



Identification of Congolese wood species by use of anatomical images and artificial intelligence

Ruben De Blaere



ISBN: 9789463579025

University Press will
generate your bar code here

Identification of Congolese wood species by use of anatomical images and artificial intelligence

Ir. Ruben De Blaere

Thesis submitted in fulfillment of the requirements for the degree
of Doctor (Ph. D.) of Bioscience engineering: natural resources

Promotors

Prof. dr. ir. Jan Van den Bulcke
UGent-Woodlab, Laboratory of Wood Technology
Department of Environment
Faculty of Bioscience Engineering
Ghent University (Belgium)

dr. ir. Hans Beeckman
Service of Wood Biology
Royal Museum for Central-Africa (Belgium)

Prof. dr. ir. Jan Verwaeren
KERMIT, Research Unit Knowledge-based Systems Mathematical and Computational Modelling in the Information Age
Department of Data analysis and mathematical modelling
Faculty of Bioscience Engineering
Ghent University (Belgium)

Prof. dr. ir. Wannes Hubau
UGent-Woodlab, Laboratory of Wood Technology
Department of Environment
Faculty of Bioscience Engineering
Ghent University (Belgium)
Service of Wood Biology
Royal Museum for Central-Africa (Belgium)

Examination-committee

Prof. dr. ir. Andre Skirtach, *Chair*
BIOMARKED
Department of Biotechnology
Faculty of Bioscience Engineering
Ghent University (Belgium)

Prof. dr. ir. Pieter De Frenne, *Secretary*
Forest & Nature Lab
Department of Environment
Ghent University (Belgium)

Prof. dr. ir. Jan Aelterman
IMEC, Centre for X-ray Tomography, i-KNOW
Department of Telecommunications and information processing
Faculty of Engineering and Architecture
Ghent University (Belgium)

dr. ir. Victor Deklerck
Department Research
Meise Botanic Garden (Belgium)

Dr. rer. nat. Andrea Olbrich
Institute of Wood Research
Johann Heinrich von Thünen Institute: Federal Research Institute for Rural Areas, Forestry and Fisheries (Germany)

Dean

Prof. dr. ir. Els Van Damme

Rector

Dr. Petra De Sutter

Woord vooraf

If you want to go fast, go alone. If you want to go far, go together.

African Wisdom

Dat gezegde is meer dan waar, en misschien nog sterker bij het schrijven van een doctoraat. Dit werk is niet enkel het resultaat van mijn inspanningen, maar van de steun en warmte die ik de afgelopen vier jaar heb mogen ervaren. Familie, vrienden, collega's, mijn promotoren, en vooral die ene bijzonder persoon die altijd aan mijn zijde stond, in dit verhaal én in mijn leven. Jullie maakten dit mogelijk.

Graag wil ik mijn collega's in het **UGent-Woodlab** bedanken. Jullie gaven me niet alleen een luisterend oor, maar ook energie, plezier, en inzicht. Brainstormen met jullie leidde vaak tot de doorbraken waar dit werk op steunt. Bedankt voor alle lunches, gesprekken en momenten van inspiratie. In het bijzonder: **Stijn Willen**, het was een voorrecht je bureaupartner te zijn; jouw technische inzicht en goedgezindheid waren goud waard. Ik zal onze precisiewerkjes, zoals het zagen van de beukenstam, nooit vergeten. **Toon Gheyle**, bedankt voor je enthousiasme dat het werk altijd lichter maakte. **Maxime Dekegeleer**, onze gesprekken over AI gaven me inzichten die dit onderzoek sterker hebben gemaakt. **Maarten Perneel**, bedankt voor je tijd, behulpzaamheid, en me een beetje extra wegwijs te maken in neurale netwerken en computer visie. **Liselotte De Ligne**, jouw begeleiding en huiselijkheid bij mijn masterproef waren fundamenteen waar ik op kon bouwen. Ik ben zeker dat anderen dit ook ervaren, en ik twijfel er geen seconde aan dat je die mooie waarden ook mee zal geven aan je kinderen. **Joris Van Acker**, jij hebt mijn liefde voor hout verder aangewakkerd en gevoed – jouw lessen, begeleiding en steun hebben mij mede gevormd en hebben me op een pad gestuurd dat ik met plezier heb mogen bewandelen. Ook wil ik graag **Katrien** en **Kasper** bedanken: jullie vriendschap en onze gezellige lunches als alumni waren telkens iets om naar uit te kijken.

Ook alle jobstudenten en stagiaires met wie ik samen heb mogen werken wil ik graag bedanken. Naast Michael Monnoye wil ik ook **Miro Cnops**, **Senne Suyckerbuyck**, en **Tibo Deckers** bedanken voor al hun hulp bij het annoteren. **Simon Vansuyt** ook jou wil ik bedanken voor de inzet en kunde waarmee je een CNC machine kon ombouwen naar een indrukwekkend robot voor hoogwaardige beeldname. Ik wens je veel succes in je toekomstige carrière.

In het **Africamuseum** wil ik graag al mijn collega's doorheen de tijd bedanken voor de fijne collegiale momenten. Ook in het Africamuseum zijn er enkele personen die ik graag extra in de verf wil zetten. Te beginnen met een persoon die het meeste praktisch werk verzet heeft voor mijn doctoraat. **Daniel Wallenus**, jouw tomeloze inzet voor SmartWoodID en de houtcollectie, je vriendelijkheid en eindeloze uren densiteitsmetingen, foto's, en scanning zijn van onschatbare waarde. Verder wil ik ook in het bijzonder **Michael Monnoye** bedanken. We hebben elkaar leren kennen als PhD student en jobstudent, maar daaruit is uiteindelijk een collegiaal gegeven mogen groeien. Je inzet voor het annoteren van de houtkenmerken en je verdere ontwikkeling als jonge ervaren houtidentificatie expert, zijn indrukwekkend en ik hoop dat we in de toekomst nog veel samen mogen werken om samen houtidentificatie nog efficiënter, accurater en toegankelijker te mogen maken voor het bestrijden van illegale houthandel. **Kévin Lievens**, onze brainstormen waren onvergetelijk. Ze waren een genot en hebben me veel bijgebracht. Ik wens je veel succes in je nieuwe uitdaging. **Eric Van Herreweghe**, al was onze tijd als collega's kort, ik kan terugkijken op je inzichten en onze carpoolgesprekken als waardevolle lessen. Ze hebben de fundamenteen gevormd voor automatisering en nieuwe onderzoeksmogelijkheden. Ook de technische ondersteuning wil ik graag bedanken met in het bijzonder: **Dieter Van Hassel**, bedankt voor je kennis en inzet rond databanken en het online beschikbaar maken van SmartWoodID. **Franck Theeten**, voor je creatieve ideeën en hulp bij deep learning. **Kristof Bollen**, voor je snelle en vriendelijke technische ondersteuning die altijd klaar stond. Bovendien wil ik ook graag de directie bedanken.

Naast mijn collega's in wil ik ook een bijzonder groepje mensen bedanken. Mensen die zich over de jaren ingezet hebben voor wetenschappelijk werk vanuit hun eigen initiatief. Ik wil graag alle vrijwilligers bedanken die me geholpen hebben bij het digitaliseren en managen van de SmartWoodID collectie en bij uitbreiding de Tervuren houtcollectie. **Cécile De Troyer**, **Frank Simoens**, **Michèle Florquin**, **Ferre Caron**, **Olayemi Razaq Salu**, **Véronique Maes**, **Philippe Quintin**, **Patrick De Snijder**, en **José Kempenaers**. Verder wil ik graag Luiza Mittrache bedanken voor de *Citizen science* initiatieven die dit mogelijk maken.

Verder wil ik ook graag iets groters bedanken dan enkele bijzonder individuen. Ik wil alle voorgaande wetenschappers bedanken, en in het bijzonder allen die hebben bijgedragen aan de twee instituten waarin ik zo hartelijk ben verwelkomd. Mijn onderzoek zou nooit mogelijk zijn geweest zonder jullie steun en inspanningen. In essentie:

I stand on the shoulders of giants.

Bernard of Chartres & sir Isaac Newton

En de grootste *giant* is daarbij de Tervuren houtcollectie. Deze impressionante collectie is de basis van mijn onderzoek, wat niet mogelijk geweest zou zijn zonder al wetenschappelijk werk dat mijn voorgangers gedaan hebben in de afgelopen eeuw. Jullie, en alle onderzoekers uit het verleden wil ik bedanken, opdat ik op jullie kennis verder mocht bouwen en dat anderen na mij dat ook mogen.

Extern wil ik ook graag **Bart Peeters** en **Filip Mollé** bedanken voor hun inzicht en vriendelijkheid om samen de automatisering uit te werken, die befaamde robotarm aan te kopen, ondanks de uitdagende administratieve procedures aan de overheid. Ook **Kristof Haneca**, **Tom De Mil**, **Steven Janssens**, **Alex Dedeckel** en vele anderen wil ik bedanken om jullie te hebben mogen leren kennen tijdens mijn doctoraat en hopelijk nog beter in de toekomst.

Ik wil graag de jury leden bedanken voor hun zorgvuldige en inzichtvolle feedback. Daarbij wil ik graag een speciaal woord van dank brengen aan **Victor De Klerck**. Uw ervaring, warmte en geloof in mij waren een bron van kracht. Uw woorden – dat inzet en gedrevenheid minstens even belangrijk zijn als punten – hebben me gedragen. U bent een voorbeeld en een inspiratie. Ik ben dankbaar dat ik met u heb mogen samenwerken en hoop dat dit in de toekomst nog vaak het geval zal zijn.

Naast al die fijne collega's wil ik nu graag in het bijzonder mijn promotoren bedanken. Elk van jullie hebben me waardevolle lessen geleerd waardoor ik niet alleen als wetenschapper, maar ook als mens heb kunnen groeien.

Ik wil **Jan Van den Bulcke** bedanken. Ik heb veel ontzag voor de manier waarop u wetenschap beoefent. Uw efficiëntie, oog voor detail en vermogen om steeds meteen tot de essentie te komen, hebben me enorm geholpen om beter te leren schrijven en informatie helder over te brengen. Ik heb genoten van onze samenwerking en hoop oprecht dat die in de toekomst verdergezet kan worden.

Jan Verwaeren, ook u wil ik graag bedanken voor de vele aangename en inzichtvolle gesprekken. Ze hebben me veel geleerd over data-analyse en over de manier waarop ik naar onderzoek kan kijken. Uw toewijding om mij steeds bij te staan met antwoorden over deep learning was cruciaal om dit doctoraat tot een goed einde te brengen. Uw vriendelijkheid en geduld heb ik bijzonder geapprecieerd.

