8-9-2022

Advancement of field-deployable, computer-vision wood identification technology

Adam Carter Wade
Mississippi State University, adamcwade80@gmail.com

Advancement of field-deployable, computer-vision wood identification technology

By

Adam Carter Wade

Approved by:
Frank C. Owens (Major Professor)
Prabu Ravindran
Alex C. Wiedenhoeft
Rubin Shmulsky (Committee Member/Graduate Coordinator)
L. Wes Burger (Dean, College of Forest Resources)

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Forest Resources
in the Department of Sustainable Bioproducts

Mississippi State, Mississippi
August 2022

Name: Adam Carter Wade

Date of Degree: August 9, 2022

Institution: Mississippi State University

Major Field: Forest Resources

Major Professor: Frank C. Owens

Title of Study: Advancement of field-deployable, computer-vision wood identification technology

Pages in Study: 133

Candidate for Degree of Doctor of Philosophy

Globally, illegal logging poses a significant threat. This results in environmental damage as well as lost profits for legitimate wood product producers and taxes for governments. A global value of $30 to $100 billion is estimated to be associated with illegal logging and processing. Field identification of wood species is fundamental to combating species fraud and misrepresentation in global wood trade. Using computer vision wood identification (CVWID) systems, wood can be identified without the need for time-consuming and costly offsite visual inspections by trained wood anatomists. While CVWID research has received significant attention, most studies have not considered the generalization capabilities of the models by testing them on a field sample, and only report overall accuracy without considering misclassifications. The aim of this dissertation is to advance the design and development of CVWID systems by addressing three objectives: 1) to develop functional, field-deployable CVWID models for Peruvian and North American hardwoods, 2) test the ability of CVWID to solve increasingly challenging problems (e.g., larger class sizes, lower anatomical diversity, and spatial heterogeneity in the context of porosity), and 3) to evaluate the generalization capabilities by testing models on independent specimens not included in training and analyzing

misclassifications. This research features four main sections: 1) an introduction summarizing each chapter, 2) a chapter (Chapter 2) developing a 24-class model for Peruvian hardwoods and testing its generalization capabilities with independent specimens not used in training, 3) a chapter (Chapter 3) on the design and implementation of a continental scale 22-class model for North American diffuse-porous hardwoods using wood anatomy-driven model performance evaluation, and 3) a chapter (Chapter 4) on the development of a 17-class models for North American ring-porous hardwoods, in particular examining the model's effectiveness in dealing with the greater spatial heterogeneity of ring-porous hardwoods.

DEDICATION

To my father, Robert Carter Wade, for his support, encouragement, and being my best friend. May he rest in peace (March 01, 2022).

To my mother, Barbara (Wendy) F. Wade, for her love, support, and sacrifices.

To my children, Star and Seth Wade, may you always follow your dreams and never give up no matter how hard life may feel.

To the rest of my family and friends near and far who believed in me.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Frank C. Owens, and committee members, Dr. Rubin Shmulsky, Dr. Alex C. Wiedenhoeft, and Dr. Prabu Ravindran, for their valuable guidance, time, and support. Their support helped me reach my goals while I was pursuing my doctorate at Mississippi State University.

I would like to express my gratitude to my colleagues, Dr. Adriana Costa and Dr. Brunela Rodrigues at the Department of Sustainable Bioproducts, for their kind assistance and ongoing support. Working with them has been a rewarding experience and a wonderful opportunity to learn and grow.

The author acknowledges the support of U.S. Department of Agriculture, Research, Education, and Economics, Agriculture Research Service, Administrative and Financial Management, Financial Management and Accounting Division Grants and Agreements Management Branch, under Agreement No. 58-0204-9-164. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION: EXECUTIVE SUMMARY

In 2018, the global trade of forest products represented a value chain of more than 550 billion U.S. dollars[1] with illegal logging accounting for 15–30% of the global timber supply chain (Food and Agriculture Organization, 2018; Nellemann, 2012). This value chain includes products such as lumber, logs, dressed lumber, veneers, furniture, pulp, wood fuel, and others (Food and Agriculture Organization, 2018). Illegal logging is the most profitable form of transnational natural resource crime and the fourth most lucrative form of transnational crime following counterfeiting, drug trafficking, and human trafficking (May, 2017).

One tool used to combat this massive problem is in-field wood identification. Conventional wood identification is largely human-based whereby a trained expert observes the anatomical features with a hand lens or a microscope (Wheeler and Baas, 1998). These observations are then compared to reference descriptions, verified specimens, field guides, or online resources such as InsideWood[2] (Wheeler and Baas, 1998). However, substantial training in wood anatomy is required to become proficient at recognizing and interpreting the anatomical patterns needed for accurate identification. A study by Wiedenhoeft et al. (2019), presents evidence suggesting that there is likely a severe shortage of wood anatomists that can identify wood at a forensic level in the U.S. and presumably worldwide. It is challenging to sustain the level of expertise needed for reliable wood identification due to individuals retiring, changing

[1] http://www.fao.org/faostat/en/#data/FO
[2] https://insidewood.lib.ncsu.edu/

jobs, or simply losing interest. Acquiring necessary proficiency would require several months to years, which does not scale with the current demand (Wiedenhoeft et al., 2019).

Automated wood identification may offer a solution to this shortage of human expertise (Ravindran et al., 2020). Computer vision wood identification (CVWID) uses a combination of feature extraction and machine learning, a method of automating decisions without instructions from a human (Hwang and Sugiyama, 2021). CVWID systems consist of hardware (digital camera, lens, and a computer) and software (feature detection and classification algorithms). The potential benefit of CVWID systems is the ability to provide accurate and reliable identifications without the need for extensive human training (Hermanson and Wiedenhoeft, 2011).

One such CVWID system is the XyloTron, an open-source, field-deployable CVWID system designed at the USDA Forest Products Laboratory (Ravindran et al., 2020). The XyloTron is comprised of the XyloScope (camera and lens) and a laptop paired with convolutional neural network (CNN) software for identification.

Published research on computer-vision based automated wood identification systems began with Khalid et al. (2008), in which the authors developed a prototype computer-based wood identification system that was capable of classifying 20 tropical Malaysian wood species with an accuracy of 95.0%, suggesting that automated identification systems might be suitable for commercial purposes. Martins et al., (2013) performed two different experiments to classify 112 different softwood and hardwood species. In the first experiment, classifiers were trained to distinguish between softwoods and hardwoods, a 2-class model. In the second experiment, classifiers were trained to distinguish between all 112 species as individual classes. They reported 98.6% and 86.0% accuracy, respectively. These results suggest a decrease in model performance as the number of classes increases. Filho et al. (2014) proposed a two-level divide-

and-conquer classification approach in which an input image is divided into several sub-images, each of which is classified independently, and the partial decisions are then combined in order to produce a final decision. When classifying 41 different wood species from Brazil, the best accuracy they achieved was 97.7%. Ravindran et al. (2018) developed a 10-class species- and a 6-class genus-level computer-vision classification model for ten CITES-listed species and other neotropical species in the family Meliaceae. For species-level and genus-level predictions, each model achieved an accuracy of 87.4% and 97.5%, respectively. This study also made an initial analysis of the misclassified images and found that they generally corresponded to human-based wood identification errors.

Previous studies on the development of CVWID models show high *in silico* accuracies (Hwang and Sugiyama, 2021); however, most studies have not tested the generalizability of their models with subsequent field-testing on independent specimens, that is, these studies have not demonstrated that their models can perform on new specimens. Most studies only report the overall accuracies of the models with little or no analysis of misclassifications. Ravindran et al., (2019) developed a 15-class model for Ghanaian timbers and were the first to perform field testing on independent specimens. They discovered a deployment gap (discrepancy between lab and field accuracies) of 25.0%, when *in silico* accuracy was 97.0%, and a field accuracy on independent specimens of 72.0%.

To advance the design and development of CVWID systems, this dissertation has three objectives: 1) to develop functional, field-deployable CVWID models for Peruvian and North American hardwoods, 2) to determine if CVWID can solve progressively more challenging problems (e.g. larger number of classes, lower anatomical diversity, and spatial heterogeneity as

it relates to porosity), and 3) to evaluate generalizability by testing models on independent specimens not used in training and analyzing misclassifications.

This dissertation contains four major sections: 1) an introduction (Chapter 1) summarizing each chapter, 2) a chapter (Chapter 2) developing a 24-class CVWID model for Peruvian hardwoods, the largest class model for Peru to date, and testing the model's generalization capabilities on independent test specimens not used training, 3) a chapter (Chapter 3) on the design and implementation of the first and largest continental scale model for North American diffuse-porous hardwoods using wood anatomy-driven model performance evaluation, and 4) a chapter (Chapter 4) on the development of one of the first and largest CVWID models for North American ring-porous hardwoods, specifically investigating how well the model handles the greater spatial heterogeneity of ring-porous hardwoods.

1.1 Summary of Chapter 2

Forests in Peru are greatly threatened by illegal logging, which is found in more than two-thirds of logging concessions and at least 58% of exported lumber (Finer et al., 2014). To help combat illegal logging in Peru, a 24-class CVWID model was trained using images of specimens from 228 Peruvian species obtained from six xylaria using the open-source, field-deployable XyloTron platform. This chapter evaluates the performance of CVWID with larger class sizes and presents the largest tested national-scale CVWID model for Peruvian woods that is ready for immediate in-country field evaluation and will be tested in Fall of 2022. The results of the specimen-level cross-validation accuracy were 97.0%. The generalization capabilities of the model were assessed by using independent test specimens, multiple hardware instantiations, and different operators of varying skill levels. When tested on independent specimens from a xylarium that did not contribute to training data, the top-1 and top-2 classification accuracies are

86.5% and 92.4%, respectively. Based on the technology's readiness for implementation in real-world field screening scenarios, this study provided evidence that the technology has the potential to help promote legal and sustainable wood value chains in Peru.

1.2 Summary of Chapter 3

North American wood products is a multi-billion-dollar industry in the United States, with total output (summation of business revenues and industry sales of hardwood products) of over 130 billion USD (Hardwood Federation, 2016). Rapid and reliable identification of wood along this value chain is essential for sustainable practices and conservation efforts to succeed, (Wiedenhoeft et al., 2019). To design a complete CVWID model for North American (NA) hardwoods, approximately 40 classes would be needed, which exceeds anything published for this region. As a way to separate the NA hardwoods into smaller number of classes, this study leverages a domain-based factor by separating the NA hardwoods by wood anatomical spatial heterogeneity as it relates to porosity. This chapter presents the first and largest continental-scale model for identifying NA diffuse-porous hardwoods and assesses model performance among classes with lower wood anatomical diversity than previous work (e.g., Chapter 2, Peru). A five-fold cross-validation strategy was used to train and evaluate the model. Images from the training specimens were split into five folds with class level stratification while maintaining mutual exclusivity at the specimen level between the folds. Each specimen contributed to images in exactly one fold, where four folds were used to train the model and the remaining fold was used for validation. The field model was trained using 100% of the images in the five-fold cross-validation and then evaluated on specimens from a separate xylarium. The cross-validation model accuracy was 95.2%, and the top-1 and top-2 accuracies were 80.6% and 90.5%, respectively, when the field model was tested on independent specimens that were not included

in the model's training. Additionally, this study assessed misclassifications by considering the anatomical properties of the woods considered, pointing out the importance of wood anatomy informing CVWID model development.

1.3 Summary of Chapter 4

Wood identification is crucial to ensuring the legality of the hardwood value chain in North America. In continuing to develop a unified North American hardwood model, this chapter expands on the work in chapter 3 by training and evaluating complementary 17-class XyloTron CVWID models for identifying North American ring-porous hardwoods. Along with evaluating model accuracy and examining misclassifications, this study focuses on how the greater spatial heterogeneity of ring-porous woods may affect model predictions. In ring-porous hardwoods, the earlywood and latewood exhibit marked differences in vessel size and arrangement. Due to this greater spatial heterogeneity, capturing an image that does not contain all the anatomical features needed for accurate identification may be possible. For example, an image may capture only latewood for fast radial growth, omitting the larger diameter earlywood vessels. Slow radial growth can produce an image primarily of earlywood, displaying the relative absence of latewood features, which readily separates the white oak group from the red oak group. Additionally, some images may lack tangentially varying characteristics, such as broad rays in Quercus. The five-fold cross-validation model's results demonstrated 98.0% accuracy, and a field model tested on independent test specimens achieved top-1 and top-2 predictions of 91.4% and 100%, respectively. When testing the model on three smaller spatial heterogeneity datasets (Slow-Growth, Fast-Growth, and Broad Rays Absent), the results suggest that spatial heterogeneity may not affect model predictions to the extent anticipated. In addition, this study

highlights the importance of cooperation between wood anatomists and machine learning experts in developing field-deployable CVWID systems.

1.4 Conclusion

Previous studies have demonstrated the theoretical potential of CVWID technology as a way to overcome the shortage of wood identification experts, help combat illegal logging, and deter fraud and misrepresentation of forest products. This dissertation evaluates how CVWID models can solve progressively more challenging problems by: 1) investigating how increasing class sizes affects CVWID models, 2) evaluating how comparatively lower wood anatomical diversity among classes impacts CVWID model performance, and 3) examining how the greater spatial heterogeneity of North American ring-porous hardwoods influences model performance. The findings suggest the following:

- It is recommended that CVWID models with large class sizes are developed, taking into account domain-based factors for informed model selection, label space design, and model predictions, while leveraging human expertise in model development, implementation, and evaluation.
- The development and evaluation of CVWID technologies require a wood anatomically-informed approach.
- The greater spatial heterogeneity of North American ring-porous hardwoods does not appear to influence the model performance as initially anticipated.

1.5 References

Food and Agriculture Organization (2018). *Global Forest Products: Facts and Figures* (Food and Agricultural Organization (FAO)). http://www.fao.org/3/ca7415en/ca7415en.pdf (Accessed June 27, 2022).

Filho, P. L. P., Oliveira, L. S., Nisgoski, S., and Britto, A. S. Jr. (2014). Forest species recognition using macroscopic images. *Mach. Vis. Appl*. 25, 1019–1031. doi: 10.1007/s00138-014-0592-7

Finer, M., Jenkins, C. N., Sky, M. A. B., and Pine, J. (2014). Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Sci. Rep*. 4:4719.

Hardwood Federation (2016). *Economic Contribution of Hardwood Products: United States*. Available online at: http://hardwoodfederation.com/resources/Documents/EIS%20States/US.pdf (accessed 26, July 2021).

Hermanson, J., and Wiedenhoeft, A. (2011). A brief review of machine vision in the context of automated wood identification systems. *IAWA J*. 32, 233–250. doi: 10.1163/22941932-90000054

Hwang, S. W., and Sugiyama, J. (2021). Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review. *Plant Methods* 17, 1–21. doi: 10.1186/s13007-021-00746-1.

Khalid, M., Lew, E., Lee, Y., Yusof, R., and Nadaraj, M. (2008). Design of an intelligent wood species recognition system. *Int. J. Simulation Syst. Sci. Technol*. 9, 9–19.

Martins, J., Oliveira, L. S., Nisgoski, S., and Sabourin, R. (2013). A database for automatic classification of forest species. *Mach. Vis. Appl*. 24, 567–578. doi: 10.1007/s00138-012-0417-5

May, C. (2017). *Transnational Crime and the Developing World* (Global Financial Integrity). https://gfintegrity.org/report/transnational-crime-and-the-developing-world/ (Accessed June 27, 2022).

Nellemann, C. (2012). Green carbon, black trade: A rapid response assessment on illegal logging, tax fraud and laundering in the world's tropical forests, in *A Rapid Response Assessment* (United Nations Environment Programme, GRID-Arendal). www.grida.no/publications/126 (Accessed June 27, 2022).

Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R., Costa, A., Wiedenhoeft, A. (2019). Image based identification of Ghanaian timbers using the XyloTron: opportunities, risks, and challenges, in *Proceedings 2019 Workshop on Machine Learning for the Developing World*, (Vancouver, BC).doi: 10.48550/arxiv.1912.00296

Ravindran, P., Costa, A. M., Soares, R., and Wiedenhoeft, A. C. (2018). Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14:25.

Ravindran, P., Thompson, B. J., Soares, R. K., and Wiedenhoeft, A. C. (2020). The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. *Front. Plant Sci*. 11:1015.

Wheeler, E. A., and Baas, P. (1998). Wood identification - A Review. *IAWA J*. 19, 241–264. doi: 10.1163/22941932-90001528

Wiedenhoeft, A. C., Simeone, J., Smith, A., Parker-Forney, M., Soares, R., and Fishman, A. (2019). Fraud and misrepresentation in retail forest products exceeds U.S. Forensic wood science capacity. *PLoS One* 14:e0219917. doi: 10. 1371/journal.pone.0219917

CHAPTER II

FIELD-DEPLOYABLE COMPUTER VISION WOOD IDENTIFICATION OF PERUVIAN TIMBERS

Ravindran, P., Owens, F. C., Wade, A. C., Vega, P., Montenegro, R., Shmulsky, R. (2021). Field-deployable computer vision wood identification of Peruvian timbers. *Front. Plant Sci*. 12:647515. doi: 10.3389/fpls.2021.647515
(Republished with permission)

2.1 Abstract

Illegal logging is a major threat to forests in Peru, in the Amazon more broadly, and in the tropics globally. In Peru alone, more than two thirds of logging concessions showed unauthorized tree harvesting in natural protected areas and indigenous territories, and in 2016 more than half of exported lumber was of illegal origin. To help combat illegal logging and support legal timber trade in Peru we trained a convolutional neural network using transfer learning on images obtained from specimens in six xylaria using the open source, field-deployable XyloTron platform, for the classification of 228 Peruvian species into 24 anatomically informed and contextually relevant classes. The trained models achieved accuracies of 97% for five-fold cross validation, and 86.5 and 92.4% for top-1 and top-2 classification, respectively, on unique independent specimens from a xylarium that did not contribute training data. These results are the first multi-site, multi-user, multi-system-instantiation study for a national scale, computer vision wood identification system evaluated on independent scientific wood specimens. We demonstrate system readiness for evaluation in real-world field screening

scenarios using this accurate, affordable, and scalable technology for monitoring, incentivizing, and monetizing legal and sustainable wood value chains.

2.2 Keywords

XyloTron, wood identification, illegal logging and timber trade, computer vision, machine learning, deep learning

2.3 Introduction

State-owned Amazonian forests cover 60% of the total area of Peru with over 15.3 million hectares of the Amazon forest being classified as natural protected areas (SERNANP, 2020) and the rest supporting diverse modes of managed production (e.g., 11 million hectares designated as Forest Logging Concessions; Kometter, 2019). However more than 68% of supervised logging concessions showed unauthorized tree harvesting from natural protected areas and indigenous territories (Finer et al., 2014), and in 2016 alone at least 58% of exported lumber was of illegal origin (SBS and GIZ, 2018). According to official data, over the past decade the volume of wood produced from illegally logged trees increased from 1.15 to 1.8 million cubic meters per annum (OSINFOR, 2015—onward).

For the last twenty years Peru has been building governance infrastructure to achieve sustainability of its forest products, facilitated by national and international policies (Office of the US Trade Representative, 2006; SERFOR, 2015) to improve the monitoring and regulation of the forest products supply chain. Oversight of this monitoring is conducted at inspection stations by government authorities such as the National Forestry and Wildlife Service, the Supervisory Agency for Forest and Wildlife Resources, the Regional Offices of Forests and Wildlife, and the National Customs Superintendency of Peru. Rapid field identification of wood can help efficiently establish due cause for further investigation (UNODC, 2016) at these

inspection stations when officials are confronted with falsified documentation. In contrast to plant identification, which is based on common botanical structures (i.e., flowers, fruits, leaves), conventional wood identification is dependent on recognizing anatomical patterns in wood and comparing them to reference descriptions or specimens. Such identifications are best performed by highly trained wood anatomists with substantial training in forensic wood identification, and are typically conducted in a laboratory, which does not meet the needs for rapid field screening at the inspection stations.

In the larger Amazonian context, two notable initiatives that enable human-based wood identification are: a mobile phone-based identification key that enables humans to identify 157 species (Gontijo et al., 2017), and the development of electronic identification keys as part of the Brazil-Colombia Amazon Cooperation Treaty Organization (OTCA, 2018). The knowledge of wood anatomical characteristics of Peruvian species conveyed in academic publications (Acevedo and Kikata, 1994; Chavesta, 2015, 2018) and industry guides/manuals (Rodriguez and Sibille, 1996; Gonzales, 2008; Ugarte and Mori, 2018) have not yet been encapsulated in similar solutions and adopted for widespread human-based wood identification in Peru. The development and uptake of these solutions at the national level in Peru has been challenging, in part, due to limited institutional wood forensics capacity, limited opportunities (university courses and infrequent workshops) for human expertise development, and mostly localized access to xylaria for comparative forensic work (the largest Peruvian xylarium, with around 8,500 samples, is housed in the National Agrarian University, Lima). It should be noted that these approaches emphasize or depend on developing human-based expertise.

To remove the need for extensive human expertise and to enable officials with only a modicum of training to identify wood, computer vision-based approaches (Khalid et al., 2008;

Ravindran et al., 2018) have been explored for automated wood identification. Souza et al. (2020) and de Andrade et al. (2020) used machine learning for macroscopic image-based identification for woods of 21 and 46 Brazilian species, respectively. Apolinario et al. (2018) used a convolutional neural network (CNN) for identification of 7 commercial Peruvian timber species using a portable microscope. Recently, the open source XyloTron system (Ravindran et al., 2020), was used to demonstrate a field deployable computer vision wood identification model for fourteen commercial Colombian woods by Arévalo et al. (2021). Among these works, it should be noted that XyloTrons have been shown to have comparable/better accuracy than expensive mass spectrometric methods (Ravindran and Wiedenhoeft, 2020), have been deployed for charcoal identification across the European Union in partnership with the Forest Stewardship Council (as noted in Wiedenhoeft, 2020), and, critically, have been field-tested for wood identification in Ghana (Ravindran et al., 2019). This field testing of a machine learning model on wholly new specimens, ideally by distinct users and using distinct instantiations of the system, especially at the scale undertaken in this work, is lacking in virtually all forensic wood identification literature, regardless of the modality, technique, or the taxa studied.

In this study, we train 24 class (228 taxa grouped into anatomically informed classes representing 57% by volume of the commercially harvested roundwood and 66% by volume of the sawn wood produced in 2019 in Peru; SERFOR, 2020) CNN based computer vision identification models of Peruvian commercial woods for the XyloTron. We use wood specimens from the MADw, SJRw, BCTw, BOFw, Tw, and FORIGw xylaria to develop five-fold cross-validated models and then train a field model using the same hyperparameter values. The field model was trained by incorporating all the images and specimens used in the cross-validation analysis but was evaluated on completely different specimens from the PACw xylarium, using

different hardware and different operators. Performance evaluation of an automated wood identification system requires expert verification of each specimen identified by the system and can be logistically challenging. Our approach using verified, mutually exclusive specimens from distinct xylaria during the training and testing phases serves as a practical surrogate for field evaluation (a first step toward real-world field deployment) and provides a useful measure of the generalization capability of the identification system. To the best of our knowledge this is the first, large-scale study of Peruvian commercial timber identification using distinct instantiations of a computer vision identification system, in this case, the XyloTron.

2.4 Materials and Methods

2.4.1 Species Selection

The 24 Peruvian woods selected for this study represent 57% by volume of the commercially harvested roundwood and 66% by volume of the sawn wood produced in 2019 in Peru (SERFOR, 2020) and are listed in Appendix A.1.1, (Supplementary Material 1). Because wood anatomy is typically accurate only to the genus level (Gasson, 2011) and given that the XyloTron operates on macroscopic anatomical variation, we included a range of wood anatomically appropriate, congeneric, Amazonian species and restricted data collection to the transverse surface of the specimens (e.g., congeneric species that are differentiable only from the tangential surface are clubbed into the same class here).

2.4.2 Sample Preparation

The transverse surface of 1,419 wood specimens from seven xylaria (Table 2.1) were polished by sanding with progressively finer-grit sandpapers (240, 400, 600, 800, 1,000, 1,500). To the extent possible, compressed air and adhesive tape were employed to remove dust from

cell lumina between each grit. This sample preparation protocol enabled the consistent an efficient preparation of wood samples for imaging.

Table 2.1 Summary of xylaria and their wood specimen contributions for model training and testing.

Institution (Xylarium acronym)	Specimen counts	Role
USDA Forest Products Laboratory, Madison collection (MADw)	501	Model Training
USDA Forest Products Laboratory, Samuel J. Record collection (SJRw)	589	Model Training
Instituto de Pesquisas Tecnologicas do Estado de SaoPaulo (BCTw)	139	Model Training
Wood Laboratory, Universidad Distrital Francisco Jose de Caldas (BOFw)	37	Model Training
Royal Museum of Central Africa (Tw)	32	Model Training
Forestry Research Institute of Ghana (FORIGw)	2	Model Training
Mississippi State University, David A. Kribs collection (PACw)	**119**	**Model Testing**

Note: The testing xylarium and specimen count is in bold case.

2.4.3 Image Collection

The XyloTron (Ravindran et al., 2020), an open-source macroscopic imaging system, was used to collect 6244 nonoverlapping RGB images of the polished transverse surfaces of specimens from 228 taxa. Each XyloTron image shows 6.35 x 6.35 mm of tissue and has dimensions 2,048 x 2,048 pixels. Each institution employed one or more unique XyloTrons to

collect images, so at least seven different hardware instantiations were employed. The details of the collected image dataset are presented in Table 2.2.

Table 2.2 Details of the image data set.

	Training data (counts)	**PACw data (counts)**	**Total (counts)**
Number of specimens	1,300	119	1,419
Number of images	5,715	529	6,244
Number of taxa	186	70	228*

Note: 1,419 specimens from 228 unique taxa were prepared and imaged. *Some taxa appeared in both data sets, thus the total number of taxa is not the sum of the training and testing taxa.

2.4.4 Label Space Design

The 228 taxa included: (i) the species of interest to the Peruvian wood value chain, and (ii) additional congeneric macroscopically inseparable species native to South America. *Brosimum* was separated into two anatomically distinguishable classes while the remaining species were grouped into genus level classes, producing 24 classes. Complete details about the class labels and their constituent taxa are provided in Appendix A.1.2 (Supplementary Material 2)

2.4.5 Model Architecture and Training

A convolutional neural network (CNN; LeCun et al., 1989) classifier, with a aResNet50 (He et al., 2016) backbone and a custom head that included batchnorm (Ioffe and Szegedy, 2015), dropout (Srivastava et al., 2014), global average and max pooling, and linear layers (Goodfellow et al., 2016), was implemented for identification (see Figures 2.1 A, B). A two-

stage (Howard and Gugger, 2020) transfer learning (Pan and Yang, 2010) methodology, comprising locking the ImageNet (Russakovsky et al., 2015) pre-trained backbone weights while training the randomly initialized weights (He et al., 2015) of the custom head followed by fine tuning the weights of the entire network, was adopted (see Figures 2.1 C, D). The Adam optimizer (Kingma and Ba, 2015) with simultaneous cosine annealing of the learning rate (maximum value of 1.8e-2) and momentum (Smith, 2018) was employed with cross-entropy loss for both the stages. Random 2,048 x 768 image patches were sampled from the training images, downsampled to 512 x 192 pixel images, and fed to the CNN in batches of size 16 with a data augmentation strategy that included horizontal/vertical flips, small rotations and cutout (Devries and Taylor, 2017). The hyperparameters were the same across all the training runs. Further details about the hyperparameter settings and training methodology can be found in Ravindran et al. (2020). The model definition, training and evaluation was performed using PyTorch (Paszke et al., 2019) and scientific Python tools (Pedregosa et al., 2011).

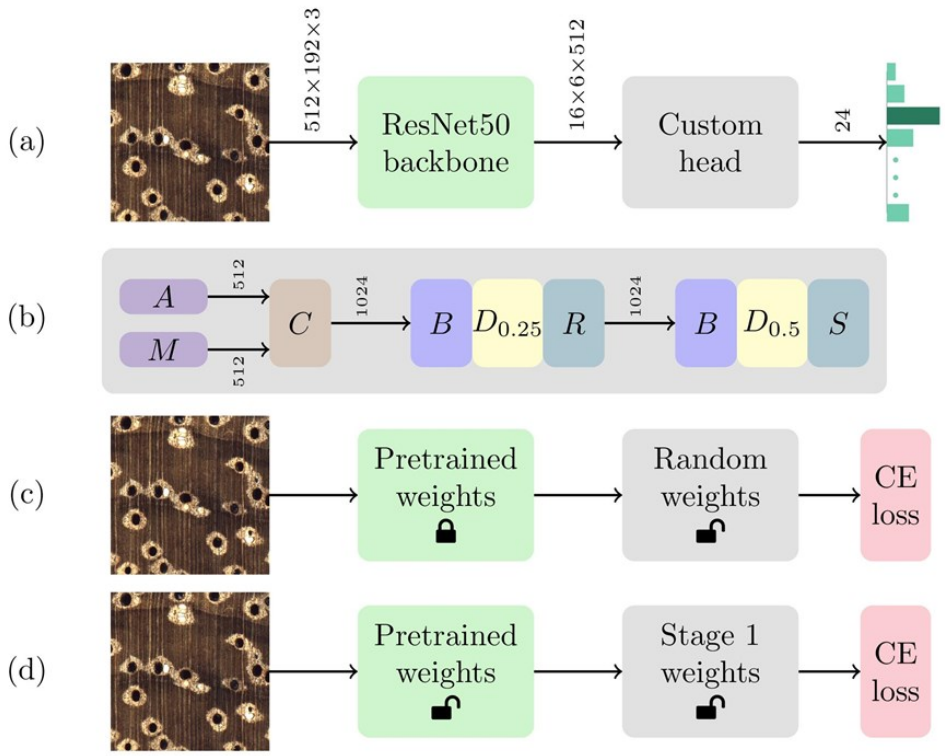

Figure 2.1 Model schematics of the CNN architecture.

Note: The CNN architecture comprises a ResNet50 backbone with a custom head. Given an input image, the network produces a 24-element vector that represents the prediction confidence for each of the 24 classes in the model. Tensor dimensions are depicted over the connections between the modules. (B) The custom head includes global average pooling (A), global max pooling (M), concatenation (C), batchnorm (B), dropout (D) and linear layers with ReLU (R) and softmax (S) activations. Dp represents a dropout layer with drop probability parameter p. Tensor dimensions are depicted over the connections between the layers. (C) The first stage of transfer learning locks (or freezes) the ImageNet pretrained weights of the ResNet50 backbone and optimizes the randomly initialized weights of the custom head using the cross-entropy (CE) loss. (D) The weights of the entire network are fine-tuned using the CE loss during the second stage of the training methodology.

2.4.6 Model Evaluation

The predictions of the trained models were evaluated at the specimen level with the predicted class for a specimen taken to be the majority of class predictions for the images contributed by the specimen. Accuracies based on the top prediction (top-1) for each specimen are reported for all the models. Additionally, the top two image-level predictions (from a

specimen) are aggregated, with equal weights, to generate the top-2 predictions for a specimen. If the true label is one of the top-2 specimen level predictions, the specimen is considered to be correctly identified.

The following two analyses were performed to evaluate model performance in this study:

(1) Training and evaluation using five-fold cross validation: Images from 1,300 specimens were split into five folds with class level stratification while maintaining mutual exclusivity at the specimen level between the folds i.e., each specimen contributed images to exactly one fold. This specimen-aware partitioning of the data into folds with distinct specimens is required for correct evaluation of a trained model's generalization capability to unseen samples. It should be noted that cross validation analysis did not include specimens from the PACw xylarium. A standard cross validation strategy, with four folds used for training and the trained model tested on the hold-out fold, was implemented and the specimen-level predictions over the five folds were accumulated to compute the accuracy (Table 2.3) and the confusion matrix (Figure 2.2).

