

L^AT_EX Miniproyecto 3 - Implementación de machine learning en clasificación.

John Acosta
Universidad de los Andes
Ingeniería Biomédica

ja.acostar1@uniandes.edu.co

Leticia Vidales
Universidad de los Andes
Ingeniería Biomédica

v.vidales@uniandes.edu.co

Juana Salazar
Universidad de los Andes
Ingeniería Biomédica

jd.salazarcl@uniandes.edu.co

1. Introducción

La clasificación de imágenes es una de las tareas fundamentales en el aprendizaje automático, que consiste en asignar una etiqueta o clase a una imagen seleccionada de un conjunto finito de categorías posibles [1]. En este proyecto, nos centramos en la clasificación automática de mamíferos a partir de imágenes, utilizando descriptores visuales y modelos de clasificación. El problema principal radica en clasificar imágenes de mamíferos en su especie correspondiente utilizando la base de datos iNaturalist, la cual contiene imágenes tomadas en condiciones reales por observadores de campo. Estas imágenes no solo incluyen vistas claras de los animales, sino también rastros, huellas y excrementos. Ahora bien, la clasificación de especies animales es de importancia para la biodiversidad, conservación, monitoreo ecológico e investigación científica, y, como señalan los biólogos Hamard & Heerah, K. [2], automatizar este proceso puede ayudar a los biólogos y conservacionistas a procesar grandes volúmenes de datos de manera eficiente y precisa, facilitando la toma de decisiones y el análisis de ecosistemas.

Asimismo, para clasificar manualmente las especies, los biólogos suelen observar varios indicadores, como la coloración del pelaje o piel. También se consideran la textura y patrones del cuerpo, como rayas o manchas, y la forma corporal, especialmente de elementos como las orejas, hocico o la cola. Además, el contexto ambiental en el que se encuentra el animal y el tamaño relativo del animal respecto al entorno son claves para una correcta identificación [3]. No obstante, el proceso de clasificación automática enfrenta desafíos, como la variabilidad intraespecífica, ya que individuos de una misma especie pueden diferir significativamente en aspectos como edad, sexo o estación del año. También existen similitudes entre especies del mismo orden, lo que puede hacer que la clasificación resulte más difícil. Las condiciones ambientales también afectan la calidad de las imágenes, como la iluminación deficiente o la presencia de obstáculos (ramas, piedras), lo que complica la identificación precisa. Un ejemplo claro es cuando se capturan imágenes nocturnas, donde la falta de color y la

distorsión de los ojos dificultan la clasificación [4].

Por otro lado, la gran cantidad de datos hace que el análisis manual sea impráctico, mientras que el machine learning ofrece una solución eficiente al permitir una clasificación consistente y escalable de grandes volúmenes de imágenes en tiempos reducidos [5]. Esta tecnología es crucial para el monitoreo de especies en peligro, proporciona una forma más eficiente de seguir su presencia y distribución. Además, puede integrarse en entornos educativos para la sensibilización sobre la biodiversidad. Por último, el objetivo de clasificar automáticamente imágenes de mamíferos es facilitar el análisis masivo y preciso de datos visuales para mejorar la conservación de especies y la toma de decisiones ecológicas, además de contribuir al desarrollo de modelos predictivos para el comportamiento de las poblaciones animales.

2. Estado del arte

La clasificación de imágenes de animales, mediante Inteligencia Artificial y técnicas de visión artificial, ha sido ampliamente investigada. Métodos como Machine Learning y Deep Learning se han utilizado para detectar y clasificar especies silvestres, contribuyendo al conocimiento ambiental y la conservación de los animales. Se han implementado algoritmos como las *Máquinas de Soporte Vectorial (SVM)* y *K-Nearest Neighbors (kNN)*, así como *Redes Neuronales Convolucionales (CNN)* [6].

Las *Máquinas de Soporte Vectorial (SVM)* son algoritmos supervisados para clasificación binaria. En su proceso, primero se entrena el clasificador con un conjunto de datos, obteniendo un conjunto de parámetros y un hiperplano que separa las clases. Luego, el clasificador asigna clases a nuevos datos según la región de clasificación [7].

El algoritmo *K-Nearest Neighbors (kNN)* se utiliza en el reconocimiento de patrones. Consiste en seleccionar un valor k , calcular la distancia entre el nuevo punto y los puntos del conjunto de entrenamiento, y asignar la clase más común entre los k vecinos más cercanos [8].

Las *Redes Neuronales Convolucionales (CNN)* son la técnica más empleada en la clasificación de animales. Es-

tas redes incluyen tres tipos de capas: convolucionales para extraer características, de pooling para reducir las dimensiones, y fully connected para realizar la clasificación final [9].

Por otro lado, los descriptores en la clasificación de animales incluyen características como tamaño, estructura, texturas, colores y patrones, como los de vacas o tigres [6]. Para evaluar el desempeño de los modelos, se usan métricas como: i) *precisión*, que mide la proporción de predicciones correctas; ii) *exactitud*, que representa el porcentaje de predicciones correctas en relación al total; iii) *recall*, que mide las predicciones correctas entre todos los casos positivos; y iv) *F1*, que combina precisión y recall para equilibrar ambas métricas [6].

Finalmente, respecto a las bases de datos existentes para la clasificación de animales, se encuentran *eBird* que abarca más de 10,000 clases de aves [10], *iNaturalist* con más de 100,000 clases de animales y plantas [11], *ZIMS* que gestiona más de 22,000 clases en la gestión de zoológicos y acuarios[12], y *Animal Diversity Web (ADW)* con más de 5,000 clases de especies de animales caracterizados como vertebrados e invertebrados [13].

3. Base de datos

En primer lugar, fue necesario realizar un submuestreo de la base de datos original, que contenía 23 clases, con el fin de agilizar el procesamiento de los algoritmos implementados y reducir la carga en nuestros equipos. El submuestreo consistió en seleccionar las clases que formarían parte de nuestro grupo de estudio, reduciendo la base de datos a 5 clases, las cuales fueron organizadas en carpetas para entrenamiento, validación y prueba. Las clases seleccionadas fueron *Canis latrans*, *Tamias striatus*, *Otospermophilus variegatus*, *Phoca vitulina* y *Sciurus carolinensis*.