Wannes Hubau, ik ben u dankbaar voor alle ondersteuning, maar in het bijzonder voor uw inzet met het oog op de toekomst. De tijd en moeite die u hebt geïnvesteerd in het voorzien van sponsoring en het samen uitschrijven van nieuwe projecten betekenen veel voor mij en mijn verdere carrière. Ik ben blij dat u me kansen biedt om hopelijk mijn onderzoek voort te zetten en dat u mijn ideeën ziet als waardevolle pistes voor toekomstig werk. Mijn dank is nu groot, maar ik weet dat die in de toekomst alleen maar verder zal groeien.

Ten slotte wil ik een zeer bijzondere promotor bedanken: **Hans Beeckman**. U hebt mij iets gegeven dat ik nooit kan teruggeven: een kans. Zonder u had ik dit avontuur nooit kunnen aangaan. U schonk een onbekende, pas afgestudeerde bio-ingenieur, die u nog niet kende, de mogelijkheid om in te stappen in een project dat u nauw aan het hart lag. Dat vertrouwen betekende alles voor mij. Het leerde me hoe belangrijk het is om jonge onderzoekers kansen te geven. De tijd die u vrijmaakte voor doelgerichte, pragmatische ondersteuning heeft me doen groeien als wetenschapper en als persoon. Ook nu u met pensioen bent, weet ik dat ik nog steeds op u kan rekenen. Ik hoop oprecht dat ik u nog vele jaren mag kennen en dat we samen nog veel kunnen realiseren—tot we de volledige Tervuren houtcollectie hebben gedigitaliseerd en alle houtsoorten ter wereld in kaart hebben gebracht.

Ten slotte wil ik ook al mijn **vrienden** en **familie** bedanken voor hun onvoorwaardelijke betrokkenheid bij mij en mijn onderzoek. Jullie oprechte interesse en enthousiasme waren voor mij een baken van motivatie en bevestiging dat mijn werk er werkelijk toe doet. Soms zag ik de ogen wel eens wegdromen wanneer ik weer te enthousiast over mijn onderzoek vertelde, maar toch bleven jullie vragen naar mijn werk. Dat vertrouwen en die warmte zijn misschien wel het mooiste geschenk dat jullie mij konden geven.

In het bijzonder wil ik mijn ouders bedanken. Mama en Papa, jullie maakten het mogelijk dat ik kon studeren en hebben mij altijd gesteund in mijn keuzes. Jullie geloof in mij heeft de basis gelegd voor alles wat ik vandaag bereikt heb. Ook mijn grootmoeder, Marie 'Mammie' Wynant, wil ik hier met veel dankbaarheid vernoemen. Dankzij jou leerde ik Hans kennen, wat de deur geopend heeft naar deze vier fantastische jaren.

En bovenal wil ik mijn vrouw, **Kim**, bedanken. Laat niemand ooit zeggen dat een doctoraat geen druk zet op een relatie; vooral dit laatste jaar was zwaar tijdens de afwerking van dit manuscript. Liefje, zonder jouw steun had ik dit nooit kunnen volbrengen. Terwijl jij zelf aan je eigen doctoraat begon en daarnaast ook nog het huishouden grotendeels op je nam, bleef je me aanmoedigen en kracht geven om door te zetten (soms lief en warm, maar altijd recht voor de raap en ongefilterd). Zelfs wanneer ik gefrustreerd mijn bureau kapot kon timmeren op het programmeren, of eens zat te klagen over mijn werk, bleef jij geduldig en begripvol. Ik weet dat dit niet altijd eenvoudig was. Voor al die steun, voor al dat geduld en voor de liefde waarmee je me door deze storm hielp, ben ik je eindeloos dankbaar.

List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AUC	Area Under the Curve
BCE	Binary Cross Entropy
BELSPO	Belgian Science Policy Office
CCE	Categorical Cross Entropy
CDF	cumulative distribution function
CITES	Convention on International Trade in Endangered Species
CNN	Convolutional Neural Networks
CROSR	Classification-Reconstruction learning for Open-Set Recognition
CT	Computed Tomography
CV	Computer Vision
DART	Direct Analysis in Real Time
DRC	Democratic Republic of the Congo
DT	Decision Tree
EUDR	European Union Deforestation Regulation
EUTR	European Union Timber Regulation
FLEGT	Forest Law Enforcement, Governance and Trade
Grad-CAM	Gradient Weighted Class Activation Mapping
IAWA	International Association of Wood Anatomists
IIIF	International Image Interoperability Framework
IUCN	International Union for Conservation of Nature
IUFRO	International Union of Forest Research Organizations
NIRS	Near-Infrared Spectroscopy
ReLU	Rectified Linear Unit
RF	Random Forests
RGB	Red Green Blue
RMCA	Royal Museum for Central-Africa
SVM	Support Vector Machines
TOFMS	Time-Of-Flight Mass Spectrometry

Summary (Dutch)

Illegale houtkap vormt een ernstige bedreiging voor bossen en kan onherstelbare schade met zich meebrengen, zeker wanneer beschermde boomsoorten worden geëxploiteerd. Naar schatting is tussen de 30 en 90 procent van het verhandelde tropisch hout illegaal gekapt. Om dit tegen te gaan zijn verschillende regelgevingen en beleidsmaatregelen ingevoerd, gericht op meer transparantie en traceerbaarheid binnen de houtketen. Voorbeelden hiervan zijn het FLEGT-actieplan, de Europese Houtverordening (EUTR), de Europese Ontbossingsverordening (EUDR), de Amerikaanse *Lacey Act* en het CITES-verdrag. Het succes van deze maatregelen hangt in grote mate af van één cruciale voorwaarde: de snelle en betrouwbare identificatie van houtsoorten, van het moment van kap tot aan transport, opslag en uitvoer.

Houtidentificatie gebeurt a.d.h.v. anatomische, chemische en genetische technieken. Van deze methoden is de anatomische analyse – gebaseerd op de structuur van houtcellen en -weefsels – het meest gebruikelijk. De IAWA heeft gestandaardiseerde kenmerken opgesteld die wereldwijd als referentie gelden. Microscopische analyse levert een hoge nauwkeurigheid op, maar vereist gespecialiseerde apparatuur, training en tijdrovende voorbereiding van stalen. Daardoor is deze methode minder geschikt voor screening in het veld. Macroscopische analyse, waarbij gekeken wordt naar kenmerken die met het blote oog of een loep zichtbaar zijn op vers gezaagde of geschuurde oppervlakken, is veel toegankelijker en goedkoper. Vooral een dwarsdoorsnede van het hout toont belangrijke kenmerken zoals de ligging van vaten, breedte van stralen en verdeling van parenchym – elementen die vaak worden gebruikt in veldgidsen.

Toch kent deze methode duidelijke beperkingen. Het aantal waarneembare kenmerken is beperkt, en er bestaat vaak grote variatie binnen dezelfde soort. Dat maakt het moeilijk om met zekerheid een soort te benoemen. De IAWA-kenmerken zijn bovendien beschrijvend van aard (“aanwezig”, “variabel”, “afwezig”), wat subtiele maar relevante verschillen kan verhullen en interpretatie subjectief maakt. Bovendien is er tot op het heden weinig onderzoek uitgevoerd naar hoe goed deze kenmerken werkelijk onderscheid tussen houtsoorten mogelijk maken, vooral in soortenrijke tropische gebieden. Daardoor blijven betrouwbare identificatiesleutels en de ontwikkeling van nieuwe herkenningssystemen achter.

Beeldherkenning via computer visie biedt hierin een veelbelovend alternatief. CNNs kunnen zelfstandig visuele kenmerken herkennen op foto's van houtdoorsneden en zo automatisch en snel soorten identificeren. Deze technologie werkt op draagbare apparaten en wordt al toegepast in de praktijk, bijvoorbeeld met het XyloTron-systeem in Ghana. Toch kent ook deze aanpak uitdagingen. De trainingsdata bestaat vaak uit perfecte stalen, terwijl hout in de praktijk beschadigd kan zijn door bijvoorbeeld scheuren, verkleuring of aantasting door insecten of schimmels. Zulke schade kan kenmerken verbergen en de nauwkeurigheid van modellen aantasten. Recente studies (zoals Ravindran et al., 2023 en Owens et al., 2024) hebben aangetoond dat beschadiging de prestaties beïnvloedt, maar er wordt nog weinig rekening gehouden met dit soort schade tijdens het trainen van modellen.

Een ander knelpunt is dat de meeste CNN-modellen uitgaan van een gesloten systeem: ze gaan ervan uit dat elk monster tot een bekende soort behoort. In regio's met veel biodiversiteit is dat niet realistisch, omdat daar ook onbekende of moeilijk te onderscheiden soorten voorkomen. Pogingen om dit op te lossen met opt-out categorieën of drempelwaarden hebben tot nu toe weinig succes gehad. Daarom groeit de belangstelling voor zogenaamde open-wereldmodellen. Een veelbelovende aanpak hierin is object-herkenning via re-identificatie: beelden worden vertaald naar zogeheten ‘*embeddings*’, een soort digitale vingerafdruk, waarbij beelden van dezelfde soort in dezelfde cluster vallen. Nieuwe beelden kunnen dan vergeleken worden met een referentiedatabase. Met technieken als *triplet learning* en binaire verificatie worden modellen getraind om subtiele verschillen tussen soorten te herkennen. Hoewel deze aanpak complexer is en nauwkeurige selectie van trainingsvoorbeelden vereist, sluit ze beter aan bij hoe experts in de praktijk te werk gaan: via vergelijking in plaats van categorisatie.

Een belangrijke uitdaging bij het opzetten van effectieve herkenningssystemen is de enorme soortenrijkdom van bomen wereldwijd, vooral in tropische gebieden zoals de DRC. Hout is bovendien een zeer variabel materiaal, beïnvloed door genetica, groeiplaats en de positie binnen de boom (bijvoorbeeld

stam versus tak). Daardoor is het lastig om uniforme herkenningsskenmerken vast te stellen. Ook worden wetenschappelijke namen regelmatig herzien, wat het samenstellen van betrouwbare databases bemoeilijkt. In de houtindustrie worden soorten bovendien vaak gegroepeerd onder handelsnamen, die niet altijd overeenkomen met botanische realiteit. Bestaande databases zoals *InsideWood*, *macroHOLZdata*, *CITESWoodID* en het *Atlas of Macroscopic Wood Identification* bieden nuttige informatie, maar bevatten vaak te weinig variatie binnen soorten en zijn niet altijd gekoppeld aan fysieke referenties. Daardoor zijn ze minder geschikt voor het trainen van robuuste AI-modellen. De beperkingen en achterliggend problematiek wordt omschreven in hoofdstuk 1 a.d.h.v. literatuur.

Om deze beperkingen tegemoet te komen, ontwikkelden we SmartWoodID – de grootste referentiedatabase van gelabelde macroscopische houtdoorsneden, specifiek bedoeld voor snelle en nauwkeurige soortherkenning in de DRC, waar illegale houtkap veel voorkomt. De database is gebaseerd op de uitgebreide Tervuren houtcollectie (ondergebracht in het Koninklijk Museum voor Midden-Afrika, Tervuren, België) en bevat meerdere, kwalitatief hoogwaardige beelden per soort, met focus op commercieel belangrijke taxa. In tegenstelling tot andere databases bevat SmartWoodID ook hout met zichtbare schade of variatie, zoals scheuren, schimmel of insectenvraat. Hierdoor zijn modellen die op deze beelden getraind zijn beter bestand tegen realistische veldomstandigheden. De bouw van deze databank wordt beschreven in hoofdstuk 2.

