Delving into the Porosity Domain Continuum in Hardwood Growth Rings: What Can We Learn from Computer Vision Wood Identification Models?

Alex C. Wiedenhoeft [iD],[a,b,c,d,e] Prabu Ravindran [iD],[a,b] Adriana Costa [iD],[c] Rubin Shmulsky,[c] and Frank C. Owens [iD] [c,*]

Hardwood porosity domains (diffuse-, semi-ring-, and ring-porosity) exist along a spectrum with some taxa embodying only one porosity domain and others spanning more than one. A cascading model scheme involving a root-level porosity classifier and second-level taxonomical classifiers might be useful for mitigating reductions in the predictive accuracy of North American computer vision wood identification (CVWID) models when the number of classes increases. Thus far, the porosity classifier has been trained on images covering the breadth of the porosity spectrum. By reducing ambiguity near the boundaries of porosity domains, training the root classifier only on taxa that are quintessentially diffuse-, semi-ring, and ring-porous might produce equivalent or better results. In this study, a two-class (diffuse- and ring-porous) model and a three-class (diffuse-, semi-ring-, and ring-porous) model were trained on specimens only from taxa with quintessentially idealized porosity and tested on specimens with and without idealized porosity. Results showed perfect predictive accuracy for both models when tested on in-model taxa but showed lower accuracy on datasets with non-ideal porosity with all misclassifications being anatomically sensible. In addition, the results showed remarkable similarities between CVWID models and humans in how they "apply" the concept of discrete porosity domains to a real-world continuum.

*Contact information: a: Center for Wood Anatomy Research, USDA Forest Service, Forest Products Laboratory, Madison, WI, USA; b: Department of Botany, University of Wisconsin, Madison, WI, USA; c: Department of Sustainable Bioproducts, Mississippi State University, Starkville, MS, USA; d: Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA; e: Departamento de Ciências Biológicas (Botânica), Universidade Estadual Paulista – Botucatu, São Paulo, Brasil; * Corresponding author: fco7@msstate.edu*

INTRODUCTION

The pervasive problem of fraud and misrepresentation in wood products trade, including the trafficking of illegally harvested materials, has encouraged the development of technology-driven wood identification methods utilizing advancements in computer vision and machine learning; DNA extraction and analysis; chemical-spectral analysis; and other areas to increase wood identification capacity and reduce reliance on human-based conventional wood identification techniques (Johnson and Laestadius 2011; Dormontt *et al*. 2015; Koch *et al*. 2015; Lowe *et al*. 2016; UNODC 2016; Beeckman *et al*. 2020). Computer-vision wood identification (CVWID) systems such as the XyloTron,

XyloPhone, AIKO, MyWood-ID, and others have shown promise as rapid, low-cost, field-deployable tools for democratizing wood identification capacity among the front-line defenders of fair-trade including customs officials and procurement personnel (Hermanson and Wiedenhoeft 2011; Tang *et al.* 2018; Ravindran *et al*. 2018, 2020, 2022a,b; Damayanti *et al.* 2019; Wiedenhoeft 2020).
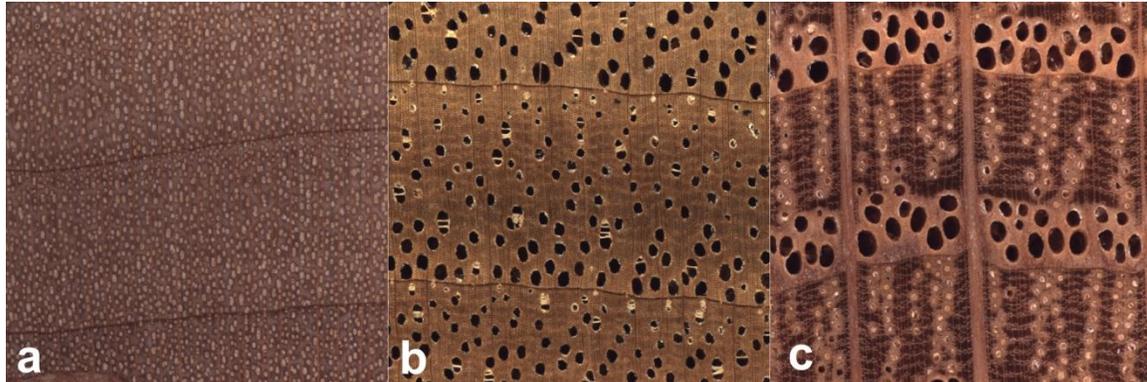

Fig. 1. Quintessential images, from left to right, of the porosity classes diffuse-porous (a: *Acer saccharinum*), semi-ring-porous (b: *Juglans cinerea*), and ring-porous (c: *Quercus nigra*). In diffuse-porous woods (a), the earlywood pores are not noticeably larger than the latewood pores (Panshin and de Zeeuw 1980). In semi-ring-porous woods (b), there is a gradual but several-fold decrease in pore size from the earlywood through the latewood (Panshin and de Zeeuw 1980). In ring-porous woods (c), larger earlywood pores abruptly transition in size to several-fold smaller latewood pores (Panshin and de Zeeuw 1980).

The XyloTron and XyloPhone CVWID systems (USDA Forest Products Laboratory, Madison, Wisconsin, USA) identify woods by analyzing macroscopic digital images of the transverse surface of wood specimens captured under magnification (Ravindran *et al*. 2018, 2020, 2022a,b; Wiedenhoeft 2020). Typically, deep learning models are trained to categorize an image into one of a pre-defined set of woods (Hwang and Sugiyama 2021; Silva *et al.* 2022). Much research to date has focused on image-level taxonomic classification, whereby a deep learning model uses features and patterns it finds in the images as a basis for making a discrimination (Hwang and Sugiyama 2021; Silva *et al.* 2022). When models are expanded by adding new classes (*i.e.*, woods from new taxa for taxonomic models), they are retrained after adding images of the new class(es) to the training dataset. The model then learns to extract features – which may be completely distinct from the features learned in the prior model – from all included images such that a penalty related to misclassification is minimized. It should be noted that these features and patterns may not necessarily correspond to the anatomical features (such as vessels, rays, growth rings, parenchyma patterns, *etc*.) that a wood anatomist might use to identify wood with a hand lens.

To achieve maximum utility, a CVWID model should be able to differentiate among many woods as opposed to merely a few – for example, hundreds as opposed to dozens. Previous research has suggested that increasing the number of classes in a CVWID model has the potential to decrease predictive accuracy (Owens *et al.* 2024). To address this concern, Owens *et al.* (2024) proposed a bi-level cascading CVWID model scheme to identify 42 classes of North American hardwoods, whereby in the first-level (root model), images were classified as one of three porosity domain classes (Fig. 1) and then a second-level, porosity-dependent taxonomic model (leaf model) classified the images into the

appropriate taxonomic class (Fig. 2). This scheme outperformed a single-level 42-class taxonomic model in terms of accuracy and number of cross-domain misidentifications (Type 3 misclassifications in this case, per Ravindran *et al.* 2021, Table 4). Owens *et al.* 2024 was the first work in CVWID to combine a wood-feature root model (porosity domain) with taxonomic leaf models (Fig. 2). Their results suggested that cascading models of this type can be useful for mitigating a reduction in accuracy when the number of classes increases, specifically by reducing the dimensionality of the feature space handled by the taxonomic leaf models.
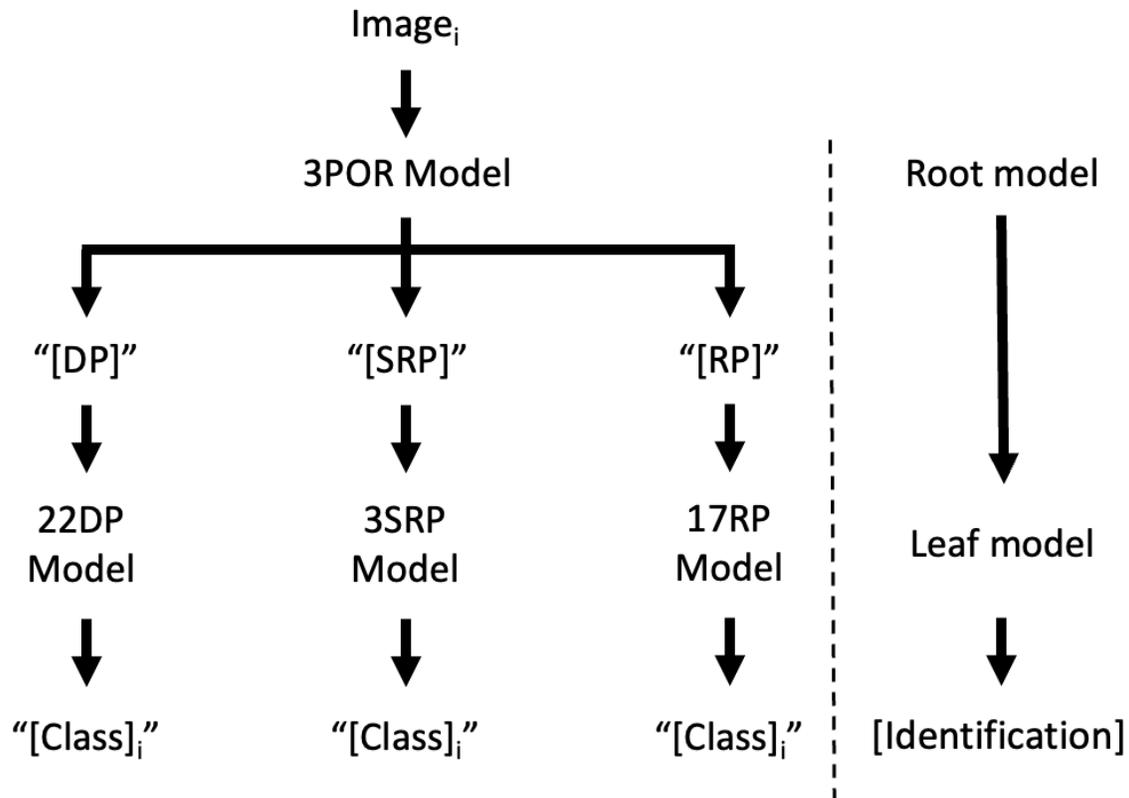