Se analizó la cantidad total de datos de cada especie en cada carpeta (ver [Tabla 1](#)) y se observó que no era posible hacer una comparación directa entre especies debido a una distribución no uniforme, lo que dificultaba determinar si la partición de los datos en las carpetas de *train*, *valid* y *test* era adecuada. Para resolver esto, se normalizó la cantidad de datos en cada carpeta, con el fin de establecer una escala comparable y permitir una evaluación justa de la distribución de cada especie en los diferentes conjuntos. Los resultados de esta normalización se muestran en la [Tabla 2](#). Con las nuevas proporciones, se determinó que la distribución era más equilibrada, con un porcentaje similar de datos para todas las especies en cada conjunto. Asimismo, se destaca la importancia de que la distribución de las especies esté equilibrada dentro de cada conjunto de datos, lo que garantiza una representación adecuada de todas las especies durante el entrenamiento y la validación, evitando sesgos hacia una especie en particular. Además, aunque la distribución de clases no tiene que ser idéntica

entre los conjuntos, sí debe ser similar para evitar sesgos en el modelo. Cada conjunto tiene un rol específico: *train* requiere más datos para un entrenamiento efectivo, mientras que *valid* y *test* contienen cantidades menores, pero suficientes, para evaluar el rendimiento del modelo y ajustar los hiperparámetros.

Table 1. Cantidad de datos de cada clase (Especie) para cada conjunto

| Especie | Train | Valid | Test |
|----------------------------|-------|-------|------|
| Canis latrans | 1286 | 429 | 429 |
| Otospermophilus variegatus | 515 | 172 | 173 |
| Phoca vitulina | 713 | 238 | 238 |
| Sciurus carolinensis | 1795 | 599 | 599 |
| Tamias striatus | 854 | 285 | 285 |

Table 2. Cantidad de datos normalizados (proporción) de cada clase (Especie) para cada conjunto

| Especie | Train | Valid | Test |
|----------------------------|-----------|-----------|----------|
| Canis latrans | 0.24908 | 0.248984 | 0.24884 |
| Otospermophilus variegatus | 0.0997482 | 0.0998259 | 0.100348 |
| Phoca vitulina | 0.138098 | 0.138131 | 0.138051 |
| Sciurus carolinensis | 0.347666 | 0.347649 | 0.347448 |
| Tamias striatus | 0.165408 | 0.165409 | 0.165313 |

Por otro lado, es importante destacar que el base de datos presenta una gran diversidad en cuanto a las características de las imágenes. Este incluye tanto avistamientos directos de las especies como evidencia indirecta, como huellas, excrementos y otros rastros. Además, las imágenes no fueron capturadas de forma uniforme, como puede evidenciarse en la [figura 1](#), algunas presentan variaciones en el brillo, otras están en escala de grises o a color, y hay diferencias en el tamaño. Para obtener información más detallada y estadísticas sobre la base de datos, se puede consultar la sección de [Anexos](#).



Figure 1. Imágenes de cada especie para cada *train*, *valid* y *test*.

4. Metodología

4.1. Descriptores

4.1.1. Unimodales

Los descriptores son elementos clave en la clasificación de imágenes, ya que permiten extraer características relevantes a partir de propiedades como la forma, la textura y el color [14]. Para evaluar su efectividad, se analizaron cuatro tipos principales: descriptor de color, descriptor de textura, descriptor combinado de color y localización (mediante pirámide espacial), y descriptor de forma (Histogram of Oriented Gradients, HOG). Los descriptores de color y textura se evaluaron utilizando un modelo de clasificación basado en k -vecinos más cercanos (k -NN), mientras que los descriptores espaciales y de forma se analizaron mediante máquinas de soporte vectorial (SVM), permitiendo además la exploración de distintos valores del parámetro de regularización C y del tipo de kernel empleado. El descriptor de color se basa en la construcción de histogramas que representan la distribución global de los colores en la imagen, siendo útil para distinguir imágenes con diferencias cromáticas notables, aunque carece de sensibilidad a la disposición espacial. El descriptor de textura, captura estructuras locales repetitivas a partir de las relaciones de intensidad entre píxeles vecinos, resultando útil para reconocer materiales o superficies. Por otro lado, el descriptor basado en la pirámide espacial incorpora información tanto de color como de localización, dividiendo la imagen en regiones jerárquicas y calculando histogramas por región, lo que permite una representación más robusta ante variaciones en la composición espacial. Finalmente, el descriptor de forma HOG se fundamenta en la distribución de gradientes de intensidad orientados, proporcionando una representación efectiva de los contornos y estructuras presentes en la imagen. En conjunto, esta evaluación comparativa permitió identificar las fortalezas y limitaciones de cada descriptor en combinación con distintos clasificadores, aportando información valiosa para el diseño de sistemas de reconocimiento visual más robustos y precisos. Experimentar con diversos descriptores es crucial para identificar que información es más útil en la clasificación de nuestras especies. Es importante destacar que el descriptor óptimo varía según el tipo de imágenes y el problema específico, ya que no existe una solución universal. Además, la combinación adecuada de descriptor y clasificador, junto con la optimización de parámetros, impacta significativamente los resultados. Por ello, la experimentación con múltiples opciones es fundamental para encontrar la combinación que mejor se ajuste a nuestros datos

4.1.2. Multimodales

Para evaluar el impacto de combinar distintos descriptores de imagen, construimos un modelo basado en la concate-

nación de vectores que representan color, forma (HOG) y textura. A partir del descriptor base, generamos nuevas representaciones al unir diferentes combinaciones de estos tres tipos de información, creando así vectores de mayor dimensionalidad. Estas representaciones se utilizaron como entrada para entrenar un clasificador de máquinas de soporte vectorial (SVM) con núcleo RBF, manteniendo constantes los parámetros de entrenamiento. Asimismo, se realizó el mismo procedimiento con clasificadores de redes neuronales (MLP) y Random Forest (RF). Posteriormente, se evaluó el rendimiento de cada combinación para cada clasificador utilizando la métrica F-score sobre un conjunto de validación. Este mismo proceso se replica con Redes neuronales y Random Forest como algoritmos clasificadores. Asimismo, se realizó un experimento ampliando la base de datos mediante la transformación de las imágenes existentes, por ejemplo, aplicando rotaciones, para esta experimentación solo se evalúa con el clasificador SVM.

