

# CONSERVACIÓN INTELIGENTE

## Descripción Metodológica

### Herramientas a Utilizar

Para el desarrollo del proyecto se utilizarán diversas herramientas, plataformas y técnicas que permitirán la construcción y aplicación de los modelos de clasificación previamente descritos.

La base de imágenes fue proporcionada mediante un disco rígido con un tamaño de aproximadamente 550 GB, la herramienta resultante permitirá que el cliente continúe con este método de almacenamiento para no incurrir en nuevos costos. Con efectos de tener mayor accesibilidad durante el trabajo las capturas y la base serán almacenadas en la nube de *Amazon Web Services* utilizando el servicio de buckets S3. Es importante aclarar que esto fue realizado para el análisis y el desarrollo del presente proyecto con el objetivo de almacenar una copia de las imágenes recibidas, el trabajo no contempla la migración de los datos de *Rewilding* a una plataforma de Cloud.

Tanto para el acceso a *AWS* como para el desarrollo del programa se optó por el lenguaje de programación *Python*, dada su amplia comunidad, las librerías disponibles especializadas para los fines de interés y los frameworks de Aprendizaje Profundo como *Pytorch* y *TensorFlow*.

En cuanto a las técnicas específicas para la construcción del modelo, se evaluarán aquellas referidas a:

- **Redes Neuronales Convolucionales (CNN):** Técnica más comúnmente utilizada para la clasificación de imágenes. Estas redes están diseñadas específicamente para procesar datos de imágenes y son capaces de aprender jerarquías de características a diferentes niveles de abstracción. Las CNN han demostrado un rendimiento excepcional en una amplia variedad de problemas de clasificación de imágenes.
- **Data Augmentation:** Implica crear versiones ligeramente modificadas de las imágenes de entrenamiento mediante rotaciones, reflejos, recortes y otros cambios. Es de utilidad para expandir la cantidad de imágenes sin animales que están etiquetadas, dado que la mayoría de esta índole fueron eliminadas.
- **Upsampling/Downsampling:** Dado el desbalance del dataset de entrenamiento entre las imágenes que efectivamente poseen animales y aquellas que no, se deberá utilizar alguna de estas técnicas para el correcto entrenamiento.
- **Transfer Learning:** La transferencia de aprendizaje implica tomar una red neuronal pre-entrenada en un conjunto de datos grande y ajustarla para un problema específico. Para el tipo de problema en cuestión, se suele utilizar modelos entrenados en base a ImageNet, una vasta base de datos de imágenes que abarca múltiples categorías. ImageNet ha sido fundamental en los avances de la visión por computadora al proporcionar un conjunto de datos estándar para entrenar modelos profundos. Al aprovechar los pesos y características aprendidas en ImageNet, se puede acelerar y mejorar el entrenamiento para tareas específicas, reduciendo la necesidad de grandes cantidades de datos propios.
- **Arquitecturas Conocidas:** En el desarrollo del modelo se seleccionarán arquitecturas de redes neuronales avanzadas como VGG16, ResNet e InceptionV3, apreciadas por su eficiencia y precisión en la clasificación de imágenes. Estas arquitecturas se adaptarán cuidadosamente a las necesidades del proyecto, considerando la naturaleza y complejidad de los datos. La decisión sobre que arquitectura utilizar se basará en pruebas de rendimiento y su compatibilidad con la infraestructura de procesamiento de datos existente.
- **Uso de Modelos Externos:** Además de las técnicas mencionadas, se contempla la integración de modelos externos pre-entrenados para alguna tarea específica, como el modelo *MegaDetector* de Microsoft, que han demostrado ser eficientes en la detección y clasificación de animales en imágenes de trampas-cámara. Estos modelos, al estar ya entrenados en grandes conjuntos de datos, pueden ofrecer una base sólida y ahorrar tiempo y recursos en el proceso de entrenamiento, si bien la robustez de los mismos sugiere que se puede lograr más precisión con un entrenamiento especializado a un bioma específico.

