases tortugas sanas o enfermas .	72
Figura 42	Curva de aprendizaje 3.	75
Figura 43	Gráficas de los conjuntos de entrenamiento y prueba:tres clases. . . .	76
Figura 44	Curva de ROC.	77
Figura 45	Clase verdadera VS predicciones (tres clases).	77
Figura 46	Componentes principales para las clases tipos de aletas en delfines .	78
Figura 47	Mapa de zonas de obtención de imágenes convencionales	79
Figura 48	Coordenadas de las imágenes convencionales y su clasificación por estado de salud.	80
Figura 49	Coordenadas de las imágenes convencionales y la clorofila.	81
Figura 50	Coordenadas de las imágenes convencionales y la temperatura	82

Índice de tablas

1	Clasificaciones de imágenes de tortugas y mamíferos marinos.	14
2	Clasificaciones de imágenes de tortugas verdes y mamíferos marinos	15
3	Clasificaciones de imágenes usando redes neuronales convolucionales con otras especies marinas.	16
4	Transformaciones de imágenes	38
5	Desempeño de los modelos de aprendizaje máquina y profundo	61
6	Experimento con validación cruzada K-Fold=5	65
7	Desempeño de los modelos de aprendizaje máquina y profundo: Clases enferma o sana	67
8	Desempeño de los modelos clases enferma o sana: Validacion cruzada K fold=5	71
9	Desempeño de los modelos de aprendizaje máquina y profundo: Clases de aletas	73
10	Desempeño de los modelos clases de aletas: Validación cruzada k-folds (k=5)	74

Capítulo 1

Introducción

La clasificación de especies marinas es un desafío importante en la investigación y conservación de la biodiversidad marina. A lo largo de los años, los científicos y los expertos en conservación han buscado métodos eficientes y precisos para identificar y clasificar diferentes especies marinas. En este sentido, el uso de métodos de aprendizaje automático y análisis de imágenes de percepción remota ha surgido como una herramienta prometedora.

El aprendizaje automático, también conocido como *machine learning*, es un campo de la inteligencia artificial que se basa en la capacidad de las computadoras para aprender y mejorar su rendimiento a través de la experiencia. Al aplicar técnicas de aprendizaje automático a la clasificación de especies marinas, tales como las redes neuronales convolucionales (*CNNs*) se pueden analizar grandes volúmenes de datos, como imágenes o registros acústicos, para identificar patrones y características distintivas que permitan diferenciar entre especies (Maglietta *et al.*, [2022](#)).

El análisis de imágenes de percepción remota es otra técnica valiosa en la clasificación de especies marinas. La percepción remota se refiere al proceso de adquirir información sobre un objeto o área sin entrar en contacto directo con él. En el contexto marino, esto implica el uso de imágenes capturadas desde satélites, aviones o vehículos no tripulados, que proporcionan una visión general de los ecosistemas marinos a gran escala. Mediante el análisis de estas imágenes, se pueden detectar características específicas de las especies, como patrones de color, forma y distribución espacial, lo que facilita su clasificación y seguimiento.

La combinación de métodos de aprendizaje automático y análisis de imágenes de percep-

ción remota permite a los científicos y conservacionistas abordar desafíos complejos en la clasificación de especies marinas. Este enfoque ofrece la posibilidad de realizar tareas de clasificación de manera más eficiente y precisa, lo que tiene importantes implicaciones para la gestión de los recursos marinos, la conservación de especies en peligro y la comprensión de los ecosistemas acuáticos en general (Wang *et al.*, 2024).

En esta era de avances tecnológicos, el desarrollo y la aplicación de técnicas de aprendizaje automático y análisis de imágenes de percepción remota en la clasificación de especies marinas abren nuevas oportunidades para aumentar nuestro conocimiento y mejorar nuestra capacidad de proteger y conservar la diversidad marina. A medida se continúa explorando y aprovechando los enfoques innovadores, se aceleran los esfuerzos para comprender, conservar y preservar los océanos y sus valiosos recursos.

1.1. Planteamiento del problema

En la actualidad las computadoras pueden interpretar, analizar y comprender imágenes del mundo real, a esta rama de la inteligencia artificial (IA) se le conoce como visión por computadora, este tipo de métodos puede enfocarse en diversas disciplinas con la finalidad de obtener información sobre objetos y reconocimientos de patrones (Wiley & Lucas, 2018). Esta investigación se enfoca principalmente en la clasificación de vertebrados marinos en el Parque Nacional Sistema Arrecifal Veracruzano (PNSAV). La mastozoología marina emplea un método esencial para identificar a las especies, se le conoce como foto-identificación (foto-Id). Las imágenes obtenidas de este proceso muestran información, basadas en las características físicas de las especies. Para los especialistas de esta área, después del proceso de foto-Id, se vuelve complejo el asignar alguna etiqueta a las imágenes, ejemplo conocer si la especie en estudio está enferma o no, esto derivado a la gran cantidad de información que se obtiene.

Basado en lo anterior esta investigación aborda la problemática de la clasificación de características de dos especies, la tortuga verde (*Chelonia mydas*) y el delfín nariz de botella (*Tursiops truncatus*). A pesar de los avances en algoritmos de clasificación supervisada, existen diversos desafíos que dificultan la aplicación de este tipo de metodologías en área como la biología marina. El obstáculo que tiene mayor presencia es la amplia variedad característica como la morfología y coloración que pueden presentar los vertebrados, incluso

dentro de la misma especie. Se propone utilizar, algoritmos de aprendizaje profundo y característica del entorno marino, extraídas a través del análisis de imágenes de percepción remota para abordar de manera más efectiva este desafío.

De esta manera, se busca aprovechar la riqueza de información visual que poseen las especies para desarrollar un sistema automatizado que contribuya significativamente a la investigación y conservación de la biodiversidad marina. Las imágenes de percepción remota (PR) obtenidas a través de satélites ofrecen una ventaja importante al proporcionar información ambiental adicional en el PNSAV. En este proyecto, se utilizan las imágenes de PR con la finalidad de comprender mejor el entorno marino donde se localizan las especies enfermas y sanas.

Las preguntas de investigación que guían este trabajo son:

1. ¿De qué manera puede mejorarse la precisión y eficiencia en la clasificación de imágenes de tortuga verde (*Chelonia mydas*) y delfín nariz de botella (*Tursiops truncatus*) en el Parque Nacional Sistema Arrecifal Veracruzano (PNSAV)?
2. ¿Qué información sobre la tortuga verde (*Chelonia mydas*) y delfín nariz de botella (*Tursiops truncatus*) sanos o enfermos se puede obtener al combinar imágenes de parámetros como la temperatura y la clorofila en el PNSAV?

1.2. Hipótesis

La eficiencia en la clasificación de tortugas verdes y delfines en el Parque Nacional Sistema Arrecifal Veracruzano (PNSAV) mejora al utilizar redes neuronales convolucionales *CNNs* entrenadas con datos diversos y representativos, lo que les permite aprender patrones específicos de estas especies marinas. La integración de conocimiento biológico complementa la información visual de las imágenes, mejorando la precisión de los modelos. Además, al analizar parámetros como la temperatura y la clorofila en el PNSAV, se pueden inferir relaciones entre la presencia de estas especies y su estado de salud, proporcionando información valiosa en su distribución o preferencia de hábitat.

1.3. Objetivos

1.3.1. General

Desarrollar modelos de clasificación de vertebrados marinos que integren imágenes adquiridas en campo, conocimiento biológico y el aprendizaje profundo (redes neuronales convolucionales).

1.3.2. Específicos

- **Clasificar** imágenes de vertebrados marinos a través de un modelo de *CNNs*.
- **Evaluar** la precisión y el rendimiento del sistema de clasificación utilizando métricas apropiadas y comparándolo con métodos de clasificación tradicionales.
- **Analizar** la relación espacial de la presencia de vertebrados marinos (sanos y enfermos) utilizando imágenes de percepción remota.

1.4. Justificación de la investigación

La clasificación automática de imágenes de vertebrados marinos y su entorno mediante métodos de aprendizaje profundo es esencial para la identificación de los vertebrados (tortuga y delfín). La clasificación de imágenes en áreas como la biología marina (fauna) puede ser costosa y lenta en términos de tiempo y propensa a errores humanos. Por lo tanto, la importancia de este trabajo radica en implementar el aprendizaje profundo para aumentar la eficiencia de las clasificaciones.

Los algoritmos de aprendizaje profundo pueden superar las limitaciones asociadas con la identificación manual, como la subjetividad y el error humano. En los últimos años, ha habido un rápido desarrollo de tecnologías de percepción remota, como satélites, drones y vehículos submarinos, que proporcionan una gran cantidad de datos e imágenes de alta resolución de los ecosistemas marinos. Estos avances tecnológicos ofrecen una oportunidad única para aprovechar el potencial del aprendizaje profundo y el análisis de imágenes en la clasificación automatizada de especies marinas.

Estos métodos propios de la inteligencia artificial son accesibles para los usuarios que

requieren procesar este tipo de información, ya que facilitan el procesamiento de imágenes, obteniendo resultados efectivos y multidisciplinarios. La clasificación automática de vertebrados marinos es importante porque preserva la biodiversidad, aprovecha los avances tecnológicos y fomenta la colaboración científica. Esta área de investigación es una herramienta útil para el estudio y la conservación de los ecosistemas marinos.

El uso de modelos de clasificación para imágenes de vertebrados marinos representa un avance significativo para la conservación de la vida marina. Esto brinda beneficios a biólogos marinos, como son identificar especies de forma más precisa y rápida, con esto se tienen resultados óptimos para comprender la distribución y comportamiento de las especie en su entorno natural.

1.5. Alcances y limitaciones

Esta tesis tiene como objetivo principal desarrollar clasificaciones de imágenes sobre vertebrados marinos y su entorno espacial, empleando métodos avanzados de aprendizaje profundo. Se realizan la creación de una arquitectura red neuronal convolucional (*CNNs*) para identificar automáticamente especies marinas a partir de imágenes obtenidas en campo. El alcance se centra en la recopilación de imágenes de las especies marinas (etiquetadas), implementar el aumento de datos, posteriormente el desarrollo de las *CNNs* y para evaluar los modelos utilizar métricas que midan la precisión y eficiencia de las mismas.

Limitaciones

- Disponibilidad de datos: La limitación principal es la disponibilidad de los datos, es decir escasez de imágenes con esto se ve afectado el modelo, sin embargo, se tiene alternativas como el aumento de datos.
- Recursos Computacionales: El tipo de modelos o métodos propios de la inteligencia artificial, requieren gran capacidad para el procesamiento y tener resultados óptimo y rápidos.
- Incorporación de nueva información: Variables ambientales, fisicoquímicos y microbiológicas, recolectadas mediante muestreos biológicos.

- Tiempo del procesamiento de las redes: El rendimiento de la computadora y la calidad de la red son factores que afectan el tiempo para procesar las CNNs. Es importante contar con un equipo con capacidad de procesamiento y conexión de wifi estable para lograr un desempeño óptimo en el entrenamiento de los modelos. .

Capítulo 2

Estado del Arte

En este capítulo se presenta una revisión de investigaciones relacionadas a la tesis realizada. Es fundamental la contextualización del estudio dentro del panorama actual, identificar aportaciones significativas para establecer la base teórica sobre la cual se construye la investigación. Es importante analizar y sintetizar de manera crítica la literatura consultada, teniendo en cuenta los principales resultados, metodologías y limitaciones. Con esto se proporciona un marco referencial para el desarrollo de la presente investigación.

2.1. Aprendizaje profundo en clasificación de imágenes

En la actualidad, la clasificación de imágenes ha tenido una gran aplicación en diversas áreas de investigación. Las aplicaciones más requeridas son con imágenes médicas, ambientales, entre otras, con esto ayudan a la identificación de diversas características en ellas. Para lograr un rendimiento óptimo en el análisis de imágenes, se requieren bases con grandes volúmenes de datos. El objetivo principal es clasificar imágenes para predecir a qué clase pertenecen. Este tipo de aplicación muchas veces se resuelve de manera convencional o por intervención del humano, sin embargo, lograr que las computadoras aprendan es un problema más complejo, por lo tanto, se han diseñado métodos como el aprendizaje profundo para resolver este tipo de tareas (Lorente *et al.*, 2021; Nichols *et al.*, 2018).

Para lograr clasificar imágenes se han desarrollado algoritmos de clasificación clásicas, sin

embargo, necesitan un procesamiento a priori para poder encontrar similitudes entre las imágenes del conjunto de datos. La era tecnológica ha desarrollado avances significativos que permiten realizar de manera automática todo el proceso de clasificación, justamente usando el aprendizaje profundo específicamente las redes neuronales convolucionales (*CNNs*) (Y. Yu, 2022; Lorente *et al.*, 2021).

La importancia del aprendizaje profundo en la clasificación de imágenes radica en la extracción automática de características importantes, con esto se pueden crear modelos con jerarquías que permiten que estos aprendan de las características básicas hasta más complejas, para facilitar la identificación de patrones en las imágenes. La versatilidad de los modelos de aprendizaje profundo se evidencia en la capacidad para adaptarse a diversas tareas de clasificación de imágenes, reconocimiento de objetos y segmentación semántica con el ajuste mínimo en las arquitecturas propuestas. (Panigrahi *et al.*, 2018)

2.1.1. Clasificación de imágenes de especies marinas

Un área de importancia para el análisis de datos es la mastozoología marina, derivado de sus diversos estudios, en particular la foto-identificación de las especies marinas, este método proporciona información importante para la conservación y gestión de los ecosistemas marinos. Este tipo de imágenes que se obtienen de estas metodologías biológicas, son significativas puesto que ayudan al desarrollo de los métodos de inteligencia artificial, para su procesamiento, como lo es el caso del aprendizaje automático y profundo. Con este tipo de herramientas computacionales se pueden reconocer patrones complejos en las imágenes de las especies. Estas tecnologías no solo realizan una óptima clasificación con exactitud acertada, sino también tienen eficacia y rapidez en la aplicación para la conservación de la vida marina.

Hoy en día los estudios derivan en colaboración de las diversas áreas del conocimiento, en esta investigación se considera los métodos computacionales para dar solución a temas en biología marina. Generalmente para la obtención de las imágenes los investigadores de la mastozoología marina ocupan la foto-identificación que es un proceso para estudiar poblaciones animales, como lo demuestra el trabajo de Valdes-Arellanes *et al.* (2011). En su estudio sobre toninas, se encontraron con el desafío de lidiar con una especie poco conocida debido a su gran movilidad y al intercambio esporádico de individuos entre diferentes

poblaciones. Para lograr su objetivo de determinar la abundancia de estas misteriosas toninas, los investigadores se embarcaron en un arduo proceso que implicó realizar 50 navegaciones en el área de estudio. Sumando todas las horas invertidas en el trabajo de campo, alcanzaron una asombrosa cantidad de 236.7 horas de esfuerzo. El proceso de foto-identificación utilizado fue el modelo *Jolly-Seber*, que permitió detectar y reconocer a ciertas toninas en diferentes puntos del área de investigación. A través de este minucioso método, pudieron recopilar datos precisos que contribuyeron a entender mejor la dinámica de esta enigmática especie.

Cabe mencionar que en el proceso de foto-identificación solo se implementa la recolección de imágenes de manera convencional y su análisis es por medio de una metodología manual-visual establecida, para esta investigación no se empleó el uso de la inteligencia artificial o algoritmos inteligentes. Este tipo de métodos clásicos para la identificación de especies es un claro ejemplo de que se tienen muchos obstáculos, como el tiempo, costos, cansancio, etc. Para captar a un individuo de cualquier especie marina, actualmente ya existen métodos automáticos en combinación con análisis manual-visual para poder colaborar en la conservación y preservación de estas especies marinas.

La dificultad para procesar imágenes de manera convencional ha quedado en el pasado, la era computacional tiene métodos que ayudan a las investigaciones de tener resultados más óptimos y rápidos, como lo presentan Ramos-Arredondo *et al.* (2014), donde desarrollaron una investigación basada en un clasificador multiespacial para imágenes de ballenas azules utilizando el método de foto-identificación, en un primer paso los investigadores obtienen las imágenes de las ballenas azules de forma manual, teniendo una clasificación mediante la aleta dorsal. Posterior a esto crean un clasificador que tome en cuenta características como la pigmentación de color, imagen de fondo, tipo de aleta, etc. Reportan que las características extraídas reportan una alta dependencia estadística y esto hace que la imagen este segmentada y se observen a través de hiperplanos. Utilizan *k*-vecinos cercano (*kNN*) y como resultado tienen tres tipos de clasificaciones por el tipo de aleta dorsal con una precisión de 71.66 % con un $k=7$.

Por otra parte Muñoz *et al.* (2014), investigaron la segmentación automática de imágenes en *RGB* utilizando el método *Fuzzy C-Means (FCM)* y la morfología matemática como apoyo para la foto-identificación de cetáceos, centrándose específicamente en la especie

de ballena azul. La investigación se basó en una base de datos que incluía un catálogo de 621 imágenes de ballenas azules en su hábitat natural. Uno de los aspectos fundamentales de la foto identificación de esta especie es la aleta dorsal, ya que suele permanecer fuera del agua durante períodos prolongados, lo que facilita la captura de fotografías para su identificación. Por lo tanto, el objetivo metodológico del estudio consistió en procesar las imágenes originales utilizando el método *FCM*. A partir de este procesamiento, se obtuvo una matriz *fuzzy* y una paleta de colores. Para continuar, fue crucial mantener la paleta de colores, aplicando operaciones de erosión y dilatación a la matriz *fuzzy*. Finalmente, se aplicó el algoritmo *FCM* para llevar a cabo la segmentación de las imágenes y se realizaron operaciones matemáticas para eliminar el ruido presente. Los resultados obtenidos en términos de precisión (*Accuracy*) para la propuesta de segmentación utilizando *Fuzzy C-Means* fueron de 93.63 %, 86.40 %, 80.53 % y 71.47 %, respectivamente. Por otro lado, para la técnica K-means, los resultados fueron de 93.56 %, 90.24 %, 81.32 % y 76.47 %. Estos resultados demostraron la efectividad de las técnicas utilizadas en la clasificación de la especie de ballena azul.

Ramos-Arredondo *et al.* (2015), basan su investigación en un clasificador de pigmentación de la ballena azul aplicando el método *SIFT* en medios no controlados, en este caso usan la foto identificación para tonalidad de la pigmentación de la piel de las ballenas, fue necesario aplicar el método de *SIFT* para extraer características. Emplean métricas como la precisión, sensibilidad, la especificidad y los valores falsos positivos. Una muestra de 500 con 5 tonalidades (clara, clara-moteada, franja, balanceada y oscura). Como resultado se obtuvo una precisión de 81.80 % para la tonalidad balanceada y 80.40 % para la clara-moteada, tomando una muestra de entrenamiento de 100. Para finalizar se aumentó la muestra a 150 y sigue la tonalidad balanceada con un 81 % y se probaron con una muestra de entrenamiento menor es decir 50, la precisión de tonalidad oscura fue de 79.80 % y 78.60 % de la tonalidad balanceada.

Dentro de la literatura relacionada, Paixão *et al.* (2018), desarrollan una investigación donde se consideran herramientas para la identificación de especies marinas empleando la inteligencia artificial, un caso particular es la detección de tortugas marinas, empleando fotos de esta especie y enfocándose específicamente en el color del caparazón, los autores de este trabajo consideran que esto es suficiente para poder realizar una clasificación, se ocupan dos métodos para obtener características basadas en color, la primera es el histo-

grama de color y la segunda es momentos de cromaticidad. Posteriormente implementan los algoritmos de clasificación los vecinos más cercanos (kNN) y máquinas de vectores de soporte (por sus siglas en Inglés *SVM de Support Vector Machine*), confirman que la combinación de algoritmos tiene mejor rendimiento, para finalizar se comprueba en este estudio que combinar la cromaticidad con los kNN , presentan mejor resultado, mientras que el histograma es menos preciso para esta aplicación.

Otro estudio relacionado es el de Hipiny *et al.* (2018), desarrollando la identificación biométrica automatizada de tortugas marinas (*Chelonia mydas*), este método permite el monitoreo de la fauna en este caso marina, sin perturbar el ecosistema. Consiste en emplear cámaras activadas por movimientos colocadas en una altura específica y apuntando hacia abajo, con esto se consigue imágenes del caparazón de estos animales, con esto se procesa el método de histograma de gradiente orientados (*HOG* por sus siglas en Inglés de *Histogram of Oriented Gradients*), teniendo como resultado el 65% de precisión del método empleado.

Gray *et al.* (2018), implementan el uso de drones con redes neuronales convolucionales para detectar tortugas marinas en imágenes, lo cual es importante para su conservación y obtención de información sobre su población. Estos proporcionan un método para evaluar las poblaciones de especies marinas obteniendo imágenes o videos que posteriormente ayudan a los investigadores a procesarlos mediante el uso de la inteligencia artificial. Uno de los métodos más utilizado es las redes neuronales que se consideran una herramienta precisa para clasificar y detectar objetos. Para este estudio de identificación de tortugas se desarrolló una red neuronal convolucional que ayudó a obtener resultados eficientes para la detección de esta especie marina. Entre los resultados encontrados se tiene una precisión del 93% de la clasificación, mientras los resultados de la curva de roc muestran un AUC de 1.

En el trabajo presentado por Dujon *et al.* (2021), se utiliza una computadora de bajo costo para entrenar una red neuronal convolucional para resolver dos tareas: (1) distinguir entre diferentes especies marinas (lobos marinos, tortugas bobas y alcatraces) y (2) rastrear simultáneamente las trayectorias de múltiples tortugas marinas en una estación de limpieza de pescado. Los resultados de este algoritmo obtuvieron un mejor rendimiento en la detección de individuos de tamaño similar que se comportan de manera constante u

ocupan hábitats uniformes. Logró detectar el 94 % de los alcatraces, el 79 % de las focas y el 75 % de las tortugas.

Höschle *et al.* (2021), se enfoca en un estudio de la conservación de la fauna marina y como han puesto en marcha el uso de la percepción remota y la computación para la detección automatizada de varias especies de ballenas con la finalidad de emplear estos métodos para la identificación de estos animales. Se emplean direcciones futuras en las que las imágenes satelitales de muy alta resolución (por sus siglas en inglés *VHR*, *Very high-resolution*) podrían usarse para abordar preguntas urgentes en la conservación de ballenas, es de gran importancia destacar la unión de la biología y la computación que suman fuerzas para implementar métodos como el aprendizaje profundo.

Fischbach y Douglas (2021), han propuesto métodos para poder procesar imágenes satelitales, donde obtienen imágenes ópticas y de radar que varían en resoluciones de píxeles de 40 m a 1 m, estas imágenes ópticas son de Landsat, Sentinel-2, Planet Labs y DigitalGlobe por último también imágenes de radar de apertura sintética (*SAR*) de Sentinel-1 y TerraSAR-X. Con este enfoque se ha estudiado el monitoreo de morsas del pacífico, donde se contaban con tres observadores que examinaron las imágenes para detectar a los animales. Obtienen resultados con las siguientes características, resolución de 4 a 10 m para la detección de las morsas. Los satélites Sentinel-2 y PlanetScope, demuestran cualidades robustas y repetibles para el monitoreo de esta especie.

En el trabajo de Correa *et al.* (2021), se estudiaron ballenas francas australes, se compararon diferentes fuentes de imágenes, como el satélite Pleiades-A1, Sentinel 2, Landsat 8, Rapid Eye y Planet Scope. Los resultados revelaron que las imágenes *VHR* del satélite Pleiades-A1, disponibles en Google Earth, ofrecieron los mejores resultados en comparación con las demás fuentes. Se concluye que este nuevo método de rastreo representa una herramienta importante para la detección de ballenas, sin descartar su utilidad para otros animales marinos. El uso de imágenes satelitales *VHR*, proporciona una capacidad mejorada para monitorear y rastrear la presencia de ballenas, lo que tiene implicaciones significativas para la conservación y gestión de estas especies.

Otro estudio relevante es el realizado por Berg *et al.* (2022), estos autores desarrollaron un análisis comparativo utilizando dos conjuntos de datos de imágenes de alta resolución de animales marinos. El método propuesto superó a la versión original de PaDiM y a otros

enfoques en términos de puntajes de las métricas F1-score y recall. En ambos conjuntos de datos, se obtuvieron resultados destacados, demostrando la eficacia y el potencial del método para la detección de animales marinos. Este estudio destaca la importancia de usar enfoques de supervisión débil, como la modificación de PaDiM, para detectar animales marinos. Los resultados respaldan la efectividad del método propuesto y su capacidad para mejorar la precisión y recuperación en imágenes de alta resolución. Estos hallazgos tienen implicaciones significativas en la conservación y monitoreo de la vida marina, al proporcionar herramientas más efectivas para identificar y rastrear especies en entornos marinos.

Khan *et al.* (2022), en esta investigación se emplea la inteligencia artificial para la identificación fotográfica de ballenas francas, la foto identificación y la inteligencia artificial están siendo utilizadas para identificar ballenas francas en peligro de extinción. Los algoritmos desarrollados por Deepsense.ai alcanzaron una precisión del 87%. Se ha creado una plataforma basada en la web para facilitar la identificación fotográfica semiautomática de ballenas. El sistema también se ha ampliado para incluir ballenas francas australes y se han incorporado catálogos de identificación de diversos países. Se espera que esta colaboración de datos e inteligencia artificial ayude en la conservación y el estudio de las poblaciones silvestres.

Vieira (2023), realizó un estudio sobre el uso de redes neuronales convolucionales para identificar individuos de mamíferos marinos, el objetivo principal de este trabajo es la obtención fotográfica para comparar las aletas dorsales de los individuos avistados en sus diversas escalas espacio temporal. Además, el autor considera mostrar el procedimiento y los obstáculos del uso de las *CNNs* para la discriminación de mamíferos marinos. Para la aplicación de las *CNNs*, se tiene un total de 1138 imágenes que tienen más 856 individuos de ballenas piloto (Calderones), con esto se hacen tres experimentos que abordan el problema de la discriminación en un número de clases mayor. A continuación se escriben los tres experimentos realizados por (Vieira, 2023) :

- Aumento de datos generados artificialmente para simular el conjunto de datos reales.
- Realizan el proceso con las arquitecturas disponibles de *CNNs*: AlexNet, DenseNet121, LeNet, MobileNet y ResNet.
- También consideran otras características para discriminar (arañazos, muescas, re-

dondez, amplitud).

Las redes neuronales convolucionales fueron capaces de identificar a los individuos de ballenas piloto. Para el segundo experimento (2), los resultados no fueron considerables, en cambio para el tercer experimento los resultados fueron bueno ya que se tiene un rendimiento bueno del modelo en las categorías. Como punto final consideran para futuros trabajos usar los árboles de decisiones para mejor rendimiento de los modelos.

Para finalizar esta sección, el estado del arte que se presenta en esta investigación proporciona una visión amplia y diversa sobre estudios aplicativos de la inteligencia artificial y el procesamiento de imágenes en la clasificación de especies marinas utilizados métodos de aprendizaje automático.