(2) Training a field model for evaluation on PACw specimens: All images in the five-fold cross-validation analysis were used to train a single model—the field model—using the same training hyperparameters. The specimen-level prediction performance of the field model was tested on 119 specimens from the PACw xylarium at Mississippi State University. The top-1 and top-2 predictions of the field model are reported in Table 2.3.

The operators and XyloTron hardware used to collect the 529 images from the PACw specimens were different from those for the training data, and the images were used to evaluate the prediction accuracy of the trained model as a proxy for in-country field testing.

All images of the misclassified specimens in the five-fold cross validation were qualitatively evaluated and the misclassified specimens were categorized into three types: (1) taxa are anatomically consistent and the test specimen is typical; (2) the test specimen is atypical—but within reasonable variation for the taxon (i.e., it is not an archetypal image for the taxon); and (3) the taxa and test specimen are anatomically typical, but not anatomically consistent with each other. Types 1 and 2 represent misidentifications that trained field inspectors are likely to make, and so are sensible. Type 3 represents misidentifications not as likely to be made by trained human field inspectors, and for which there is no clear anatomical explanation.

2.5 Results

The cross-validated specimen-level identification accuracy (accumulated over the five folds) was 97%. The field model had top-1 and top-2 specimen-level accuracies of 86.5 and 92.4% when tested on the PACw specimens. The cross-validation confusion matrix is shown in Figure 2.2, and the predictive performance of the models is summarized in Table 2.3.

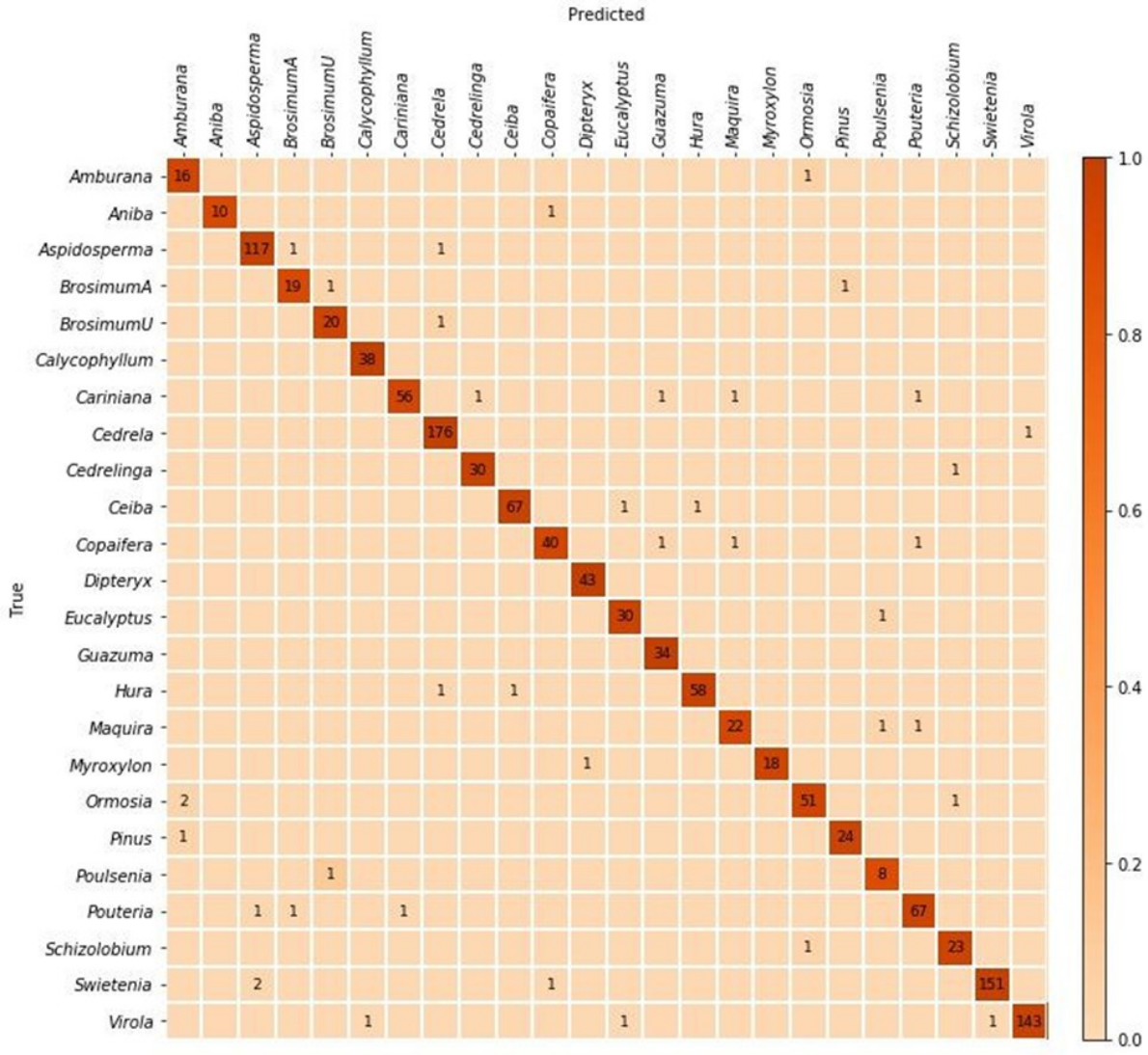

Figure 2.2 Confusion matrix for the top-1 predictions of the five-fold cross-validation models.

Note: The specimen-level accuracy accumulated over the five folds was 97%. The majority of misclassifications are between anatomically similar woods.

Table 2.3 Predictive accuracies for the trained models and the corresponding number of specimen-level prediction errors.

	Accuracy (%)	**Number of specimens misclassified**
Predictions on cross-validation folds	97%	39/1,300
Top-1 prediction on PACw specimens	86.5%	16/119
Top-2 prediction on PACw specimens	92.4%	9/119

Figure 2.3 presents examples of each of the three types of misclassifications, which are summarized and reported in Table 2.4. When comparing two wood anatomically similar taxa (Type 1 misclassification, Figures 2.3 A, B) the misclassification is sensible—both woods are characterized by vessels with similar grouping, arrangement, and of similar diameter, with lozenge-aliform-to-confluent axial parenchyma, and narrow, abundant rays. In Figures 2.3 C, D (an example of Type 2 misclassification) the similarities between the atypical specimen of class Virola (Virola surinamensis; Figure 2.3 C) and class Swietenia (Swietenia macrophylla; Figure 2.3 D) include prominent marginal parenchyma, roughly similar vessel diameters, similar vessel grouping and arrangement, and absence of axial parenchyma in the body of the growth ring. An example of anatomically disparate misclassification (Type 3 misclassification) is shown in Figures 2.3 E, F where the apotracheal banded parenchyma and much smaller vessels of class Cariniana (Cariniana pyriformis; Figure 2.3 E) present a pattern not at all similar to the human eye to the larger vessels and vasicentric axial parenchyma of class Cedrelinga (Cedrelinga cateniformis; Figure 2.3 F).

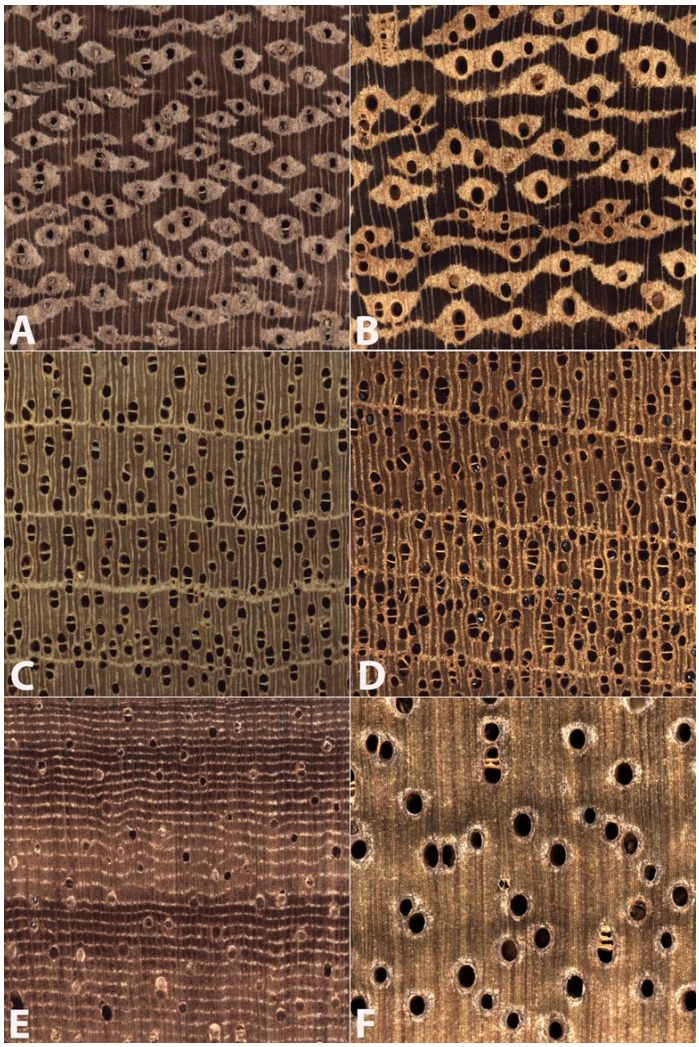

Figure 2.3 Images of the transverse surface of test specimens (A, C, E) and exemplars of the class to which they were assigned (B, D, F).

Note: All images are 6.35 mm on a side. An anatomically representative specimen of class Amburana (A) was misclassified as the anatomically similar class Ormosia (B). An anatomically atypical specimen of class Virola (C) was classified as class Swietenia (D). An anatomically typical specimen of class Cariniana (E) was misclassified as the wood anatomically disparate class Cedrelinga (F).

Table 2.4 Number and proportion of misclassified specimens from Figure 2.2 when categorizing into one of three misclassification types.

Misclassification type	Number of misclassified specimens	Proportion of misclassified specimens
Taxa are anatomically consistent, test specimen typical (Type 1)	13	0.333
Test specimen atypical for its taxon* (Type 2)	11	0.282
Taxa and test specimen are not anatomically consistent (Type 3)	15	0.385
Total	39	1.0

Note: Types 1 and 2 are consistent with wood anatomy and the kind of misidentifications likely to be made by human field inspectors. Type 3 misclassifications are inconsistent with macroscopic wood anatomy and would not be expected to be made by a human inspector. *But within reasonable variation for the taxon.

2.6 Discussion

The top-1 specimen-level accuracy of the field model was approximately 10 percentage points lower than the cross-validation accuracy while the top-2 specimen-level accuracy was over 90% — a level which is arguably sufficient to establish probable cause and initiate a full forensic investigation when fraud or misrepresentation is detected. The generalization capability of machine learning wood identification models must be evaluated on specimens that were not used to train the model. Additionally, real world systems deployed at scale must also be robust to the skills of operators (with different levels of training) and variations in system instantiations. The prediction accuracies reported above were obtained using training and testing datasets that were mutually exclusive at the specimen level. We maintained specimen-level mutual exclusivity

of specimens across folds for cross validation analysis, and likewise xylaria specimen mutual exclusivity for field model evaluation. Additionally, the performance evaluation metrics were obtained using data collected at multiple sites and by multiple operators using different instantiations of the XyloTron system.

Our approach of testing models on specimens from a xylarium that did not contribute data to model training was employed as a logistically manageable, practically useful surrogate for real-world field testing. The ultimate test of any automated wood identification system is in-country field testing, but the main logistical challenge is the requirement of a wood identification expert for validation of the specimens being tested. Prior field testing by Ravindran et al. (2019) of a pilot XyloTron model for Ghanaian woods showed a 25%[3] drop in identification accuracy when comparing results on xylarium specimens to testing on field specimens. Such losses of accuracy of computer vision models when tested on wholly new datasets have been found by research in other domains of computer vision (Recht et al., 2018, 2019; Zech et al., 2018). The drop in performance shown in Ravindran et al. (2019) and in this study could be attributed to a combination of many factors such as differences in the quality of specimen surface preparation; differences in subtle anatomical patterns present in xylarium specimens as compared to material currently in trade; differences between green and dry wood; and slight variations in operator use of the equipment or the equipment itself. A well-designed field-testing strategy for evaluating automated wood identification systems must incorporate these factors in a context-specific manner. For example, given that the XyloTron platform is intended as a field screening rather than a forensic tool, a testing protocol that incorporated taking multiple images per specimen of

[3] The reported value of 25% replaces the mistake value of 10% reported in the published paper.

multiple specimens per shipment/consignment, etc., should yield reliable, robust results when characterizing the shipment at large, rather than any single piece of wood.

Our top-2 specimen-level accuracy was computed with equal weights for the top-2 image-level predictions, but for practical deployment a weighting scheme should be chosen in a context dependent fashion that takes into account factors such as the taxa-aware cost of making an incorrect identification, the anatomical similarity of the taxa being considered, the number of specimens to field screen per shipment, and the calibration of the model predictions (Niculescu-Mizil and Carauna, 2005; Guo et al., 2017). By including top-2 specimen level accuracy, we provide a window into the performance of the model and how such a model could be deployed. For example, the XyloTron platform's classification software, xyloinf (Ravindran et al., 2020), provides the confidence value and an exemplar image for each class for the top-3 predictions per image, plus the sum of the confidences for the remaining N-3 classes in a given model of N classes. An operator thus has access not only to the ranked results, but also the confidence of a prediction and an exemplar image for human evaluation. This opens an interesting avenue for future research into the real-world deployment of computer vision wood identification systems (and other modalities) for maximum practical effect by incorporating human judgment (e.g., visual matching of an image from a field specimen to reference exemplar images for human approval and for flagging Type 3 misclassifications) or comparison of top-k results to some affirmative claim (e.g., a shipping manifest or transit permit). Even as field screening and forensic tools grow in power and sensitivity, it is critical to ensure that users of those tools are guided in how to achieve best practical effect with the tools at hand.

The uptake of computer vision and machine learning for automated wood identification is accelerating (Ravindran et al., 2018, 2019, 2020; de Andrade et al., 2020; Souza et al., 2020;

Arévalo et al., 2021) and the real-world adoption of these systems is critically dependent on rigorous validation metrics and methodologies underlying any well-considered field-deployment framework. An easy first step toward rigorous validation is to enforce specimen-level separation between the training and testing splits (as in this work) rather than only image-level separation (most prior works). As affordable mobile phone adaptations (Tang et al., 2018; Wiedenhoeft, 2020) democratize access to these automated technologies, for wider impactful adoption it is critical that they be rigorously evaluated on external validation data. For this work, the next obvious steps will be testing the field model on specimens in Peruvian xylaria; folding in the PACw specimens to train a new field model to test in Peruvian xylaria; folding in the specimens from the Peruvian xylaria to iterate a new field model; and then, taking that model into the real-world and conducting the necessary field-testing coupled with independent forensic validation of the field tested specimens, an approach that should be applied to all modalities (Dormontt et al., 2015) in forensic wood science.

2.7 Summary

We provided the largest tested computer vision wood identification model for Peruvian woods that is ready for immediate in-country field evaluation on the XyloTron platform. We demonstrated the utility and practicality of our model by evaluation using completely new specimens with independent hardware instantiations and different users, emphasized the critical need for specimen-level control of training and testing splits, and laid out a clear, iterative plan for augmenting the existing model. It is our hope that this work can be deployed within Peru to prevent illegally logged material from entering trade, and to support the trade in legal timber.

2.8 Data Availability Statement

A minimal data set can be obtained by contacting the corresponding author and the trained models will be made available, but the full data set used in the study is protected for up to 5 years by a CRADA between FPL, UW-Madison, and FSC.

2.9 Author Contributions

PV and RM provided the Peruvian timber market context and contributed the initial draft of the introduction. FO and RS provided access to and supervised data acquisition from the PACw test specimens. AWa prepared and imaged the PACw specimens. RM, FO, and AWi established the wood anatomical scope of the study. PR implemented the machine learning pipelines for the study. PR and AWi conducted data analysis, synthesis, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

2.10 Funding

This work was supported in part by a grant from the US Department of State via Interagency Agreement No. 19318814Y0010 to AWi and in part by research funding from the Forest Stewardship Council to AWi. PR was partially supported by a Wisconsin Idea Baldwin Grant. We wish to acknowledge the support of U.S. Department of Agriculture (USDA), Research, Education, and Economics (REE), Agriculture Research Service (ARS), Administrative and Financial Management (AFM), Financial Management and Accounting Division (FMAD) Grants and Agreements Management Branch (GAMB), under Agreement No. 58-0204-9-164, specifically for support of AWa, FO, and RS. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

2.11 Acknowledgments

We wish to gratefully acknowledge the specimen preparation and imaging efforts of Nicholas Bargren, Karl Kleinschmidt, Caitlin Gilly, Richard Soares, Adriana Costa, Emmanuel Ebanyenle, Asi Ebeheakey, Kofi Abban, Ophilious Lambog, and Flavio Ruffinatto. The feedback provided by the reviewers and the editor helped improve the manuscript and is gratefully acknowledged.

2.12 References

Acevedo, M., and Kikata, Y. (1994). *Atlas de Maderas del Peru*. Peru: Universidad Nacional Agraria La Molina.

Apolinario, M. P. E., Huamán Bustamante, S. G., and Orellana, G. C. (2018). "Deep learning applied to identification of commercial timber species from Peru," in Proceedings of the 2018 IEEE XXV *International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, (Peru: IEEE).

Arévalo, R., Pulido, E. N. R., Solórzano, J. F. G., Soares, R., Ruffinatto, F., Ravindran, P., et al. (2021). *Image based Identification of Colombian Timbers using the XyloTron: a Proof of Concept International Partnership (Identificación de Maderas Colombianas Utilizando el Xylotron: Prueba De Concepto de una Colaboración Internacional)*. Colombia: Universidad Distrital Francisco Jose de Caldas.

Chavesta, M. (2015). *Atlas Anatómico de Maderas del Perú II*. Lima: Universidad Nacional Agraria La Molina

Chavesta, M. (2018). *Atlas Anatómico de Maderas del Perú III*. Lima: Universidad Nacional Agraria La Molina.

de Andrade, B. G., Basso, V. M., and de Figueiredo Latorraca, J. V. (2020). Machine vision for field-level wood identification. *IAWA J*. 41, 681–698. doi: 10.1163/ 22941932-bja10001

Devries, T., and Taylor, G. W. (2017). *Improved Regularization of Convolutional Neural Networks with Cutout*. New York: Cornell University

Dormontt, E. E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., et al. (2015). Forensic timber identification: it's time to integrate disciplines to combat illegal logging. *Biol. Conserv*. 191, 790–798. doi: 10.1016/j.biocon.2015.06.038

Finer, M., Jenkins, C. N., Sky, M. A. B., and Pine, J. (2014). Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Sci. Rep*. 4:4719.

Gasson, P. (2011). How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. *IAWA J*. 32, 137–154. doi: 10.1163/22941932-90000049

Gontijo, A. B., Rodrigues, J. S., Santana, C. S., and Azevedo, N. (2017). "have de Identificação "madeiras comerciais do Brasil" (Versão Android)," *in I Fórum de Anatomistas de Madeira da Amazônia (Famazon)* (Brazil: Desafios E Perspectivas Para Os Próximos Anos).

Gonzales, E. V. (2008). *Identificación Organoléptica y Macroscópica de Maderas Comerciales*. CITEmadera: Lima.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: The MIT Press.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning* (New York, NY: ACM).

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *Proceedings of the 2015 International Conference on Computer Vision* (Santiago).

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (Las Vegas).

Howard, J., and Gugger, S. (2020). Fastai: a layered API for deep learning. *Information* 11:108. doi: 10.3390/info11020108

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning* (France).

Khalid, M., Lew, E., Lee, Y., Yusof, R., and Nadaraj, M. (2008). Design of an intelligent wood species recognition system. *Int. J. Simulation Syst. Sci. Technol*. 9, 9–19.

Kingma, D., and Ba, J. (2015). "Adam: a method for stochastic optimization," in *Proceedings of 2015 International Conference on Learning Representations* (San Diego).

Kometter, R. (2019). *Asistencia Técnica para el Análisis del Funcionamiento del Modelo de Conceciones Forestales Maderables desde la Perspectiva Técnica. Evaluación del modelo de concesiones forestales con fines maderables,que compila los análisis legal, técnicoy económicofinanciero, así como las propuestas normativas para el fortalecimiento del modelo. PROYECTO*. Washington, DC: USAID.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Niculescu-Mizil, A., and Carauna, R. (2005). "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning* (New York, NY: ACM).

Office of the US Trade Representative (2006). *United States - Peru Trade Promotion Agreement*. Washington, DC: Office of the US Trade Representative.

OSINFOR (2015). *onward. OSINFOR-SIGO. In: Supervision Agency for Wildlife Resources*. Available online at: https://observatorio.osinfor.gob.pe/Estadisticas/ Home/Reportes/9. (Accessed September 1, 2020).

OTCA (2018). *Clave de Identificación Electrónica de Especies Maderables de la Amazonia*. Peru: OCTA.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowledge Data Eng*. 22, 1345–1359.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., and Chanan, G. (2019). Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural Inform. Process. Syst*. 2019, 8026–8037.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., and Grisel, O. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res*. 12, 2825–2830.

Ravindran, P., and Wiedenhoeft, A. C. (2020). Comparison of two forensic wood identification technologies for ten Meliaceae woods: computer vision versus mass spectrometry. *Wood Sci. Technol*. 46:1163.

Ravindran, P., da Costa, A. M., Soares, R., and Wiedenhoeft, A. C. (2018). Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14:25.

Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R. (2019). "Image based identification of Ghanaian timbers using the XyloTron: opportunities, risks and challenges," in *Proceedings 2019 Workshop on Machine Learning for the Developing World* (New York: Cornell University).

Ravindran, P., Thompson, B. J., Soares, R. K., and Wiedenhoeft, A. C. (2020). The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. *Front. Plant Sci*. 11:1015.

Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2018). *Do CIFAR-10 Classifiers Generalize to CIFAR-10?*. New York: Cornell University.

Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2019). *Do ImageNet Classifiers Generalize to ImageNet?*. New York: Cornell University.

Rodriguez, M., and Sibille, A. M. (1996). *Manual de Identificación de Especies Forestales de la Subregión Andina. Proyecto PD 150/91 Rev. I(I). "Identificación y Nomenclatura de las Maderas Tropicales Comerciales en la Subregión Andina*.". Peru: ITTO.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet: large scale visual recognition challenge. *Int. J. Comput. Vis*. 115, 211–252. doi: 10.1007/s11263-015-0816-y

SBS and GIZ (2018). *Sectorial Assessment of Exposure to Money Laundering and Terrorist Financing Risks in the Peruvian Timber Sector*. Peru: Superintendency of Banking, Insurance and AFP.

SERFOR (2015). *Ley Forestal y de Fauna Silvestre N 29763 y sus Reglamentos. Bosques Productivos para la Vida*. Peru: SERFOR.

SERFOR (2020). Available online at: http://sniffs.serfor.gob.pe/estadistica/es

SERNANP (2020). *Nuestras Áreas Naturales Protegidas*. Peru: SERNANP.

Smith, L. (2018). *A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay*. New York: Cornell University.

Souza, D. V., Santos, J. X., Vieira, H. C., Naide, T. L., Nisgoski, S., and Oliveira, L. E. S. (2020). An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. *Wood Sci. Technol*. 54, 1065–1090. doi: 10.1007/s00226-020-01196-z

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res*. 15, 1929–1958.

Tang, X. J., Tay, Y. H., Siam, N. A., and Lim, S. C. (2018). "MyWood-ID: automated macroscopic wood identification system using smartphone and macro-lens," in *Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems* (New York, NY: ACM).

Ugarte, J., and Mori, I. (2018). *Guía Para la Identificación de la Madera de 50 Especies Forestales del Perú*. Peru: CITEmadera.

UNODC (2016). *Best Practice Guide for Forensic Timber Identification*. New York: UNDOC.

Wiedenhoeft, A. C. (2020). The XyloPhone: toward democratizing access to highquality macroscopic imaging for wood and other substrates. *IAWA J*. 41, 699–719. doi: 10.1163/22941932-bja10043

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., and Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 15:e1002683. doi: 10.1371/journal.pmed.1002683

CHAPTER III

TOWARDS SUSTAINABLE NORTH AMERICAN WOOD PRODUCT VALUE CHAINS, PART 1: COMPUTER VISION IDENTIFICATION OF DIFFUSE-POROUS HARDWOODS

Ravindran, P., Owens, F. C., Wade, A. C. Shmulsky, R., Wiedenhoeft, A. C. (2022). Towards sustainable North American wood product value chains, Part I: Computer vision identification of diffuse porous hardwoods. *Front. Plant Sci*. 12:104536. doi: 10.3389/fpls.2021.758455 (Republished with permission)

3.1 Abstract

Availability of and access to wood identification expertise or technology is a critical component for the design and implementation of practical, enforceable strategies for effective promotion, monitoring and incentivization of sustainable practices and conservation efforts in the forest products value chain. To address this need in the context of the multi-billion-dollar North American wood products industry 22- class, image-based, deep learning models for the macroscopic identification of North American diffuse porous hardwoods were trained for deployment on the open-source, field-deployable XyloTron platform using transverse surface images of specimens from three different xylaria and evaluated on specimens from a fourth xylarium that did not contribute training data. Analysis of the model performance, in the context of the anatomy of the woods considered, demonstrates immediate readiness of the technology developed herein for field testing in a human-in-the-loop monitoring scenario. Also proposed are

strategies for training, evaluating, and advancing the state-of-the-art for developing an expansive, continental scale model for all the North American hardwoods.

3.2 Keywords

Wood identification, illegal logging and timber trade, XyloTron, computer vision, machine learning, deep learning, diffuse porous hardwoods, sustainable wood products

3.3 Introduction

North American hardwoods are utilized in a multitude of applications including furniture (household, office, and institutional), construction and remodeling (e.g., flooring, millwork, and kitchen cabinets), and industrial products (e.g., pallets, access mats, and crossties). In 2016, the total US output1 of hardwood products was US$135.6 billion including US$39.8 billion in exports (Hardwood Federation, 2016). Proper identification of hardwoods along this value chain is essential for ensuring that contractual obligations have been met, detecting and preventing commercial fraud (Wiedenhoeft et al., 2019), determining appropriate drying schedules (Simpson, 1991), deciding on suitable methods of chemical treatment (Kirker and Lebow, 2021), and assessing the condition of in-service structures (Ross and White, 2014). Whether in the context of in-service wood or new wood-based products, identification of the material is germane both in an engineering context, and in terms of interrogating or verifying claims of legality and/or sustainability of the wood in a final product. Material identification is a necessary requirement for the design of practical strategies for designing, monitoring, and incentivizing sustainable wood product value chains.

Legality and sustainability of wood and wood-based products are two disparate concepts, the former being a matter of jurisdiction and legislation and thus essentially referring to de facto

claims or criteria, whereas the latter is a topic of scholarly, practical, economic, and environmental debate (Giovannoni and Fabietti, 2013; Magnus Boström et al., 2015). For wood and woodbased products, legality can be governed by international treaties (e.g., the Convention on the International Trade in Endangered Species of Flora and Fauna [CITES, 27 U.S.T. §1087]) and by national laws and policies (e.g., the United States' Lacey Act [18 U.S.C. §42-43; 16 U.S.C. §3371-3378]) and wood identification can play a critical role in enforcement. Sustainability is a more elusive concept and legitimate disagreements as to what constitutes sustainability can occur between otherwise similarly minded parties (Miller and Bush, 2015; Ruggerio, 2021). In addition to the conceptual or theoretical differences that may exist between the principles and details subtending sustainability criteria, there is also the question of real-world implementation and enforcement of sustainability measures along supply chains (Bush et al., 2015; Chappin et al., 2015; Dieterich and Auld, 2015) to ensure that a product labelled as sustainable is in fact sustainably sourced. Confirming the sustainability of a consumer product may not be possible by testing the final product, but rather may depend more upon the supply chain and sustainability regime employed to produce and guarantee that product claim. Disproving sustainability, however, can sometimes happen readily by testing consumer products, for example by determining that the wood used in a product is from a threatened or protected species (Wiedenhoeft et al., 2019), or from a region with a high overall prevalence of unmanaged forest harvest. For establishing claims of legality and sustainability for wood-products there is a critical need for developing and scaling wood identification capacity.

Presently, wood identification is primarily performed by wood anatomy experts who have spent months or years training to acquire this skill; who typically carry out this function in a laboratory setting; and whose accuracy depends on the ability to recognize and distinguish a

wood specimen's anatomical features and interpret them in the context of established methods (e.g., dichotomous keys, multiple entry keys, comparison to reference specimens) for wood identification (Wheeler and Baas, 1998). Despite the efficacy of such human-based anatomical identification, trained experts are rare, competence varies, and overall capacity for this task in the United States (Wiedenhoeft et al., 2019)–and presumably globally–is critically limited. For example, respondents to the proficiency test in Wiedenhoeft et al. (2019), when confronted with US domestic woods, demonstrated in-laboratory accuracies (with access to the full gamut of traditional wood identification resources such as light microscopy, reference specimens, keys, online resources, etc.) ranging from as low as 7% of the 28 specimens to as high as 86%-when considering only the specimens attempted, accuracies ranged from 25 to 92% (Wiedenhoeft et al., 2019). There is the expectation that macroscopic field identification would achieve substantially lower accuracies (Wiedenhoeft, 2011; Ruffinatto et al., 2015).

To overcome the dearth of human expertise in wood identification, various teams have developed computer vision-based systems which can be implemented in the laboratory or in the field (Khalid et al., 2008; Martins et al., 2013; Filho et al., 2014; Figueroa-Mata et al., 2018; Ravindran et al., 2018, 2019, 2021; Damayanti et al., 2019; de Andrade et al., 2020; Ravindran and Wiedenhoeft, 2020; Souza et al., 2020). Even with microscopic inspection and complete access to reference collections, human based wood identification is typically accurate only to the genus level with reliable species-level identification being rare (Gasson, 2011). Machine learning, on the other hand, either alone (Martins et al., 2013; Filho et al., 2014; Barmpoutis et al., 2017; Kwon et al., 2017, 2019; Rosa da Silva et al., 2017; Figueroa-Mata et al., 2018; Ravindran et al., 2018, 2019, 2020, 2021; de Geus et al., 2020; Hwang et al., 2020; Ravindran and Wiedenhoeft, 2020; Souza et al., 2020; Fabija´nska et al., 2021) or in combination with

human expertise (Esteban et al., 2009, 2017; He et al., 2020), has shown promise that species-level identification might be possible, when the woods in question allow resolution at this granularity. Recent work involving the open-source XyloTron platform (Ravindran et al., 2020) has shown promise for real-time, field-deployable, screening-level wood identification (Ravindran et al., 2019, 2021; Ravindran and Wiedenhoeft, 2020; Arévalo et al., 2021) with the hardware to transition to smartphone-based systems now available (Tang et al., 2018; Wiedenhoeft, 2020). Affordability and democratization make computer vision wood identification (CVWID) an attractive technology for robust, multi-point monitoring of the full sustainable wood products value chain from producers to consumers. While multiple platforms for imaging biological specimens in natural history collections are available (e.g., Hedrick et al., 2020; Pearson et al., 2020; von Baeyer and Marston, 2021), it should be noted that the XyloTron, XyloPhone, and similar systems for CVWID have been designed for affordability, field screening, human-in-the-loop deployment, and also have the potential (especially given the comparative affordability of the XyloPhone system) for crowd-sourcing data collection, citizen-science efforts (Goëau et al., 2013), and use in secondary education, all of which have the potential to enrich image datasets if images can be vetted and curated.