### Metodologías a Implementar

En el desarrollo del proyecto, se exploraron tres arquitecturas de vanguardia en el ámbito de las redes neuronales profundas: VGG16, InceptionV3 y ResNet.

- **VGG16:** Desarrollada por el Visual Graphics Group de la Universidad de Oxford, esta arquitectura se destaca por su simplicidad, utilizando exclusivamente capas convolucionales de 3x3 seguidas de capas de pooling para reducción de dimensionalidad. A pesar de su estructura aparentemente simple, VGG16 consiste en 16 capas y ha demostrado una eficacia excepcional en la clasificación y localización de imágenes.
- **InceptionV3:** Originaria de Google, la arquitectura Inception es conocida por su módulo "Inception", que permite a la red realizar convoluciones a diferentes escalas simultáneamente. Esta característica permite a InceptionV3 captar patrones en diferentes niveles de granularidad y ha mostrado ser particularmente eficiente en la gestión de recursos computacionales.

- **ResNet:** Desarrollada por Microsoft Research, ResNet introdujo la noción de "conexiones residuales". Estas conexiones permiten que las activaciones salten algunas capas y se sumen a las activaciones de capas posteriores, facilitando el entrenamiento de redes muy profundas al abordar el problema de desvanecimiento del gradiente. ResNet ha establecido varios récords en precisión de clasificación, siendo una elección preferida para tareas que requieren una profundidad considerable.

Para maximizar el potencial de estas arquitecturas en el contexto del proyecto, se implementó una estrategia de transferencia de aprendizaje utilizando modelos previamente entrenados en el dataset ImageNet, un vasto repositorio de imágenes etiquetadas que ha sido pilar en el avance del aprendizaje profundo en la visión por computadora. La idea detrás de este enfoque es beneficiarse de las características genéricas aprendidas por las redes en conjuntos de datos extensos y adaptarlas a las particularidades del dataset específico del proyecto. Esta combinación busca capitalizar la rica representación de ImageNet con la especificidad del conjunto de datos en estudio, con el objetivo de alcanzar una performance de clasificación óptima.

**Seteo de Experimentación**

Una vez los datos han sido depurados, estandarizados, validados y están en condiciones para su manipulación, se inicia la fase experimental del proyecto. Se cuenta con una base de datos etiquetada compuesta por 6049 registros verificables, aptos para su utilización en las fases de entrenamiento y validación.

En el contexto de aprendizaje automático y, en particular, en el ámbito del procesamiento de imágenes, se puede representar imágenes con tensores a través de sus dimensiones: alto, ancho y canales de color. Por ejemplo, una imagen a color de tamaño 150x150 píxeles se podría representar con un tensor de dimensiones y tres canales de color: rojo, verde y azul. Para la implementación y gestión eficiente del proyecto, la base de datos consolidada contendrá la ruta a la imagen, el tensor generado y las etiquetas respectivas de la imagen.

Dicha base es segmentada en conjuntos de entrenamiento y prueba, en una proporción del 80% y 20%, respectivamente. Dado que existe desigualdad en todos los casos de predicción, se implementarán mecanismos de balanceo en la separación en subconjuntos para asegurar una distribución equitativa de los datos durante la segmentación. Este mecanismo es conocido como Stratified Split sobre la clase objetivo. Además, esta partición se realiza con una semilla para que todos los modelos se entrenen y hagan pruebas sobre los mismos datos y que la aleatoriedad no favorezca a ninguno.

El conjunto de prueba se preservará inalterado hasta la conclusión de la fase experimental, momento en el que se utilizará por única vez para evaluar el rendimiento del modelo. Este procedimiento asegura que dicho conjunto refleje de manera fidedigna la realidad, al no estar sujeto a modificaciones ni haber sido empleado durante el entrenamiento.