Así la propuesta novedosa del trabajo consiste en evaluar sistemáticamente estas combinaciones de descriptores mediante una concatenación simple, manteniendo los hiperparámetros de los clasificadores constantes para analizar el impacto relativo de cada tipo de información. A través de este enfoque, se evidencia que agregar más descriptores no siempre mejora de forma considerable el desempeño del modelo, pero sí permite identificar combinaciones particularmente efectivas.

4.2. Modelo de expertos

Como alternativa al uso de descriptores de gran dimensionalidad junto con información adicional, se propone una estrategia basada en modelos especializados para cada tipo de descriptor. En este enfoque, se entrenarán clasificadores SVM, MLP y RF por separado para cada uno de los descriptores considerados: color, forma y textura.

Cada clasificador realizará sus respectivas predicciones y, posteriormente, se evaluará qué combinación modelo-descriptor ofrece el mejor rendimiento con base en métricas de evaluación. A partir de estos resultados, se seleccionarán las mejores combinaciones, y se construirá un modelo experto mediante un promedio ponderado de dichas métricas.

Para ello, se llevarán a cabo tres experimentaciones. En las dos primeras se ajustarán los hiperparámetros de los clasificadores (SVM, RF y MLP). En la tercera, además de modificar los hiperparámetros, se ampliará la base de datos mediante la transformación de las imágenes existentes, por ejemplo, aplicando rotaciones. Cada experimentación dará lugar a un modelo, el cual será evaluado utilizando la métrica F-score. Los hiperparámetros para cada experimentación se observan respectivamente en las [Figura 2](#), [Figura 3](#) y [Figura 4](#)

✓ Experimento 1 - Hiperparámetros manejados durante toda la entrega

Parámetros:

RF: n_estimators=100, random_state=30

MLP: hidden_layer_sizes=(100,), max_iter=1000, random_state=30

SVM: random_state=42

Figure 2. Hiperparámetros para el Experimento 1

Experimento 2 - Hiperparámetros escogidos

Parámetros:

RF: n_estimators=200, max_depth=10, random_state=42

MLP: hidden_layer_sizes=(150,50), max_iter=1500, alpha=0.001, random_state=42

SVM: random_state=42, kernel='poly', degree=4, c=5, gamma='scale'

Figure 3. Hiperparametros para el Experimento 2

```
kernels = ['linear', 'rbf']
c_values = [1.0, 10.0]
```

Figure 4. Hiperparametros para el Experimento 3

5. Resultados

5.1. Descriptores

5.1.1. Unimodales

En esta sección se presentan tanto resultados cualitativos como cuantitativos. Los resultados cualitativos permiten visualizar cómo se representa cada imagen según los distintos descriptores (ver [Figura 5](#)). Por otro lado, los resultados cuantitativos nos permiten evaluar la efectividad de cada descriptor en tareas de clasificación de especies. Para ello, se utilizaron dos algoritmos de clasificación: k-nearest neighbors (*KNN*) y máquinas de soporte vectorial (*SVM*), calculando métricas como F1-score, precisión y cobertura, tanto por especie como en promedio general. A partir de estos experimentos, se compararon los distintos descriptores con el objetivo de identificar cuál ofrece el mejor rendimiento y, por lo tanto, será utilizado en el modelo final de clasificación.

Los resultados obtenidos para la experimentaciones realizadas con el algoritmo *K-means* puede observarse en la [Tabla 3](#) y para *SVM* en la [Tabla 4](#). Se condensan los resultados para cada especie, descriptor y métrica de evaluación de rendimiento. En base a los resultados obtenidos, se determinó que el mejor rendimiento se obtuvo al utilizar el descriptor de forma (HOG) en conjunto con el clasificador de máquinas de soporte vectorial (*SVM*) con kernel RBF y parámetro C=10. Este método alcanzó un promedio de F1-score de 0.66, siendo esta métrica especialmente relevante en nuestro caso al considerar el balance entre precisión y cobertura en todas las clases. Un valor de F1 promedio por encima de 0.6 puede considerarse aceptable en contextos de clasificación multiclase con clases visualmente similares, como en este caso.

Dentro de esta misma configuración, la especie que obtuvo el mayor F1-score individual fue *Phoca vitulina*, lo

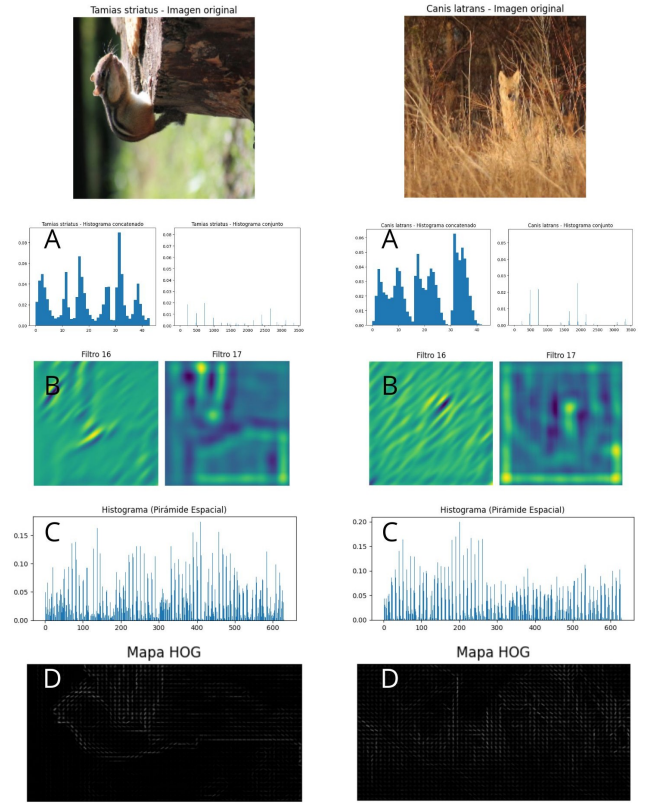


Figure 5. Ejemplo de imágenes correspondientes a dos especies de la base de datos (*Tamias striatus* y *Canis latrans*), descritas mediante los diferentes descriptores evaluados. A. Descriptores basados únicamente en color, utilizando tanto histogramas concatenados como histogramas conjuntos. B. Representación de la textura mediante la aplicación de filtros. C. Histograma generado a partir de la pirámide espacial, que conserva información de color junto con la localización espacial. D. Mapa HOG, correspondiente al descriptor de forma basado en gradientes orientados.

que sugiere que esta clase es más fácilmente reconocible mediante descriptores de forma, posiblemente debido a características visuales más consistentes o distintivas en comparación con el resto.