Putting forth a field-deployable computer vision model for the identification of commercially important North American hardwoods requires on the order of 50 classes, which far exceeds anything published to date for this region, either at the naked eye level (Wu et al., 2021) or using macroscopic images (Lopes et al., 2020). Increasing the number of classes in a model has the potential to influence model accuracy (Bilal et al., 2018; Shigei et al., 2019), and unpublished work on the expansion of a 15-class Ghanaian timber model (Ravindran et al., 2019), using the same model training methodology, to 39 and 43 classes showed a reduction in

model accuracy. While these data might suggest a negative relationship between number of classes and accuracy, the literature does not provide consensus on how increasing the number of classes impacts the performance of classification models. Abramovich and Pensky (2019) suggest that increasing the number of classes could positively influence model accuracy while other sources suggest, in general, an inverse relationship (e.g., Bilal et al., 2018; Shigei et al., 2019). Whether additional classes improve or reduce model accuracy undoubtedly depends on multiple factors including the degree to which the additional classes are similar to each other and to those already in the model. Greatly increasing the number of classes is presumed to have a non-trivial effect on model accuracy; thus, larger multiclass models should be handled with care, paying close attention to factors that might negatively impact model performance. An option for building practical, high performing models with a large number of classes is to leverage domain-based factors for informed model selection, label space design, and filtering of the model predictions, thus taking advantage of human expertise in determining the breadth and scope of the model implementation, evaluation, and deployment.

In the case of North American hardwoods, one such factor, commonly used for human-based macroscopic identification, that could affect accuracy might be wood anatomical spatial heterogeneity as it relates to porosity (IAWA, 1989; Ruffinatto et al., 2015). Classically ring-porous woods exhibit large and abrupt differences in vessel diameter and often in parenchyma patterns between earlywood and latewood. In addition, the macroscopic appearance of vessel and parenchyma patterns in the latewood can vary greatly among specimens exhibiting slow growth, medium growth, and fast growth. In cases of fast-grown ring-porous specimens, the growth rings can be so wide that images captured at the macroscopic level might include nothing but latewood, completely excluding earlywood features important for identification. This greater

spatial heterogeneity of ring-porous woods contrasts with the lesser spatial heterogeneity of classically diffuse-porous woods, which exhibit little macroscopic anatomical variation both between and within growth rings regardless of variations in radial growth rate. As shown in Figure 3.1, the radial growth rate of a ring-porous wood imparts greater spatial heterogeneity at the macroscopic scale (Figures 3.1 B, D, F) compared to the lower spatial heterogeneity of a diffuse-porous wood growing at similar radial growth rates (Figures 3.1 A, C, E).

This study presents the design and implementation of 22-class deep learning models for image-based, macroscopic identification of North American diffuse porous hardwoods. The main highlights of this study include:

- Providing the first continental scale model for the identification of an important set of North American hardwoods, which is the largest wood identification model reported across all available wood identification technologies (Schmitz et al., 2020);
- Reporting on the first multi-site, multi-operator, multi-instantiation study of computer vision identification for North American woods that has been evaluated using a practical field-testing surrogate (Ravindran et al., 2020);
- Using wood anatomy-driven label space design (the grouping and partition of species into classes) and model performance evaluation;
- Establishing a strong baseline using a simple machine learning methodology for the quantitative comparison of advances in wood identification across all modalities; and,
- Discussing practical strategies for field-testing and model deployment for empowering sustainability and conservation efforts in wood product value chains.

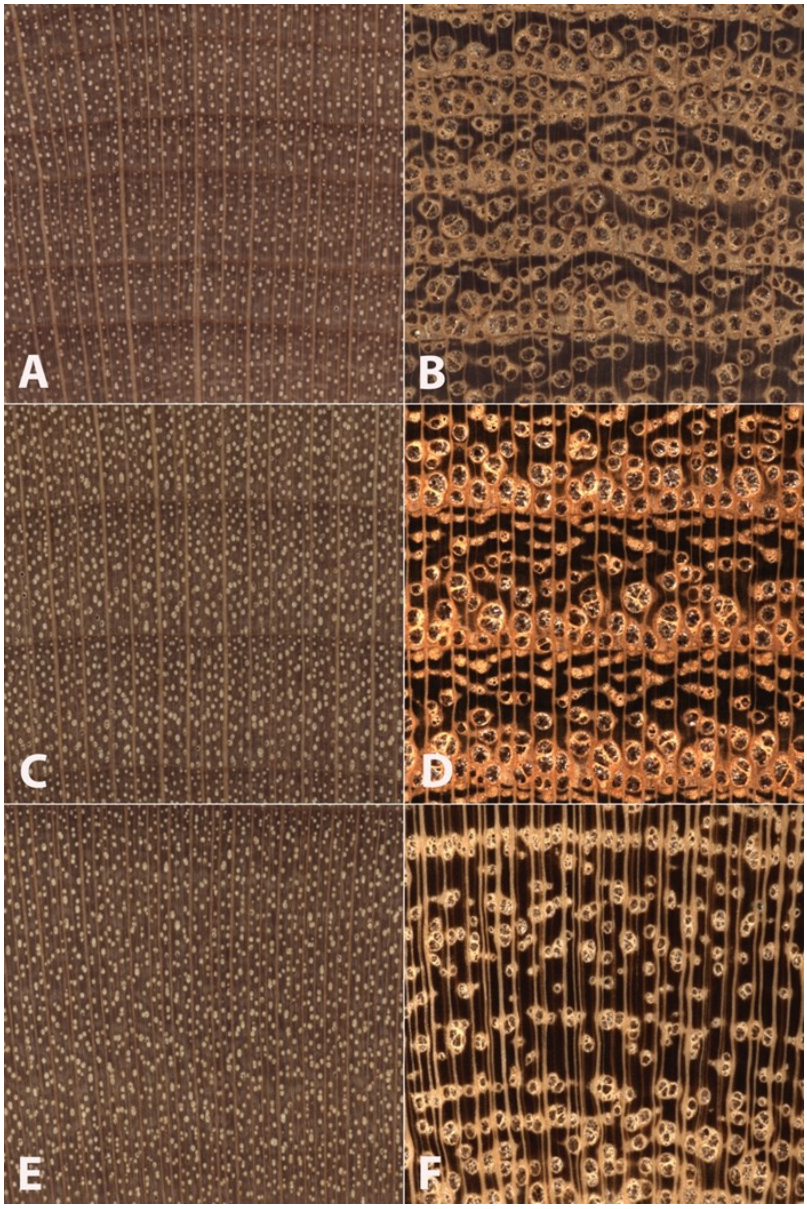

Figure 3.1 Images of transverse surfaces of *Betula alleghaniensis* (A, C, E) *and Robinia pseudoacacia* (B, D, F) showing similar slow-growth conditions (A, B) medium-growth conditions (C, D), and faster-growth conditions (E, F).

Note: *Betula alleghaniensis* shows comparatively lesser wood anatomical spatial heterogeneity than *Robinia pseudoacacia*. The nearly three complete growth rings in C and D present wood anatomical detail sufficient to facilitate an identification. The slow growth in A and B and partial growth rings in E and F demonstrate the comparatively lesser spatial heterogeneity of the diffuse porous *Betula alleghaniensis.* In *Robinia pseudoacaia* there a is lack of latewood characters in the slow-grown image (B), and only latewood anatomy in F. By contrast, *Betula alleghaniensis* shows substantially similar anatomy across the three images (A, C, E).

3.4 Materials and Methods

3.4.1 Dataset Details

3.4.1.1 Taxa and Sample Selection

105 unique species from 24 prominent genera of North American diffuse porous woods were selected based on the commercial importance and specimen availability among four scientific wood collections. The four wood collections and details of their specimen contributions are summarized in Table 3.1.

Table 3.1 The four xylaria providing wood specimen images for the data sets used to train and test the wood identification models.

Institution (Xylarium acronym)	Specimen counts	Role
USDA Forest Products Laboratory, Madison collection (MADw)	410	Model Training
USDA Forest Products Laboratory, Samuel J. Record collection (SJRw)	77	Model Training
Royal Museum of Central Africa (Tw)	17	Model Training
Mississippi State University (PACw)	284	Model Testing

Note: The MADw, SJRw and Tw specimens contributed images exclusively to the training data set, while the test data set was obtained from only the PACw specimens.

3.4.1.2 Sample Preparation and Imaging

The transverse surfaces of 788 wood specimens from the selected taxa were progressively sanded from coarse to fine grit (240, 400, 600, 800, 1000, 1500) with dust removal from cell lumina using compressed air and adhesive tape when possible. The prepared surfaces were imaged using multiple instantiations of the XyloTron system (Ravindran et al., 2020) to produce

a data set with 6393 non-overlapping images. The 2048 x 2048-pixel images obtained with the XyloTron had a linear resolution of 3.1 microns/pixel and each image shows 6.35 mm 6.35 mm of tissue. The sample preparation and image collection were done by multiple operators with varying levels of wood anatomy expertise and specimen preparation experience (undergraduate students, graduate students, postdoctoral researchers, and technical specialists). A summary of the collected dataset is provided in Table 3.2.

Table 3.2 Image data set summary.

	Training (counts)	**Testing (counts)**	**Total (counts)**
Number of xylaria	3	1	4
Number of taxa	98	69	105*
Number of specimens	504	284	788
Number of images	5184	1209	6393

Note: 788 specimens from 105 unique taxa (belonging to 24 genera) were prepared and imaged to produce 6393 images for training and testing the classification models. *The total number of taxa does not equal the sum of the training and testing counts as not all species comprising each class were present in both the training and testing data sets.

3.4.1.3 Label Assignment

Wood identification is typically accurate only to the genus level when the full gamut of light microscopic characters is employed (Gasson, 2011). For the taxa in this study, a combination of supra-generic, generic, and sub-generic granularity for classification is appropriate for macroscopic wood identification. To facilitate machine learning, the taxa were grouped into 22 classes based on their macroscopic anatomical similarity in the following manner:

1. The genera *Aesculus, Alnus, Arbutus, Betula, Carpinus, Fagus, Frangula, Liquidambar, Liriodendron, Magnolia, Nyssa, Ostrya, Oxydendrum, Platanus, Populus, Rhamnus, Salix*, and *Tilia* were assigned to 18 genus-level classes (with genus names as labels).
2. The genus *Acer* was split into two classes, "hard" and "soft," with labels "AcerH" and "AcerS," respectively, as within North American Acer, hard maple (*A. saccharum*) is separable from the soft maples (e.g., *A. macrophyllum, A. saccharinum, A. rubrum*) based on ray widths observed macroscopically and microscopically (Panshin and de Zeeuw, 1980; Hoadley, 1990).
3. Species from the genera *Crataegus, Malus, Prunus, Pyrus, and Sorbus* were grouped into one class, with the label "Fruitwood," with the exception of *Prunus serotina* which was its own class with the label "Prunus" as *P. serotina* is wood anatomically distinct from the other fruitwoods.

A listing of the 105 taxa, their class labels and their training/testing set membership can be found in Appendix A.2.1 (Supplementary Material 3).

3.4.2 Machine Learning Details

3.4.2.1 Model Architecture and Training

While multiple deep learning architectures for image classification exist (e.g., Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015; Huang et al., 2017), we employed a convolutional neural network (CNN; LeCun et al., 1989) with a ResNet34 (He et al., 2016) backbone and a custom 22-class classifier head (see Figure 3.2), based on prior success using this architecture for wood identification (e.g., Ravindran et al., 2019, 2021). The CNN backbone was initialized with ImageNet (Russakovsky et al., 2015) trained weights and He

weight initialization (He et al., 2015) was employed for the custom classifier head. In the first stage of training, the backbone weights were frozen, and the weights of the custom head were optimized. The weights of the entire network were fine-tuned during the second training stage. For both the stages, the Adam optimizer (Kingma and Ba, 2015) with a two-phase simultaneous cosine annealing (Smith, 2018) of the learning rate and momentum was employed. Each mini-batch (of size 16) was composed of 2048 x 768 pixel random image patches extracted from each of 16 images, down-sampled to 512 x 192 pixels, randomly augmented using horizontal/vertical flips, small rotations, and cutout (Devries and Taylor, 2017), and input to the network. Complete details about the architecture and the adopted two-stage (Howard and Gugger, 2020) transfer learning (Pan and Yang, 2010) training methodology can be found in Ravindran et al. (2019) and Arévalo et al. (2021). Models with a ResNet50 backbone were also trained and evaluated, with the results presented in Appendix A.2.2 (Supplementary Material 4). Scientific Python tools (Pedregosa et al., 2011) and the PyTorch deep learning framework (Paszke et al., 2019) were used for model definition, training, and evaluation.

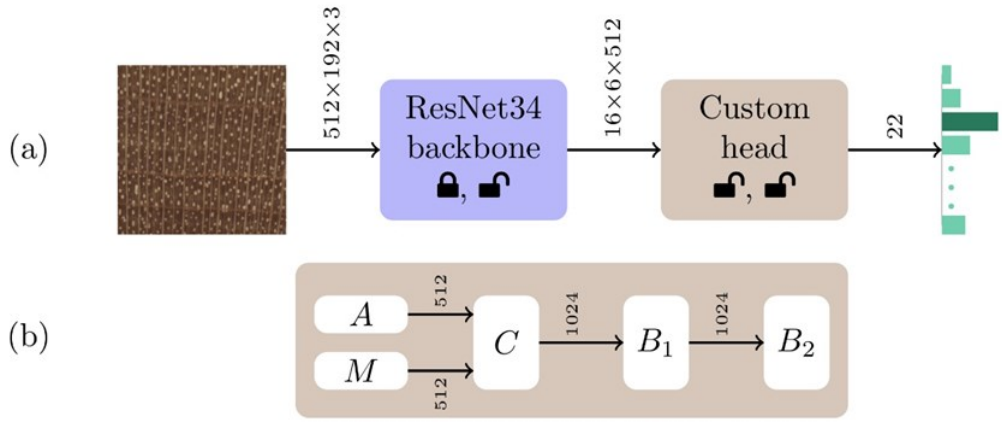

Figure 3.2 Model schematic of the CNN architecture.

Note: (a) The CNN architecture for our 22-class wood identification models consisted of a ResNet34 backbone with a custom classifier head. The custom head shown in (b) is comprised of global average (A) and max (M) pooling (Goodfellow et al., 2016) layers that are concatenated (C) to form a 1024-vector. This is followed by two fully-connected blocks (B1, B2) each with batchnorm (Ioffe and Szegedy, 2015) and dropout (Srivastava et al., 2014) layers. The dropout layers had parameters p=0.5 and p=0.25 in the B1 and B2 blocks, respectively. ReLU activation was used in B1, while B2 had a softmax activation. The status of the weights of the backbone and custom head, whether they are modified or not during the two stages of training, are represented by the lock and unlock symbols respectively.

3.4.2.2 Model Evaluation

The predictive performance of the trained models was evaluated using specimen level top-k accuracies with k = 1 and k = 2. The top-1 prediction for a specimen was the majority of the class predictions for the images contributed by the specimen. The top-2 prediction for a specimen was obtained by equally weighted voting of the top-2 image level predictions for the images contributed by the specimen and the specimen was considered correctly identified if its true class was one of the top-2 predicted classes. The specimen level top-1 and top-2 performance of the trained models were evaluated using fivefold cross-validation (5184 images from 504 specimens; MADw, SJRw, and Tw collections) and an independent test set (1209 images from 284 specimens; PACw collection). The PACw images: (i) were obtained by a different operator using a different instantiation of the XyloTron, (ii) were not used to train the

field or cross validation models, and (iii) serve as a valid, practical proxy for real field testing (Ravindran et al., 2021). Each PACw specimen contributed up to five images for evaluation and this maximum number of images per specimen was fixed before any model evaluation was performed i.e., the number of images per PACw test specimen was not tuned. Specifically, the following analyses were performed:

1. Five-fold cross-validation analysis was performed with label stratified folds and specimen level separation between the folds i.e., each specimen contributed images to exactly one fold. Specimen level mutual exclusivity between the folds is necessary for the valid evaluation of any machine learning based classifier for wood identification (e.g., Ravindran et al., 2019, 2020, 2021 and as discussed in Hwang and Sugiyama, 2021). Model predictions over the five folds were aggregated to compute the (top-1) prediction accuracy and a confusion matrix.
2. The (mean) top-1 and top-2 predictive performance of the five trained models from the cross-validation analysis on the PACw data was computed. It should be noted that each of the five models was trained on four folds (80%) of the training data.
3. All the images from the cross-validation analysis (i.e., 100% of the training data) were used to train a separate model (field model) which was then evaluated on the independent PACw data. The top-1 and top-2 prediction accuracy and the confusion matrix were computed to evaluate the efficacy of the field model.

3.4.2.3 Misclassified Specimens

All images of the misclassified specimens in the five-fold cross validation model and field model were evaluated and reported as in Ravindran et al. (2021), assigning each to one of

three types of misclassification: (1) taxa were anatomically consistent, and the test specimen was typical; (2) the individual test specimen was atypical for the taxon (i.e., it is not an archetypal specimen for the taxon); or, (3) the taxa and test specimen were anatomically typical, but the classes are not anatomically consistent with each other, and errors of this type would not be expected to be made by a human identifier. It is important to note that these attributions are made on a specimen basis, so while Types 1 and 3 are mutually exclusive, the remaining combinations are possible (e.g., class A misclassified as class B with 5 such misclassifications could show all Type 1, all Type 2, all Type 3, combinations of Types 1 and 2 or Types 2 and 3, but never a combination of Type 1 and Type 3).

3.5 Results

The specimen level prediction accuracies for the cross-validation and field models are presented in Table 3.3. While the cross-validation accuracy was 95.2%, the (mean) top-1 and top-2 accuracies were 73.5 and 85.1%, respectively, when the models were tested on the PACw test specimens. The top-1 accuracy of the field model was 80.6%, and the top-2 accuracy was 90.5%. Figures 3.3, 3.4 display the confusion matrices for the cross-validation (accumulated over the five folds) and field models, respectively.

Table 3.3 Specimen level model prediction accuracies.

Training and evaluation details	**Top-k**	**Accuracy (%)**
Five-fold cross-validation	k=1	**95.2**
Trained using four folds, tested on PACw*	k=1	73.5
	k=2	85.1
Field model trained using all five folds, tested on PACw	k=1	**80.6**
	k=2	90.5

Note: *The mean top-1 and top-2 prediction accuracies over the five models are reported with the standard deviations 4.5% and 4.1% respectively. Accuracies in bold are those for which a confusion matrix is provided.

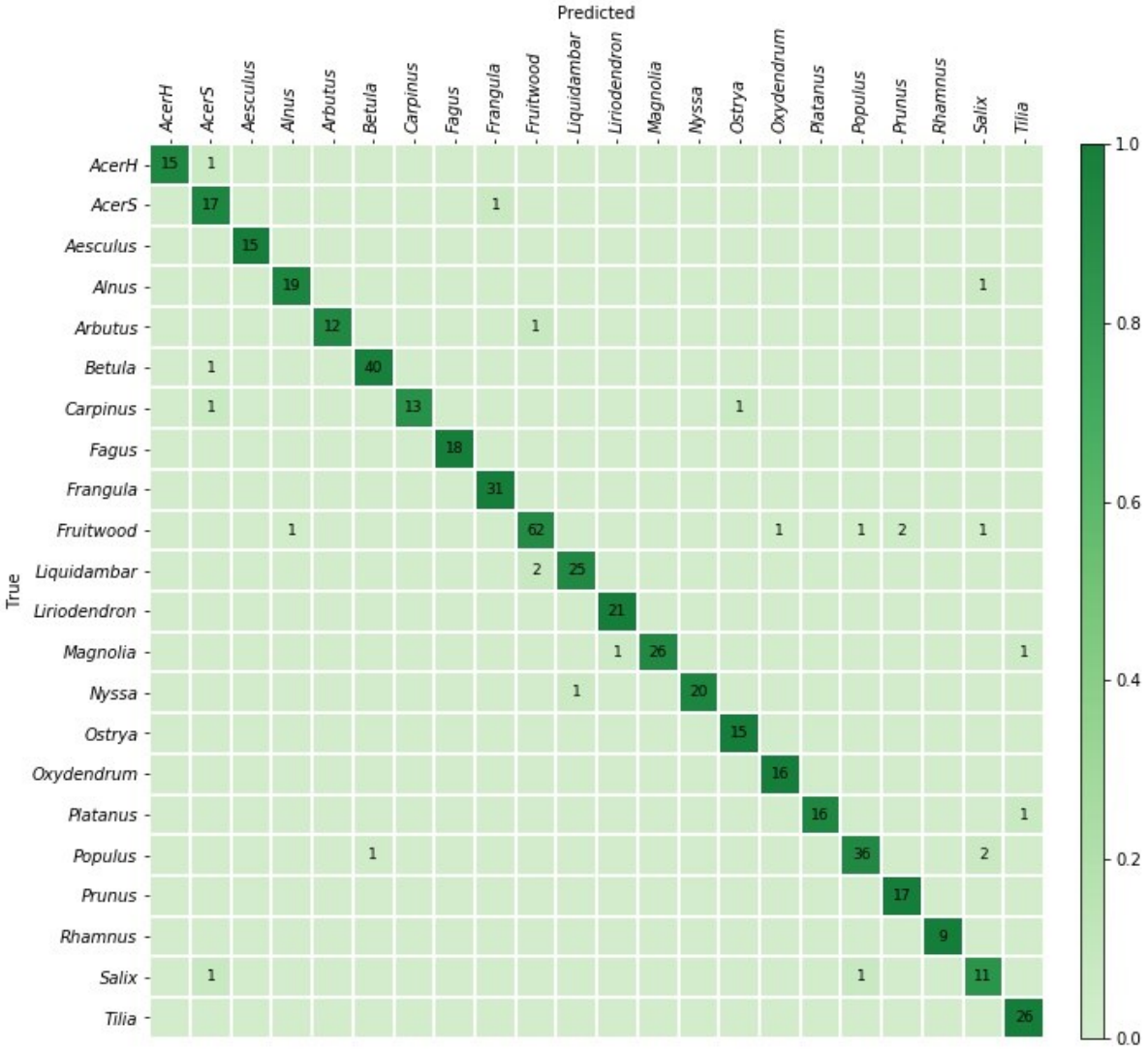

Figure 3.3 Confusion matrix for the cross-validation model predictions on 504 specimens.

Note: The specimen-level top-1 prediction accuracy accumulated over the five folds was 95.2%.

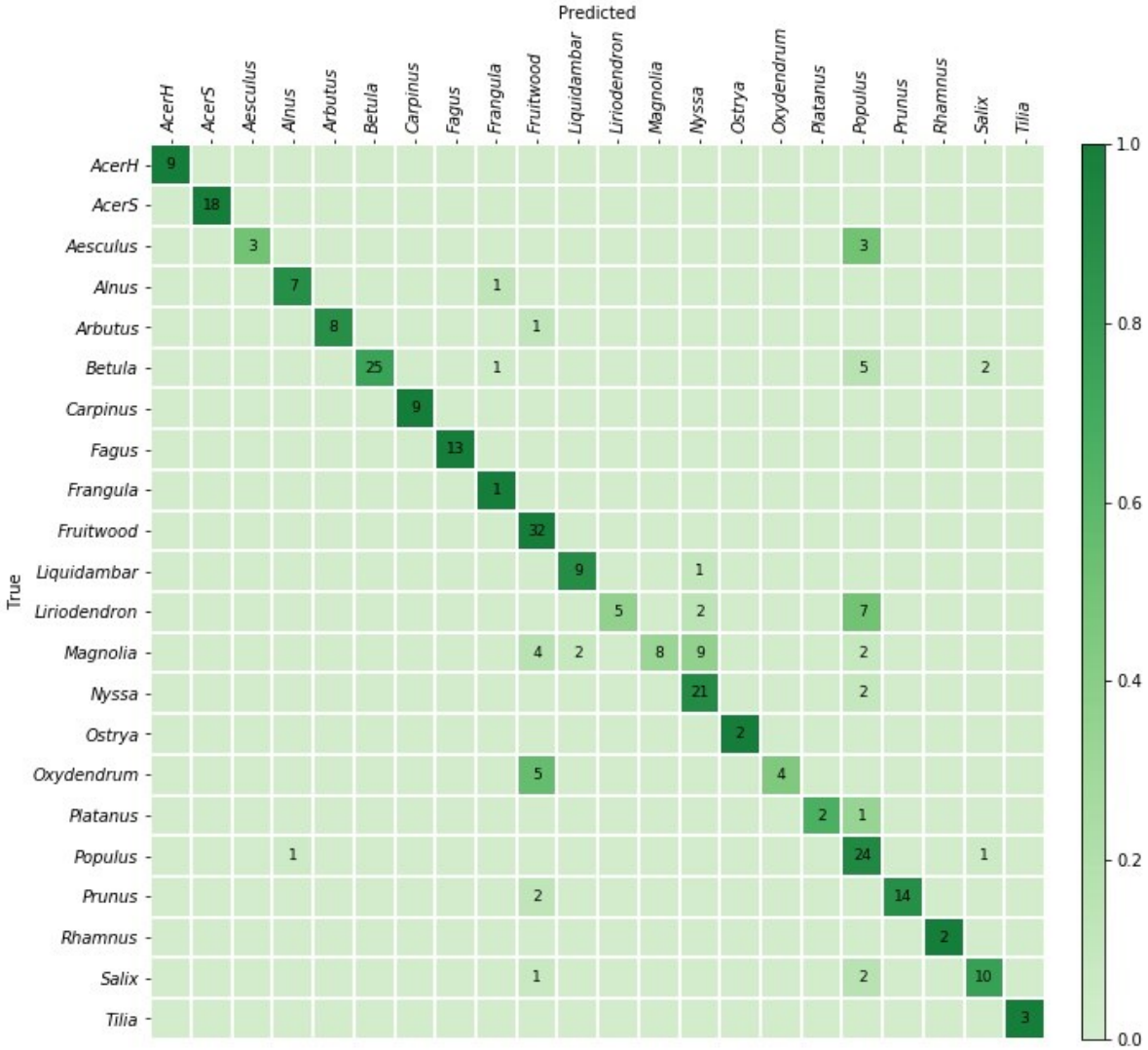

Figure 3.4 Confusion matrix for the field model predictions on 284 PACw specimens.

Note: The top-1 and top-2 specimen-level accuracies were 80.6% and 90.5%, respectively.

Figure 3.5 presents example images of Type 1, Type 2, and Type 3 misclassifications, and summary of misclassification data for both the five-fold cross-validation model and the field model are presented in Table 3.4.

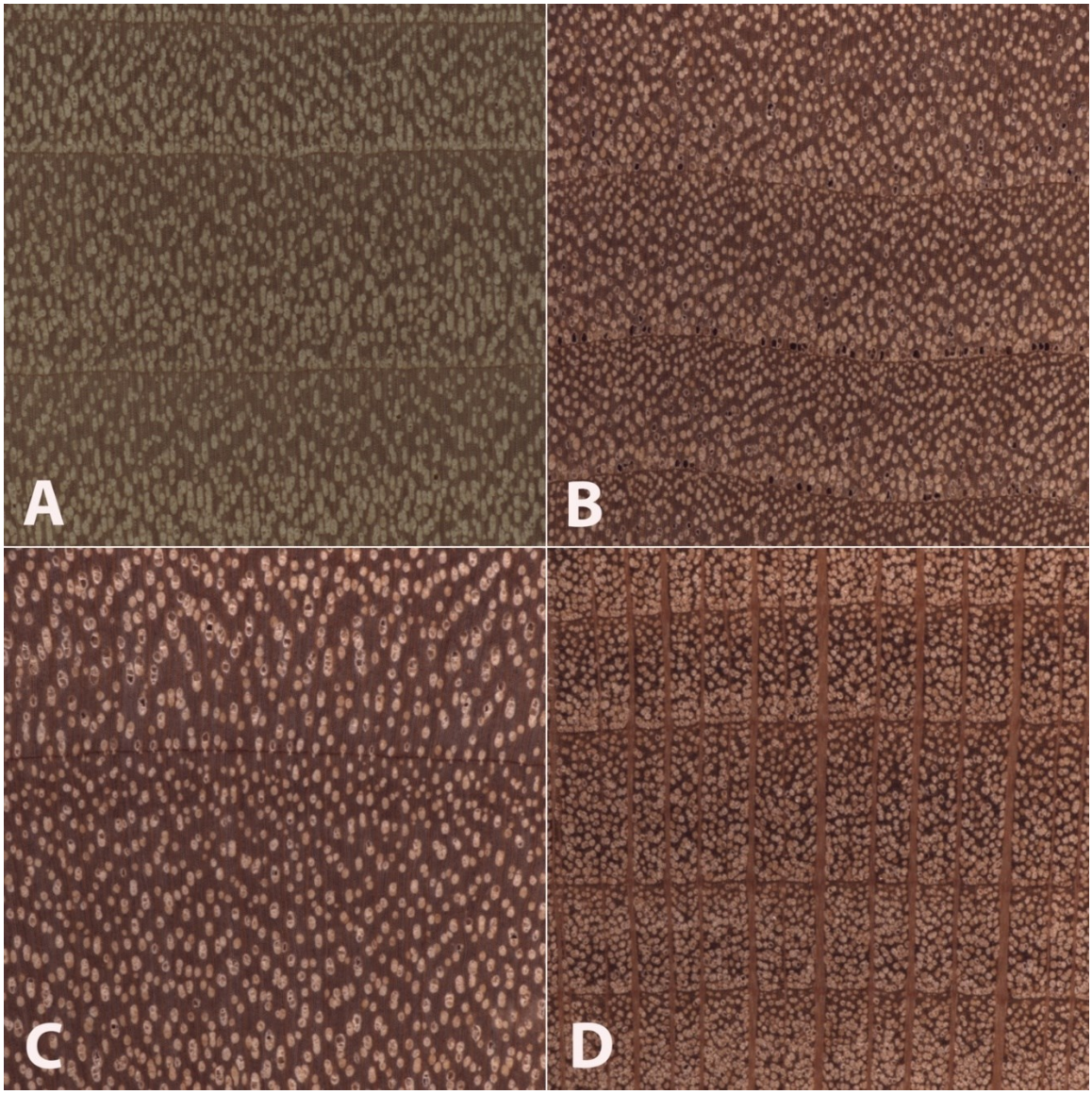

Figure 3.5 Images of the transverse surface of test specimens (B, C, D) and an exemplar (A) of the class (Populus).

Note: All images are 6.35mm on a side. An anatomically representative specimen of Salix scouleriana (B) was misclassified as the wood anatomically similar class Populus (A), a Type 1 misclassification. An anatomically atypical specimen of Betula nigra (C) was classified as (A), a Type 2 misclassification. An anatomically typical specimen of Platanus occidentalis (D) was misclassified as the anatomically disparate class (A), a Type 3 misclassification. Note the anatomical similarities between A and B, and to a lesser extent A and C, and the anatomical dissimilarity between A and D, especially with regard to the wide rays in D.

Table 3.4 Number and proportion of misclassified specimens from Figure 3.4 by type of misclassification.

Misclassification type	**Number of misclassified specimens**	**Proportion of 55 misclassified (of 284 total) specimens**
Taxa are anatomically consistent, test specimen typical (Type 1)	34	0.618 (0.12)
Test specimen atypical for its taxon* (Type 2)	10	0.182 (0.035)
Taxa and test specimen are not anatomically consistent (Type 3)	11	0.20 (0.039)
Total	55	1.0 (0.194)

Note: Types 1 and 2 are consistent with wood anatomy and are expected errors made by human field inspectors. Type 3 errors are inconsistent with macroscopic wood anatomy and would not be expected to be made by a human inspector.