Fig. 2. The cascading model scheme adapted from Owens *et al.* 2024, whereby an image of the transverse surface of a wood specimen is first classified into a porosity class – diffuse-porous (DP), semi-ring-porous (SRP), or ring-porous (RP) – by the 3POR (root) model and then classified by the second-level taxonomic (leaf-level) model that corresponds to that porosity class: 22DP, 3SRP, or 17RP. The predictions of the cascaded models yield a class assignment for an image. The accuracy of this cascading model scheme was 85.7% overall and 80.3%, 100%, and 91.4% for diffuse-porous, semi-ring-porous, and ring-porous, woods respectively.

The novelty of training a root classifier using wood anatomical domain space rather than taxonomic domain space (Owens *et al.* 2024) tacitly opened the question of whether a root-level classifier should be trained with biologically variable representations of each class (as done in that study), or with a more narrowly circumscribed dataset of quintessential representations of each class. That is, does a root-level classifier show better generalizability when trained with "messy" data, or if idealized (*i.e.*, textbook-like) images representing each class result in better generalizability to new specimens and taxa. The latter is more akin to how we train humans to identify wood (Panshin and de Zeeuw 1980; Hoadley 1990), whether using microscopic (Wheeler *et al.* 1989; Richter and Dallwitz 2000; Richter *et al.* 2004) or macroscopic (Miller *et al.* 2002; Wiedenhoeft 2011;

Ruffinatto *et al.* 2015; Florsheim *et al.* 2020; Arévalo *et al.* 2021; Arévalo and Wiedenhoeft 2022; Ministerio del Ambiente [MINAM] 2022) characters. Wood identification textbooks or feature lists commonly define each porosity domain by relative pore size and distribution – diffuse-porosity characterized by little or no conspicuous change in pore size throughout the ring, semi-ring-porosity exhibiting a gradual several-fold change in pore size, and ring-porous exhibiting an abrupt several-fold change in pore size from the earlywood to the latewood – followed by exemplar images, such as those in Fig. 1.

When training human beings in this way, students quickly learn that the porosity of some taxa does not always fall discretely into these categories. For example, the genus *Carya* can intergrade between ring- and semi-ring-porous (Fig. 3, e, f), and the genus *Populus* can intergrade between semi-ring-like porosity and diffuse-porous (Fig. 3, b, c). The distinction in *Populus* is one more of vessel frequency in the earliest earlywood compared to the latest latewood than one of a several-fold change in diameter over the ring; it is not semi-ring-porous *sensu stricto*. There can also be individual variations in porosity typology between and within trees, even within taxa that generally exhibit idealized manifestations of a porosity domain. Wood anatomists typically make porosity classifications based on idealized definitions of those discrete categories even when the porosity under observation is less than quintessential (Costa, Owens, Wiedenhoeft, personal observation), and the tacitly known spectrum of porosity is itself a continuum, not three discrete character states. This problem of discretizing as distinct characters variability that is inherently continuous and quantitative is not new to research in biology (Thiele 1993; Goloboff *et al.* 2006; Parins-Fukuchi 2018).

Given that porosity is in actuality a continuum from diffuse- to semi-ring- to ring-porous, and further given that all woods must be forced into one of three character states along this continuum, it may well be critical to test explicitly whether deeper-level classifiers are positively or negatively impacted by incorporating real-world (*vs.* idealized) variability, given the comparative paucity of reference specimens available in wood identification research, broadly. We are far from a world in which wood anatomy datasets are defined by *i* images of *c* classes, where *i* and *c* are large – that is to say, ImageNet for wood is not likely on the horizon, especially given the quality of some CVWID datasets (Ravindran and Wiedenhoeft 2022). Hence, the ability to train accurate wood anatomical domain classifiers (*e.g.*, the porosity continuum) with modest datasets can be beneficial for constructing highly skilled, cascaded CVWID models for large label spaces.

This work explored the definition of wood porosity in the context of CVWID, as follows:

1. The root-level 3-class porosity classification model from Owens *et al.* (2024) was retrained on a smaller, constrained dataset comprised *only* of taxa that typically exhibit idealized ("textbook") manifestations of diffuse-, semi-ring-, and ring-porosity. The retrained model was evaluated on independent specimens from taxa that were included in the constrained training dataset (hereafter, "in-model taxa") and also on a dataset comprised of images from taxa that were excluded from the training dataset due to the strict definition of porosity (hereafter, "out-of-model taxa"). The bi-level porosity-conditioned model scheme of Owens *et al.* (2024) was revisited by replacing the root-level model with a model trained on a dataset comprising only idealized manifestations of diffuse- semi-ring-, and ring-porosity.

2. A porosity classification model with a 2-class label space (diffuse-porous and ring-porous) was trained on a dataset comprising taxa that typically exhibit idealized diffuse- and ring-porosity and evaluated on a dataset that included the out-of-model

diffuse-, semi-ring, and ring-porous taxa. This analysis provides insight into how a CVWID model "generalizes" the concept of porosity and suggests future avenues for designing label spaces that incorporate discretized feature classes for CVWID models.

EXPERIMENTAL

Materials
Specimens and images

The datasets for this study were compiled from images of specimens used in Owens *et al.* (2024).

Table 1. Porosity Domain Membership of Woods Used in this Study

Diffuse-porous	Semi-ring-porous	Ring-porous
*AcerH (hard maples)	*Diospyros	*Asimina
*AcerS (soft maples)	*JuglansC (*J. cinerea*)	Carya
*Aesculus	*JuglansN (*J. nigra*)	*Castanea
*Alnus		Catalpa
Arbutus		*Celtis
*Betula		Cladrastis
*Carpinus		*Fraxinus
Fagus		*Gleditsia
Frangula		*Gymnocladus
*Fruitwood (comprising:)		*Maclura
• *Crataegus* spp.		*Morus
• *Malus* spp.		*QuercusR (red oaks)
• *Prunus* spp. (except *P. serotina*)		*QuercusW (white oaks)
• *Pyrus* spp.		*Robinia
• *Sorbus* spp.		*Sassafras
*Liquidambar		UlmusH (hard elms)
*Liriodendron		*UlmusS (soft elms)
*Magnolia		
*Nyssa		
*Ostrya		
*Oxydendrum		
Platanus		
Populus		
Prunus (*P. serotina* only)		
*Rhamnus		
Salix		
*Tilia		
All specimens imaged for this study were also used in Owens *et al.* 2024. Woods marked with an asterisk tend to exhibit quintessential diffuse-, semi-ring-, or ring-porosity. Woods without an asterisk (out-of-model taxa) commonly exhibit porosity that intergrade with a different porosity domain or otherwise differ from a quintessential representation of that porosity.		

