Wood Science and Technology (2022) 56:1567–1583
https://doi.org/10.1007/s00226-022-01404-y

ORIGINAL

Can quantitative wood anatomy data coupled with machine learning analysis discriminate CITES species from their look-alikes?

Shoujia Liu[1,2] · Tuo He[1,2] · Jiajun Wang[1] · Jiabao Chen[1,2] · Juan Guo[1,2] · Xiaomei Jiang[1,2] · Alex C. Wiedenhoeft[3,4,5,6,7] · Yafang Yin[1,2]

Received: 30 January 2022 / Accepted: 5 August 2022 / Published online: 12 August 2022

Abstract
Due to increasing global trade of timber commodities and illegal logging activities, wood species listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) appendices are facing extinction, and their international trade has been banned or is under supervision. Reliable and applicable species-level discrimination methods have become urgent to protect global forest resources and promote the legal trade of timbers. This study aims to discriminate CITES-listed species from their look-alikes in international trade using quantitative wood anatomy (QWA) data coupled with machine learning (ML) analysis. Herein, the QWA data of 14 CITES-listed species and 15 of their look-alike species were collected from microscope slide collection, and four ML classifiers, J48, Multinomial Naïve Bayes, Random Forest, and SMO, were used to analyze the QWA data. The results indicated that ML classifiers exhibited better performance than traditional wood identification methods. Specifically, Multinomial Naïve Bayes outperformed other classifiers, and successfully discriminated CITES-listed *Pterocarpus* species from their look-alike species with an accuracy of 95.83%. Furthermore, the discrimination accuracy was affected by the combinations of wood anatomical features, and combinations with fewer features included could result in higher accuracy at the species level. In conclusion, the QWA data coupled with ML analysis could unlock the potential of wood anatomy to discriminate CITES species from their look-alikes for forensic applications.

✉ Tuo He
tuohe@caf.ac.cn

Extended author information available on the last page of the article

Introduction

In the last decades, illegal logging and associated trade have become a primary concern globally, as they are a major cause of deforestation, biodiversity loss, and ecosystem imbalance (Brancalion et al. 2018; Richardson and Peres 2016; Young et al. 1996). Globally, it is estimated that 15%–30% of the timbers in the international trade are illegally harvested, accounting for \$51 to \$152 billion annually (INTERPOL 2021). The international community has adopted law enforcement measures to combat the illegal trade of timbers (Gulbrandsen and Humphreys 2006), emphasizing the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), which came into effect on July 1, 1975, to guarantee that wild animal and plant survival is not threatened by international trade (Goldsmith 1978). CITES lists species in one of three appendices (I, II, III) based on the degree of protection needed to regulate their trade. As of the close of the latest Conference of the Parties (CoP) of CITES, the 18th, held in August 2019, more than 520 tree species have been listed in CITES appendices, including a number of commercially important timbers (CITES 2019), for example, *Dalbergia*, *Pterocarpus*, *Cedrela,* and *Swietenia*.

Dalbergia is endemic to tropical and subtropical regions of Asia, Africa, and the Americas, and is commonly used for furniture, musical instruments, handicrafts, and medicine (Ruffinatto and Crivellaro 2019; Wick 2019). All *Dalbergia* species were listed in CITES Appendix II at CoP 17 in 2016, except *Dalbergia nigra,* which was previously listed in CITES Appendix I. *Pterocarpus* is also distributed in tropical areas and *P. santalinus*, *P. erinaceus,* and *P. tinctorius* are listed in CITES Appendix II due to their over-exploitation for furniture, musical instruments, and high-grade decorative materials (Da Silva et al. 2019; Dumenu 2019). *Cedrela* has 17 recognized species distributed from Mexico through Central America and into northern South America (Pennington et al. 2010) and *Cedrela* spp. growing in the neotropics were uplisted to CITES Appendix II at CoP18 due to the over-exploitation of their timbers (CITES 2019). Similarly, three *Swietenia* species, i.e., *S. mahagoni*, *S. macrophylla*, and *S. humilis* have been commercially dominant in the neotropical wood trade for centuries. *S. mahagoni* is one of the most valuable and traded timber species globally, though it is largely replaced in the market by *S. macrophylla* in modern times as *S. mahagoni* was largely commercially extirpated (Corneliusa et al. 2004; Rodan et al. 1992). This pattern of shifting from historically valuable to progressively lesser-known timbers as the abundance of historical timber dwindles is a driver of the changing needs for wood identification tools to promote the legal and sustainable trade of timbers and improve compliance with regulations.

Traditional wood identification is a laborious process performed by experienced wood anatomists observing and interpreting macroscopic and microscopic characters of wood (Wheeler and Baas 1998; Yin et al. 2016), which is time-consuming and prone to human subjectivity (Wiedenhoeft et al. 2019). International conservation mechanisms like CITES require field-level identification of potentially illegal shipments, but in many jurisdictions, subsequent enforcement decisions depend upon forensic determinations, thus quantitative and conclusive data are critically needed.

Quantitative wood anatomy (QWA) quantifies wood anatomical features to address research questions related to plant function, growth, and response to the environment (Beeckman 2016; He et al. 2021; von Arx et al. 2016). In addition to these ecophysiological uses, QWA can be used to study subtle anatomical variations of key characters and by using multivariate statistical analysis to eliminate human bias in the field of wood identification. Gasson et al. (2010) distinguished *Dalbergia nigra* from congeners using QWA, principal components analysis, and Naïve Bayes classification, demonstrating that Naïve Bayes combined with specific wood anatomical characters correctly classified all eight *D. nigra* specimens. He et al. (2020b) used QWA data and machine learning (ML) models to discriminate between three *Swietenia* species, where the overall accuracy of support vector machine (SVM) reached 91.4%, and in contrast to new wood identification methods such as computer vision and DNA barcoding, QWA can be performed on a traditional laboratory basis and does not require the investment of new and expensive instruments (Hwang and Sugiyama 2021; Jiao et al. 2020; Lowe et al. 2016). Although these prior successes are encouraging, the effects of QWA features on discrimination accuracy of CITES-listed species from their look-alikes and the resulting discrimination capacity especially at a relatively larger scale have not been further investigated for forensic application.