En contraste, el descriptor que mostró el peor desempeño fue el de textura, con un F1-score promedio cercano a 0.2 o incluso inferior, lo que indica una baja capacidad discriminativa para esta tarea en particular. Este bajo rendimiento puede deberse a la falta de variabilidad informativa en las texturas presentes entre las especies analizadas.

Otro aspecto relevante observado a lo largo de los experimentos es que el uso de *SVM* frente a *KNN* mejora sistemáticamente los resultados promedio para los distintos descriptores. En particular, el uso de kernel RBF (frente a kernel lineal) permite una mejor adaptación a la complejidad de los datos, lo cual se refleja en un mejor F1-score. Además, el aumento del parámetro C también tiene

Table 3. Comparación del rendimiento de descriptores de color y textura con el algoritmo de *k-means* para la métricas de *precisión*, *cobertura* y *F-1*. El mejor descriptor obtenido puede ser observado en negrilla y corresponde a la métrica *F-1* para el histograma conjunto de la especie *Tamias latrans* para el descriptor de color.

| | | | Especie | | | | | Promedio |
|------------|---------------------|------------------------|----------------------|----------------|----------------------------|-----------------|---------------|----------|
| Descriptor | | | Sciurus carolinensis | Phoca vitulina | Otospermophilus variegatus | Tamias striatus | Canis latrans | |
| Precisión | Color | Histograma concatenado | 0.4218750 | 0.20382166 | 0.5462963 | 0.50623053 | 0.36734694 | 0.409 |
| Cobertura | | | 0.5034965 | 0.18604651 | 0.4957983 | 0.54257095 | 0.25263158 | 0.396 |
| F-1 | | | 0.4590861 | 0.1945289 | 0.5198238 | 0.52377115 | 0.29937630 | 0.399 |
| Precisión | Histograma conjunto | 0.44064386 | 0.3030303 | 0.58247423 | 0.5106383 | 0.35123967 | 0.438 | |
| Cobertura | | 0.51048951 | 0.23255814 | 0.47478992 | 0.56093489 | 0.29824561 | 0.415 | |
| F-1 | | 0.47300216 | 0.26315789 | 0.52314815 | 0.53460621 | 0.32258065 | 0.423 | |
| Precisión | Textura | | 0.1333 | 0.3333 | 0.1429 | 0.3125 | 0.2222 | 0.2288 |
| Cobertura | | | 0.2 | 0.125 | 0.1429 | 0.3333 | 0.2 | 0.2002 |
| F-1 | | | 0.16 | 0.1818 | 0.1429 | 0.3226 | 0.2105 | 0.2036 |

Table 4. Comparación del rendimiento de descriptores de pirámide espacial y HOG con el algoritmo de *SVM* (*Máquina de soporte vectorial*) para las métricas de *precisión*, *cobertura* y *F-1*. Todos los descriptores observados corresponden a el hiperparámetro $K=RBF$. El mejor descriptor obtenido puede ser observado en negrilla y corresponde a la métrica de *F-1* para el descriptor de HOG con $C = 10$ para la especie de *Phoca vitulina*.

| | | Hiperparámetros | Especie | | | | | Promedio |
|------------|-------------------|-----------------|----------------------|----------------|----------------------------|-----------------|---------------|-------------|
| Descriptor | | C | Sciurus carolinensis | Phoca vitulina | Otospermophilus variegatus | Tamias striatus | Canis latrans | |
| Precisión | Pirámide espacial | 1 | 0.52 | 0.72 | 0.79 | 0.66 | 0.63 | 0.66 |
| Cobertura | | | 0.84 | 0.64 | 0.06 | 0.22 | 0.63 | 0.48 |
| F-1 | | | 0.65 | 0.68 | 0.12 | 0.34 | 0.63 | 0.48 |
| Precisión | Pirámide espacial | 10 | 0.68 | 0.71 | 0.62 | 0.71 | 0.68 | 0.68 |
| Cobertura | | | 0.78 | 0.70 | 0.40 | 0.59 | 0.71 | 0.64 |
| F-1 | | | 0.72 | 0.70 | 0.49 | 0.64 | 0.69 | 0.65 |
| Precisión | HOG | 1 | 0.56 | 0.69 | 0.96 | 0.79 | 0.61 | 0.72 |
| Cobertura | | | 0.82 | 0.67 | 0.16 | 0.34 | 0.66 | 0.53 |
| F-1 | | | 0.67 | 0.68 | 0.27 | 0.48 | 0.63 | 0.55 |
| Precisión | HOG | 10 | 0.64 | 0.73 | 0.82 | 0.68 | 0.68 | 0.71 |
| Cobertura | | | 0.77 | 0.79 | 0.40 | 0.52 | 0.70 | 0.64 |
| F-1 | | | 0.70 | 0.76 | 0.54 | 0.59 | 0.69 | 0.66 |

un efecto positivo sobre el rendimiento del modelo, al permitir un ajuste más estricto a los datos de entrenamiento, siempre y cuando no se incurra en sobreajuste.

