AUTOMATED PEST IMAGE RECOGNITION TO ASSIST IN PROTECTING AMERICAN FORESTS

BY CHRISTOPHER MARAIS

INTRODUCTION

Bark beetle is the common name for beetle species that fall into the *Scolytinae* clade. Some of these species can cause extensive tree mortality through infestations. Infestations often happen due to invasive bark beetle species or due to the environment changing as a result of climate change. The post-COVID era makes the importance of disease prevention obvious. This is also true for bark beetle infestations just as it is for human infectious diseases. However, to prevent something it first needs to be detected and identified. In this study we propose a machine learning image recognition model capable of identifying non-native bark beetle species in Florida to allow for more effective prevention of bark beetle infestations.

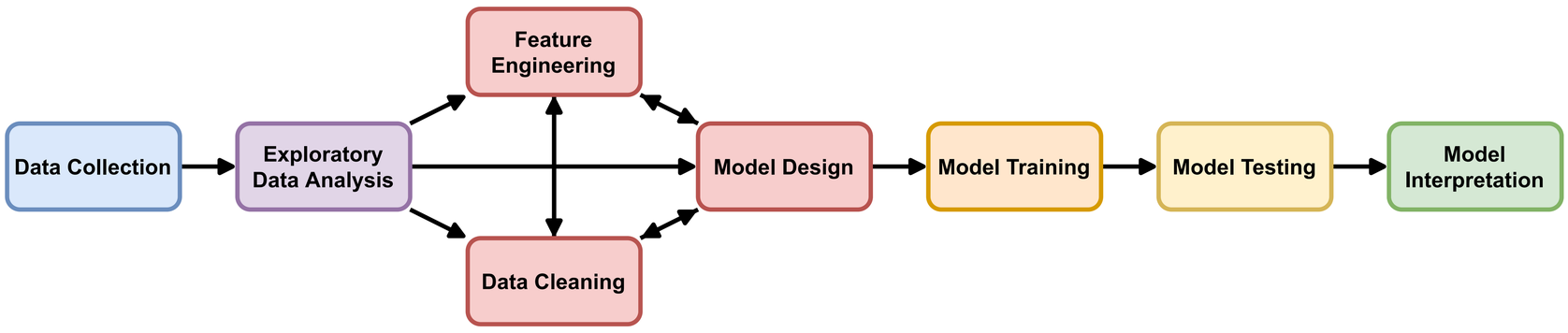
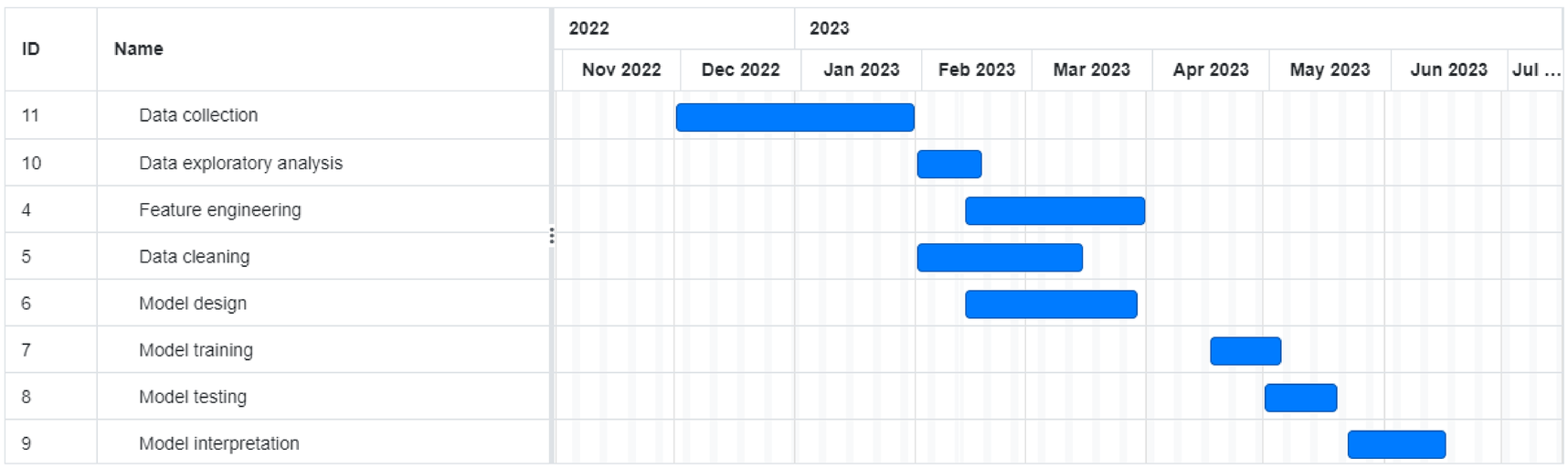


Figure 1: Project flow pipeline for creating and evaluating a machine learning model capable of identifying bark beetles from images.