Training datasets comprised images of specimens from the MADw and SJRw collections housed at the USDA Forest Products Laboratory in Madison, Wisconsin, and from the Tw collection at the Royal Museum of Africa. Images for the independent testing datasets were of specimens sourced exclusively from Mississippi State University's David A. Kribs (PACw) and teaching collections.
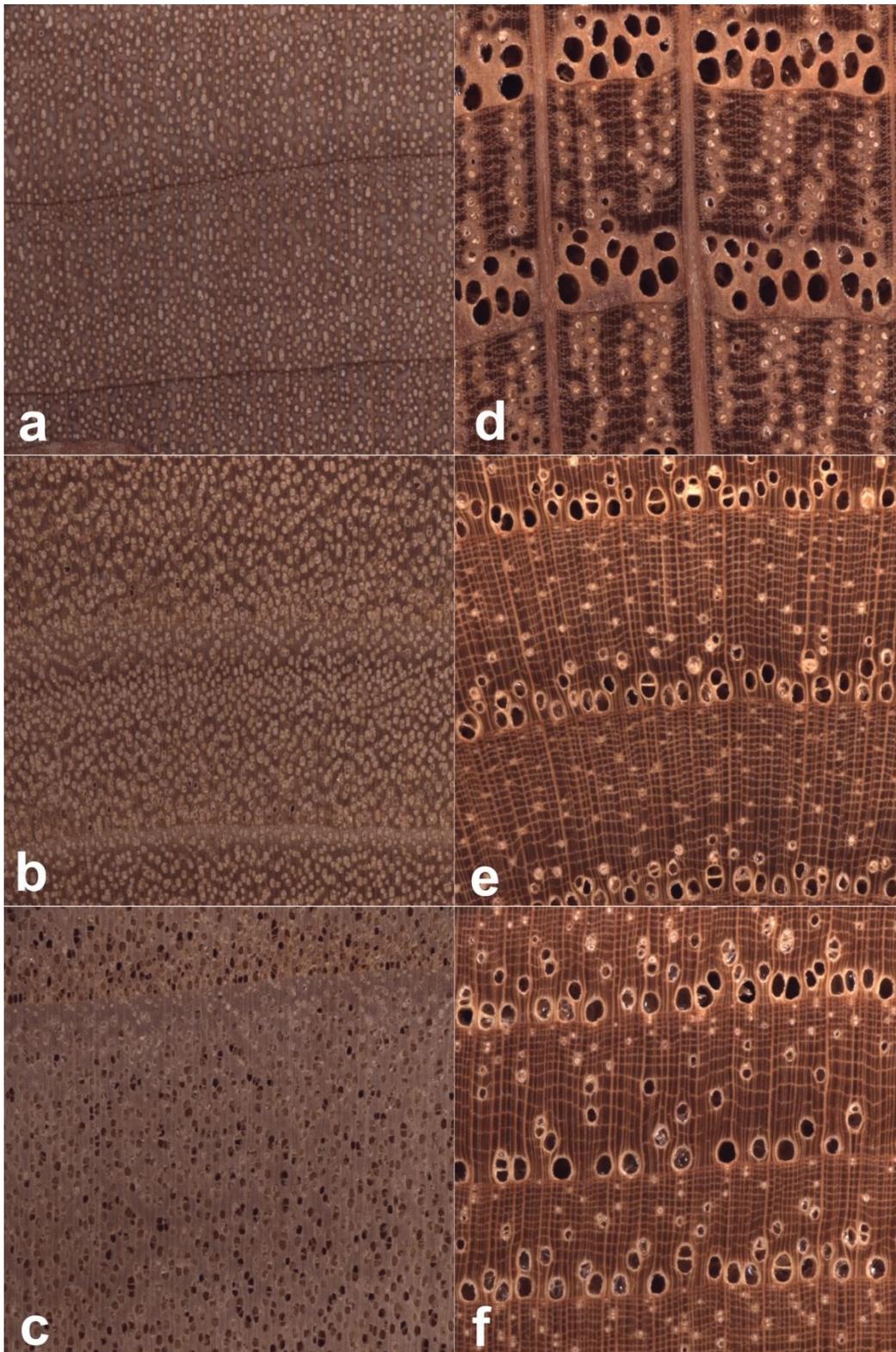
bioresources.cnr.ncsu.edu

Fig. 3. Examples of idealized v. non-idealized diffuse- and ring-porosity. *Acer saccharinum* (a, idealized diffuse-porous) and two specimens of *Populus deltoides* (non-idealized, b is more diffuse, and c is more semi-ring-porous). *Quercus nigra* (d, idealized ring-porous) and two specimens of *Carya tomentosa* (non-idealized, e is more ring-porous, and f is more semi-ring-porous).

Specimens used in the training and testing datasets were mutually exclusive. All specimens were at moisture contents consistent with ambient indoor conditions presumed to be ~5% to 9%, depending on geographical location and time of year. Commonly, specimens were heartwood, though sapwood images were also included. Specimens with a sapwood-heartwood transition were not used.

Specimens were prepared for imaging by polishing the transverse surface on a benchtop disc sander using progressively finer sandpaper grits of 80, 180, 240, 400, 600, 800 and 1500. Sanded surfaces were imaged using the XyloTron system. Images were non-overlapping, each capturing approximately 6.35 mm × 6.35 mm of tissue – all wood images in this manuscript are 6.35mm on a side. Image resolution was 2048 × 2048 pixels with a spatial resolution of ~3.1 microns per pixel.

Table 1 shows the taxa imaged in this study broken down by porosity class as determined by consensus of Wiedenhoeft, Costa, and Owens (personal observations). Those marked with an asterisk were regarded as idealized ring, semi-ring or diffuse-porous woods characterized by consistent textbook manifestations of those patterns (Fig. 3a,d). Woods without an asterisk (out-of-model taxa) were deemed to exhibit porosity that tended to intergrade or be confused with others (Fig. 3 b,c,e,f, and see comments in Table 1). As in previous publications by the same authors, italicized font is used herein to reference botanical entities, while all CVWID classes are written without italics (*e.g.*, the genus *Carya vs.* the CVWID class Carya).

Label spaces

Descriptions of the two label spaces used in this study appear in Table 2. 2POR-IDL is a two-class label space trained on images of only the taxa from Owens *et al.* (2024) that met the strictest definition of diffuse- and ring-porous (those marked with an asterisk in the left and right columns of Table 1, respectively). 3POR-IDL is a three-class label space including images from the taxa in 2POR-IDL plus images of the asterisked semi-ring-porous taxa in the middle column of Table 1 that met the strictest definition of semi-ring-porous. These label spaces, and the corresponding trained models, hereafter will be referred to as the "idealized label spaces" and the "idealized models" to contrast them with the 3POR model (Fig. 2) in Owens *et al.* (2024).

Table 2. Descriptions of the Label Spaces used in this Study

Label Space	Description
2POR-IDL	A 2-class label space of *idealized diffuse-* and *idealized ring-porous* woods trained with only a subset of the woods from Owens *et al.* 2024 that met strict definitions of diffuse- and ring-porous, respectively (only the diffuse- and ring-porous woods marked with an asterisk in Table 1).
3POR-IDL	A 3-class label space of *idealized diffuse-*, *idealized semi-ring-*, and *idealized ring-porous* woods trained with the diffuse- and ring-porous woods from 2POR-IDL plus a third class of idealized semi-ring-porous woods (the second column in Table 1).

Training and testing datasets

Training image, specimen, taxa, and label counts for 2POR-IDL and 3POR-IDL appear in Table 3. The training datasets contained images only from taxa deemed to exhibit idealized diffuse, ring and semi-ring porosity (in-model taxa). In order to avoid the

potential confounding effects of class imbalance, the authors chose a subset of images so that each of the three porosity classes had roughly the same number of images.

Testing image, specimen, taxa, and label counts for 2POR-IDL and 3POR-IDL appear in Table 4. Unlike the training datasets, the testing datasets contained not only images from the in-model idealized diffuse, semi-ring and ring-porous taxa but also images from taxa the model did not encounter during training (out-of-model taxa). A complete list of taxa included in the testing and training datasets can be found in Tables S1 and S2 in the Appendix.

Table 3. Training Dataset Details by Label Space

Detail	2POR-IDL (counts)	3POR-IDL (counts)
Images	5816	2716
Specimens	692	746
Taxa	114	117
Labels	2	3
These wood specimens were sourced from the USDA Forest Products Laboratory's MADw and SJRw collections and the Royal Museum of Africa's Tw collection. Images of the specimens from these xylaria were used **only** for training and **not** for testing. Only woods classified in Table 1 as idealized were used in training. The 3POR-IDL model was trained on idealized woods from all three columns in Table 1 while the 2POR-IDL model was trained only on idealized woods from the left and right columns.		

Table 4. Test Dataset Details by Label Space

Detail	2POR-IDL (in-model counts)	2POR-IDL (out-of-model counts)	3POR-IDL (counts)	3POR-IDL (out-of-model counts)
Images	1453	823	1584	692
Specimens	329	181	357	153
Taxa	76	36	79	33
Labels	2	2	3	3
These wood specimens were sourced from Mississippi State University's PACw and teaching collections. Images of the specimens from these xylaria were used only for testing and not for training. Both the 2POR-IDL and 3POR-IDL models were tested on images of both idealized and non-idealized woods from all three columns in Table 1.				

