Literature Review

Christopher Marais

FNR5608 Research Planning

28 October 2022

Introduction

Bark beetle is a common name used to refer to hundreds of different beetle species that fall into the Scolytinae and Platypodinae subfamilies. Some of these species can cause much more extensive tree mortality and have large scale damaging effects on forests than others. Distinguishing between these species is therefore an important component in responding to possible harmful infestations.

Identifying the species to which an individual bark beetle belongs to by visual classification is a difficult task. Bark beetles are very small at only a few millimeters in length and have very similar visual characteristics between species such as size and shape. The alternative of classifying species molecularly is comparatively slow and expensive to perform on scale with hundreds of beetles.

In this study we aim to produce a machine learning model capable of classifying bark beetle genus and species from microscopic images as well as beetle size. This model will be used to test multiple hypotheses. Firstly, we hypothesize that size is the most informative visual factor in classifying beetle species. Secondly, we hypothesize that it will have better classification accuracy when using a hierarchical model that first classifies on the genus level before classifying on a species level.

This research may provide clues into which features are the most informative for distinguishing between beetle species. It will also provide insights into which beetles are the most difficult to distinguish from one another be it for a lack of data of a lack of visual variation between species.

Bark and Ambrosia Beetles

Bark and ambrosia beetles refer to beetles in the Curculionidae family, more specifically to beetles in the Scolytinae and Platypodinae subfamilies [1]. This includes over 7000 different species of beetles with approximately 250 genera that make up 26 different tribes [2]. All bark and ambrosia beetles are wood boring beetles that bore into many different species of woody plants globally. Even though as a group these beetles have a global range, some species are limited to particular regions and also attack only particular tree or plant species [3,4].

Most bark beetle species are not considered pests; however, some can cause large scale damage to forests [5] . The two main factors that lead to tree mortality through bark beetle attacks are increased winter temperatures and invasive bark beetle species [6]. It is hypothesized that climate change reduces bark beetle development and temperature-induced mortality which in turn allows them to develop bigger and more damaging swarms in areas where these beetles were previously scarcer [5]. Increased trade and travel in the modern world also allow for more cross contamination and introduction of invasive species between regions.

Bark beetles can have a substantial impact ecologically by destroying habitats and affecting multiple organisms [7]. Damage caused to forests also have a notable economic impact on the pulp, paper, and timber industries [8]. It is clear that the risk is high when it comes to destructive species of bark and ambrosia beetles and that this risk is increasing with the rise of climate change and trade between countries. Even though there are techniques that forest managers can employ to counter unwanted bark and ambrosia beetle infestations they often have little time to execute them in time to counter an attack [5]. There have been several attempts to predict bark beetle outbreaks, however, these models tend to be contextual to a specific area or bark beetle species [9,10]. To build supervised predictive machine learning models on the spread of invasive species it would be beneficial have accurate data.

An essential part of recording species or tribe specific data is naturally to identify samples correctly. This is not trivial when working with bark and ambrosia beetles as they tend to be smaller than 8 mm in length. This small size makes it hard to classify species by eye and often microscopes are required to investigate the smaller intricate features that differ between tribes and species [11]. This same limitation also makes it so that some training or experience is required to classify beetles reliably. An alternative method and more accurate way of classifying bark beetle species is to use DNA barcoding[12]. However, this is not a scalable way to identify large amounts of beetles collected from multiple traps from multiple forests.

In a previous program the United States Department of Agriculture had more than 300 traps in 22 different states. This program was aimed at early detection and rapid response of bark and ambrosia beetles [13]. This program relied on three taxonomists to correctly identify bark and ambrosia beetles which posed as a bottleneck in the processing of samples. A more standardized alternative to visually classifying samples would be to make use of an automated image recognition system. This can potentially lower the costs and decrease the bottleneck in processing samples. The power of image recognition for species classification has already been shown in mobile applications such as iNaturalist[14]. Some other examples include classification models aimed at classifying carabid beetles and ants[15–17]. A lot of work has also been done on many other insects or small objects proving that this field has a lot of potential[18–20]. However, no such studies have yet been done for bark and ambrosia beetles.