When considering top-1 accuracy of the field model, 9 classes showed no misclassifications when input into the trained model for field testing with PACw specimens: Acer (hard), Acer (soft), Carpinus, Fagus, Frangula, Fruitwood, Ostrya, Rhamnus, and Tilia, with the other 13 classes showing at least one specimen misclassification (Figure 3.4). Of the 55 misclassified specimens, 80% were Type 1 or Type 2 misclassifications, with only 20% being anatomically inconsistent (Type 3) misclassifications (Table 3.4). While specimens from 13 classes were misclassified, they were attributed only to 7 classes: Alnus, Frangula, Fruitwood, Liquidambar, Nyssa, Populus, and Salix (Figure 3.4). Seven classes neither contributed nor drew misclassifications: Acer (hard), Acer (soft), Carpinus, Fagus, Ostrya, Rhamnus, and Tilia.

3.6 Discussion

For a field-deployable image-based CVWID model for North American diffuse porous hardwoods to make the greatest real-world impact in law enforcement, industrial compliance, and supply chain verification, it is critical to establish the ways in which the model succeeded in identifying the woods and to dissect the ways in which it failed. Prior work in the field of CVWID has largely limited its analysis of results to reports of overall model accuracy (e.g., Martins et al., 2013; Filho et al., 2014; Rosa da Silva et al., 2017; Figueroa-Mata et al., 2018; Ravindran et al., 2019; de Geus et al., 2020; Souza et al., 2020) with comparatively little prior work addressing wood anatomical details of the misclassifications (Lens et al., 2020; Ravindran et al., 2021). More detailed analyses of the types of misclassifications can yield insights that improve the state-of-the-art in the performance and interpretability of CVWID technologies.

3.6.1 Accuracy of Cross-Validation and Field Models

Top-1 cross-validation accuracy (Table 3.3, row 1) was ~22 points higher than when the same fivefold models were tested with the PACw specimens (Table 3.3, row 2). The increase in top-1 performance of the field model (trained on 100% of the training data) when compared to the five-fold models trained on 80% of the data suggests that the wood anatomy variability captured within the full training dataset contributes to a field model with better predictive power. Moreover, this suggests that the wood anatomical data space may not have been fully represented by 80% of the data, and in fact even the field model (trained with 100% of the data) may not fully represent the wood anatomical data space. One contributor to a richer data space is provision of a representative and robust selection of specimens from which images can be captured. The question of how top-k specimen level accuracy varies with the number of image

level predictions used to compute the specimen level prediction is an open problem [but see Supplementary Material 4 in Appendix for the impact of the number of images per specimen (1–5) on model prediction accuracy], but certainly should be informed by deployment context and the wood anatomy of classes in the model. Top-k accuracy can also be informative in a field-deployed CVWID system when done in a human-in-the-loop context where a human user can make a visual comparison of the unknown to reference images of the top-k predictions. Here the number of image-level predictions used to derive a specimen level prediction was fixed *a priori*, but for a practical system this should be informed by model calibration (Niculescu-Mizil and Carauna, 2005; Guo et al., 2017), inter- and intra-class anatomical variability of the woods in the model (Ravindran et al., 2018), and probably adaptively based on predictions being performed.

3.6.2 Analysis of Misclassifications

When considering a confusion matrix (e.g., Figure 3.4), the off-diagonal results are misclassifications, and can further be evaluated as the propensity for an input class to be misclassified, and/or the propensity for a predicted class to pull or draw misclassifications, each of which can display any of the three misclassification types (1, 2, 3), or combinations thereof, excluding Type 1 + Type 3, as they are mutually exclusive. To codify this concept, the terms "source" and "sink" misclassifications are introduced, where the input misclassified specimens are sources (i.e., the sum of the off-diagonal predictions for each row), and the classes that draw misclassifications are sinks (i.e., the sum of the off-diagonal predictions for each column). For example, in a confusion matrix with four classes A, B, C, and D (Figure 3.6), the on-diagonal cells (e, j, o, t) are correct predictions. For class B, i + k + l would be the source misclassifications, and f + n + r would be its sink misclassifications. If classes A and B were anatomically similar, source misclassification f and sink misclassification i would both be Type 1

misclassifications. If A and C were anatomically disparate, source misclassification g and sink misclassification m would both be Type 3 misclassifications. The anatomical characteristics of the classes and test images therefore determine which type of misclassification is found in each cell, and this finer grained analysis of the misclassifications may assist in designing cost-aware loss functions for improved training (Elkan, 2001; Chung et al., 2016) in the future, making more robust inferences about model performance, and possibly using these insights to inform protocols for real-world model deployment.

		Predicted class			
		A	**B**	**C**	**D**
True class	**A**	e	f	g	h
	B	i	j	k	l
	C	m	n	o	p
	D	q	r	s	t

Figure 3.6 Example 4-class confusion matrix, with classes A, B, C, and D.

Note: Correct predictions are on the main diagonal (e, j, o, t, shown shaded) and off-diagonal cells are the misclassifications. Sums of off-diagonal elements along a row (column) are the source (sink) misclassifications for the class.

Table 3.5 presents a summary of the analysis of source/sink misclassifications by the field model for the confusion matrix in Figure 3.4. With regard to source misclassifications, it is noteworthy that in three of thirteen classes with misclassifications–Aesculus, Liriodendron, and Magnolia (yellow cells)–half or more of the source specimens are misclassified. Of particular note in source misclassifications is the class Liriodendron (green cell), which accounts for over 63% (7 of 11) of all Type 3 source misclassifications, though it contributes only 14 of 284 (~5%)

specimens to the entire test data set. Of the seven classes showing sink misclassifications, three are responsible for more than 85% - Fruitwood, Nyssa, and Populus (blue cells). Fruitwood is a composite multi-generic class (see Supplementary Material 3 in Appendix) but interestingly contributes no source misclassifications while drawing nearly a quarter of sink misclassifications.

Table 3.5 A class-wise assessment of misclassifications for the top-1 misclassified specimens in the field model.

	Class-wise proportion of all specimens (source) or all misclassified specimens (sink)							
	Source				Sink			
Class (n specimens)	**Type 1**	**Type 2**	**Type 3**	**Total**	**Type 1**	**Type 2**	**Type 3**	**Total**
AcerH (9)								
AcerS (9)								
Aesculus (6)	0.500			0.500				
Alnus (8)		0.125		0.125	0.018			0.018
Arbutus (9)	0.111			0.111				
Betula (33)	0.091	0.091	0.061	0.243				
Carpinus (9)								
Fagus (13)								
Frangula (1)						0.018	0.018	0.036
Fruitwood (32)					0.182	0.036	0.018	0.236
Liquidambar (10)	0.100			0.100	0.036			0.036
Liriodendron (14)		0.143	0.500	0.643				
Magnolia (25)	0.440	0.120	0.120	0.680				
Nyssa (23)	0.043			0.043	0.145	0.018	0.055	0.218
Ostrya (2)								
Oxydendrum (9)	0.111			0.111				
Platanus (3)			0.333	0.333				
Populus (26)	0.038			0.038	0.182	0.091	0.127	0.400
Prunus (16)	0.063			0.063				
Rhamnus (2)								
Salix (13)	0.077			0.077	0.018	0.018	0.018	0.055
Tilia (3)								

Note: Source misclassification proportions are the based on the total number of input specimens (n=284). Sink misclassification proportions are based on the total number of misclassified specimens (n=55). Dark grey indicates a class for which there were neither source nor sink misclassifications; light grey indicates the absence of misclassifications in either source or sink; colored cells are proportions of note and are discussed in the text.

The inter-class variability is largely limited to variations in the vessels and the rays, as the diffuse-porous North American woods we included have comparatively limited macroscopically visible variation in axial parenchyma patterns. In Figure 3.5, the Type 3 misclassification between Populus (A) and Platanus (D) suggests that the model's feature detection is perhaps less sensitive to ray width and abundance than a human identifier would be, as the rays in Platanus are much wider and less numerous than the abundant, uniseriate rays in Populus. A human identifier would be expected to note this distinct difference with little trouble. Similarly, in Figure 3.4, seven Liriodendron are misidentified as class Populus, which would appear to be another instance of the feature detection either failing to detect or the classifier failing to weight the wider rays of Liriodendron sufficient to make a correct classification, an error that would not be expected of human identifier. Tools adapted from research on feature visualization (e.g., Zeiler and Fergus, 2014; Olah et al., 2017; Qin et al., 2018) and model interpretability (e.g., Chen et al., 2020) may enable further understanding of the misclassifications and spur richer methodologies that guide the CNN to emphasize human recognized features.

3.6.3 On Datasets and Architectures for Computer Vision Wood Identification

In this work strict adherence to specimen level splits was maintained to encourage learning of generalizable features (vs. memorizing the dataset) and for model evaluation based on specimen identification which is the desired real-world capability. This practically relevant constraint means that despite combining data from three xylaria at multiple institutions, our dataset is still modest in size–even though we have hundreds of images per class, there are only tens of unique representatives (the specimens) per class. Unlike other datasets (e.g., Horn et al., 2018), images used in CVWID are fully composed of the wood tissue being imaged and do not have a foreground and background. Additionally, for the classes considered in this study the

wood anatomical spatial heterogeneity is low. Given these characteristics of CVWID data, though our ResNet34 based model trained on the modest sized dataset (by sampling random patches with a fixed size) yields a practically useful model, the interplay between inter- and intra-class wood anatomical feature variability, dataset size, architecture depth (or capacity), and hyperparameter optimization is yet largely unexplored (an area that we are actively exploring– Supplementary Material 4 in Appendix provides results for a ResNet50 based model trained with the same epoch budget that suggests that our dataset size may be insufficient to leverage the higher capacity afforded by the deeper ResNet50 architecture).

Unique scientifically collected and properly identified specimens are a limited resource, typically found only in xylaria, many of which are underfunded, effectively closed, or gone altogether, though the World Forest ID project (Gasson et al., 2021) is a noteworthy effort in opposition to this trend. The intent of the open-source XyloTron (Ravindran et al., 2020) and XyloPhone (Wiedenhoeft, 2020) projects is the democratization of CVWID technology to enable research groups across the world to contribute to a frequently updated and globally relevant standardized wood dataset, but finding the resources to establish, curate, and maintain such a repository remains a challenge. Crowdsourcing technology may aid in the construction of such curated datasets but paucity of expertise in vetting non-scientific specimens (Wiedenhoeft et al., 2019) must be adequately addressed to optimally leverage citizen science resources such as Pl@ntNet (Goëau et al., 2013).

3.6.4 Towards Real Field Evaluation

Model evaluation with a surrogate for field testing, i.e., specimens from a xylarium not used for model training, was a first step towards real field testing which is the gold standard for evaluating any wood identification technology. The polished specimens used to train the models

reflect a different surface preparation to what occurs in the field, but prior work with the XyloTron on Ghanaian woods (Ravindran et al., 2019) demonstrated a similar deployment gap (drop in accuracy from the cross validation to field testing results) even though field specimens were prepared by knife-cut of the transverse surface (as described in Wiedenhoeft, 2011). Based on these results with Ghanaian woods, it is expected that the trained models described herein can be deployed effectively in a human-in-the-loop setting for field testing where the top predictions of the model along with exemplar images for the predicted classes are presented to the user for verification of the predictions (e.g., as in the xyloinf interface for the XyloTron platform of Ravindran et al., 2020). To derive maximum insights enabling real deployment, any performance metric must be evaluated in the contexts of taxonomic ambiguity, discriminative anatomical features among the woods, and commercially or practically relevant granularity to facilitate the formulation of practical, useful models. To make best use of such models, strategies for deploying them along wood product value chains to promote sustainability should consider context-specific requirements for each use-case. The performance of our trained models (in cross-validation, surrogate, and future field-testing scenarios) can also serve as a strong baseline for developing and comparing future state-of-the-art models or systems.

3.7 Conclusion

Employing practical, wood anatomy-driven strategies for the development and evaluation of CVWID technologies, we presented the first continental-scale, image-based identification model for North American diffuse porous hard woods. Ongoing work tackles the development of a complementary model for the ring porous North American hardwoods and a unified North American hardwood identification model. Operationalization of CVWID technologies with market-relevant scale will require the rigorous exploration of machine learning architecture and

hyperparameters, model training paradigms, performance evaluation protocols, and evidence-based deployment strategies. This work is a first step towards the realization of such a practical, field-deployable, wood identification technology with the potential to inform and impact strategies for the promotion, monitoring, and monetization of sustainable North American and global wood product value chains, and for enabling biodiversity and conservation efforts.

3.8 Data Availability Statement

The datasets presented in this article are not immediately available, but a minimal data set can be obtained by contacting the corresponding author; the full data set used in the study is protected for up to 5 years by a CRADA between FPL, UW-Madison, and FSC. Requests to access the datasets should be directed to corresponding author.

3.9 Author Contributions

FO and RS provided access to and supervised data acquisition from the PACw test specimens. AWa prepared and imaged the PACw specimens. AWa, FO, and AWi established the wood anatomical scope of the study. PR implemented the machine learning pipelines for the study. PR and AWi conducted data analysis and synthesis. PR, AWi, and FO wrote the manuscript. All authors provided actionable feedback that improved the presentation of the manuscript.

3.10 Funding

This work was supported in part by a grant from the United States Department of State via Interagency Agreement number 19318814Y0010 to AWi and in part by research funding from the Forest Stewardship Council to AWi. PR was partially supported by a Wisconsin Idea Baldwin Grant. The authors wish to acknowledge the support of United States Department of

Agriculture (USDA), Research, Education, and Economics (REE), Agriculture Research Service (ARS), Administrative and Financial Management (AFM), Financial Management and Accounting Division (FMAD) Grants and Agreements Management Branch (GAMB), under Agreement No. 58-0204-9-164, specifically for support of AWa, FO, and RS.

3.11 Acknowledgments

The early support and consultation provided by Robert J. Ross for this study is gratefully acknowledged. We also wish to thank Nicholas Bargren, Karl Kleinschmidt, Caitlin Gilly, Richard Soares, Adriana Costa, and Flavio Ruffinatto for specimen preparation and imaging efforts.

3.12 References

Abramovich, F., and Pensky, M. (2019). Classification with many classes: challenges and pluses. *J. Multivariate Anal*. 174:104536. doi: 10.1016/j.jmva.2019.104536

Arévalo, R., Pulido, E. N. R., Solórzano, J. F. G., Soares, R., Ruffinatto, F., Ravindran, P., et al. (2021). Image based identification of Colombian timbers using the XyloTron: a proof of concept international partnership. *Colombia Forestal* 24, 5–16. doi: 10.14483/2256201X.16700

Barmpoutis, P., Dimitropoulos, K., Barboutis, I., Nikos, G., and Lefakis, P. (2017). Wood species recognition through multidimensional texture analysis. *Comput. Electron. Agric*. 144, 241–248. doi: 10.1016/j.compag.2017.12.011

Bilal, A., Jourabloo, A., Ye, M., Liu, X., and Ren, L. (2018). Do convolutional neural networks learn class hierarchy? IEEE *Trans. Visu. Comput. Graph*. 24, 152–162. doi: 10.1109/TVCG.2017.2744683

Bush, S. R., Oosterveer, P., Bailey, M., and Mol, A. P. J. (2015). Sustainability governance of chains and networks: a review and future outlook. J. Clean. Prod. 107, 8–19. doi: 10.1016/j.jclepro.2014.10.019

Chappin, M. M. H., Cambré, B., Vermeulen, P. A. M., and Lozano, R. (2015). Internalizing sustainable practices: a configurational approach on sustainable forest management of the Dutch wood trade and timber industry. *J. Clean. Prod*. 107, 760–774. doi: 10.1016/j.jclepro.2015.05.087

Chen, Z., Bei, Y., and Rudin, C. (2020). Concept whitening for interpretable image recognition. *Nat. Mach. Intell*. 2, 772–782. doi: 10.1038/s42256-020-00265-z

Chung, Y.-A., Lin, H.-T., and Yang, S.-W. (2016). "Cost-aware pre-training for multiclass cost-sensitive deep learning," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, New York, NY, 1411–1417.

Damayanti, R., Prakasa, E., Krisdianto, Dewi, L. M., Wardoyo, R., Sugiarto, B., et al. (2019). LignoIndo: image database of Indonesian commercial timber. *IOP Conf. Ser*. 374, 012057. doi: 10.1088/1755-1315/374/1/012057

de Andrade, B. G., Basso, V. M., and de Figueiredo Latorraca, J. V. (2020).Machine vision for field-level wood identification. *IAWA J*. 41, 681–698. doi: 10.1163/ 22941932-bja10001

de Geus, A., Silva, S. F. D., Gontijo, A. B., Silva, F. O., Batista, M. A., and Souza, J. R. (2020). An analysis of timber sections and deep learning for wood species classification. *Multimed. Tools Appl*. 79, 34513–34529. doi: 10.1007/s11042- 020-09212-x

Devries, T., and Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv* [Preprint] arXiv:1708.04552,

Dieterich, U., and Auld, G. (2015). Moving beyond commitments: creating durable change through the implementation of Asia Pulp and Paper's forest conservation policy. *J. Clean. Prod*. 107, 54–63. doi: 10.1016/j.jclepro.2014.07. 084

Elkan, C. (2001). "The foundations of cost-sensitive learning," in *Proceedings of the 17th International Joint Conference on Artificial intelligence*, Vol. 2, (San Francisco, CA: Morgan Kaufmann Publishers Inc), 973–978.

Esteban, L. G., de Palacios, P., Conde, M., Fernandez, F. G., Garcia-Iruela, A., and Gonzalez-Alonso, M. (2017). Application of artificial neural networks as a predictive method to differentiate the wood of Pinus sylvestris L. and Pinus nigra Arn subsp. salzmannii (Dunal) Franco. *Wood Sci. Technol*. 51, 1249–1258. doi: 10.1007/s00226-017-0932-7

Esteban, L. G., Fernández, F. G., de Palacios, P., Romero, R. M., and Cano, N. N. (2009). Artificial neural networks in wood identification: the case of two Juniperus species from the Canary Islands. *IAWA J*. 30, 87–94. doi: 10.1163/ 22941932-90000206

Fabija´nska, A., Danek, M., and Barniak, J. (2021). Wood species automatic identification from wood core images with a residual convolutional neural network. *Comput. Electron. Agric*. 181:105941. doi: 10.1016/j.compag.2020. 105941

Figueroa-Mata, G., Mata-Montero, E., Valverde-Otárola, J. C., and Arias-Aguilar, D. (2018). "Automated image-based identification of forest species: challenges and opportunities for 21st century Xylotheques," in *Proceedings of the 2018 IEEE International Work Conference on Bioinspired Intelligence*, (San Carlos, CA), 1–8. doi: 10.1109/IWOBI.2018.8464206

Filho, P. L. P., Oliveira, L. S., Nisgoski, S., and Britto, A. S. Jr. (2014). Forest species recognition using macroscopic images. *Mach. Vis. Appl*. 25, 1019–1031. doi: 10.1007/s00138-014-0592-7

Gasson, P. (2011). How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. *IAWA J*. 32, 137–154. doi: 10.1163/22941932-90000049

Gasson, P. E., Lancaster, C. A., Young, R., Redstone, S., Miles-Bunch, I. A., Rees, G., et al. (2021). WorldForestID: addressing the need for standardized wood reference collections to support authentication analysis technologies; a way forward for checking the origin and identity of traded timber. *PLANTS People Planet* 3, 130–141. doi: 10.1002/ppp3.10164

Giovannoni, E., and Fabietti, G. (2013). "What is sustainability? A review of the concept and its applications," in *Integrated Reporting*, eds C. Busco, M. Frigo, A. Riccaboni, and P. Quattrone (Cham: Springer), doi: 10.1007/978-3-319-02168-3_2

Goëau, H., Bonnet, P., Joly, A., Baki´c, V., Barbe, J., Yahiaoui, I., et al. (2013). "Pl@ntNet mobile app," in. *Proceedings of the 21st ACM international conference on Multimedia*, (New York, NY), 423–424. doi: 10.1145/2502081. 2502251

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: The MIT Press.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning*, (New York, NY), 1321–1330.

Hardwood Federation (2016). *Economic Contribution of Hardwood Products: United States*. Available online at: http://hardwoodfederation.com/resources/ Documents/EIS%20States/US.pdf (accessed 26, July 2021).

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the 2015 International Conference on Computer Vision*, (Santiago), doi: 10. 1109/ICCV.2015.123

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas,NV), 770–778. doi: 10.1109/CVPR.2016.90

He, T., Lu, Y., Jiao, L., Zhang, Y., Jiang, X., and Yin, Y. (2020). Developing deep learning models to automate rosewood tree species identification for CITES designation and implementation. *Holzforschung* 74, 1123–1133. doi: 10.1515/ hf-2020-0006

Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., et al. (2020). Digitization and the future of natural history collections. *BioScience* 70, 243–251. doi: 10.1093/biosci/biz163

Hoadley, R. B. (1990). *Identifying Wood: Accurate Results With Simple Tools*. Newtown, CT: Taunton Press, 223.

Horn, G. V., Aodha, O. M., Song, Y., Cui, Y., Sun, C., Shepard, A., et al. (2018). "The inaturalist species classification and detection dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Salt Lake City, UT), 8769–8778. doi: 10.1109/CVPR.2018.00914

Howard, J., and Gugger, S. (2020). Fastai: a layered API for deep learning. *Information* 11:108. doi: 10.3390/info11020108

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 2261–2269. doi: 10.1109/CVPR.2017.243

Hwang, S.-W., Kobayashi, K., and Sugiyama, J. (2020). Detection and visualization of encoded local features as anatomical predictors in crosssectional images of Lauraceae. *J. Wood Sci*. 66:16. doi: 10.1186/s10086-020-01864-5

Hwang, S. W., and Sugiyama, J. (2021). Computer vision-based wood identification and its expansion and contribution potentials in wood science: a review. *Plant Methods* 17:47. doi: 10.1186/s13007-021-00746-1

IAWA (1989). IAWA Committee (eds. Wheeler, E.A., Baas, P., Gasson, P.), 1989. IAWA list of microscopic features for hardwood identification. *IAWA Bull*. 10, 219–332. doi: 10.1002/fedr.19901011106

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, (Mountain View, CA), 448–456.

Khalid, M., Lew, E., Lee, Y., Yusof, R., and Nadaraj, M. (2008). Design of an intelligent wood species recognition system. *Int. J. Simul. Syst. Scie. Technol*. 9, 9–19.

Kingma, D., and Ba, J. (2015). "Adam: a method for stochastic optimization," in *Proceedings of the 2015 International Conference on Learning Representations*, (San Diego, CA).

Kirker, G. T., and Lebow, S. T. (2021). "Chapter 15: wood preservatives," in *Wood Handbook—Wood as an Engineering Material. General Technical Report FPLGTR- 282*, ed. R. J. Ross (Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory), 26.

Krizhevsky, A., Sutskever, I., and Hinton, G. F. (2012). "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, (Red Hook, NY), 1097–1105.

Kwon, O., Lee, H., Yang, S.-Y., Kim, H., Park, S.-Y., Choi, I.-G., et al. (2019). Performance enhancement of automatic wood classification of Korean softwood by ensembles of convolutional neural networks. *J. Korean Wood Sci. Technol*. 47, 265–276.

Kwon, O., Lee, H. G., Lee, M.-R., Jang, S., Yang, S.-Y., Park, S.-Y., et al. (2017). Automatic wood species identification of Korean softwood based on convolutional neural networks. *Korean Wood Sci. Technol*. 45, 797–808.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Lens, F., Liang, C., Guo, Y., Tang, X., Jahanbanifard, M., da Silva, F. S. C., et al. (2020). Computer-assisted timber identification based on features extracted from microscopic wood sections. *IAWA J*. 41, 660–680. doi: 10.1163/22941932-bja10029

Lopes, D., Burgreen, G., and Entsminger, E. (2020). North American hardwoods identification using machine-learning. *Forests* 11:298. doi: 10.3390/f11030298

Magnus Boström, M., Jönsson, A. M., Lockie, S., Mol, A. P. J., and Oosterveer, P. (2015). Sustainable and responsible supply chain governance: challenges and opportunities. *J. Clean. Prod*. 107, 1–7. doi: 10.1016/j.jclepro.2014.11.050

Martins, J., Oliveira, L. S., Nisgoski, S., and Sabourin, R. (2013). A database for automatic classification of forest species. *Mach. Vis. Appl*. 24, 567–578. doi: 10.1007/s00138-012-0417-5

Miller, A. M. M., and Bush, S. R. (2015). Authority without credibility? Competition and conflict between ecolabels in tuna fisheries. *J. Clean. Prod*. 107, 137–145. doi: 10.1016/j.jclepro.2014.02.047

Niculescu-Mizil, A., and Carauna, R. (2005). "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning*, (New York, NY), 625–632. doi: 10.1145/1102351.1102430

Olah, C., Mordvintsev, A., and Schubert, L. (2017). Feature Visualization. *Distill* 2:11. doi: 10.23915/distill.00007

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng*. 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Panshin, A. J., and de Zeeuw, C. (1980). *Textbook of WOOD Technology: Structure, Identification, Properties, and Uses of the Commercial Woods of the United States and Canada. McGraw-Hill Series in Forest Resources*, 4th Edn. New York, NY: McGraw-Hill Book Co.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural Inform*. Proc. Syst. 2019, 8026–8037.

Pearson, K. D., Nelson, G., Aronson, M. F. J., Bonnet, P., Brenskelle, L.,Davis, C. D., et al. (2020). Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70, 610–620. doi: 10.1093/biosci/biaa044

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res*. 12, 2825–2830.

Qin, Z., Yu, F., Liu, C., and Chen, X. (2018). How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. *Math. Foundat. Comput*. 1, 149–180. doi: 10.3934/mfc.2018008

Ravindran, P., Costa, A., Soares, R., andWiedenhoeft, A. C. (2018). Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14:25. doi: 10.1186/s13007-018-0292-9

Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R., et al. (2019). "Image based identification of Ghanaian timbers using the XyloTron: opportunities, risks, and challenges," in *Proceedings 2019 Workshop on Machine Learning for the Developing World*, (Vancouver, BC).

Ravindran, P., Owens, F. C., Wade, A. C., Vega, P., Montenegro, R., Shmulsky, R., et al. (2021). Field-deployable computer vision wood identification of Peruvian timbers. *Front. Plant Sci*. 12:647515. doi: 10.3389/fpls.2021.647515

Ravindran, P., Thompson, B. J., Soares, R. K., and Wiedenhoeft, A. C. (2020). The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. *Front. Plant Sci*. 11:1015. doi: 10.3389/fpls.2020.01015

Ravindran, P., and Wiedenhoeft, A. C. (2020). Comparison of two forensic wood identification technologies for ten Meliaceae woods: computer vision versus mass spectrometry. *Wood Sci. Technol.* 54, 1139–1150. doi: 10.1007/s00226- 020-01178-1

Rosa da Silva, N., De Ridder, M., Baetens, J. M., and den Bulcke, J. V. B. (2017). Automated classification of wood transverse cross-section micro-imagery from 77 commercial Central-African timber species. *Ann. For. Sci*. 74:30. doi: 10. 1007/s13595-017-0619-0

Ross, R. J., and White, R. H. (2014). *Wood Condition Assessment Manual: USDA Forest Service Forest Products Laboratory General Technical Report FPL-GTR- 234*, 2nd Edn. Madison, WI: U.S. Dept. of Agriculture, doi: 10.2737/FPL-GTR-234

Ruffinatto, F., Crivellaro, A., and Wiedenhoeft, A. C. (2015). Review of macroscopic features for hardwood and softwood identification and a proposal for a new character list. *IAWA J*. 36, 208–241. doi: 10.1163/22941932-00000096

Ruggerio, C. A. (2021). Sustainability and sustainable development: a review of principles and definitions. *Sci. Total Environ*. 786:147481. doi: 10.1016/j. scitotenv.2021.147481

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet: large scale visual recognition challenge. *Int. J. Comput. Vis*. 115, 211–252. doi: 10.1007/s11263-015-0816-y

Schmitz, N., Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J. W. B., Cabezas, J. A., et al. (2020). *Overview of Current Practices in Data Analysis for Wood Identification. A Guide for the Different Timber Tracking Methods. Global Timber Tracking Network, GTTN Secretariat*. Braunschweig: European Forest Institute and Thünen Institute,

Shigei, N., Mandai, K., Sugimoto, S., Takaesu, R., and Ishizuka, Y. (2019). "Landuse classification using convolutional neural network with bagging and reduced categories," in *Proceedings of the International MultiConference of Engineers and Computer Scientists* 2019, (Hong Kong).

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv* [Preprint] arXiv:1409.1556,

Simpson, W. T. (1991). *Dry Kiln Operator's Manual, USDA Forest Service, Agriculture Handbook.* Madison, WI: U.S. Dept. of Agriculture, 188.

Smith, L. (2018). A disciplined approach to neural network hyper-parameters: part 1 – learning rate, batch size, momentum, and weight decay. *arxiv* [Preprint] arXiv:1803.09820,

Souza, D. V., Santos, J. X., Vieira, H. C., Naide, T. L., Nisgoski, S., and Oliveira, L. E. S. (2020). An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. *Wood Sci. Technol*. 54, 1065–1090. doi: 10.1007/s00226-020 01196-z

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res*. 15, 1929–1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, (Boston, MA), 1–9. doi: 10.1109/CVPR.2015.7298594

Tang, X. J., Tay, Y. H., Siam, N. A., and Lim, S. C. (2018). "MyWood-ID: automated macroscopic wood identification system using smartphone and macro-lens," in *Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems*, (New York, NY), 37–43. doi: 10.1145/3293475.3293493

von Baeyer, M., and Marston, J. M. (2021). Best practices for digitizing a wood slide collection: the Bailey-Wetmore Wood Collection of the Harvard University Herbaria. *Q. Int*. 593–594, 50–59. doi: 10.1016/j.quaint.2020.08.053

Wheeler, E. A., and Baas, P. (1998). Wood Identification -A Review. *IAWA J*. 19, 241–264. doi: 10.1163/22941932-90001528

Wiedenhoeft, A. C. (2011). *Identification of Central American Woods (Identificacion de las Especies Maderables de Centroamerica)*. Madison, WI: Forest Products Society, 167.

Wiedenhoeft, A. C. (2020). The XyloPhone: toward democratizing access to high quality macroscopic imaging for wood and other substrates. *IAWA J*. 41, 699–719. doi: 10.1163/22941932-bja10043

Wiedenhoeft, A. C., Simeone, J., Smith, A., Parker-Forney, M., Soares, R., and Fishman, A. (2019). Fraud and misrepresentation in retail forest products exceeds U.S. Forensic wood science capacity. *PLoS One* 14:e0219917. doi: 10.1371/journal.pone.0219917

Wu, F., Gazo, R., Haviarova, E., and Benes, B. (2021). Wood identification based on longitudinal section images by using deep learning. *Wood Sci. Technol*. 55, 553–563. doi: 10.1007/s00226-021-01261-1

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science*, Vol. 8689, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer), doi: 10.1007/978-3-319-10590-1_53

CHAPTER IV

TOWARDS SUSTAINABLE NORTH AMERICAN WOOD PRODUCT VALUE CHAINS, PART 2; COMPUTER VISION IDENTIFICATION OF RING-POROUS HARDWOODS

Ravindran, P., Wade, A. C., Owens, F. C., Shmulsky, R., and Wiedenhoeft, A. C. (2022). Towards sustainable North American wood product value chains, Part 2: Computer vision identification of ring-porous hardwoods. *Can. J. For. Res*., cjfr-2022-0077. doi: 10.1139/cjfr-2022-0077. (Republished with permission)

4.1 Abstract

Wood identification is vitally important for ensuring the legality of North American hardwood value chains. Computer vision wood identification (CVWID) systems can identify wood without necessitating costly and time-consuming offsite visual inspections by highly trained wood anatomists. Previous work by Ravindran et al. presented macroscopic CVWID models for identification of North American diffuse porous hardwoods from 22 wood anatomically informed classes using the open-source XyloTron platform. This manuscript expands on that work by training and evaluating complementary 17-class XyloTron CVWID models for the identification of North American ring porous hardwoods – woods that display spatial heterogeneity in earlywood and latewood pore size and distribution and other radial growth-rate related features. Deep learning models trained using 4045 images from 452 ring-porous wood specimens from four xylaria demonstrated 98% five-fold cross-validation accuracy. A field model trained on all the training data and subsequently tested on 198 specimens drawn from two additional xylaria achieved top-1 and top-2 predictions of 91.4% and 100%,

respectively, and images devoid of earlywood, latewood, or broad rays did not greatly reduce the prediction accuracy. This study advocates for continued cooperation between wood anatomy and machine learning experts for implementing and evaluating field-operational CVWID systems.