Los estudios consultados se centran en otras especies marinas, como son ballenas, peces o morsas, sin embargo, todas las aplicaciones de clasificación e identificación de estas especies comparten el uso de la inteligencia artificial y el procesamiento de imágenes, esto es relevante para la zona de estudio, ya que se demuestra que este tipo de métodos se puede adaptar a diversas especies incluyendo tortugas y delfines en el Parque Arrecifal Veracruzano. En este sentido el uso específicamente de las *CNNs* demuestra efectividad en términos de clasificación de imágenes (características específicas), lo cual es beneficioso para la zona de estudio ya que no se han implementado este tipo de investigaciones.

Pese a que existen diferentes trabajos sobre identificación de especies marinas en la zona de estudio, en esta investigación se tienen nuevas complejidades ya que las especies en estudio tiene diferentes características y solo se enfocan en dos vertebrados marinos, el delfín y la tortuga, así como el uso de métodos de la inteligencia artificial, para resolver problemas de clasificación e identificación con imágenes convencionales de las especies, así como el uso de la percepción remota para combinar imágenes de ciertos parámetros ambientales que caractericen el hábitat de las especies en estudio.

A continuación, se visualizan las Tablas 1, 2 y 3, en la primera y segunda se encuentra de manera resumida la información de la literatura consultada y en la tercera otras investigaciones con enfoque al estudio de peces y de manera general usando aprendizaje profundo y algunas arquitecturas preentrenadas.

Tabla 1: Clasificaciones de imágenes de tortugas y mamíferos marinos.

Autores	Especie	Métodos	Resultados	Aumento de datos
Ramos-Arredondo <i>et al.</i> , 2014	Ballenas Azules	(<i>kNN</i>)	Clasificaciones por el tipo de aleta dorsal con una precisión de 71.66% con un $K=7$.	Sin aumento de datos
Muñoz <i>et al.</i> , 2014	Ballena Azul	<i>Fuzzy C-Means (FCM)</i>	Segmentación utilizando <i>Fuzzy C-Means</i> precisión de 93.63%, 86.40%, 80.53% y 71.47%, respectivamente.	Sin aumento de datos
Ramos-Arredondo <i>et al.</i> , 2015	Ballena Azul	<i>SIFT</i>	Precisión de 81.80% para la tonalidad balanceada y 80.40% para la clara-moteada, tomando una muestra de entrenamiento de 100	100 imágenes aumentadas
Paixão <i>et al.</i> , 2018	Tortugas marinas	<i>KNN y SVM</i>	Identificación precisa sobre la cromaticidad del caparazón	Sin aumento de datos
Hipiny <i>et al.</i> , 2018	Tortugas verde	histograma de gradiente orientados <i>HOG</i>	Resultado el 65% de precisión del método empleado.	Sin aumento de datos
Gray <i>et al.</i> , 2018	Tortugas marinas	CNNs	Resultados eficientes para la detección de esta especie marina.	Sin aumento de datos
Dujon <i>et al.</i> , 2021	Lobos marinos, tortugas bobas y alcatraces	CNNs	Logró detectar el 94% de los alcatraces, el 79% de las focas y el 75% de las tortugas.	Sin aumento de datos
Höschle <i>et al.</i> , 2021	Ballenas	Aprendizaje profundo	Identificación de imágenes con alta resolución	Sin aumento de datos
Fischbach y Douglas, 2021	Morsas	Monitoreo espacial	Mejores imágenes por Sentinel-2 y PlanetScope	Sin aumento de datos
Correa <i>et al.</i>	Ballenas francas	Comparación de obtención de imágenes del satélite	El uso de imágenes satelitales <i>VHR</i> es óptimo para el rastreo de ballenas	Sin aumento de datos

Tabla 2: Clasificaciones de imágenes de tortugas verdes y mamíferos marinos

Autores	Especie	Métodos	Resultados	Aumento de datos
<i>Berg et al.</i>	Animales marinos	PaDiM	Probar una nueva propuesta del software para mejores resultados	Sin aumento de datos
<i>Khan et al.</i>	Ballenas francas	Deepsense.ai	Precisión del 87%. Se ha creado una plataforma basada en la web para facilitar la identificación fotográfica semi-automática de ballenas.	Sin aumento de datos
Vieira	Ballenas pilotos	CNNs preentrenadas	Precisión buena para la clasificación de las imágenes	Con aumento de datos

Tabla 3: Clasificaciones de imágenes usando redes neuronales convolucionales con otras especies marinas.

Autores	Especie	Métodos	Resultados
(Rathi <i>et al.</i> , 2018)	Peces	CNNs	Propuesta de la clasificación automatizada para diversas especies de peces utilizando CNNs, consiguiendo un 96.29% de precisión.
(Hridayami <i>et al.</i> , 2019)	Peces	CNNs VGG19	Clasificaciones precisas de las imágenes con mezclas de GAR y filtro de Canny, en ambos casos consiguiendo un resultado 94.4% y 92.4% de precisión. Mientras que en la combinación de imágenes los resultados fueron más bajos 80.4% y 75.6%.
(Štifanić <i>et al.</i> , 2020)	Peces	CNNs	Clasificaciones óptimas, aplicando RMSprop, teniendo como resultados imágenes de alta calidad.
(Paraschiv <i>et al.</i> , 2022)	Peces	Short CNNs	Sistema de clasificación efectivo que permita identificar y categorizar diferentes especies de peces utilizando una cantidad limitada de datos.
(Zhou <i>et al.</i> , 2023)	Peces e invertebrados	CNNs ResNet 50	Resultados precisos para imágenes submarinas, con precisión entre 97% y 99%.
(Liu <i>et al.</i> , 2023)	Peces domésticos	red Adam-Dropout-CNN (AD-CNN)	Se muestran resultados de la tasa de reconocimiento AD_CNN de 96.97% considerando factores como la luz no uniforme.

Capítulo 3

Marco Teórico

En este capítulo se describe la información necesaria para el desarrollo de esta investigación, partiendo de dos enfoques: conceptual y teórico.

3.1. Ecosistemas marinos

Los vertebrados marinos se encuentran distribuidos a lo largo y ancho de los vastos espacios marinos en todo el mundo. Los océanos y mares brindan un hábitat diversos a estas especies. Cada una ocupa un lugar específico en el ecosistema marino, interactuando con otros organismos y desempeñando roles vitales en el equilibrio y la salud de estos espacios. Estos ecosistemas marinos abarcan desde las aguas superficiales, donde se encuentran las algas y el fitoplancton que realizan la fotosíntesis y sustentan la base de la cadena alimentaria, hasta las profundidades abisales donde existen criaturas peculiares adaptadas a condiciones extremas de presión y falta de luz (Rodríguez & Ruíz, 2010).

La variedad de los ecosistemas marinos permite que las especies marinas se distribuyan en diferentes regiones y hábitat, como arrecifes de coral, manglares, estuarios y zonas costeras. Estos espacios albergan una inmensa diversidad biológica y son hogar de especies únicas y especializadas. Sin embargo, es importante destacar que los espacios marinos y las especies que los habitan enfrentan amenazas significativas, como la sobreexplotación pesquera, la contaminación, la destrucción de hábitats costeros y el cambio climático. Estas presiones pueden tener impactos negativos en la biodiversidad y en la salud de los ecosistemas marinos, poniendo en peligro la supervivencia de muchas especies marinas

(Mancera-Pineda *et al.*, 2013). Por tanto, la conservación y el manejo sostenible de los espacios marinos y las especies que los habitan se vuelven fundamentales para preservar la salud y la diversidad de los océanos. Además, es esencial promover la conciencia y la acción colectiva para proteger estos espacios y garantizar que las futuras generaciones puedan disfrutar y beneficiarse de su belleza y recursos naturales.

3.2. Zona de estudio

El Parque Nacional Sistema Arrecifal Veracruzano (PNSAV) es una de las comunidades de arrecifes de coral más importantes del Golfo de México. El PNSAV se sitúa frente a los municipios de Veracruz, Boca del Río y Antón Lizardo (Figura 1). Sus coordenadas 19°02'24" - 19°16'00" norte y 95°46'19" - 96°12'01" oeste. Es, un área natural protegida denominada Parque Nacional. Tiene una extensión de 52,238 ha, luego se modificó la extensión a 65,516 ha (Martínez-Serrano *et al.*, 2018).

Otras características del PNSAV son (Morales, 2022; Martínez-Serrano *et al.*, 2018):

- La temperatura superficial más alta se registra próximas a los arrecifes y en las zonas con menor profundidad (27.8, 29.7 °C).
- Su clima es considerado como cálido húmedo a subhúmedo con presencia de lluvias en verano.
- Cuenta con una marea diurna y su amplitud es menor a los 15 cm.

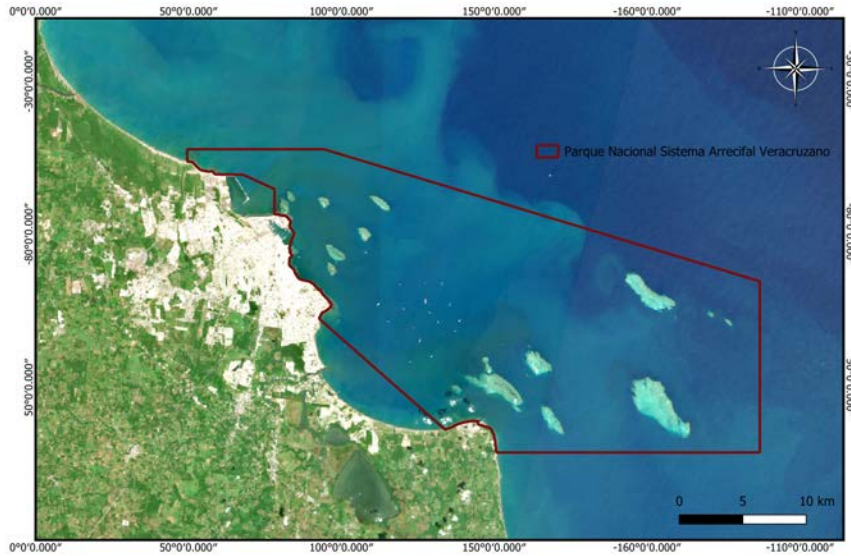


Figura 1: Mapa: Polígono del Parque Nacional Sistema Arrecifal Veracruzano (PNSAV). Fuente: (Elaboración propia).

3.3. Vertebrados marinos

Para comprender mejor el objetivo de esta investigación es importante hacer hincapié en las especies marinas, estas son organismos que habitan en los océanos, mares, y otros cuerpos de agua salada. Incluyen una amplia variedad de seres vivos, desde microorganismos y algas hasta peces, mamíferos marinos, tortugas marinas, crustáceos, moluscos y muchos otros. Para esta investigación se realizó un estudio sobre dos especies específicas como son los delfines nariz de botella (*Tursiops Truncatus*) y la tortuga verde (*Chelonia mydas*).

Delfín (*Tursiops truncatus*) conocido también como nariz de botella, tursión o tonina (Figura 2) y fue descrito por primera vez por Montagü en 1821 (Würsing *et al.*, 2000, como lo cita, Ruíz, 2014). Es un delfín de tamaño promedio, con cabeza robusta, cuenta con una mandíbula inferior sobresaliente a la superior. Su aleta dorsal es grande, ancha y ligeramente curva hacia atrás (Jefferson y Lynn, 1994, como se citó, en Ruíz, 2014). Tiene una aleta caudal y dos aletas pectorales. En cuanto a la coloración tienen un tono gris oscuro en la parte dorsal y por el vientre se considera un gris más claro o blanco, cuentan con una línea oscura que va desde el ojo hasta la aleta pectoral (Ruíz, 2014). El delfín nariz de botella (*Tursiops truncatus*) se encuentra en aguas marinas de tem-

peraturas templadas y tropicales, a lo largo de las costas alrededor del mundo. Una de sus características distintivas es su capacidad para adaptar su alimentación mediante la plasticidad conductual, lo que les permite ajustar sus preferencias alimenticias según la temporada y el entorno geográfico en el que se encuentren (Martínez-Serrano *et al.*, 2011; Ortega-Ortiz *et al.*, 2019).



Figura 2: Visualización de las características de la especie. Fuente: (Ruíz,2014, pág. 17).

Las tortugas verdes (*Chelonia mydas*) (Figura 3), son reptiles y cuentan con una concha que les ayuda a nadar con rapidez y esta se divide en dos partes: Dorsal o carpacho. Esta tortuga es la única especie del género *Chelonia* dentro de la familia *Cheloniidae*, se considera la segunda especie más grande del mundo, entre sus características se pueden considerar que miden 120 cm de largo y pueden llegar a pesar hasta 225 kg. (Mera-Ortiz, 2016; Suárez, 2020; Gómez, 2021). Tienen un Caparazón ovalado, cuatro pares de escudo laterales, de cabeza redonda, con escamas definidas y pequeñas, pico romo, un par de escamas frontales, cuatro pares de escamas postorbitales. Tienen una uña en cada aleta, su coloración es negra en el dorso de las crías y ventralmente blancas y en adultos presentan tonalidad amarillenta ventralmente (Gómez, 2021).



Figura 3: Visualización de las características de la especie. Fuente: Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV.

3.4. Clasificación de imágenes

La aplicación de la clasificación de imágenes actualmente es requerida, por las diversas áreas disciplinares, debido a que una de las ventajas de la identificación o reconocimiento de patrones en una imagen es la cantidad de información que se puede procesar en un tiempo menor. El tipo de información que tienen las imágenes ayudan a la toma de decisiones para los diversos problemas que puedan presentarse en las investigaciones (Azlan *et al.*, 2019; Luo, 2021).

El desarrollo de la clasificación de imágenes se basa de manera general en preprocesar las imágenes, extraer sus características y entrenar esos datos con algún algoritmo de clasificación (Luo, 2021). El tipo de características que se pueden extraer a partir de una imagen digital están relacionadas con el color, la textura, la forma, entre otros aspectos.

En el análisis de imágenes algunos de los métodos tradicionales que se emplean son los que pertenecen a la categoría del aprendizaje supervisado, por ejemplo, las máquinas de soporte vectorial, árboles de decisión, bosques aleatorios, naive bayes, regresión logística, redes neuronales, etc (Azlan *et al.*, 2019).

Tiempo más tarde se presenta una rama especializada del aprendizaje automático, lo nombran aprendizaje profundo ya que se basa en las redes neuronales artificiales y la visión artificial, contando con un método que ha mostrado buen rendimiento en la tarea de clasificación de imágenes, este método son las redes neuronales convolucionales (*CNNs* por *Convolutional Neural Network*). En las *CNNs*, no se necesita el proceso de extracción de características de la imagen a priori, puesto a que el método hace todo para facilitar este tipo de tareas que para los usuarios puede ser laborioso (Deepa & Lalwani, 2019). Es decir, dentro del proceso de aprendizaje de las *CNNs* se contempla la extracción de características “óptimas” para el modelo de clasificación generado.

A continuación, se explican los métodos y sus componentes a configurar para el procesamiento de clasificar imágenes mediante las redes neuronales convolucionales.

3.5. Aprendizaje automático

El aprendizaje automático (*Machine Learning*) es una subdisciplina de la inteligencia artificial (IA) que tiene una gran variedad de métodos que apoyan diversas áreas de estudio para la toma de decisiones. Entre las tareas que puede resolver se incluyen, reconocimiento de patrones, diagnóstico, predicciones, etc. Durante las últimas décadas, el aprendizaje automático, ha revolucionado a la tecnología de la información, con toda la información disponible, existen motivos para decir que el análisis inteligente de datos está presente en la vida cotidiana y es algo necesario para el progreso de la tecnología en las diversas áreas de estudio (Nilsson, 1998; Smola y Vishwanathan, 2008). El aprendizaje automático se puede clasificar en cuatro tipos: aprendizaje supervisado, aprendizaje semi supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. Entre ellos, el aprendizaje supervisado ha demostrado ser el más exitoso en comparación con otros (Ren *et al.*, 2023).

En esta subdisciplina de la IA existen actualmente diversos paradigmas sobre el aprendizaje automático como ya se mencionó anteriormente, los más comunes y empleados en diversas aplicaciones son el supervisado y el no supervisado, en el primero se necesita a priori el conocimiento de las etiquetas del conjunto de datos, mientras que, en el no supervisado, el proceso de entrenamiento no depende de las etiquetas. Para esta investigación se emplearán algoritmos de aprendizaje supervisado, cuya finalidad es obtener una función que, al tener la muestra de los datos de entrenamiento y sus etiquetas, aproxime la relación entre las variables de entrada y la salida. El aprendizaje supervisado se utiliza para resolver problemas de clasificación y regresión (Stack, 2021).

Dentro del aprendizaje automático se encuentra el aprendizaje profundo, es una rama del campo de la inteligencia artificial (IA) que se centra en el entrenamiento de redes neuronales artificiales para aprender y realizar tareas específicas. Se basa en la idea de que las redes neuronales con múltiples capas pueden aprender características y patrones complejos de manera automática a partir de grandes conjuntos de datos (Centeno, 2019).

A diferencia de los enfoques tradicionales de aprendizaje automático, que a menudo requieren que los humanos diseñen manualmente características o reglas específicas para resolver un problema, el aprendizaje profundo permite que los modelos aprendan automáticamente cómo representar la información a partir de los datos de entrada. Esto significa que las redes neuronales profundas pueden descubrir características y relaciones

complejas en los datos sin una programación explícita (Murphy, 2012).

El aprendizaje profundo ha impulsado avances significativos en muchas áreas, como el reconocimiento de imágenes y de voz, el procesamiento del lenguaje natural, la visión por computadora, la traducción automática, entre otras. Algunas de las arquitecturas más comunes utilizadas en el aprendizaje profundo son las redes neuronales convolucionales (CNNs), las redes neuronales recurrentes (RNNs) y las redes neuronales generativas adversariales (GAN) (Centeno, 2019; Padial, 2019).

Es importante destacar que el aprendizaje profundo requiere grandes cantidades de datos de entrenamiento y poder computacional significativo, ya que el entrenamiento de redes neuronales profundas puede ser computacionalmente costoso. Sin embargo, su capacidad para aprender representaciones complejas de los datos ha llevado a avances notables en muchos campos y continúa siendo objeto de investigación y desarrollo activo.

En la vida cotidiana se presentan problemas para entrenar modelos, que, en diversas situaciones, se incluyen parámetros desconocidos y a su vez estos parámetros deben determinarse basándose en las observaciones conocidas. A lo anterior se le conoce como entrenar el modelo. En este caso se requiere una división del conjunto de datos que ayude a tener predicciones más certeras (Gholamy *et al.*, 2018).

Para la generación de modelos de aprendizaje automático y profundo, se necesita el proceso de separar el conjunto de datos reales, en tres partes, datos de entrenamiento, validación y prueba, a continuación, se mencionan.

Los datos de entrenamiento (para entrenar el modelo), son requeridos para la búsqueda de los parámetros $\hat{\theta}$ los cuales generan un resultado al modelo \hat{y}_i y estos se aproximan de manera precisa al valor real y_i . El modelo se adapta a este conjunto de datos (datos de entrenamiento), con la finalidad de aprender patrones y relaciones de los datos de entrada y salida. Es importante contar con una cantidad representativa de datos para garantizar las predicciones de manera más precisas (Nichols *et al.*, 2018; Medar *et al.*, 2017).

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^n L(y_i, \hat{y}_i) \quad (3.1)$$

Donde:

$\hat{\theta}$: Parámetros estimados del modelo.

\hat{y}_i : Valor estimado por el modelo para la observación i-ésima

y_i : Valor real de la observación i-ésima

$$L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

La validación se emplea para realizar el ajuste de hiperparámetros y la evaluación del rendimiento del modelo durante el primer paso (entrenamiento). En este caso al modelo se le realiza una evaluación en el conjunto de validación, para la medición de su rendimiento y con esto hacer el ajuste de los hiperparámetros para la mejora de su capacidad (Nichols *et al.*, 2018; Medar *et al.*, 2017).

Los datos de prueba se utilizan para evaluar el rendimiento del modelo propuesto, pero no deben emplearse para ajustar los parámetros del modelo. Este procedimiento se lleva a cabo una vez completado el entrenamiento y la validación. El conjunto de datos de prueba se utiliza al final para obtener una medida objetiva de qué tan bien el modelo puede generalizar los datos, proporcionando una estimación real del rendimiento esperado. (Nichols *et al.*, 2018).

3.5.1. Métricas para la evaluación del modelo

Al evaluar los modelos de reconocimiento o clasificación de imágenes, se utilizan varias métricas de rendimiento para cuantificar la calidad y la eficiencia del modelo. Estas métricas se utilizan para medir el rendimiento general del modelo en cuanto a precisión y capacidad para predecir los datos. Algunas de las métricas más comunes incluyen exactitud, precisión, medida F1 y matriz de confusión, etc.

La matriz de confusión, se considera una herramienta para medir los errores del modelo es decir, para evaluar y comprobar el rendimiento de las clasificaciones. Es una matriz $n \times n$ asociada a un clasificador, esta a su vez tiene información sobre las clases predichas y las reales. Como se muestra en la Figura 4, se tienen las entradas de dicha matriz, verdaderos positivos (VP), falso negativo (FN), falso positivo (FP) y verdadero negativo (VN) (Ting, 2010; Visa *et al.*, 2011).

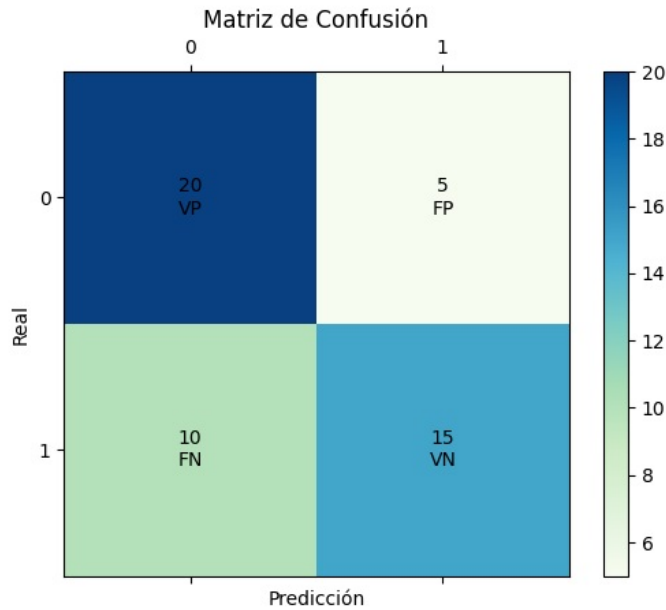


Figura 4: Matriz de confusión. Fuente: Elaboración propia.

Donde:

- Verdaderos positivos (VP): Representa el número de muestras correctamente clasificadas como positivas por el modelo.
- Falsos positivos (FP): Indica el número de muestras incorrectamente clasificadas como positivas por el modelo.
- Verdaderos negativos (VN): Representa el número de muestras correctamente clasificadas como negativas por el modelo.
- Falsos negativos (FN): Indica el número de muestras incorrectamente clasificadas como negativas por el modelo.

Otras métricas de gran importancia para la evaluación del modelo tal como lo mencionan (Visa *et al.*, 2011; Pikabea, 2022):

Exactitud (*Accuracy*) calcula cuantas predicciones correctas, genera el modelo para el conjunto de datos de prueba. se considera una métrica fiable para medir el rendimiento del modelo.

$$\text{Exactitud} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}} \quad (3.2)$$

Precisión calcula los valores predichos correctamente y que resultaron ser positivo esto

determinará si el modelo tiene fiabilidad.

$$\text{Precisión} = \frac{\text{VP}}{\text{VP} + \text{FP}} \quad (3.3)$$

Exhaustividad (*Recall*) en esta métrica se calculan los casos positivos reales que se encontraron en el modelo de manera correcta.

$$\text{Exhaustividad} = \frac{\text{VP}}{\text{VP} + \text{FN}} \quad (3.4)$$

Medida F1(score F1) es basada en la media armónica de precisión, usar esta métrica es acertada cuando la precisión es igual a la exhaustividad (Recall).

$$F1 = 2 \cdot \frac{(\text{Precisión} \cdot \text{Exhaustividad})}{\text{Precisión} + \text{Exhaustividad}} \quad (3.5)$$

La última métrica importante es la curva característica operativa del receptor (ROC por sus siglas en inglés de *Receiver Operating Characteristic*) (Figura 5) es un método que se utiliza para la evaluación del rendimiento de los modelos de clasificación. Sus resultados los muestra gráfica y cuantitativamente mediante el valor del área bajo la curva (AUC), se calcula utilizando la especificidad y sensibilidad, es decir 1-especificidad que se considera los falsos positivos esto en el eje de las x , mientras que para el eje y está la sensibilidad (Nahm, 2022).

$$\text{ROC}(\text{test}) = \text{Sensibilidad} \cdot (1 - \text{Especificidad}) \quad (3.6)$$

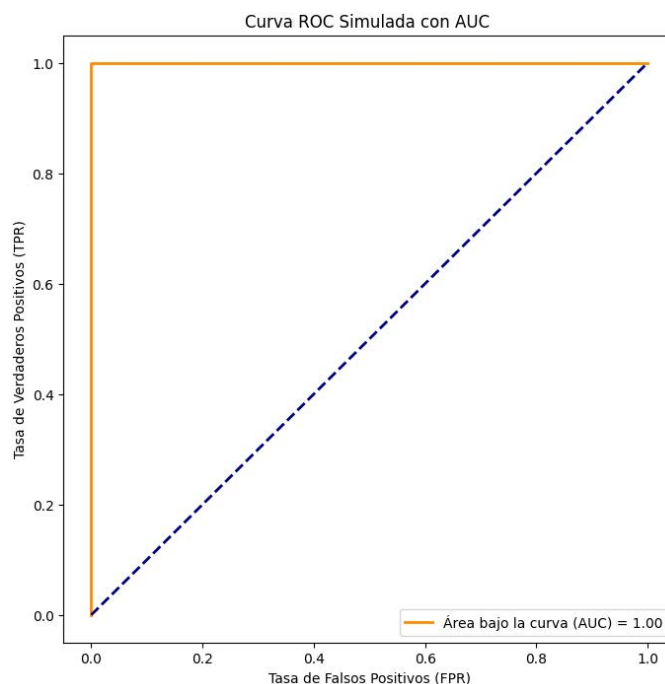


Figura 5: Curva ROC. Fuente: Elaboración propia.

3.6. Redes Neuronales

Las neuronas se pueden representar como modelos neuronales biológicos y componentes conceptuales para circuitos con capacidad de resolver procesos computacionales. Las redes neuronales tienen gran importancia para las diversas áreas multidisciplinares de estudio, aunque principalmente está concebida desde la parte biológica, con la finalidad de mostrar la funcionalidad que tiene en el cerebro y de sus procesos neuronales. Las redes neuronales biológicas constan de cuatro partes (Figura 6): dendritas, cuerpo celular y axón. El proceso que lleva a cabo la red neuronal biológica es que las neuronas naturales reciben las señales mediante la sinapsis que se ubican en las dendritas, cuando esta señal es captada las neuronas se activan y proceden a emitir una señal por medio del axón, el proceso sigue con la transferencia de más señales y hasta enviar la señal a otras sinapsis y dar paso a la activación de otras neuronas (McCulloch y Pitts, 1943; Prieto *et al.*, 2000; Gershenson, 2003).