Methods

Machine learning models

For each label space, a ResNet34 convolutional neural network (CNN) with a pretrained backbone and a custom head classifier was trained utilizing two-stage transfer learning, whereby the custom head was trained first followed by finer adjustments to both the custom head and the backbone. Prior research (Owens *et al.* 2024) showed that this network depth was sufficient for accurate predictions. Data augmentation was employed during model training including vertical/horizontal flips, cutouts, and slight rotations. In all stages of training, random patches of 2048×768 pixels were extracted from each image and downsampled to 512×192 before being fed to the model in batches of 16. An Adam optimizer with cosine annealing of the momentum and learning rate was used to adjust model weights. Scientific Python tools and PyTorch were employed to train, define, and evaluate the models. Additional details about model architecture and training can be found in Ravindran *et al.* (2019) and Arévalo *et al.* (2021).

Model predictions were made at the specimen level and determined in the same way for both in-model and out-of-model taxa. Multiple images (up to 5) of each specimen were fed into the model. After the model predicted the porosity domain of each image, the majority prediction among the images from a specimen became the prediction for the specimen. A specimen was deemed correctly identified if the top specimen-level prediction matched the porosity class to which the specimen belongs in Table 1 whether or not images from that taxon has been encountered during model training.

RESULTS AND DISCUSSION

2POR-IDL Predictions on In-model and Out-of-model Taxa

When the 2POR-IDL model was tested on the in-model (idealized) diffuse- and ring-porous taxa, all 329 specimens were correctly predicted. When the model was tested on 153 specimens of out-of-model diffuse- and ring-porous taxa, only one specimen of *Cladrastis lutea*, classified in Table 1 as ring-porous, was misclassified as diffuse-porous (Table 5). Figure 4 compares an image from a correctly predicted specimen of *Cladrastis lutea* (Fig. 4a) with an image of the incorrectly predicted specimen (Fig. 4b). The latter image exhibits a diffuse-porous pattern atypical of *Cladrastis*; thus, this prediction is a Type 2 misclassification per Ravindran *et al.* (2021). In this regard, this is a misclassification only because humans (especially including the authors) have asserted that the class Cladrastis is ring-porous, but as a biological entity, the genus *Cladrastis* does not always exhibit idealized ring-porosity, and sometimes exhibits near diffuse-porosity, at least in some growth rings and/or individuals in this dataset. Given this result, it is plausible that this taxon should have been excluded, or at least that atypical images had been culled. Such biological variability is known and accounted for by expert wood anatomists, but by both non-expert humans and CWVID models trained on idealized images, one would expect such misclassifications (see below where this is addressed more specifically).
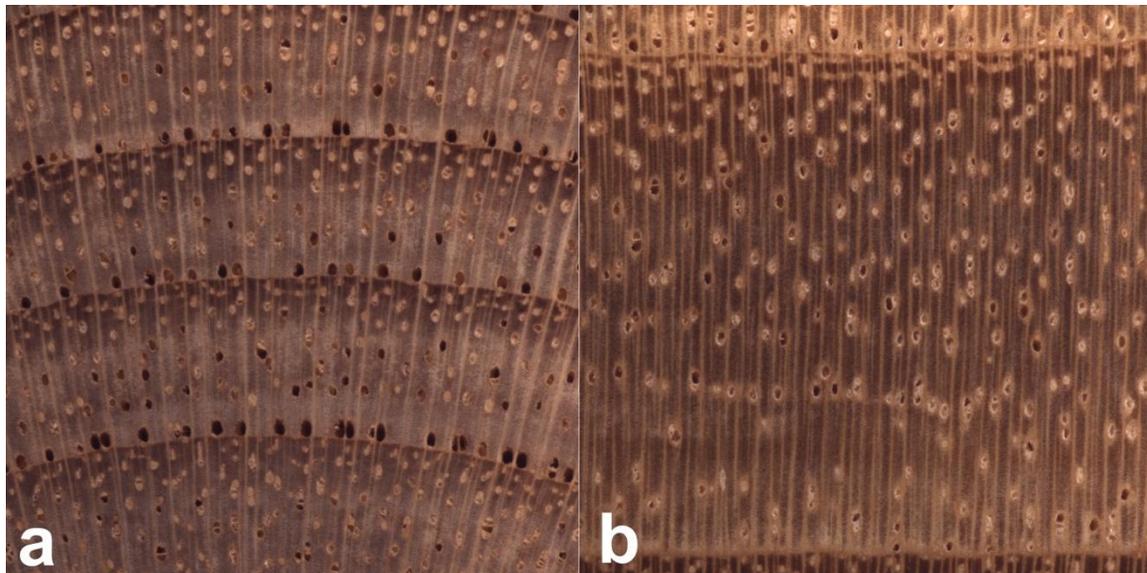

Fig. 4. A test dataset image of *Cladrastis lutea* classified by 2POR-IDL as ring-porous (a) compared with an image of the same taxon classified by the model as diffuse-porous (b). Image (b) appears diffuse-porous, thus this is a Type 2 misclassification (Ravindran *et al.* 2021).

PEER-REVIEWED ARTICLE   *bioresources.cnr.ncsu.edu*

Table 5. Specimen-Level 2POR-IDL Model Predictions when Tested on Out-of-Model Diffuse- and Ring-Porous taxa (*i.e.*, Taxa with Non-Idealized Porosity not Used in Model Training)

Out-of-model Taxa	Total Specimens	2POR-IDL Model Predictions	
		Class, consistent with porosity assignment (count)	Misclassification Type (count), class
Arbutus menziesii	6	DP (6)	
Arbutus xalapensis	3	DP (3)	
Carya aquatica	3	RP (3)	
Carya cordiformis	5	RP (5)	
Carya glabra	9	RP (9)	
Carya illinoinensis	3	RP (3)	
Carya laciniosa	9	RP (9)	
Carya ovata	11	RP (11)	
Carya tomentosa	5	RP (5)	
Catalpa speciosa	4	RP (4)	
Cladrastis lutea	4	RP (3)	Type 2 (1) DP
Fagus grandifolia	13	DP (13)	
Frangula purshiana	1	DP (1)	
Platanus occidentalis	3	DP (3)	
Populus angustifolia	2	DP (2)	
Populus balsamifera	2	DP (2)	
Populus deltoides	5	DP (5)	
Populus fremontii	1	DP (1)	
Populus grandidentata	8	DP (8)	
Populus heterophylla	2	DP (2)	
Populus tremuloides	4	DP (4)	
Populus triochocarpa	2	DP (2)	
Prunus americana	3	DP (3)	
Prunus serotina	16	DP (16)	
Rhamnus crocea	2	DP (2)	
Salix laevigata	1	DP (1)	
Salix lasiandra	4	DP (4)	
Salix nigra	4	DP (4)	
Salix scouleriana	4	DP (4)	
Sorbus americana	2	DP (2)	
Sorbus decora	1	DP (1)	
Ulmus americana	8	RP (8)	
Ulmus thomasii	3	RP (3)	
Counts for consistent predictions are in the third column, and the misclassification Type (per Ravindran *et al.* 2021) is in the fourth column. DP = diffuse-porous, RP = ring-porous. Total n = 153.			

To gain insight into how the model might classify woods that were neither diffuse- nor ring-porous, 2POR-IDL was also tested on 28 specimens from three semi-ring-porous taxa. The results are shown in Table 6. The model identified all the semi-ring-porous woods

bioresources.cnr.ncsu.edu

as ring-porous except for four specimens of *Juglans nigra*, which were classified as diffuse-porous. Figure 5 compares exemplar images of the specimens that were identified as ring- porous with those that were identified as diffuse-porous.