In this study, a wider range of CITES-listed commercially important timber species and their look-alike species not listed in CITES were targeted and their QWA data were collected for statistical analysis. The specific aims were as follows: (1) to determine the feasibility of ML analysis for discriminating CITES species from their look-alikes at the species level; (2) to explore the effects of wood anatomical features on the identification accuracy of four ML classifiers for optimizing anatomical features selection used for species discrimination; (3) to propose a conclusive and applicable method for wood identification that requires less expertise in forensic wood identification.

Materials and methods

Specimens preparation and image collection

Fourteen commercially important CITES-listed species, including six *Dalbergia*, three *Pterocarpus*, three *Swietenia*, two *Cedrela*, and fifteen look-alike species which were often mixed with CITES-listed species in trade, were selected for this study. A total of 141 specimens representing these 29 species from 13 genera were collected from the USDA Forest Products Laboratory Wood Collection of Madison (MADw), the Samuel J. Record Collection (SJRw), and the Wood Collection of the Chinese Academy of Forestry (CAFw) (Table S1). Transmitted light micrographs of 4096 pixels × 2160 pixels of two planes of section (transverse and tangential) of prepared slides were collected with a digital camera mounted on a Leica DFC-425C microscope. For each specimen, five non-overlapping images were randomly captured with a 2.5 × objective for transverse sections and a 5 × objective for tangential sections, with a total of ten images collected from each specimen. Considering that CITES-listed species are often disguised as their look-alikes in the real world due to similar wood anatomical features (Richter et al. 2014), we divided the CITES-listed species and their look-alikes into six groups based on the similarity of wood anatomical features and the possibility of mislabeling in the documents submitted to the customs departments (Table 1).

Measurement of wood anatomical features

Five quantitative wood anatomical features of vessels and rays were measured, as described in Table 2. The wood anatomical features were measured using Image J 2.0 (Schindelin et al. 2015). Ten measurements were obtained randomly for each specimen, and the average value of vessel diameter (VD), ray height (RH), and ray width (RW) was calculated.

Statistical analysis

Four supervised ML algorithms, i.e., J48, Multinomial Naïve Bayes, Random Forest, and SMO, were deployed in the Waikato Environment for Knowledge Analysis (WEKA) (Hornik et al. 2009). J48 is based on a top-down strategy to classify the instances sequentially, and it stops computing when all instances have the same classification criteria (Frank et al. 2009; Smith and Frank 2016). Multinomial Naïve Bayes assumes independence between features among the groups and selects the optimal category marker based on prior probability and misjudgment loss (Jiang et al. 2013). Random Forest is a collection of unpruned classification or regression trees extracted from bootstrap training data specimens, using random character selection in the tree induction process (Liaw and Wiener 2002). SMO is based on statistical learning theory to improve the generalization ability of the model by seeking the minimum structured risk so that the model can obtain better statistical laws with fewer statistical specimens (Keerthi and Gilbert 2002; Shevade et al. 2000). To ensure that the errors are representative of the entire data set, specimen-level threefold cross-validation for each model was used and accuracy is reported as the average over the three folds (Refaeilzadeh et al. 2009).

The QWA data of all species was input to WEKA, and the accuracy of correct identification was calculated (Smith and Frank 2016). Additionally, the QWA data of different combinations of wood anatomical features from six groups were used as inputs, respectively, and the accuracy of successfully predicted species was output.

Table 1 CITES-list species and their look-alikes with category used in this study

Category	CITES-listed species	Look-alike species			
Group 1	*Pterocarpus erinaceus*	*Pterocarpus dalbergioides*	*Pterocarpus indicus*	*Pterocarpus angolensis*	*Pterocarpus macrocarpus*
Group 2	*Pterocarpus santalinus Pterocarpus tinctorius*	*Pterocarpus soyauxii*	*Ormosia microphylla*	*Baikiaea plurijuga*	
Group 3	*Dalbergia cochinchinensis Dalbergia tucurensis Dalbergia sissoo Dalbergia stevensonii Dalbergia retusa Dalbergia latifolia*	*Bobgunnia madagascariensis*	*Platymiscium pinnatum*	*Burkea africana*	
Group 4	*Cedrela odorata*	*Guarea cedrata*	*Entandrophragma cylindricum*	*Khaya anthotheca*	
Group 5	*Cedrela fissilis*	*Carapa guianensis*	*Carapa procera*		
Group 6	*Swietenia mahagoni Swietenia macrophylla Swietenia humilis*	*Guarea cedrata*	*Carapa guianensis*		

Results and discussion

Wood anatomical features and QWA data of CITES-listed species and their look-alikes

Figure 1 presents the micrographs of three planes of section of CITES-listed species *Pterocarpus erinaceus* and its look-alikes, as an example showing the similarity of these species concerning wood anatomical features. Micrographs of other CITES-listed species and their look-alikes are presented in Fig. S1-S5 in Supplementary Information, which illustrates the difficulty in separating these species correctly by examining the qualitative wood anatomical features.

Table 3 lists the QWA data of selected anatomical features for 29 species. Regarding the Group 1, *P. erinaceus* had a smaller VD value than the other species and equivalent VF, RF and RW values with *P. macrocarpus*. Specifically, in Group 4, *C. odorata* showed much lower RF than *Guarea cedrata*, and lower RH values in comparison with *Entandrophragma cylindricum and Khaya anthotheca.* Similarly, in Group 5, *C. fissilis* showed much lower VF and RH values than *Carapa guianensis* and *C. procera.* In Group 6, three *Swietenia* species showed much lower RH values than *C. guianensis*, and a much higher RW value than *G. cedrata*. Although the QWA data of the species from different group exhibited a degree of divergence, it remains challenging to discriminate CITES-listed species from their look-alikes without further statistical analysis, especially for those species within one group.