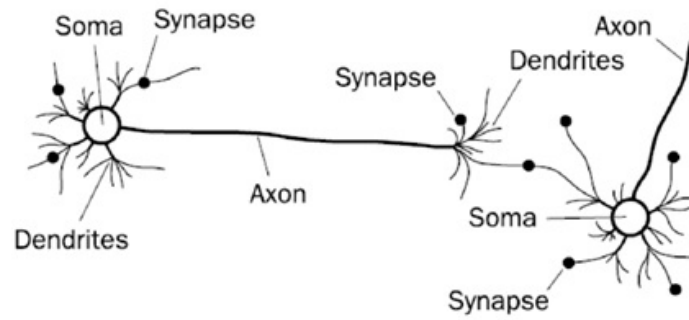


Figura 6: Ejemplo de red neuronal biológica. (Fuente: Makinde *et al.*, 2012 .)

3.7. Red neuronal artificial

Una red neuronal artificial (*ANNs* por sus siglas en inglés) es un algoritmo computacional que se inspira en el sistema nervioso para su funcionamiento. Su objetivo principal es imitar la capacidad de aprendizaje del sistema nervioso, permitiendo así que aprenda a reconocer patrones y asociaciones entre los valores de un conjunto de variables predictivas (entradas) y los estados que dependen de esos valores (salidas) (Trujillo *et al.*, 2004).

Se consideran un modelo matemático complejo compuesto por múltiples elementos de procesamiento organizados en niveles dentro de un sistema computacional, que a su vez tiene numerosos elementos simples. Estos elementos de procesamiento están altamente interconectados y procesan información utilizando su estado dinámico como respuesta a entradas externas. Estas unidades de proceso, llamadas nodos, forman un conjunto que se asemeja a las neuronas debido a su interconexión a través de un entramado de relaciones, similares a las conexiones sinápticas en el sistema nervioso. La señal se propaga a través de la red, comenzando desde los nodos de entrada y avanzando hasta los nodos de salida, donde se genera una respuesta en forma de nivel de activación. Los valores de salida ofrecen una predicción del resultado basada en las variables de entrada (Matich, 2001; Trujillo *et al.*, 2004).

Matemáticamente la *ANNs* se definen como:

$$\text{ANN}_j = \sum_{i=1}^N x_i \omega_{i,j} + \theta_j \quad (3.7)$$

Donde:

x_i : Parámetros de entrada

$\omega_{i,j}$: Pesos sinápticos

θ_j : El umbral

N : El número de parámetros de entrada

A continuación, se muestra la Figura 7, que representa una red neuronal artificial.

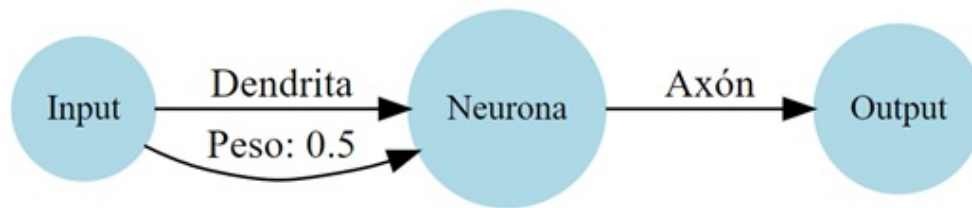


Figura 7: Ejemplo de una red neuronal artificial. Fuente: (Elaboración propia).

3.8. Redes neuronales convolucionales (*Convolutional Neural Networks*)

Las redes neuronales convolucionales (*CNNs*) son un tipo especial de red neuronal diseñado específicamente para procesar datos estructurados en forma de una cuadrícula. Estos datos pueden ser series temporales, que se representan como una cuadrícula unidimensional con muestras tomadas a intervalos de tiempo regulares, o imágenes, que se visualizan como una cuadrícula n -dimensional de píxeles. Las redes convolucionales han demostrado ser altamente efectivas en una variedad de aplicaciones.

El término red neuronal convolucional indica que estas redes utilizan una operación matemática llamada convolución, que es una forma especializada de realizar operaciones lineales. En resumen, las redes convolucionales son redes neuronales que sustituyen la multiplicación de matrices generales por la convolución en al menos una de sus capas, lo cual les confiere ventajas notables en el procesamiento de datos estructurados en forma de cuadrícula (Goodfellow *et al.*, 2016).

Una de las principales ventajas de las *CNNs* en comparación con otros métodos anteriores es su capacidad para identificar automáticamente características relevantes sin requerir

supervisión humana. Las *CNNs* se aplican en diversos campos, como la visión por computadora, el procesamiento de voz y el reconocimiento facial. Las *CNNs* están inspiradas en la estructura de las neuronas presentes en el cerebro humano y animal, específicamente en la corteza visual de los gatos. Utilizan conexiones locales y pesos compartidos para aprovechar eficientemente la información de entrada en formato 2D, 3D y 4D como imágenes. Esta estructura permite reducir la cantidad de parámetros necesarios, simplificando el proceso de entrenamiento y acelerando el rendimiento de la red. A diferencia de las redes neuronales totalmente conectadas, las *CNNs* se centran en regiones específicas de una imagen en lugar de procesarla en su totalidad, lo que les permite extraer correlaciones locales (Alzubaidi *et al.*, 2021).

Para explicar el proceso de manera entendible, se describe genéricamente la arquitectura de una *CNNs* (Figura 8): como primera parte se tiene la entrada que posteriormente pasa a una capa de convolución, para continuar a las capas de agrupación y a las capas completamente conectadas. El cálculo del rendimiento para el modelo se obtiene con la función de pérdida, empleando la propagación hacia adelante utilizando el conjunto de entrenamiento, y los parámetros se actualizan con el valor de pérdida utilizando la propagación hacia atrás con algoritmos como el descenso de gradiente, *ADAM*, entre otros (Yamashita *et al.*, 2018; Hernández *et al.*, 2022).

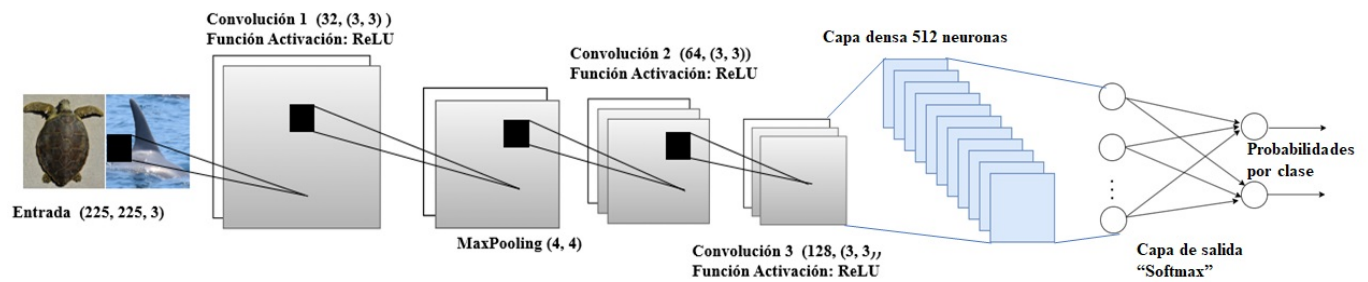


Figura 8: Ejemplo del proceso de una *CNNs*. Fuente (Elaboración propia).

Un componente importante de las *CNNs* es la convolución que es la operación matemática que se realizan sobre dos vectores o matrices. En teoría de señal, la convolución ocurre entre dos señales continuas unidimensionales (Bonilla, 2020). Matemáticamente, una convolución se representa de la siguiente forma: $(x*w)(a)$, donde x y w son funciones (Teuwen & Moriakov, 2020).

Para el caso continuo,

$$(w * w)(a) = \int_{-\infty}^{\infty} x(t)w(a - t) dt \quad (3.8)$$

Donde:

$x(t)$: Es la entrada

w : El filtro o kernel

Para el caso discreto se visualiza de la siguiente manera:

$$(w * w)(a) = \sum_a x(t)w(t - a) \quad (3.9)$$

3.9. Componentes de la red neuronal convolucional

Las CNNs se componen de varias capas, cada una de ellas cumplen una función específica en el proceso de aprendizaje y extracción de características de las imágenes de entrada. Es importante tener el conocimiento de las funciones y componentes claves de una CNNs, a continuación se describen cada una de las capas de este modelo de aprendizaje profundo.

3.9.1. Capa de entrada (Input layer)

Se trata de la primera capa que la red es la que recibe a los datos de entrada. Las neuronas en esta capa tienen una característica o atributo. Un ejemplo son las imágenes, cada neurona en la capa de entrada es la representación de un píxel de la imagen. Para esta capa no se necesita ningún cálculo, simplemente pasa los datos de entrada a la siguiente capa de la red (Zhang *et al.*, 2019).

3.9.2. Capa de convolución (Convolution layer)

Se utiliza para la extracción de las características que contiene la imagen. Existen dos niveles de extracción la baja que solo puede obtener bordes, líneas y esquinas, mientras el alta obtiene características adicionales. Esta es una matriz de 3×3 o 5×5 es decir de $k \times k$ haciendo referencia al kernel (k), donde k puede tomar cualquier valor impar. También con este método se consigue tener mejor nitidez de la imagen (López, 2019; Zhang *et al.*, 2019).

Un ejemplo de convolución para obtener bordes verticales y horizontales se puede ver en la Figura 9.

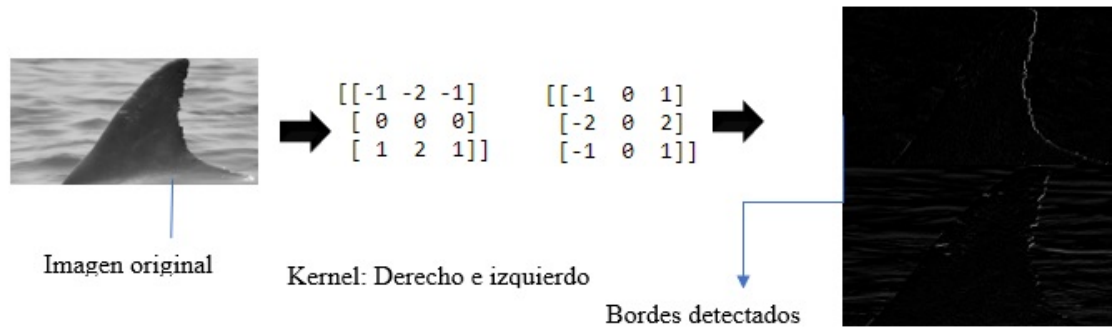


Figura 9: Ejemplo de la convolución para bordes. Fuente (Elaboración propia).

3.9.3. Función de activación

Es un parte fundamental para el proceso de la red neuronal convolucional ya que se necesita una función de activación, cuando se tiene una o varias entradas, esto define la salida de una neurona en la red indicada (López, 2019). De acuerdo con la literatura consultada las funciones más importantes para este tipo de método son: La sigmoide, la tangente hiperbólica y la ReLU, siendo esta última la más recomendable puesto que es más rápida que las primeras dos funciones. En la Figura 10 se observan las gráficas de estas funciones.

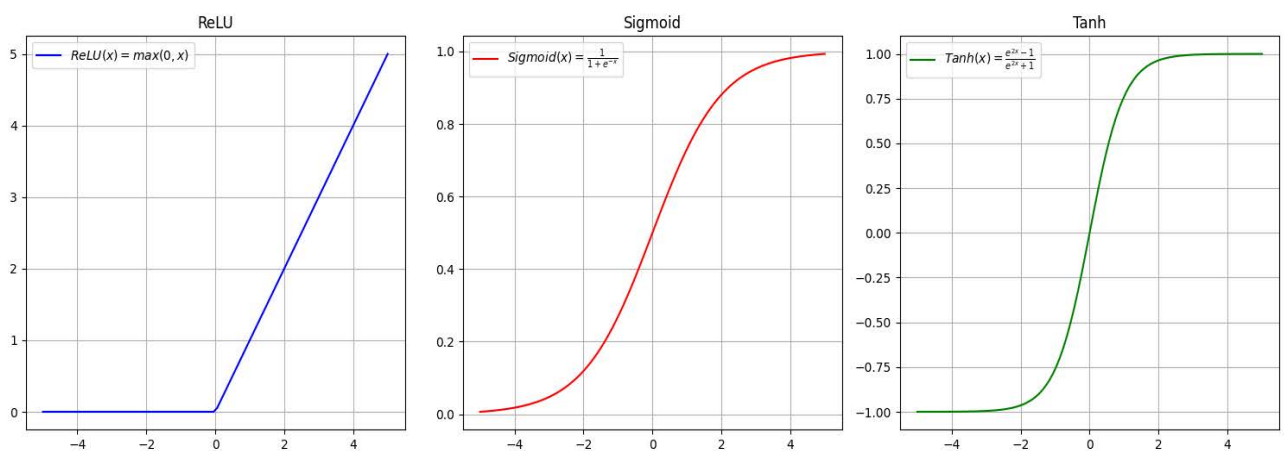


Figura 10: Gráfico de las funciones de activación . Fuente (Elaboración propia).

3.9.4. Capa de agrupación (Pooling layer)

Es utilizada en *CNNs*, generalmente su tarea es tomar el mapa de características a un tamaño más pequeño (es decir se realizan submuestreos). Otro aspecto es que esta capa determina la parte más importante que es el desempeño del modelo, también ayuda a la reducción de la dimensión espacial, esto a su vez interviene en la reducción del costo computacional de la representación de la imagen, minorizando el número de parámetros (Özdemir, 2023).

3.9.5. Capa totalmente conectada (Fully connected layer)

Generalmente la capa de agrupación se aplana, en otras palabras, se convierte en una matriz, unidimensional de vectores y esta se conecta a una o más capas completamente conectadas, que también se les conoce como capas densas. Cuando se tienen las características que se obtienen por las capas de convolución y el submuestreo por las de agrupación, entonces un subconjunto de capas completamente conectadas se asigna a la red final, esta capa tiene que ir de la mano de la función no lineal, alguna de las ya mencionadas anteriormente (Yamashita *et al.*, 2018).

3.9.6. Capa de salida (Output layer)

La capa final de la *CNNs* produce la salida deseada, su tarea principal es mostrar el resultado que se obtuvo de las capas anteriores, como ejemplo ver la Figura 11 donde se tiene 95 % de probabilidad de que la imagen de entrada sea un delfín, contra un 5 % que no lo sea (Zhang *et al.*, 2019).

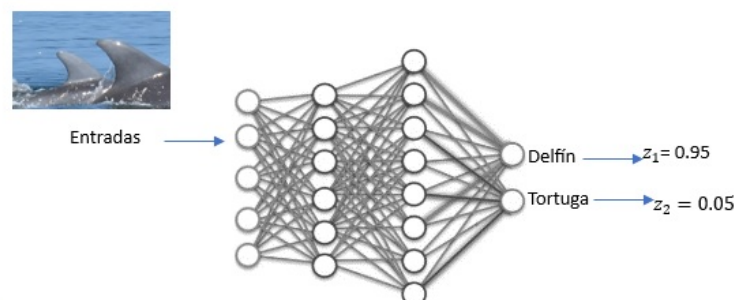


Figura 11: Ejemplo de la capa de salidas. Fuente (Elaboración propia).

3.9.7. Función de pérdida (Loss Function)

Es una medida que se utiliza en algoritmos de aprendizaje automático y aprendizaje profundo para cuantificar qué tan bien el modelo predice los valores deseados en comparación con los valores reales (etiquetas). La función de pérdida calcula la diferencia entre las predicciones del modelo y las etiquetas reales y genera un valor numérico que representa el error de predicción (Tian *et al.*, 2022). Para esta investigación se utilizó entropía cruzada (cross-entropy), la fórmula es:

$$H(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (3.10)$$

Donde:

N es el número de muestras en el conjunto de datos.

y_i es la etiqueta real de la muestra i (0 o 1).

\hat{y}_i es la predicción del modelo para la muestra i (valor entre 0 y 1).

La función \log es el logaritmo natural.

3.9.8. Optimización

Para la optimización de las *CNNs* se empleó la estimación del momento adaptativo (*ADAM*) es considerado un algoritmo de optimización que se basa en el gradiente de primer orden con funciones objetivos estocásticas. Es un método sencillo para aplicar y tiene invariancia a los cambios de escala de los gradientes. Es aplicable cuando se tiene volúmenes de datos grandes y también para objetivos no estacionarios. Es un método que calcula la tasa de aprendizaje adaptativo para los parámetros (Kingma & Ba, 2014).

Algunas ventajas (Kingma & Ba, 2014):

- Actualización de parámetros son invariantes
- Los hiperparámetros están limitados por los pasos
- No se requiere un objetivo estacionario
- Funciona con pocos gradientes
- Procesa un recocido

Fórmula para actualizar los parámetros empleando *ADAM*

Inicializar los momentos

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla J(\theta_t) \quad (3.11)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (\nabla J(\theta_t))^2 \quad (3.12)$$

m_0 y v_0 , son inicializados como vectores en 0.

Corregir el sesgo que pueda presentarse

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.13)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.14)$$

Se actualizan los parámetros

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (3.15)$$

Donde:

$J(\theta)$: Función de pérdida

$\nabla J(\theta_t)$: Gradiente de la función

t : El tiempo

α : Tasa de aprendizaje

m_t : Primer momento

v_t : Segundo momento

ϵ : Pequeño término para evitar la división por cero

3.10. Arquitecturas conocidas de *CNNs*

Las arquitecturas de redes neuronales convolucionales (*CNNs*) han revolucionado el campo de la visión por computadora y el reconocimiento de imágenes. Desde la introducción de LeNet (LeCun *et al.*, 1998) y su equipo, hasta el desarrollo de EfficientNet (M. Tan &

Le, 2019), se han creado diversas arquitecturas con el objetivo de mejorar la precisión y eficiencia en tareas de clasificación de imágenes. Estas arquitecturas, como AlexNet (Krizhevsky *et al.*, 2012), VGGNet (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy *et al.*, 2015), ResNet (K. He *et al.*, 2016), DenseNet (Huang *et al.*, 2017), MobileNet (Howard *et al.*, 2017) han sido diseñadas para abordar desafíos específicos, utilizando capas convolucionales, submuestreo, conexiones residuales, entre otros elementos, para extraer características relevantes y lograr un rendimiento excepcional en la clasificación de imágenes. A continuación se observa la arquitecturas de LeNet 5 (Figura 12).

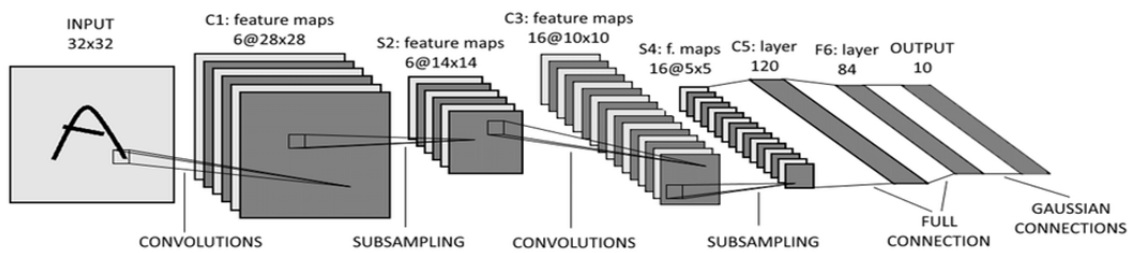


Figura 12: Arquitectura LeNet 5. Fuente: (Golovko *et al.*, 2017).

3.11. Clases desbalanceadas

Se consideran datos desbalanceados si las muestras de clases no son iguales o cercanas para este estudio se tienen datos con clases cercanas (Figura 13). Los datos son una gran herramienta para comprender el descubrimiento del conocimiento a partir de análisis que ayudan y refuerzan el proceso de estos para tomar decisiones en las diversas aplicaciones que lo requieren. Sin embargo, los datos reales, pueden presentar el desequilibrio de clases, esto afecta cuando se trabajan modelos de clasificación supervisada, ya que el desempeño del algoritmo se puede ver perjudicado, es decir pueden presentar exactitud no favorable (H. He y Garcia, 2009; L. Yu y Zhou, 2021).

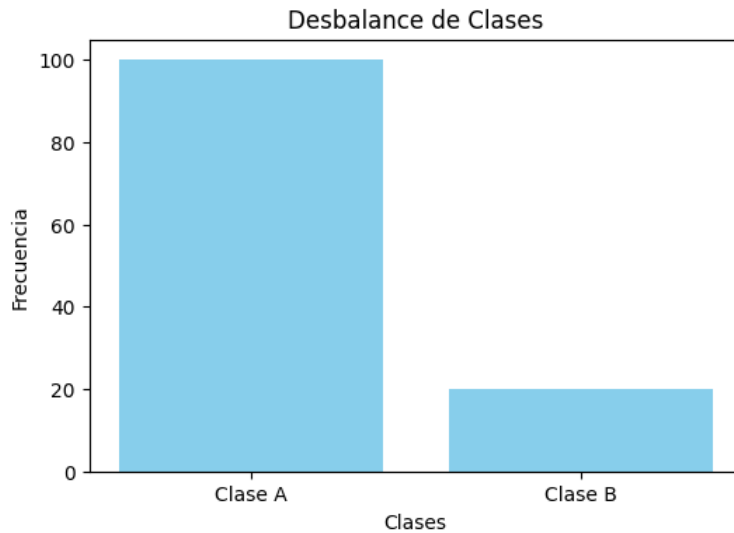


Figura 13: Ejemplo de clases desbalanceadas. Fuente (Elaboración propia).

3.11.1. Aumento de datos









El aumento de datos proporciona una solución al problema de tener pequeñas cantidades de datos, como ejemplo cuando se tienen imágenes que requieren tiempo o disponibilidad del objeto de estudio para obtenerlas. Este tipo de método engloba procesos para mejorar el tamaño y la calidad de los conjuntos de datos de entrenamiento, para beneficiar a los modelos de aprendizaje profundo. Los datos aumentados ayudan a la disminución del error de entrenamiento, también representa un conjunto más completo de posibles puntos y minimiza la distancia entre el conjunto de entrenamiento y validación. Además, este método da paso a ser una opción para resolver el problema de desbalance de clases mediante el submuestreo (Yorioka *et al.*, 2020; Shorten y Khoshgoftaar, 2019). Para fines de esta investigación se emplearon diversas deformaciones geométricas a las imágenes, que se pueden consultar en el apartado de metodología.

El manipular imágenes de forma básica se refiere a las transformaciones como rotación, zoom, relleno de píxeles, entre otros. Este tipo de técnicas modifican las imágenes directamente y son más fáciles para su procesamiento. Es importante mencionar que la región que se observa fuera del límite de la imagen tiene que ser constante en 0, pues muchas veces será negro después del proceso, el objeto (imagen) no debe moverse fuera del marco (Yang *et al.*, 2023).

En esta investigación se empleó la manipulación básica para el aumento de datos. Para visualizar las transformaciones se ejemplifican con una imagen del delfín nariz de botella

(Tabla 4).

Tabla 4: Transformaciones de imágenes

Transformación	Resultado
Escalado de píxeles	
Rotación aleatoria	
Desplazamiento horizontal aleatorio	
Desplazamiento vertical aleatorio	
Inclinación de imagen	
Zoom en las imágenes	
Reflejo horizontal	
Relleno de píxeles en deformaciones	

3.12. Validación cruzada (*Cross Validation*)

Es una técnica empleada en el aprendizaje automático y profundo para la evaluación del rendimiento de los modelos. La *CV* divide el conjunto de datos en subconjuntos pequeños, que se les conoce como “*folds*” o particiones, esto se realiza para entrenar el modelo en varios pliegues y generar la evaluación (Figura 14). Este proceso se realiza de manera iterativa, para cada subconjunto (Refaeilzadeh *et al.*, 2009; White y Power, 2023).

La importancia de *CV* es la estimación para verificar el comportamiento del modelo en los datos no vistos, es decir, datos que aun no se utilizan en el entrenamiento. Este procedimiento es importante ya que evita el sobreajuste (*overfitting*) en los modelos y con esto tener la capacidad de generar nuevos datos (Yadav y Shukla, 2016).

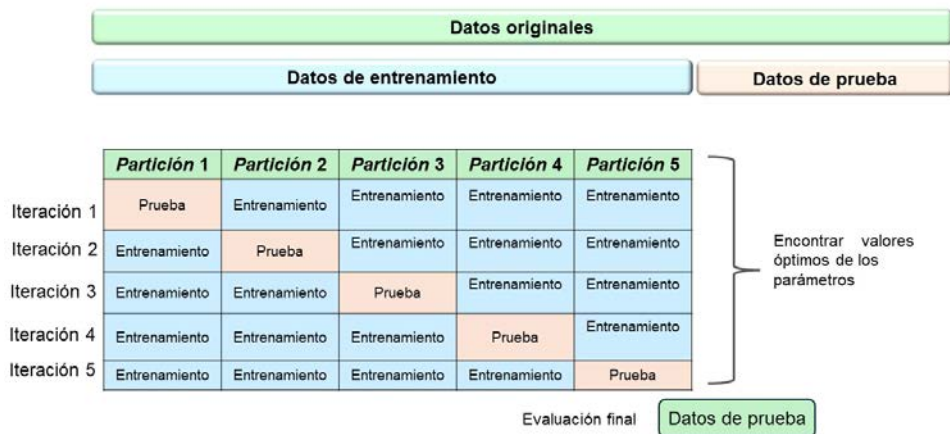


Figura 14: Ejemplo del proceso de validación cruzada. Fuente (Elaboración propia).

La validación cruzada puede ser expresada matemáticamente como:

$$CV_{(k)} = \frac{1}{K} \sum_{k=1}^K L(\text{modelo entrenado en } D_{-k}, D_k) \quad (3.16)$$

Donde:

D : Conjunto de datos.

K : Número de particiones en CV.

D_k : Conjunto de datos utilizado en la prueba k-ésima iteración.

D_{-k} : Conjunto de datos restante utilizado en el entrenamiento de la k-ésima iteración.

L : Función de pérdida evaluar el modelo.

3.13. Análisis de componentes principales (*PCA*)

El análisis de componentes principales (*PCA* por sus siglas en inglés) es una técnica de estadística y también usada en el aprendizaje automático para reducir dimensionalidad en un conjunto de datos. El objetivo principal es obtener un conjunto de nuevas variables a las que se le denomina componentes principales, se captura la mayor cantidad de variabilidad en los datos originales. A estas nuevas variables son consideradas como combinaciones lineales de las variables originales y son ortogonales entre sí (Peña, 2002).