5.1.2. Multimodales

Como se observa en la [Table 5](#), la combinación de los tres tipos de descriptores (color, forma y textura) obtuvo el mayor F-score (0.6972). Las combinaciones de color + textura y forma + textura fueron las mejores para los tres clasificadores. Aunque las diferencias no son drásticamente significativas, se aprecia que la peor combinación se obtuvo para Forma + Textura. En cuanto a la longitud del descriptor, se observa que los vectores más largos suelen garantizar un mayor rendimiento aunque no tan significativamente respecto a las combinaciones de Color + Forma, esto sugiere

que una representación más completa puede aportar beneficios si los tipos de información son complementarios. Respecto a la clasificación por clase, se observó que las clases con diferencias visuales más marcadas en forma y color fueron mejor clasificadas por el modelo final. No obstante, esto puede variar ligeramente según la implementación y la máquina utilizada, ya que el rendimiento de la baseline mostró una oscilación entre 0.62 y 0.68, según el procesador del computador en que se corre el presente proyecto.

Por otro lado, en la [Table 6](#) se observa la experimentación realizada con la base de datos aumentada. Aunque esta solo fue posible realizarla con el clasificador SVM por cuestiones de tiempo, se evidencia una mejora significativa frente a la experimentación realizada sin el aumento. La mejor métrica obtenida también fue obtenido a

Table 5. F-score obtenido para diferentes combinaciones de descriptores utilizando como clasificadores SVM, MLP y RF. La mejor combinación de descriptores corresponde a el uso de los tres descriptores para los tres clasificadores, resaltando el mejor F-score para el clasificador SVM. En negrilla se resaltan las mejores métricas obtenidas, las cuales corresponden a la combinación de los tres descriptores para los tres clasificadores.

| Clasificador | Hiperparámetros | Combinaciones | | | |
|--------------|--|---------------|-----------------|-----------------|-------------------------|
| | | Color + Forma | Color + Textura | Forma + Textura | Color + Forma + Textura |
| SVM | kernel = "rbf" C = 10.0 | 0.6943 | 0.6577 | 0.6558 | 0.6972 |
| MLP | Hidden layer sizes = (100,) max_iteraciones = 1000 random_state = 30 | 0.6347 | 0.6312 | 0.5865 | 0.6542 |
| RF | n_estimators = 100 random_state = 30 | 0.6698 | 0.6685 | 0.6176 | 0.6762 |

Table 6. F-score obtenido con la base de datos ampliada para diferentes combinaciones de descriptores utilizando el clasificador SVM. En negrilla se resalta la mejor métrica obtenida

| Clasificador | Hiperparámetros | Combinaciones | | | |
|--------------|----------------------------|---------------|-----------------|-----------------|-------------------------|
| | | Color + Forma | Color + Textura | Forma + Textura | Color + Forma + Textura |
| SVM | kernel = "rbf" C = 10.0 | 0.7505 | 0.6507 | 0.6449 | 0.7555 |

través de la combinación de los tres descriptores.

5.2. Modelo de expertos

En esta sección se presentan los resultados obtenidos a partir de las tres etapas de experimentación, en las que se evaluaron clasificadores SVM, MLP y RF aplicados de forma independiente a los descriptores de color, forma y textura y posteriormente, se construyó un modelo experto mediante un promedio ponderado de las mejores combinaciones obtenidas. Observe en la [Tabla 7](#) para el primer experimento, para el segundo y [Tabla 8](#) y la [Tabla 9](#) para el último.

El mejor resultado general se obtuvo en la primera experimentación, donde el modelo experto construido a partir del clasificador Random Forest (RF) alcanzó un F-score de 0.6735. Esta combinación superó al resto de las configuraciones evaluadas, tanto en el uso individual de descriptores como en el modelo experto. Además, en esta misma experimentación, RF también logró los mejores resultados individuales para los tres descriptores, destacando especialmente en el descriptor de color con un F-score de 0.6609 y en textura con 0.6552.

En cuanto a la influencia de los hiperparámetros, se identificó que ciertas configuraciones afectaron significativamente el rendimiento. En el caso del MLP, el ajuste del tamaño de las capas ocultas y el parámetro de regularización (alpha) fue determinante para mejorar los resultados, como se evidenció en la segunda experimentación. De forma similar, en RF, el número de árboles (n_estimators) y la profundidad máxima (max_depth) fueron factores que impactaron directamente en el desempeño del clasificador.

No todas las combinaciones mostraron un rendimiento adecuado. En particular, el uso de SVM sobre el descriptor de textura en la primera experimentación resultó en un F-score de apenas 0.4147, lo cual representa el peor desempeño registrado entre todas las combinaciones. Esto sugiere que este clasificador tiene limitaciones para modelar correctamente las características asociadas a la textura en este contexto.

Asimismo, se resalta que la ampliación de la base de datos permite obtener un mejor resultado con un F-score de 0.7254, sin embargo mejor identificador de color para C = 10 y kernel rbf.

5.3. Método final

Para el desarrollo del modelo final, se implementó una estrategia de aumento de datos sobre el conjunto original de imágenes, con el fin de mejorar la generalización y robustez del clasificador ante variaciones comunes como rotaciones, cambios de escala y ruido visual (figura 6). Este conjunto aumentado fue utilizado para entrenar un descriptor multimodal que combina tres tipos de información visual: color, forma y textura.

El modelo sigue una secuencia de procesamiento estructurada que comienza con la normalización de la imagen de entrada. Cada imagen es redimensionada y convertida a formato RGB si es necesario. Luego, se extraen tres descriptores de manera independiente: un histograma piramidal de color, un descriptor HOG (Histogram of Oriented Gradients) para capturar la forma, y un histograma de textura basado en la respuesta a un banco de filtros aplicado a la im-

Table 7. F-score obtenido en el primer experimento para los descriptores de color, forma y textura para los clasificadores de *SVM*, *MLP*, *RF*. La mejor combinación de modelo experto corresponde al clasificador *RF*, siendo a su vez este el mejor clasificador para los tres descriptores. En negrilla se resaltan las mejores métricas obtenidas para cada clasificador.

| Clasificador | Hiperparámetros | Descriptores | | | Modelo experto |
|--------------|--|---------------|---------------|---------|----------------|
| | | Color | Forma | Textura | |
| SVM | random_state = 42 | 0.4809 | 0.5638 | 0.4147 | 0.6387 |
| MLP | Hidden layer sizes = (100,) max_iteraciones = 1000 random_state = 30 | 0.6205 | 0.5897 | 0.4193 | 0.6572 |
| RF | n_estimators = 100 random_state = 30 | 0.6609 | 0.6093 | 0.6552 | 0.6735 |