4.2 Key words

Wood identification, illegal logging and timber trade, XyloTron, computer vision, machine learning, deep learning, diffuse porous hardwoods, ring porous hardwoods, sustainable wood products

4.3 Introduction

Wood identification can be of vital importance for designing, monitoring, and establishing sustainable wood product value chains and for ensuring legality under laws and policies governed by international treaties (e.g., the Convention on the International Trade in Endangered Species of Flora and Fauna) as well as national laws and policies (e.g., the United States' Lacey Act., and 2012 Illegal Logging Prohibition Act of Australia). Wood identification is traditionally performed by wood anatomy experts in a laboratory setting and relies on the ability of human experts to recognize and differentiate anatomical features. Recently, in order to tackle the paucity of traditional wood identification expertise (Wiedenhoeft et al., 2019), computer vision wood identification (CVWID) systems have been applied both in the laboratory and in the field to address the challenge of identifying wood without a trained expert's eye (Khalid et al., 2008; Martins et al., 2013; Filho et al., 2014; Figueroa-Mata et al., 2018; Ravindran et al., 2018, 2019; Damayanti et al., 2019, de Andrade et al., 2020; Ravindran and Wiedenhoeft, 2020; Souza et al., 2020; Ravindran et al., 2021). The open-source XyloTron platform (Ravindran et al., 2020, 2021) has shown potential for real-time, field-deployable,

screening-level wood identification (Ravindran et al., 2019; Ravindran and Wiedenhoeft, 2020; Arévalo et al., 2021; Ravindran et al., 2021), and, with the XyloPhone (Wiedenhoeft, 2020), it is possible to move from laptop-based devices to smartphones for field deployment. Both the XyloTron and XyloPhone platforms provide an imaging system that enable the capture of macroscopic features (Miller et al., 2002; Ruffinatto et al., 2015) suitable for wood identification.

Designing high-performing, scalable CVWID systems requires understanding wood anatomy and how that anatomy influences the training, performance and deployability of convolutional neural networks (CNN) (Ravindran et al., 2022) or other machine learning based models (de Geus et al., 2021). Hwang and Sugiyama (2021) report the classification accuracy of numerous CNN models used in wood identification studies, with most prior works demonstrating a high *in silico* accuracy of 90% and better with similar performance across different architectures, but most of those studies do not report any subsequent model testing on new, unique specimens, so their real-world applicability is unknown. It may be the case that for CVWID the number of classes, number of training images (coverage of anatomical variation), quality of specimen surface preparation (visibility of anatomical features), quality of images (clarity of anatomical features), the size of the area imaged vis-à-vis the scale of diagnostic anatomical features, and the degree to which the anatomical features among the classes are similar are all likely important factors for CNN architecture design and eventual field performance of trained models. For this reason, it is vitally important to attempt to evaluate how wood anatomy at a range of scales affects imaging and CVWID model performance.

Ravindran et al. (2022) estimated that approximately 40 classes of North American hardwoods need to be included in a field-deployable computer vision model for the North American market, a number substantially greater than anything previously published for this

region, either in terms of macroscopic images (Lopes et al., 2020, 10 classes) or at the naked-eye level (Wu et al., 2021, 11 classes). As noted in Ravindran et al. 2022, the influence of class number on CVWID models is unknown, especially for North American hardwoods, where there are, broadly speaking, two wood anatomically distinct groups of woods – the diffuse-porous woods and the ring-porous woods. They therefore used a fundamental domain-specific factor, porosity, to inform taxa selection and label space design. In general, diffuse-porous woods show less wood anatomical spatial heterogeneity with regard to radial growth rate, growth ring domains (earlywood vs. latewood), and physiological age of the wood (Ravindran et al., 2022). Diffuse-porous woods of North America also show comparatively lower overall wood anatomical variability (e.g. axial parenchyma patterns, vessel arrangement, and ray width and frequency), than, for example, diffuse-porous tropical woods (e.g. de Andrade et al., 2020; Arevalo et al., 2021), or compared to the latewood of ring-porous North American woods (Fig. 4.1). Ravindran et al. (2022) therefore separated the North American hardwoods into two groups: the diffuse-porous woods of the earlier work and the ring-porous woods addressed herein.

Unlike diffuse-porous hardwoods, ring-porous hardwoods, by definition, show dramatic differences between earlywood and latewood within a growth ring and among species (Fig. 4.1). Due to the spatial heterogeneity displayed by ring-porous woods, it is possible, depending on the area of tissue captured and the respective sizes of the earlywood and latewood regions, to obtain an image that does not exhibit all the anatomical characteristics that typify the wood. Fast radial growth can result in images that show only latewood (Fig. 4.1-C), that is, only the latter-formed portion of a single growth ring. Tangentially varying features (e.g., broad rays in *Quercus*; Fig. 4.1-B) may be absent in some images. Slow radial growth can produce an image that is primarily earlywood (Fig. 4.1-D). The impact of such spatial heterogeneity as reflected in test images is

unknown and unexplored. An initial work purporting to use CVWID to classify ten ring-porous North American hardwoods did not appear to consider spatial heterogeneity related to wood anatomy (Lopes et al., 2020). Further, the apparently subpar image quality of that dataset was first questioned (Wiedenhoeft, 2020), and, later, the machine learning analysis and underlying dataset were demonstrated to be inherently flawed based on data hygiene for CVWID inference (Ravindran and Wiedenhoeft, 2022).

In this study, we develop a CVWID model to identify 17 classes of North American ring-porous woods using the XyloTron platform and a convolutional neural network. In addition to performance evaluation for accuracy and domain-informed examination of model misclassifications, we investigate the influence of wood anatomical spatial heterogeneity of ring-porous woods on specimen level model predictions and discuss how other forms of wood anatomical heterogeneity are thus potentially capable of influencing model performance in field deployment settings. Finally, we propose a path for future research for developing a robust, highly accurate, field-deployable, unified North American hardwood model.

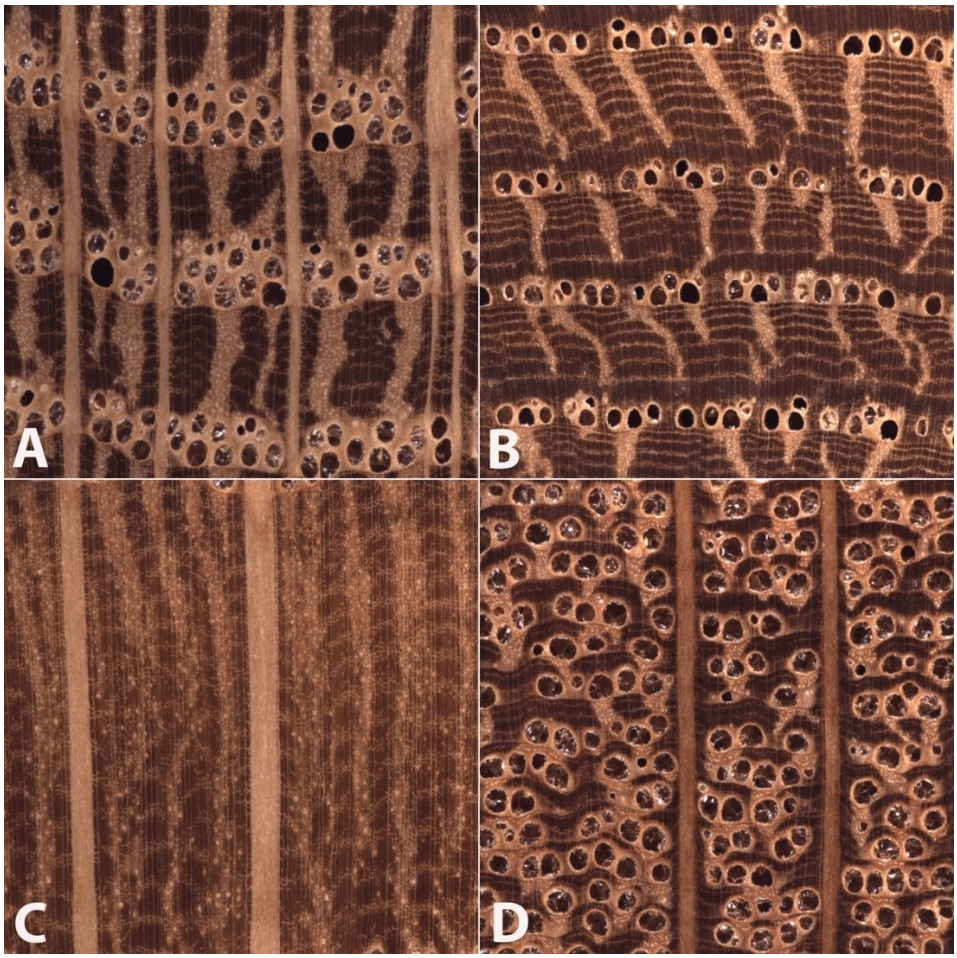

Figure 4.1 Images of the transverse surfaces of *Quercus alba* specimens under varying growth conditions: slow-, medium-, and fast-growth.

Note: Images A and B show medium-growth with approximately three complete growth rings. Image B lacks broad rays which are necessary for identifying *Quercus*. Images C and D are each missing important anatomical features that would allow for accurate identification. As a result of fast radial growth, image C shows a partial, latewood-only growth ring, thus not demonstrating ring porosity. Due to the Slow-Growth conditions, image D displays the relative absence of latewood features, precluding the ready separation of the white oak group from the red oak group. Note also in D, the ring-porous character of the wood is less obvious as a result of the closely spaced growth rings. Each image represents 6.35mm of tissue on a side.

4.4 Material and Methods

4.4.1 Dataset Details

4.4.1.1 Taxa and Sample Selection

We selected 68 North American ring-porous hardwood species from 15 prominent genera based on their commercial importance, botanical relevance, and specimen availability from five scientific wood collections and forensically verified specimens from a wood anatomical teaching collection. Table 4.1 summarizes the details of these six collections and their specimen contributions.

Table 4.1 Summary of xylaria and wood specimen contributions for model training and testing.

Institution (Xylarium acronym)	Specimen counts	Role
USDA Forest Products Laboratory, Madison collection (MADw)	314	Model Training
USDA Forest Products Laboratory, Samuel J. Record collection (SJRw)	94	Model Training
USDA Forest Products Laboratory, Forest Stewardship Council *Quercus* specimen collection (fscquercus)	28	Model Training
Royal Museum of Central Africa (Tw)	16	Model Training
Mississippi State University, David A. Kribs collection (PACw)	**192**	**Model Testing**
Mississippi State University, Department of Sustainable Bioproducts Teaching collection (MSUtw)	**6**	**Model Testing**

Note: Four xylaria, one teaching collection (MSUtw), and a set of scientifically collected, georeferenced stem discs (fscquercus) provided specimens for image datasets for the training and testing of the wood identification models. MADw, SJRw, fscquercus, and Tw specimens contributed solely to the training dataset. In contrast, the independent test dataset was obtained from specimens acquired from the PACw and MSUtw specimens, the class-level identifications of the latter confirmed by laboratory analysis.

4.4.1.2 Sample Preparation and Imaging

The transverse surface of 650 wood specimens was polished using sanding discs with progressively finer abrasive grit (240, 400, 600, 800, 1000, 1500). Between each grit, compressed air and adhesive tape were used to remove dust from the cell lumina to the extent possible. It should be noted that the aqueous polishing method of Barbosa et al. (2021) is not suitable for entire xylarium specimens, as it would tend to damage historic specimen labels,

induce swelling-related checking, cause extractive movement or staining, and/or a combination of all the above. Our progressive sanding protocol provided a repeatable method for consistently preparing uniform specimen surfaces for imaging. Multiple non-overlapping images of the transverse surface of each wood sample were captured with the XyloTron platform (Ravindran et al., 2020). Each image had a resolution of 2048 × 2048 pixels and captured an area of tissue that measured 6.35mm × 6.35mm with a linear resolution of 3.1 microns/pixel. Multiple instantiations of the XyloTron system along with multiple operators with varying degrees of experience in sample preparation and knowledge of wood anatomy (undergraduate students, graduate students, postdoctoral researchers, and technical specialists) were utilized for sample preparation and image capture. The resulting images were subsequently curated for image quality and the presence of representative anatomical characteristics. Table 4.2 shows a summary of the collected datasets.

Table 4.2 Summary of image datasets.

	Training (counts)	**Testing (counts)**	**Total (counts)**
Number of collections	4	2	6
Number of taxa*	64	40	68
Number of specimens	452	198	650
Number of images	4045	936	4981

Note: 650 specimens from 68 unique taxa (from 15 genera) were polished and imaged, resulting in 4981 images: 4045 for training, and 936 for testing the classification models. Supplementary Material 5 details the membership of classes and training/testing datasets.

4.4.1.3 Label Assignment

According to Gasson (2011), the light microscopic identification of wood specimens is generally accurate only to the genus-level. In this study, we categorized the selected taxa into a combination of generic and sub-generic classes based on the similarity of macroscopic anatomical characteristics to facilitate machine learning and for use on the XyloTron platform. We grouped the taxa into 17 classes in the following ways:

1. The genera *Asimina, Carya, Castanea, Catalpa, Celtis, Cladrastis, Fraxinus, Gleditsia, Gymnocladus, Maclura, Morus, Robinia*, and *Sassafras* were each assigned to a genus-level class.

2. The genera *Quercus* and *Ulmus* were each split into two classes. *Quercus* classes were labeled “QuercusR” (red) and “QuercusW” (white) corresponding to the commercial red and white oak groups, which are anatomically distinguishable on the transverse surface on the basis of differences in latewood pore diameter and distribution. *Ulmus* classes were labeled “UlmusS” (soft) and “UlmusH” (hard) based on commercial grouping and continuous and discontinuous

row(s) of earlywood vessels, respectively, and differences in the mean radial and tangential earlywood vessel diameter (Wheeler et al., 1989).

Although class names include genus names, we follow a convention of not italicizing the class names so that we can distinguish when we are discussing genera or species (which are italicized) versus class names.

Supplementary Material 5 in Appendix contains a list of the 68 taxa, their class labels, and their training and testing dataset membership.

4.4.1.4 Spatial Heterogeneity Datasets

In addition to the 936 images that comprised the main testing dataset shown in Table 4.2, three smaller datasets were collected to evaluate the effects of spatial heterogeneity on model accuracy (hereafter the "spatial heterogeneity datasets," Table 4.3). From the 192 PACw specimens imaged for the main test dataset, 38 specimens were selected that exhibited 1) especially slow radial growth (narrow, closely-spaced growth rings), 2) especially fast growth (wide growth rings), or 3) large areas devoid of broad rays (in *Quercus*). These specimens were reimaged in areas that contained entirely earlywood (to generate the Slow-Growth dataset), virtually no earlywood (to generate the Fast-Growth dataset), or that lacked broad rays (to generate the Broad Rays Absent dataset). The three resulting datasets thus each lacked at least one characteristic wood anatomical feature used by human identifiers to characterize the woods in question.

The classes included in the Slow-Growth dataset are Carya, Cladrastis, Gleditsia, Morus, QuercusR, and QuercusW. The Fast-Growth dataset included classes Catalpa, Cladrastis, Morus, QuercusR, QuercusW, Robinia, and UlmusS. Specimens in QuercusW were the only ones to

display images lacking broad rays. Not all classes were included in these datasets due to the absence of specimens in some classes featuring distinctly slow-or fast-growth. Table 4.3 summarizes the number of images contained in each dataset.

Table 4.3 Summary of spatial heterogeneity image datasets.

	No. of Images	**No. of classes**	**No. of Specimens**
Slow-Growth (earlywood only)	101	6	23
Fast-Growth (latewood only)	44	7	10
Broad-Rays Absent	44	1	6

4.4.2 Machine Learning Details

4.4.2.1 Model Architecture and Training

Prior work (e.g., Ravindran et al., 2019, 2020, 2021, Arevalo et al., 2021) has demonstrated the effectiveness of using a two-stage (Howard and Gugger, 2020) transfer learning (Pan and Yang, 2010) approach for training strong baseline convolutional neural network (CNN) models for CVWID. This training approach was employed here to learn the weights of a ResNet34-based CNN with a custom classifier head that can handle 17 classes. The custom classifier head consisted of global average and global maximum pooling layers which were concatenated and fed through two fully-connected layers with batch-normalization (Ioffe and Szegedy, 2015) and dropout (Srivastava et al., 2014) in sequence. This was followed by a softmax layer that produced class prediction distribution over the 17 classes. In the first stage of training, the ImageNet (Russakovsky et al., 2015) pretrained weights of the backbone were frozen and only the weights of the custom head were learned. During the second stage, the

weights in both the backbone and the head were finetuned. Data augmentation, that included reflections, rotations and CutOut (DeVries and Taylor, 2017), was performed during training. The learning rate hyperparameter was estimated using the one-cycle policy of (Smith, 2018) and was annealed (Howard and Gugger ,2020) when training using the Adam optimizer (Kingma and Ba, 2017). Details about the model architecture, training methodology, hyperparameter optimization, and data augmentation can be found in Ravindran et al., 2022. PyTorch (Paszke et al., 2019) and scientific Python tools (Pedregosa et al., 2011) were used for model definition, training, and evaluation. Additional details of a ResNet 50 model trained and evaluated identically to the ResNet34 model are available in Supplementary Material 6 in the Appendix.

4.4.2.2 Model Evaluation

The following analyses were conducted for evaluation of trained models:

(1) Training and evaluation were performed using five-fold cross-validation analysis with class level stratification folds along with specimen-level separation among the five folds (i.e., images of each specimen contributed images to exactly one fold). For valid assessment of any wood identification machine learning-based classifier, it is necessary to conduct specimen level mutual exclusivity between the folds (e.g., Ravindran et al., 2019, 2020, 2021, as discussed in Hwang and Sugiyama, 2021, and in Ravindran and Wiedenhoeft, 2022). A confusion matrix and the corresponding top-1 and top-2 prediction accuracies were computed by consolidating the model predictions over the five folds. It should be reiterated that the cross-validation analysis did not include images from the PACw or MSUtw datasets.

(2) The five models from the cross-validation analysis, each trained using a different 80% split of the training data, were also tested on the PACw + MSUtw dataset. The top-1 and top-2 accuracies for this analysis are also presented.

(3) A field model was trained using all the images from the cross-validation analysis (i.e., 100% of the training data) and evaluated on images from the PACw + MSUtw dataset. A confusion matrix and the top-1 and top-2 prediction accuracies were computed to assess the utility of the field model.

The predicted top-1 class for a specimen was taken to be the majority of class predictions for the images contributed by the specimen. The top-2 image-level predictions for a specimen were generated with equal weight voting: if a specimen's true class was one of the top-2 predicted classes, the specimen was considered correctly identified.

4.4.2.3 Misclassification Analysis

Images from all misclassified specimens from the field models were evaluated and assigned to one of three types of misclassification as reported in detail in Ravindran et al. 2022, and we also adopt their source and sink misclassification analysis as implemented therein.

4.4.2.4 Spatial Heterogeneity Evaluation

The impact of spatial heterogeneity on model performance was evaluated using the three datasets obtained from the PACw specimens (see section 4.4.1.4). Table 4.4 lists the classes and the number of specimens per class that comprise each of the spatial heterogeneity datasets (Slow-Growth, Fast-Growth, and Broad Rays Absent).

Table 4.4 Summary of the number of specimens and their class labels included in each of the spatial heterogeneity datasets.

Slow-Growth Dataset		**Fast-Growth Dataset**		**Broad-Rays Absent Dataset**	
Class	**No. of Specimens**	**Class**	**No. of Specimens**	**Class**	**No. of Specimens**
Carya	11	Catalpa	2	QuercusW	6
Cladrastis	1	Cladrastis	1		
Gleditsia	1	Morus	1		
Morus	1	QuercusR	2		
QuercusR	3	QuercusW	1		
QuercusW	6	Robinia	1		
		UlmusS	2		

Note: Not all classes are included in these datasets due to the lack of specimens characterized as slow-growth or fast-growth.

4.5 Results

The top-1 prediction accuracy for the specimen level cross-validation model was 98.0%. When tested on the PACw + MSUtw dataset, the top-1 and top-2 cross-validation accuracies were 91.9% and 98.3%, respectively. The field model top-1 accuracy was 91.4%, and the top-2 accuracy was 100%. Table 4.5 shows the summary of the cross-validation (accumulated over the five folds) and field model's prediction accuracies. Confusion matrices for the cross-validation and field models are shown in Figures 4.2 and 4.3, respectively.

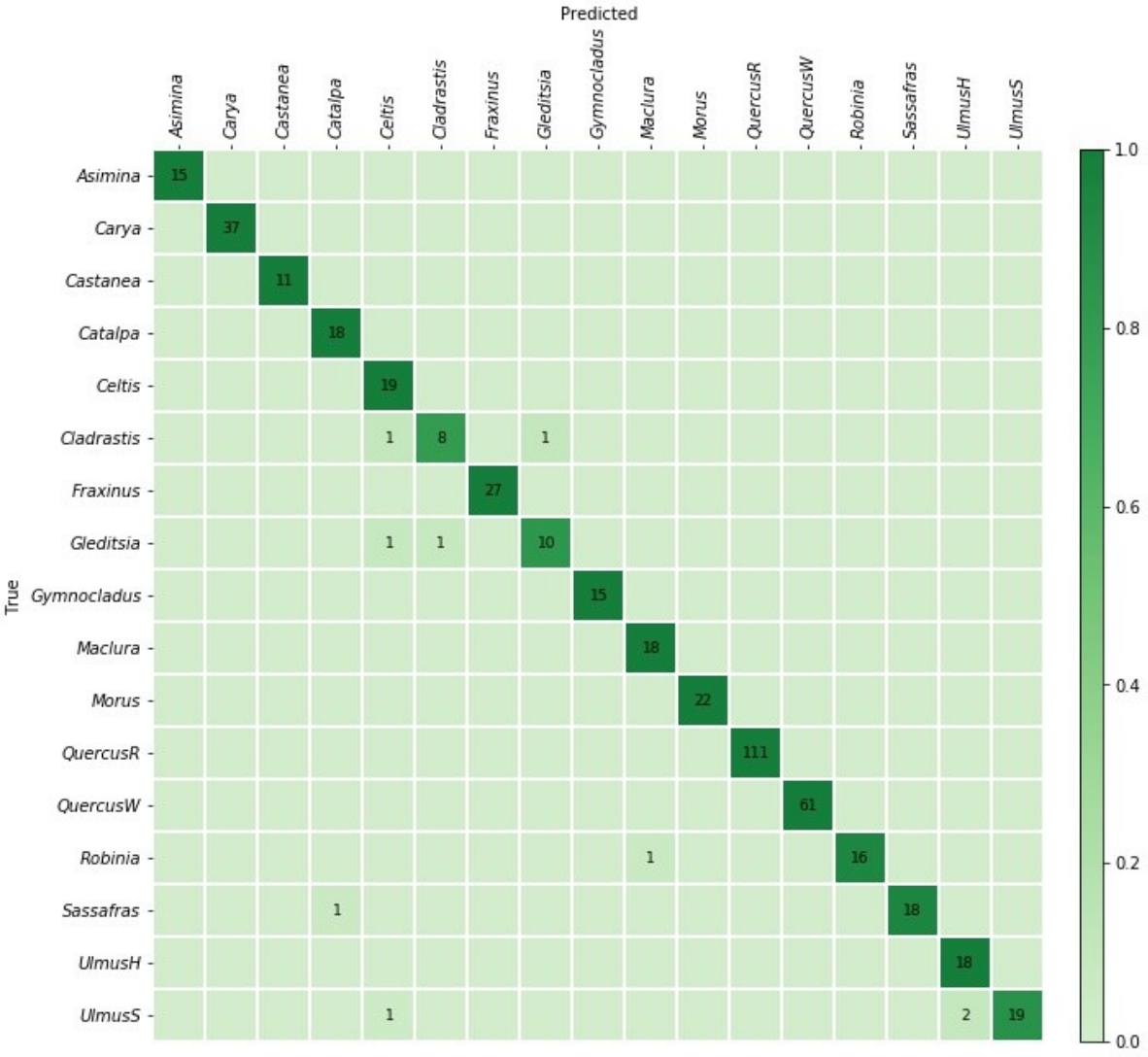

Figure 4.2 Confusion matrix for the cross-validation model.

Note: Top-1 predictions on 452 specimens (accumulated over five folds), with a specimen-level accuracy of 98.01%.

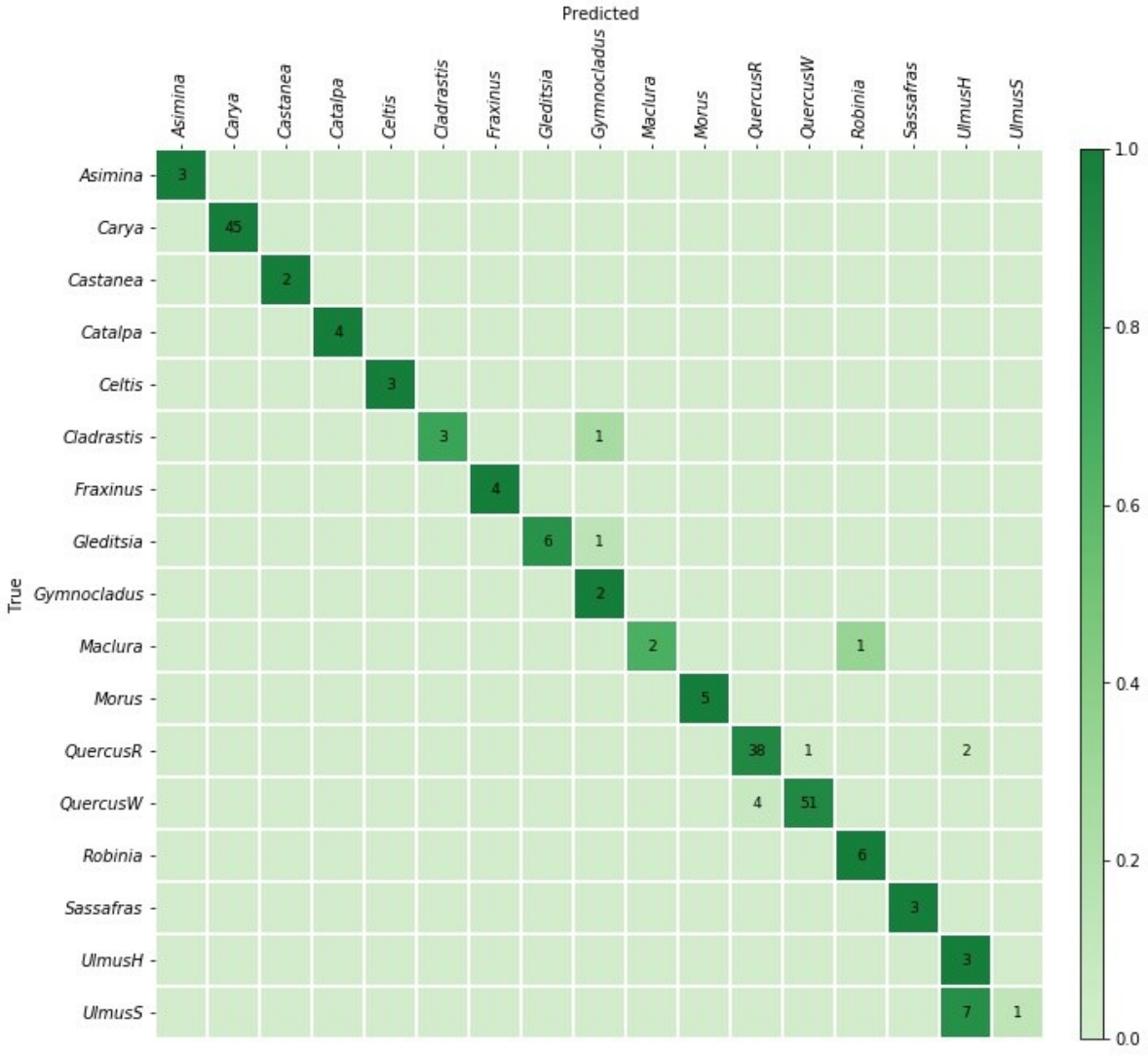

Figure 4.3 Confusion matrix for the field model's top-1 predictions.

Note: 198 specimens in the PACw + MSUtw dataset. Specimen-level accuracies for top-1 and top-2 predictions were 91.41% and 100%, respectively.

Table 4.5 Training and testing specimen level model prediction accuracies.

Training and evaluation details	**Top-k**	**Accuracy (%)**
Five-fold cross-validation	k=1	**98.0**
Trained using four folds, tested on PACw + MSUtw	k=1	91.9
	k=2	98.3
Field model trained using all five folds, tested on PACw + MSUtw	k=1	**91.4**
	k=2	100

Note**:** The training and testing dataset accuracies were computed using 452 (across the five folds) and 198 specimens respectively. Confusion matrices are provided for the percentages in bold.

Example images of Type 1 and Type 3 misclassifications from the field model's confusion matrix (Fig. 4.3) are shown in Figure 4.4. A summary of misclassification data for the field model is presented in Table 4.6.

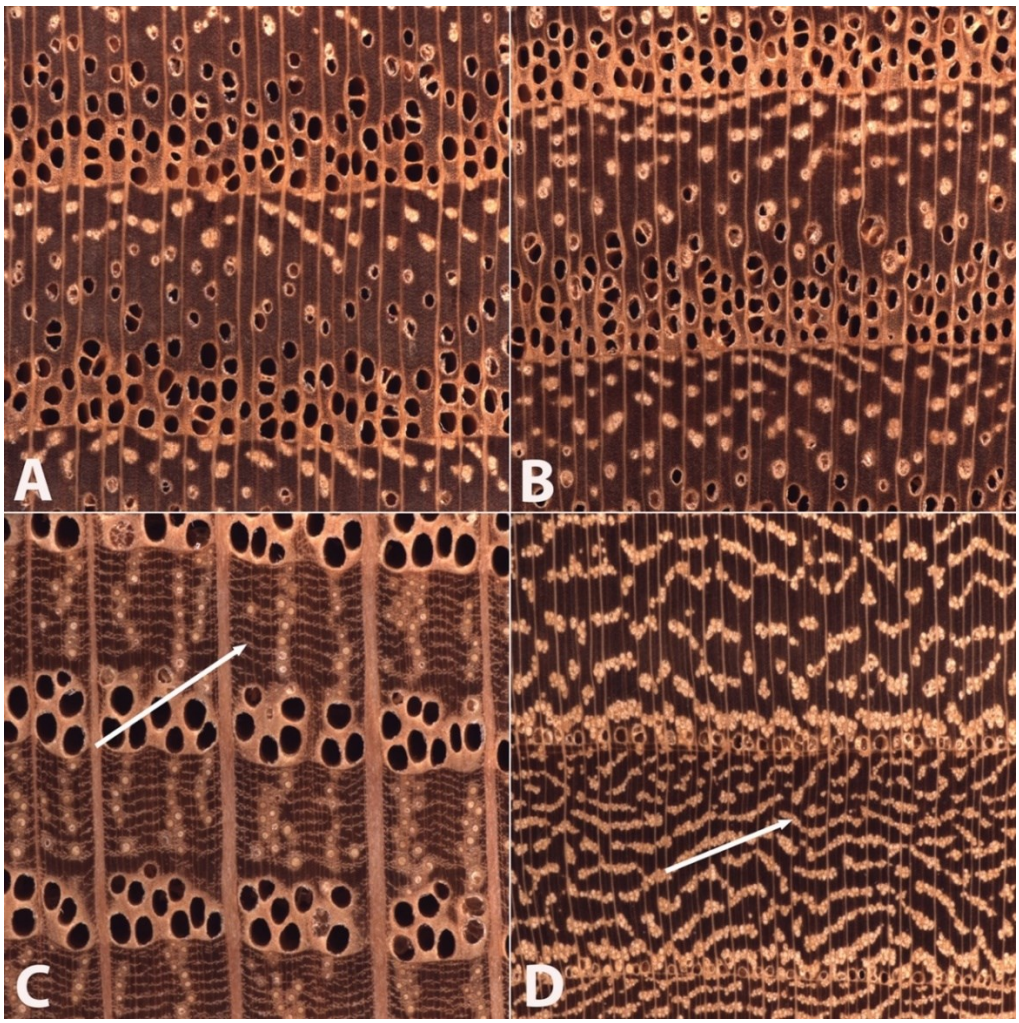

Figure 4.4 Images of the transverse surface of test specimens from classes Gleditsia and QuercusR (A and C) and exemplar images from classes Gymnocladus (B) and UlmusH (D).