Table 2 Descriptions of measured and calculated features

Wood Anatomical Feature	Acronym	*n*	Description
Vessel Diameter (μm)	VD	50	Mean tangential diameter of vessels; measured at the widest point of the vessel and including the vessel wall over five transverse fields of view per specimen
Vessel Frequency (/mm^2)	VF	5	Mean number of vessels per mm^2, count the number of vessels in 11.27 mm^2 on five fields of view per transverse section per specimen
Ray Width (μm)	RW	50	Mean width of rays on five tangential fields of view per specimen
Ray Height (μm)	RH	50	Mean height of rays (tails included) on five tangential fields of view per specimen
Ray Frequency (/mm)	RF	5	The number of rays crossing a 1 mm straight line on five tangential fields of view per specimen

The performance of different classifiers based on combinations of wood anatomical features

Two challenges faced by traditional wood identification methods are that species-level resolution is rarely achieved, and it takes much time to train an experienced wood identification expert. The QWA method is more suitable than traditional identification methods for inexperienced people, who only need to collect quantitative anatomical data referring to established guidelines using commonly available measurement tools and feed those data into developed models.

Wood is a complex biological structure of many cell types, the structure of a typical hardwood is composed of fibers, vessels, parenchyma and rays (Wiedenhoeft and Miller 2005). In this study, we selected five important features in relation to vessels and rays, which are more easily quantifed than other features like axial parenchyma patterns. The overall success rate of discrimination of 29 species using four classifiers is shown in Fig. 2. The highest accuracy is reported by Multinomial Naïve Bayes, though it is relatively dissatisfactory. The four classifiers were used to discriminate the species of each group with 30 combinations of wood anatomical features, and the results are presented in Fig. 3. The Multinomial Naïve Bayes classifier performed better than the other three classifiers in Groups 1–5. Specifically, in Group 2, the discrimination accuracy of Multinomial Naïve Bayes reached an accuracy over 94% in a four-class model. Gasson et al. (2010) also reported that Naïve Bayes classifier could discriminate *D. nigra* from other *Dalbergia* species with QWA data. Random Forest performed the best in Group 6, which is a collection of different decision trees that are generated based on the initial data randomly selected from the training set (Chen et al. 2018). The identification accuracy of SMO presented here is somewhat inferior to the previously reported species-level wood identification using quantitative anatomical (He et al. 2020b) and genetic data for a range of organisms including tropical trees (Hartvig et al. 2015; He et al. 2019). Herein, we demonstrated that Naïve Bayes classifier showed the best performance when discriminating CITES-listed species from their look-alikes using QWA data.

Effects of wood anatomical features on species identification

Gasson et al. (2010) showed that more anatomical features led to better classification results in traditional wood identification for *Dalbergia nigra*. In the context of QWA, the combination of three or more wood anatomical features generally resulted in higher accuracy than two or one features, however, the combination with the most features did not always perform the best. In Group 1, a four-feature combination (VD, VF, RH, and RW) performed the best (81.82%) when discriminating *P. erinaceus* from its look-alikes, and the highest identification rate was achieved by the other combination of four features without RF included, as RF of all taxa were similar, perhaps as a result of the uniform presence of storied rays, which is consistent

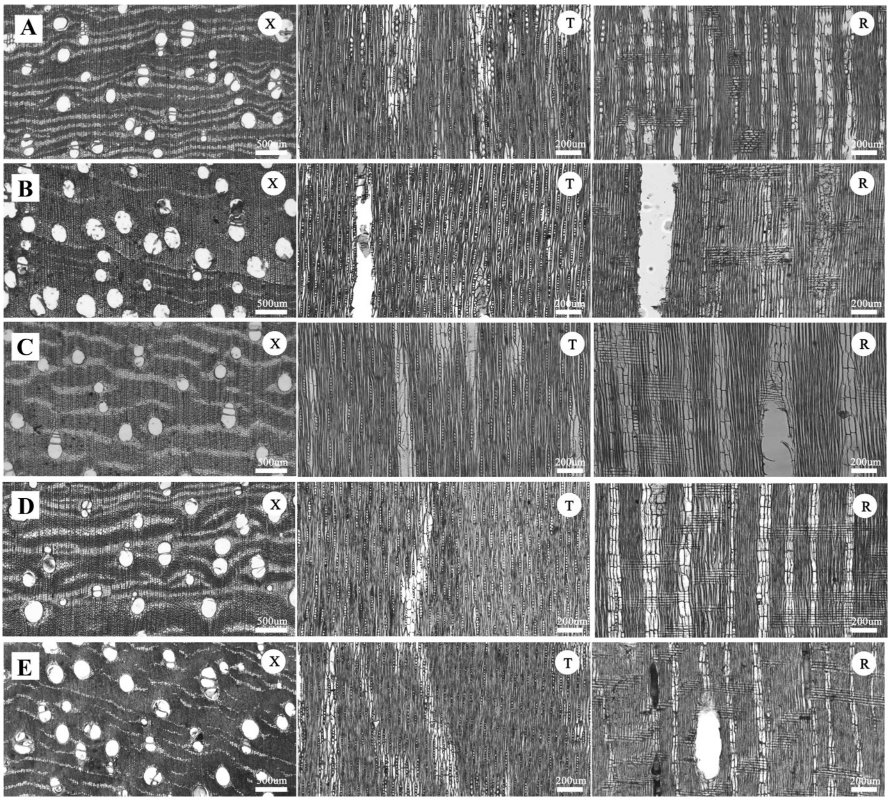

Fig. 1 Light micrographs of the woods included in Group 1. X-transverse, T-tangential, and R-radial sections of *Pterocarpus erinaceus* (**A**), *Pterocarpus dalbergioides* (**B**), *Pterocarpus indicus* (**C**), *Pterocarpus angolensis* (**D**) and *Pterocarpus macrocarpus* (**E**) showing the similarity of the five species concerning wood anatomical features