ID	Name	2022		2023						
		Nov 2022	Dec 2022	Jan 2023	Feb 2023	Mar 2023	Apr 2023	May 2023	Jun 2023	Jul ...
11	Data collection									
10	Data exploratory analysis									
4	Feature engineering									
5	Data cleaning									
6	Model design									
7	Model training									
8	Model testing									
9	Model interpretation									

Figure 2: Projected timeline of engineering and interpreting an image recognition machine learning model.

MODEL DESIGN

Machine learning (ML) is a commonly used method in automating image classification. The convolutional neural network (CNN) architecture has specifically shown to be extremely effective at accurate image classification. One drawback of a conventional CNN is that it does not leverage the existing hierarchies that classes commonly form a part of. In this study we will use and compare a Branch Convolutional Neural Network (B-CNN) to a conventional CNN. The B-CNN can leverage the existing hierarchies in our classes to ostensibly improve classification accuracy.

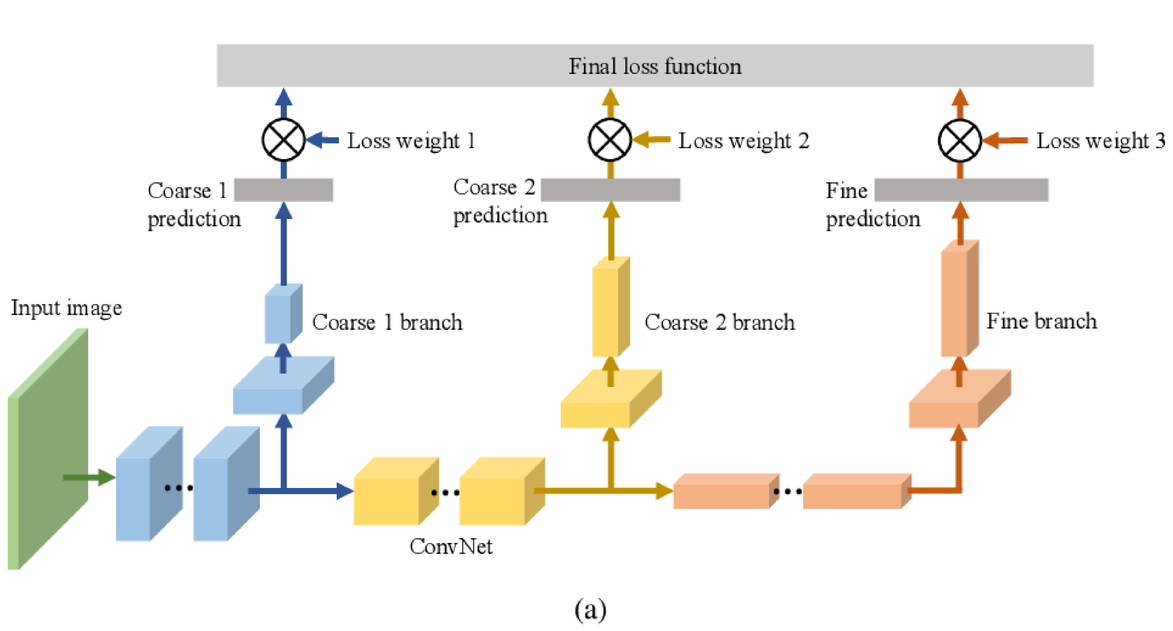


Figure 3: *"(a) Architecture of Branch Convolutional Neural Network (B-CNN). The network at the bottom can be an arbitrary ConvNet. There can be multiple branch networks and each of them outputs a coarse prediction. The final loss function is a weighted summation of all coarse losses. (b) A sample hierarchical label tree where classes are taken from CIFAR-100 dataset." - Zhu, X., & Bain, M. (2017). B-CNN: Branch Convolutional Neural Network for Hierarchical Classification. ArXiv, abs/1709.09890.*

PROBLEM STATEMENT

- Bark beetles are difficult to identify visually due to their small size (1-5mm in length)
- Bark beetles also share very similar visual characteristics and training is required to effectively identify different species visually through a microscope
- Identifying bark beetles through molecular methods such as DNA sequencing is also not viable as this is expensive to perform on scale
- The current process of detecting bark beetles is time consuming with a long delay between trap deployment and identification
- Samples are currently classified by hand by the United States Department of Agriculture (USDA) leading to identification fatigue and human error
- Data management is lacking as samples discarded and not stored for backreferencing

AIM AND HYPOTHESES

- **Aim:** Produce a machine learning model capable of identifying native bark beetle genera and species from microscopic images and size measurements
- **Hypotheses:**
  1. Size is the most informative factor in classifying species
  2. A hierarchical convolutional neural network is more accurate than a standard convolutional neural network at classifying images as genera and species
  3. The Machine learning model is significantly more consistent, and faster at classifying beetles accurately than the current process
  4. Adding a broad range of beetle image data to the training data improves classification accuracy for identifying native bark beetle species and genera