Table 6. 2POR-IDL Specimen-Level Model Predictions when Tested on Out-of-Model Semi-Ring-Porous Taxa

Out-of-model Taxa	Total Specimens	2POR-IDL Model Predictions (count)
Diospyros virginiana	3	RP (3)
Juglans cinerea	12	RP (12)
Juglans nigra	13	RP (9) DP (4)
DP = diffuse-porous, RP = ring-porous. Total n = 28. Note that there cannot be correct or incorrect predictions, as the test specimens exhibit semi-ring-porosity, thus accuracy of prediction is not applicable.		
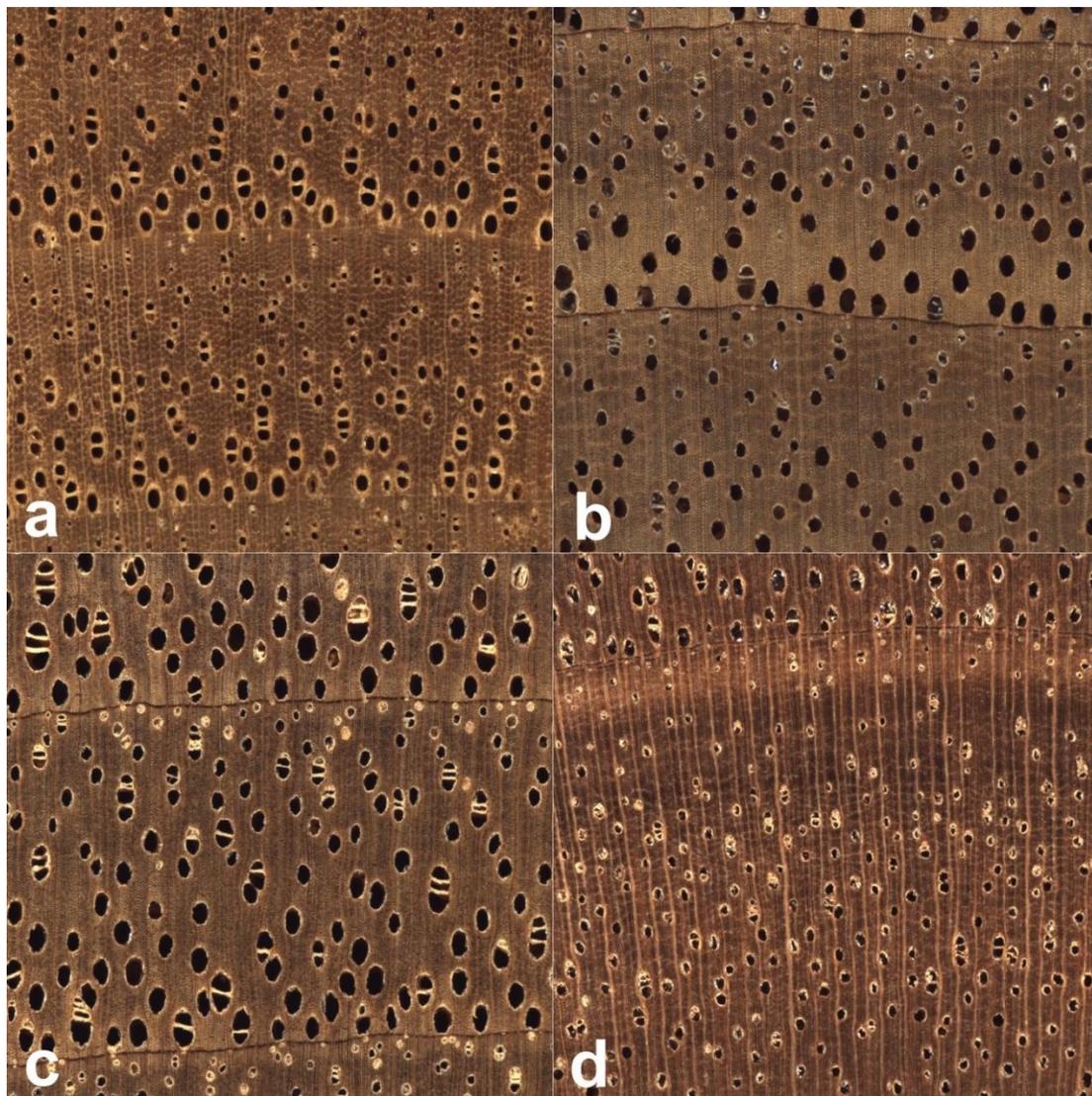
Fig. 5. Images (a, *Diospyros virginiana*), (b, *Juglans cinerea*) and (c*, Juglans nigra*) were all classified by the 2POR-IDL model as ring-porous, while another specimen of *Juglans nigra* (d) was classified as diffuse-porous. The three other *Juglans nigra* specimens the model classified as diffuse-porous exhibited similar, comparatively diffuse-porous structure.

3POR-IDL Predictions on In-model and Out-of-model Taxa

When the 3POR-IDL model was tested on the in-model (idealized) diffuse-, semi-ring-, and ring-porous taxa, all 357 specimens were correctly predicted. When the model was tested on 153 specimens of out-of-model (non-idealized) diffuse- and ring-porous taxa, several predictions differed from the human classification (Table 7). The model predicted 13 out of the 45 specimens of *Carya* to be semi-ring-porous (similar to what is shown in Fig. 3 e, f), two out of the 26 specimens of *Populus* to be ring-porous, and both of the *Rhamnus* specimens to be ring-porous.

Table 7. 3POR-IDL Specimen-Level Model Predictions when Tested on Out-of-Model Diffuse- and Ring-Porous Taxa (*i.e.*, Taxa with Non-Idealized Porosity that Were Not Used in Model Training)

| Out-of-model Taxa	Total Specimens	3POR-IDL Model Predictions	
		Class, consistent with porosity assignment (count)	Misclassification Type (count), class
Arbutus menziesii	6	DP (6)	
Arbutus xalapensis	3	DP (3)	
Carya aquatica	3		Type 2 (3), SRP
Carya cordiformis	5	RP (5)	
Carya glabra	9	RP (9)	
Carya illinoinensis	3	RP (1)	Type 2 (2), SRP
Carya laciniosa	9	RP (5)	Type 2 (4), SRP
Carya ovata	11	RP (8)	Type 2 (3), SRP
Carya tomentosa	5	RP (4)	Type 2 (1), SRP
Catalpa speciosa	4	RP (4)	
Cladrastis lutea	4	RP (4)	
Fagus grandifolia	13	DP (13)	
Frangula purshiana	1	DP (1)	
Platanus occidentalis	3	DP (3)	
Populus angustifolia	2	DP (2)	
Populus balsamifera	2	DP (1)	Type 3 (1), RP
Populus deltoides	5	DP (4)	Type 3 (1), RP
Populus fremontii	1	DP (1)	
Populus grandidentata	8	DP (8)	
Populus heterophylla	2	DP (2)	
Populus tremuloides	4	DP (4)	
Populus triochocarpa	2	DP (2)	
Prunus americana	3	DP (3)	
Prunus serotina	16	DP (16)	
Rhamnus crocea	2		Type 3 (2), RP
Salix laevigata	1	DP (1)	
Salix lasiandra	4	DP (4)	
Salix nigra	4	DP (4)	
Salix scouleriana	4	DP (4)	
Sorbus americana	2	DP (2)	
Sorbus decora	1	DP (1)	
Ulmus americana	8	RP (8)	
Ulmus thomasii	3	RP (3)	
Counts for consistent predictions are in the third column, and the misclassification Type (per Ravindran *et al.* 2021) is in the fourth column. DP = diffuse-porous, SRP = semi-ring-porous RP = ring-porous. Total n = 153.			
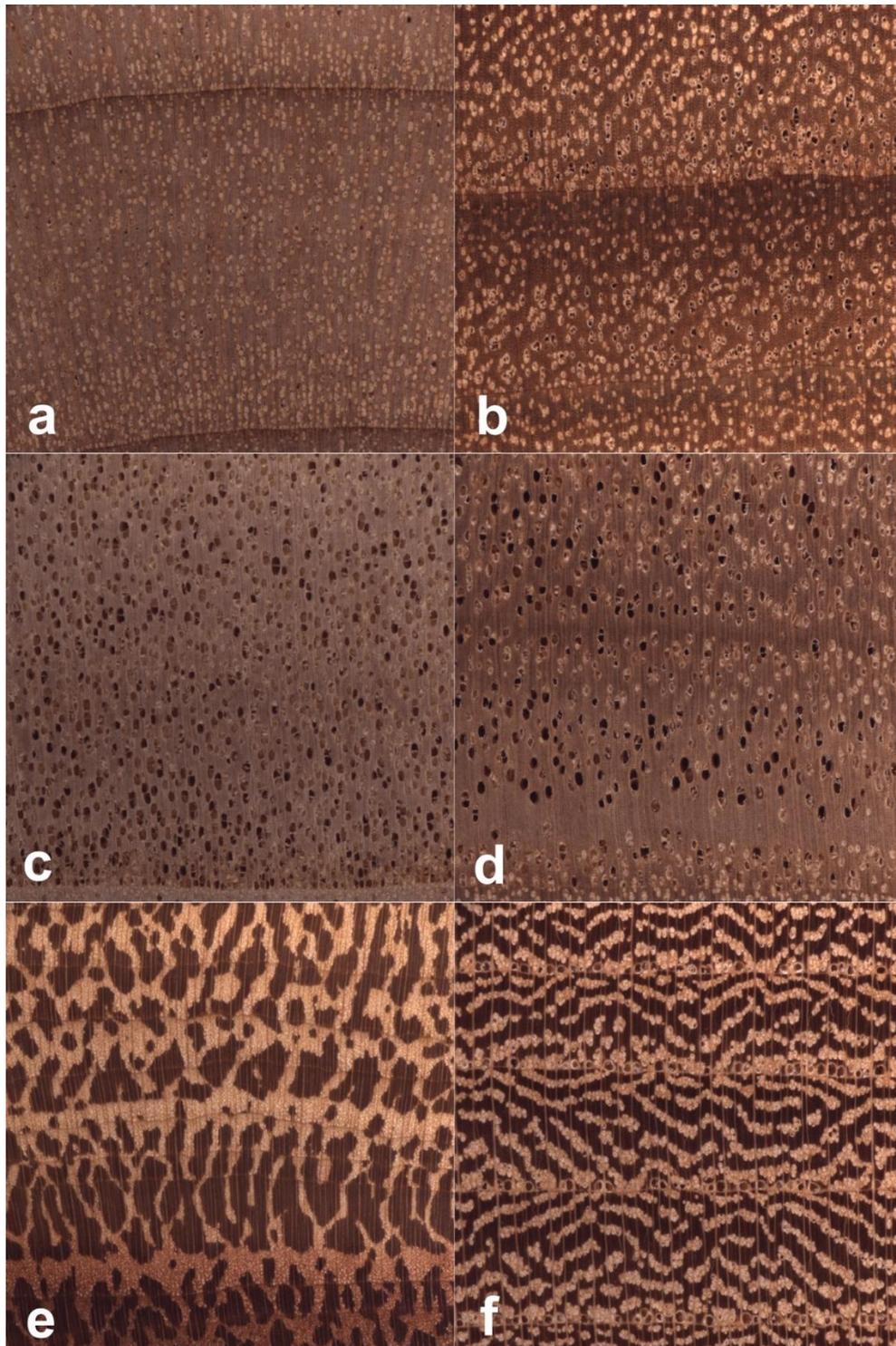
Fig. 6. Woods classified by the 3POR-IDL model: Two specimens of *Populus balsamifera* (a and b) with (a) classified as diffuse-porous and (b) classified as ring-porous, the latter a Type 3 misclassification. Two specimens of *Populus deltoides* (c and d) with (c) classified as diffuse-porous and (d) classified as ring-porous, the latter a Type 3 misclassification. (e) shows a specimen of *Rhamnus* that was incorrectly classified as ring-porous despite being a diffuse-porous wood with dendritic pore arrangement, a Type 3 misclassification. (f) shows a specimen of *Ulmus thomasii* that was correctly classified as ring-porous, despite its lack of continuous, large earlywood vessels.