Table 8. F-score obtenido en el segundo experimento para los descriptores de color, forma y textura para los clasificadores de *SVM*, *MLP*, *RF*. La mejor combinación de modelo experto corresponde al clasificador *MLP*. En negrilla se resaltan las mejores métricas obtenidas para cada clasificador, mientras que en cursiva se destaca la mejor métrica obtenida para cada descriptor.

| Clasificador | Hiperparámetros | Descriptores | | | Modelo experto |
|--------------|--|---------------|---------------|---------------|----------------|
| | | Color | Forma | Textura | |
| SVM | random_state = 42 kernel = "poly" degree = 4 C = 5 gamma = "scale" | 0.4763 | 0.5971 | 0.4608 | 0.6396 |
| MLP | Hidden layer sizes = (150, 50) max_iteraciones = 1500 random_state = 42 alpha = 0.001 | 0.6193 | 0.6033 | 0.4208 | 0.6685 |
| RF | n_estimators = 200 max_depth = 10 random_state = 42 | 0.5522 | 0.4865 | <i>0.5367</i> | 0.5797 |

Table 9. F-score obtenida con la base de datos ampliada para los tres descriptores con clasificación SVM.

| Clasificador | Hiperparámetros | Descriptores | | | 2*Modelo experto |
|--------------|----------------------------|--------------|--------|---------|------------------|
| | | Color | Forma | Textura | |
| SVM | C = 10 kernel = "rbf" | 0.7149 | 0.6482 | 0.6128 | 0.7254 |
| SVM | C = 1 kernel = "linear" | 0.6922 | 0.5827 | 0.5244 | 0.6255 |

agen. Estos vectores de características son posteriormente concatenados para formar un único descriptor multimodal de alta dimensión.

Finalmente, este vector combinado se introduce en un clasificador previamente entrenado (el mejor modelo identificado en los experimentos previos, con SV), que predice la clase correspondiente a la imagen. El funcionamiento general del modelo puede verse resumido en la Figura de Overview, que ilustra las etapas de entrada, procesamiento de descriptores y salida de predicción. Obteniendo las métricas de la [Tabla 10](#)

Table 10. Métricas de evaluación del modelo final

| Métrica | Cobertura | Precisión | F-score |
|--------------|-----------|-----------|---------|
| Valor | 0.72146 | 0.83445 | 0.758 |

6. Discusión

6.1. Descriptores

6.1.1. Unimodales

El modelo baseline corresponde al uso del descriptor HOG aplicado a la especie *Phoca vitulina*. Este resultado es co-

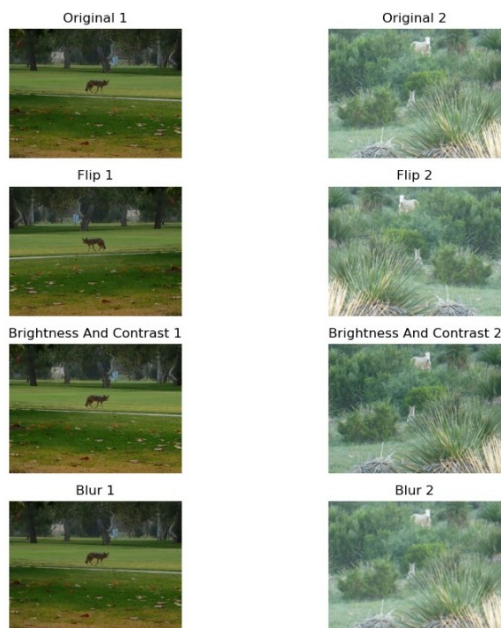


Figure 6. Ejemplo visual de algunas transformaciones realizadas a las imágenes de la base de datos

herente, ya que HOG se centra en capturar las formas y contornos presentes en una imagen, y *Phoca vitulina* posee una silueta claramente distinta al resto de especies evaluadas. Su forma ovalada y compacta contrasta con los cuerpos más alargados, con cola peluda y orejas erguidas, que presentan especies como *Canis latrans* o *Tamias striatus*. Esta diferencia morfológica hace que los gradientes de orientación capturados por HOG sean especialmente efectivos para esta especie en particular.

En cuanto a las métricas obtenidas, los resultados cuantitativos muestran una mejora significativa al utilizar el algoritmo *SVM* en lugar de *k-means*. Mientras que *k-means* arrojó valores de F1-score promedio entre 0.20 y 0.44, el uso de *SVM* permitió alcanzar resultados en un rango superior (0.48–0.72), dependiendo del descriptor utilizado. Esta diferencia se debe a que *SVM* es un clasificador más flexible, capaz de manejar conjuntos de datos no linealmente separables y de ajustar su comportamiento mediante hiperparámetros clave, como el tipo de kernel y el parámetro de penalización (C). Por el contrario, *k-means* es un algoritmo no supervisado con menor capacidad de adaptación, limitado principalmente por el número de clústeres definidos. No obstante, *SVM* también implica un riesgo mayor de sobreajuste si no se realiza una selección cuidadosa de los hiperparámetros.

A lo largo de los experimentos, fue posible contrastar el desempeño de varios descriptores junto con diferentes clasificadores. En términos generales, el descriptor de forma (HOG) combinado con *SVM* fue el que ofreció mejores re-

sultados, tanto a nivel promedio como por clase individual. Aunque podría parecer que el descriptor por sí solo explica el rendimiento, nuestros resultados indican que la elección del clasificador tiene un peso igualmente importante. En particular, *SVM* superó sistemáticamente al clasificador *KNN* en todas las combinaciones probadas.

Este comportamiento destaca una idea central: el desempeño del modelo depende de la interacción entre el descriptor y el clasificador. Un descriptor pobre, que no capture adecuadamente las diferencias entre clases, limitará el rendimiento sin importar qué clasificador se utilice. De igual manera, un clasificador simple o mal ajustado no podrá explotar todo el potencial de un descriptor informativo. Por ello, ni el descriptor ni el clasificador deben evaluarse de forma aislada; es el ajuste conjunto entre ambos lo que determina la calidad del modelo final.