Table 3 Quantitative wood anatomy (QWA) data of selected anatomical features for 29 species

Category	Species	*n*	Vessel diameter (µm)	Vessel frequency	Ray frequency	Ray width (µm)	Ray hight (µm)
Group 1	*Pterocarpus erinaceus*	3	121 ± 41^{b}	7.5 ± 3.4^{b}	12.3 ± 1.3^{a}	11 ± 2.6^{b}	136 ± 12^{a}
	Pterocarpus dalbergioides	3	204 ± 78^{a}	3.5 ± 0.4^{c}	13.2 ± 2.1^{a}	12 ± 12^{b}	146 ± 4^{a}
	Pterocarpus indicus	5	178 ± 16^{a}	3.4 ± 0.8^{c}	11.3 ± 1.1^{a}	15 ± 4^{ab}	157 ± 17^{a}
	Pterocarpus angolensis	5	223 ± 34^{a}	2.9 ± 0.3^{c}	10.0 ± 2.6^{a}	19 ± 2^{a}	144 ± 18^{a}
	Pterocarpus macrocarpus	6	199 ± 19^{a}	12.9 ± 1.7^{a}	13.2 ± 1.5^{a}	14 ± 2^{b}	154 ± 18^{a}
Group 2	*Pterocarpus santalinus*	4	138 ± 7^{b}	5.9 ± 1.0^{b}	12.1 ± 0.9^{a}	13 ± 1^{c}	126 ± 17^{c}
	Pterocarpus tinctorius	6	115 ± 11^{b}	7.1 ± 2.7^{b}	12.7 ± 2.1^{a}	12 ± 3^{c}	129 ± 7^{c}
	Pterocarpus soyauxii	6	252 ± 20^{a}	1.6 ± 0.5^{b}	12.9 ± 1.9^{a}	16 ± 2^{c}	162 ± 16^{c}
	Ormosia microphylla	3	135 ± 15^{b}	3.2 ± 0.9^{b}	7.7 ± 0.3^{b}	33 ± 5^{a}	347 ± 22^{a}
	Baikiaea plurijuga	5	61 ± 8^{c}	59.9 ± 11.6^{a}	10.8 ± 1.2^{ab}	24 ± 3^{b}	229 ± 31^{b}
Group 3	*Dalbergia tucurensis*	5	189 ± 30^{a}	2.9 ± 0.6^{b}	11.4 ± 1.0^{ab}	22 ± 1^{ab}	136 ± 8^{c}
	Dalbergia sissoo	5	156 ± 4^{ab}	4.3 ± 0.5^{b}	12.9 ± 1.7^{a}	19 ± 2^{bcd}	129 ± 4^{c}
	Dalbergia stevensonii	4	135 ± 11^{bc}	3.4 ± 1.4^{b}	11.5 ± 0.7^{ab}	20 ± 2^{bcd}	138 ± 17^{c}
	Dalbergia cochinchinensis	6	166 ± 10^{ab}	2.5 ± 0.5^{b}	10.2 ± 0.7^{bc}	20 ± 4^{bcd}	143 ± 6^{c}
	Dalbergia retusa	6	161 ± 15^{ab}	1.8 ± 0.2^{b}	12.7 ± 1.3^{a}	15 ± 1^{d}	140 ± 11^{c}
	Dalbergia latifolia	5	167 ± 14^{ab}	3.8 ± 0.6^{b}	9.7 ± 0.7^{bc}	27 ± 2^{a}	158 ± 13^{bc}
	Platymiscium pinnatum	4	160 ± 25^{ab}	5.9 ± 3.2^{b}	10.9 ± 1.6^{ab}	16 ± 4^{cd}	189 ± 13^{ab}
	Bobgunnia madagascariensis	4	87 ± 7^{d}	15.2 ± 5.6^{a}	10.0 ± 1.4^{bc}	21 ± 1^{bc}	147 ± 23^{c}
	Burkea africana	5	105 ± 12^{cd}	11.7 ± 1.2^{a}	8.1 ± 0.7^{c}	24 ± 3^{ab}	216 ± 33^{a}
Goup 4	*Cedrela odorata*	5	183 ± 11^{a}	3.9 ± 1.2^{a}	4.8 ± 0.2^{b}	56 ± 9^{a}	326 ± 38^{b}
	Entandrophragma cylindricum	3	149 ± 3^{bc}	8.8 ± 1.8^{b}	4.9 ± 0.4^{b}	53 ± 6^{a}	435 ± 38^{a}
	Guarea cedrata	5	133 ± 8^{c}	12.9 ± 2.9^{ab}	7.8 ± 1.2^{a}	25 ± 3^{b}	325 ± 33^{b}
	Khaya anthotheca	3	171 ± 26^{ab}	7.1 ± 3.3^{b}	5.4 ± 1.4^{b}	55 ± 14^{a}	438 ± 30^{a}

Table 3 (continued)

Category	Species	*n*	Vessel diameter (µm)	Vessel frequency	Ray frequency	Ray width (µm)	Ray hight (µm)
Group 5	*Cedrela fissilis*	5	191 ± 30^{a}	2.9 ± 0.2^{b}	5.2 ± 0.2^{b}	33 ± 9^{a}	324 ± 51^{b}
	Carapa guianensis	6	155 ± 10^{b}	6.8 ± 1.3^{a}	5.8 ± 0.6^{b}	41 ± 13^{a}	537 ± 84^{a}
	Carapa procera	5	96 ± 22^{c}	8.3 ± 3.4^{a}	7.1 ± 1.2^{a}	48 ± 7^{a}	566 ± 52^{a}
Group 6	*Swietenia macrophylla*	6	159 ± 5^{a}	8.3 ± 1.1^{bc}	5.1 ± 0.6^{bc}	47 ± 8^{a}	383 ± 29^{b}
	Swietenia mahagoni	6	117 ± 13^{b}	12.4 ± 1.0^{a}	5.6 ± 0.2^{bc}	56 ± 8^{a}	308 ± 20^{b}
	Swietenia humilis	6	152 ± 12^{a}	10.1 ± 2.8^{ab}	4.7 ± 0.2^{c}	52 ± 10^{a}	375 ± 26^{b}
	Carapa guianensis	6	155 ± 10^{a}	6.8 ± 1.3^{c}	5.8 ± 0.6^{b}	41 ± 13^{a}	537 ± 84^{a}
	Guarea cedrata	5	133 ± 8^{b}	12.9 ± 2.9^{a}	$7.\ 8 \pm 1.2^{a}$	25 ± 3^{b}	325 ± 33^{b}