Note: Images A and B shows the type 1 misclassification where a specimen of Gleditsia was misclassified to the anatomically similar class Gymnocladus. An anatomically typical specimen of the class QuercusR (C) was misclassified as the anatomically disparate class UlmusH (D), a Type 3 misclassification. Note the anatomical similarities between A and B and the anatomical dissimilarity between C and D, especially with regard to the difference in ray size, earlywood vessel diameter and arrangement. Also, in images C and D, the red arrows indicate a possible comparison of banded parenchyma in the latewood of QuercusR (C) and ulmiform latewood vessel arrangement in UlmusH (D). This comparison of anatomical features may have accounted for the misclassification of QuercusR specimens as UlmusH specimens. Each image represents 6.35mm of tissue on a side.

Table 4.6 Number and proportion of misclassified specimens from Figure 4.3 by type of misclassification.

Misclassification types	Number of misclassified specimens (of 198 total specimens)	Proportion of the 17 misclassified (of 198 total) specimens
Taxa are anatomically consistent, test specimen typical (Type 1)	15	0.88 (0.08)
Test specimen atypical for its taxon* (Type 2)	0	0
Taxa and test specimen are not anatomically consistent (Type 3)	2	0.12 (0.01)
Total	17	1 (0.09)

Note: Types 1 and 2 are consistent with wood anatomy and expected errors made by human field inspectors. Type 3 errors are inconsistent with macroscopic wood anatomy and would not be expected to be made by a human inspector. * But within a reasonable range for the taxon.

For the top-1 accuracy of the field model, 11 classes showed zero source misclassifications on the PACw + MSUtw dataset: Asimina, Carya, Castanea, Catalpa, Celtis, Fraxinus, Gymnocladus, Morus, Robinia, Sassafras, and UlmusH. At least one source misclassification was shown in the remaining 6 classes (Fig. 4.3), with 17 misclassified specimens of 198 test specimens in total. Six classes provided source misclassifications, and those misclassified specimens were attributed to the five following classes: Gymnocladus, QuercusR, QuercusW, Robinia, and UlmusH. There were 5 classes that drew sink misclassifications: Gymnocladus, QuercusR, QuercusW, Robinia, and UlmusH. Eight classes showed neither source nor sink misclassifications: Asimina, Carya, Castanea, Catalpa, Celtis, Fraxinus, Morus, and Sassafras. Table 4.6 summarizes the number and proportions of misclassification types. Fifteen of the seventeen (88.2%) misclassifications were Type 1. There

were only 2 out of 17 (11.8%) misclassifications that were of Type 3, and there were zero Type 2 misclassifications.

When tested against the three spatial heterogeneity datasets, prediction accuracy of the field model remained nearly unchanged at 91.3% in the case of the Slow-Growth dataset and fell by 11.4% for the Fast-Growth dataset and 8.3% for the Broad Rays Absent (QuercusW) dataset. Of the Slow-Growth dataset, a specimen from class Cladrastis was predicted as Gymnocladus and a specimen from the class QuercusW was predicted as QuercusR. Within the Fast-Growth dataset, two specimens from the class UlmusS were predicted as UlmusH. The Broad Rays Absent dataset, which consisted of the class QuercusW, had one of six specimens misclassified as QuercusR. Table 4.7 summarizes the accuracies for the three spatial heterogeneity datasets when tested with the field model. A comparison of the test specimen and an example image of the predicted class of each spatial heterogeneity dataset is shown in Figure 4.5.

Table 4.7 Specimen-level field model performance metrics on spatial heterogeneity datasets.

Dataset	**Accuracy**
Slow-Growth (earlywood only)	91.3%
Fast-Growth (latewood only)	80.0%
Broad-Rays Absent (only in QuercusW)	83.3%

Note: Top-1 field model accuracy on typical images was 91.41%.

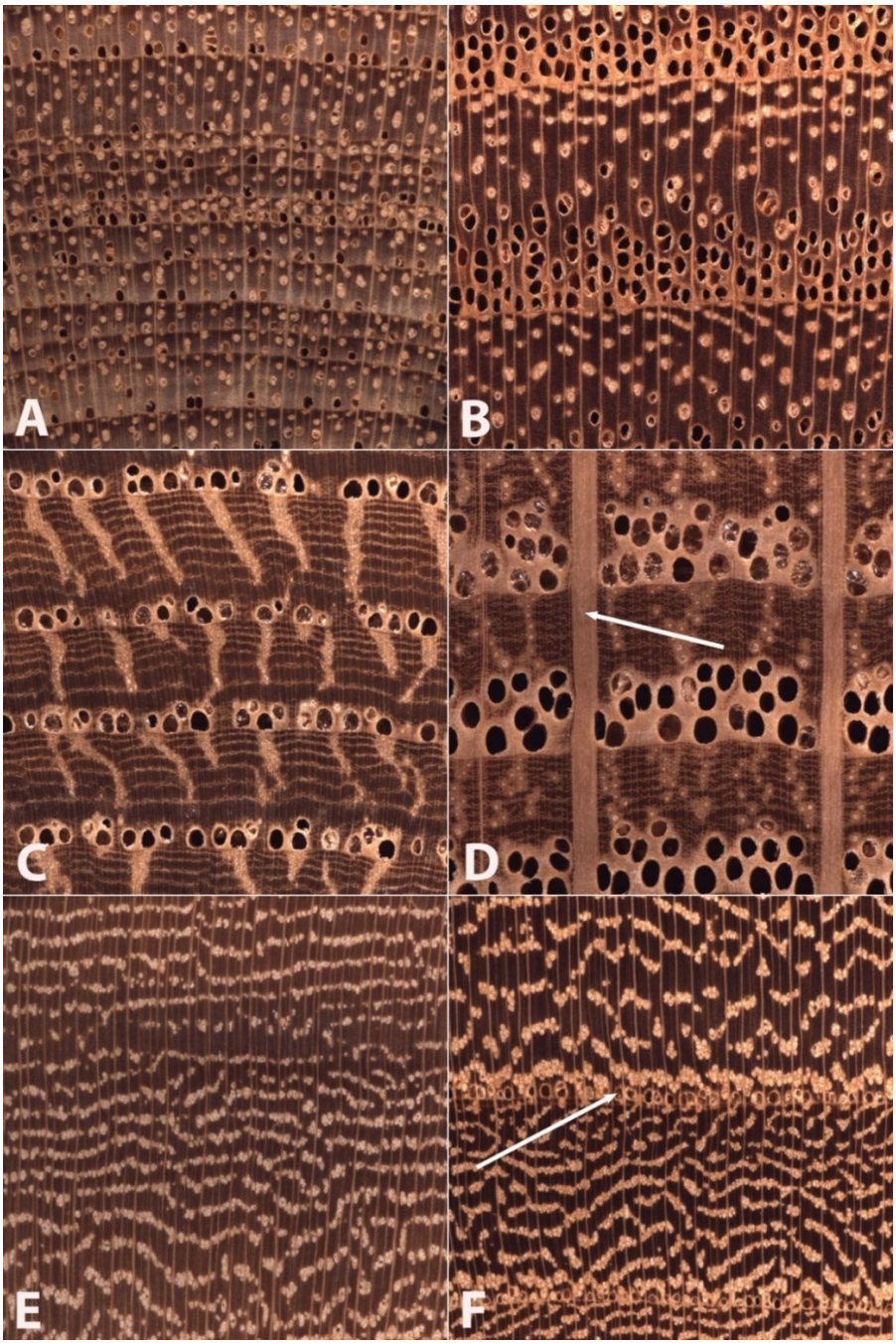

Figure 4.5 Images of the transverse surface of test specimens from classes Cladrastis, QuercusW, and UlmusS (A, C, and E) along with exemplar images from classes Gymnocladus (B), QuercusR (D), and UlmusH (F).

Note: Image pairs A and B, C and D, and E and F illustrates the misclassification within the slow-growth, no broad rays, and fast-growth spatial heterogeneity datasets.

4.6 Discussion

4.6.1 Deployment Gap of Cross-Validation and Field Testing

The deployment gap, the drop in accuracy (Ravindran et al., 2021) between the top-1 cross-validation and field-testing accuracy when tested on PACw and MSUtw specimens, was 6.6%. Previous studies by Ravindran et al. (2019, 2021) found deployment gaps of 25.0% and 10.5%, respectively, and in the diffuse-porous North American hardwoods, a deployment gap of 14.6% was reported (Ravindran et al., 2022). Research in other fields of computer vision have found a comparable loss in accuracy when models are tested on completely new datasets (Recht et al., 2018, 2019; Zech et al., 2018). According to Recht et al. (2019), there is a strong likelihood that models will struggle to generalize to images that present greater challenges than those in the original dataset. Other factors described in Ravindran et al. (2021) that might influence this deployment gap include minor variations in the anatomical patterns between xylarium specimens and the wood currently available in the market, differences between green and dry wood, variability in operator skills, and/or systematic differences imparted by different instantiations of imaging equipment. Using CVWID models in human-in-the-loop scenarios, such as the *xyloinf* classification software (Ravindran et al., 2020), provides users with the top predictions and exemplar images for the predicted classes, permitting the incorporation of human judgment. Additionally, organoleptic characters unavailable to the CVWID system, such as odor, luster, and taste, could serve as initial indicators of wood identity as well as assist in identification of Type 3 misclassifications by visual comparison of field images with representative images.

CVWID is typically formulated as an inductive learning problem where a model ($f: \mathcal{X} \rightarrow \mathcal{L}$) that maps images (from $\mathcal{X}$) to labels (from $\mathcal{L}$) is learned using labeled training data and the

quality of the trained model is evaluated on its capability to generalize to unseen test data. The test data are assumed to be drawn independent and identically distributed (i.i.d.) from the same distribution as the training data i.e., if P_t and P_d represent the training and deployment data distributions then $P_t(\mathcal{X}) = P_d(\mathcal{X})$ and $P_t(\mathcal{L}) = P_d(\mathcal{L})$. When the i.i.d. assumption is violated and distributional shifts between the training and testing/deployment data exist, *in silico* model performance will not translate to commensurate real-world field performance. Two types of distributional shifts can influence the real-world performance of deployed CVWID models: covariate shift ($P_t(\mathcal{X}) \neq P_d(\mathcal{X})$) and semantic shift ($P_t(\mathcal{L}) \neq P_d(\mathcal{L})$).

Differences in wood anatomy, sample preparation, imaging parameters, and operator skill can be sources of covariate shift. In this work, the use of the XyloTron platform to image progressively sanded wood specimen surfaces enabled the capture of consistent image data thereby minimizing covariate shifts due to specimen preparation and imaging. Our study pooled data from different wood collections (using multiple operators for specimen preparation and imaging) to train a model which was then tested on specimens from a different xylarium that did not contribute to the training data (a logistically manageable surrogate for real-world field testing), thereby enforcing separation between the training and testing datasets. The surrogate field testing approach used here (similar to Ravindran et al., 2022) is an initial attempt to incorporate covariate shifts due to operator skill (while following the progressive sample preparation protocol) in the evaluation of trained models. In the context of covariate shifts in relation to spatial heterogeneity of ring-porous woods, the Slow-Growth, Fast-Growth and Broad Rays Absent datasets were used to evaluate model performance with respect to the positioning of the imaging sensor on the specimen surface.

While our data set is the largest (in terms of number of images and unique specimens) for the considered classes and leads to models that are deployable in a human-in-the-loop setting, the observed deployment gap for the field model suggests the need for larger datasets for training models that better capture the inter- and intra-class wood anatomical variations. Our models (like most prior works but see Apolinario et al., 2019) were trained and evaluated based on a closed-world assumption i.e., there is no semantic shift, and the test specimen belongs to one of the classes the model was trained to identify. Augmented models that include a larger set of woods along with a "catch all" out-of-distribution class and/or reporting prediction uncertainties can be practical approaches for handling semantic shifts in the data distribution. Elucidating the interplay of dataset sizes, model capacities, and distributional shifts, especially relaxation of sample preparation protocols (e.g., using knife cuts instead of progressive sanding, or sanding to a coarser grit [thus involving fewer steps and less time]) and the closed-world assumption (Scheirer et al., 2013), is likely to be an important challenge in the realization of general field-deployable CVWID systems. We expect the exploration of these ideas (e.g., Mahdavi and Carvalho, 2021; Vaze et al., 2021; Yang et al., 2021) to be a fertile area for future work.

4.6.2 Spatial Heterogeneity

According to Panshin and de Zeeuw (1980) and Hoadley (1990), the initial discriminant macroscopic character commonly used in North American hardwood identification is porosity (ring-porous vs. diffuse-porous). The second character frequently invoked in such wood identification keys within the ring-porous hardwoods is the characteristic presence of wide rays in Quercus. By definition, vessels in ring-porous woods will display dramatic and abrupt changes in diameter between earlywood and latewood, as well as changes in the parenchyma patterns

from earlywood to latewood. It has been found that specimens with varying growth rates (slow-growth, medium-growth, and fast-growth (Fig. 4.1)) can have an impact on the appearance of both these features such that some macroscopic images of some fast-grown specimens may not capture the earlywood vessels, some images of Slow-Growth specimens may not capture latewood features, and some images of Quercus may lack wide rays. Our spatial heterogeneity datasets explicitly tested the influence of these features (or their lack), and the results suggest that this may not affect model predictions as strongly as anticipated (no change for Slow-Growth, 11.4% reduction for Fast-Growth, and 8.3% reduction for Broad Rays Absent). In the Slow-Growth dataset, a specimen from class Cladrastis was misclassified as class Gymnocladus, a Type 1 misclassification that was also observed in the five-fold dataset. Of the Fast-Growth dataset, all specimens in class UlmusS were predicted as UlmusH. This outcome was likely caused by the absence of the earlywood zone, where the prominent, continuous and sometimes multiple rows of earlywood vessels would have served to separate UlmusS from UlmusH. Interestingly, none of the images of QuercusW captured without broad rays (e.g. Fig. 4.1-B) was mistaken for Castenea despite the latter's striking resemblance to the former in absence of this distinguishing feature. Figure 4.5 shows an example comparison of misclassifications in each of the three spatial heterogeneity datasets.

4.6.3 Analysis of Misclassifications

Table 4.8 presents a summary of the misclassifications in the field model based on the confusion matrix in Figure 4.3. Source misclassification proportions were calculated on a per class basis (n = number of specimens in each class) and sink misclassification proportions were calculated as a percentage of the total number of misclassified specimens across all classes (n =

17). Our analysis found eight classes that have neither source nor sink misclassifications; they include Asimina, Carya, Castanea, Catalpa, Celtis, Fraxinus, Morus, and Sassafras. As for classes exhibiting Type 1 source misclassifications, they ranged from a 2.4% in QuercusR to 87.5% in UlmusS. The only Type 3 source misclassification came from QuercusR, where 4.9% of the specimens were misclassified as UlmusH. Among the five classes with sink misclassifications, UlmusH and QuercusR account for over 60% of the Type 1 errors. It is noteworthy that UlmusH was the only class to show a Type 3 misclassification, with 11.8% of specimens classified as Quercus R.

Table 4.8 Proportions of misclassifications in the top-1 predictions in the field model by class.

	Proportion of misclassified specimens by class.							
	Source				**Sink**			
Class (no. specimens)	Type 1	Type 2	Type 3	Total	Type 1	Type 2	Type 3	Total
Asimina (3)								
Carya (45)								
Castanea (2)								
Catalpa (4)								
Celtis (3)								
Cladrastis (4)	0.25			0.25				
Fraxinus (4)								
Gleditsia (7)	0.14			0.14				
Gymnocladus (2)					0.12			0.12
Maclura (3)	0.33			0.33				
Morus (5)								
QuercusR (41)	0.02		0.05	0.07	0.24			0.24
QuercusW (55)	0.07			0.07	0.06			0.06
Robinia (6)					0.06			0.06
Sassafras (3)								
UlmusH (3)					0.41		0.12	0.53
UlmusS (8)	0.88			0.88				

Note: Source misclassification proportions are based on the total number of specimens per class (shown in parentheses next to the class name). Sink misclassification proportions are based on the total number of misclassified specimens across all classes (n=17). Dark grey indicates a class for which there were neither source nor sink misclassifications; light grey indicates the absence of misclassifications in either source or sink; colored cells are proportions of note discussed in the text.

When analyzing the field-model confusion matrix (Fig. 4.3), two of 49 QuercusR specimens were classified as UlmusH. As noted above, this is classified as a Type 3 misclassification, but in reviewing the misclassified images, it might be that the features recognized by the model are emphasizing the relative similarity between the wavy bands of latewood parenchyma in QuercusR and the ulmiform latewood vessel arrangement in UlmusH,

and thus failing to emphasize the importance of earlywood vessel arrangement, ray abundance, and ray width. A human identifier would be unlikely to mistake these features for each other and thus would be unlikely to confuse these woods. This comparison can be seen in Figure 4.4 images C and D.

4.6.4 Toward a Unified Model for North American hardwoods.

. In continuity with our previous study on diffuse-porous hardwoods (Ravindran et. al., 2022), we are working to develop a unified CVWID model for North American commercial woods that covers the entire spectrum of porosity patterns. In addition to ring- and diffuse-porous woods, a unified model must incorporate semi-ring-porous woods into the label space design. As CVWID for semi-ring-porous woods is still unexplored, it represents a type of label space heterogeneity that requires parsing before a unified model can be developed.

As future studies lead us closer to model unification, it becomes increasingly important to evaluate model performance in a way that assesses accuracy and most closely approximates model deployment in the field. While CVWID models are commonly subjected to five-fold cross validation, the extent to which such models developed from inadequate (at best) and potentially misleading (at worst) datasets can misrepresent accuracy in the absence of testing with independent specimens has been demonstrated (Ravindran and Wiedenhoeft, 2022). To be sure, testing models with specimens from disparate xylaria not used for model training has been a useful and convenient surrogate for field testing, but the gold standard for CVWID model evaluation remains on-the-ground testing of actual commercial material.

As the number of classes increases to cover more of the commercial hardwood species in North America (somewhere around 40+), we expect to see the overall frequency of

misclassifications increase, as well as the frequency of Type 3 errors. Whereas a human trained in wood identification would rarely (if ever) mistake a ring-porous wood for a diffuse-porous wood, a 40+ class CVWID model might. For this reason, it is important to develop large, unified models in such a way as to reduce or eliminate those types of errors. In addition to incorporating domain expertise in designing the label space, this improvement could possibly be accomplished by varying CNN depth, applying penalty weights for out-of-genus misclassifications, or even nesting models inside others

4.7 Conclusions

. The CVWID model presented here is one of the first, and the largest, to be developed for ring-porous hardwoods in North America. A 17-class model was trained using 4045 images captured from 452 specimens of ring-porous woods from four xylaria to determine how well the model handles spatial heterogeneity. A five-fold cross validation showed a 98.0% accuracy while a field model tested on 198 specimens drawn from two additional xylaria achieved top-1 and top-2 predictions of 91.4% and 100%, respectively. Results tested on three smaller spatial heterogeneity datasets also showed that images devoid of earlywood, latewood, or broad rays did not greatly reduce prediction accuracy. This study emphasizes the continued importance of allowing wood anatomy to inform CVWID model creation and evaluation and advises against relying solely on CNN architecture to increase accuracy. In an ongoing study, we are working on developing a combined model for North American diffuse-porous and ring-porous hardwoods that will also examine how semi-ring-porous hardwoods (such as *Juglans*) affect the predictions of the model and the possibilities of getting computer-vision models to predict classes in a more anatomically informed way.

4.8 Acknowledgments

The authors wish to gratefully acknowledge the specimen preparation and imaging efforts of Nicholas Bargren, Karl Kleinschmidt, Caitlin Gilly, Richard Soares, Adriana Costa, and Flavio Ruffinatto.

4.9 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.10 Author Contributions

FO and RS provided access to and supervised data acquisition from the PACw and MSUtw test specimens. AW prepared and imaged the PACw and MSUtw specimens and prepared the initial draft of the paper. AW, FO, and ACW established the wood anatomical scope of the study. PR implemented the machine learning pipelines for the study. PR and ACW conducted data analysis and synthesis. PR, ACW, AW, and FO wrote the paper. All authors contributed actionable feedback that improved the presentation of the paper. AW: Adam Wade, ACW: Alex C. Wiedenhoeft.

4.11 Funding

This work was supported in part by a grant from the US Department of State via Interagency Agreement number 19318814Y0010 to ACW and in part by research funding from the Forest Stewardship Council to ACW. PR was partially supported by a Wisconsin Idea Baldwin Grant. The authors wish to acknowledge the support of U.S. Department of Agriculture (USDA), Research, Education, and Economics (REE), Agriculture Research Service (ARS), Administrative and Financial Management (AFM), Financial Management and Accounting Division (FMAD) Grants and Agreements Management Branch (GAMB), under Agreement No.

58-0204-9-164, specifically for support of AW, FO, RS. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

4.12 Data Availability Statement

A minimal dataset can be obtained by contacting the corresponding author, but the full training dataset used in the study is protected for up to five years by a CRADA between FPL, UW-Madison and FSC.

4.13 Code Availability

The software apps for image dataset collection and trained model deployment along with the weights of the trained model will be made available at https://github.com/fpl-xylotron.

4.14 References

Apolinario, M. P. E., Urcia Paredes, D. A., and Huaman Bustamante, S. G. (2019). Open Set Recognition of Timber Species Using Deep Learning for Embedded Systems. *IEEE Lat. Am. Trans*. 17, 2005–2012. doi: 10.1109/TLA.2019.9011545.

Arévalo, R., Pulido R., E. N., Solórzano G., J. F., Soares, R., Ruffinatto, F., Ravindran, P., et al. (2021). Imaged based identification of colombian timbers using the xylotron: A proof of concept international partnership. *Colomb. For*. 24, 5–16. doi: 10.14483/2256201X.16700.

Barbosa, A. C. F., Gerolamo, C. S., Lima, A. C., Angyalossy, V., and Pace, M. R. (2021). Polishing entire stems and roots using sandpaper under water: An alternative method for macroscopic analyses. *Appl. Plant Sci*. 9. doi: 10.1002/aps3.11421.

Damayanti, R., Prakasa, E., Krisdianto, Dewi, L. M., Wardoyo, R., Sugiarto, B., et al. (2019). LignoIndo: Image database of Indonesian commercial timber. in *IOP Conference Series: Earth and Environmental Science* (IOP Publishing), 012057. doi: 10.1088/1755-1315/374/1/012057.

de Andrade, B. G., Basso, V. M., and de Figueiredo Latorraca, J. V. (2020). Machine vision for field-level wood identification. *IAWA J*. 41, 681–698. doi: 10.1163/22941932-bja10001.

de Geus, A. R., Backes, A. R., Gontijo, A. B., Albuquerque, G. H. Q., and Souza, J. R. (2021). Amazon wood species classification: a comparison between deep learning and pre-designed features. *Wood Sci. Technol*. 55, 857–872. doi: 10.1007/s00226-021-01282-w.

DeVries, T., and Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. doi: 10.48550/arxiv.1708.04552.

Figueroa-Mata, G., Mata-Montero, E., Valverde-Otarola, J. C., and Arias-Aguilar, D. (2018). Automated Image-based Identification of Forest Species: Challenges and Opportunities for 21st Century Xylotheques. in 2018 *IEEE International Work Conference on Bioinspired Intelligence, IWOBI 2018 - Proceedings* (Institute of Electrical and Electronics Engineers Inc.). doi: 10.1109/IWOBI.2018.8464206.

Filho, P. L. P., Oliveira, L. S., Nisgoski, S., and Britto, A. S. (2014). Forest species recognition using macroscopic images. *Mach. Vis. Appl*. 25, 1019–1031. doi: 10.1007/S00138-014-0592-7/TABLES/8.

Gasson, P. (2011). How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially cites. *IAWA J*. 32, 137–154. doi: 10.1163/22941932-90000049.

Hoadley, R.B. (1990). Identifying wood: accurate results with simple tools. Taunton Press. p. 223.

Howard, J., and Gugger, S. (2020). Fastai: A layered API for deep learning. *Inf.* 11, 108. doi: 10.3390/info11020108.

Hwang, S. W., and Sugiyama, J. (2021). Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review. *Plant Methods* 17, 1–21. doi: 10.1186/s13007-021-00746-1.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. in *32nd International Conference on Machine Learning, ICML 2015* (International Machine Learning Society (IMLS)), 448–456. doi: 10.48550/arxiv.1502.03167.

Khalid, M., Lee, E. L. Y., Yusof, R., and Nadaraj, M. (2008). Design of an Intelligent Wood Species Recognition System. *Int. J. Simul. Syst. Sci. Technol*. 9, 9–19.

Kingma, D. P., and Ba, J. L. (2015). Adam: A method for stochastic optimization. in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (International Conference on Learning Representations, ICLR). doi: 10.48550/arxiv.1412.6980.

Lopes, D. J. V., Burgreen, G. W., and Entsminger, E. D. (2020). North American hardwoods identification using machine-learning. *Forests* 11, 298. doi: 10.3390/f11030298.

Mahdavi, A., and Carvalho, M. (2021). A Survey on Open Set Recognition. in *Proceedings - 2021 IEEE 4th International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2021* (Institute of Electrical and Electronics Engineers Inc.), 37–44. doi: 10.1109/AIKE52691.2021.00013.

Martins, J., Oliveira, L. S., Nisgoski, S., and Sabourin, R. (2013). A database for automatic classification of forest species. *Mach. Vis. Appl*. 24, 567–578. doi: 10.1007/s00138-012-0417-5.

Miller, R., Wiedenhoeft, A., and Ribeyron, M. J. (2002). *CITES Identification Guide – Tropical Woods*. Environment Canada, Canada.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng*. 22, 1345–1359. doi: 10.1109/TKDE.2009.191.

Panshin, A.J., de Zeeuw, C. (1980). *Textbook of Wood Technology: Structure, identification, properties, and uses of the commercial woods of the United States and Canada*. 4th ed. McGraw-Hill Series in Forest Resources. New York, McGraw-Hill Book Co.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems* (Neural information processing systems foundation). doi: 10.48550/arxiv.1912.01703.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res*., 12(85):2825−2830.

Ravindran, P., and Wiedenhoeft, A. C. (2020). Comparison of two forensic wood identification technologies for ten Meliaceae woods: computer vision versus mass spectrometry. *Wood Sci. Technol.* 54, 1139–1150. doi: 10.1007/s00226-020-01178-1.

Ravindran, P., and Wiedenhoeft, A. C. (2022). Caveat emptor: On the Need for Baseline Quality Standards in Computer Vision Wood Identification. *Forests* 13, 632. doi: 10.3390/f13040632.

Ravindran, P., Costa, A., Soares, R., and Wiedenhoeft, A. C. (2018). Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14, 1–10. doi: 10.1186/s13007-018-0292-9.

Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R. (2019). Image Based Identification of Ghanaian Timbers Using the XyloTron: Opportunities, Risks and Challenges. in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (Vancouver, Canada). doi: 10.48550/arxiv.1912.00296.

Ravindran, P., Owens, F. C., Wade, A. C., Shmulsky, R., and Wiedenhoeft, A. C. (2022). Towards Sustainable North American Wood Product Value Chains, Part I: Computer Vision Identification of Diffuse Porous Hardwoods. *Front. Plant Sci*. 12, 3306. doi: 10.3389/fpls.2021.758455.

Ravindran, P., Owens, F. C., Wade, A. C., Vega, P., Montenegro, R., Shmulsky, R., et al. (2021). Field-Deployable Computer Vision Wood Identification of Peruvian Timbers. *Front. Plant Sci*. 12, 940. doi: 10.3389/fpls.2021.647515.

Ravindran, P., Thompson, B. J., Soares, R. K., and Wiedenhoeft, A. C. (2020). The XyloTron: Flexible, Open-Source, Image-Based Macroscopic Field Identification of Wood Products. *Front. Plant Sci*. 11, 1015. doi: 10.3389/fpls.2020.01015.

Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2018). Do CIFAR-10 Classifiers Generalize to CIFAR-10? doi: 10.48550/arxiv.1806.00451.

Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet? in *36th International Conference on Machine Learning*, ICML 2019 (International Machine Learning Society (IMLS)), 9413–9424. doi: 10.48550/arxiv.1902.10811.

Ruffinatto, F., Crivellaro, A., and Wiedenhoeft, A. C. (2015). Review of macroscopic features for hardwood and softwood identification and a proposal for a new character list. *IAWA J.* 36, 208–241. doi: 10.1163/22941932-00000096.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis*. 115, 211–252. doi: 10.1007/s11263-015-0816-y.

Scheirer, W. J., De Rezende Rocha, A., Sapkota, A., and Boult, T. E. (2013). Toward open set recognition. IEEE Trans. *Pattern Anal. Mach. Intell*. 35, 1757–1772. doi: 10.1109/TPAMI.2012.256.

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. doi: 10.48550/arxiv.1803.09820.

Souza, D. V., Santos, J. X., Vieira, H. C., Naide, T. L., Nisgoski, S., and Oliveira, L. E. S. (2020). An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. *Wood Sci. Technol*. 54, 1065–1090. doi: 10.1007/s00226-020-01196-z.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res*. 15, 1929–1958.

Vaze, S., Han, K., Vedaldi, A., and Zisserman, A. (2021). Open-Set Recognition: a Good Closed-Set Classifier is All You Need? doi: 10.48550/arxiv.2110.06207.

Wheeler, E. A., LaPasha, C. A., and Miller, R. B. (1989). Wood anatomy of elm (Ulmus) and hackberry (Celtis) species native to the united states1. *IAWA J*. 10, 5–26. doi: 10.1163/22941932-90001106.

Wiedenhoeft, A. C. (2020). The xylophone: Toward democratizing access to high-quality macroscopic imaging for wood and other substrates. *IAWA J*. 41, 699–719. doi: 10.1163/22941932-bja10043.

Wiedenhoeft, A. C., Simeone, J., Smith, A., Parker-Forney, M., Soares, R., and Fishman, A. (2019). Fraud and misrepresentation in retail forest products exceeds U.S. Forensic wood science capacity. *PLoS One* 14, e0219917. doi: 10.1371/journal.pone.0219917.