El *PCA* ayuda a la exploración de estructuras subyacentes de los datos, para la identificación de patrones y la reducción del ruido. Al pasar los datos a una menor dimensión facilita la interpretación de estos (Peña, 2002).

Para la parte grafica el *PCA* muestra las observaciones (filas de la matriz) en un espacio bidimensional definido por los primeros dos componentes principales (Karamizadeh *et al.*, 2013). Cada observación se presenta como un punto en el plano cartesiano y los colores que representa cada punto o individuo es para formar grupos o clases de estas (Figura 46).

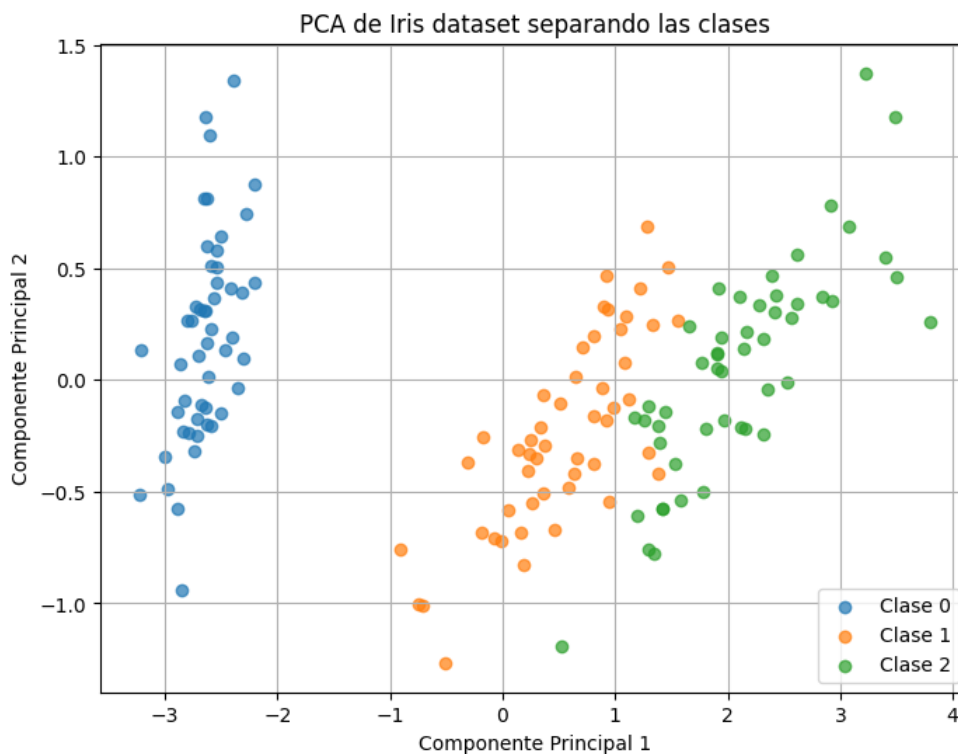


Figura 15: Ejemplo gráfico *PCA*. Fuente (Elaboración propia).

Capítulo 4

Conjunto de datos

En este capítulo se describen las imágenes empleadas para realizar el proceso de *CNNs*, como primera parte se tienen los datos recolectados en campo y posteriormente las obtenidas por los sensores remotos.

4.1. Descripción de la base de datos

En este trabajo de investigación se utilizaron imágenes como la fuente principal de datos, el conjunto de imágenes que corresponden a dos especies marinas: la tortuga verde (*Chelonia myda*) y el delfín nariz de botella (*Turpsion Truncatus*). En el caso de la tortuga verde, se cuenta con imágenes del caparazón, perfiles y la cabeza. En cuanto a los delfines, se disponen de imágenes de las aletas dorsales tanto del lado izquierdo como del derecho. En total, el álbum de imágenes consta de 210 imágenes de tortugas (45.79%) y 238 imágenes de delfines (54.21%) ver la Figura 16. Las imágenes de los vertebrados marinos fueron proporcionadas por la Dra. Ibiza Martínez Serrano, de la Universidad Veracruzana. Obtenidas del proyecto “programa de monitoreo biológico basado en especies indicadoras de integridad ecológicas en el PNSAV”.

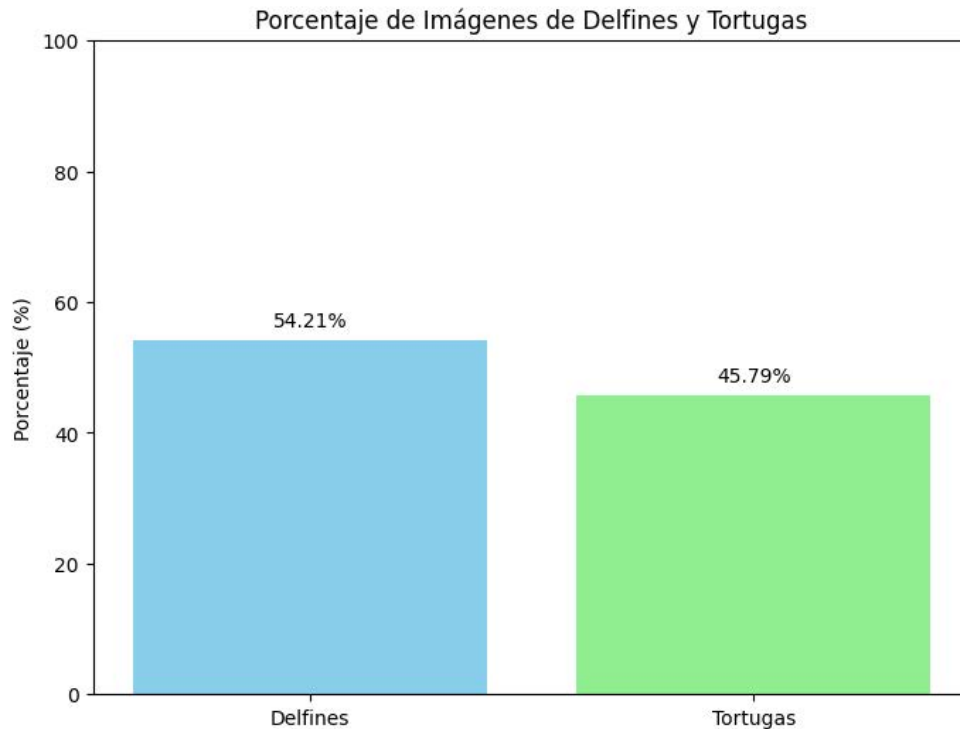


Figura 16: Frecuencia de imágenes originales. Fuente (Elaboración propia).

4.2. Recopilación de datos

4.2.1. Foto-identificación

Las imágenes fueron proporcionadas por especialista en fauna marina, para la obtención de estas imágenes utilizaron la foto-identificación que es considerado un método fundamental para el estudio de residencia, el volumen de poblaciones, movimientos y organización social de los mamíferos marinos. Se considera un método no invasivo y de presupuesto accesible para la identificación de individuos. En este método se realiza trabajo de campo (embarcaciones), con la finalidad de obtener fotografías de características particulares, según la especie que se investiga (Figura 17). La foto-identificación tiene un gran avance en las últimas décadas debido al desarrollo de la tecnología (cámaras digitales con mejor resolución). Este método se considera confiable, debido a que ayuda a comprender la ecología de las especies marinas y con esto se logra realizar acciones para la conservación y manejo de estas (Gómez-Salazar *et al.*, 2011).

Este método es utilizado para lograr visualizar la presencia de marcas naturales en el cuerpo de las especies, las marcas que se consideran son muescas, cicatrices, marcas de dientes, mutilaciones y coloración, se considera un método de gran utilidad para el mo-

nitoreo de especies que permite conocer la preferencia de hábitat (Gómez-Salazar *et al.*, 2011; Ruíz, 2014; Gómez, 2021).

El proceso de la foto-identificación inicia con el recorrido marítimo, se establecen temporadas, describen el tipo de eslora y motor, así como la lancha. Posteriormente a esto el tipo de recorrido y el tiempo en los que se buscará a la especie, así como considerar paradas específicas para la observación (Martínez-Serrano, 2005).

Otros aspectos importantes:

- Registro de fecha
- Hora
- Posición geográfica
- Condiciones ambientales

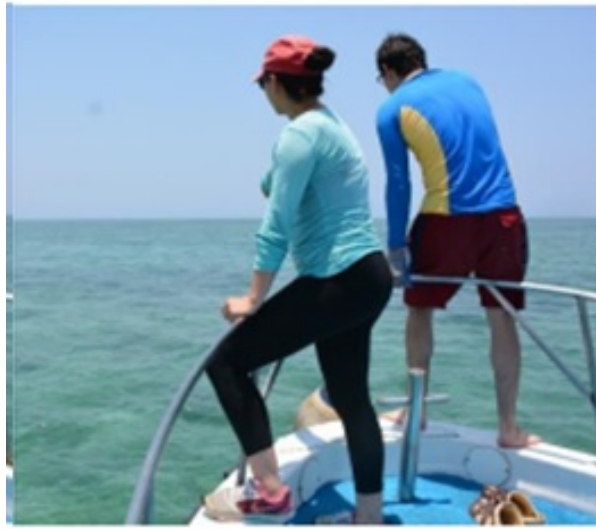


Figura 17: Foto-identificación. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

Cuando la especie objetivo es captada, se realiza un acercamiento para captar la fotografía (cámara especial). Con estas fotografías se realiza el proceso de foto identificación. Estas tienen un proceso de revelado, posterior a esto se toman en cuenta solo las mejores fotografías de acuerdo con su calidad. También se seleccionan las características como, el enfoque, la muestra de aleta dorsal, fracciones anterior y posterior del cuerpo del animal, con todo esto se logra crear catálogos (Figura 18). Otra característica de las imágenes es que para ambas especies fueron captadas en aguas claras. En la siguiente sección se describen las clasificaciones y algunas características en particular como son, marcas en las aletas de los delfines o rastros de tumores en las tortugas.

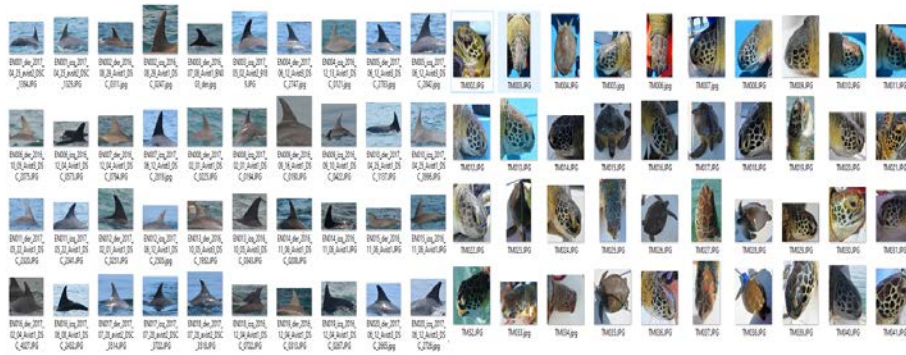


Figura 18: Álbum de imágenes de los vertebrados marinos. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

En la Figura 19 se observa el mapa que muestra las coordenadas de cada una de las imágenes obtenidas (polígono del PNSAV).

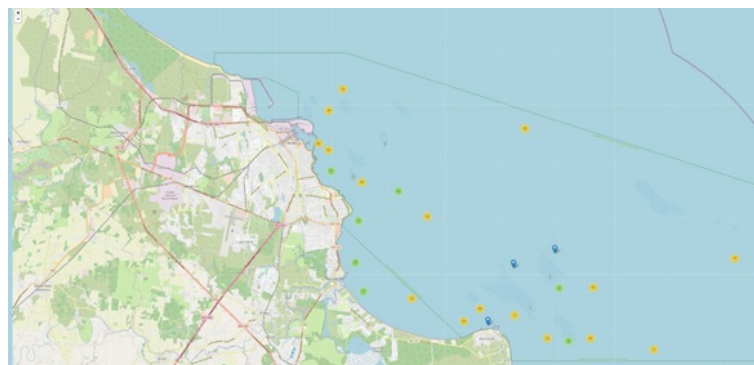


Figura 19: Coordenadas geográficas de las imágenes. Fuente (Elaboración propia).

Para complementar el estudio se hizo uso de imágenes de percepción remota con la fina-

lidad de observar el entorno en el Parque Sistema Arrecifal Veracruzano, esto sirve para conservar y preservar el ecosistema marino. En este caso, los datos proporcionados por la percepción remota se utilizan para evaluar las relaciones espaciales, temporales y ambientales del PNSAV y la distribución de los vertebrados marinos. Estas técnicas permiten la vigilancia de la biodiversidad en todas las áreas costeras y la contribución al estudio ecológico a largo plazo (Kavanaugh *et al.*, 2021).

El impacto de los satélites en los océanos es de gran ayuda para la observación de los ecosistemas marinos, justamente para la obtención de información propia de los océanos, los tipos de variables que se pueden obtener son por ejemplo las empleadas en (Ho & Liu, 2023) como la clorofila, la temperatura de la superficie, la altura de la superficie del mar, el viento de la superficie del mar, la salinidad del mar, el hielo marino y la corriente de la superficie del mar.

Para esta investigación se usaron imágenes del entorno marino del PNSAV, considerando, el entorno en su estado normal, temperatura y clorofila, para poder relacionar las imágenes convencionales (especies) y su hábitat, otro aspecto importante fue la fecha de extracción de estas, comprendido un periodo del 2016 al 2019. La finalidad de este proceso es estudiar la relación del entorno marino y la distribución de las especies. Estas imágenes se obtuvieron de las misiones Sentinel-2 y 3.

4.2.2. Satélites

Para extraer las imágenes del entorno marino se utilizaron datos georreferenciados, que fueron recolectados durante los muestreos que realizaron el equipo de investigación biológica en el PNSAV. Para fines de esta investigación se implementó el uso de la misión Sentinel-2 y 3.

La misión Sentinel-2 (Figura 20), es destinada para observación de la tierra implementada por la Agencia Espacial Europea (ESA), es el único programa de vigilancia medioambiental, Sentinel 2 tiene una cámara multiespectral de alta resolución, con 13 bandas espectrales, abarca un campo de visión de 290 km de anchura y sobrevuelos. Además, esta misión es empleada para verificar y alertar la contaminación de lagos y aguas costeras (“Copernicus Data Space Ecosystem”, 2023)



Figura 20: Misión Sentinel-2. Fuente (“Copernicus Data Space Ecosystem”, 2023).

Para el caso de Sentinel-2, se utilizó el nivel 2A, debido a la corrección atmosférica que tiene, es decir presentan una mejor resolución en las imágenes y con esto se puede observar mejor el área de estudio, cabe mencionar que también se indicó un rango de [5-10 %] de nubosidad (European Space Agency, 2023).

Las imágenes extraídas en este nivel son en la categoría color verdadero (True Color), basado en las bandas, B4: rojo, B3: verde y B2: azul (Figura 21). Esto con la finalidad de tener imágenes generales del entorno marino (“Copernicus Data Space Ecosystem”, 2023). El color verdadero se obtiene de la siguiente manera:

$$\text{Color Verdadero} = \begin{bmatrix} B4 \\ B3 \\ B2 \end{bmatrix} \quad (4.1)$$

Para la creación de la matriz tridimensional a cada banda se asigna a un canal de color: B4 (roja), B3 (verde) y B2 (Azul). El resultado de la dimensión de la matriz es la misma que la de la imagen original, pero ahora con tres canales de color.



Figura 21: PNSAV: Color verdadero (RGB). Fuente (Sentinel-2).

La misión Sentinel-3 (Figura 22), tiene como principal objetivo la medición de la topografía

de la superficie, el color de los océanos y la tierra, con alta fiabilidad y precisión, con finalidad de tener un monitoreo ambiental y climático más completo (European Space Agency, 2023).



Figura 22: Misión Sentinel-3 . Fuente(“Copernicus Data Space Ecosystem”, 2023).

En el caso de Sentinel-3, se utilizó el instrumento SLSTR L1B (Radiómetro Térmico de Océano y Tierra por Explorador de Mares), este instrumento ayuda a la medición de radiación térmica en las superficies de la tierra y océanos. Entre sus bandas espectrales incluye el infrarrojo térmico y visible. También para esta colección de imágenes se indicó un rango de [5-10 %] de nubosidad (“Copernicus Data Space Ecosystem”, 2023).

Las imágenes extraídas de este instrumento son en la categoría temperatura del brillo 7, que ayuda a medir temperaturas de diversas áreas y por último se utilizó OLCI (Instrumento de Color del Océano y Tierra) para medir cuestiones como clorofila en la superficie marina (Figura 23). Para obtener la imagen de temperatura se ocupan las Bandas: 12 que representa la reflectancia del infrarrojo de onda corta, la 11 de onda corta y 10 de onda media, su fórmula de cálculo es (“Copernicus Data Space Ecosystem”, 2023):

$$\text{Índice} = \frac{B12 - B11}{B11 - B10} \quad (4.2)$$

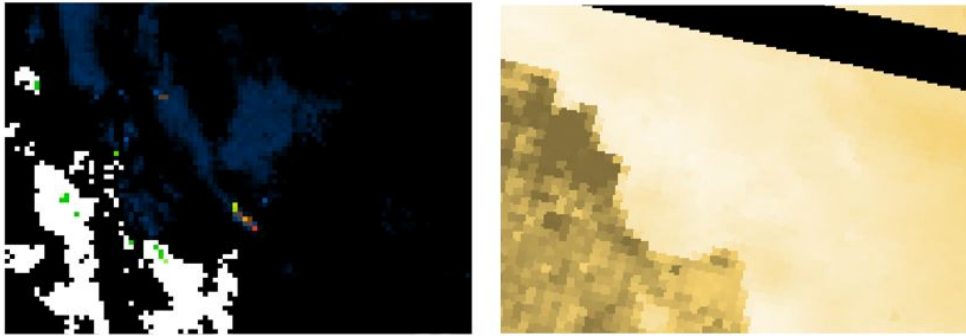


Figura 23: Imágenes de temperatura y clorofila PNSAV. Fuente(Sentinel-3).

4.3. Diseño de la base de datos

Con las imágenes convencionales y percepción remota se diseñó la base de datos, para el caso de las imágenes de percepción remota se obtuvieron de acuerdo con las coordenadas proporcionadas en cada especie, con la finalidad de tener información para comprender la relación de los parámetros y el estado de salud de las especies.

Capítulo 5

Metodología

Para corroborar y comprobar, objetivos e hipótesis planteadas se plantea el siguiente esquema metodológico (Figura 24).

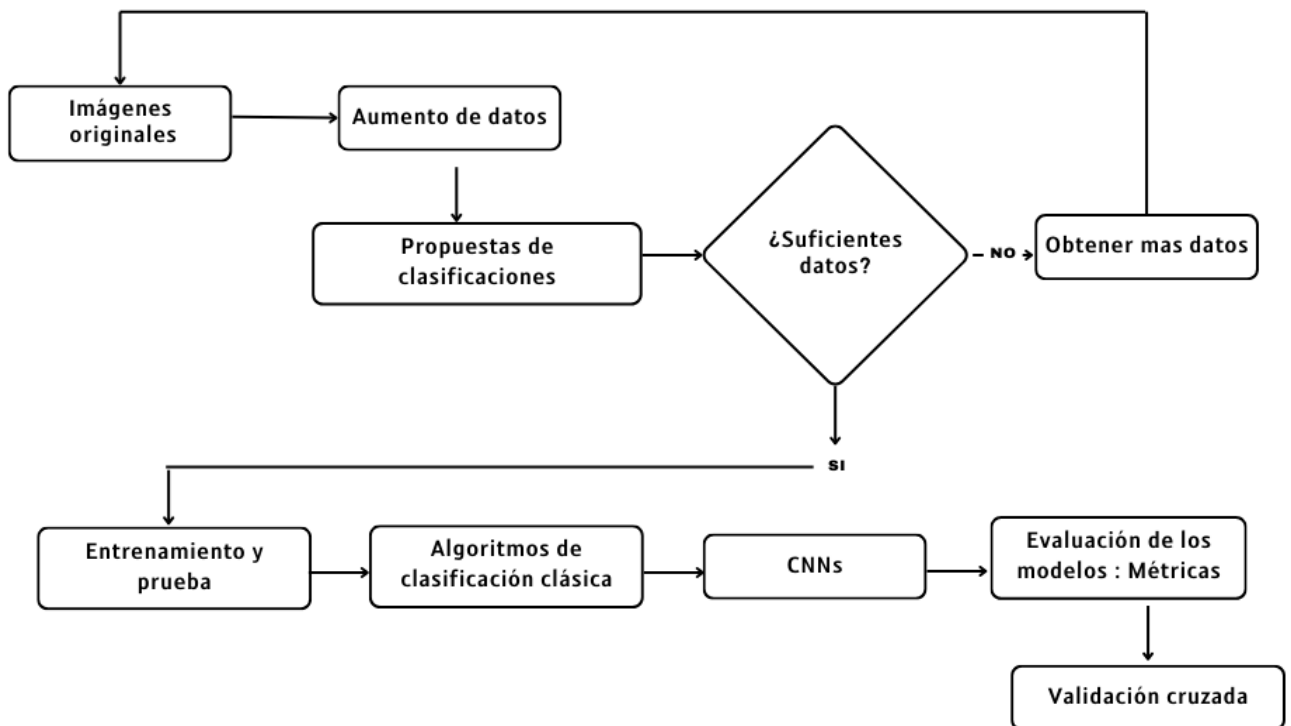


Figura 24: Proceso metodológico

Para esta investigación se empleó el proceso metodológico, que comienza con la obtención de imágenes originales de las especies en estudio. Con la finalidad de procesar redes neuronales convolucionales fue necesario utilizar técnicas de aumento de datos para fortalecer las clases que se ocuparon para este trabajo. Posteriormente se proponen las clasificaciones y se verifican si cuentan con los suficientes datos para continuar el proceso. Si los datos no son suficientes, el proceso tiene que regresar al inicio para tener más imágenes mediante el aumento de datos.

Para continuar el proceso, se procede a la división de los datos en conjunto de entrenamiento y prueba. Los datos de entrenamiento se utilizan para entrenar los modelos de vecinos mas cercanos, random forest y redes neuronales convolucionales. Al entrenar los modelos se aplican las métricas de evaluación, precisión, F1 score y la curva de ROC, esto es para ver el rendimiento de los modelos.

Para finalizar se realiza la validación cruzada para verificar que los modelos son generalizables y no presentan sobreajuste a los datos de entrenamiento. Para la parte espacial se analiza la distribución de las especies, utilizando métodos geoespaciales para comprender el hábitat, distribución y su relación con la salud de los especies.

5.1. Definición de la metodológica

A continuación se describe la propuesta metodológica para llevar a cabo esta investigación (Figura 24).

5.1.1. Descripción de las imágenes originales

En el capítulo de datos se menciona que se tienen imágenes de dos especies, tortugas y delfines, ambos conjuntos de datos fueron recolectados en mar abierto (PNSAV). Las imágenes originales tienen diferentes tamaños (píxeles) por lo que se tuvo que homogenizar dicha característica, para poder procesar los modelos, posteriormente. Fue fundamental el proceso de las imágenes originales, como primero punto importante es la definición tamaño esta será a conveniencia de cada usuario, para este ejercicio se definieron a un tamaño 225×225 (para ambas categorías). En el siguiente paso se pueden visualizar las nuevas dimensiones del conjunto de imágenes, así como el total y la entrada de color, RGB, finalmente se realizan las clases donde delfines pertenece a la clase 0 y tortugas a la clase 1, se unen ambas asignaciones y se obtiene la matriz final para trabajar los métodos de aprendizaje de máquina y profundo.

5.1.2. Aumento de datos

Para finalidad de esta investigación se realizó un aumento del conjunto de imágenes, se considera útil cuando los datos son limitados, para poder entrenar un modelo de red neuronal convolucional o algún proceso basado en aprendizaje profundo. Es importante destacar que para cada conjunto de imágenes, se aumentó en cinco la cantidad de imágenes originales que lo conformaban.

Se implementaron modificaciones como:

- Escalado de píxeles: El rango empleado para escalar los píxeles fue de 10 %.
- Rango de rotación aleatoria: Los grados empleado para la rotación de las imágenes fue entre 10 a 60 grados (de manera aleatoria)
- Desplazamiento horizontal aleatorio: El rango de desplazamiento fue de -20 a 20 píxeles.

- Desplazamiento vertical aleatorio: El rango de desplazamiento fue de -20 a 20 píxeles.
- Inclinación de imagen: En esta transformación se usó un rango de 10 a 60 grados.
- Zoom en las imágenes: El rango para el zoom fue de 1 a 10 veces el tamaño original.
- Reflejo horizontal: La imagen original era reflejo normal y se aplicó el horizontal.
- Relleno de píxeles en deformaciones: Se empleó el relleno de píxeles en borde de manera aleatoria.

Este tipo de proceso ayuda a la *CNNs*, a reconocer objetos en diferentes perspectivas, además de introducir la variabilidad en el conjunto de entrenamiento, así evitar el sobreajuste, para este caso particularmente también fue importante tener el equilibrio en las clases (se resuelve aumentando datos en aquellas clases minoritarias). En la Figura 25 se observar un ejemplo de algunas imágenes procesadas con los parámetros anteriormente mencionados.

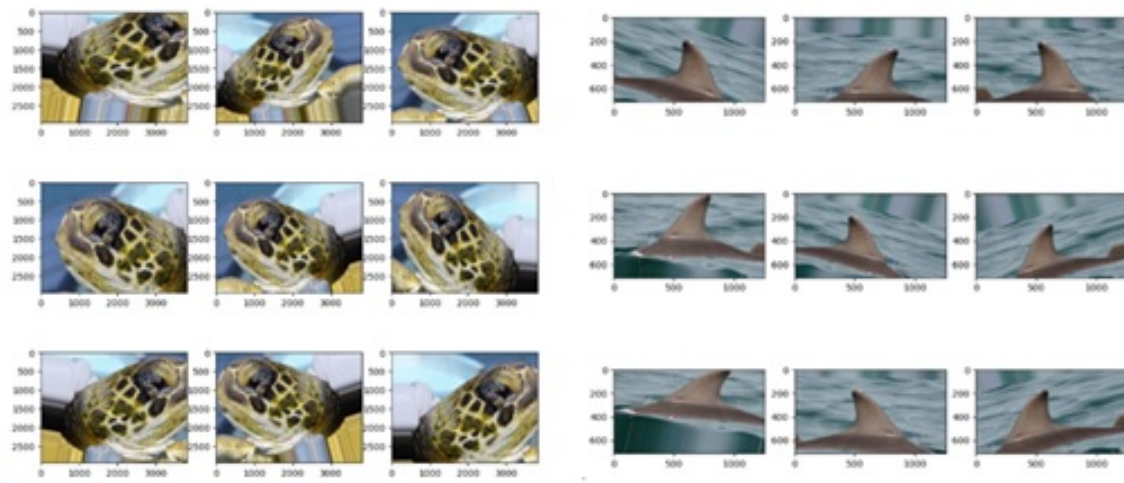


Figura 25: Imágenes con deformaciones geométricas para el aumento de datos. Fuente (Elaboración propia).