Image classification with machine learning

In the past 10 years deep learning and more specifically convolutional neural networks (CNN) have dominated image machine learning applications[21]. Many different variants of this deep learning architecture have been explored, but the core concepts persist as a reliable way of training machine learning models on image data. CNNs offer some simple advantages over standard artificial neural networks (ANNs). The convolutional layers allow them to reduce the number of parameters when compared to ANNs. CNNs are also translation invariant whereas ANNs are sensitive to different translations of images[22]. One augmented version of the CNN worth mentioning is the hierarchical Branch Convolutional Neural Network (B-CNN) architecture[23]. This version of the CNN is adapted to have multiple output layers making it capable to classify on orders of class resolution and specificity. An expansion of this hierarchical approach known as the Hierarchical Bilinear Convolutional Neural Network (HB-CNN) takes a multi-task learning approach to produce even better results[24]. This shows that these hierarchical algorithms are capable to ingest predefined hierarchical information to improve performance.

One common hurdle when classifying small entities is that there are often multiple objects in a single image that need to be classified individually. The standard way of overcoming this is by either taking individual images of samples or to perform object detection in an image and then break up a single image into multiple small images that each can be classified [25]. Object detection itself is also a form of image classification or more specifically pixel classification. The purpose of object detection is to identify the pixels in an image that belong to a specific entity. Examples may include non-neural network-based methods, such as Histogram of Oriented Gradients (HOG) feature description, or a neural network-based, such as in the case of the You Only Look Once (YOLO) algorithm[26,27]. One thing that is vital when using these automated object detection methods is that the noise in the background of the image, the complexity, and size of the objects all contribute to the efficacy of the algorithm being used [28].

A commonly mentioned criticism of deep learning is that the operations of the system are not certain to the user. This has led to a rising interest in the field of explainable artificial intelligence (XAI)[29,30]. This field is concerned with the ways in which machine learning models are interpreted. For computer vision tasks using deep learning models XAI is often implemented by identifying the most important pixels in an image to the model in use [31]. There are many different tools in which to do this, but most tools make use of the trained parameters of a network and some defined input image to identify the pixels that are most critical to the model's outcome[29]. These pixels can then be visualized, analyzed, and reviewed to understand the model. By highlighting the features or pixels on a model image it is possible to gain insight into the biases about the training data and the model. This way the "black box" element of deep neural networks can be reduced.

References

1. T.H. Atkinson Bark and Ambrosia beetle pages: Home. https://www.barkbeetles.info/.

2. Pistone, D., Gohli, J. & Jordal, B. H. Molecular phylogeny of bark and ambrosia beetles (Curculionidae: Scolytinae) based on 18 molecular markers. *Syst Entomol* **43**, 387–406 (2018).

3. Kirkendall, L. R., Biedermann, P. H. W. & Jordal, B. H. Evolution and Diversity of Bark and Ambrosia Beetles. *Bark Beetles: Biology and Ecology of Native and Invasive Species* 85–156 (2015) doi:10.1016/B978-0-12-417156-5.00003-4.

4. KnÍŽek, M. & Beaver, R. Taxonomy and Systematics of Bark and Ambrosia Beetles. *Bark and Wood Boring Insects in Living Trees in Europe, a Synthesis* 41–54 (2007) doi:10.1007/978-1-4020-2241-8_5.

5. Bark Beetles in a Changing Environment on JSTOR. https://www.jstor.org/stable/20113226#metadata_info_tab_contents.

6. Krokene, P. Conifer Defense and Resistance to Bark Beetles. *Bark Beetles: Biology and Ecology of Native and Invasive Species* 177–207 (2015) doi:10.1016/B978-0-12-417156-5.00005-8.

7. Morris, J. L. *et al.* Managing bark beetle impacts on ecosystems and society: priority questions to motivate future research. *Journal of Applied Ecology* **54**, 750–760 (2017).

8. Grégoire, J. C., Raffa, K. F. & Lindgren, B. S. Economics and Politics of Bark Beetles. *Bark Beetles: Biology and Ecology of Native and Invasive Species* 585–613 (2015) doi:10.1016/B978-0-12-417156-5.00015-0.