En situaciones donde el conjunto de datos de entrenamiento exhibe un desbalance significativo entre las clases, se emplearán estrategias de Data Augmentation con el propósito de potenciar la capacidad de generalización del modelo y equilibrar la distribución de las clases. Estas técnicas involucran la aplicación de transformaciones sutiles a las imágenes, incluyendo ajustes en luminosidad y contraste, rotaciones, introducción de perturbaciones aleatorias y reflexiones. El objetivo de estas transformaciones es enriquecer el conjunto de datos con variantes únicas, previniendo así el overfitting al evitar la simple duplicación de ejemplares existentes.

En cuanto al conjunto de entrenamiento, este se subdivide en datos para entrenamiento y para validación, aplicando técnicas de Cross Validation. Con el objetivo de optimizar la capacidad de generalización del modelo, evitando el sobreajuste a los datos de entrenamiento, mediante la selección correcta de hiper-parámetros.

Las métricas utilizadas para la evaluación de los modelos de clasificación mencionados son:

- **Accuracy:** Proporción de predicciones correctas sobre el total de predicciones.

Es una métrica fundamental para evaluar el rendimiento general del modelo. La clasificación se considerará satisfactoria de lograrse una exactitud mayor a la precisión en la tarea presentada por la situación actual, la cual es la exactitud esperada para un humano.
- **Recall:** Con el objetivo de medir la proporción de ejemplos positivos que el modelo ha identificado correctamente.

En proyectos de conservación es crucial asegurar que no se omitan avistamientos de especies, lo cual es fundamental para la identificación precisa de falsos negativos. Dada la naturaleza imperativa de la preservación de la biodiversidad, se da prioridad al recall para maximizar la detección de individuos y garantizar que ningún espécimen relevante sea inadvertido en el análisis.
- **Precisión:** Esta métrica se centra en los Falsos Positivos. La precisión mide la proporción de identificaciones correctas frente al total de identificaciones positivas realizadas por el modelo. Una precisión alta significa que hay una menor proporción de Falsos Positivos.