Si bien el uso de HOG con *SVM* produjo resultados sólidos, también es importante explorar estrategias más integradoras. Evaluar descriptores por separado aporta claridad analítica, pero en escenarios reales es recomendable combinar múltiples fuentes de información. La integración de características como forma, color, textura y distribución espacial podría construir descriptores más robustos, capaces de compensar las debilidades individuales. Por ejemplo, aunque la textura tuvo un bajo rendimiento cuando se utilizó por sí sola, podría aportar información valiosa en combinación con descriptores de forma o color. Este enfoque multimodal podría mejorar la clasificación especialmente en especies visualmente similares, donde ningún descriptor aislado logra representar todas sus diferencias relevantes.

Finalmente, al observar las imágenes mal clasificadas, se identifican ciertos patrones: muchas de ellas comparten poses similares, fondos visualmente complejos o condiciones de iluminación variables. Estas características dificultan la extracción de información discriminativa, lo que sugiere que mejorar el preprocesamiento de imágenes o incorporar mecanismos de atención espacial también podrían ser líneas prometedoras para fortalecer el modelo.

6.1.2. Multimodales

Los resultados obtenidos indican que la combinación de descriptores visuales ofrece un beneficio moderado pero consistente en el rendimiento del modelo. El F-score más alto (0.6972) se logró al combinar color, forma (HOG) y textura, lo que sugiere que estos tres tipos de información aportan características complementarias. Cada descriptor captura un aspecto distinto: el color distingue clases con paletas específicas, la forma aporta estructura útil en siluetas bien definidas, y la textura agrega detalles finos relevantes en objetos complejos. Sin embargo, la mejora respecto a combinaciones dobles (como color + forma) no fue drástica, lo que sugiere que ciertos descriptores, como la textura, pueden aportar información redundante o poco relevante según la especie.

En el contexto de las especies asignadas a nuestro grupo, se evidenció que algunos descriptores no aportaron mejoras significativas respecto al rendimiento del baseline. Además, la considerable variabilidad intraclasal observada en ciertas categorías —como en el caso de las dos especies de ardillas morfológicamente similares— dificultó la identificación de patrones consistentes por parte de los descriptores. En estos casos, el modelo tendió a apoyarse en atributos menos informativos, como el color en imágenes con baja saturación, lo que incrementó los errores de clasificación. Otro caso de nuestras especies, es la foca, que fue más fácil de distinguir gracias a características de color y forma, que son muy discriminantes entre las demás especies asignadas (ardillas, lobos y zorros). Esto refuerza la idea de que la efectividad de un descriptor no reside únicamente en la cantidad de información añadida, sino en su capacidad para complementar otros y su relevancia específica para las clases que se buscan clasificar. Finalmente, se evidencia que el amplificar la base de datos brinda métricas significativamente superiores.

6.2. Modelo de expertos

Al analizar los resultados obtenidos con los distintos clasificadores, se observa que no todos lograron superar el valor del baseline. En el primer experimento, únicamente el modelo experto del clasificador Random Forest (RF) logró superar el baseline, obteniendo un f-score de 0.6735 y 0.6657, respectivamente. En contraste, en el segundo experimento, solo el modelo experto del clasificador Multilayer Perceptron (MLP) logró mejorar el rendimiento base, con un f-score de 0.6685. Estos resultados reflejan una alta variabilidad en el desempeño de los clasificadores dependiendo de la configuración de sus hiperparámetros, lo que impide generalizar su comportamiento. Respecto a los descriptores utilizados, se evidenció que su rendimiento no fue consistente entre los distintos clasificadores. Esto sugiere que la efectividad de los descriptores está fuertemente influenciada por la arquitectura del modelo y su capacidad para capturar patrones relevantes en los datos. Asimismo, se resalta que la amplificación de la base de datos resulta significativamente útil para entrenar el modelo.

6.3. Método final

El modelo final alcanzó una cobertura (recall) de 0.72146, una precisión de 0.83445 y un F-score de 0.758. Esto indica que el modelo es más preciso que exhaustivo: tiende a acertar cuando clasifica, pero aún deja escapar algunos casos verdaderos. En aplicaciones del mundo real, esto podría ser deseable en contextos donde los falsos positivos tienen un mayor costo que los falsos negativos (por ejemplo, en sistemas de detección de especies protegidas o amenazas específicas), pero puede no ser ideal en tareas donde la detección completa es crítica.

Entre las principales limitaciones del modelo se encuen-

tra la alta variabilidad intraclasal de ciertas categorías, lo que dificulta encontrar patrones comunes, así como la posible redundancia de algunos descriptores en clases con información visual poco diferenciada. Además, la dependencia de procesamiento manual para la extracción de descriptores limita su escalabilidad en entornos con grandes volúmenes de datos. Como posibles mejoras, se plantea la integración de técnicas de selección de características para reducir dimensionalidad y ruido, y la incorporación de modelos más avanzados como redes convolucionales, que puedan aprender representaciones más ricas directamente desde las imágenes sin necesidad de descriptores manuales.

7. Conclusiones

Finalmente, el presente trabajo demostró la viabilidad y efectividad del uso de técnicas de aprendizaje automático para la clasificación automática de especies de mamíferos a partir de imágenes. En este sentido, la exploración de los distintos descriptores visuales combinados con clasificadores como SVM, Random Forest y redes neuronales, permitió identificar algoritmos eficaces para esta tarea. Asimismo, se destaca la combinación del descriptor de HOG con una máquina de soporte vectorial con kernel RBF y parámetro de penalización de $C = 10$ ofreció resultados que se destacan, alcanzando un F1-score promediado de 0.66. Adicionalmente, se evidenció y se señaló que la combinación de múltiples descriptores aporta un valor adicional al modelo para su clasificación, a causa de que integran características complementarias. Esta estrategia alcanzó su mejor rendimiento tras aumentar la base de datos con nuevas imágenes, lo que, en consecuencia, se obtuvo un F1-score de 0.7555. Por otra parte, los modelos expertos permitieron evaluar el rendimiento de los clasificadores entrenados individualmente para cada tipo de descriptor. En decir, el modelo basado en Random Forest se destacó en la primera etapa experimental, superando incluso al modelo base. No obstante, posterior al aumento de la base de datos los resultados de los modelos expertos no se destacaron a causa de que no alcanzaron un valor mayor a los multimodales. En síntesis, la interacción entre el tipo de descriptor, el clasificador y los parámetros empleados tiene un impacto determinante en la calidad de la clasificación. Aunque los resultados obtenidos son prometedores, persisten desafíos como la variabilidad intraespecífica, las condiciones ambientales adversas y la similitud visual entre ciertas especies. Estas limitaciones abren nuevas líneas de investigación, como la incorporación de técnicas más avanzadas de preprocesamiento, el uso de arquitecturas profundas, o la integración de mecanismos de atención visual.