Wu, F., Gazo, R., Haviarova, E., and Benes, B. (2021). Wood identification based on longitudinal section images by using deep learning. *Wood Sci. Technol*. 55, 553–563. doi: 10.1007/s00226-021-01261-1.

Yang, J., Zhou, K., Li, Y., and Liu, Z. (2021). Generalized Out-of-Distribution Detection: A Survey. doi: 10.48550/arxiv.2110.11334.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., and Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. 15. doi: 10.1371/journal.pmed.1002

APPENDIX A

SUPPLEMENTARY MATERIAL

A.1 Supplementary Material for Chapter 2

A.1.1 Supplementary Material 1: The 24 Peruvian Woods Selected

Table A.1 Volume of round wood of selected Peruvian species

VOLUME OF ROUND WOOD OF SELECTED SPECIES		
GENUS	**TOTAL VOLUME (m3)**	**% of Total Volume Analyzed**
Dipteryx	143,592.12	19.52%
Cedrelinga	83,436.12	11.34%
Cariniana	76,283.40	10.37%
Guazuma	70,056.20	9.52%
Maquira	67,583.58	9.19%
Copaifera	53,754.59	7.31%
Brosimum	40,141.17	5.46%
Eucalyptus	34,717.20	4.72%
Virola	24,172.74	3.29%
Myroxylon	19,035.69	2.59%
Hura crepitans	18,831.12	2.56%
Aspidosperma	16,922.09	2.30%
Calycophyllum	16,675.37	2.27%
Amburana	16,313.51	2.22%
Schizolobuim	12,970.92	1.76%
Pinus	11,743.42	1.60%
Poulsenia	8,943.72	1.22%
Cedrela	6,575.55	0.89%
Ormosia	4,804.40	0.65%
Aniba	4,531.20	0.62%
Iryanthera	2,463.04	0.33%
Swietenia	1,376.99	0.19%
Chorisia	576.89	0.08%
Pouteria		0.00%
	735,501.07	

VOLUME OF ROUND WOOD OF SELECTED SPECIES COMPARED TO TOTAL VOLUME 2019		
GENUS	**TOTAL VOLUME (m3)**	**% of Total Volume Produced**
Dipteryx	143,592.12	11.09%
Cedrelinga	83,436.12	6.45%
Cariniana	76,283.40	5.89%
Guazuma	70,056.20	5.41%
Maquira	67,583.58	5.22%
Copaifera	53,754.59	4.15%
Brosimum	40,141.17	3.10%
Eucalyptus	34,717.20	2.68%
Virola	24,172.74	1.87%
Myroxylon	19,035.69	1.47%
Hura crepitans	18,831.12	1.45%
Aspidosperma	16,922.09	1.31%
Calycophyllum	16,675.37	1.29%
Amburana	16,313.51	1.26%
Schizolobuim	12,970.92	1.00%
Pinus	11,743.42	0.91%
Poulsenia	8,943.72	0.69%
Cedrela	6,575.55	0.51%
Ormosia	4,804.40	0.37%
Aniba	4,531.20	0.35%
Iryanthera	2,463.04	0.19%
Swietenia	1,376.99	0.11%
Chorisia	576.89	0.04%
Pouteria		0.00%
Analyzed volume	735,501.07	57%
Total volume 2019	**1,294,501.64**	

Table A.2 Volume of lumber of selected Peruvian species

VOLUME OF LUMBER OF SELECTED SPECIES		
GENUS	**TOTAL VOLUME (m3)**	**% of Total Volume Analyzed**
Cedrelinga	80,760.62	20.55%
Dipteryx	58,773.05	14.95%
Cariniana	43,265.26	11.01%
Copaifera	33,851.90	8.61%
Guazuma	27,711.82	7.05%
Brosimum	26,156.52	6.65%
Virola	23,140.46	5.89%
Hura crepitans	16,847.34	4.29%
Calycophyllum	11,719.18	2.98%
Eucalyptus	9,465.67	2.41%
Aspidosperma	9,150.05	2.33%
Pinus	7,571.08	1.93%
Myroxylon	6,890.35	1.75%
Amburana	5,907.81	1.50%
Schizolobuim	5,713.90	1.45%
Poulsenia	5,646.56	1.44%
Maquira	5,611.33	1.43%
Cedrela	5,187.25	1.32%
Aniba	3,417.98	0.87%
Ormosia	2,641.89	0.67%
Iryanthera	1,618.11	0.41%
Chorisia	1,378.06	0.35%
Swietenia	645.38	0.16%
Pouteria		0.00%
	393,071.57	

VOLUME OF LUMBER OF SELECTED SPECIES COMPARED TO TOTAL VOLUME 2019		
GENUS	**TOTAL VOLUME (m3)**	**% of Total Volume Produced**
Cedrelinga	80,760.62	13.54%
Dipteryx	58,773.05	9.85%
Cariniana	43,265.26	7.25%
Copaifera	33,851.90	5.67%
Guazuma	27,711.82	4.65%
Brosimum	26,156.52	4.38%
Virola	23,140.46	3.88%
Hura crepitans	16,847.34	2.82%
Calycophyllum	11,719.18	1.96%
Eucalyptus	9,465.67	1.59%
Aspidosperma	9,150.05	1.53%
Pinus	7,571.08	1.27%
Myroxylon	6,890.35	1.16%
Amburana	5,907.81	0.99%
Schizolobuim	5,713.90	0.96%
Poulsenia	5,646.56	0.95%
Maquira	5,611.33	0.94%
Cedrela	5,187.25	0.87%
Aniba	3,417.98	0.57%
Ormosia	2,641.89	0.44%
Iryanthera	1,618.11	0.27%
Chorisia	1,378.06	0.23%
Swietenia	645.38	0.11%
Pouteria		0.00%
Analyzed volume	393,071.57	66%
Total volume 2019	**596,547.14**	

A.1.2 Supplementary Material 2: Class Composition Details

Across the training and testing data sets a total of 188 taxa were partitioned into 24 classes based on wood anatomy similarities. For model training the taxa were grouped into genus level classes. It should be noted that some of the species included in the test dataset were not part of the training dataset. The species of the genus *Brosimum* were grouped into two classes: BrosimumA (anatomy similar to *Brosimum alicastrum*) and BrosimumU (anatomy similar to *Brosimum utile*).

In the table below, the column “Accession Taxon" lists the taxa of the specimen as they are recorded in the xylarium (and used for label space design) while the column “Verified Taxon" lists the taxa as verified using the online data base http://www.worldoraonline.org/.

We would like to highlight and provide clarification on two specific cases of taxonomic flux encountered during the verification:

- The correct taxonomic designation of *Copaifera chodatiana* is *Guibortia chodatiana*. There was exactly one *Copaifera chodatiana* specimen that contributed four images to our data set. In our five-fold cross validation experiments, this specimen (when its images were part of the testing fold) was misclassified (both at top-1 and top-2 predictions), which is the expected correct behavior.
- Images from exactly one specimen of *Eucalyptus trachyphloia*, from the PACw xylarium, was included in our testing data set. The correct taxonomic designation of *Eucalyptus trachyphloia* is *Corymbia trachyphloia*. The wood anatomy of *Corymbia trachyphloia*, at the hand lens level, is consistent with that of

Eucalyptus. This specimen was correctly identified by our model as class Eucalyptus at the top-2 prediction.

In conclusion, our label space design and accuracy results are both valid and consistent after accounting for taxonomic flux. The authors would like to gratefully acknowledge the reviewer who suggested accounting for taxonomic flux issues.

Table A.3 The class labels and their constituent taxa.

Label	**Accession Taxon**	**Verified Taxon**	**Train**	**Test**
Amburana	*Amburana cearensis*	*Amburana cearensis*	X	X
Aniba	*Aniba bracteata*	*Aniba bracteata*	X	
Aniba	*Aniba canelilla*	*Aniba canellila*	X	
Aniba	*Aniba hostmanniana*	*Aniba hostmanniana*	X	
Aniba	*Aniba nearparurensis*	*Aniba panurensis*		X
Aniba	*Aniba riparia*	*Aniba riparia*	X	
Aniba	*Aniba rosaeodora*	*Aniba rosaeodora*	X	
Aniba	*Aniba trinitatis*	*Aniba citrifolia*	X	
Aspidosperma	*Aspidosperma album*	*Aspidosperma album*	X	
Aspidosperma	*Aspidosperma centrala*	*Aspidosperma album*	X	
Aspidosperma	*Aspidosperma cruentum*	*Aspidosperma desmanthum*	X	
Aspidosperma	*Aspidosperma cylindrocarpon*	*Aspidosperma cylindrocarpon*	X	
Aspidosperma	*Aspidosperma desmanthum*	*Aspidosperma desmanthum*	X	
Aspidosperma	*Aspidosperma macrocarpon*	*Aspidosperma macrocarpon*	X	
Aspidosperma	*Aspidosperma marcgravianum*	*Aspidosperma excelsum*	X	
Aspidosperma	*Aspidosperma megalocarpon*	*Aspidosperma megalocarpon*	X	
Aspidosperma	*Aspidosperma melanocalix*	*Aspidosperma spruceanum*	X	
Aspidosperma	*Aspidosperma parvifolium*	*Aspidosperma parvifolium*	X	
Aspidosperma	*Aspidosperma polyneuron*	*Aspidosperma polyneuron*	X	
Aspidosperma	*Aspidosperma rigidum*	*Aspidosperma rigidum*	X	
Aspidosperma	*Aspidosperma* sp.	*Aspidosperma* sp.		X
BrosimumA	*Brosimum alicastrum*	*Brosimum alicastrum*	X	
BrosimumA	*Brosimum colimbianum*	*Brosimum alicastrum*		X
BrosimumA	*Brosimum latifolium*	*Brosimum alicastrum*		X
BrosimumA	*Brosimum terrabanum*	*Brosimum alicastrum*		X
BrosimumU	*Brosimum utile*	*Brosimum utile*	X	
Calycophyllum	*Calycophyllum acreanum*	*Calycophyllum megistocaulum*	X	
Calycophyllum	*Calycophyllum candidissimum*	*Calycophyllum candidissimum*	X	X
Calycophyllum	*Calycophyllum multiflorum*	*Calycophyllum multiflorum*	X	X
Calycophyllum	*Calycophyllum obovatum*	*Calycophyllum obovatum*	X	
Calycophyllum	*Calycophyllum spruceanum*	*Calycophyllum spruceanum*	X	X
Amburana	*Amburana cearensis*	*Amburana cearensis*	X	X
Aniba	*Aniba bracteata*	*Aniba bracteata*	X	
Aniba	*Aniba canelilla*	*Aniba canellila*	X	
Aniba	*Aniba hostmanniana*	*Aniba hostmanniana*	X	

Table A.3 (continued)

Label	Accession Taxon	Verified Taxon	Train	Test
Aniba	*Aniba nearparurensis*	*Aniba panurensis*		X
Aniba	*Aniba riparia*	*Aniba riparia*	X	
Aniba	*Aniba rosaeodora*	*Aniba rosaeodora*	X	
Aniba	*Aniba trinitatis*	*Aniba citrifolia*	X	
Aspidosperma	*Aspidosperma album*	*Aspidosperma album*	X	
Aspidosperma	*Aspidosperma centrala*	*Aspidosperma album*	X	
Aspidosperma	*Aspidosperma cruentum*	*Aspidosperma desmanthum*	X	
Aspidosperma	*Aspidosperma cylindrocarpon*	*Aspidosperma cylindrocarpon*	X	
Aspidosperma	*Aspidosperma desmanthum*	*Aspidosperma desmanthum*	X	
Aspidosperma	*Aspidosperma macrocarpon*	*Aspidosperma macrocarpon*	X	
Aspidosperma	*Aspidosperma marcgravianum*	*Aspidosperma excelsum*	X	
Aspidosperma	*Aspidosperma megalocarpon*	*Aspidosperma megalocarpon*	X	
Aspidosperma	*Aspidosperma melanocalix*	*Aspidosperma spruceanum*	X	
Aspidosperma	*Aspidosperma parvifolium*	*Aspidosperma parvifolium*	X	
Aspidosperma	*Aspidosperma polyneuron*	*Aspidosperma polyneuron*	X	
Aspidosperma	*Aspidosperma rigidum*	*Aspidosperma rigidum*	X	
Aspidosperma	*Aspidosperma* sp.	*Aspidosperma* sp.		X
BrosimumA	*Brosimum alicastrum*	*Brosimum alicastrum*	X	
BrosimumA	*Brosimum colimbianum*	*Brosimum alicastrum*		X
BrosimumA	*Brosimum latifolium*	*Brosimum alicastrum*		X
BrosimumA	*Brosimum terrabanum*	*Brosimum alicastrum*		X
BrosimumU	*Brosimum utile*	*Brosimum utile*	X	
Calycophyllum	*Calycophyllum acreanum*	*Calycophyllum megistocaulum*	X	
Calycophyllum	*Calycophyllum candidissimum*	*Calycophyllum candidissimum*	X	X
Calycophyllum	*Calycophyllum multiflorum*	*Calycophyllum multiflorum*	X	X
Calycophyllum	*Calycophyllum obovatum*	*Calycophyllum obovatum*	X	
Calycophyllum	*Calycophyllum spruceanum*	*Calycophyllum spruceanum*	X	X
Cariniana	*Cariniana domestica*	*Cariniana domestica*	X	
Cariniana	*Cariniana estrellensis*	*Cariniana estrellensis*	X	
Cariniana	*Cariniana excelsa*	*Cariniana estrellensis*	X	
Cariniana	*Cariniana exigua*	*Cariniana pyriformis*		X
Cariniana	*Cariniana legalis*	*Cariniana legalis*	X	X
Cariniana	*Cariniana micrantha*	*Cariniana micrantha*	X	
Cariniana	*Cariniana pyriformis*	*Cariniana pyriformis*	X	X
Cedrela	*Cedrela angustifolia*	*Cedrela angustifolia*	X	
Cedrela	*Cedrela fissilis*	*Cedrela fissilis*	X	X
Cedrela	*Cedrela lilloi*	*Cedrela angustifolia*	X	
Cedrela	*Cedrela montana*	*Cedrela montana*	X	
Cedrela	*Cedrela odorata*	*Cedrela odorata*	X	X
Cedrela	*Cedrela rosei*	*Cedrela rosei*		X
Cedrela	*Cedrela* sp.	*Cedrela* sp.		X
Cedrela	*Cedrela whitfordii*	*Cedrela odorata*		X
Cedrelinga	*Cedrelinga cateniformis*	*Cedrelinga cateniformis*	X	
Cedrela	*Cedrela odorata*	*Cedrela odorata*	X	X
Cedrela	*Cedrela rosei*	*Cedrela rosei*		X
Cedrela	*Cedrela sp.*	*Cedrela sp.*		X
Cedrela	*Cedrela whitfordii*	*Cedrela odorata*		X

Table A.3 (continued)

Label	Accession Taxon	Verified Taxon	Train	Test

Cedrelinga	*Cedrelinga cateniformis*	*Cedrelinga cateniformis*	X	
Ceiba	*Ceiba pentandra*	*Ceiba pentandra*	X	
Ceiba	*Ceiba samauma*	*Ceiba samauma*	X	
Ceiba	*Ceiba speciosa*	*Ceiba speciosa*	X	
Ceiba	*Chorisia insignis*	*Ceiba insignis*	X	
Ceiba	*Chorisia integrifolia*	*Ceiba insignis*	X	
Ceiba	*Chorisia speciosa*	*Ceiba speciosa*	X	
Copaifera	*Copaifera aromatica*	*Copaifera aromatica*	X	
Copaifera	*Copaifera canime*	*Copaifera canime*	X	
Copaifera	*Copaifera chiriquensis*	*Copaifera sp.*	X	
Copaifera	*Copaifera chodatiana*	*Guibourtia chodatiana*	X	
Copaifera	*Copaifera guianensis*	*Copaifera guyanensis*	X	
Copaifera	*Copaifera langsdorffii*	*Copaifera langsdorffii*	X	
Copaifera	*Copaifera langsdorfii*	*Copaifera langsdorffii*	X	
Copaifera	*Copaifera majorina*	*Copaifera majorina*	X	
Copaifera	*Copaifera martii*	*Copaifera martii*	X	
Copaifera	*Copaifera multijuga*	*Copaifera multijuga*	X	
Copaifera	*Copaifera officinalis*	*Copaifera officinalis*	X	
Copaifera	*Copaifera panamensis*	*Copaifera panamensis*	X	
Copaifera	*Copaifera religiosa*	*Copaifera religiosa*	X	
Copaifera	*Copaifera reticulata*	*Copaifera reticulata*	X	X
Copaifera	*Copaifera trapezifolia*	*Copaifera trapezifolia*	X	
Dipteryx	*Dipteryx alata*	*Dipteryx alata*	X	
Dipteryx	*Dipteryx cf-ferrea*	*Dipteryx micrantha*	X	
Dipteryx	*Dipteryx magnifica*	*Dipteryx magnifica*	X	
Dipteryx	*Dipteryx micrantha*	*Dipteryx micrantha*	X	
Dipteryx	*Dipteryx odorata*	*Dipteryx odorata*	X	X
Dipteryx	*Dipteryx oleifera*	*Dipteryx oleifera*	X	
Dipteryx	*Dipteryx polyphylla*	*Dipteryx polyphylla*	X	
Dipteryx	*Dipteryx punctata*	*Dipteryx punctata*	X	
Dipteryx	*Dipteryx rosea*	*Dipteryx rosea*	X	
Dipteryx	*Dipteryx sp.*	*Dipteryx sp.*	X	
Eucalyptus	*Eucalyptus acmenioides*	*Eucalyptus acmenoides*		X
Eucalyptus	*Eucalyptus botryoides*	*Eucalyptus botryoides*		X
Eucalyptus	*Eucalyptus camaldulensis*	*Eucalyptus camaldulensis*	X	
Eucalyptus	*Eucalyptus capitellata*	*Eucalyptus capitellata*		X
Eucalyptus	*Eucalyptus cornuta*	*Eucalyptus cornuta*		X
Eucalyptus	*Eucalyptus dawsonii*	*Eucalyptus dawsonii*		X
Eucalyptus	*Eucalyptus deglupta*	*Eucalyptus deglupta*	X	X
Eucalyptus	*Eucalyptus delegatensis*	*Eucalyptus delegatensis*		X
Eucalyptus	*Eucalyptus diversicolor*	*Eucalyptus diversicolor*		X
Eucalyptus	*Eucalyptus eugenoides*	*Eucalyptus eugenioides*		X
Eucalyptus	*Eucalyptus gigantea*	*Eucalyptus globulus*		X
Eucalyptus	*Eucalyptus gomphocephala*	*Eucalyptus gomphocephala*		X
Eucalyptus	*Eucalyptus delegatensis*	*Eucalyptus delegatensis*		X
Eucalyptus	*Eucalyptus diversicolor*	*Eucalyptus diversicolor*		X
Eucalyptus	*Eucalyptus eugenoides*	*Eucalyptus eugenioides*		X

Table A.3 (continued)

Label	**Accession Taxon**	**Verified Taxon**	**Train**	**Test**
Eucalyptus	*Eucalyptus gigantea*	*Eucalyptus globulus*		X

Eucalyptus	*Eucalyptus gomphocephala*	*Eucalyptus gomphocephala*		X
Eucalyptus	*Eucalyptus grandis*	*Eucalyptus grandis*	X	
Eucalyptus	*Eucalyptus hemiphloia*	*Eucalyptus moluccana*		X
Eucalyptus	*Eucalyptus longifolia*	*Eucalyptus elata*		X
Eucalyptus	*Eucalyptus macrorhyncha*	*Eucalyptus macrorhyncha*		X
Eucalyptus	*Eucalyptus maculata*	*Corymbia maculata*		X
Eucalyptus	*Eucalyptus marginata*	*Eucalyptus marginata*	X	X
Eucalyptus	*Eucalyptus muelleriana*	*Eucalyptus muelleriana*		X
Eucalyptus	*Eucalyptus obliqua*	*Eucalyptus obliqua*	X	X
Eucalyptus	*Eucalyptus pilularis*	*Eucalyptus pilularis*	X	X
Eucalyptus	*Eucalyptus propinqua*	*Eucalyptus propinqua*		X
Eucalyptus	*Eucalyptus redunea*	*Eucalyptus redunca*		X
Eucalyptus	*Eucalyptus regnans*	*Eucalyptus regnans*	X	
Eucalyptus	*Eucalyptus robusta*	*Eucalyptus robusta*	X	
Eucalyptus	*Eucalyptus rostrata*	*Eucalyptus camaldulensis*		X
Eucalyptus	*Eucalyptus rudderi*	*Eucalyptus rudderi*		X
Eucalyptus	*Eucalyptus saligna*	*Eucalyptus saligna*	X	X
Eucalyptus	*Eucalyptus siderophloia*	*Eucalyptus siderophloia*		X
Eucalyptus	*Eucalyptus sieberiana*	*Eucalyptus sieberi*		X
Eucalyptus	*Eucalyptus tereticornis*	*Eucalyptus tereticornis*		X
Eucalyptus	*Eucalyptus trachyphloia*	*Corymbia trachyphloia*		X
Guazuma	*Guazuma crinita*	*Guazuma crinita*	X	
Guazuma	*Guazuma guazuma*	*Guazuma ulmifolia*	X	
Guazuma	*Guazuma rosea*	*Guazuma crinita*	X	
Guazuma	*Guazuma* sp.	*Guazuma* sp.	X	
Guazuma	*Guazuma tomentosa*	*Guazuma ulmifolia*	X	
Guazuma	*Guazuma ulmifolia*	*Guazuma ulmifolia*	X	
Hura	*Hura crepitans*	*Hura crepitans*	X	
Maquira	*Maquira calophylla*	*Maquira calophylla*	X	
Maquira	*Maquira coriacea*	*Maquira coriacea*	X	
Maquira	*Maquira costaricana*	*Maquira guianensis*	X	
Maquira	*Maquira guianensis*	*Maquira guianensis*	X	
Maquira	*Maquira sclerophylla*	*Maquira sclerophylla*	X	
Maquira	*Maquira* sp.	*Maquira* sp.	X	
Myroxylon	*Myroxylon balsamum*	*Myroxylon balsamum*	X	
Myroxylon	*Myroxylon balsamun*	*Myroxylon balsamum*		X
Myroxylon	*Myroxylon peruiferum*	*Myroxylon peruiferum*	X	
Myroxylon	*Myroxylon toluiferum*	*Myroxylon balsamum*	X	
Ormosia	*Ormosia amazonica*	*Ormosia amazonica*	X	
Ormosia	*Ormosia avilensis*	*Ormosia avilensis*	X	
Ormosia	*Ormosia coccinea*	*Ormosia coccinea*	X	X
Ormosia	*Ormosia colombiana*	*Gongylolepis colombiana*	X	
Ormosia	*Ormosia costulata*	*Ormosia costulata*	X	
Ormosia	*Ormosia coutinhoi*	*Ormosia coutinhoi*	X	
Ormosia	*Ormosia dasycarpa*	*Ormosia monosperma*	X	
Ormosia	*Ormosia avilensis*	*Ormosia avilensis*	X	
Ormosia	*Ormosia coccinea*	*Ormosia coccinea*	X	X

Table A.3 (continued)

Label	**Accession Taxon**	**Verified Taxon**	**Train**	**Test**
Ormosia	*Ormosia colombiana*	*Gongylolepis colombiana*	X	

Ormosia	*Ormosia costulata*	*Ormosia costulata*	X	
Ormosia	*Ormosia coutinhoi*	*Ormosia coutinhoi*	X	
Ormosia	*Ormosia dasycarpa*	*Ormosia monosperma*	X	
Ormosia	*Ormosia elata*	*Ormosia elata*	X	
Ormosia	*Ormosia excelsa*	*Ormosia excelsa*	X	X
Ormosia	*Ormosia flava*	*Ormosia flava*	X	X
Ormosia	*Ormosia isthmensis*	*Ormosia isthmensis*	X	
Ormosia	*Ormosia krugii*	*Ormosia krugii*	X	
Ormosia	*Ormosia larecajana*	*Ormosia larecajana*	X	
Ormosia	*Ormosia lignivalis*	*Ormosia lignivalvis*	X	
Ormosia	*Ormosia macrocalyx*	*Ormosia macrocalyx*	X	
Ormosia	*Ormosia monosperma*	*Ormosia monosperma*	X	
Ormosia	*Ormosia nobilis*	*Ormosia nobilis*	X	
Ormosia	*Ormosia pacycarpa*	*Ormosia pachycarpa*		X
Ormosia	*Ormosia panamensis*	*Ormosia panamensis*	X	X
Ormosia	*Ormosia paraensis*	*Ormosia paraensis*	X	
Ormosia	*Ormosia smithii*	*Ormosia smithii*	X	
Ormosia	*Ormosia stipularis*	*Ormosia stipularis*	X	
Ormosia	*Ormosia toledoana*	*Ormosia macrocalyx*	X	
Pinus	*Pinus ayacahuite*	*Pinus ayacahuite*	X	
Pinus	*Pinus caribaea*	*Pinus caribaea*	X	
Pinus	*Pinus hartwegii*	*Pinus hartwegii*	X	
Pinus	*Pinus montezumae*	*Pinus hartwegii*	X	
Pinus	*Pinus occidentalis*	*Pinus montezumae*	X	X
Pinus	*Pinus oocarpa*	*Pinus oocarpa*	X	
Pinus	*Pinus patula*	*Pinus patula*	X	
Pinus	*Pinus pseudostrobus*	*Pinus montezumae*	X	
Pinus	*Pinus tropicalis*	*Pinus tropicalis*	X	
Poulsenia	*Poulsenia armata*	*Poulsenia armata*	X	X
Pouteria	*Pouteria aff-gongrijpii*	*Pouteria gongrijpii*	X	
Pouteria	*Pouteria aff-reticulata*	*Planchonella reticulata*	X	
Pouteria	*Pouteria bilocularis*	*Pouteria bilocularis*	X	
Pouteria	*Pouteria caimito*	*Pouteria caimito*	X	
Pouteria	*Pouteria cuspidata*	*Pouteria cuspidata*	X	
Pouteria	*Pouteria decorticans*	*Pouteria decorticans*	X	
Pouteria	*Pouteria engleri*	*Pouteria engleri*	X	
Pouteria	*Pouteria filipes*	*Pouteria filipes*	X	
Pouteria	*Pouteria furcata*	*Pouteria furcata*	X	
Pouteria	*Pouteria glomerata*	*Pouteria glomerata*	X	
Pouteria	*Pouteria gomphiifolia*	*Pouteria gomphiifolia*	X	
Pouteria	*Pouteria guianensis*	*Pouteria guianensis*	X	
Pouteria	*Pouteria gutta*	*Pouteria torta*	X	
Pouteria	*Pouteria hispida*	*Pouteria hispida*	X	
Pouteria	*Pouteria jariensis*	*Pouteria jariensis*	X	
Pouteria	*Pouteria krukovii*	*Pouteria krukovii*	X	
Pouteria	*Pouteria lasiocarpa*	*Pouteria caimito*	X	
Pouteria	*Pouteria laurifolia*	*Pouteria caimito*	X	

Table A.3 (continued)

Label	**Accession Taxon**	**Verified Taxon**	**Train**	**Test**
Pouteria	*Pouteria guianensis*	*Pouteria guianensis*	X	
Pouteria	*Pouteria gutta*	*Pouteria torta*	X	

Pouteria	*Pouteria hispida*	*Pouteria hispida*	X	
Pouteria	*Pouteria jariensis*	*Pouteria jariensis*	X	
Pouteria	*Pouteria krukovii*	*Pouteria krukovii*	X	
Pouteria	*Pouteria lasiocarpa*	*Pouteria caimito*	X	
Pouteria	*Pouteria laurifolia*	*Pouteria caimito*	X	
Pouteria	*Pouteria macrocarpa*	*Pouteria multiflora*	X	
Pouteria	*Pouteria macrophylla*	*Pouteria macrophylla*	X	
Pouteria	*Pouteria multiflora*	*Pouteria multiflora*		X
Pouteria	*Pouteria ovata*	*Pouteria ramiflora*	X	
Pouteria	*Pouteria plicata*	*Pouteria plicata*	X	
Pouteria	*Pouteria procera*	*Pouteria procera*	X	
Pouteria	*Pouteria ramiflora*	*Pouteria ramiflora*	X	
Pouteria	*Pouteria reticulata*	*Pouteria reticulata*	X	
Pouteria	*Pouteria salicifolia*	*Pouteria salicifolia*	X	
Pouteria	*Pouteria solimoesensis*	*Pouteria hispida*	X	
Pouteria	*Pouteria surumuensis*	*Pouteria surumuensis*	X	
Pouteria	*Pouteria torta*	*Pouteria torta*	X	
Pouteria	*Pouteria torta-glabra*	*Pouteria torta*	X	
Pouteria	*Pouteria trichopoda*	*Pouteria hispida*	X	
Pouteria	*Pouteria trilocularis*	*Pouteria trilocularis*	X	
Pouteria	*Pouteria vestita*	*Sarcaulus vestitus*	X	
Schizolobium	*Schizolobium amazonicum*	*Schizolobium amazonicum*	X	
Schizolobium	*Schizolobium excelsum*	*Schizolobium parahyba*		X
Schizolobium	*Schizolobium parahyba*	*Schizolobium parahyba*	X	
Schizolobium	*Schizolobium parahybum*	*Schizolobium parahyba*		X
Schizolobuim	*Schizolobuim excelsum*	*Schizolobium parahyba*		X
Swietenia	*Swietenia macrophylla*	*Swietenia macrophylla*	X	X
Swietenia	*Swietenia sp.*	*Swietenia sp.*		X
Virola	*Virola bicuhyba*	*Virola bicuhyba*	X	
Virola	*Virola calophylla*	*Virola calophylla*	X	X
Virola	*Virola calophylloidea*	*Virola calophylloidea*	X	
Virola	*Virola carinata*	*Virola carinata*	X	
Virola	*Virola cuspidata*	*Virola elongata*		X
Virola	*Virola elongata*	*Virola elongata*	X	
Virola	*Virola flexuosa*	*Virola flexuosa*	X	
Virola	*Virola gardneri*	*Virola gardneri*	X	
Virola	*Virola gracilis*	*Virola surinamensis*	X	
Virola	*Virola guatemalensis*	*Virola guatemalensis*	X	
Virola	*Virola koschnyi*	*Virola koschnyi*	X	
Virola	*Virola loretensis*	*Virola loretensis*	X	X
Virola	*Virola macrocarpa*	*Virola macrocarpa*	X	
Virola	*Virola melinonii*	*Virola michelii*	X	
Virola	*Virola merendonia*	*Virola koschnyi*		X
Virola	*Virola michelii*	*Virola michelii*	X	
Virola	*Virola molissima*	*Virola mollissima*	X	X
Virola	*Virola multicostata*	*Virola multicostata*	X	

Table A.3 (continued)

Label	**Accession Taxon**	**Verified Taxon**	**Train**	**Test**
Virola	*Virola multinervia*	*Virola multinervia*	X	
Virola	*Virola melinonii*	*Virola michelii*	X	
Virola	*Virola merendonia*	*Virola koschnyi*		X

Virola	*Virola michelii*	*Virola michelii*	X	
Virola	*Virola molissima*	*Virola mollissima*	X	X
Virola	*Virola multicostata*	*Virola multicostata*	X	
Virola	*Virola multinervia*	*Virola multinervia*	X	
Virola	*Virola officinalis*	*Virola officinalis*	X	X
Virola	*Virola oleifera*	*Bicuiba oleifera*	X	
Virola	*Virola pavonis*	*Virola pavonis*	X	
Virola	*Virola sebifera*	*Virola sebifera*	X	X
Virola	*Virola sp.*	*Virola sp.*	X	
Virola	*Virola surinamensis*	*Virola surinamensis*	X	X
Virola	*Virola uaupensis*	*Virola elongata*		X
Virola	*Virola venosa*	*Virola venosa*	X	

A.2 Supplementary Material for Chapter 3

A.2.1 Supplementary Material 3: Class Composition Details

The 105 unique taxa were assigned to 22 anatomically relevant classes. The table below lists the taxa used and their class membership, along with their inclusion in the training/testing data set. With the exception of \Fruitwood", all other classes contain species from exactly one genus. The genus *Acer* was split into two classes, namely \AcerH" and \AcerS".