Exemplar images in Fig. 6 (a-d) contrast correct and incorrect predictions of out-of-model taxa by 3POR-IDL. Fig. 6e shows an exemplar image of a *Rhamnus* specimen that was misclassified as ring-porous, and Fig. 6f shows an image of *Ulmus thomasii* that was correctly placed in the ring-porous class despite the absence of one or more continuous tangential rows of large earlywood vessels. *Rhamnus* was an included taxon, but *Ulmus thomasii*, a hard elm, was an out-of-model taxon.

Evaluation of Idealized Model Performance

Both 2POR-IDL and 3POR-IDL accurately predicted all specimens of the in-model taxa on which they were trained (N=329 and 357, respectively). Selecting specimens only from taxa that commonly exhibit idealized porosity (*e.g.*, *Quercus*, *Fraxinus*, *Acer*, and *Betula*) – as opposed to taxa whose porosity tend to intergrade (*e.g.*, *Carya*, *Catalpa*, *Populus*, and *Salix*) – likely contributed to the models' perfect performance by reducing variability among the three classes in both training and testing.

When testing on out-of-model taxa, 2POR-IDL performed well, classifying 152 out of 153 specimens into their presumed domains of diffuse- and ring-porous. Upon closer examination, the one specimen of *Cladrastis* (a taxon classified by wood anatomists as ring-porous in Table 1) that the model divergently classified as diffuse-porous exhibited an atypical pore distribution that a wood anatomist could have classified as diffuse-porous had s/he been instructed to ignore the taxon label (Fig. 5b), and especially if they did not take note of the earlywood pores at the top of the image. This suggests that the CVWID model has learned to discern subtle differences in porosity that exist along a spectrum, much as a human would.

2POR-IDL had not been trained on images of specimens belonging to the idealized semi-ring-porous class, so all semi-ring-porous test specimens were, by definition, out of model. Table 6 offers two interesting observations. First, most of the semi-ring-porous specimens (all three of the *Diospyros virginiana*, all 12 of the *Juglans cinerea,* and nine of the 13 *Juglans nigra*) were classified as ring-porous. More closely associating semi-ring-porous structure with ring-porous structure is reminiscent of the perceived proximity many dichotomous wood identification keys exhibit when woods are separated early in the key by "earlywood pores not conspicuously larger than latewood pores" – namely, the diffuse-porous condition – and "earlywood pores conspicuously larger than latewood pores," which would include both ring- and semi-ring-porous woods (Panshin and de Zeeuw 1980; Hoadley 1990; Arévalo and Wiedenhoeft 2022). This result suggests that, like humans, the CVWID model more closely relates semi-ring- to ring-porosity. Second, as with the previously mentioned specimen of *Cladrastis*, the four *Juglans nigra* specimens divergently classified as diffuse-porous exhibited an atypical pore distribution that a wood anatomist could have classified as diffuse-porous had s/he been instructed to ignore the taxon label (Fig. 5d). This is further evidence to suggest that the CVWID model has learned to discern subtle difference in porosity that exist along a spectrum much as a human would.

The out-of-model testing results for 3POR-IDL were similarly interesting. The model classified 17 out of 153 specimens differently than did the wood anatomists (Table 7). Fourteen of those divergent classifications were from the genus *Carya*, which is known to exhibit porosity that intergrades between ring- and semi-ring-porous (Hoadley 1990). Upon closer examination, specimens that exhibited porosity closer to ring-porous were classified by the model as ring-porous while those exhibiting semi-ring-porous structure were classified as semi-ring-porous (as shown in Fig. 3e,f). Like the previously mentioned specimens of *Cladrastis* and *Juglans*, these model predictions seemed justifiable based on

the anatomy appearing in the images, and are thus instances of Type 2 misclassifications.

The remaining divergent predictions by 3POR-IDL include two specimens of *Populus* classified as ring-porous (Fig. 6b,d) and two specimens of *Rhamnus* classified as ring-porous (Fig. 6e). There seems to be little evidence in the pore size and distribution of the images to suggest that these predictions were anatomically justified – these are thus characterized as Type 3 misclassifications, errors that neither a human nor the CVWID model should make. Perhaps the model mistook the darker zone in the latewood of the middle growth ring in Fig. 6b, the two (atypical) horizontal zones devoid of pores in Fig. 6d, and the (typical) light and dark regional contrast made by the dendritic pore arrangement to be indicators of ring-porosity (6e). As the features identified by the CNN do not necessarily correspond to the anatomical features, determining the cause of the divergent classifications is at best speculative.

The degree to which test specimen sorting conforms to its designated porosity class seems to depend less on the CNN's ability to distinguish among domains of idealized porosity and more to do with the conformity of the specimens to those ideals. In this regard, the present results suggest that the CNN can learn porosity domains roughly as well as human field screeners after a weeklong wood identification workshop (Wiedenhoeft, personal observation). The misclassifications by this model, while understandable in the context of the anatomical variation in the test specimens, render the model distinctly less useful than the 3-POR model of Owens *et al.* (2024), which when tested on the same test dataset, misclassified only a single specimen. The hypothesis that using quintessential specimens to define the classes might give rise to greater classification accuracy was incorrect, suggesting that CNNs as used in this and the prior studies are able to select features that correctly predict class membership even in the face of anatomical variability, an encouraging result for CVWID models in general.

Implications for Training Multi-Level Computer-Vision Models

Prior work (including our own) has claimed a comparative absence of bias in CVWID models compared to human identifiers, but when experienced wood anatomists define the label space for a model, it influences the performance of the resulting model. This attempt at being "unbiased" is itself imperfect because, as has been shown, which taxa are included in a class influence the performance of otherwise identical models (the 3-POR of Owens *et al.* (2024), compared to 3POR-IDL of this work). In addition to the influence of label space design on model results, it is worth restating assertions made in the past that the kinds of errors made by a model may be more or less severe in practice: for example, confusing a ring-porous wood with a semi-ring porous wood may have lesser practical consequence than confusing a ring-porous wood with a diffuse-porous wood. When using a CVWID model to predict anatomical features in a cascading model scheme such as in Owens *et al.* (2024), such misclassifications can have significant downstream consequences for the necessary taxonomic breadth of lower-level models (*e.g.*, if a porosity classifier regularly classifies some *Carya* as ring-porous and some as semi-ring-porous, unless those distinctions are at a species level where all of *Carya* species X is always ring-porous and all of *Carya* Y is always semi-ring-porous, all *Carya* would have be included in a Carya class in the taxonomic models of both ring-porous and semi-ring-porous woods).