5.1.3. Clasificaciones

Propuestas de clasificaciones: en esta investigación se realizaron clasificaciones, basadas en imágenes de vertebrado marinos.

Primera propuesta: clasificación por especies, es decir que el algoritmo reconozca cuando las imágenes pertenecen a Tortugas o Delfines para el caso de la primera especie se tiene

imágenes de su caparazón y para la segunda de su aleta dorsal (Figura 26). Segunda

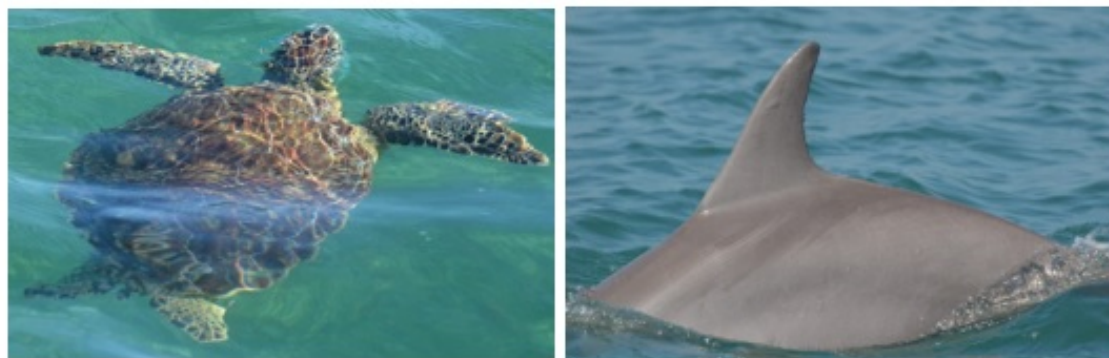


Figura 26: Imágenes de tortuga y delfín. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

propuesta: Basada solo en las imágenes de tortugas, en esta clasificación se entrenó el modelo en términos de la salud de la especie para identificar si tiene neoplasias provocadas por fibropapilomatosis (FP), a continuación, se presenta una imagen de un individuo de tortuga verdes con presencia de tumores (a) de la Figura 27 y el (b) hace referencia a una tortuga sana.

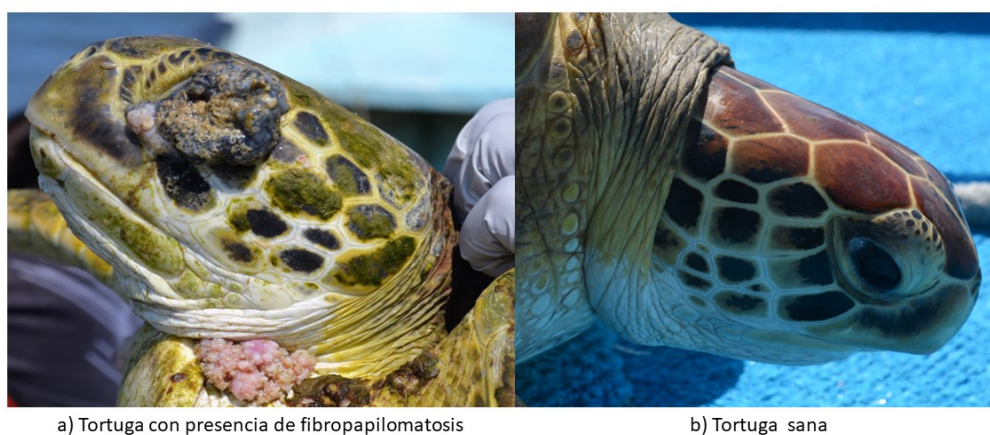


Figura 27: Imágenes de tortugas. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

Tercera propuesta: Categorías basadas en la foto-identificación de las aletas dorsales, con esto se pueden visualizar distintos patrones para identificar a los individuos muestreados, asimismo ver si tienen alguna característica como mordeduras o alguna herida realizada por otra especie o humanos, para esta investigación se tienen tres clasificaciones de aletas, enteras, típicas y otras, ver las Figura 29, Figura 30 y Figura 31.

Para realizar esta selección de categorías, se utilizó el manual para la investigación de campo y laboratorio de actividades de *Sarasota Dolphin Research Program*, 2016. En este manual se tienen diversas categorías basadas en la aleta dorsal de la especie, a continuación, se muestran en la Figura 28:

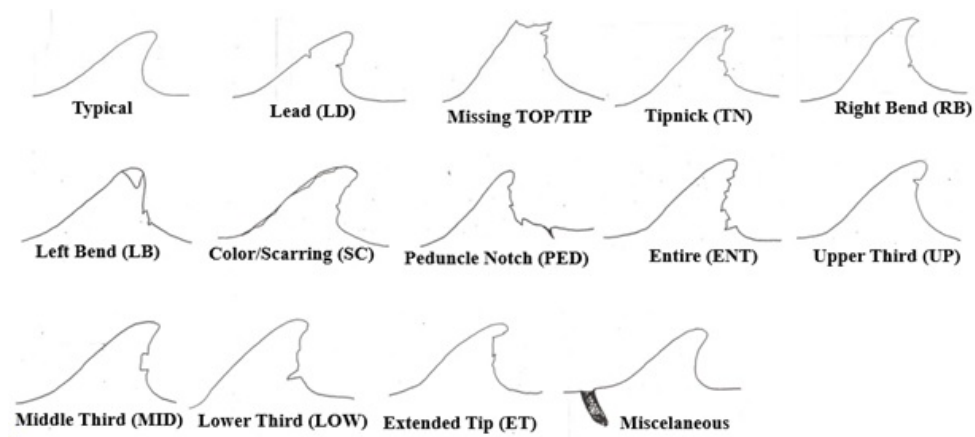


Figura 28: Clasificaciones de aletas dorsales. Fuente (Sarasota Dolphin Research Program, 2006).

Para esta investigación se tienen las siguientes:

Completo (*Entire*)



Figura 29: Aletas dorsales completas. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

Típica (*Typical*)



Figura 30: Aletas dorsales típica. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

Otros: Para esta tercera clase se agrupan aquellas imágenes que se clasificaban en otros tipos de aletas (*Extended Tip*, *Middle Thrid*, *Missing Top* y *Tipnick*), sin embargo, se agruparon en una sola categoría, debido a que presenta una clase minoritario entre 2 a 20 imágenes (Figura 31).



Figura 31: Otras aletas dorsales. Fuente (Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV).

5.1.4. División de datos

Es importante mencionar que, para la aplicación de los métodos de clasificación de imágenes, se dividen los datos en entrenamiento con 80 % de estas. El modelo se entrena para reconocer patrones y asociaciones en los datos, dando como resultado las predicciones para visualizar nuevas instancias. En este caso el modelo hace iteraciones de los datos y con esto evalúa su rendimiento basado en la función de pérdida y las métricas elegidas.

El 20 % restante se ocupa para la prueba y se utilizan estos datos después que el modelo ha sido entrenado, para verificar su rendimiento. Con esto se proporciona la estimación de cómo es el desempeño de datos no vistos.

De manera general la división de datos en estos conjuntos, ayudan a que el modelo tenga una probabilidad baja de sufrir sobreajuste.

5.1.5. Algoritmos de clasificación (Aprendizaje de máquina)

En esta sección se describen los algoritmos supervisados que se emplearon para verificar su precisión en la clasificación de las imágenes, para extraer los datos y trabajar con estos algoritmos fue necesario, extraer la capa de representación en la *CNNs*, este tipo de capa procesa operaciones de convolución, activación y agrupación, específicamente para obtener las características más relevantes de las entradas. Las salidas que se obtienen son de las capas más profundas que contienen información semántica sobre la imagen original.

Los algoritmos supervisados que se emplearon para fines de esta investigación fueron los que se describen a continuación.

Bosques aleatorios (*RF* por sus siglas en inglés)

Tal como lo menciona Breiman (2001) Los *RF* consisten en múltiples árboles de decisión, donde cada árbol se basa en un conjunto de datos aleatorios. A medida que aumenta el número de árboles en el bosque, entonces el error de generalización tiende a estabilizarse. El error es fundamental, ya que de él depende la calidad de los árboles individualmente y su relación entre sí. En este proceso, los árboles se entrenan de manera independiente con una muestra aleatoria, lo que ayuda a reducir el sobreajuste. Finalmente, las predicciones de cada árbol se fusionan para tener un resultado más confiable.

RF es un algoritmo que tiene un rendimiento de aprendizaje, a diferencia de los árboles de decisión los *RF* tiene una estimación más precisa en la tasa de error (Schonlau y Zou, 2020).

Proceso del *RF* :

- Combinación de múltiples arboles de decisiones
- Aplicación del Booststrap
- Características del muestreo
- Promedio

- Robustez frente al sobreajuste

k Vecinos más cercanos (*k-NN* por sus siglas en inglés)

Es un método de aprendizaje supervisado, con resultados aceptables en tareas de clasificación y regresión. En la clasificación los *k-NN*, se desarrollan cuando un objeto se clasifica en función de la mayoría de las clases de sus vecinos más cercanos, donde “*k*” es el número de vecinos considerados. Para el caso de regresión los *k-NN*, realizan la predicción del valor sobre un tributo objetivo calculando el promedio de los valores de los “*k*” vecinos más cercanos. Este modelo de clasificación se basa en la suposición de que los objetos similares tienden a estar en la misma clase o al menos tener los valores cercanos. (P.-N. Tan *et al.*, 2014).

5.1.6. Red Neuronal Convolutiva (*CNNs*)

Para esta investigación se propone una arquitectura de *CNNs*, con las siguientes características:

Entrada que es equivalente a las clases que se requieran estudiar sea el caso (delfín, tortuga). Se definen cinco capas con filtros de, 32, 64, 128, 256 y 512, con la finalidad de que cada filtro aprenda diferentes patrones y características en los datos de entrada. El Kernel utilizado fue 3×3 , lo que significa que se considera una vecindad de 9 vecinos (tres píxeles de ancho y tres de alto).

MaxPooling: para este caso se empleó un tamaño de ventana (kernel) 2×2 , lo que indica que, después del proceso de convolución se realiza una operación de agrupación con ese tamaño de ventana, lo cual ayuda a reducir la cantidad de parámetros en la red, proporciona invariancia espacial y fija las características más destacadas.

Para la capa de activación se empleó la función ReLU, con esta se introduce la no linealidad en el modelo, activa las neuronas en caso de que la entrada es positiva deja pasar el valor, siendo el caso contrario (negativa) establecerá la salida en cero.

La capa densa empleada fue 512 neuronas, aquí cada neurona tiene conexiones con todas las salidas de la capa anterior y con todas las entradas de la capa siguiente. Las 512 neuronas desarrollarán una activación lineal de las entradas y activará a su vez la función

de activación en este caso la Softmax, que se emplea para convertir las salidas de las neuronas en probabilidades.

Hiperparámetros considerados para esta arquitectura:

- Padding (same): Se empleó para generar relleno y las salidas tengan el mismo tamaño que la imagen original. Con este parámetro se logra que la convolución no minimice la dimensión espacial de la entrada.
- Learning rate: Este es un hiperparámetro ya que ayuda a determinar el tamaño de pasos que el algoritmo requiere para el ajuste de los pesos en el modelo, durante el procedimiento de la optimización, entre menor sea el valor se logran pasos más pequeños y hace que la optimización sea más precisa, pero puede que todo el proceso sea más lento.
- Optimizador: ADAM
- Función de pérdida: `sparse_categorical_crossentropy`

A continuación, observamos la propuesta de arquitectura de *CNNs* (Figura 29).

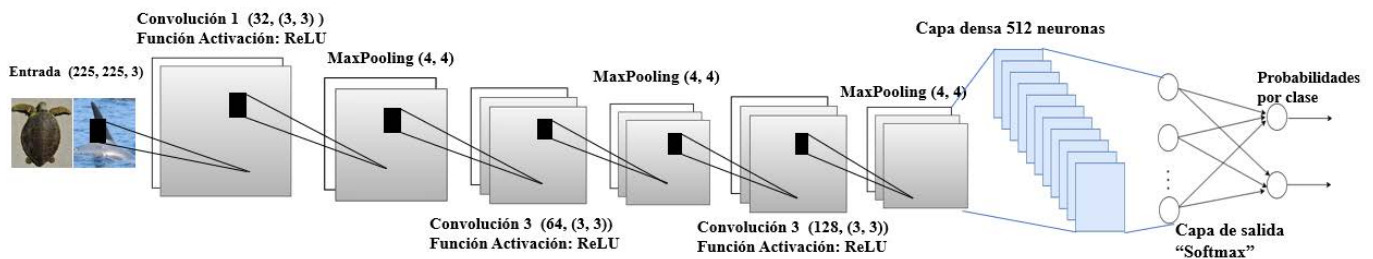


Figura 32: Esquema de arquitectura de *CNNs* propuesta. Fuente: Elaboración propia.

Cuando el modelo ha sido probado con los respectivos hiperparámetros, se evalúan las métricas como son la precisión que mide el rendimiento de la clasificación dada por el modelo y el F1 score, esta métrica ayuda a evaluar el rendimiento del modelo basado en su capacidad de equilibrio de la precisión y Recall, se recomienda cuando existen clases desbalanceadas.

Métricas empleadas:

- Matriz de confusión
- F1 score

- Accuracy
- Curva de Roc

Para finalizar este capítulo, es importante mencionar que se probaron arquitecturas pre entrenadas como, VGG19, GoogleNet y ResNet50.

5.1.7. Validación cruzada

Para evitar problemas de sobreajuste en los modelos de aprendizaje profundo y automático se emplea la validación cruzada. Para todos los modelos se empleó un número de particiones $k = 5$ para garantizar la generalización de los resultados. La validación cruzada se realizó dividiendo el conjunto de datos en k subconjuntos. Luego, se realizaron $k = 5$ iteraciones de entrenamiento y evaluación, utilizando un subconjunto diferente como conjunto de prueba en cada iteración y el resto como conjunto de entrenamiento. Los resultados finales se promediaron sobre las k iteraciones para obtener una estimación más precisa del rendimiento de cada modelo.

5.1.8. Distribución espacial de las especies en el PNSAV

En la distribución espacial de las especies, se utilizaron imágenes de percepción remota para relacionar las coordenadas geoespaciales de las especies y posteriormente observar su posición geográfica. Se emplearon técnicas de análisis espacial para determinar cómo se distribuyen las especies enfermas y sanas en relación con variables ambientales clave, como la temperatura y la concentración de clorofila en el entorno. Estos datos permitieron estudiar las posibles relaciones entre la salud de las especies y su entorno, lo que proporciona información valiosa para comprender la ecología y el comportamiento de las especies en el ecosistema estudiado.

Capítulo 6

Resultados y Discusión

En este capítulo se muestran los experimentos desarrollados para probar la viabilidad del método empleado (*CNNs*), así como los resultados más relevantes obtenidos de la aplicación . Para dar un estudio más completo, también se consideraron redes neuronales convolucionales preprocesadas, así como los métodos de clasificación tradicional (*random forest* y *k* vecinos más cercanos).

6.1. Etapa experimental

- Proceso de aplicación de la arquitectura: cinco capas, Maxpooling (2 x 2 en cada capa), learning rate (0.001), padding(same), kernel 3 x 3, optimizador ADAM, la función de pérdida (`sparse_categorical_crossentropy`).
- Desarrollo de la misma arquitectura, empleando el aumento de datos: Por cada imagen se aumentaban cinco, con las transformaciones establecidas: Rotación, zoom, relleno de pixeles, entre otros.
- Aplicación de las redes preentrenadas: VGG19, GoogleNet, ResNet50 .
- Aplicacion de los algoritmos clásicos de clasificacion.
- Métrica de evaluación fueron: Precisión (*Accuracy*), F1 score, matriz de confusión y curva de ROC.

6.2. Resultados

En este capítulo, se presentan a detalle los resultados obtenidos, basados en la aplicación del aprendizaje profundo para la clasificación de imágenes.

6.2.1. Resultados dos clases: Delfín o tortuga

Tabla 5: Desempeño de los modelos de aprendizaje máquina y profundo

Arquitecturas	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	2	5	3min	0.01	0.99	0.98
VGG19_AugD	2	3	23min	0.00	1	1
GoogleNet_AugD	2	3	19min	0.10	0.98	0.96
ResNet_AugD	2	3	24min	1.05	1	1
CNN_1AugD	2	5	10min	0.08	0.99	1
CNN_1	3	5	7min	0.04	0.98	0.93
VGG19	3	3	28min	0.00	0.99	0.93
GoogleNet	3	3	20min	0.69	0.87	0.74
ResNet	3	3	30min	0.01	0.99	0.99
CNN_1AugD	3	5	10min	0.07	0.99	1
Algoritmos de Clasificación						
K_NNAugD	2				0.97	0.97
RF_AugD	2				0.98	0.98
K_NNAugD	3				0.89	0.88
RF_AugD	3				0.90	0.89

Tabla 5, proporciona la comparación sobre arquitecturas preentrenadas y una propuesta de red neuronal convolucional (CNN_1), considerando elementos importantes para su procesamiento, como la capa de entrada, el maxpooling, la capa de convolución, la función de activación, la capa densa y la de salida, en cuanto a los hiperparámetros, se utilizaron el Padding (same), learning rate (0.0001), kernel 3 x 3 y para optimizar se empleó el método de ADAM, para todos los modelos.

Se observa que las primeras cinco filas pertenecen al experimento para dos clases, siendo la CNN_1 la propuesta de arquitectura con cinco capas y una exactitud de 0.99 es decir el rendimiento del modelo basado en las predicciones correctas. También para esta arquitectura (CNN_1AugD) se entrenó el mismo modelo, pero con datos aumentados dando una precisión de 0.99 del rendimiento y un F1 score de 1, es decir que se tiene un equilibrio

entre la identificación correcta de instancias positivas y la minimización de falsos positivos y falsos negativos. En el caso de las redes preentrenadas (con datos aumentados) VGG19, GoogleNet y ResNet50, tuvieron rendimientos altos en precisión, sin embargo, el tiempo de proceso fue mas tardado. Para esta aplicación el área bajo la curva es igual a uno lo que indica buen rendimiento del modelo, mientras que la matriz de confusión solo detecta un caso real como delfín y lo predice como tortuga (Figura 34).

Se realizó el entrenamiento de las mismas arquitecturas la diferencia radica en el número de clases, ahora se procesaron tres (tortugas, delfines y otras especies marinas), los resultados muestran mas eficiencia en la CNN_1, VGG19 y ResNet50, mientras que para CNN_1AugD la precisión para las tres clases no fue baja, pero existe un rango considerable entre los resultados de las otras arquitecturas. La GoogleNet fue quien presentó los valores más bajo en precisión para la clasificación de las tres clases.

Se entrenaron algoritmos de clasificación clásica para ambas clases y utilizando el aumento de datos, en términos generales, se obtuvieron resultados óptimos, siendo k vecinos mas cercanos para tres clases el modelo que presentó una precisión mas baja que el resto.

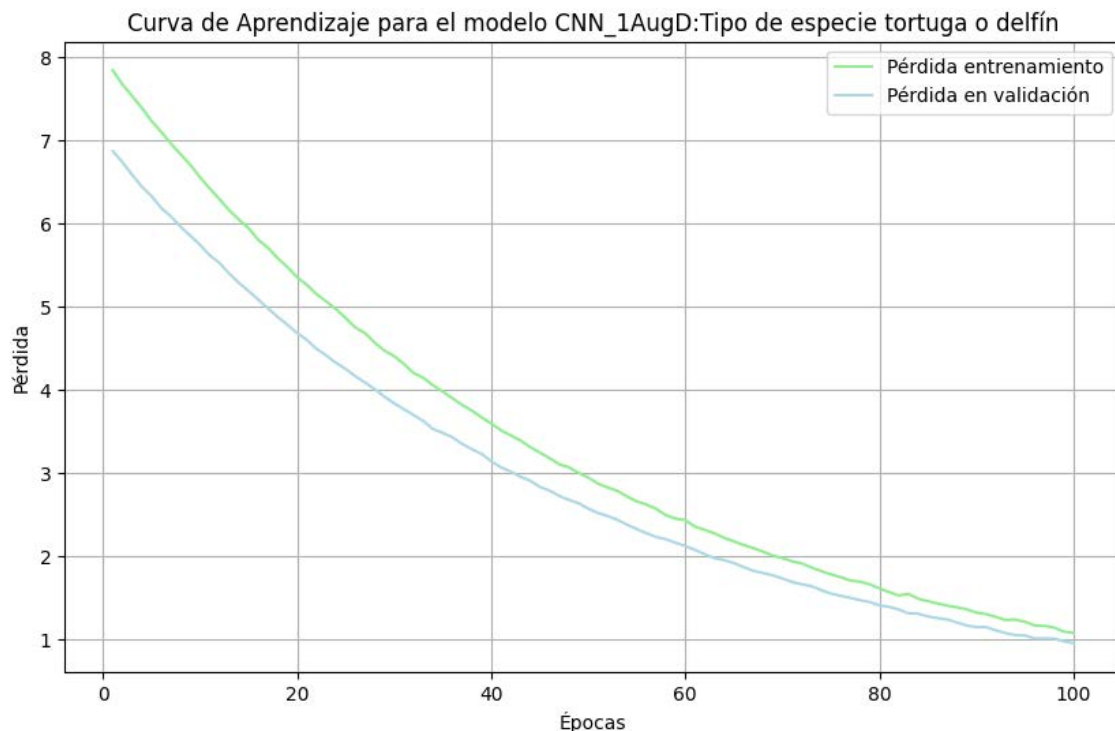


Figura 33: Curva de aprendizaje 1.

El error para el modelo en el conjunto de entrenamiento es de 0.08 y para la validación

es de 0.07. Esto índice que el modelo tiene un buen rendimiento en ambos conjuntos. El cambio de los errores a lo largo de las épocas disminuye y se estabilizan, es decir el modelo logró aprender correctamente. En este caso la curva de validación esta por debajo de la de entrenamiento, indica que el modelo esta generalizando bien las clasificaciones.

Matriz de confusión en el conjunto de entrenamiento aumentado

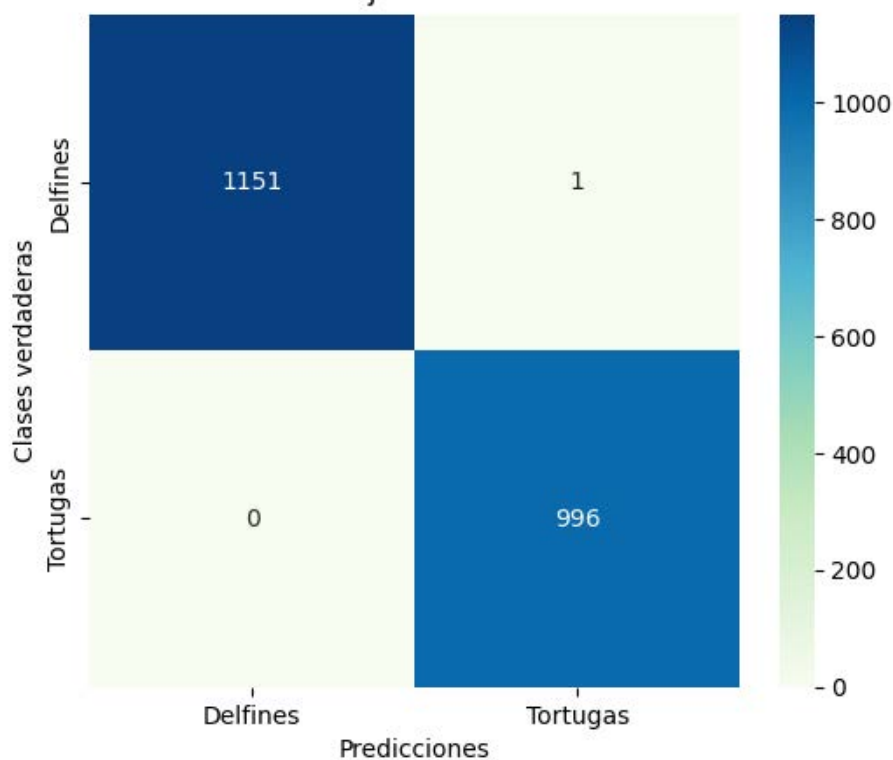


Figura 34: Clasificaciones basada en las predicciones del modelo

La Figura 34 muestra la matriz de confusión, en la diagonal principal se observan las predicciones correctas del modelo. Para el caso de esta investigación se tienen 1151 imágenes que fueron clasificadas correctamente como delfines y 996 como tortugas. Mientras que los valores fuera de la diagonal representan los errores de clasificación. Para este caso se tiene una sola imagen que en su clase real es delfín y la predicción la reconoce como tortuga.

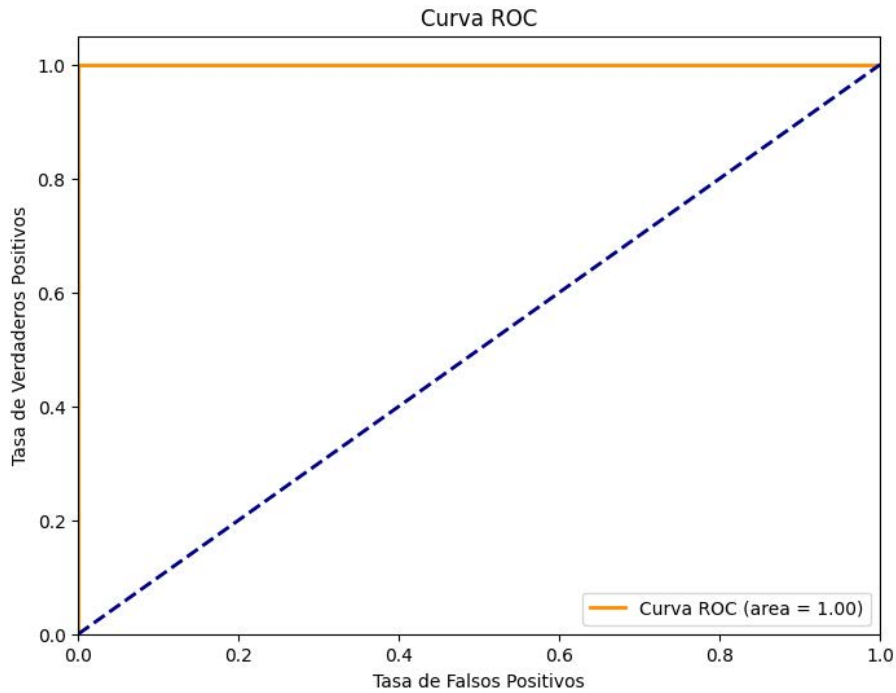


Figura 35: Área bajo la curva

La curva de ROC (Figura 35), muestra un modelo con capacidad óptima para identificar los verdaderos positivos de las imágenes de los delfines, mientras que muestra una tasa baja de falsos positivos, en imágenes incorrectamente clasificadas de delfines como tortugas. Para este problema de clasificación de dos especies, se tiene un AUC igual a uno, lo que indica un rendimiento bueno del modelo evaluado.

