Colombian wildlife classification

# Authors

Henry Gardner, Juliana Gómez Consuegra, Jacqueline Lam and Eunice Ngai

Emails: henrygardner, julianagc, jacq, eunicen @berkeley.edu

Feel free to contact the authors if you have any questions or comments.

# Abstract

This paper presents a computer vision approach to classifying Colombian wildlife through camera traps in the Orinoquia region. Utilizing a dataset originally consisting of over 100,000 images from 50 cameras, we aim to classify the following six animal species or genus: Peccary genus (collared and white-lipped), Black Agouti, Spotted Paca, Dasypus Species, South American Coati, and Bos Species. Significant variation in class sizes, colored and monochrome images, animal movement, and depth variation presented difficult challenges that were addressed through various preprocessing and feature extraction techniques. We developed models using Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Neural Networks (NN) to classify data into these six classes. The SVM model on the HOG feature vector demonstrated the highest performance, achieving 77\% accuracy on the testing set. While there is potential for further improvement, these results represent a significant step towards automating a traditionally manual task that contributes to wildlife preservation.

# Folder Structure

## 0\_data

Here you will find the labels for each of our test sets:

* dual\_class\_images.csv and dual\_class\_images: images containing more than one class per image were thus discarded.
* test\_downsampled.csv: contains the labels for our downsampled test set.
* train\_downsampled.csv: contains the labels for our downsampled train set.
* val\_downsampled.csv: contains the labels for our downsampled validation set.

You will also find the following folders:

* Deprecated: contains the initial steps we took in our project to find datasets and load them into GCP
* Sample\_images: contains images for some of the classes, as part of our initial EDA for the project proposal.

## 1\_proposal

Contains the files for our initial proposal for class 281 - computer vision.

* 2 Minute Overview Notes: contains the elevator pitch for our project.
* 281-Team-1-2 Project Proposal (doc and pdf): the initial proposal.

## 2\_colab\_notebooks

All of the code we used for our project is stored here, in the order used in our pipeline:

* 0: EDA
* 1: Data splitting
* 2: Data preprocessing
* 3: Feature EDA
* 4: Feature extraction, scaling and dimensionality reduction
* 5. Classifiers (LDA, SVM, NN)
* 6: Verification of train labels by visualizing images per class

As well as a folder, Archive, which includes previous iterations of our code.

## 3\_output

Includes the deliverables for our project:

* Final\_report.pdf: The paper written for our project
* Presentation: The slide deck with our project findings.

## \*\_references

Includes scientific papers we read to research for relevant features for our project, as well as sample papers from past students, and the project rubric

## \*\_team\_notes

Includes information that guided our project:

* Agenda 281-team-1-2: our weekly meeting minutes
* Data\_lineage: how the data is stored in GCP
* Notes\_and\_Observations: of odd things we noticed in our data along the way.