Using the 2POR-IDL CVWID model to interrogate the hardwood porosity continuum has yielded interesting but not entirely surprising results – that semi-ring-porous woods are classified as ring-porous when they exhibit quintessential semi-ring-porosity (Fig. 5a,b,c) and, when they exhibit less distinct semi-ring-porosity, the model classifies

them as diffuse-porous (Fig. 5d). Based on an analysis of the disparately identified specimens, the predictions made by 2POR-IDL are similar to what a non-expert human wood identifier would make. Ravindran *et al.* (2021) identified three types of misclassifications; Table 8 explicitly states the expectations for whether human experts in the laboratory, trained human field agents, or the XyloTron used in field screening would commit each type of misclassification. It is important to be explicit about these expectations, so that technology adoptee expectations are managed appropriately – if scientists or developers of wood identification methods oversell the utility, reliability, scale at which the technology can be deployed, or breadth of wood products to which a technique can be applied, they do a disservice to themselves, the sub-field in which their technology exists, and to wood identification overall. Setting scientifically defensible expectations, then conducting ongoing research to determine possible limitations of a technology is a necessary part of responsible development.

Table 8. Expected Distribution of Misclassification by Type, as Committed by Human Experts in a Laboratory, Trained Human Field Agents, and the Xylotron in a Field Screening Setting

Misclassification	Human-mediated		CVWID
	Expert	Field Agent	XyloTron
Type 1	No	Yes	Yes
Type 2	No	Yes	Yes
Type 3	No	No	No

Using Cascading Models to Reduce Domain-Space Dimensionality

In this paper emphasizing North American hardwoods, a root classifier wood porosity model (diffuse-, semi-ring-, and ring-porous) based on idealized images as was compared to prior work that used an anatomically and taxonomically "messier" classifier. The idealized classifier underperformed compared to a classifier trained with more anatomical (and taxonomic) variability. It is noteworthy that for woods worldwide, such a root-level classifier would likely do comparatively little to reduce domain-space dimensionality of at least one of the submodels, as the great majority of woods worldwide are diffuse-porous (Wheeler 2011). A set of deeper feature-based classifiers (*e.g.*, predominant parenchyma type and ray width/frequency) following the porosity classifier are likely to contribute greater resolution, at least until CVWID methods successfully implement true anatomical feature detection and quantification (*e.g.*, via semantic segmentation).

CONCLUSIONS

1. The 2POR-IDL and 3POR-IDL models both achieved perfect accuracy when tested on independent specimens of the same in-model taxa on which they were trained.

2. When tested on independent specimens from out-of-model (non-ideal) taxa, 2POR-IDL and 3POR-IDL showed lower accuracy than Owens *et al.* (2024), indicating that training on quintessential images did not improve generalization, even though most misclassifications were anatomically sensible.

3. A misclassification analysis revealed remarkable similarities between CVWID models and humans in how they apply the concept of discrete porosity domains to a real-world continuum.

4. When creating labeled datasets for CVWID, label assignments for images are typically made using taxon-level knowledge, and not at the specimen or image level. The effect of this "weak supervision" on model interpretability and its interplay with the ambiguity inherent during discretization of a continuum are still largely unexplored for the variety of wood anatomical features used in traditional wood identification, and this remains an important question for CVWID models as well.

ACKNOWLEDGMENTS

REFERENCES CITED

Arévalo, R., Pulido R., E. N., Solórzano G., J, F, Soares, R., Ruffinatto, F., Ravindran, P. and Windenhoeft, A. C. (2021). "Imaged based identification of Colombian timbers using the Xylotron: A proof of concept international partnership," *Colombia Forestal*

24(1), 5-16. DOI: 10.14483/2256201X.16700

Arévalo, R., and Wiedenhoeft, A. C. (2022). *Identification of Central American, Mexican, and Caribbean Woods* (General technical report FPL-GTR-190), U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI. DOI: 10.2737/FPL-GTR-293

Beeckman, H., Jolivet-Blanc, C., Boeschoten, L., Braga, J. W. B., Cabezas, J. A., Chaix, G., Crameri, S., Degen, B., Deklerck, V., Dormontt, E., *et al*. (2020). "Overview of current practices in data analysis for wood identification: A guide for the different timber tracking methods," in: N. Schmitz (ed.), *GTTN, Global Timber Tracking Network: GTTN Secretariat: European Forest Institute and Thunen Institute.* 143 pp. (https://orfeo.belnet.be/bitstream/handle/internal/12570/Beeckman%2520et%2520al%2520GTTN_2020_DataAnalysisGuide.pdf?sequence=1andisAllowed=y)

Damayanti, R., Prakasa, E., Dewi, L. M., Wardoyo, R., Sugiarto, B., Pardede, H. F., Riyanto, Y., Astutiputri, V. F., Panjaitan, G. R., Hadiwidjaja, M. L., *et al*. (2019). "LignoIndo: Image database of Indonesian commercial timber," *IOP Conf. Series: Earth Environ. Sci.* 374(1), 012057. DOI: 10.1088/1755-1315/374/1/012057

Dormontt, E. E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J. C., and Koch, G. (2015). "Forensic timber identification: It's time to integrate disciplines to combat illegal logging," *Biological Conservation* 191, 790-798. DOI: 10.1016/j.biocon.2015.06.038

Florsheim, S. M. B., Ribeiro, A. P., Longui, E. L., de Andrade, I. M., Sonsin-Oliveira, J., Chimelo, J. P., Soares, R. K., Gouveia, T. C., and Marques, V. N. (2020). *Identificação macroscópica de madeiras comerciais do estado de São Paulo* [*Macroscopic identification of commercial woods of the state of São Paulo*], Instituto Florestal. (Original work published in Portuguese).

Goloboff, P. A., Mattoni, C. I., and Quinteros, A. S. (2006). "Continuous characters analyzed as such," *Cladistics* 22, 589-601. DOI: 10.1111/j.1096-0031.2006.00122.x

Hermanson, J. C., and Wiedenhoeft, A. C. (2011). "A brief review of machine vision in the context of automated wood identification systems," *IAWA Journal* 32(2), 233-250. DOI: 10.1163/22941932-90000054

Hoadley, R. B. (1990). *Identifying Wood- Accurate Results with Simple Tools*, Taunton Press, Newtown, CT, USA.

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

Johnson, A., and Laestadius, L. (2011). "New laws, new needs: The role of wood science in global policy efforts to reduce illegal logging and associated trade," *IAWA Journal* 32(2), 125-136. DOI: 10.1163/22941932-90000048

Koch, G., Haag, V., Heinz, I., Richter, H., and Schmitt, U. (2015). "Control of international traded timber — The role of macroscopic and microscopic wood identification against illegal logging," *Journal of Forensic Research* 6, article 317. DOI: 10.4172/2157-7145.1000317

Lowe, A. J., Dormontt, E., Bowie, M., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoeft, A. C., Yin, Y., and Sasaki, N. (2016). "Opportunities for improved transparency in the timber trade through scientific verification," *BioScience* 66 (11), 990-998. DOI: 10.1093/biosci/biw129

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

Ministerio del Ambiente (MINAM). (2022). Guía para la elaboración de planes de manejo forestal con fines de conservación en el marco de contratos de cesión en uso para sistemas agroforestales y recuperación de ecosistemas forestales y otros ecosistemas de vegetación silvestre (Manual No. 001-2022-MINAM/VMDERN/DGOT). Retrieved from https://cdn.www.gob.pe/uploads/document/file/4158691/87769%2000%20COOP%20ALEMANA%20MANUAL%20-%20web.pdf.pdf

Owens, F., Ravindran, P., C., Costa, Shmulsky, R., and Wiedenhoeft, A. C. (2024). "Predicting hardwood porosity domains: Toward cascading computer-vision wood identification models," *BioResources* 19(4), 9741-9772. DOI: 10.15376/biores.19.4.9741-9772

Panshin, A. J. and De Zeeuw, C. (1980). *Textbook of Wood Technology*, Fourth Edition, McGraw-Hill, New York.

Parins-Fukuchi, C. (2018). "Use of continuous traits can improve morphological phylogenetics," *Systematic Biology* 67(2), 328-339. DOI: 10.1093/sysbio/syx072

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), article 25. DOI: 10.1186/s13007-018-0292-9

Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R., Costa, A., and Wiedenhoeft, A. (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 (arXiv:1912.00296v1)

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," *Frontiers in Plant Science* 11, article 1015. DOI: 10.3389/fpls.2020.01015

Ravindran, P., Owens, F. C., Wade, A. C., Vega, P., Montenegro, R., Shmulsky, R., and Wiedenhoeft, A. C. (2021). "Field-deployable computer vision wood identification of Peruvian timbers," *Frontiers in Plant Science* 12, article 647515. DOI: 10.3389/fpls.2021.647515

Ravindran, P., Owens, F. C., Wade, A. C., Shmulsky, R., and Wiedenhoeft, A. C. (2022a). "Towards sustainable North American wood product value chains, Part I: Computer vision identification of diffuse-porous hardwoods," *Frontiers in Plant Science* 12, article 758455. DOI: 10.3389/fpls.2021.758455

Ravindran P., Wade A. C., Owens F. C., Shmulsky R., and Wiedenhoeft A. C. (2022b). "Towards sustainable North American wood product value chains, part 2: Computer vision identification of ring-porous hardwoods," *Can. J. Forest Res.* 52, 1014-1027. DOI: 10.1139/cjfr-2022-0077.