Table A.4 The class labels and their constituent taxa.

Label	**Taxon**	**Train**	**Test**
AcerH	*Acer saccharum*	X	X
AcerS	*Acer macrophyllum*	X	X
AcerS	*Acer negundo*	X	X
AcerS	*Acer rubrum*	X	X
AcerS	*Acer saccharinum*	X	X
Aesculus	*Aesculus californica*	X	X
Aesculus	*Aesculus glabra*	X	X
Aesculus	*Aesculus hippocastanum*	X	
Aesculus	*Aesculus octandra*	X	X
Alnus	*Alnus incana*		X
Alnus	*Alnus rhombifolia*	X	X
Alnus	*Alnus rubra*	X	X
Alnus	*Alnus rugosa*	X	
Alnus	*Alnus serrulata*	X	
Alnus	*Alnus tenuifolia*	X	
Arbutus	*Arbutus menziesii*	X	X
Arbutus	*Arbutus texana*	X	
Arbutus	*Arbutus xalapensis*		X
Betula	*Betula alleghaniensis*	X	
Betula	*Betula lenta*	X	X
Betula	*Betula nigra*	X	X
Betula	*Betula occidentalis*	X	X
Betula	*Betula papyrifera*	X	X
Betula	*Betula populifolia*	X	X
Carpinus	*Carpinus caroliniana*	X	X
Fagus	*Fagus grandifolia*	X	X
Frangula	*Frangula purshiana*		X

Table A.4 (continued)

Label	Taxon	Train	Test
Frangula	*Rhamnus californica*	X	
Frangula	*Rhamnus caroliniana*	X	
Frangula	*Rhamnus frangula*	X	
Frangula	*Rhamnus lanceolata*	X	
Frangula	*Rhamnus purshiana*	X	
Fruitwood	*Rhamnus tomentella*	X	
Fruitwood	*Crataegus aestivalis*	X	X
Fruitwood	*Crataegus assurgens*	X	
Fruitwood	*Crataegus calpodendron*		X
Fruitwood	*Crataegus compacti*	X	
Fruitwood	*Crataegus cordata*	X	
Fruitwood	*Crataegus cuneiformis*	X	
Fruitwood	*Crataegus douglasii*	X	X
Fruitwood	*Crataegus macracantha*	X	
Fruitwood	*Crataegus mollis*	X	X
Fruitwood	*Crataegus nitida*	X	
Fruitwood	*Crataegus rivularis*	X	X
Fruitwood	*Crataegus rotundifolia*	X	
Fruitwood	*Crataegus spathulata*	X	X
Fruitwood	*Crataegus succulenta*	X	
Fruitwood	*Crataegus tomentosa*	X	
Fruitwood	*Malus angustifolia*	X	X
Fruitwood	*Malus baccata*	X	
Fruitwood	*Malus coronaria*	X	X
Fruitwood	*Malus domestica*	X	
Fruitwood	*Malus fusca*		X
Fruitwood	*Malus pumila*	X	X
Fruitwood	*Malus rivularis*	X	
Fruitwood	*Malus sp*	X	
Fruitwood	*Prunus americana*	X	X
Fruitwood	*Prunus angustifolia*	X	X
Fruitwood	*Prunus avium*	X	
Fruitwood	*Prunus caroliniana*	X	X
Fruitwood	*Prunus emarginata*	X	X
Fruitwood	*Prunus myrtifolia*	X	X
Fruitwood	*Prunus nigra*	X	
Fruitwood	*Pyrus ioensis*	X	

Table A.4 (continued)

Label	Taxon	Train	Test
Fruitwood	*Sorbus americana*	X	X
Fruitwood	*Sorbus aucuparia*	X	
Fruitwood	*Sorbus decora*	X	X
Liquidambar	*Liquidambar styraciflua*	X	X
Liriodendron	*Liriodendron tulipifera*	X	X
Magnolia	*Magnolia acuminata*	X	X
Magnolia	*Magnolia fraseri*	X	X
Magnolia	*Magnolia grandiflora*	X	X
Magnolia	*Magnolia macrophylla*	X	X
Magnolia	*Magnolia tripetala*	X	X
Magnolia	*Magnolia virginiana*	X	X
Nyssa	*Nyssa aquatica*	X	X
Nyssa	*Nyssa biflora*	X	
Nyssa	*Nyssa ogeche*	X	X
Nyssa	*Nyssa sylvatica*	X	X
Nyssa	*Nyssa sylvatica-var-biflora*		X
Ostrya	*Ostrya virginiana*	X	X
Oxydendrum	*Oxydendrum arboreum*	X	X
Platanus	*Platanus occidentalis*	X	X
Populus	*Populus angustifolia*	X	X
Populus	*Populus balsamifera*	X	X
Populus	*Populus deltoides*	X	X
Populus	*Populus fremontii*	X	X
Populus	*Populus grandidentata*	X	X
Populus	*Populus heterophylla*	X	X
Populus	*Populus tremuloides*	X	X
Populus	*Populus trichocarpa*	X	X
Prunus	*Prunus serotina*	X	X
Rhamnus	*Rhamnus cathartica*	X	
Rhamnus	*Rhamnus crocea*	X	X
Salix	*Salix laevigata*	X	X
Salix	*Salix lasiandra*	X	X
Salix	*Salix nigra*	X	X
Salix	*Salix nuttellii*	X	
Salix	*Salix scouleriana*	X	X
Tilia	*Tilia americana*	X	X
Tilia	*Tilia americana-var-heterophylla*		X

Table A.4 (continued)

Label	Taxon	Train	Test
Tilia	Tilia caroliniana	X	
Tilia	*Tilia floridana*	X	
Tilia	*Tilia heterophylla*	X	
Tilia	*Tilia pubescens*	X	

A.2.2 Supplementary Material 4: Further Results

Effect of images per specimen on specimen level accuracy:

In the PACw (test) dataset, the number of images provided by any specimen ranged from 1 to 5 depending on the physical dimensions of the wood specimen. A specimen-level prediction was obtained by applying a majority voting rule to the model predictions on the images contributed by the specimen. In the main manuscript we reported the specimen level accuracy using up to $n = 5$ images per specimen – if a specimen contributed, say, 3 images then the majority rule was applied to the 3 image level predictions. Here we provide additional results of an experiment that explored the impact of n images (for the values 1, 2, 3, 4, 5) on specimen-level predictive accuracy of the model.

Specifically:

For n in [1,2,3,4]:

For r in [1,2,3, … ,10]:

$D(n, r) = \emptyset$

For each specimen $S(i)$ in PACw:

$S(i, n, r) =$ Randomly select n images from specimen $S(i)$.

$D(n, r) \leftarrow D(n, r) \cup S(i, n, r)$

Compute $Acc(n, r)$ = Prediction accuracy of field model on $D(n, r)$.

In Figure A.1, for each value of n the mean, minimum and maximum accuracies over the ten repeats (i.e., over the values $Acc(n,1), Acc(n,2), Acc(n,3), \ldots, Acc(n,10)$) are plotted.

For $n = 5$, the dataset remains the same over all the 10 repeats, as that is the maximum number of images per specimen, so all 10 repeats are identical. The performance metrics are also reported in Table A.5. For a particular value of n, the accuracy range is 5%.

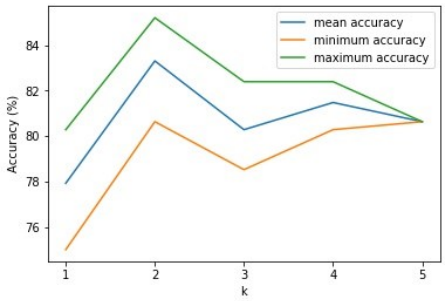

Figure A.1 Accuracy as a function of the number of images used to obtain a specimen-level prediction.

Table A.5 The mean, standard deviation, minimum, and maximum specimen level accuracies as the number of images per specimen varies.

n	Mean	Standard deviation	Minimum	Maximum
1	77.9	1.5	75	80.3
2	83.3	1.1	80.6	85.2
3	80.3	1.2	78.5	82.4
4	81.5	0.71	80.3	82.4
5	80.6	0.0	80.6	80.6

All values are percentages.

Performance of model with a ResNet50 backbone:

A model with a ResNet50 backbone was trained with the same methodology used for the ResNet34 based model described in the manuscript. The confusion matrices for the cross-validation analysis and for the prediction of the field model on the PACw specimens are

presented in figures A.2 and A.3 respectively. The prediction accuracies are presented in Table A.6. The accuracy of the ResNet50 *field model* is ~5% points lower than the ResNet34 model.

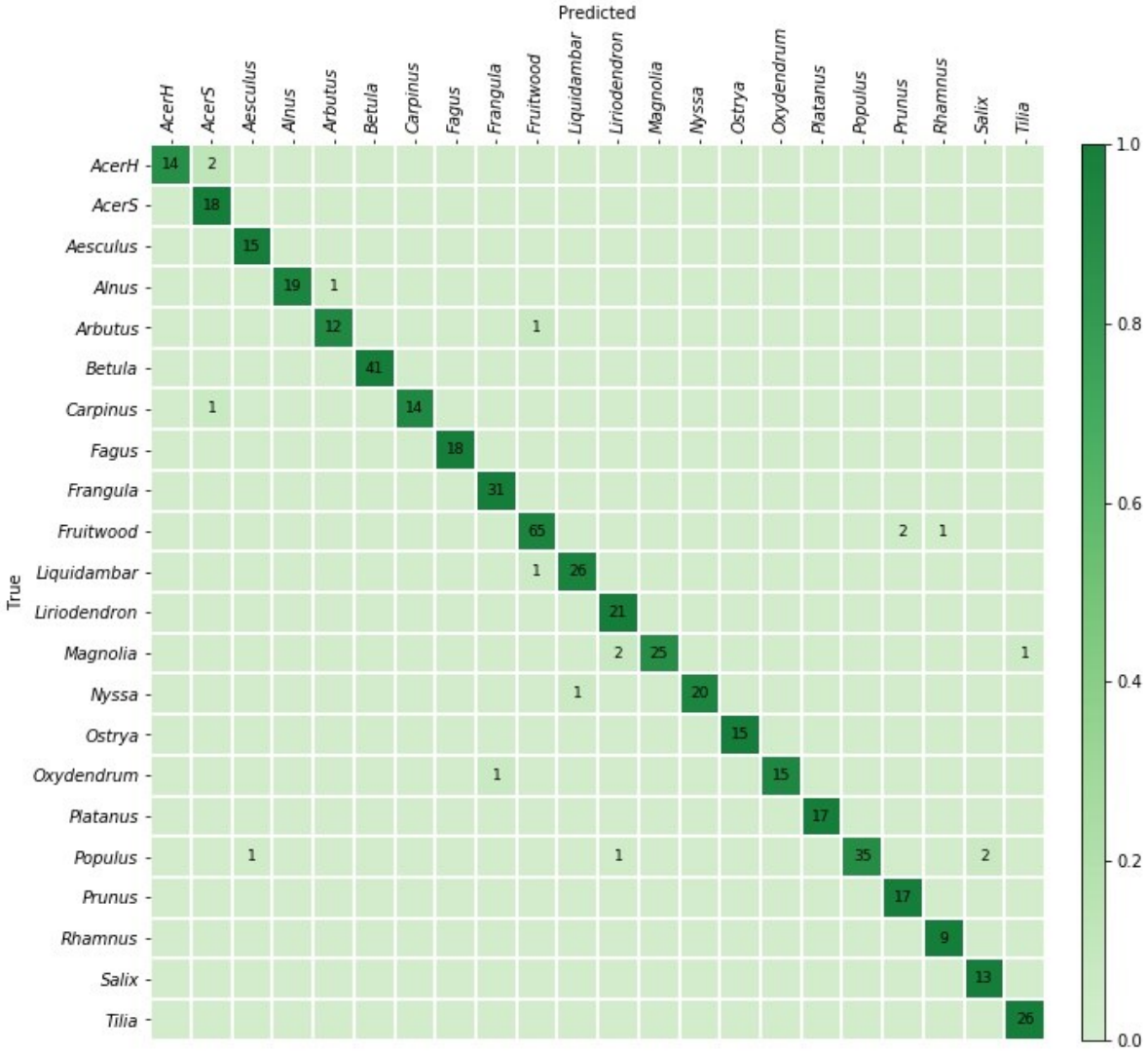

Figure A.2 Cross validation confusion matrix for the ResNet50-backbone based model.

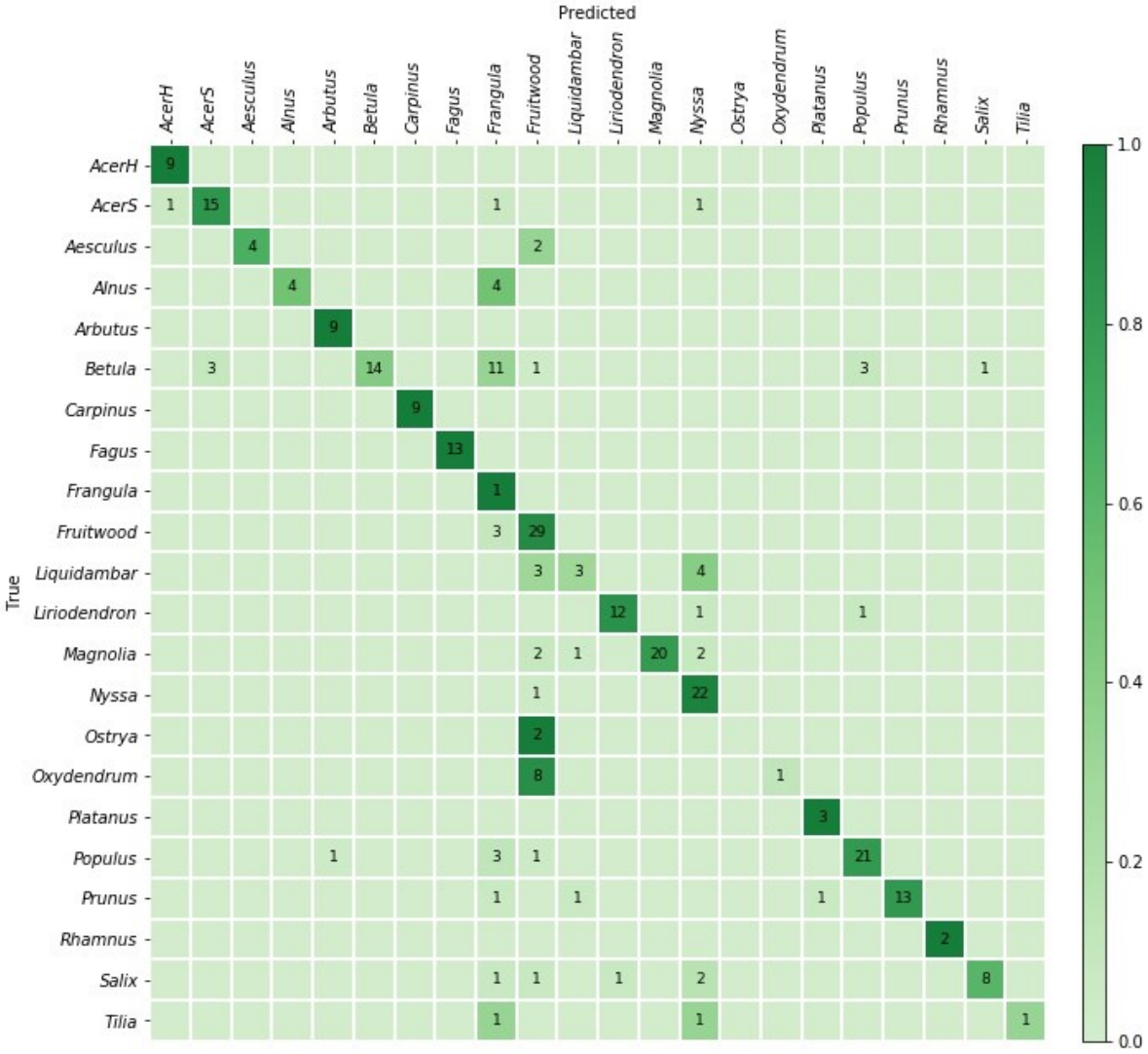

Figure A.3 Confusion matrix of predictions from the ResNet50 based model on PACw data.

Table A.6 Specimen level prediction accuracies of ResNet50 based model.

Training and evaluation details	Top-k	Accuracy (%)
Five-fold cross-validation	k=1	96.43
Trained using four folds, tested on PACw	k=1	76.13*
	k=2	86.83*
Trained using all five folds, tested on PACw	k=1	75
	k=2	85.56

Note: * Mean accuracies over the 5 cross-validation models.

Performance of ResNet34 and ResNet50 models after stage 1 of transfer learning:

A 2-stage transfer learning methodology was adopted for training the models presented in the manuscript. The results presented in Tables 3.3 (in main manuscript) and A.6 (in this supplement) are for the ResNet34 and ResNet50 models at the end of two stages of the training.

In the first stage of our transfer learning training methodology only the weights of the custom head are trained (i.e., the backbone layers initialized with ImageNet pre-trained weights are frozen). Here we present the performance of the ResNet34 and ResNet50 models at the end of the first stage of training in the same format as in Tables A.7 and A.8.

Table A.7 Specimen level prediction accuracies of ResNet34 based model at the end of first training stage

Training and evaluation details	Top-k	Accuracy (%)
Five-fold cross-validation	k=1	94.25
Trained using four folds, tested on PACw	k=1	71.62*
	k=2	84.01*
Trained using all five folds, tested on PACw	k=1	79.93
	k=2	89.08

Note: * Mean accuracies over the 5 cross-validation models.

Table A.8 Specimen level prediction accuracies of ResNet50 based model at the end of first training stage

Training and evaluation details	Top-k	Accuracy (%)
Five-fold cross-validation	k=1	95.63
Trained using four folds, tested on PACw	k=1	76.83*
	k=2	86.97*
Trained using all five folds, tested on PACw	k=1	73.24
	k=2	82.39

Note: * Mean accuracies over the 5 cross-validation models.

Fine tuning (i.e., the second stage of training) improves the performance of the model by up to 3% points. It is of interest that as the amount of data is increased by 20%, the performance of the ResNet50 drops – a behavior that may be attributed to the higher capacity of the ResNet50 (more weights) in the context of the relatively limited number of specimens in our dataset compared to other machine learning datasets in biology (e.g., Horn et al. 2018). Despite our dataset tapping into one of the richest collections of North American diffuse porous specimens, it may be the case that larger high-quality datasets may be needed to leverage the increased capacity of the ResNet50 for improved predictive performance. Another option to explore is to unfreeze only the top layers of the ResNet50 backbone in stage 2 and try to estimate a sweet spot between the available dataset and the (trained) capacity of the architecture. In this work we have explored both the ResNet34 and ResNet50 architectures under an “equal epoch budget” regime and we leave the exploration of the capacity-dataset size tradeoff, fine-grade hyperparameter optimization, and field deployment of the model presented here to future work.

A.3 Supplementary Material for Chapter 4

A.3.1 Supplementary Material 5: Class Composition Details

The 68 unique taxa were assigned to 17 anatomically relevant classes. The table below lists the taxa used and their class membership, along with their inclusion in the training/testing data set. The manuscript provides details for the label space design.

Table A.9 The class labels and their constituent taxa.

Label	**Taxon**	**Train**	**Test**
Asimina	*Asimina triloba*	X	X
Carya	*Carya aquatica*	X	X
Carya	*Carya cordiformis*	X	X
Carya	*Carya glabra*	X	X
Carya	*Carya illinoensis*	X	
Carya	*Carya illinoinensis*		X
Carya	*Carya laciniosa*	X	X
Carya	*Carya myristiciformis*	X	
Carya	*Carya ovata*	X	X
Carya	*Carya texana*	X	
Carya	*Carya tomentosa*	X	X
Castanea	*Castanea dentata*	X	X
Castanea	*Castanea pumila*	X	
Catalpa	*Catalpa bignonioides*	X	
Catalpa	*Catalpa speciosa*	X	X
Celtis	*Celtis laevigata*	X	
Celtis	*Celtis occidentalis*	X	X
Celtis	*Celtis reticulata*	X	
Celtis	*Celtis sp*		X
Cladrastis	*Cladrastis kentukea*	X	
Cladrastis	*Cladrastis lutea*	X	X
Fraxinus	*Fraxinus americana*	X	
Fraxinus	*Fraxinus nigra*	X	X
Fraxinus	*Fraxinus oregona*	X	
Fraxinus	*Fraxinus pennsylvanica*	X	X
Fraxinus	*Fraxinus quadrangulata*	X	X
Gleditsia	*Gleditsia aquatica*	X	
Gleditsia	*Gleditsia triacanthos*	X	X
Gymnocladus	*Gymnocladus dioica*	X	X
Maclura	*Maclura pomifera*	X	X

Table A.9 (continued)

Label	Taxon	Train	Test
Maclura	*Maclura sp*		X
Morus	*Morus alba*	X	X
Morus	*Morus rubra*	X	X
QuercusR	*Quercus arkansana*	X	
QuercusR	*Quercus coccinea*	X	X
QuercusR	*Quercus ellipsoidalis*	X	
QuercusR	*Quercus falcata*	X	X
QuercusR	*Quercus georgiana*	X	
QuercusR	*Quercus ilicifolia*	X	
QuercusR	*Quercus incana*	X	
QuercusR	*Quercus laevis*	X	
QuercusR	*Quercus laurifolia*	X	X
QuercusR	*Quercus marilandica*	X	X
QuercusR	*Quercus myrtifolia*	X	
QuercusR	*Quercus nigra*	X	X
QuercusR	*Quercus palustris*	X	
QuercusR	*Quercus phellos*	X	X
QuercusR	*Quercus rubra*	X	
QuercusR	*Quercus shumardii*	X	X
QuercusR	*Quercus texana*	X	
QuercusR	*Quercus velutina*	X	X
QuercusW	*Quercus alba*	X	X
QuercusW	*Quercus bicolor*	X	X
QuercusW	*Quercus lyrata*	X	X
QuercusW	*Quercus macrocarpa*	X	X
QuercusW	*Quercus michauxii*	X	
QuercusW	*Quercus montana*	X	X
QuercusW	*Quercus stellata*	X	
Robinia	*Robinia neo-mexicana*	X	
Robinia	*Robinia pseudoacacia*	X	X
Sassafras	*Sassafras albidum*	X	X
Sassafras	*Sassafras sp*		X
UlmusH	*Ulmus alata*	X	
UlmusH	*Ulmus crassifolia*	X	
UlmusH	*Ulmus serotina*	X	
UlmusH	*Ulmus thomasii*	X	X
UlmusS	*Ulmus americana*	X	X

Table A.9 (continued)

Label	Taxon	Train	Test
UlmusS	*Ulmus rubra*	X	

A.3.2 Supplementary Material 6: Further Results

Performance of model with a ResNet50 backbone:

A model with a ResNet50 backbone was trained with the same methodology used for the ResNet34 based model described in the manuscript. The confusion matrices for the cross-validation analysis and for the prediction of the field model on the PACw specimens are presented in figures A.4 and A.5 respectively. The prediction accuracies are presented in Table A.10.

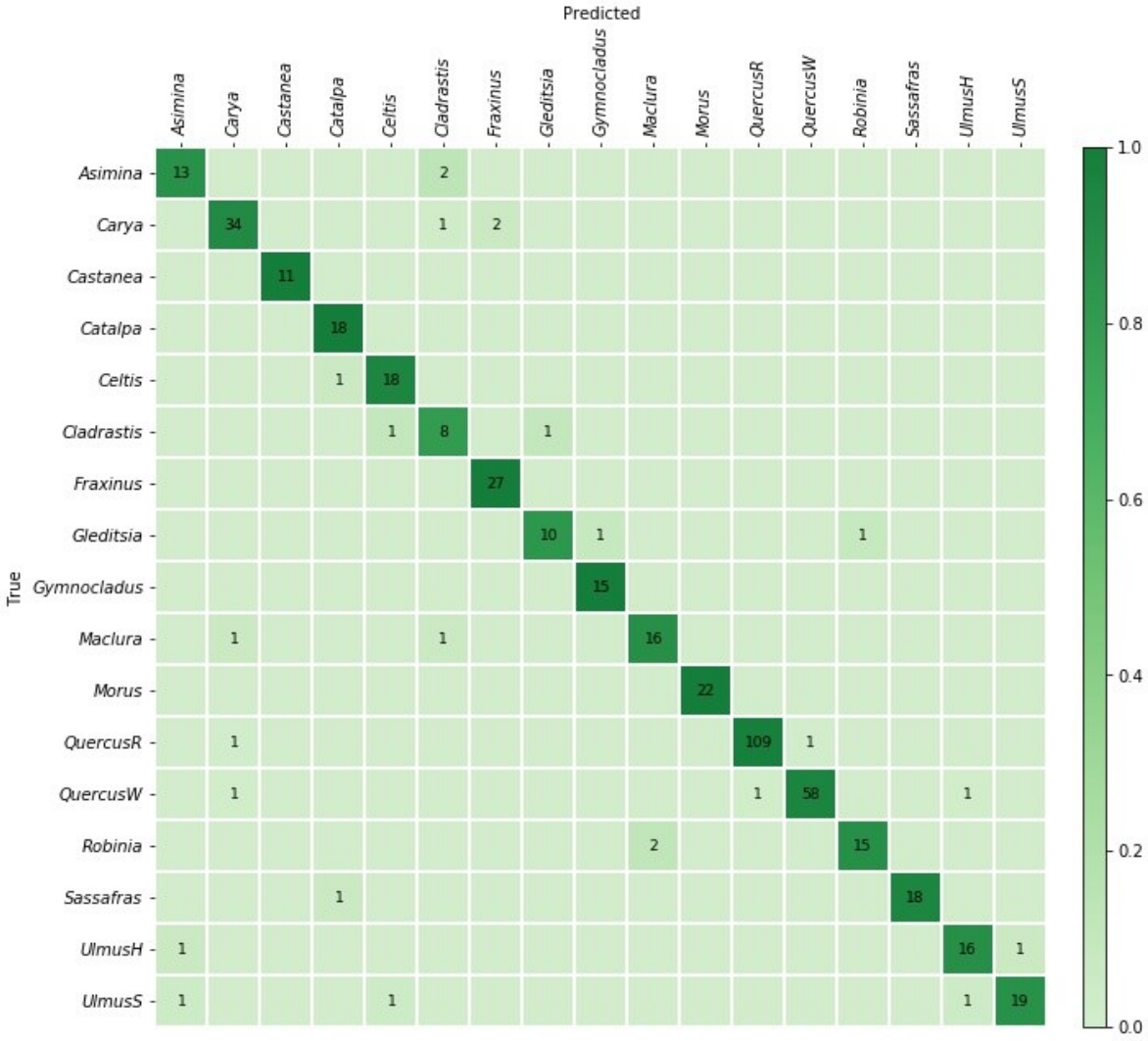

Figure A.4 Cross validation confusion matrix for the ResNet50-backbone based model.

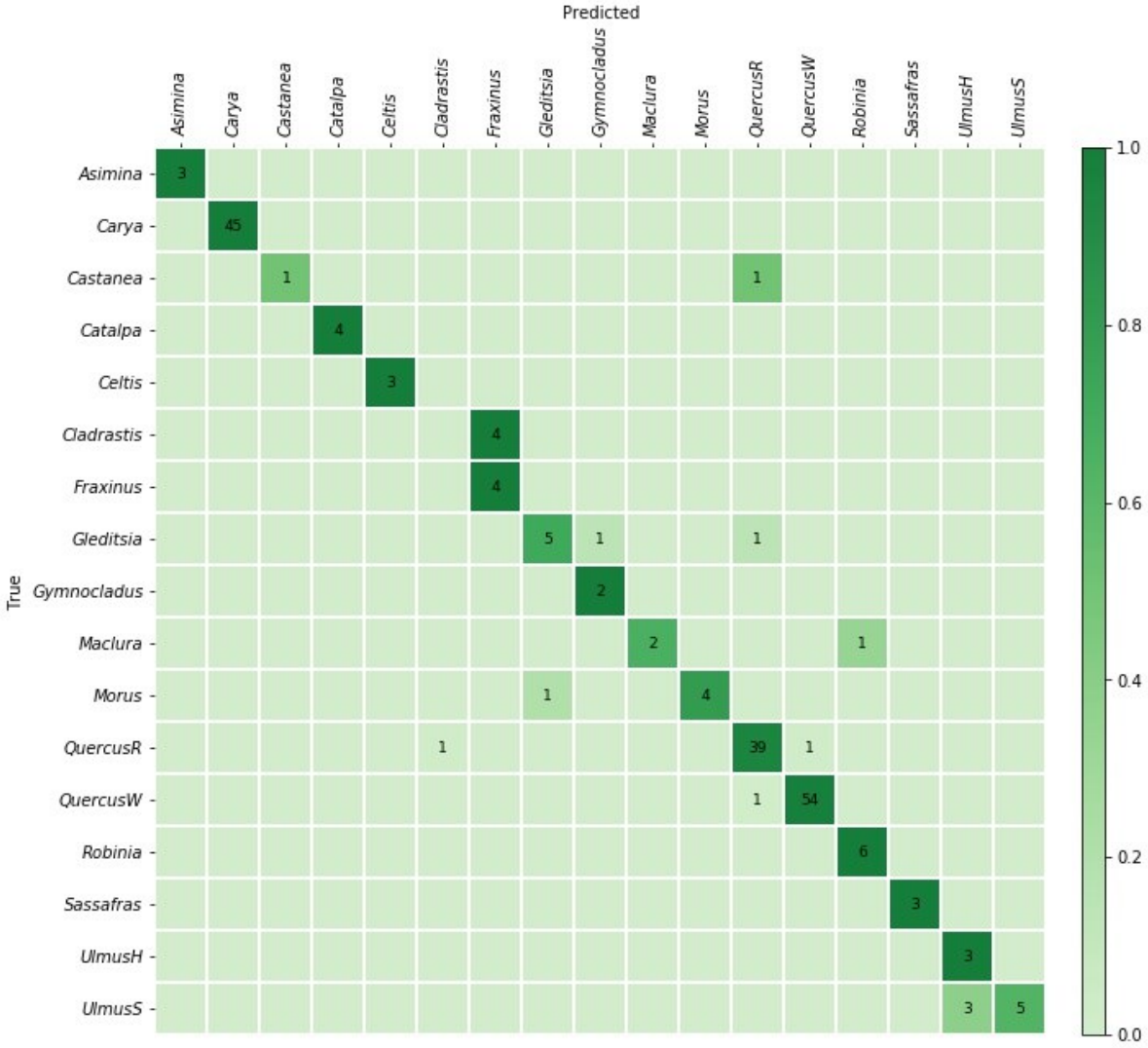

Figure A.5 Confusion matrix of predictions from the ResNet50 based model on PACw data.

Table A.10 Specimen level prediction accuracies of ResNet50 based model. * Mean accuracies over the 5 cross-validation models

Training and evaluation details	**Top-k**	**Accuracy (%)**
Five-fold cross-validation	k=1	94.47
Trained using four folds, tested on PACw	k=1	88.38*
	k=2	95.35*
Trained using all five folds, tested on PACw	k=1	92.42
	k=2	97.47