6.2.2. Validación cruzada: Delfín o tortuga.

Los resultados, de la Tabla 6, muestra valores de pérdida bajos, se considera adecuados, ya que indican que el modelo obtuvo predicciones óptimas para el conjunto de entrenamiento. Para la columna de exactitud indica la precisión del modelo en el conjunto de entrenamiento, es decir la proporción de predicciones correctas. Para las dos columnas restantes se presentan los valores de la validación es decir un conjunto de datos que el modelo no ha visto durante el entrenamiento. En términos generales la validación cruzada realiza una evaluación mas exacta de la capacidad que tiene el modelo para generalizar y contribuye a la medición de la robustez en datos no vistos.

Tabla 6: Experimento con validación cruzada K-Fold=5

Arquitecturas <i>CNNs</i>	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	2	5	6min	0.01	0.96	0.94
VGG19_AugD	2	3	40min	0.00	1	1
GoogleNet_AugD	2	3	29min	0.09	0.97	0.96
ResNet_AugD	2	3	24min	0.13	1	1
CNN_1AugD	2	5	8min	0.10	0.98	0.96
CNN_1	3	0 5	6min	0.05	0.97	0.94
VGG19_AugD	3	3	38min	0.01	0.98	0.94
GoogleNet_AugD	3	3	28min	0.55	0.83	0.69
ResNet_AugD	3	3	30min	0.04	0.97	0.95
CNN_1AugD	3	5	5min	0.09	0.89	0.88
Algoritmos de Clasificación						
K_NNAugD	2				0.95	0.94
RF_AugD	2				0.97	0.97
K_NNAugD	3				0.85	0.86
RF_AugD	3				0.91	0.89

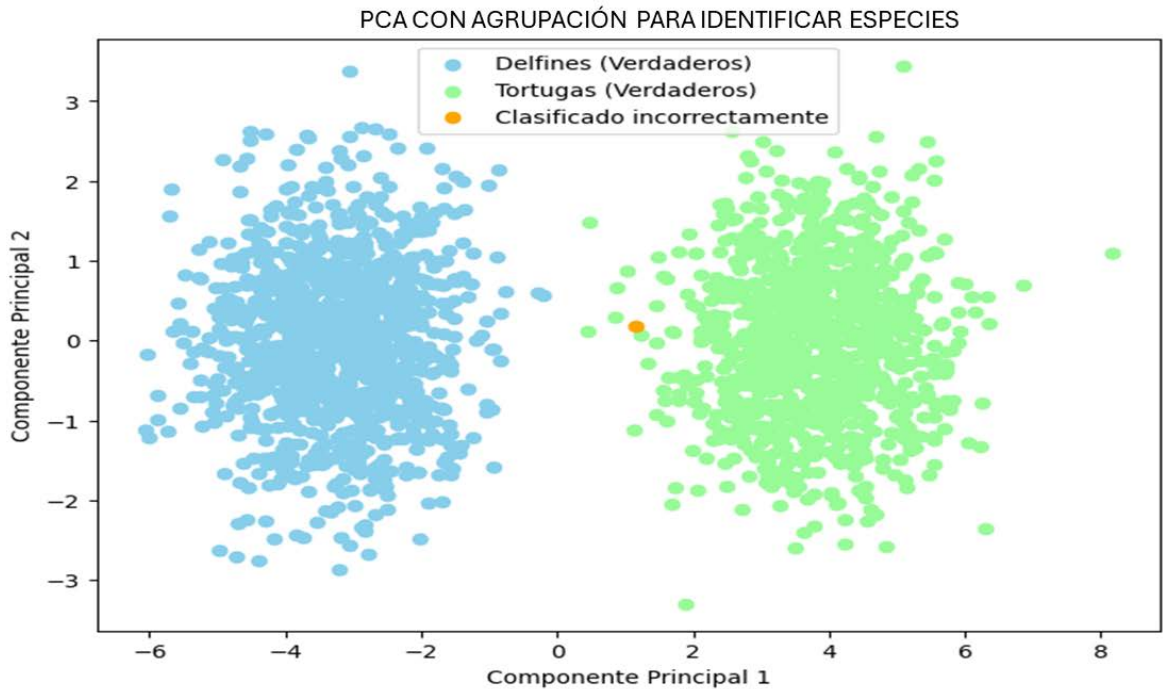


Figura 36: Componentes principales para las clases tortugas o delfines

La distribución de los puntos en el gráfico de PCA con agrupaciones (Figura 36), muestra las imágenes de acuerdo con las características de cada especie. La separación de los puntos azules (delfines) de los verdes (tortugas), claramente indican que se tienen dos clases con características distintivas que el PCA muestra en sus dos principales componentes.

El punto naranja, representa una imagen clasificada como delfín en su clase verdadera, pero predicha como tortuga por el modelo. Esto sugiera que, en términos de las imágenes consideradas para la reducción de las dimensiones, esta imagen comparte similitud con las tortugas que con los delfines.

La discrepancia que se presenta en la clasificación real y la predicción reflejada por el modelo se debe a características muy específicas de la imagen que se consideran mas similar al grupo de tortugas, lo que lleva a tener una clasificación en términos de falsos negativos. La exploración de características específicas influyen en esta predicción incorrecta ya que puede proporcionar información valiosa sobre las imágenes de estas especies y así mejorar los modelos en trabajos futuros.

6.2.3. Resultados dos clases: Tortugas enfermas o sanas

Tabla 7: Desempeño de los modelos de aprendizaje máquina y profundo: Clases enferma o sana

Arquitecturas CNNs	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	2	5	2min	0.42	0.83	0.55
CNN_1AugD	2	5	5min	0.08	0.87	0.86
VGG19_AugD	2	3	25min	0.23	0.99	1
GoogleNet_AugD	2	3	19min	0.12	0.96	0.96
ResNet_AugD	2	3	28min	1.01	0.99	1
Algoritmos de Clasificación						
K_NN_AugD	2	-	-	-	0.79	0.79
RF_AugD	2	-	-	-	0.81	0.80

La Tabla 7, muestra el rendimiento de diferentes modelos de clasificación en imágenes de dos especies marinas (tortugas y delfines), basado en su estado de salud. Es importante mencionar que para esta investigación se ocupó el aumento de datos, a excepción de la arquitectura CNN_1 que fue procesada con los datos originales.

De manera general se puede observar que el modelo CNN_1, procesado con los datos originales tiene una exactitud de 0.83, si embargo el F1 score se reporta bajo, debido al desbalance de datos en los grupos, asimismo por la falta de disponibilidad de datos. Los modelos con mejores resultados para la clasificación de estas imágenes son, CNN_1AugD, VGG19_AugD, GoogleNet_AugD y ResNet_AugD, que son arquitecturas de redes neuronales convolucionales con aumento de datos. De estas arquitecturas todas presentan una exactitud en un rango de 0.80- 0.99, es decir que estos modelos tienen un buen rendimiento, el F1 de score para algunos casos alcanza el 1, esto indica que se tienen equilibrio en las clases evaluadas, por lo tanto, la clasificación del modelo es buena. La diferencia de los modelos es el tiempo de ejecución siendo las redes pre-entrenadas, las que mayor tiempo tardaron en procesar la clasificación.

Para los algoritmos de clasificación clásica, K_NN_AugD y RF_AugD, tienen valores aceptables en las métricas evaluadas, aunque de acuerdo con la literatura consultadas, estos algoritmos tienen una capacidad menor para reconocer a profundidad ciertas características de las imágenes.

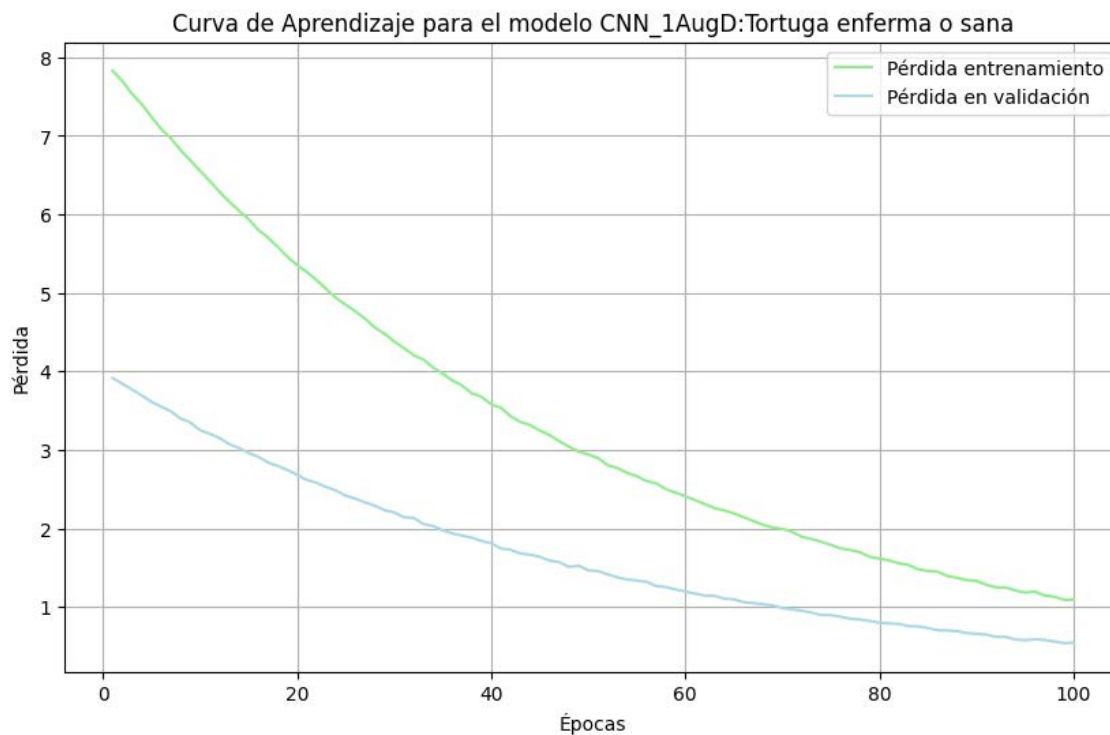


Figura 37: Curva de aprendizaje 2.

Para la Figura 37, la curva de aprendizaje de pérdida de entrenamiento y validación se estabilizan en cien épocas, esto indica que el modelo ha convergido y aprendió de manera efectiva tanto para los datos de entrenamiento como en la validación. La diferencia entre los valores es pequeña ($0.08-0.04=0.04$), lo que indica que el modelo generaliza bien a los nuevos datos. Para este caso la pérdida disminuye rápidamente al principio y luego se estabiliza, es decir el modelo aprende rápidamente. Se considera una estabilidad ya que las curvas son suaves es decir no hay fluctuaciones significativas.

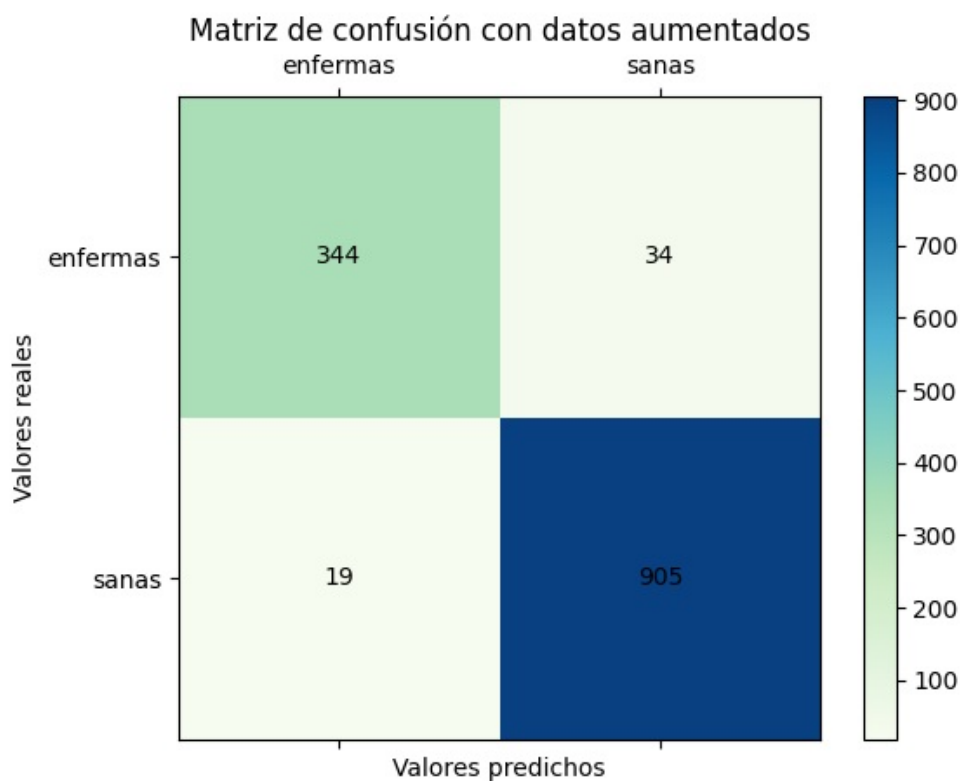


Figura 38: Matriz de confusión del conjunto de entrenamiento con datos aumentados.

La Figura 38, muestra en la diagonal principal las predicciones correctas, se tienen 344 imágenes de tortugas enfermas y 905 sanas. Mientras que los valores fuera de la diagonal representan los errores de clasificación. Para este caso se tiene 34 imágenes que son sanas y se clasifican como enfermas y 19 enfermas que se clasifican como sanas.

La Figura 39, muestra algunas imágenes basadas en los resultados del modelo CNN1 con datos aumentados. Las imágenes etiquetadas con el a), su clase verdadera es tortugas sanas y el algoritmo los reconoce como tortugas enfermas. Las imágenes del b), su etiqueta verdadera es enfermas y el algoritmo las reconoce como sanas, por ejemplo la imagen 1 de la fila dos, claramente la tortuga tiene un tumor en el ojo, aún así se reconoce como sana. Esto puede deberse a factores la geometría o forma de la cabeza de la especie.



Figura 39: Predicciones.

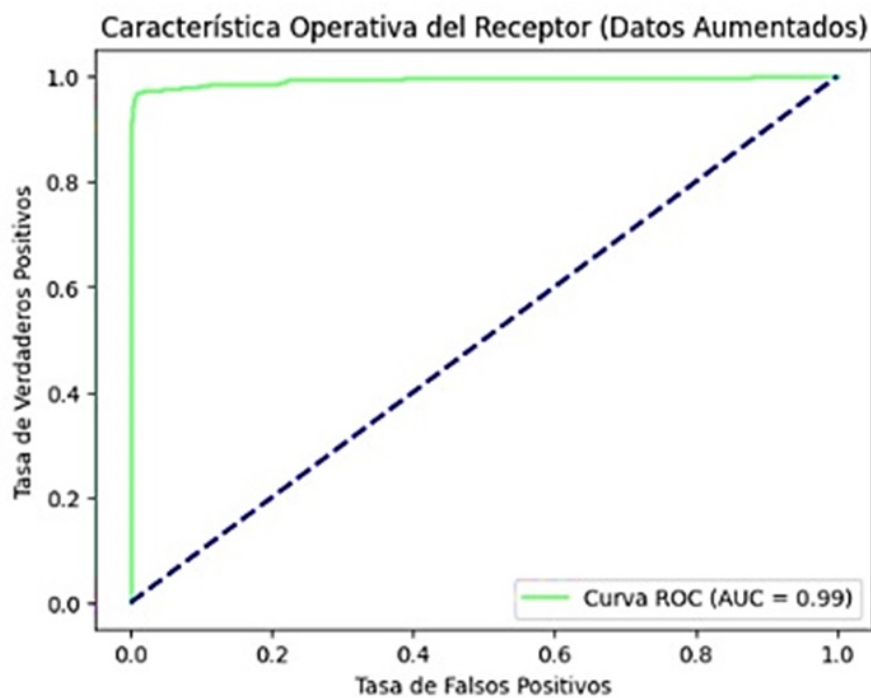


Figura 40: Curva de ROC con datos aumentados.

La curva de ROC (Figura 40), muestra un modelo con capacidad óptima para identificar los verdaderos positivos de las imágenes de los delfines, mientras que muestra una tasa baja de falsos positivos, en imágenes incorrectamente clasificadas enfermas como sanas. Para este problema de clasificación de estado de salud, se tiene un AUC igual a 0.99, lo

que indica un rendimiento bueno del modelo evaluado.

6.2.4. Validación cruzada: Tortugas enfermas o sanas.

Tabla 8: Desempeño de los modelos clases enferma o sana: Validacion cruzada K fold=5

Arquitecturas CNNs	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	2	5	4min	0.41	0.79	0.50
CNN_1AugD	2	5	8min	0.09	0.88	0.87
VGG19AugD	2	3	27min	0.25	0.99	0.99
GoogleNetAugD	2	3	20min	0.13	0.94	0.96
ResNetAugD	2	3	25min	1.00	98	0.9
Algoritmos de Clasificación						
K_NNAugD	2	-	-	-	0.78	0.78
RFAugD	2	-	-	-	0.80	0.81

La Tabla 8, muestra la evaluación de las arquitecturas mencionadas anteriormente, pero ahora utilizando la validación cruzada con $k = 5$, para la robustez de los resultados. Cabe mencionar que estos resultados son los valores promedios para cada modelo. En el modelo CNN_1, se tiene una pérdida de 0,41 y una exactitud de 0.79, mientras que F1 score 0.50, lo que indica que el desbalance de las clases afecta el rendimiento de dicho modelo. Para mejorar los resultados se aplicó el aumento de datos, CNN_1AugD, tiene una exactitud de 0.88 y F1 score de 0.87, mientras que los modelos pre entrenados igualmente tienen resultado óptimo, sin embargo, los tiempos de ejecución son altos a diferencia de la arquitectura propuesta CNN_1AugD. Para los algoritmos clásicos se tiene valores más bajos en términos de la exactitud y F1 score.

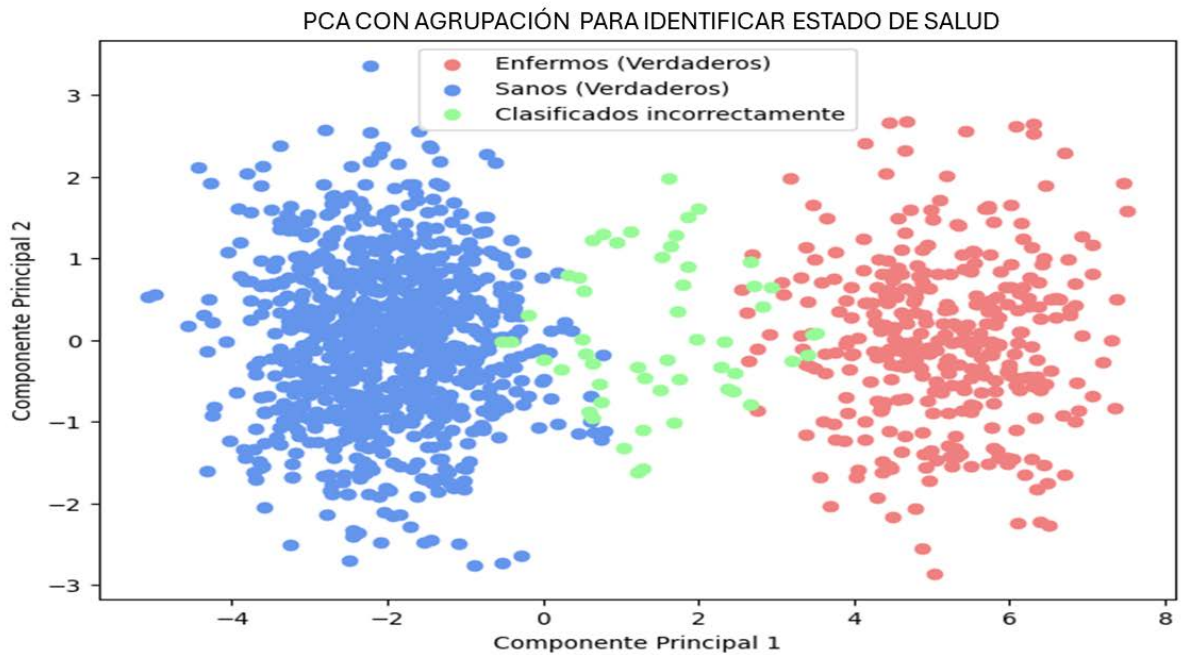


Figura 41: Componentes principales para las clases tortugas sanas o enfermas

En la (Figura 41), tenemos el gráfico PCA con agrupaciones, los puntos azules representan tortugas consideradas como sanas y los puntos rojos tortugas enfermas. La distribución que se visualiza en el espacio bidimensional indica como las características en las imágenes logran separar las clases en estudio.

En los puntos verdes se tienen las clasificaciones incorrectas, es decir son imágenes de tortugas que fueron predichas erróneamente. De los cincuenta y tres puntos verdes claros, treinta y cuatro corresponden a tortugas enfermas clasificadas como sanas y diecinueve corresponden a tortugas sanas que se predicen como enfermas.

Estos errores indican que se pueden tener mejoras en el modelo para detectar patrones más específicos en la imagen. Por ejemplo, tener treinta y cuatro tortugas enfermas clasificadas como sanas sugiere que el modelo no este considerando algunas características que indican la enfermedad.

Las características específicas que conducen a este tipo de clasificaciones erróneas pueden ayudar a mejorar la exactitud del modelo en el futuro, lo que es importante para aplicaciones como la detección temprana de enfermedades en la vida marina.

6.2.5. Resultados tres clases: Aletas enteras, otras y típicas

Tabla 9: Desempeño de los modelos de aprendizaje máquina y profundo: Clases de aletas

Arquitecturas CNNs	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	3	5	9min	0.97	0.52	0.29
CNN_1AugD	3	5	12min	0.11	0.96	0.97
VGG19AugD	3	3	27min	1.01	0.99	1
GoogleNetAugD	3	3	24min	0.15	0.96	0.96
ResNetAugD	3	3	30min	1.14	1	1
Algoritmos de Clasificación						
K_NNAugD	3	-	-	-	0.85	0.86
RFAugD	3	-	-	-	0.88	0.88

La Tabla 9, proporciona los resultados de los modelos evaluados para la clasificación de aletas de los delfines nariz de botella. Primero se entrena una propuesta de modelo CNN_1, con los datos originales para tres clases, su rendimiento en tiempo fue de 9 minutos, pérdida de 0.97, exactitud de 0.52 y un F1 score de 0.29, lo cual indica que este modelo no es óptimo derivado, al desbalance de clases y falta de datos en las clases.

Para mejorar los resultados se utilizó el aumento de datos para el modelo CNN_1AugD, obteniendo resultados en la pérdida de 0.11, exactitud de 0.96 y F1 score de 0.97. Para los siguientes modelos se utilizaron redes pre entrenadas, como VGG19_AugD, GoogleNet_AugD y ResNet_AugD, también se obtuvieron resultados sobresalientes para las métricas evaluadas. Las diferencias radican en los tiempos de ejecución, las redes preentrenadas tuvieron mayor tiempo de ejecución, mientras que la arquitectura propuesta (CNN1) procesar la clasificación de las tres clases fue en menor tiempo.

Para finalizar igualmente se evaluaron los algoritmos de clasificación clásica, ambos modelos presentan resultados aceptables para la clasificación de imágenes con datos aumentados.

6.2.6. Validación cruzada: Aletas enteras, otras y típicas.

Tabla 10: Desempeño de los modelos clases de aletas: Validación cruzada k-folds (k=5)

Arquitecturas CNNs	Clases	Capas	Tiempo	Pérdida	Exactitud	F1 score
CNN_1	3	5	11min	0.98	0.51	0.31
CNN_1AugD	3	5	18min	0.01	0.98	0.98
VGG19AugD	3	3	29min	1.1	0.99	1
GoogleNetAugD	3	3	26min	0.14	0.95	0.96
ResNetAugD	3	3	31min	1.16	1	1
Algoritmos de Clasificación						
K_NNAugD	3	-	-	-	0.85	0.87
RFAugD	3	-	-	-	0.89	0.88

La Tabla 10, muestra los resultados de la clasificación por aletas, pero utilizando validación cruzada k-fold igual a 5. Nuevamente se evalúa el modelo CNN_1 con los datos originales, con un promedio de exactitud de 0.51 y F1 score, de 0.31, es decir la falta de datos afecta el modelo y su rendimiento. En comparación con el mismo modelo CNN_1_AugD, con datos aumentados, tuvo en promedio 0.98 en exactitud y 0.98 en F1 score. También se evaluaron las arquitecturas pre entrenadas, con datos aumentados, teniendo como resultados buena exactitud y F1 score. Estas tienen el tiempo de ejecución mas amplio a diferencia de la arquitectura propuesta. Para los algoritmos clásico en promedio presentan un rendimiento aceptable.

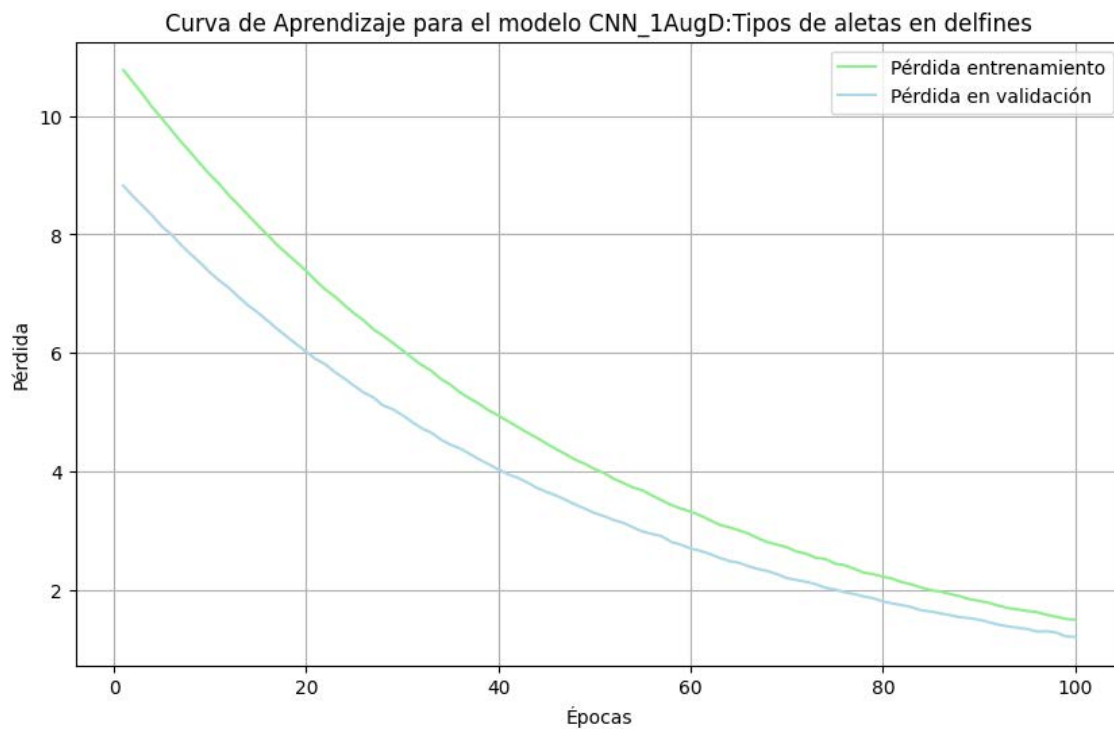


Figura 42: Curva de aprendizaje 3.