For each group, values with different superscripts (a, b and c) are statistically different. n is the number of specimens per taxon with the t-test results in Table 3. In Group 2, three-feature (VD, RH, and RP) and five-feature combinations achieved the highest identification accuracy (95.83%). The results of Group 4 indicated that two or more features with VD, RH, or RW included showed higher accuracy (87.50%) than the others. Interestingly, two- to five-feature combinations showed equivalent accuracy (94.12%) when classifying the species in Group 5.

These results indicated that the rule of "the more features, the higher the accuracy" in traditional wood identification is not always applicable to wood identification using QWA data, but of course this depends on the type of QWA features for measurement and analysis. We assume that the QWA data of some wood anatomical features are frequently correlated (Piermattei et al. 2020), and the correlations between wood anatomical data call for future analysis, especially given that some of the methods employed in this study (e.g., Multinomial Naïve Bayes classifier) explicitly assume independence between characters. It is noteworthy that the data of only five wood anatomical features were collected in this study. When the QWA data of more features from a large number of microscope slides are collected and the QWA dataset is scaled up, it is possible that combining fewer features yields equal or higher accuracy with significant time- and labor-saving for wood identification using QWA data, although those combinations of the most effective features are likely to differ as the class composition for the models changes.

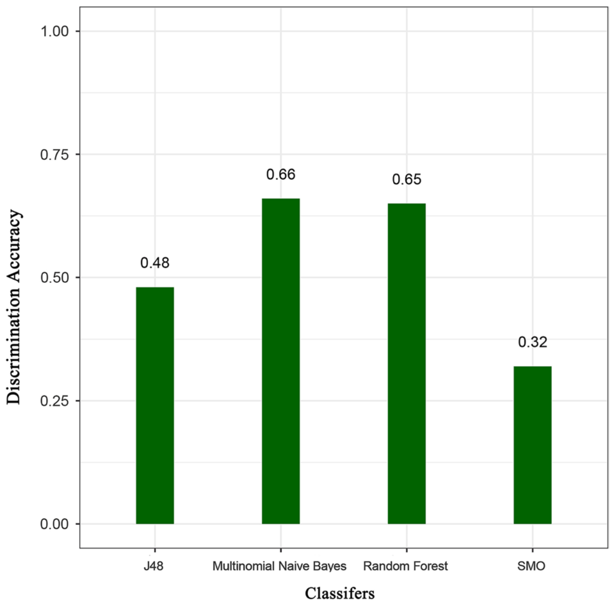

Fig. 2 Discrimination accuracy of four classifiers when seperating 29 species based on five features

Discrimination of CITES-listed species from their look-alikes

Figure 4 shows the confusion matrices reporting the highest performing supervised classifiers for the six groups. The predicted species are in the columns, the true species are on the left, and the correct predictions are on the diagonal. All specimens of *P. tinctorius*, *D. stevensonii*, *D. retusa*, *C. odorata*, and *C. fissilis* are discriminated from their look-alikes by Multinomial Naïve Bayes, and all *S. mahagoni* specimens could be successfully classified using Random Forest. In Group 1, 33% ($n=3$) of the specimens from *P. erinaceus* were misidentified as *P. macrocarpus*. In Group 2, although 25% of the specimens ($n=4$) of *P. santalinus* were misidentified as *P. soyauxii*, the other species were correctly predicted. In Group 3–5, all CITES-listed species were successfully identified from similar species that were non-CITES. Moreover, in Group 4, *C. odorata* and *G. cedrata* are all correctly classified, however, *Khaya anthotheca* and *Entandrophragma cylindricum* were mispredicted as each other. In Group 6, *S. mahagoni* could be discriminated from their look-alikes. He et al. (2020b) used QWA data coupled with ML analysis to discriminate three *Swietenia* species, specifically, between *S. mahagoni* and *S. macrophylla*. However, these three species are all CITES-listed species, and discrimination between them using QWA data did not provide a relevant evidence for combating illegal trade of CITES-listed species, but for cultural, commercial, and scientific interests. As within-genus intra-specific variation is typically as large as inter-specific variation, species-level resolution is usually impossible in traditional wood identification. Generally speaking, genus-level identification is achieved by experienced anatomists (Gasson et al. 2011), but this can be affected by subjective factors. If QWA data could be collected accurately and efficiently and fed into ML models for species

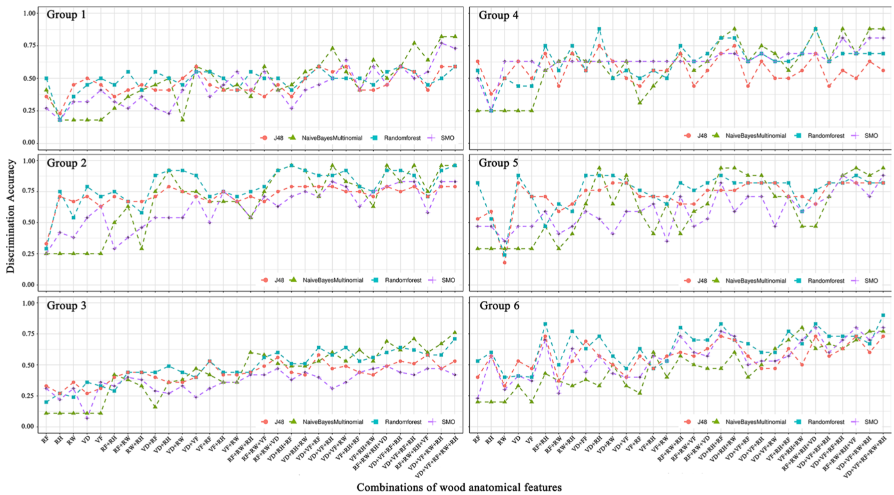

Fig. 3 Accuracy of the best-performing classifiers to discriminate CITES-listed species from their look-alikes using different combinations of wood anatomical features for Group 1–6. VD:Vessel Diameter; VF: Vessel Frequency; RW: Ray Width; RH:Ray Height; RF: Ray Frequency

identification, it has the potential to reduce human subjectivity in the species decision-making stage of wood identification.