9. Ortiz, S. M., Breidenbach, J. & Kändler, G. Early Detection of Bark Beetle Green Attack Using TerraSAR-X and RapidEye Data. *Remote Sensing 2013, Vol. 5, Pages 1912-1931* **5**, 1912–1931 (2013).

10. Rammer, W. & Seidl, R. Harnessing Deep Learning in Ecology: An Example Predicting Bark Beetle Outbreaks. *Front Plant Sci* **10**, 1327 (2019).

11. Hulcr, J., Atkinson, T. H., Cognato, A. I., Jordal, B. H. & McKenna, D. D. Morphology, Taxonomy, and Phylogenetics of Bark Beetles. *Bark Beetles: Biology and Ecology of Native and Invasive Species* 41–84 (2015) doi:10.1016/B978-0-12-417156-5.00002-2.

12. Jordal, B. H. & Kambestad, M. DNA barcoding of bark and ambrosia beetles reveals excessive NUMTs and consistent east-west divergence across Palearctic forests. *Mol Ecol Resour* **14**, 7–17 (2014).

13. Rabaglia, R., Duerr, D., Acciavatti, R. & Ragenovich, I. Early Detection and Rapid Response for Non-Native Bark and Ambrosia Beetles. (2008).

14. Horn, G. van *et al.* The iNaturalist Species Classification and Detection Dataset. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 8769–8778 (2017) doi:10.48550/arxiv.1707.06642.

15. Hansen, O. L. P. *et al.* Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecol Evol* **10**, 737–747 (2020).

16. Marques, A. C. R. *et al.* Ant genera identification using an ensemble of convolutional neural networks. *PLoS One* **13**, e0192011 (2018).

17. Boer, M. J. A. & Vos, R. A. Taxonomic Classification of Ants (Formicidae) from Images using Deep Learning. *bioRxiv* 407452 (2018) doi:10.1101/407452.

18. Parmezan, A. R. S., Souza, V. M. A., Seth, A., Žliobaitė, I. & Batista, G. E. A. P. A. Hierarchical classification of pollinating flying insects under changing environments. *Ecol Inform* **70**, 101751 (2022).

19. Tresson, P., Carval, D., Tixier, P. & Puech, W. Hierarchical Classification of Very Small Objects: Application to the Detection of Arthropod Species. *IEEE Access* **9**, 63925–63932 (2021).

20. Martineau, M. *et al.* A survey on image-based insect classification. *Pattern Recognit* **65**, 273–284 (2017).

21. Chen, L. *et al.* Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sensing 2021, Vol. 13, Page 4712* **13**, 4712 (2021).

22. O'Shea, K. & Nash, R. An Introduction to Convolutional Neural Networks. (2015) doi:10.48550/arxiv.1511.08458.

23. Zhu, X. & Bain, M. B-CNN: Branch Convolutional Neural Network for Hierarchical Classification.

24. Zhang, X. *et al.* Hierarchical bilinear convolutional neural network for image classification. *IET Computer Vision* **15**, 197–207 (2021).

25. Zou, Z., Shi, Z., Guo, Y., Ye, J. & Member, S. Object Detection in 20 Years: A Survey. (2019) doi:10.48550/arxiv.1905.05055.

26. Dalal, N., Histograms, B. T. & Triggs, B. Histograms of Oriented Gradients for Human Detection. 886–893 (2005) doi:10.1109/CVPR.2005.177ï.

27. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *ArXiv* **2016-December**, arXiv:1506.02640 (2015).

28. Hu, P. & Ramanan, D. Finding tiny faces. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017-January**, 1522–1530 (2017).

29. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy 2021, Vol. 23, Page 18* **23**, 18 (2020).

30. Samek, W., Wiegand, T. & Müller, K.-R. EXPLAINABLE ARTIFICIAL INTELLIGENCE: UNDERSTANDING, VISUALIZING AND INTERPRETING DEEP LEARNING MODELS.

31. Vermeire, T., Brughmans, D., Goethals, S., de Oliveira, R. M. B. & Martens, D. Explainable image classification with evidence counterfactual. *Pattern Analysis and Applications* **25**, 315–335 (2022).