Para el caso del modelo sobre clasificación de aletas, la diferencia entre los valores para ambos conjuntos es pequeña lo que sugiere que el modelo generaliza correctamente. En este caso después de las cien épocas, es posible que el modelo aún no tenga convergencia, pero la tendencia de la disminución de ambos errores es buena. Un error de 0.11 en entrenamiento y 0.09 en validación se consideran correctos (Figura 42).

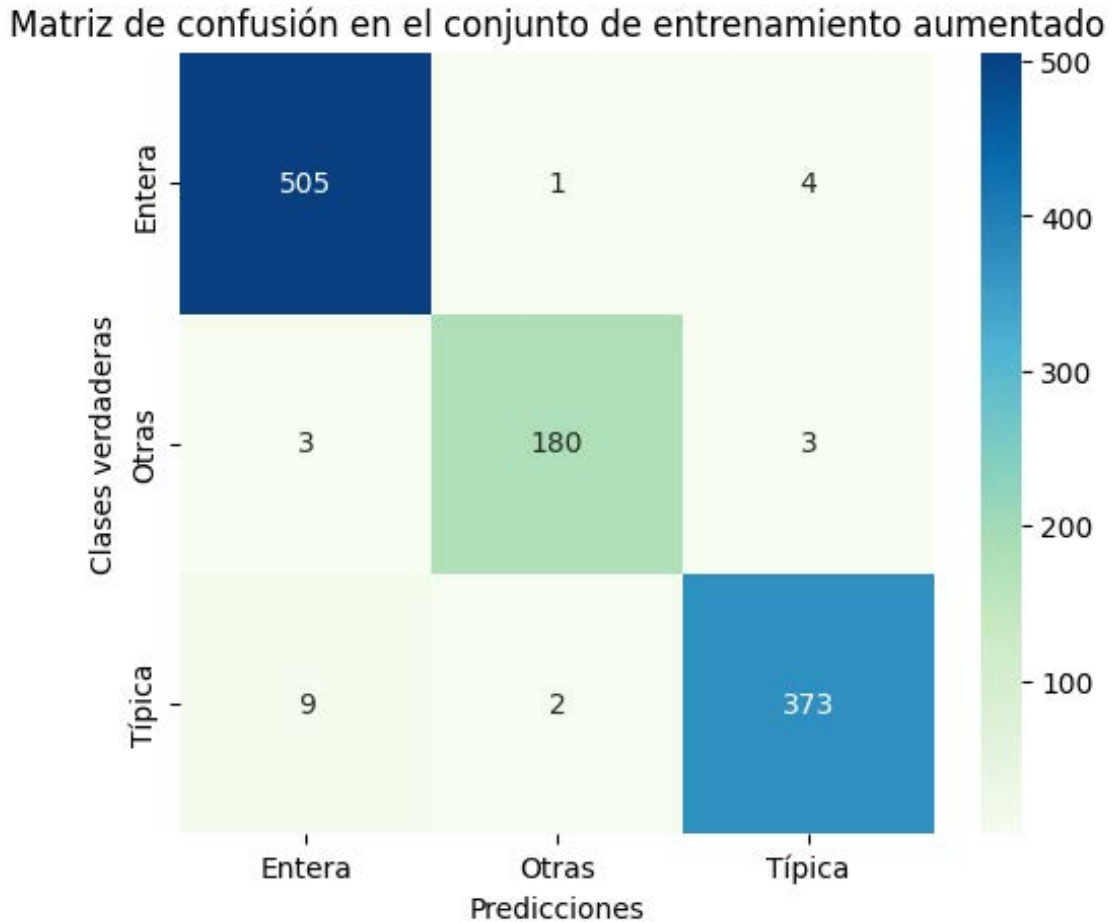


Figura 43: Gráficas de los conjuntos de entrenamiento y prueba: tres clases.

La Figura 43, muestra la matriz de confusión, en el primer cuadrante (color azul intenso) representan quinientos cinco delfines con aletas en la clasificación entera. Para el cuadrante intermedio de la diagonal ciento ochenta delfines con otro tipo de aletas. El último cuadrante de la diagonal, muestra treientos setenta y tres animales con aletas en la clasificación típica. El modelo propuesto, CNN1 con datos aumentados, muestra en las predicciones como ejemplo a tres delfines que se identifican como otras tipos de aletas pero la predicción lo coloca como entera. Otro caso es la clase verdadera típica con nueve individuos que la predicción detecta como entera y finalmente cuatro individuos que su clase verdadera es entera y se predicen como típica

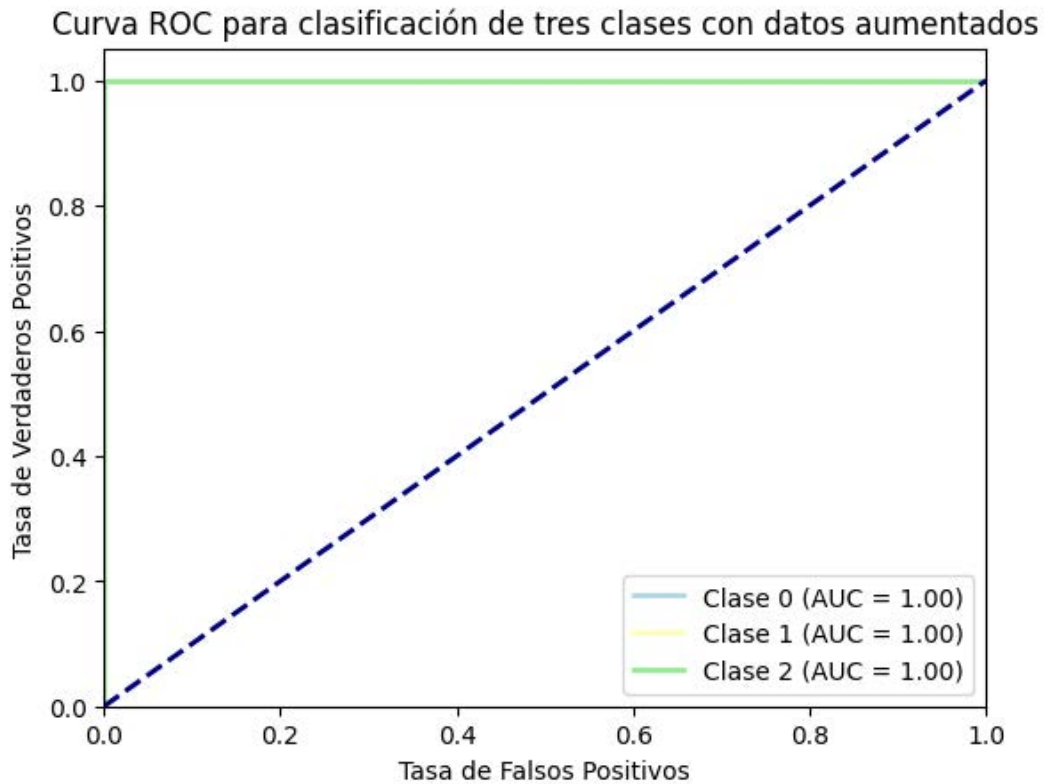


Figura 44: Curva de ROC.

La curva de ROC (Figura 44), muestra un modelo con capacidad óptima para identificar los verdaderos positivos de las imágenes de los delfines y sus tipos de aletas. Para este problema de clasificación de tipos de aletas, se tiene un AUC igual a uno para todas las clases, lo que indica un rendimiento bueno del modelo evaluado.



Figura 45: Clase verdadera VS predicciones (tres clases).

En la Figura 45 se observan algunas de las predicciones realizadas por el modelo. En la primera imagen, es de otro tipo de aletas, la predicción la coloca como entera. En la segunda imagen, que es típica, el resultado la clasifica como entera. En la tercera, cuya

clase verdadera es típica, se clasifica como otra. En la cuarta imagen, que es típica, se clasifica como entera. Finalmente, en la quinta imagen, es entera, su predicción es típica.

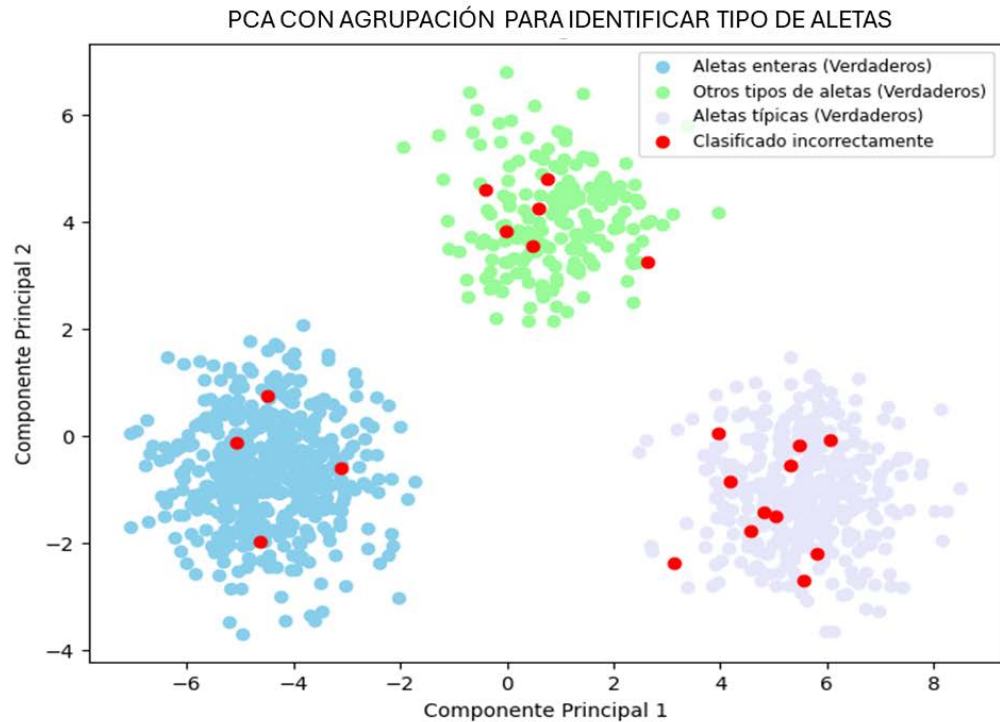


Figura 46: Componentes principales para las clases tipos de aletas en delfines

Para el gráfico PCA de tipos de aletas en delfines nariz de botella, la nube de puntos en color morado representa aletas típicas, los puntos verdes otros tipos de aletas, y los puntos azules aletas enteras. La distribución que tienen los puntos en el espacio bidimensional muestras cómo se agrupan las aletas de acuerdo con los patrones identificados por el modelo (Figura 46),.

Los puntos rojos dentro de cada agrupación corresponden a los individuos mal clasificados. En total se tienen veintitrés aletas que fueron predichas incorrectamente. Este tipo de errores de clasificación se deben a diversas razones, como la similitud en ciertas características que presentan las aletas entre grupos, la variabilidad natural de las aletas e incluso las limitaciones que presenta el modelo utilizado.

La exploración de las características específicas puede proporcionar información valiosa sobre este tipo de aplicaciones. La identificación de patrones erróneos también es considerada para mejorar la formulación de selección de características para modelos de clasificación que se apliquen a futuro.

6.2.7. Resultados de las imágenes satelitales y distribución de los vertebrados marinos

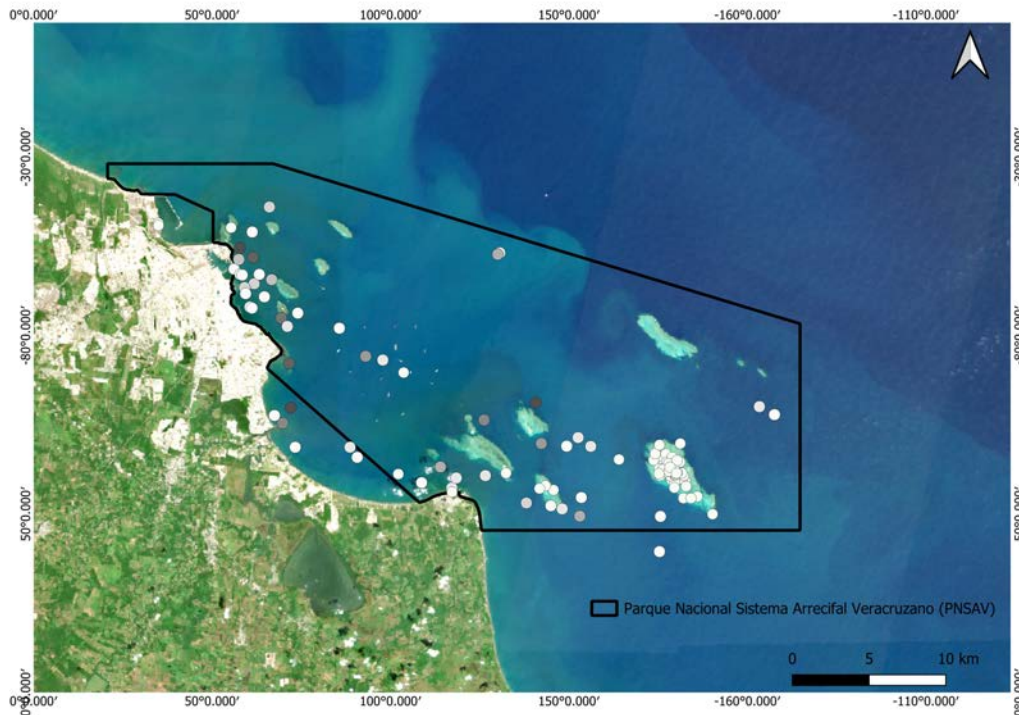


Figura 47: Mapa de zonas de obtención de imágenes convencionales

En la Figura 47, se visualiza el mapa del Polígono del Parque Nacional Sistema Arrecifal Veracruzano, se observa los diferentes sitios de la zona de estudio donde se obtuvieron las imágenes de los delfines y tortugas verdes, con el propósito de entender la distribución de estas especies en el PNSAV. La gama de colores, van desde los puntos más oscuros hasta lo más blanco, es decir entre mas intenso el color (oscuro) se indica una mayor observación y actividad de los delfines y tortugas.

Los puntos blancos indican las áreas que se tomaron menos fotografías, esto se debe a que el sitio tiene menor presencia de estas especies. Estas áreas de menor densidad de fotografía podrían representar zonas con menor actividad humana, esto es una gran oportunidad para explorar nuevas zonas para futuras investigaciones.

De manera general, el mapa brinda información de la estratificación de la distribución de delfines y tortugas en el PNSAV. La intensidad de los puntos no solo destaca la presencia relativa de las especies, sino también, la importancia de hábitats claves.

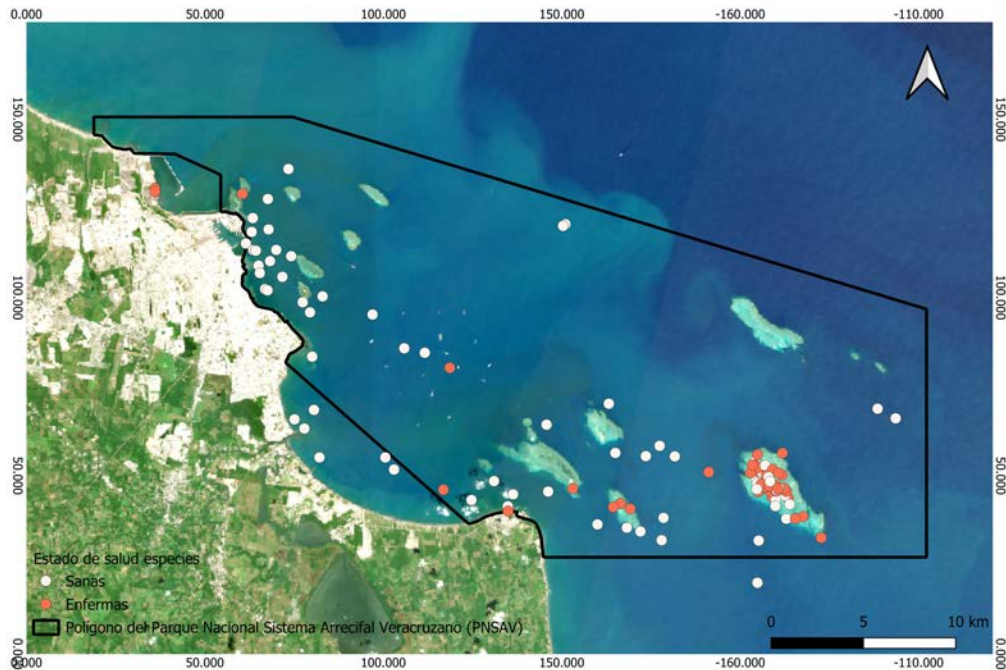


Figura 48: Coordenadas de las imágenes convencionales y su clasificación por estado de salud.

El mapa del PNSAV (Figura 48), indica la presencia de las dos especies de vertebrados marinos: Delfín y tortuga verde. Los puntos blancos representan a los animales considerados como sanos (todos los delfines caen en esta categoría). Mientras que los puntos naranjas representan a los animales enfermos (las tortugas son las más afectadas). Se considera de importancia la identificación de esta clasificación debido a que ambas especies son consideradas como indicadores biológicos, ya que son esenciales para evaluar la salud de los ecosistemas marinos.

Los puntos naranjas, identifican animales enfermos como se menciona en el párrafo anterior. Esta categoría es importante para los esfuerzos de conservación, ya que con esto se identifican zonas críticas donde la salud de poblaciones de delfines y tortugas pueden estar comprometidas, cabe mencionar que en esta investigación la especie, afectada con la enfermedad de fibropapilomatosis es la tortuga verde. Además de presenciar los puntos naranjas como animales enfermos, también indica posible presencia de contaminantes en el mar, que podrían afectar negativamente a estas especies. Por otra parte, se tiene a los puntos blancos que visibilizan las áreas que predominan los animales sanos.

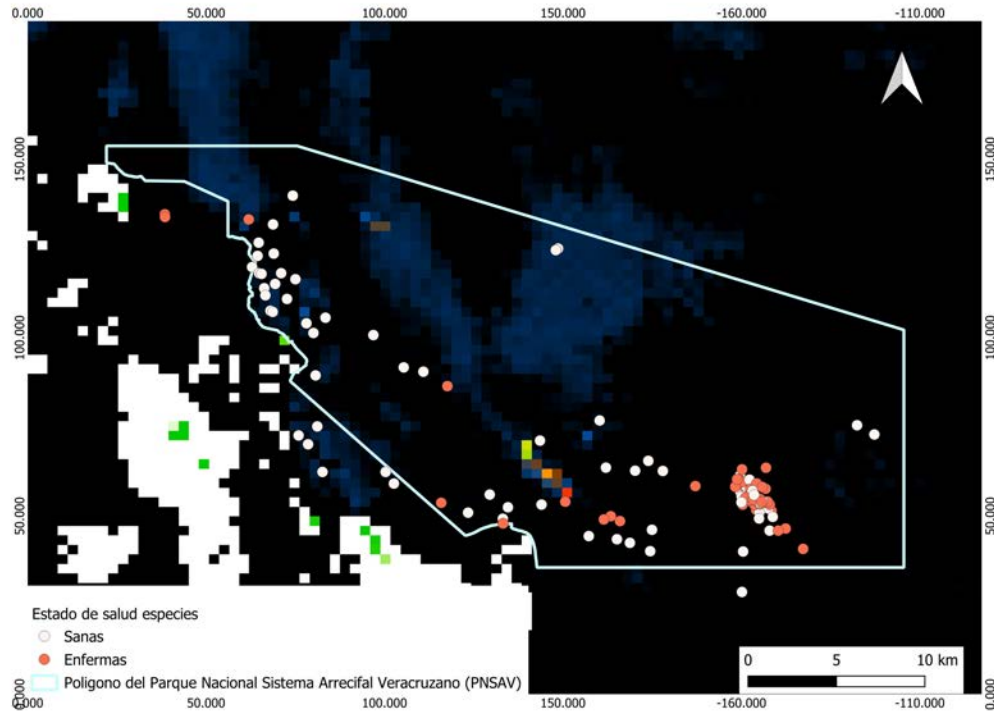


Figura 49: Coordenadas de las imágenes convencionales y la clorofila.

El mapa sobre la distribución de delfines y tortugas en el PNSAV y su relación con la clorofila (Figura 49), se observan algunos puntos naranjas y blancos se concentran alrededor de áreas donde se registra un rango de 1.5 a 3 de clorofila. Este tipo de información puede tener una relación entre la calidad del agua (medida por niveles de clorofila) y la salud de las especies en estudio.

En la zona norte (borde del polígono), se observa una presencia elevada de clorofila (punto verde), específicamente en un rango de 3 a 4. Esta área puede ser importante, ya que la concentración de este parámetro puede tener impactos diversos en la salud y el hábitat de los vertebrados. En estudios posteriores sería importante verificar como la presencia de clorofila en la periferia del polígono puede tener relación directa con la salud de las especies.

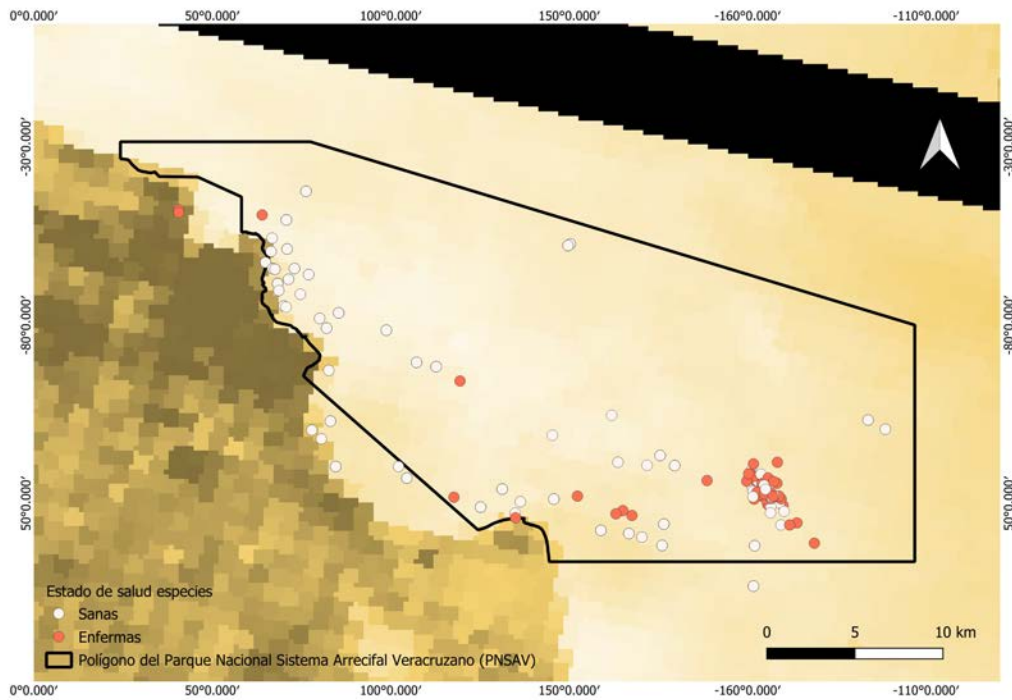


Figura 50: Coordenadas de las imágenes convencionales y la temperatura

El estudio espacial de especies en el Parque Nacional Sistema Arrecifal Veracruzano (PN-SAV), es una herramienta importante para investigaciones que ayuden a identificar factores que afecten a las especies y el ecosistema. La concentración de puntos naranjas indica existencia de otras posibles amenazas o condiciones adversas que impacten en la salud de las especies, específicamente en tortugas. La importancia de este tipo de análisis ayuda a los especialistas en mastozoología marina a presenciar áreas que pueden ser potenciales para investigaciones biológicas mas profundas, con esto detectar las causas subyacentes de la enfermedad en los animales.

El análisis espacial de especies en el PNSAV es crucial para la investigación y la comprensión de los efectos de la temperatura en la salud de las especies. La concentración de puntos naranjas indica especies enfermas que pueden ser afectadas por las extremas temperaturas. Este tipo de análisis es importante para los especialistas en mastozoología marina a identificar áreas para investigaciones biológicas, lo que permite detectar causas subyacentes de las enfermedades en los vertebrado marinos y tomar medidas para proteger su entorno.

En la Figura 50, se observa que tanto puntos naranjas y blancos se encuentran en un

mismo rango de temperatura del mar (20 y 27 grados Celsius). La temperatura es un factor importante y fundamental en la ecología y la salud de tortugas y delfines. Si este parámetro no se encuentra entre el rango ideal de las especies, pueden afectar aspectos como fisiología, reproducción, migración y relaciones tróficas.

Las imágenes satelitales sobre temperatura en el agua, amplían la comprensión y puntualizan la importancia de investigaciones para proteger y conservar la biodiversidad en el PNSAV.

6.3. Discusión

En la primera sección de resultados se evalúan diferentes arquitecturas de CNNs entre ellas, algunas pre entrenadas y una propuesta diseñada para esta investigación. También se emplearon algoritmos de clasificación clásica, con la finalidad de diferenciar imágenes convencionales del delfines y tortugas en varias propuestas de clasificación. La arquitectura propuesta (CNN_1AugD) con datos aumentados muestra un rendimiento óptimo para realizar esta tarea, con una precisión de 99% y F1 score de 1.

De acuerdo con (Vieira, 2023) el uso de métodos de aumento de datos, son factibles para cuando no se tienen disponibilidad suficiente de estos, logrando tener resultados óptimos al entrenar las arquitecturas. Sin embargo, para esta investigación a pesar de obtener resultados buenos para el modelo, uno de los obstáculos presentes fue el tiempo de procesamiento debido al tipo de máquina utilizada.

Por otra parte, el utilizar modelos de clasificación clásica para procesar imágenes son viables, ya que para esta investigación se obtuvieron resultados aptos, tal como lo mencionan (Muñoz *et al.*, 2014; Paixão *et al.*, 2018), que estos métodos ayudan al procesamiento de clasificación dando resultados óptimos. Finalmente, autores como (Dujon *et al.*, 2021; Gray *et al.*, 2018; Khan *et al.*, 2022), consideran una oportunidad buena, al utilizar la inteligencia artificial y modelos como las redes neuronales convolucionales para la clasificación de imágenes, esto con la finalidad de resolver problemas en diferentes áreas tal como es la mastozoología marina.