When discriminating CITES-listed *Dalbergia* species from their look-alikes, the best-performing ML classifier exhibited the accuracy of 90.56%, performing much better than experienced wood anatomists (He et al. 2020a). The six *Dalbergia* species selected are commercially important in international trade and are often mislabeled as non-CITES species in documents submitted to customs officials (Saunders and Reeve 2014). The Multinomial Naïve Bayes classifier achieved the highest accuracy (95.83%) when separating CITES-listed *P. santalinus* and *P. tinctorius* from their look-alikes. These results demonstrate that QWA data can be used to discriminate CITES-listed species from look-alikes within this group with quantitative evidence at a high certainty, providing a laboratory-based forensically relevant method.

Applicability of QWA for forensic wood identification

Traditional wood identification methods typically only reach genus-level identification and require wood anatomists with decades of experience, who are nonetheless prone to inevitable subjectivity. Moreover, the scarcity of these professionals remains a challenge in large-scale scenarios such as the battle against illegal logging and associated trade (Johnson and Laestadius 2011; Wiedenhoeft et al. 2019). Compared with traditional methods, the QWA data in this study is measured from only two sections (transverse and tangential) without the need to make slides from all three sections of wood, which increases the efficiency and reduces the necessary training for human-mediated data collection.

The QWA data of commercially important CITES-listed species in this study were collected from scientific wood collections, which serve as potential repositories of reference data for forensic wood identification. At present, there are more than 180 scientific xylaria in the world, housing a total of more than 1.5 million wood specimens and a large number of microscope slides (IAWA 2016; ITTO 2016). The use of QWA data for forensic wood identification helps to showcase the tremendous value of wood slides “sleeping” in wood collections around the world. The digitization of microscope slides shows the potential to provide valuable QWA data for statistical analysis (Wheeler 2011), but only if the spatial scale and specimen metadata regarding vouchers, geographic origin, etc. are available to validate the images. With the development of artificial intelligence, ML approaches are widely used in classification tasks (Arganda-Carreras et al. 2017; Frank et al. 2009; Sebastiani 2002), including wood identification (He et al. 2019, 2020a, 2020b; Ravindran et al. 2018, 2020). Some field-deployable applications based on macroscopic images have been developed to discriminate CITES-listed species, such as XyloTron (Ravindran et al. 2018, 2019, 2021). However, microscopic images contain more abundant and detailed structural features information than macroscopic images, making QWA + ML approaches in the laboratory more promising for species-level wood identification based on wood anatomical features. QWA data could enhance such efforts by measuring wood cell geometric characters and subjecting them to ML models for species classification. Not

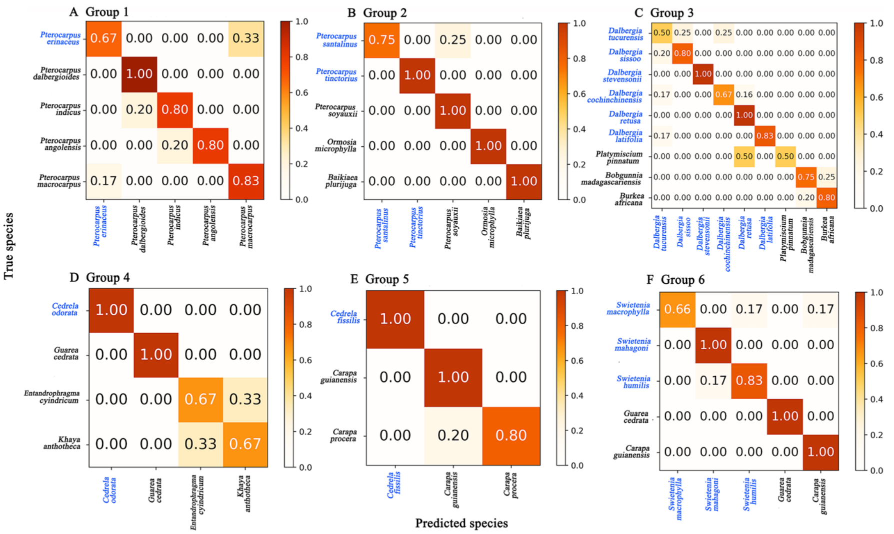

Fig. 4 Confusion matrices generated by the best-performing classification predictions of the best classifiers. **A** - **E**: Multinomial Naïve Bayes; **F** : Random Forest. Species in blue text are CITES-listed

surprisingly, compared to becoming an expert in traditional wood identification, it is easier to measure selected QWA characters following standard wood anatomical protocols and to input the QWA data to user-friendly software, though learning to prepare specimens and collect such data is not a trivial matter. The greatest advantage of QWA approaches for forensic wood identification is that the analysis and classification are done non-subjectively by the ML algorithm rather than by a human.