References

- [1] Miao, Z., Gaynor, K. M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M. S., McInturff, A., Bowie, R. C. K., Nathan, R., Yu, S. X., & Getz, W. M. (2019). Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-44565-w> 1
- [2] Hamard, Q., Pham, M., Cazau, D., & Heerah, K. (2024). A deep learning model for detecting and classifying multiple marine mammal species from passive acoustic data. *Ecological Informatics*, 84, 102906. <https://doi.org/10.1016/j.ecoinf.2024.102906> 1
- [3] J. Lenzi, A. F. Barnas, A. A. ElSaid, T. Desell, R. F. Rockwell, and S. N. Ellis-Felege, “Artificial intelligence for automated detection of large mammals creates path to upscale drone surveys,” *Scientific Reports*, vol. 13, no. 1, Jan. 2023, doi: 10.1038/s41598-023-28240-9. 1
- [4] Z. Xu, T. Wang, A. K. Skidmore, and R. Lamprey, “A review of deep learning techniques for detecting animals in aerial and satellite images,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103732, Mar. 2024, doi: 10.1016/j.jag.2024.103732. 1
- [5] M. S. Norouzzadeh et al., “Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, Jun. 2018, doi: 10.1073/pnas.1719367115. 1
- [6] B. M. Cristian, “Detección y clasificación automática de animales silvestres mediante visión artificial y machine learning,” 2022. 1, 2
- [7] V. Manuel et al., “Capacidad predictiva de las Máquinas de Soporte Vectorial. Una aplicación en la planificación financiera,” *Revista Cubana de Ciencias Informáticas*, vol. 13, no. 3, pp. 59–75, 2019, Accessed: Apr. 07, 2025. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992019000300059 1
- [8] S. Dhanabal and S. Chandramathi, “A review of various knearest neighbor query processing techniques,” *International Journal of Computer Applications*, vol. 31, no. 7, pp. 14–22, 2011. 1
- [9] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: a Brief Review,” *Computational Intelligence and Neuroscience*, vol. 2018, no. 1, pp. 1–13, Feb. 2018, doi: <https://doi.org/10.1155/2018/7068349>. 2
- [10] Cornell Lab of Ornithology, “eBird - Discover a new world of birding,” *Ebird.org*, 2019. <https://ebird.org/home> 2
- [11] iNaturalist, “iNaturalist.org,” *iNaturalist.org*, 2019. <https://www.inaturalist.org/> 2
- [12] Species 360, “Home,” *Species360*. <https://species360.org/> 2
- [13] P. Myers, R. Espinoza, C. S. Parr, T. Jones, S. Hammond, and T. A. Dewey, “ADW: Home,” *Animaldiversity.org*, 2019. <https://animaldiversity.org/> 2
- [14] L. Cristina and S. Nunes, “Intelligent retrieval and classification in three-dimensional biomedical images — A systematic mapping,” *Computer Science Review*, vol. 31, pp. 19–38, Feb. 2019, doi: <https://doi.org/10.1016/j.cosrev.2018.10.003>. 3

Anexos

Estadísticas de Imágenes para Train

| Especie | Escala de Grises | Color | Tamaño Promedio | Horizontal | Vertical | Cuadrada |
|----------------------------|------------------|-------|-----------------|------------|----------|----------|
| Canis latrans | 20 | 1266 | 734.5 × 591.0 | 1018 | 213 | 55 |
| Otospermophilus variegatus | 0 | 515 | 656.7 × 532.8 | 402 | 108 | 5 |
| Phoca vitulina | 0 | 713 | 720.8 × 529.5 | 644 | 48 | 21 |
| Sciurus carolinensis | 0 | 1795 | 700.4 × 637.2 | 1094 | 650 | 51 |
| Tamias striatus | 0 | 854 | 692.9 × 578.0 | 623 | 211 | 20 |

Estadísticas de Imágenes para Validation

| Especie | Escala de Grises | Color | Tamaño Promedio | Horizontal | Vertical | Cuadrada |
|----------------------------|------------------|-------|-----------------|------------|----------|----------|
| Canis latrans | 9 | 420 | 736.3 × 592.2 | 343 | 71 | 15 |
| Otospermophilus variegatus | 0 | 172 | 661.5 × 543.0 | 131 | 38 | 3 |
| Phoca vitulina | 1 | 237 | 696.2 × 512.1 | 211 | 23 | 4 |
| Sciurus carolinensis | 0 | 599 | 703.2 × 642.6 | 359 | 223 | 17 |
| Tamias striatus | 0 | 285 | 687.8 × 574.7 | 204 | 71 | 10 |

Estadísticas de Imágenes para Test

| Especie | Escala de Grises | Color | Tamaño Promedio | Horizontal | Vertical | Cuadrada |
|----------------------------|------------------|-------|-----------------|------------|----------|----------|
| Canis latrans | 14 | 415 | 728.0 × 581.5 | 350 | 66 | 13 |
| Otospermophilus variegatus | 0 | 173 | 647.7 × 536.6 | 152 | 21 | 0 |
| Phoca vitulina | 1 | 237 | 719.4 × 531.3 | 209 | 29 | 0 |
| Sciurus carolinensis | 0 | 599 | 699.5 × 641.8 | 351 | 228 | 20 |
| Tamias striatus | 0 | 285 | 686.7 × 582.2 | 197 | 76 | 12 |

Figure 7. Tabla de características comparativas de imágenes para cada especie en cada carpeta de *train*, *valid* y *test*.