In hoofdstuk 3 onderzochten we de in welke mate de 31 gestandaardiseerde kenmerken bij 601 houtsoorten identificatie mogelijk maken, gebruikmakend van de SmartWoodID-database. Hoewel deze kenmerken helpen om binnen kleine taxonomische groepen te onderscheiden, bleken ze op bredere schaal onvoldoende onderscheidend. Modellen die alleen op deze kenmerken gebaseerd waren, behaalden slechts ongeveer 50% nauwkeurigheid op genusniveau bij 56 Congolese commerciële houtsoorten. Dit benadrukt de noodzaak om ook de andere houtanatomische kenmerken, zichtbaar op andere vlakken en bij grotere microscopische vergroting, te gebruiken en te combineren met andere methoden.

Hoofdstukken 4 en 5 onderzochten CNN-modellen die getraind werden op beelden van houtdoorsneden. Deze modellen bleken veel beter in staat om fijne kleur- en textuurpatronen te herkennen – informatie die experts intuïtief gebruiken, maar niet in vaste kenmerken wordt vastgelegd. CNNs presteerden aanzienlijk beter dan modellen op basis van handmatige geannoteerde kenmerken, met nauwkeurigheden boven de 85% en in meer dan 95% van de gevallen stond het juiste genus in de top zes voorspellingen. Dit bevestigt dat visuele data meer diagnostische waarde heeft dan eerder werd aangenomen, en dat AI deze informatie effectief kan benutten.

Omdat in de praktijk vaak een combinatie van visuele indrukken en gestandaardiseerde kenmerken wordt gebruikt, onderzochten we in hoofdstuk 6 of deze twee benaderingen gecombineerd konden worden. Het opnieuw rangschikken van CNN-voorspellingen met behulp van handmatige kenmerken leverde voor sommige soorten verbeteringen op, maar bij andere – waaronder belangrijke zoals *Khaya* – leidde het tot lagere nauwkeurigheid. Dit wijst op het belang van zorgvuldige afstemming bij het combineren van methodes.

We onderzochten ook hoe de samenstelling van de trainingsdata de prestaties beïnvloedt. Modellen presteerden beter naarmate er meer verschillende monsters en grotere scanoppervlakken beschikbaar waren. Modellen getraind op onbeschadigde beelden behaalden de hoogste sensitiviteit (90,5%), gevolgd door gemengde (88,4%) en beschadigde beelden (79,1%). Analyse met Grad-CAM liet zien dat modellen zich vooral richten op intacte structuren, wat het belang van zorgvuldig monsterbeheer en beeldkwaliteit benadrukt.

Tot slot keken we in hoofdstuk 5 naar herkenningstrategieën voor open-wereldtoepassingen (scenario's waarin niet alle mogelijke soorten vooraf in het model zijn opgenomen). Binaire verificatie, waarbij beelden vergeleken worden met referentievoorbeelden in plaats van direct geclassificeerd, bleek robuust en effectief – zelfs bij onbekende soorten. Deze aanpak is bijzonder geschikt voor situaties waarin de vraag niet is "welke soort is dit?" maar "komt deze soort overeen met wat is opgegeven?". In vrijwel alle gevallen presteerde deze methode minstens zo goed als of beter dan traditionele classificatie.

We onderzochten ook *triplet learning*, waarbij beelden worden omgezet in numerieke vectoren die het anatomisch patroon vastleggen. Deze vectoren kunnen vervolgens worden vergeleken of ingevoerd in eenvoudige technieken zoals *nearest neighbour* of XGBoost. De eerste resultaten waren gemengd – mogelijk door suboptimale training – maar de aanpak blijft veelbelovend, vooral voor toekomstige systemen die meerdere soorten data (zoals DNA of chemische profielen) combineren.

Dit onderzoek biedt een directe vergelijking van houtidentificatie a.d.h.v. traditionele houtkenmerken en CNN-classificatie op houtsoorten in de DRC. Het benadrukt het belang van open-wereld benaderingen voor de ontwikkeling van betrouwbare, schaalbare en toegankelijke identificatiesystemen. De bevindingen bieden waardevolle handvatten voor de bestrijding van illegale houtkap en vormen de basis voor de volgende generatie AI-ondersteunde houtherkenningstools.

Summary (English)

Illegal logging significantly impacts forests, posing a high risk of irreversible damage, particularly when exploiting populations of protected species. Thirty to ninety percent of traded tropical timber is estimated to have been harvested illegally. In response, a range of regulatory frameworks has been established to enhance transparency and traceability within the timber supply chain. These include the FLEGT Action Plan, the EUTR and EUDR, the U.S. Lacey Act, and CITES. The effectiveness of these measures, however, hinges on a single critical capability: the rapid and accurate identification of wood species across all stages of the supply chain—from forest harvest sites to transport hubs, storage facilities, and ports.

Wood identification methods encompass anatomical, chemical, and genetic techniques. Of these, anatomical assessment—based on the observation of wood's cellular and tissue structures—remains the most widely applied. The IAWA provides standardized anatomical features that underpin global anatomical assessment practices. While microscopic analysis offers high taxonomic resolution, it is constrained by the need for specialized equipment, expert training, and intensive sample preparation, limiting its utility for frontline enforcement. Macroscopic anatomical assessment, by contrast, relies on features visible to the naked eye or a hand lens and provides a more accessible, low-cost alternative applicable to freshly cut or sanded surfaces. Cross-sectional views are especially informative, revealing diagnostic traits such as vessel arrangement, ray width, and parenchyma distribution—features used in field keys and identification guides.

Despite its operational simplicity, macroscopic anatomical assessment is limited by the relatively small number of observable features and significant intra-species variation, which can undermine diagnostic accuracy. Furthermore, the categorical nature of IAWA descriptors (e.g., "present," "variable," "absent") may obscure subtle but taxonomically relevant variation and introduce subjectivity in interpretation. Importantly, the actual discriminatory power of these features—especially in species-rich tropical regions—has not been systematically evaluated at scale, constraining the development of reliable identification keys and limiting the benchmarking of emerging data-driven models.

CV offers a compelling alternative by automating wood species recognition through image analysis. CNNs, in particular, can extract diagnostic features directly from macroscopic cross-sectional images, enabling rapid and accurate assessments. These models can operate on portable devices and have already been piloted in enforcement scenarios, such as with the XyloTron system in Ghana. However, existing computer vision applications face key limitations. Training and testing data are often derived from pristine specimens, raising concerns about robustness in real-world conditions, where samples may exhibit cracks, insect damage, discoloration, or fungal decay. These factors can obscure anatomical features and degrade model performance. Recent studies (e.g., Ravindran et al. 2023; Owens et al. 2024) have begun to quantify the impact of occlusion of anatomical information on classification accuracy, but do not tackle the influence of including damage during training of models.

Moreover, most CNN-based models adopt a multiclass classification approach, which assumes that all test samples belong to a fixed set of known species. This closed-world assumption limits their applicability in biodiverse regions, where unknown or closely related species may occur. Attempts to mitigate this limitation through opt-out classes or confidence thresholds have shown limited success. In response, the field is increasingly turning to open-world recognition frameworks. One promising approach is object re-identification, which encodes images as embedding vectors (anatomical fingerprints) within a learned feature space, where samples from the same species form clusters. Identification is then performed by comparing a query image to a reference database, allowing recognition of both known and novel species. Training strategies such as triplet learning and binary verification promote the learning of discriminative, species-specific representations. While more complex to implement—requiring careful sample selection and robust loss functions—re-identification approaches offer generalizability to timbers beyond the taxonomic scope of the training data, and align more closely with expert practices, which often rely on comparative rather than categorical judgments.

A central challenge in building effective wood identification systems is the immense diversity of tree species worldwide, particularly in tropical regions like the DRC. Wood is a highly variable biological material, influenced by genetics, environmental conditions, and intra-tree location (e.g., trunk vs. branches, pith vs. bark). This complexity makes it difficult to define consistent diagnostic criteria. Additionally, taxonomic classifications are frequently revised, complicating database curation. Commercial trade further obscures species-level identification by grouping timbers under broad trade names based on physical properties rather than botanical identity. Current databases, such as InsideWood, macroHOLZdata, and CITESWoodID, and the Atlas of Macroscopic Wood Identification, offer valuable resources but may not cover the necessary variability that wood anatomical feature can portray within a species. InsideWood, for example, compiles species descriptions and images but often generalizes from limited specimens and lacks traceability to physical references. This constrains assessments of intra-specific variation and undermines the reliability of training data for machine learning models. The limitations and underlying problems are described in Chapter 1, based on literature.

To address these limitations, we built SmartWoodID, the largest reference database of annotated macroscopic cross-sectional images designed to support rapid and accurate wood identification in the DRC, a hotspot of illegal logging. SmartWoodID draws on the extensive Tervuren wood collection and includes multiple high-quality images per species, prioritizing economically important taxa. Unlike other databases, SmartWoodID intentionally includes specimens with natural variation and surface defects (e.g., cracks, fungal stains, insect damage) to better represent real-world conditions. This ensures that models trained on the dataset are more resilient to the variability encountered in field applications. The construction of this database is described in Chapter 2.

In Chapter 3, we systematically evaluated the diagnostic utility of 31 standardized macroscopic features across 601 timber species using the SmartWoodID dataset. While useful for narrowing identifications within small taxonomic scopes, these features exhibited limited discriminatory power at broader scales. Predictive models based solely on expert-defined features achieved only ~50% genus-level accuracy across 56 commercial Congolese genera, with significant anatomical overlap and large candidate sets required for confident identifications. These findings highlight the need to reassess the diagnostic validity of traditional descriptors and suggest that future research should explore both improvements to feature-based methodologies and complementary techniques to enhance field applicability.

To investigate whether visual information not captured by standard descriptors could improve identification, Chapters 4 and 5 explored CNN models trained on raw cross-sectional images. These models preserved nuanced patterns of colour and texture that experts intuitively use but which are not codified in existing feature sets. CNNs achieved substantially better performance than feature-based models, with precision, recall, and accuracy all exceeding 85% at the genus level. The correct genus was among the top six predictions in over 95% of test cases. These results affirm that raw visual data contains richer diagnostic information than codified features alone and that CV can effectively harness this information.

Recognizing that real-world identifications often integrate anatomical descriptors and visual impressions, Chapter 6 examined whether expert-defined features could be used to refine CNN predictions. Re-ranking top-k CNN outputs using feature data led to modest improvements for some genera but reduced accuracy for others, including priority genera such as *Khaya*. This indicates that while hybrid approaches have potential, their implementation must be carefully tailored to avoid counterproductive effects.