Conclusion

Wood identification is usually achieved by traditional methods requiring qualitative wood anatomy and is prone to subjectivity. This study used QWA data to discriminate CITES-listed species from their look-alikes through ML analysis. The results demonstrated that QWA data coupled with ML analysis is superior to traditional wood identification methods at the species level when discriminating CITES-listed species from their look-alikes. We also concluded that the types and combinations of wood anatomical features had effects on discrimination accuracy when using QWA data for wood identification. This study demonstrated the feasibility of QWA data coupled with ML analysis to discriminate CITES-listed species from their look-alikes for forensic wood identification, thus providing a practical tool for wood anatomy laboratories worldwide to assist in efforts to combat illegal logging and associated trade.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00226-022-01404-y.

Acknowledgements This work was supported by the Project of Natural Science Foundation of Beijing (6224064) and the Project of Chinese Academy of Forestry (CAFYBB2021ZD002). We would like to express our gratitude to Dr. Lichao Jiao, Mr. Jie Wang, Mr. Yang Lu and Miss. Qiulu Meng for their help in participating in the discussion.

Funding Project of Natural Science Foundation of Beijing, 6224064, Tuo He, Project of Chinese Academy of Forestry, CAFYBB2021ZD002, Yafang Yin

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, Seung HS (2017) Trainable weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 33:2424–2426. https://doi.org/10.1093/bioinformatics/btx180

Beeckman H (2016) Wood anatomy and trait-based ecology. IAWA J 37:127–151. https://doi.org/10.1163/22941932-20160127

Brancalion PHS, Almeida DRA, Vidal E, Molin PG, Sontag VE, Souza S, Schulze MD (2018) Fake legal logging in the Brazilian amazon. Sci Adv. https://doi.org/10.1126/sciadv.aat1192

Chen W, Zhang S, Li RW, Shahabi H (2018) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naive bayes tree for landslide susceptibility modeling. Sci Total Environ 644:1006–1018. https://doi.org/10.1016/j.scitotenv.2018.06.389

CITES (2019) Appendices I, II and III. Retrieved from https://www.cites.org/eng/app/appendices.php. Accessed 26 November 2019

Cornelius JP, Wightman KE, Grogan JE, Ward SE (2004) TROPICAL ECOSYSTEMS | Swietenia (American Mahogany). Encyclopedia of Forest Sciences. Elsevier, pp 1720–1726. https://doi.org/10.1016/B0-12-145160-7/00190-3

Da Silva JAT, Kher MM, Soner D, Nataraj M (2019) Red sandalwood (*Pterocarpus santalinus* L. f.): biology, importance, propagation and micropropagation. J For Res 30:745–754. https://doi.org/10.1007/s11676-018-0714-6

Dumenu WK (2019) Assessing the impact of felling export ban and CITES designation on exploitation of African rosewood (*Pterocarpus erinaceus*). Biol Cons 236:124–133. https://doi.org/10.1016/j.biocon.2019.05.044

Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L (2009) Weka-A machine learning workbench for data mining. Springer, Boston, In Data mining and knowledge discovery handbook. https://doi.org/10.1007/978-0-387-09823-4_66

Gasson P, Miller R, Stekel DJ, Whinder F, Zieminska K (2010) Wood identification of *Dalbergia nigra* (CITES Appendix I) using quantitative wood anatomy, principal components analysis and naive bayes classification. Ann Bot 105:45–56. https://doi.org/10.1093/aob/mcp270

Gasson P, Baas P, Wheeler E (2011) Wood anatomy of CITES-listed tree species. IAWA J 32:155–198. https://doi.org/10.1163/22941932-90000050

Goldsmith E (1978) The convention on international trade in endangered species of wild fauna and flora. J Med Primatol 7:122–124. https://doi.org/10.1159/0004597967

Gulbrandsen LH, Humphreys D (2006) International initiatives to address tropical timber logging and trade. A Report for the Norwegian Ministry of the Environment. FNI Report.

Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I (2015) The use of DNA barcoding in identification and conservation of rosewood (Dalbergia spp). PLoS One 10:e0138231. https://doi.org/10.1371/journal.pone.0138231

He T, Jiao LC, Yu M, Guo J, Jiang XM, Yin YF (2019) DNA barcoding authentication for the wood of eight endangered *Dalbergia* timber species using machine learning approaches. Holzforschung 73:277–285. https://doi.org/10.1515/hf-2018-0076

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

He T, Marco J, Soares R, Yin YF, Wiedenhoeft AC (2020b) Machine learning models with quantitative wood anatomy data can discriminate between *Swietenia macrophylla* and *Swietenia mahagoni*. Forests 11:36–49. https://doi.org/10.3390/f11010036

He T, Jiao LC, Guo J, Yin YF (2021) Wood Informatics: History of Development, Application, and Prospective Trend. Chinese J Wood Sci Technol 35(4):15–24

Hornik K, Buchta C, Zeileis A (2009) Open-source machine learning: R meets Weka. Comput Statistics 24:225–232. https://doi.org/10.1007/s00180-008-0119-7

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

IAWA (2016) Index Xylariorum 4.1. https://www.iawa-website.org/uploads/soft/Abstracts/Index%20Xylariorum%204.1.pdf.

INTERPOL (2021) Forestry Crime Fact sheet 2021. https://www.interpol.int/content/download/17367/file/Forestry%20Crime%20Fact%20sheet%202021.pdf.

ITTO (2016) Biennial review and assessment of the world timber situation. Yokohama, Japan.