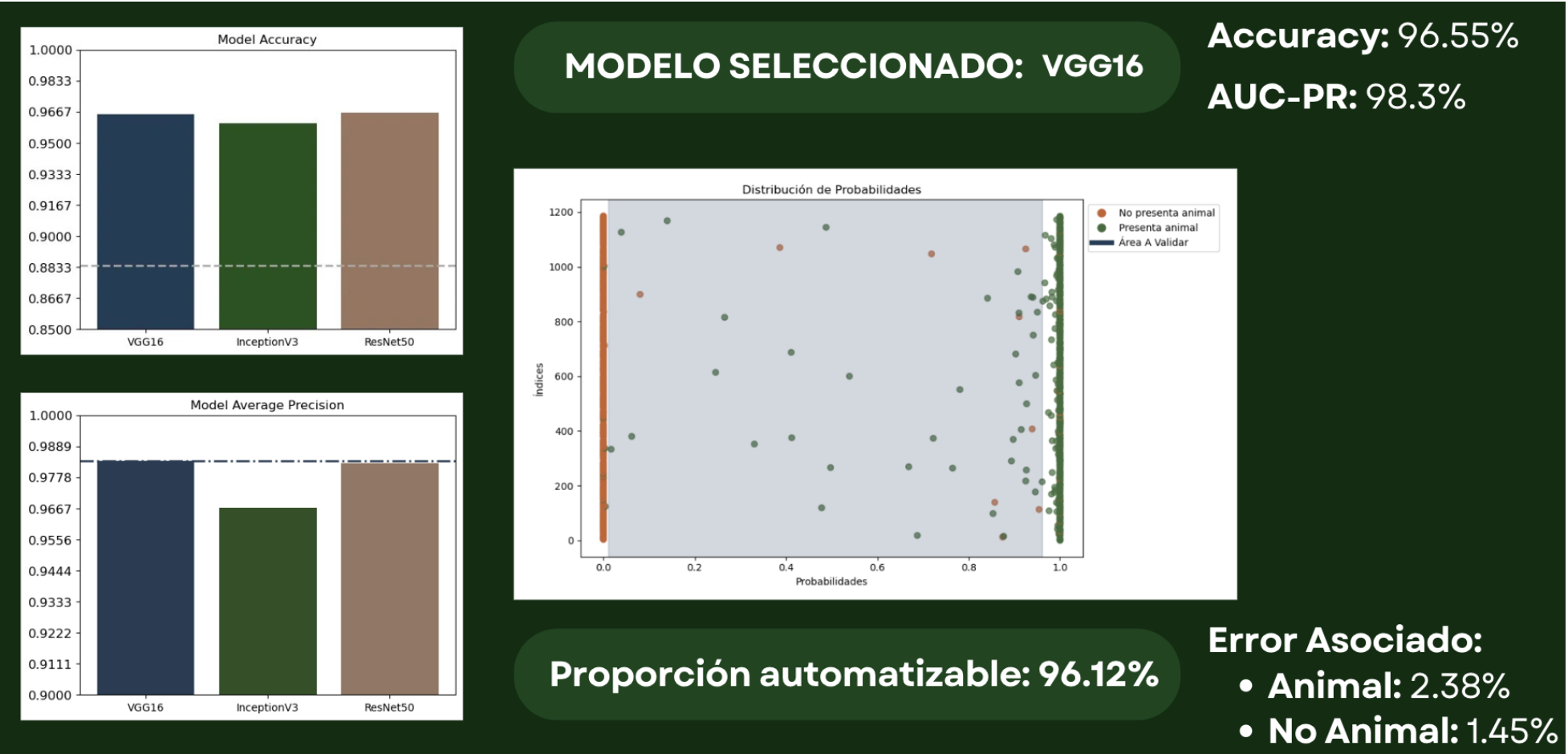
Al priorizar la precisión, se minimiza la incidencia de falsos positivos, lo cual es esencial para evitar el desvío de recursos de conservación hacia especies no objetivo.
- **Average Precision (AUC-PR):** Para obtener una visión holística del comportamiento del modelo a lo largo de diferentes umbrales de decisión, se estudiará la curva Precisión-Recall y, en particular, el área bajo esta curva (AUC-PR). Un valor de AUC-PR más alto indica un rendimiento general superior en la clasificación, teniendo en cuenta tanto la Precisión como el Recall en cada umbral.

En resumen, los modelos se evalúan maximizando el Average Precision dado un accuracy mayor o igual a la precisión humana identificada por la tarea. Así, los modelos buscarán maximizar su poder de clasificación siendo capaces de mejorar la exactitud de un humano. Finalmente, el modelo elegido será aquel que logre el máximo Average Precision.