The study also addressed critical factors in building effective training databases. Empirical analyses showed that increasing specimen representation and scan area improved CNN performance, underscoring the value of capturing anatomical variability. Models trained on pristine image patches achieved higher recall (90.5%) than those trained on mixed (88.4%) or damaged (79.1%) patches. Grad-CAM visualizations confirmed that CNNs consistently focused on intact anatomical structures, further supporting the emphasis on high-quality specimen preparation and imaging during database construction.

Chapter 5 evaluated scalable identification strategies for open-world contexts. Binary verification emerged as a particularly promising approach, comparing query images to reference samples to generate similarity scores rather than fixed labels. This method performed robustly, even for species not included in training, and proved effective in practical scenarios where the goal is to verify the plausibility of declared identities rather than assign definitive species labels. Binary verification matched or outperformed multiclass models in ranking the correct genus among top candidates for the 56 Congolese genera studied.

We also evaluated triplet learning, which transforms images into numerical vectors representing anatomical patterns. These embeddings can be directly compared or fed into lightweight classifiers such as nearest-neighbour or XGBoost. Although initial performance was suboptimal—likely due to suboptimal selection of hard training examples—the approach remains promising, particularly for integrating multimodal data (e.g., DNA, chemical signatures) into unified identification systems.

In conclusion, this study provides the first direct comparison of expert-defined feature-based keys, CNN classifiers, and re-identification models under realistic, field-like conditions in the DRC. It emphasizes the need for open-world recognition frameworks and hybrid strategies to create robust, scalable, and interpretable systems for timber identification. The findings have direct implications for international efforts to combat illegal logging and lay the groundwork for next-generation, AI-enabled wood identification tools tailored to the operational realities of enforcement.

Table of Contents

Woord vooraf	2
List of abbreviations	4
Summary (Dutch).....	5
Summary (English)	8
Table of Contents	11
Chapter 1: General introduction	15
1.1 Illegal logging: A major threat to global forests and sustainability	15
1.1.1 The importance of forests in tropical regions.....	15
1.1.2 Law enforcement.....	16
1.2 State-of-the-art in wood identification techniques	17
1.3 Scalable wood Identification: assessment of available techniques for field-application and knowledge gaps	20
1.3.1 Diagnostic information for rapid identification in the field	21
1.3.2 The need for reference databases.....	23
1.3.3 Unexplored areas in literature regarding macroscopic cross-sectional identification applications.....	24
1.4 Structure of the PhD	26
1.5 Structure of the data in each chapter	31
Chapter 2: SmartWoodID—an image collection of large end-grain surfaces to support wood identification systems	33
2.1 Abstract	33
2.2 Introduction	34
2.3 Material and methods	35
2.3.1 Delineating a set area for the first edition	35
2.3.2 Species selection	37
2.3.3 Collecting the specimens	38
2.3.4 Sanding.....	39
2.3.5 Scanning.....	39
2.3.6 Annotating (updated after publication)	39
2.3.7 Quality control.....	41
2.3.8 Technical description of the database and functionalities	42
2.4 Results.....	43
2.4.1 Taxonomic coverage (updated after publication)	43
2.4.2 Geographical coverage	43
2.4.3 Threatened status.....	44

2.4.4	Economic value	46
2.5	Discussion	47
2.5.1	The DRC and its representativeness	47
2.5.2	Opportunities of the SmartWoodID database	47
2.6	Conclusion.....	49
2.7	Data availability.....	49
2.8	Acknowledgements	49
Chapter 3: Evaluating expert-defined anatomical features for rapid identification of Congolese tree species		50
3.1	Abstract	50
3.2	Introduction	51
3.3	Materials and Methods	53
3.3.1	Dataset Description	53
3.3.2	Macroscopic feature annotations	53
3.3.3	Clustering analysis	54
3.3.4	Classification Models	54
3.4	Results.....	58
3.4.4	Clustering analysis	58
3.4.5	Classification models	65
3.5	Discussion	73
3.6	Conclusion.....	78
3.7	Acknowledgements	78
3.8	Availability of code and metadata.....	78
Chapter 4: Evaluating the effect of anomalous images on CV-based wood identification models		79
4.1	Abstract	79
4.2	Introduction	0
4.3	Material and Methods	1
4.3.2	Dataset Description	1
4.3.3	Model architecture	2
4.3.4	Classification of image patches by presence or absence of anomalies.....	3
4.3.5	Classification of image patches by genus	3
4.4	Results.....	5
4.4.1	Classification of patches by anomalies.....	5
4.4.2	Classification of patches by genus	8
4.5	Discussion	15

4.6	Conclusion.....	17
4.7	Acknowledgments.....	17
4.8	Availability of code and metadata.....	17
Chapter 5: Expert-defined vs. deep learning approaches for scalable and generalizable timber identification.....		18
5.1	Abstract	18
5.2	Introduction	19
5.3	Material and methods.....	21
5.3.2	Dataset Description	21
5.3.3	CNN Backbone Architecture	21
5.3.4	Classification Techniques	21
5.3.5	Balancing the dataset	24
5.3.6	Data augmentation	24
5.3.7	Hard mining for Object Re-Identification approaches	25
5.3.8	Standardizing CNN output for model comparison	26
5.3.9	Model evaluation	27
5.4	Results.....	29
5.4.2	Comparison of multiclass classification models	29
5.4.3	Comparison of CNN-based approaches	32
5.4.4	Top-k classification accuracy analysis	35
5.4.5	Generalization to unseen taxa	36
5.5	Discussion	38
5.6	Conclusion.....	41
5.7	Acknowledgements	41
5.8	Availability of code and metadata.....	41
Chapter 6: Combining expert-defined and CV-extracted macroscopic cross-sectional features for enhanced wood identification		42
6.1	Abstract	42
6.2	Introduction	43
6.3	Material and methods.....	44
6.3.1	Dataset Description	44
6.4	Results.....	46
6.5	Discussion	51
6.6	Conclusion.....	53
6.7	Acknowledgements	53
6.8	Availability of code and metadata.....	53

7	General discussion	54
7.1	Evaluating expert-defined macroscopic cross-sectional features for identification of Congolese tree species	54
7.2	The importance of colour and subtle patterns for identification	55
7.3	Refining CV-Based wood identification using expert-defined features	55
7.4	Reflecting on future research for wood identification.....	56
7.4.1	Database construction	57
7.4.2	Model development	61
	Funding	69
	Bibliography.....	70
Chapter 8:	Supplementary materials	87
8.1	CNNs for image classification	87
8.2	Definitions of classification performance metrics: Accuracy, Precision, Recall, and F1-score	154
8.2.1	Accuracy	154
8.2.2	Precision	154
8.2.3	Recall.....	154
8.2.4	F1-Score.....	154
8.3	Model performance during training.....	155
8.4	Correlations between class-specific changes in recall	157
8.4.1	Across trained models and model-specific training metrics.....	157
8.4.2	Within trained models (between anomaly-free and anomalous test data) and model-specific training metrics	158
	Curriculum Vitae.....	175

Chapter 1: General introduction

1.1 Illegal logging: A major threat to global forests and sustainability

Environmental crimes—illegal activities that exploit natural resources and harm ecosystems—are among the most pressing global challenges (van Uhm, 2024; White, 2018). They contribute to biodiversity loss by driving deforestation, habitat destruction, and the decline of vulnerable species through illegal logging, poaching, and land degradation (FAO, 2022; Gibbs and Boratto, 2017; Lirëza and Koçi, 2023). These activities disrupt entire ecosystems, reducing species richness and threatening key species that maintain ecological stability. Furthermore, these crimes undermine sustainable development by depleting essential natural resources such as timber, freshwater, and soil fertility, which local communities and economies rely on for their livelihoods, jeopardizing the well-being of future generations (Gabris, 2025). Protecting forests, wildlife, and other natural resources is therefore crucial to preserving ecological balance and ensuring long-term environmental sustainability. (DeFries et al., 2007; Inatimi, 2023). Among environmental crimes, illegal logging is the most profitable, representing 50 to 152 billion USD per year (equivalent to 10-30 % of the total global timber trade (Nellemann and INTERPOL Environmental Crime Programme (eds), 2012)), posing a high risk of irreversible damage, particularly when targeting threatened species (Lowe et al., 2016; Tacconi et al., 2016). It leads to widespread deforestation and endangers the survival of vulnerable tree species.

1.1.1 The importance of forests in tropical regions

The problem is particularly severe in tropical regions, where forests play an indispensable role in regulating climate and sustaining species richness (Stokstad, 2014). Tropical forests sequester and store vast amounts of carbon, acting as one of the planet's most important carbon sinks and helping to mitigate the effects of anthropogenic climate change (Lewis et al., 2015; Watson et al., 2000). Forest degradation not only releases stored carbon into the atmosphere but also reduces the planet's capacity to sequester future emissions, intensifying global warming (Mitchell et al., 2017). Estimates suggest that 30% to 90% of traded tropical timber is harvested illegally, making illegal logging a major driver of forest loss (Hirschberger, 2008;

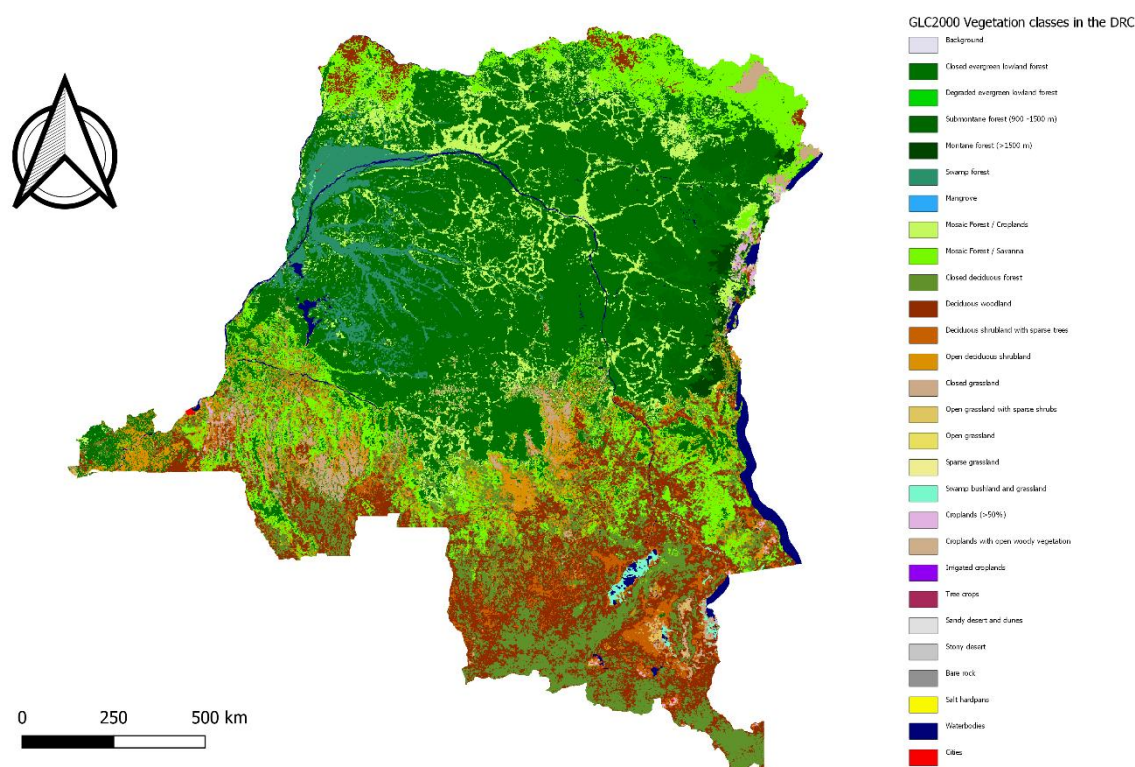


Figure 1.1: Map of the Democratic Republic of the Congo, showing the land area cover in vegetation classes according to GlobalLandCover map 2000, adapted from ("Global Land Cover 2000 database. European Commission, Joint Research Center" 2000)