Ravindran, P., and Wiedenhoeft, A. C. (2022). "Caveat emptor: On the need for baseline quality standards in computer vision wood identification," *Forests* 13, article 632. DOI: 10.3390/f13040632

Richter, H. G., Grosser, D., Heinz, I., and Gasson, P. E. (2004). "IAWA list of microscopic features for softwood identification," *IAWA Journal*, 25(1), 1-70.

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

Silva, J. L., Bordalo, R., Pissarra, J., and de Palacios, P. (2022). "Computer vision-based wood identification: A review," *Forests* 13(12), 2041. DOI: 10.3390/f13122041

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* (CIIS '18). Association for Computing Machinery, New York, NY, USA, 37–43. DOI: 10.1145/3293475.3293493

Thiele, K. (1993). "The holy grail of the perfect character: The cladistic treatment of morphometric data," *Cladistics* 9(3), 275-304. DOI: 10.1006/clad.1993.1020

United Nations Office on Drugs and Crime (UNODC) (2016). *Best Practice Guide for Forensic Timber Identification*, New York.

Wheeler, E., Baas, P., and Gasson, P. (1989). "IAWA list of microscopic features for hardwood identification," *IAWA Journal* 10, 219-332. DOI: 10.1163/22941932-90000496

Wheeler, E. A. (2011). "Inside wood – A Web resource for hardwood anatomy," *IAWA Journal* 32(2), 199-211. DOI: 10.1163/22941932-90000051

Wiedenhoeft A. C. (2011). *Identification of Central American Woods* (Identificacion de las Especies Maderables de Centroamerica). Forest Products Society, Madison, WI. Publication #7215-11

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

APPENDIX

Table S1. Taxa Used in Model Training (Idealized Taxa)

Acer macrophyllum	*Crataegus rivularis*	*Nyssa biflora*	*Quercus stellata*
Acer negundo	*Crataegus rotundifolia*	*Nyssa ogeche*	*Quercus texana*
Acer rubrum	*Crataegus spathulata*	*Nyssa sylvatica*	*Quercus velutina*
Acer saccharinum	*Crataegus succulenta*	*Ostrya virginiana*	*Rhamnus californica*
Acer saccharum	*Crataegus tomentosa*	*Oxydendrum arboreum*	*Robinia neo-mexicana*
Aesculus californica	**Diospyros virginiana*	*Prunus americana*	*Robinia pseudoacacia*
Aesculus glabra	*Fraxinus americana*	*Prunus angustifolia*	*Sassafras albidum*
Aesculus hippocastanum	*Fraxinus nigra*	*Prunus avium*	*Tilia americana*
Aesculus octandra	*Fraxinus oregona*	*Prunus caroliniana*	*Tilia caroliniana*
Alnus rhombifolia	*Fraxinus pennsylvanica*	*Prunus emarginata*	*Tilia floridana*
Alnus rubra	*Fraxinus quadrangulata*	*Prunus myrtifolia*	*Tilia heterophylla*
Alnus rugosa	*Gleditsia aquatica*	*Prunus nigra*	*Tilia pubescens*
Alnus serrulata	*Gleditsia triacanthos*	*Pyrus ioensis*	*Ulmus rubra*
Alnus tenuifolia	*Gymnocladus dioica*	*Quercus alba*	
Asimina triloba	**Juglans cinerea*	*Quercus arkansana*	
Betula alleghaniensis	**Juglans nigra*	*Quercus bicolor*	
Betula lenta	*Liquidambar styraciflua*	*Quercus coccinea*	
Betula nigra	*Liriodendron tulipifera*	*Quercus ellipsoidalis*	
Betula occidentalis	*Maclura pomifera*	*Quercus falcata*	
Betula papyrifera	*Magnolia acuminata*	*Quercus georgiana*	
Betula populifolia	*Magnolia fraseri*	*Quercus ilicifolia*	
Carpinus caroliniana	*Magnolia grandiflora*	*Quercus incana*	
Castanea dentata	*Magnolia macrophylla*	*Quercus laevis*	
Castanea pumila	*Magnolia tripetala*	*Quercus laurifolia*	
Celtis occidentalis	*Magnolia virginiana*	*Quercus lyrata*	
Celtis reticulata	*Malus angustifolia*	*Quercus macrocarpa*	
Crataegus aestivalis	*Malus baccata*	*Quercus marilandica*	
Crataegus assurgens	*Malus coronaria*	*Quercus michauxii*	
Crataegus compacti	*Malus domestica*	*Quercus montana*	
Crataegus cordata	*Malus pumila*	*Quercus myrtifolia*	
Crataegus cuneiformis	*Malus rivularis*	*Quercus nigra*	
Crataegus douglasii	*Malus* sp.	*Quercus palustris*	
Crataegus macracantha	*Morus alba*	*Quercus phellos*	
Crataegus mollis	*Morus rubra*	*Quercus rubra*	
Crataegus nitida	*Nyssa aquatica*	*Quercus shumardii*	

Taxa noted with * were included in 3POR-IDL, but not 2POR-IDL.

bioresources.cnr.ncsu.edu

Table S2. Taxa Used in Model Testing

Acer macrophyllum	*Celtis occidentalis*	*Malus angustifolia*	*Quercus alba*
Acer negundo	*Celtis* sp.	*Malus coronaria*	*Quercus bicolor*
Acer rubrum	*Cladrastis lutea*	*Malus fusca*	*Quercus coccinea*
Acer saccharinum	*Crataegus aestivalis*	*Malus pumila*	*Quercus falcata*
Acer saccharum	*Crataegus calpodendron*	*Morus alba*	*Quercus laurifolia*
Aesculus californica	*Crataegus douglasii*	*Morus rubra*	*Quercus lyrata*
Aesculus glabra	*Crataegus mollis*	*Nyssa aquatica*	*Quercus macrocarpa*
Aesculus octandra	*Crataegus rivularis*	*Nyssa ogeche*	*Quercus marilandica*
Alnus incana	*Crataegus spathulata*	*Nyssa sylvatica*	*Quercus montana*
Alnus rhombifolia	*Diospyros virginiana*	*Nyssa sylvatica* var. *biflora*	*Quercus nigra*
Alnus rubra	*Fagus grandifolia*	*Ostrya virginiana*	*Quercus phellos*
Arbutus menziesii	*Frangula purshiana*	*Oxydendrum arboreum*	*Quercus shumardii*
Arbutus xalapensis	*Fraxinus nigra*	*Platanus occidentalis*	*Quercus velutina*
Asimina triloba	*Fraxinus pennsylvanica*	*Populus angustifolia*	*Rhamnus crocea*
Betula lenta	*Fraxinus quadrangulata*	*Populus balsamifera*	*Robinia pseudoacacia*
Betula nigra	*Gleditsia triacanthos*	*Populus deltoides*	*Salix laevigata*
Betula occidentalis	*Gymnocladus dioica*	*Populus fremontii*	*Salix lasiandra*
Betula papyrifera	*Juglans cinerea*	*Populus grandidentata*	*Salix nigra*
Betula populifolia	*Juglans nigra*	*Populus heterophylla*	*Salix scouleriana*
Carpinus caroliniana	*Liquidambar styraciflua*	*Populus tremuloides*	*Sassafras albidum*
Carya aquatica	*Liriodendron tulipifera*	*Populus trichocarpa*	*Sassafras* sp.
Carya cordiformis	*Maclura pomifera*	*Prunus americana*	*Sorbus americana*
Carya glabra	*Maclura* sp.	*Prunus angustifolia*	*Sorbus decora*
Carya illinoinensis	*Magnolia acuminata*	*Prunus caroliniana*	*Tilia americana*
Carya laciniosa	*Magnolia fraseri*	*Prunus emarginata*	*Tilia americana* var. *heterophylla*
Carya ovata	*Magnolia grandiflora*	*Prunus myrtifolia*	*Ulmus americana*
Carya tomentosa	*Magnolia macrophylla*	*Prunus serotina*	*Ulmus thomasii*
Castanea dentata	*Magnolia tripetala*		
Catalpa speciosa	*Magnolia virginiana*		