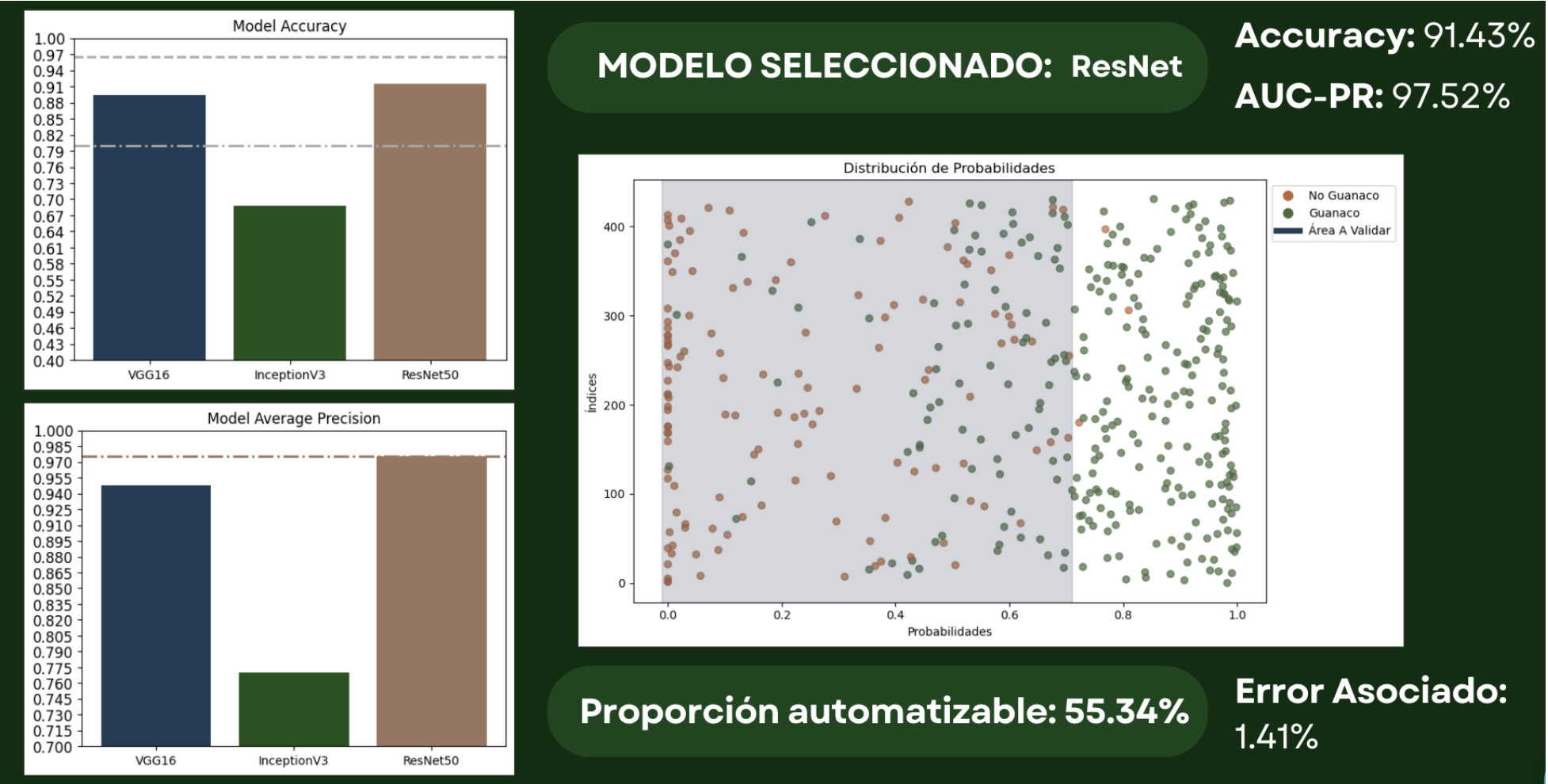
Siendo  $P_t$  la precisión humana de la tarea  $t$ . Los humanos son capaces de identificar una presencia animal en su 96.6%. La tarea de detección de especies tiene un accuracy humano del 96.6% y de cantidades es el 90%.

Además, las métricas de Recall y Precisión serán utilizadas para definir los thresholds de validación. Es decir, la confianza en la predicción necesaria para llegar a valores máximos en Recall o Specificity, dependiendo el caso. Esta minimización se ve limitada por el costo beneficio que implica validar