Hoare, 2015; Magrath et al., 2009). Among these regions, the Congo Basin stands out as a crucial stronghold for climate stability and biodiversity conservation (Shapiro et al., 2021). As the second-largest tropical rainforest in the world, it spans approximately 178 million hectares (Mayaux et al., 2013) and plays a pivotal role in carbon sequestration, acting as a significant and stable carbon sink in aboveground biomass—sequestering 0.66 tonnes of carbon per hectare per year over the three decades leading up to 2015 (Dargie et al., 2017; Hubau et al., 2020). However, illegal logging threatens this critical biome, not only accelerating forest degradation but also disrupting indigenous communities that depend on these forests for their livelihoods (Aleman et al., 2017; Mulvagh, 2006; Piabuo et al., 2021; Réjou-Méchain et al., 2021). This problem is especially pertinent in the DRC. Approximately half of the rainforest area in the Congo Basin is located within the boundaries of the DRC (Potapov et al., 2012), and the DRC features the highest area of annual forest cover loss compared to other Central-African countries ((Rome), 2010; Lawson, 2014). Other forest types in the DRC, such as the dry deciduous Miombo woodlands, are also being overexploited, particularly for tree species that are currently—or soon may be—threatened with extinction (CITES, 2022a, 2019). In southern DRC, Miombo woodlands cover nearly 23% of the national forest area and dominate the former Katanga province (Kabulu Djibu et al., 2008; Potapov et al., 2012) (see Figure 1.1). These forests face increasing anthropogenic pressure from agricultural expansion, fuelwood collection, charcoal production, and rapid urbanisation (Hourticq and Carole Megevand, 2013; Münkner et al., 2015; Potapov et al., 2012). This has led to loss in species diversity and abundance, reduced availability of non-wood forest products, declining access to bushmeat, and negative climatic effects such as altered rainfall patterns (Barima et al., 2011; Kazadi and Kaoru, 1996; Malaisse, 1997). A particularly alarming trend is the illegal exploitation of *Pterocarpus tinctorius* Welw., a high-value timber species. Once used primarily in traditional medicine and as a dye (Augustino and Hall, 2008), it has become a target for unsustainable logging driven by demand in the non-Congolese luxury furniture market (Hong et al., 2020). In areas such as Kasenga territory, this has shifted local labour away from subsistence farming toward illicit logging, accelerating forest fragmentation and degradation (Cabala Kaleba et al., 2017). Given its ecological significance, protecting tropical regions like the Congo Basin, and the DRC in particular, is paramount to mitigating climate change, preserving biodiversity, and ensuring the sustainability of tropical forest ecosystems worldwide.

1.1.2 Law enforcement

To combat this issue, a complex framework of international and national regulations has been established. At the international level, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) plays a central role (UNEP-WCMC (Comps.), 2022). CITES regulates the trade of protected species through a tiered system of Appendices: Appendix I prohibits commercial trade in species threatened with extinction; Appendix II restricts trade in species that are not currently threatened but could become so without strict regulation; and Appendix III covers species protected at the request of a Party that requires international cooperation to prevent unsustainable exploitation (UNEP-WCMC (Comps.), 2022). Importantly, CITES also extends protections to so-called “look-alike” species, which resemble listed taxa closely enough to be easily confused in trade (Gasson et al., 2011). Under CITES, importers are required to declare both the botanical identity and geographic origin of imported wood (Wiedenhoeft et al., 2019).

Complementing CITES, a variety of policy measures have been adopted at regional and national levels to improve forest law enforcement and governance. In the European Union, the *FLEGT Action Plan* was launched in 2003 (Ayed, 2006; EC, 2003; Jonsson et al., 2015). The plan evolved into the FLEGT licensing system, operationalized through bilateral voluntary partnership agreements with exporting countries, to ensure only legally harvested timber enters the EU market (European Commission, 2019; Jonsson et al., 2015). To further strengthen the system, the EUTR came into force in 2013, shifting responsibility to importers and first-time suppliers to exercise due diligence in minimizing the risk of illegality (Parliament, 2023; Tegegne et al., 2018; Union, 2010). More recently, the EUDR expanded this framework by imposing stricter requirements on deforestation-free supply chains (Köthke et al., 2023; Parliament, 2023). Together, these instruments prohibit placing illegally harvested timber on the EU market, mandate risk assessment and mitigation measures, and require traceability back to the country, region, or concession of harvest (Lowe et al., 2016).

In the United States, the *Lacey Act*—originally enacted in 1900 and amended in 2008 to cover plants and plant products—prohibits the trade of illegally sourced timber (Alexander, 2014; Lowe et al., 2016). This law requires identification at the genus–species level and specification of the country of harvest, and violations include importing, exporting, transporting, selling, or acquiring plants in violation of any domestic or foreign law, as well as falsifying records or mislabelling products.

Comparable legislation is enforced in other parts of the world. Australia regulates illegal timber through the *Illegal Logging Prohibition Act 2012* and *Regulation 2012*, which criminalize the import or processing of illegally harvested timber and require businesses to conduct due diligence on supply chains (Lowe et al., 2016; World Resources Institute, 2024). The framework obliges importers and processors to collect information on product species, origin, and harvest location, assess risks of illegality, and maintain written records. Following a mandatory ten-year review, the original act was amended in 2024 through the *Illegal Logging Prohibition Amendment (Strengthening Measures to Prevent Illegal Timber Trade) Act 2024* and the *Illegal Logging Prohibition Rules 2024*. These were enforced starting March 2025, with a six-month transition period. The reforms introduced two streamlined risk-assessment pathways (certified vs. non-certified timber), a repeat due diligence exception for imports from the same supplier within twelve months, and strengthened monitoring through timber testing technologies and mandatory pre-import notices. Enforcement has also been tightened. The reforms expand audit powers, establish strict liability offenses alongside fault-based ones, and increase penalties. Public disclosure of non-compliance further raises reputational risks.

Canada similarly enforces the *Wild Animal and Plant Protection and Regulation of International and Interprovincial Trade Act* (1992), which prohibits the import or possession of illegally harvested plants and imposes penalties for misrepresentation of plant identity or origin (Government of Canada, 2025; Lowe et al., 2016).

Despite these legal frameworks, enforcement remains a major challenge. Enforcing these mechanisms is crucial to curbing illegal logging and ensuring forests remain a vital resource for future generations (Gasson et al., 2021; Piabuo et al., 2021). The enforcement of timber trade regulations depends on the ability to accurately verify both the species and origin of traded wood (Lowe et al., 2016). Species verification ensures that the declared taxon matches the wood being sold, which is critical where regulations apply to particular taxa (e.g., *Dalbergia* spp. under CITES). Origin verification, in contrast, establishes the geographic source of the timber, since legality is tied to compliance with harvesting laws in the jurisdiction of harvest. Trade documents do not always reflect the actual timber being sold enabling illegal timber trade and fraud, and underscoring the need to verify claims of legality and ensure compliance with regulatory frameworks.

1.2 State-of-the-art in wood identification techniques

Wood identification refers to the process of determining the botanical taxon of a given sample with the highest possible certainty based on a set of diagnostic features (Jacobs and Baker, 2018). Identification follows a hierarchical refinement process, iteratively adjusting the classification to achieve the highest possible taxonomic specificity. Ideally, a sample is identified at the species level, but when species-level resolution is not feasible, it may be assigned to a broader genus or functional group representing similar timber types. Identification consists of two key steps: feature extraction (or assessment) and classification. Feature extraction involves isolating diagnostically relevant characteristics from raw data. These features can later be used in classification models to distinguish between taxa or other categories, depending on the analytical context. Classification, in turn, uses these extracted features to predict class probabilities and assign specimens to one or more taxa based on the highest probability.

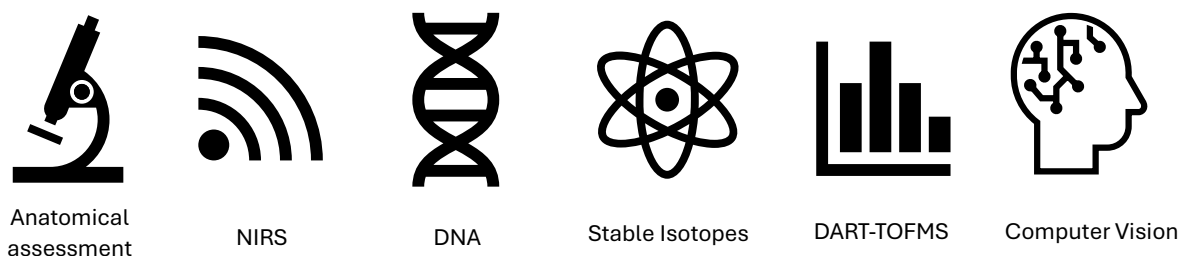


Figure 1.2: Overview of methods for extracting diagnostic information for taxonomic identification or origin tracking, based on Schmitz et al. (2019) (Schmitz et al., 2019). From left to right: Anatomical assessment; Near-Infrared Spectroscopy (NIRS); Genetic methods; Stable isotope ratio analysis; DART-TOFMS; Computer Vision-based wood identification.