Jiang L, Cai Z, Zhang H, Wang D (2013) Naive bayes text classifiers: a locally weighted learning approach. J Exp Theor Artif Intel 25:273–286. https://doi.org/10.1080/0952813X.2012.721010

Jiao LC, Lu Y, He T, Guo J, Yin YF (2020) DNA barcoding for wood identification: global review of the last decade and future perspective. IAWA J 41:620–643. https://doi.org/10.1163/22941932-bja10041

Johnson A, Laestadius L (2011) New laws, new needs: The role of wood science in global policy efforts to reduce illegal logging and associated trade. IAWA J 32:125–136. https://doi.org/10.1163/22941932-90000048

Keerthi SS, Gilbert EG (2002) Convergence of a generalized SMO algorithm for SVM classifier design. Mach Learn 46:351–360. https://doi.org/10.1023/A:1012431217818

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2:18–22

Lowe AJ, Dormontt EE, Bowie MJ, Degen B, Gardner S, Thomas D, Clarke C, Rimbawanto A, Wiedenhoeft AC, Yin YF, Sasaki N (2016) Opportunities for improved transparency in the timber trade through scientific verification. Bioscience 66(11):990–998. https://doi.org/10.1093/biosci/biw129

Pennington TD, Muellner AN, Wise R (2010) A monograph of *Cedrela* (Meliaceae). DH Books, Sherborne, UK

Piermattei A, Von Arx G, Avanzi C, Fonti P, Gärtner H, Piotti A, Urbinati C, Vendramin GG, Büntgen U, Crivellaro A (2020) Functional relationships of wood anatomical traits in *Norway Spruce*. Front Plant Sci 11:683–698. https://doi.org/10.3389/fpls.2020.00683

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

Ravindran P, Thompson BJ, Soares RK, Wiedenhoeft AC (2020) The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. Front Plant Sci 11:1015. https://doi.org/10.3389/fpls.2020.01015

Ravindran P, Owens FC, Wade AC, Vega P, Muro RM, Shmulsky R, Wiedenhoeft AC (2021) Field-deployable computer vision wood identification of Peruvian timbers. Front Plant Sci 12:940. https://doi.org/10.3389/fpls.2021.647515

Ravindran P, Ebanyenle E, Ebeheakey AA, Abban KB, Lambog O, Soares R, Costa A, Wiedenhoeft AC (2019) Image based identification of Ghanaian timbers using the XyloTron: Opportunities, risks and challenges. In: NeurIPS workshop on machine learning for the developing world (ML4D): challenges and risks.

Refaeilzadeh P, Tang L, Liu H (2009) Cros-Valid Encycl Database Syst 5:532–538. https://doi.org/10.1007/978-1-4899-7993-3_565-2

Richardson VA, Peres CA (2016) Temporal decay in timber species composition and value in Amazonian logging concessions. PLoS ONE 11:e0159035. https://doi.org/10.1371/journal.pone.0159035

Richter HG, Gembruch K, Koch G (2014) CITESwoodID: descriptions, illustrations, identification, and informationretrieval. In English, French, German, and Spanish. Version: 19th August 2019. delta-intkey.com

Rodan BD, Newton AC, Veríssimo A (1992) Mahogany conservation: status and policy initiatives. Environ Conserv 19:331–338. https://doi.org/10.1017/S0376892900031453
Ruffinatto F, Crivellaro A (2019) Atlas of macroscopic wood identification: with a special focus on timbers used in Europe and CITES-listed species. Springer Nature, Switzerland AG. https://doi.org/10.1007/978-3-030-23566-6
Saunders J, Reeve R (2014) The EU timber regulation and CITES. Energy, Environment Resources PP EER, 8.
Schindelin J, Rueden CT, Hiner MC, Eliceiri KW (2015) The ImageJ ecosystem: An open platform for biomedical image analysis. Mol Reprod Dev 82:518–529. https://doi.org/10.1002/mrd.22489
Sebastiani F (2002) Machine learning in automated text categorization. ACM Comput Surv 34:1–47. https://doi.org/10.1145/505282.505283
Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Networks 11:1188–1193. https://doi.org/10.1109/72.870050
Smith TC, Frank E (2016) Introducing machine learning concepts with WEKA. In: Mathé Ewy, Davis Sean (eds) Statistical Genomics. Springer New York, New York, NY, pp 353–378. https://doi.org/10.1007/978-1-4939-3578-9_17
Von Arx G, Crivellaro A, Prendin AL, Cufar K, Carrer M (2016) Quantitative wood anatomy-practical guidelines. Front Plant Sci 7:781. https://doi.org/10.3389/fpls.2016.00781
Wheeler E (2011) Inside Wood–A web resource for hardwood anatomy. IAWA J 32:199–211. https://doi.org/10.1163/22941932-90000051
Wheeler EA, Baas P (1998) Wood identification - A review. IAWA J 19:241–264. https://doi.org/10.1163/22941932-90001528
Wick E (2019) Impacts of the Convention on International Trade in Endangered species of Wild Fauna and Flora (CITES) on various stakeholders in the music industry. America.
Wiedenhoeft AC, Miller RB (2005) Structure and function of wood. In: Rowell RM (ed) Handbook of wood chemistry and wood composites. CRC Press, Florida, pp 9–33
Wiedenhoeft AC, Simeone J, Smith A, Parker-Forney M, Soares R, Fishman A (2019) Fraud and misrepresentation in retail forest products exceeds US forensic wood science capacity. PLoS ONE 14:e0219917. https://doi.org/10.1371/journal.pone.0219917
Yin Y, Jiang X, Yuan L (2016) Identification manual of endangered and precious timber species common in trades. Science Press, Beijing
Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413–418. https://doi.org/10.1016/0169-5347(96)10045-8

Authors and Affiliations

Shoujia Liu[1,2] · Tuo He[1,2] · Jiajun Wang[1] · Jiabao Chen[1,2] · Juan Guo[1,2] · Xiaomei Jiang[1,2] · Alex C. Wiedenhoeft[3,4,5,6,7] · Yafang Yin[1,2]

1 Department of Wood Anatomy and Utilization, Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing 100091, China

2 Wood Collections, Chinese Academy of Forestry, Beijing 100091, China

3 Center for Wood Anatomy Research, Forest Products Laboratory, USDA Forest Service, Madison, WI 53726, USA

[4] Department of Botany, University of Wisconsin, Madison, WI 53706, USA

[5] Department of Forestry and National Resources, Purdue University, West Lafayette, IN 47907, USA

[6] Ciências Biológicas (Botânica), Universidade Estadual Paulista, Botucatu, São Paulo 18610-034, Brazil

[7] Department of Sustainable Biomaterials, Mississippi State University, Starkville, MS 39759, USA