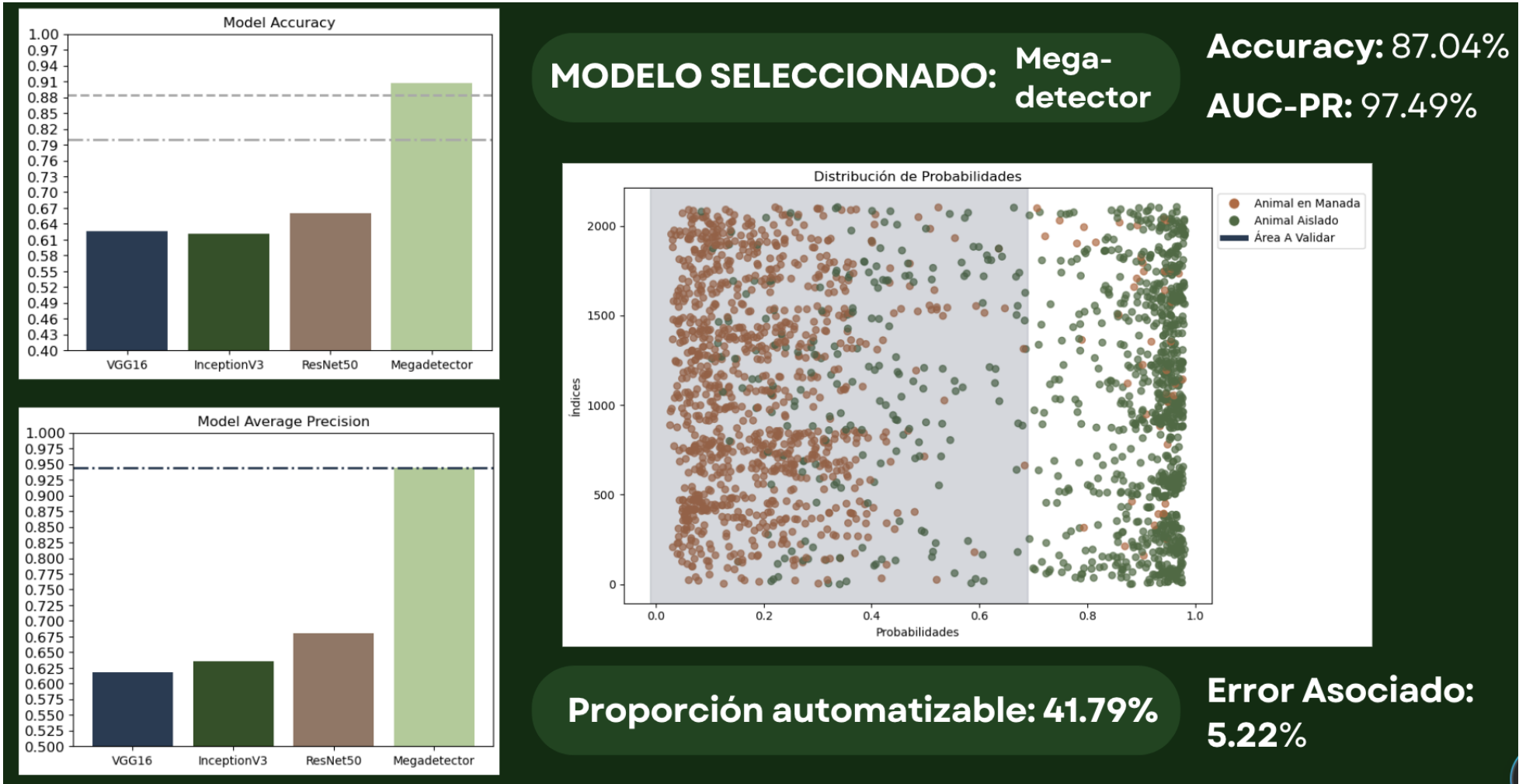
Tarea 1: Detección de Presencias



Tarea 2: Detección de Especies



Tarea 3: Recuento de Cantidades



## Contacto

Para más información sobre el proyecto se pone a disposición el [Repositorio de GitHub](#) utilizado.

Abril Noguera	Lucas Arbues	Ignacio De Achaval
<div>anoguera@itba.edu.ar</div> <div><a href="#">LinkedIn</a></div> <div><a href="#">Github</a></div>	<div>larbues@itba.edu.ar</div> <div><a href="#">LinkedIn</a></div> <div><a href="#">Github</a></div>	<div>ideachaval@itba.edu.ar</div> <div><a href="#">LinkedIn</a></div> <div><a href="#">Github</a></div>