Various approaches to extract stable diagnostic patterns are being studied. An overview is presented in Figure 1.2, based on Schmitz et al. (2019) (Schmitz et al., 2019). One of the most established methods involves the study of wood anatomy (Koch et al., 2015; Richter and Dallwitz, 2000; Wheeler, 2011). Wood is an anisotropic material composed of multiple tissues that function together to support tree growth and physiological processes. Tracheids, the main cell elements of softwoods and present in some hardwoods, conduct water transport and provide mechanical support. Vessels facilitate the transport of water and mineral solutes in hardwoods from the roots to the leaves, enabling photosynthesis. Rays serve as horizontal transport pathways between the pith and bark. Parenchyma cells act as pathways for metabolic products, short-distance transport, and storage. Those pathways form a complex three-dimensional network that distributes resources throughout the tree. Fibres provide structural strength, allowing trees to withstand external forces such as wind and gravity. The unique characteristics of these tissues—whether individually, in combination, or as part of larger anatomical patterns—contain diagnostic information that can be used to differentiate species and even determine geographic origin based on local environmental influences. (Butterfield, 2012; Niemz et al., 2023b)

Traditionally, wood identification has relied on anatomical assessment, where the structure of the cells, cell walls and tissues is visualized using partially destructive techniques such as microtomy, sanding, or laser ablation (Arzac et al., 2018; Fukuta et al., 2016; Guo et al., 2021; Spiecker et al., 2000; Tardif and Conciatori, 2015). Processed samples are then examined through microscopic observation across different planes and magnifications to identify species-specific diagnostic features (Niemz et al., 2023b; NS, 1989; Tardif and Conciatori, 2015). The three primary anatomical planes considered in wood identification are the transverse plane, radial plane, and tangential plane (see Figure 1.3), each providing a distinct perspective on wood structure (Butterfield, 2012). Their orientation is determined by the natural growth pattern of the tree. The transverse plane (also known as the axial plane, cross-sectional plane, or end-grain) is obtained by cutting the wood perpendicular to the trunk's length. This reveals a circular cross-section, exposing growth rings, vessel arrangements, and other structural features essential for species differentiation. The radial plane (vertical longitudinal section through the centre) is obtained by cutting vertically along the trunk's length, passing through the pith. The tangential plane (vertical longitudinal section away from the centre) is obtained by cutting vertically along the trunk's length but not through the pith, following the curvature of the growth rings. Traditionally, these planes are studied separately to analyse specific features. Modern imaging technologies, such as X-ray CT, enables researchers to examine wood anatomy across all three planes in a three-dimensional dynamic, interactive manner, revealing how structures change throughout the tree and how they are interconnected (Van den Bulcke et al., 2009). Additionally, some diagnostic features (e.g. fluorescence of heartwood, or water/ethanol extracts) are only visible at specific magnifications or by using specific wavelengths of the electromagnetic spectrum, necessitating customized visualization techniques to enhance accuracy in species identification (NS, 1989; Price et al., 2021). To ensure consistency in wood anatomical assessment, standardized feature sets have been established by the IAWA (Angyalossy et al., 2016; Committee, 2004; Gasson et al., 2011; NS, 1989; Ruffinatto et al., 2015; Wheeler, 2011). These descriptors provide a systematic framework for comparing species, improving both reliability and reproducibility. The long-established practice of

anatomical assessments continues to be one of the most fundamental approaches in forensic and regulatory applications (Wheeler and Baas, 1998).

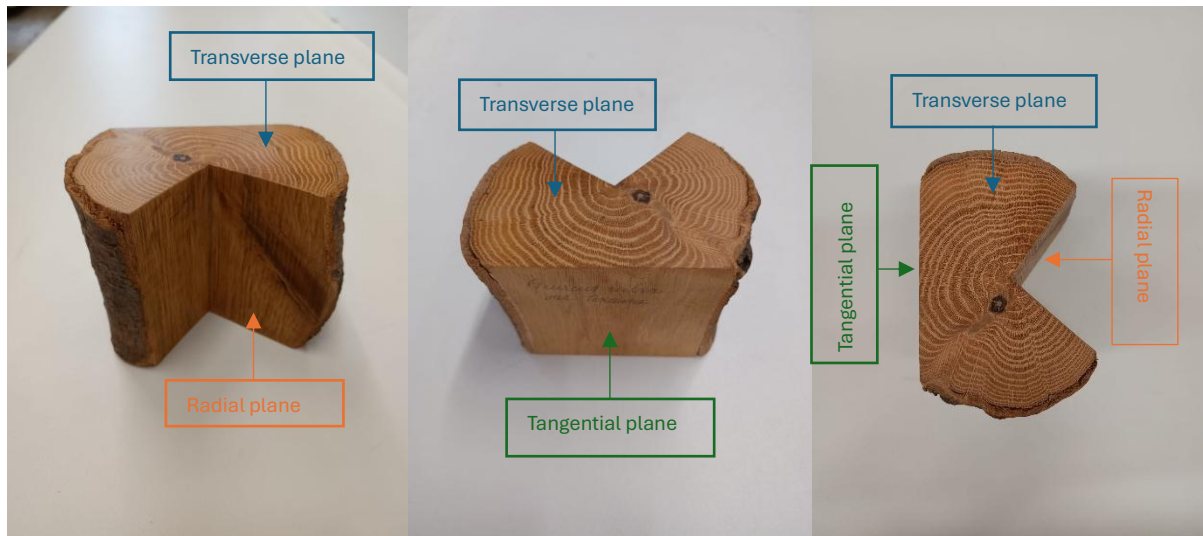


Figure 1.3: The three primary anatomical planes considered in wood identification: transverse plane (or cross-sectional plan), radial plane, and tangential plane. ©RubenDeBlaere(first author)

Several alternative techniques have been developed in recent decades for identifying the botanical taxa and determining geographic provenance, by leveraging chemical composition, genetic markers, and stable isotope ratios. NIRS is primarily used to analyse the chemical and physical properties of wood, but it has potential for species differentiation (Deklerck, 2019; Lowe et al., 2016; Tsuchikawa et al., 2003). This technique operates by measuring the absorption of near-infrared light (800–2500 nm), which interacts with high-molecular-weight compounds such as cellulose, hemicellulose, lignin, and extractives (Tsuchikawa et al., 2003; Tsuchikawa and Kobori, 2015). Because different species exhibit distinct absorption patterns, NIRS can be used to distinguish between species or even subspecies. Genetic techniques use deoxyribonucleic acid (DNA)-based approaches for species identification and origin determination. Genetic methods are particularly effective at distinguishing closely related species through techniques such as DNA-barcoding (Jiao et al., 2020, 2019) and, in some cases, even tracing individual logs throughout the supply chain using DNA fingerprinting (Lowe et al., 2010). Stable isotope analysis provides insights into the geographic origin of timber rather than species identification (Lin et al., 2024). This technique examines the ratios of naturally occurring stable isotopes in wood, which vary based on geographical location, climate conditions, soil composition, and local geology (Camin et al., 2017; Horacek et al., 2009). Since trees absorb elements from their surrounding environment, their isotopic composition reflects their growth location, allowing researchers to estimate timber provenance (Dormontt et al., 2015; Kagawa and Leavitt, 2010). Mass spectrometry serves as a valuable analytical tool for wood identification, with DART-TOFMS being extensively studied (Cody et al., 2005). This method produces chemical fingerprint for wood by ionizing low-molecular-weight compounds (<1000 Daltons) through thermal desorption. A single wood sliver is placed in an open stream of excited helium atoms, which ionize the chemical compounds present in the sample. These ionized molecules are then propelled into the mass spectrometer, where their mass-to-charge (m/z) ratios are measured and shown to the user in real time (DART). The technique operates on the principle of TOFMS, in which heavier molecules take longer to travel through the instrument, allowing for their precise differentiation based on mass. DART-TOFMS has shown significant potential for both species identification and geographic provenance determination, making it a promising tool for forensic and regulatory applications in timber trade monitoring (Deklerck, 2019). These alternative identification techniques, while still evolving, provide valuable complementary tools to traditional wood anatomical assessment, expanding the capacity to identify timber species and trace their origins. As research continues and reference databases grow, these methods are expected to play an increasingly critical role in timber trade regulation and enforcement.

The field of wood identification has increasingly explored the use of AI to automate the extraction of diagnostic patterns, aiming to enhance objectivity, scalability, and accuracy beyond traditional expert-driven methods (Hwang and Sugiyama, 2021; Silva et al., 2022). Image-based analysis of wood anatomy has proven particularly effective, demonstrating strong performance in distinguishing between timbers with highly similar anatomical structures (Owens et al., 2024; Ravindran et al., 2021, 2020, 2018; Rosa da Silva et al., 2017; Wu et al., 2021). In CV-based wood identification, the input data typically consists of images derived from various imaging modalities, including light microscopy (JANSEN et al., 1998), scanning electron microscopy (Baas and Werker, 1981; Jansen et al., 2001, 2000; Jansen and Smets, 1998), and X-ray CT (Dierickx et al., 2024). These images may capture one or multiple anatomical sections—such as cross, radial, or tangential views (Rosa da Silva et al., 2017)—or even 3D structures (Dierickx et al., 2024), depending on the imaging setup. CV-based techniques are being integrated into tools that enable non-experts to verify whether traded timber matches documentation, providing a practical mechanism for flagging suspicious cases for further forensic analysis. This approach has already shown promise in real-world deployments, such as in Ghana (Ravindran et al., 2019). A range of systems has emerged, including those using custom-designed microscopes for standardized image capture to improve model consistency (Ravindran et al., 2020), as well as smartphone-based applications that offer portability and ease of use in the field (Tang and Tay, 2019; Wiedenhoeft, 2020).

CV-based approaches to wood identification have explored using classical machine learning pipelines using hand-crafted features such as Local Phase Quantization (Rosa da Silva et al., 2022, 2017), Local Binary Patterns (Dormontt et al., 2015; Souza et al., 2020), Gray-Level Co-occurrence Matrices (Bremananth et al., 2009), or edge descriptors like Histogram of Oriented Gradients (Sugiarto et al., 2017). These descriptors were then classified with algorithms such as SVM (Joachims, 2002), RF (Biau and Scornet, 2016), or gradient boosting (Bentéjac et al., 2021). While interpretable and computationally efficient, such approaches were limited by their reliance on predefined features, which cannot fully capture the anatomical variation of wood.

Deep learning has since enabled more powerful alternatives (Hwang and Sugiyama, 2021; Silva et al., 2022). Early methods using multilayer perceptron treated images as flattened vectors, ignoring spatial structure and requiring large parameter counts, which often led to overfitting on small datasets. More recently, Vision Transformers have set benchmarks in general computer vision by employing self-attention mechanisms that capture global dependencies from the earliest layers (Gufran et al., 2023; Ye et al., 2024). However, their success depends heavily on large-scale pretraining or extensive labelled data, conditions rarely available in wood identification tasks.

CNNs remain the most effective and practical solution under these constraints. Their hierarchical architecture learns feature representations directly from images: shallow layers extract local patterns such as edges, vessels, and rays, while deeper layers integrate these into higher-level anatomical structures (Hwang and Sugiyama, 2021; Silva et al., 2022). This local-to-global progression and efficiency in small-to-moderate data regimes, makes CNNs particularly well suited to automating wood identification. CNNs can capture subtle structural differences that are imperceptible to human observers, substantially enhancing the discriminatory power of macroscopic imagery. For these reasons, CNNs were chosen as the basis for the models developed in this dissertation. The general principles of AI, deep learning, and the internal mechanisms of CNNs for image classification are outlined on further detail in Supplementary materials: CNNs for image classification.