Finalmente se tiene evidencia que el aprendizaje profundo, especialmente la arquitectura propuesta para esta investigación (CNN1_AugD), para las clasificaciones propuestas, tiene una eficiencia alta, se destaca también los resultados de las arquitecturas pre entrenadas y los algoritmos clásicos. Se resaltan los logros y desafíos presentados durante el proceso de este proyecto, proporcionando cimientos sólidos para futuras investigaciones en el área de la biología marinas en colaboración con la inteligencia artificial.

Los resultados obtenidos de las imágenes satelitales y clasificaciones de vertebrados marinos en el PNSAV muestran información precisa e importante para los investigadores. El uso de misiones como la Sentinel es de uso valioso para este tipo de investigación, tal como lo mencionan (Fischbach y Douglas, 2021; Correa *et al.*, 2021), que son sensores

remotos que tienen utilidad precisa para abordar la vida marina, desde el monitoreo hasta la detección del hábitat en general.

Esta investigación a comparación de los autores mencionados muestra la distribución de los delfines y tortugas verdes. Otros aspectos para destacar es la zona de estudio, ya que por primera vez se realiza una investigación basada en inteligencia artificial y percepción remota. En otras investigaciones que se pueden consultar en el apartado de estado del arte, la mayoría de los investigadores utilizan imágenes de ballenas o morsas.

Por otra parte, los resultados se abordan desde una perspectiva espacial y su relación con el estado de salud de ambas especies, ya que estos animales son cruciales para evaluar la salud del ecosistema. Los parámetros evaluados son un inicio para investigar la conexión que existe entre la calidad del agua y el hábitat con la distribución de especies. Finalmente se menciona que las imágenes satelitales son importantes para entender la ecología y salud de las especies en esta área de estudio, ya que proporcionan información adicional y de suma importancia para la conservación de la biodiversidad marina.

Capítulo 7

Conclusiones

En esta investigación, se evaluaron modelos de aprendizaje profundo, específicamente redes neuronales convolucionales, para realizar tareas de clasificación de imágenes de vertebrados marinos. Para el primer experimento se tienen dos clases delfines o tortugas. Se creó una arquitectura de red neuronal convolucional (CNN_1) que para el primer experimento se utilizaron los datos originales y se demostró un rendimiento del modelo de 99 % de exactitud y un F1 score de 98 % para la clasificación de las dos clases y para el caso de las tres clases, se obtuvo un rendimiento óptimo. Para finalidad de tener una aplicación mas certera de las CNNs, se utilizó el aumento de datos, para mejorar el rendimiento especialmente en la arquitectura CNN_1AugD, que obtuvo una precisión de 99 % y un F1 score del 100 %.

La validación cruzada con $k = 5$, es un método que ayuda a evaluar la robustez de los modelos, para esta aplicación se destaca la eficacia de la arquitectura CNN_1AugD, 98 % y un 96 % de F1 score, para las tres clases 89 % y 88 %, mientras que para el resto de los experimentos también los resultados fueron buenos. Para la clasificación en términos de estado de salud de los vertebrados marinos, los modelos CNN_1AugD, VGG19_AugD y ResNet_AugD demostraron rendimientos aceptables, es importante mencionar que los experimentos anteriores fueron procesados con el aumento de datos. De manera general para todos los experimentos los algoritmos clásicos como, K_NN_AugD y RF_AugD, también presentan resultados aceptables en este tipo de clasificación.

Esta investigación también se clasificaron los tipos de aletas de los delfines en tres clases: Aletas enteras, otras y típicas. La arquitectura CNN_1, se procesó con los datos originales,

debido al desbalance de clases y la falta de datos, los resultados no fueron óptimos. Para la mejora de los resultados se implementó el aumento de datos en las tres clases, con esto se obtuvo mejor rendimiento de los modelos en términos de exactitud y F1 score. Para esta clasificación las arquitecturas pre entrenadas como VGG19_AugD, GoogleNet_AugD y ResNet_AugD, obtuvieron resultados sobresalientes, aunque la dificultad que presentan es el tiempo de proceso que fueron mas tardados. Los algoritmos clásicos K_NNAugD y RFAugD también proporcionaron evaluaciones aceptables.

La combinación y uso de las arquitecturas de red neuronal convolucional, el aumento de datos y la validación cruzada, son efectivas para la clasificación de imágenes de vertebrados marinos, abordando desafíos como la falta de datos y desbalance de clases. Los resultados obtenidos muestran que este tipo de metodologías son un enfoque prometedor para la identificación y clasificación de especies marinas a partir de imágenes convencionales.

El estudio de imágenes satelitales y la clasificación de vertebrados marinos en el PNSAV muestran información importante sobre las especies. Los estratos del mapa visualizan las áreas con mayor actividad de las especies, siendo áreas de importancia ecológica. La clasificación de animales sanos o enfermos, específicamente la presencia de fibropapilomatosis en tortugas verdes, muestran zonas críticas para la conservación y no buena salud del mar. La relación de la clorofila y la distribución de los vertebrados se vincula a la calidad del agua y la salud de las especies. Por su parte la temperatura del agua se identifica como un factor crucial ya que, si no es el ideal para las especies, puede afectar su ecología y salud. El uso de los métodos computacionales y la percepción remota son importantes en colaboración de la biología marina, puesto ayuda a la protección y conservación de biodiversidad en el área de estudio.

7.1. Reflexión de la investigación

Para esta investigación se tienen aplicaciones del aprendizaje profundo para la clasificación de imágenes de vertebrados marinos, donde se verifica que la inteligencia artificial tiene capacidad para procesar este tipo de datos de manera eficiente y precisa. Algunas características de las imágenes de las especies marina representan desafíos significativos que el aprendizaje profundo aborda con ayuda de la identificación de patrones. Se destaca la importancia de el entendimiento sobre el entorno espacial de los vertebrados marinos.

Incluir información espacial en el análisis permite no solo la clasificación de imágenes de las especies, sino también la comprensión de su preferencia de hábitat. Esto tiene implicaciones directas para la conservación marina y la gestión de ecosistemas acuáticos. Este estudio también tiene posibilidad de aplicarse en la monitorización y protección de la biodiversidad marina. Por la parte tecnológica, la investigación basada en clasificar imágenes de vertebrados marinos destaca la importancia de colaboración interdisciplinaria. El complementar la tecnología de la información y aprendizaje profundo es esencial para abordar los desafíos específicos de este campo y lograr avances significativos.

7.2. Contribución de la investigación

Esta investigación contribuye a diversos puntos, a continuación, se presentan los más importantes:

- Avances en la clasificación de imágenes de especies marinas en combinación con la inteligencia artificial, en la zona de estudio, con esto se obtienen resultados mas precisos y eficientes para identificar a estos animales, superando limitaciones de métodos tradicionales.
- Incorporación de información espacial, es importante puesto da una perspectiva mas completa de la vida marina, ya que no solo permite clasificar las especies mediante imágenes, también con las imágenes espaciales, se comprende la preferencia de hábitat y su relación con el ecosistema.
- El trabajo interdisciplinario, entre personas expertas en el área de mastozoología marina, con expertos de las ciencias computacionales, para promover investigación que abordan desafíos específicos y con esto acelerar las soluciones con innovación en el área.
- Facilidad en la automatización de las clasificaciones de imágenes, que reduce el trabajo manual y promueve de manera óptima el proceso de análisis, permitiendo a los biólogos trabajar con grandes volúmenes de datos y realizar estudios mas exhaustivos.

7.3. Trabajos a futuro

Se plantea el refinamiento de los modelos de aprendizaje profundo e incorporación de otras técnicas avanzadas para el procesamiento digital de imágenes. Otro aspecto que se considerará a futuro es la fusión de datos cuantitativos de factores ambientales, para proporcionar mas información para comprender mejor el entorno marino y su relación con las especies. El desarrollo de plataformas para lograr un monitoreo de la fauna, se utilizarán métodos computacionales que ayuden al desarrollo de este tipo de investigación. La colaboración entre investigadores y la implementación de redes de sensores para la recopilación de datos podría enriquecer enormemente los conjuntos de datos, proporcionando información más detallada y diversa para mejorar la calidad de la investigación.

- Refinar las técnicas de aprendizaje profundo utilizando métodos de regularización, para el uso de arquitecturas mas complejas.
- Fusionar datos cuantitativas de los parámetros biológicos.
- Utilizar métodos como el aprendizaje por refuerzo para tener eficiencia en el monitoreo de los datos.
- Emplear técnicas de visualización de datos avanzados como la realidad virtual y aumentada.
- Uso de tecnología submarina para la obtención de datos.
- Identificar el tipo de enfermedad en las especies aplicando inteligencia artificial.

Bibliografía

- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [Epub 2021 Mar 31]. *J Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Azlan, M., Hazirah, N., Halim, A., Amalia, N., & Ahmad, I. (2019). A Study on Image Classification Based on Deep Learning and TensorFlow. *International Journal of Engineering Research and Technology*, 12(4), 563-569.
- Berg, P., Santana Maia, D., Pham, M.-T., & Lefèvre, S. (2022). Weakly Supervised Detection of Marine Animals in High Resolution Aerial Images. *Remote Sensing*, 14, 339. <https://doi.org/10.3390/rs14020339>
- Bonilla, C. (2020). *Redes convolucionales* [Trabajo de fin de grado]. Universidad de Sevilla.
- Breiman, L. (2001). Bosques aleatorios. *Aprendizaje automático*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Centeno, F. (2019). *Deep Learning*.
- Copernicus Data Space Ecosystem [Accessed on [fecha de acceso]]. (2023).
- Correa, A. A., Quoos, J. H., Barreto, A. S., Groch, K. R., & Eichler, P. (2021). Use of satellite imagery to identify southern right whales (*Eubalaena australis*) on a Southwest Atlantic Ocean breeding ground. *Marine Mammal Science*.
- Deepa & Lalwani, K. N. (2019). Image Classification and Text Extraction using Machine Learning. *2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 680-684. <https://doi.org/10.1109/ICECA.2019.8821936>
- Dujon, A. M., Ierodionou, D., Geeson, J. J., Arnould, J. P., Allan, B. M., Katselidis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals in UAV

- imagery: effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Conservation*, 7.
- European Space Agency. (2023). Sentinel Mission [Accesado on {fecha de acceso}].
- Fischbach, A. S., & Douglas, D. C. (2021). Evaluation of Satellite Imagery for Monitoring Pacific Walruses at a Large Coastal Haulout. *Remote Sensing*, 13, 4266. <https://doi.org/10.3390/rs13214266>
- Gershenson, C. (2003). *Artificial Neural Networks for Beginners*.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*, (1209). https://scholarworks.utep.edu/cs_techrep/1209
- Golovko, V., Egor, M., Brich, A., & Sachenko, A. (2017). A Shallow Convolutional Neural Network for Accurate Handwritten Digits Classification. *Communications in Computer and Information Science*, 673, 77-85. https://doi.org/10.1007/978-3-319-54220-1_8
- Gómez, H. (2021). *Foto-identificación de Tortuga Verde (Chelonia mydas) en el Parque Nacional Sistema Arrecifal Veracruzano* [Tesis de maestría, Facultad de Biología, Universidad Veracruzana].
- Gómez-Salazar, C., Trujillo, F., & Whitehead, H. (2011). Photo-Identification: A Reliable and Noninvasive Tool for Studying Pink River Dolphins (*Inia geoffrensis*). *Aquatic Mammals*, 37, 472-485.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [Recuperado de <http://www.deeplearningbook.org>]. MIT Press.
- Gray, P., Fleishman, A., Klein, D., Mckown, M., Bézy, V., Lohmann, K., & Johnston, D. (2018). A Convolutional Neural Network for Detecting Sea Turtles in Drone Imagery. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210x.13132>
- He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>

- Hernández, D. L., Sánchez, J. P., León, E., Frausto, J., & González, J. (2022). Análisis de arquitectura de redes neuronales convolucionales aplicadas a la detección de cenicilla polvorienta en cultivos de jitomate. *COMENI-2021, VII CONGRESO NACIONAL DE RIEGO DRENAJE Y BIOSISTEMAS*.
- Hipiny, I., Ujir, H., Mujahid, A., & Yahya, N. (2018). Towards automated biometric identification of sea turtles (*Chelonia mydas*). *Journal of ICT Research and Applications*, 12, 256-266. <https://doi.org/10.5614/2Fitbj.ict.res.appl.2018.12.3.4>
- Ho, C.-R., & Liu, A. (2023). Preface: Remote Sensing Applications in Ocean Observation. *Remote Sensing*, 15(415), 415. <https://doi.org/10.3390/rs15020415>
- Höschle, C., Cubaynes, H. C., Clarke, P. J., Humphries, G., & Borowicz, A. (2021). The Potential of Satellite Imagery for Surveying Whales. *Sensors*, 21, 963. <https://doi.org/10.3390/s21030963>
- Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*.
- Hridayami, P., Putra, I. K. G. D., & Wibawa, K. S. (2019). Fish Species Recognition Using VGG16 Deep Convolutional Neural Network. *Journal of Computing Science and Engineering*, 13(3), 124-130.
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2017.243>
- Karamizadeh, S., Abdullah, S., Manaf, A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*. <https://doi.org/10.4236/jsip.2013.43B031>
- Kavanaugh, M. T., Bell, T., Catlett, D., Cimino, M. A., Doney, S. C., Klajbor, W., Messié, M., Montes, E., Muller-Karger, F. E., Otis, D., Santora, J. A., Schroeder, I. D., Triñanes, J., & Siegel, D. A. (2021). SATELLITE REMOTE SENSING AND THE MARINE BIODIVERSITY OBSERVATION NETWORK: CURRENT SCIENCE AND FUTURE STEPS. *Oceanography*, 34(2), 62-79. <https://www.jstor.org/stable/27085039>
- Khan, C., Blount, D., Parham, J., & et al. (2022). Inteligencia artificial para la identificación fotográfica de ballenas francas: de la competencia de ciencia de datos a la

- colaboración mundial. *Mammalian Biology*, 102, 1025-1042. <https://doi.org/10.1007/s42991-022-00253-3>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86, 2278-2324. <https://doi.org/10.1109/5.726791>
- Liu, K., Wang, S., Wu, Y., & Zhang, W. (2023). Underwater Inhomogeneous Light Field Based on Improved Convolutional Neural Net Fish Image Recognition. *Open Journal of Applied Sciences*, 13(7). <https://doi.org/10.4236/ojapps.2023.137086>
- López, F. (2019). *Clasificación de imágenes usando redes neuronales convolucionales* [Tesis de Maestría]. Universidad Autónoma Metropolitana, Unidad Azcapotzalco.
- Lorente, Ò., Riera, I., & Rana, A. (2021). Image Classification with Classic and Deep Learning Techniques.
- Luo, L. (2021). *J. Phys.: Conf. Ser.* 2083, 032054.
- Maglietta, R., Carlucci, R., Fanizza, C., Dimauro, G., & Maglietta, R. (2022). Machine Learning and Image Processing Methods for Cetacean Photo Identification: A Systematic Review. *IEEE Access*, 10, 1-1. <https://doi.org/10.1109/ACCESS.2022.3195218>
- Makinde, F., Ako, C., Orodu, O., & Asuquo, I. (2012). Prediction of Crude Oil Viscosity Using Feed-forward Back-propagation Neural Network (FFBPNN). *Petroleum and Coal*, 54.
- Mancera-Pineda, J. E., Gavio, B., & Lasso-Zapata, J. (2013). PRINCIPALES AMENAZAS A LA BIODIVERSIDAD MARINA. *Actualidades Biológicas*, 35(99), 111-133. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-35842013000200001&lng=en&tlng=es
- Martínez-Serrano, I. (2005). *Asociaciones entre individuos de ballena azul (Balaenoptera musculus) en el suroeste del Golfo de California, México* [Tesis de Maestría]. Departamento de Pesquería y Biología Marina, Instituto Politécnico Nacional, Centro Interdisciplinario de Ciencias Marinas.

- Martínez-Serrano, I., Serrano, A., Heckel, G., & Schramm, Y. (2011). Distribución y ámbito hogareño de toninas (*Tursiops truncatus*) en Veracruz, México. *Ciencias Marinas*, 37, 379-392.
- Martínez-Serrano, I., Suárez-Domínguez, E., Cobos-Silva, J., Bello-Sánchez, E., Ramos-Díaz, A., Robledo-Catalina, M., Chamlaty-Fayad, Y., Mazaba-Lara, J., & Valencia, A. (2018). *Programa de Monitoreo Biológico basado en Especies Indicadoras de Integridad Ecológica en el PNSAV. Informe técnico 2016-2018* (inf. téc.) (164 pp. + anexos). Administración Portuaria Integral de Veracruz, Universidad Veracruzana.
- Matich, D. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*.
- McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Neurons Activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Medar, R., Rajpurohit, V. S., & Rashmi, B. (2017). Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning. *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1-6. <https://doi.org/10.1109/ICCUBEA.2017.8463779>
- Mera-Ortiz, M. (2016). *Foto-identificación de Tortuga blanca (Chelonia mydas) en el campamento Tamul, Moon Palace Resorts, Quintana Roo* [Tesis de maestría, Instituto de Ciencias Biológicas, UNICACH].
- Morales, G. N. A. (2022). *Monitoreo acústico de tiburones en el Parque Nacional Sistema Arrecifal Veracruzano, México* [Tesis de maestría, Maestría en Ciencias Biológicas, Facultad de Biología, Universidad Veracruzana] [Fecha de examen de grado: 30 enero 2022].
- Muñoz, P., Cabrera, D., Carvajal-Gaméz, B., Gallegos-Funes, F., & Gendron, D. (2014). Segmentación automática de imágenes en RGB, utilizando el método de Fuzzy C-means (FCM), de la morfología matemática para la ayuda de foto identificación de cetáceos. *Research in Computing Science*, 74, -.
- Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective* [Library of Congress Cataloging-in-Publication Information. ISBN 978-0-262-01802-9]. Massachusetts Institute of Technology.
- Nahm, F. (2022). Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians [Epub 2022 Jan 18]. *Korean Journal of Anesthesiology*, 75(1), 25-36. <https://doi.org/10.4097/kja.21209>

- Nichols, J., Chan, H., & Baker, M. (2018). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, 11. <https://doi.org/10.1007/s12551-018-0449-9>
- Nilsson, N. (1998). *Introduction to Machine Learning: An Early Draft of Proposed Textbook* (inf. téc.). Robotics Laboratory, Stanford University. Stanford, CA.
- Ortega-Ortiz, C., Wonneberger, E., Martínez-Serrano, I., Kono-Martínez, T., Villegas Zurita, F., Paredes, L., Llamas-González, M., Olivos, A., Liñán-Cabello, M., & Verduzco-Zapata, M. (2019). Consequences Potentially Related to a Meteorological Event on a Resident Group of Bottlenose Dolphins (*Tursiops truncatus*) from the Mexican Pacific. *Aquatic Mammals*, 45, 99-105. <https://doi.org/10.1578/AM.45.1.2019.99>
- Özdemir, C. (2023). Avg-topk: A new pooling method for convolutional neural networks. *Expert Systems with Applications*, 223, 119892. <https://doi.org/10.1016/j.eswa.2023.119892>
- Padial. (2019). Técnicas de programación "Deep Learning": simulacro o realización artificial de la inteligencia. *Naturaleza y Libertad*, (12), Universidad de Málaga, ISSN: 2254-9668.
- Paixão, W., Paixão, T., Costa, M., Andrade, J., Garcia Pereira, F., & Komati, K. (2018). Texture Classification of Sea Turtle Shell Based on Color Features: Color Histograms and Chromaticity Moments. *International Journal of Artificial Intelligence & Applications*, 9, 55-67. <https://doi.org/10.5121/ijaia.2018.9205>
- Panigrahi, S., Nanda, A., & Swarnkar, T. (2018). Enfoque de aprendizaje profundo para la clasificación de imágenes. *2018 Segunda Conferencia Internacional sobre Ciencia de Datos y Análisis de Negocios (ICDSBA)*, 511-516. <https://doi.org/10.1109/ICDSBA.2018.00101>
- Paraschiv, M., Padrino, R., Casari, P., Bigal, E., Scheinin, A., Tchernov, D., & Fernández Anta, A. (2022). Classification of Underwater Fish Images and Videos via Very Small Convolutional Neural Networks. *Journal of Marine Science and Engineering*, 10, 736. <https://doi.org/10.3390/jmse10060736>
- Peña, D. (2002). Análisis de Datos Multivariantes.
- Pikabea, I. (2022). *Revisión de modelos de aprendizaje en el ámbito de pregunta-respuesta*.

- Prieto, R., Herrera, A., Pérez, J., & Padrón-Godínez, A. (2000). El Modelo Neuronal de McCulloch y Pitts: Interpretación Comparativa del Modelo.
- Ramos-Arredondo, R., Carvajal-Gómez, B., Gallegos-Funes, F., & Gendron, D. (2014). Multi-spatial Classifier for Blue Whale Images using Photo-identification Method. *Research in Computing Science*, 82, 31-40. <https://doi.org/10.13053/rcs-82-1-3>
- Ramos-Arredondo, R., Carvajal-Gómez, B., Gallegos-Funes, F., & Gendron, D. (2015). Clasificador de pigmentación de la ballena azul aplicando el método SIFT en medios no controlados. *Research in Computing Science*, 91, 69-79. <https://doi.org/10.13053/rcs-91-1-6>
- Rathi, D., Sreedevi, I., & Jain, S. (2018). Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning. *Nombre de la revista o conferencia*, Páginas.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Validación cruzada. En L. Liu & M. Özsu (Eds.), *Enciclopedia de sistemas de bases de datos*. Springer. https://doi.org/10.1007/978-0-387-39940-9_565
- Ren, Z., Wang, S., & Zhang, Y. (2023). Weakly Supervised Machine Learning. https://www.researchgate.net/publication/370342034_Weakly_supervised_machine_learning
- Rodríguez, J., & Ruíz, J. (2010). Conservación y protección de ecosistemas marinos: conceptos, herramientas y ejemplos de actuaciones. *Ecosistemas*, 19(2), 5-23.
- Ruíz, I. (2014). *Desplazamiento de toninas (Tursiops truncatus) en la costa central de Veracruz, México* [Tesis de maestría, Facultad de Biología, Universidad Veracruzana].
- Sarasota Dolphin Research Program. (2006). *Manual for Field Research and Laboratory Activities* [Available through www.sarasotadolphin.org].
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3-29. <https://doi.org/10.1177/1536867X20909688>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.

- Smola, A., & Vishwanathan, S. (2008). *Introduction to Machine Learning*.
- Stack, P. (2021). *Métodos de aprendizaje supervisado y no supervisado para la estimación de microestructuras cerebrales en datos DWMR* [Tesis de maestría, Centro de Investigación de Matemáticas, A.C.].
- Štifanić, D., Štifanić, J., Car, Z., & Čep, R. (2020). Use of Convolutional Neural Network for Fish Species Classification. *Journal of Maritime Transportation Science*, 59, 131-142. <https://doi.org/10.18048/2020.59.08>.
- Suárez, F. (2020). Línea base de la tortuga verde (*Chelonia mydas*) en el Área de Conservación Ambiental Estuario de Virrilá, Sechura.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper with Convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv preprint arXiv:1905.11946*.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2014). *Introduction to Data Mining*. Pearson Education Limited. www.pearsoned.co.uk
- Teuwen, J., & Moriakov, N. (2020). Convolutional Neural Networks. En *Handbook of Medical Image Computing and Computer-Assisted Intervention*. Elsevier. <https://doi.org/10.1016/B978-0-12-816176-0.00025-9>
- Tian, Y., Su, D., Lauria, S., & Liu, X. (2022). Recent advances on loss functions in deep learning for computer vision. *Neurocomputing*, 497, 129-158. <https://doi.org/10.1016/j.neucom.2022.04.127>
- Ting, K. (2010). Matriz de Confusión. En *Enciclopedia de Aprendizaje Automático y Minería de Datos*.
- Trujillo, J., March, J., & Sorribas, A. (2004). Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina. *Med Clin (Barc)*, 122(Supl 1), 59-67.
- Valdes-Arellanes, M. P., Serrano, A., Heckel, G., Schramm, Y., & Martínez-Serrano, I. (2011). Abundancia de dos poblaciones de toninas (*Tursiops truncatus*) en el norte de Veracruz, México [Recibido: 05 enero 2010; aceptado: 28 mayo 2010]. *Revista Mexicana de Biodiversidad*, 82, 227-235.

- Vieira, J. (2023). *Usage of Convolutional Neural Networks for Identifying Marine Mammal Individuals* [Tesis de maestría, Universidade da Madeira].
- Visa, S., Ramsay, B., Ralescu, A. L., & van der Knaap, E. (2011). Confusion Matrix-based Feature Selection. En S. Visa, A. Inoue & A. L. Ralescu (Eds.), *MAICS* (pp. 120-127).
- Wang, Z., Zhang, J., Bai, L., Chang, H., Chen, Y., Zhang, Y., & Tao, J. (2024). A Deep Learning Based Platform for Remote Sensing Images Change Detection Integrating Crowdsourcing and Active Learning. *Sensors*, *24*, 1509. <https://doi.org/10.3390/s24051509>
- White, J., & Power, S. (2023). k-Fold Cross-Validation Can Significantly Over-Estimate True Classification Accuracy in Common EEG-Based Passive BCI Experimental Designs: An Empirical Investigation. *Sensors (Basel)*, *23*(13), 6077. <https://doi.org/10.3390/s23136077>
- Wiley, V., & Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. *International Journal of Artificial Intelligence Research*, *2*(1), 22. <https://doi.org/10.29099/ijair.v2i1.42>
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78-83. <https://doi.org/10.1109/IACC.2016.25>
- Yamashita, R., Nishio, M., Do, R., *et al.* (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, *9*, 611-629. <https://doi.org/10.1007/s13244-018-0639-9>
- Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2023). Image Data Augmentation for Deep Learning: A Survey.
- Yorioka, D., Kang, H., & Iwamura, K. (2020). Data Augmentation For Deep Learning Using Generative Adversarial Networks. *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 516-518. <https://doi.org/10.1109/GCCE50665.2020.9291963>
- Yu, L., & Zhou, N. (2021). Survey of Imbalanced Data Methodologies. *arXiv preprint arXiv:2104.02240*.

- Yu, Y. (2022). Deep Learning Approaches for Image Classification. *EITCE 2022: 2022 6th International Conference on Electronic Information Technology and Computer Engineering*. <https://doi.org/https://doi.org/10.1145/3573428.3573691>
- Zhang, D., Wang, Xu & Chen. (2019). Research on Scene Classification Method of High-Resolution Remote Sensing Images Based on RFPNet. *Applied Sciences*, 9, 2028. <https://doi.org/10.3390/app9102028>
- Zhou, Z., Yang, X., Ji, H., & Zhu, Z. (2023). Mejora de la precisión de la clasificación de peces e invertebrados utilizando redes neuronales convolucionales residuales [https://doi.org/10.1093/icesjms/fsad041]. *ICES Journal of Marine Science*, 80(5), 1256-1266.