1.3 Scalable wood Identification: assessment of available techniques for field-application and knowledge gaps

The scale of global timber production complicates regulatory oversight. In 2022, industrial roundwood, sawn wood, veneer, and plywood production amounted to approximately 2,651 million m³, with tropical species contributing significantly to these figures, with ~16% of logs, ~15% of sawnwood, ~51% of veneer, ~38% of plywood (ITTO, 2021). Timber is frequently processed at different locations worldwide, making traceability increasingly difficult. Given the vast volumes of timber moving through complex global supply

chains there is a need for identification systems that are scalable. Researchers have increasingly focused on methodologies that expedite identification while also making it more scalable across global supply chains (Brack et al., 2002; Hoare, 2015; Johnson and Laestadius, 2011; Tacconi, 2012). The effectiveness of identification methods in a laboratory setting does not necessarily translate to scalability for widespread enforcement and trade monitoring (Spiecker et al., 2000; Tardif and Conciatori, 2015). Regarding feasibility in the field, methodologies should have low costs for initial purchase, maintenance, and consumables. In addition, it is favourable if the methodologies involve tools that are robust and have a long service life. In addition, feasibility entails a low expertise barrier during use, so the method can be applied by non-experts. Finally, execution speed is also essential due to the large volumes of timber that necessitate rapid assessment to systematically cover enough wood.

1.3.1 Diagnostic information for rapid identification in the field

NIRS could become a screening tool, due to the machinery being rather straightforward and fast, but requires further development to become a common method in forensic research (Dormontt et al., 2015). DART-TOFMS, demonstrated speed and automation potential in extracting diagnostic information from wood. However, these methods come with high costs for purchasing equipment and require controlled laboratory conditions to prevent contamination, which could otherwise compromise identification accuracy. Additionally, DART-TOFMS is not a standalone technique; instead, it serves to refine broad identifications initially made through other techniques (Cody et al., 2005; Deklerck, 2019; Dormontt et al., 2015). Engineered wood products are difficult to identify using DART-TOFMS due to processing steps, such as the incorporation of adhesives, resins, and chemical preservatives (e.g., copper-based compounds, borates, creosote) or even non-wood materials introducing interference. Genetic methods face similar challenges in field applicability. While ongoing efforts aim to develop portable solutions, extracting high-quality DNA from timber remains difficult (Jiao et al., 2020, 2019, 2012; Michael Höltnen et al., 2012). Especially due to heat treatments frequently applied for drying, which degrades DNA (Jiao et al., 2020; Michael Höltnen et al., 2012). Nevertheless, as sequencing technologies and stable extraction procedure from dry wood advance, costs are decreasing, making genetic approaches increasingly viable for forensic applications and combating illegal logging (Deklerck, 2019; Lu et al., 2024). The application of stable isotope analysis in forensic wood identification is still relatively underexplored (Dormontt et al., 2015; Lin et al., 2024).

Wood anatomy presents significant opportunities for large-scale applications (Beeckman et al., 2020; Gasson, 2011). Anatomical structures serve as a stable form of diagnostic information, remaining largely preserved despite the transformative processes that roundwood undergoes when converted into commonly traded wood products. Visualizing wood anatomy can be achieved through various techniques, each with distinct advantages and limitations for field applications. The traditional approach involves cutting wood along its three principal orientations using sharp blades to expose anatomical structures (Tardif and Conciatori, 2015). This method allows for two primary modes of study: thin sectioning, where the excised tissue is observed under a microscope with transmitted light, or direct surface examination, where the exposed surface is analysed under reflected light (either natural light or enhanced via external light sources). Qualitative thin sections are typically cut to a thickness of around 14 μm , allowing anatomical features to be observed clearly under transmitted light microscopy (JANSEN et al., 1998). These sections are bleached to enhance visibility of the cell walls and stained to increase contrast between different tissue types. This technique enables the observation of most, if not all, of the anatomical features listed by the IAWA. However, certain fine details, such as pit vestures, are more effectively visualized using scanning electron microscopy (SEM) (Baas and Werker, 1981; Jansen et al., 2001, 2000; Jansen and Smets, 1998). While cutting approaches are relatively fast and cost-effective, they also present several challenges. An important limitation of cutting is the dependence on wood moisture content and density. As wood dries, it becomes increasingly difficult to cut, particularly in dense species, making it challenging to obtain clean surfaces. Irregular or torn cuts may obscure anatomical details (on the surface directly) (Ravindran et al., 2023) and complicate the extraction of clean tissue samples for thin sectioning. To mitigate this issue, softening techniques—such as boiling, storing in glycerol, or continuously wetting the wood—are often employed (Tardif and Conciatori, 2015)(Spiecker et al., 2000; Tardif and Conciatori, 2015; von Arx et

al., 2016). However, these additional steps increase both the complexity and time required, reducing the practicality of cutting-based methods for fieldwork. Producing high-quality thin sections also demands specialized expertise, followed by additional steps such as staining, fixing, and mounting samples on glass slides, which further increase costs and preparation time. Another drawback of cutting is the difficulty in achieving precise orientations, particularly in field conditions. While microtomes in laboratory settings allow for precise sectioning, field applications typically rely on handheld knives. This introduces several limitations. Cutting naturally begins at the edge of the wood sample for practical handling, meaning the central regions of the wood often remain unsampled, introducing a minor bias. Additionally, manual cutting tends to produce concave surfaces, which can distort anatomical features and complicate quantitative analysis, especially when preparing thin sections.

Sanding offers a more accessible and practical alternative for field applications (Arzac et al., 2018). Unlike thin sectioning, which allows for transmitted light microscopy, sanding only permits examination under reflected light. The sanding process involves the sequential use of progressively finer grits of sandpaper to expose cellular structures (Spiecker et al., 2000). Like cutting, it requires only inexpensive and straightforward equipment, such as manual or motorized sanders and sandpaper, and allows for rapid processing. An advantage compared to cutting is the fact that the surface quality is not impaired by the density of the material, with only minor added processing time as a result. Additionally, sanding is particularly advantageous for preparing large surface areas, making it a widely employed technique in dendrochronological studies. However, one of its primary limitations is the need for precise and consistent handling, as inconsistent sanding can obscure anatomical details and compromise analysis. Achieving high-quality results through manual manipulation of sanding tools is challenging, as it demands exceptionally steady manual control to ensure uniform surface preparation (Spiecker et al., 2000). Recent advancements in robotics, have demonstrated that automating the sanding process can overcome these limitations, producing surfaces of exceptionally high quality that facilitate detailed quantification of wood anatomical features (Van den Bulcke et al., 2025). Field application of robotic sanding remains limited due to the need for stable electricity and controlled conditions. However, small handheld tools enable sanding of small sections in under two minutes, achieving quality comparable to robotic systems and providing an effective method for rapid and scalable wood identification in the field.

Other advanced techniques, such as laser ablation or diamond fly-cutting, provide high precision but come with significant limitations (Fukuta et al., 2016; Guo et al., 2021; Spiecker et al., 2000). While effective for exposing fine anatomical structures, the techniques require specialized equipment, involve high costs, and are limited to small sample areas, restricting feasibility for field applications (Spiecker et al., 2000). In contrast, cutting and sanding remain the most affordable and widely accessible methods for wood anatomical analysis (Spiecker et al., 2000).

Despite its diagnostic power, traditional wood anatomical assessment faces challenges related to expertise, limiting scalability. Skilled specialists are required to accurately interpret anatomical features, and the increasing demand for large-scale timber identification necessitates further advancements in automation and accessibility. To reduce the expertise barrier, limited ranges of features have been proposed to serve as preliminary screening tools (Richter et al., 2017; Ruffinatto et al., 2019, 2015; Ruffinatto and Crivellaro, 2019). These methods prioritize broad taxonomic identification based on key anatomical features that are easy to visualize and assess, enabling rapid assessment of potentially suspicious cargo. The cross-section is particularly valuable for rapid field assessment. Firstly, among the three principal planes of wood—cross, radial, and tangential—the cross-section (or end-grain) is the simplest to locate, as it is characterized by the presence of ring-like growth patterns formed by secondary growth. In contrast, identifying the radial and tangential sections requires first locating the cross-section and then determining their orientation based on the wood's structure and grain direction. Secondly, the cross-section reveals numerous anatomical features that can be examined with the naked eye or a hand lens, facilitating on-site wood identification without the need for microscopes. Thirdly, radial and tangential sections are often processed and refined for use as aesthetic outer layers, whereas the cross-section is generally not utilized for this purpose. This is particularly relevant given that most wood identification methodologies involve some degree of destructive sampling, such as extracting a small splinter, sawing off

a section, cutting the surface with a knife, or sanding. Non-destructive methods, such as X-ray Computer Tomography, offer an alternative; however, their practical application in the field is limited due to the need for highly precise imaging, costly equipment, radiation shielding, complex image processing, and the challenge of selecting appropriate resolutions to clearly visualize multiple anatomical features (each requiring its own resolution range for accurate anatomical assessment) (Dierickx et al., 2024). Consequently, the cross-section is well-suited for sanding and cutting, offering a minimally invasive method for rapid wood identification without significantly damaging finished products like furniture.

1.3.2 The need for reference databases

Several databases have been developed for wood identification using cross-sectional macroscopic anatomy, relying on visual comparison, IAWA features, or a combination of both. Many of these have been implemented in field-deployable applications that integrate identification keys with visual aids to support user interpretation. Notable examples include *macroHOLZdata* (Richter et al., 2017), *CITESWoodID* (Koch et al., 2011; Weerth, 2024), the Malaysian Timber Council's *Wood Wizard* (Malaysian Timber Council, 2018), and the *Atlas of Macroscopic Wood Identification* (Ruffinatto et al., 2019). The most complete online database for wood identification is InsideWood, a wood anatomy reference, research and teaching tool, containing wood anatomical descriptions of wood based on the IAWA Lists of Microscopic Features for Hardwood and Softwood Identification accompanied by a collection of photomicrographs (NS, 1989; Wheeler, 2011; Wheeler et al., 2020). This database has a global scope, incorporating timber samples from around the world. It includes over 9,400 wood anatomical descriptions of both fossil and modern woody dicotyledons, representing more than 10,000 species across 200 plant families, and is accompanied by over 50,000 images showcasing both microscopic and macroscopic features (Wheeler, 2011). While it serves as a valuable reference resource, the anatomical descriptions are generalized and often compiled from varying numbers of individual specimens. However, no direct link is provided between the descriptions and specific reference specimens, making it impossible to verify the original source material or assess intraspecific variation with accuracy. Furthermore, certainly not all species are represented.

Wood identification is particularly challenging due to the inherent complexity of wood as a biological material. Wood is a highly variable natural material influenced by genetic and environmental factors (Downes and Drew, 2008; Stackpole et al., 2011; Wodzicki, 2001). Trees have evolved over millions of years, diversifying into a vast number of species, many of which exhibit similar diagnostic characteristics (Beech et al., 2017). This makes distinguishing between closely related species challenging, even for experts. Moreover, diagnostic features not only vary between individual trees of the same species but also within a single tree—depending on the organ (e.g., stem, branches, roots), the radial position from pith to bark, and the vertical position along the height of the tree. The number of species in tropical regions is especially high, increasing the challenge within tropical regions such as the Congo basin (Ifo et al., 2016; Partnership., 2005). Furthermore, the taxonomic classification of tree species is continuously updated as advances in plant phylogenetics refine taxonomic relationships (Denk et al., 2017; Mishler, 2000; Wiley and Lieberman, 2011; Yang et al., 2022). These ongoing revisions can lead to inconsistencies, further complicating identification. Beyond biological complexity, industrial processing introduces additional challenges. The timber industry often groups multiple species under broad trade names based on shared functional properties rather than strict taxonomic distinctions (Chudnoff, 1984; Mark et al., 2014). While practical for trade, this practice introduces additional ambiguity, making it difficult to identify the exact species composition of a given timber product (Mboma et al., 2022). Currently, hundreds of tree species are commercially traded worldwide, with an even greater number used locally, highlighting the immense diversity that wood identification systems must account for (Chudnoff, 1984; Council and Organization, 2012; Mark et al., 2014; Richter and Dallwitz, 2000; tropicaux, 1979). Those challenges underscore the need for solid reference databases of the diagnostic features, that encompass enough reliable reference specimens to consider biological variations.

It is the paucity of large databases that cover the variability of diagnostic features, which is the main obstacle currently faced when developing wood identification methodologies (Cody et al., 2005; Deklerck, 2019; Dormontt et al., 2015; Ravindran et al., 2018; Silva et al., 2022). This paucity of large-quality datasets

stems from the difficulty of acquiring sufficient wood specimens that give a faithful representation of all species and their variability. Those wood specimens can be gathered by collecting specimens in targeted field expeditions, active timber harvest sites, lumber mills or other sites in the field. While such endeavours may faithfully capture the current data distribution, they can be logistically challenging and expensive to accomplish at large scale. A second source of information is institutional wood collections that have the advantage of having specimens readily available and that are, in some cases, the result of century-long collecting efforts. This makes them fit for rapidly building reference databases by extracting different types of diagnostic features. However, relatively few wood collections meet the essential criteria for establishing a robust reference database, particularly regarding collection size and specimen reliability. Most collections include only a limited number of specimens per species, typically focusing on vouchered samples verified by expert botanists based on traits (e.g. leaves, fruits, flowers, roots). While these specimens are representative of the species as a whole, they often fail to capture the full range of variation in diagnostic features that can occur within a species. The vast number of tree species makes it difficult to capture the full variability of wood anatomy with limited numbers of specimens per species. Among the most extensive and scientifically valuable wood collections is that of the Naturalis Biodiversity Centre in Leiden, The Netherlands. With a long-standing history in wood anatomical research, Naturalis houses the world's largest scientific wood collection, comprising approximately 125,000 specimens representing tree species from across the globe (Naturalis Biodiversity center, 2025). The second-largest collection is maintained by the USDA Forest Service at the University of Wisconsin–Madison, United States. This collection contains over 103,000 specimens, including approximately 25,000 samples with corresponding herbarium material stored in the Wisconsin State Herbarium (University of Wisconsin–Madison, 2025). Another important reference for wood identification is the Thünen Institute for Wood Research in Hamburg, Germany. Serving as a centre of competence on the origin of timber, wood samples can be determined at the genus or species level, and the geographical origin of the wood can be determined for various tree species (Johann Heinrich von Thünen-Institut, 2025a). Forensic research is based on the scientific wood collection of the Thünen Institute encompassing 35,000 wood samples from 11,300 species (Johann Heinrich von Thünen-Institut, 2025b). The third largest is the Tervuren Wood Collection of the Royal Museum for Central Africa (RMCA, Belgium), founded in 1898 to demonstrate the importance of African tropical timber for economic purposes. During the first half of the 20th century, the economic purpose has been gradually extended with a much broader scientific interest. Not only tropical species and lower taxa (subspecies and varieties) with commercial value but also any tropical African tree species and lower taxa that could be of interest in comparative wood anatomy or for the study of ethnographic objects were collected. From the middle of the 20th century and onwards, wood specimens from other continents were also incorporated in the collection (Beeckman, 2007, 2003; RMCA, 2019). Today, the wood collection has become the Belgian scientific reference collection for wood, containing ca. 81 000 specimens from 13 533 species and lower taxa with accompanying microtome sections, ca. 20 500 sets of thin sections in the three principal directions (Beeckman, 2007; Deklerck, 2019; RMCA, 2019). Most of the species and lower taxa are represented by multiple samples, each from a different specimen. The Tervuren Wood Collection holds 26 604 specimens of DRC tree species and lower taxa, which encompasses 30% of the total collection, thereby offering the most complete collection of reference material for wood identification of >2000 woody species and lower taxa from the DRC (timber trees, small trees, shrubs, dwarf shrubs and lianas) (Beeckman, 2007). The Tervuren wood collection presents a unique opportunity to develop a robust reference database for wood identification, essential for combating illegal logging in the DRC. By leveraging its vast database of tree species and lower taxa with potential timber applications, the collection supports efforts to protect the Congo Basin—one of the world's most critical carbon sinks and biodiversity hotspots—thereby contributing to climate change mitigation and sustainable forest management.

1.3.3 Unexplored areas in literature regarding macroscopic cross-sectional identification applications

To facilitate wood identification in species-rich contexts, dichotomous and multi-entry keys offer a straightforward means of interpreting wood anatomy without requiring users to memorize distinct diagnostic features across a wide range of timbers. These tools guide users through the anatomical

assessment process by narrowing down potential species based on observed traits (Brazier and Franklin, 1961; Ilic, 1993; Richter et al., 2017). Available in both printed and digital formats, they provide a structured pathway from standardized anatomical features to a shortlist of likely taxa, improving field accessibility (Barefoot and Hankins, 1982; Gregory, 1980; LaPasha and Wheeler, 1987; Malaysian Timber Council, 2018; Vander Mijnsbrugge and Beeckman, 1992). Although their simplicity contributes to their popularity, keys also have clear limitations. Their effective use still depends on a foundational understanding of wood anatomy, as users must accurately recognize diagnostic features—limiting their scalability. They often fail to accommodate intra-specific anatomical variation and do not fully exploit the richness of anatomical data. Categorical feature states (e.g., present, variable, or absent) are inherently subjective, with thresholds that lack standardized definitions. For example, what one user considers a ‘present’ feature may be scored as ‘variable’ by another, leading to divergent identification pathways and potentially excluding the correct species early in the process. While multi-entry keys offer more flexibility than dichotomous ones, they still risk generating misleading results if the underlying database does not adequately capture natural variation. More critically, the limited number of macroscopic diagnostic features constrains the ability of keys to distinguish among species in taxonomically diverse or morphologically convergent groups. Combined with observer bias, this scarcity of features can result in partial or inaccurate identifications. Despite their widespread use, no systematic, quantitative assessment has been conducted to evaluate the real-world accuracy of macroscopic features for identification in species-rich contexts. Empirical evaluation is crucial to objectively determine the resolution enabled by applied methodologies.

In this context, raw visual information retains diagnostic patterns that are lost when converting the anatomy into the expert-defined, standardized codified features. This has long been leveraged in pure visual keys, which offer intuitive, user-friendly tools by presenting curated reference images for direct comparison (Kirchoff et al., 2008). These visuals allow users to identify wood based on observable traits without requiring extensive anatomical expertise. Building on this foundation, the growing need for rapid, accurate, and scalable wood identification has accelerated the development of fully automated systems using AI (Hwang and Sugiyama, 2021; Silva et al., 2022). Despite advances, CV-based wood identification remains an evolving field, challenged by the biological variability of wood and the need for large, diverse datasets. Given these challenges, it is essential to evaluate the performance of CV-based methods—especially in species-rich, high-diversity contexts—and to compare them with traditional approaches based on expert-defined anatomical features. Such comparisons are critical to understanding their relative strengths, limitations, and potential for integration into practical identification workflows.

Wood identification presents some unique CV-related challenges that are rarely encountered in other domains. While advancements in CV have improved model robustness to variations in lighting, resolution, and image quality (e.g., blur) (Shorten and Khoshgoftaar, 2019), wood often exhibits physical anomalies that obscure key anatomical features and complicate classification (Goodell and Nielsen, 2023; Niemz et al., 2023a; Schmidt, 2006). As a biological material, wood is subject to degradation from disease, infestation, and physical stress. Insects and marine borers can damage wood, by removing wood material, and fungi and bacteria can cause discoloration and decay (Goodell and Nielsen, 2023; Schmidt, 2006). Furthermore, wood can crack, especially during drying (Niemz et al., 2023a). These alterations can obscure diagnostic structures, hinder DNA extraction, and even change wood chemistry, thereby complicating both visual and laboratory-based identification. However, the impact of these anomalies on CV-based classification remains largely unexplored. Owens et al. (2024) is the only study to date that systematically tested how CNN predictions are affected by digital perturbations mimicking real-world wood degradation (Owens et al., 2024). Most other work has relied on defect-free specimens (Hwang and Sugiyama, 2021; Ravindran et al., 2021; Silva et al., 2022), overlooking the imperfections typically encountered in applied contexts. To ensure reliable field deployment, it is essential to evaluate how such anomalies influence model predictions—and to develop mitigation strategies that improve classification resilience under realistic conditions.

Beyond the challenges presented of anomalies on wood, model design—particularly the classification strategy—also plays a critical role in the performance and interpretability of CV-based wood identification. Despite its importance, this aspect remains understudied. Most existing approaches rely on a single

strategy: multiclass classification, in which CNNs assign each image to one of a fixed set of predefined labels, such as species or commercial timber names (Ravindran et al., 2021, 2018; Silva et al., 2022). However, this method assumes a closed set, limiting recognition to species included in the training data and struggling with unknown samples in real-world applications (Sünderhauf et al., 2018; Wilber et al., 2013). While adding an "unknown" class can help, it remains an imperfect solution (Entezari and Saukh, 2020; Geifman and El-Yaniv, 2019). Moreover, these models require balanced datasets, yet collecting diverse, high-quality timber specimens is costly and time-consuming. As a result, most studies use small datasets with limited species diversity, reducing model generalizability (Hwang and Sugiyama, 2021; Silva et al., 2022). This makes it risky to assume CNN-based wood identification can be directly applied in the field, as classifying between learned timbers is not the same as identifying or verifying the timber species of field samples. To overcome this limitation, other domains, such as facial and vehicle recognition, have adopted open-world approaches like object re-identification (Kumar et al., 2020; Schroff et al., 2015). Instead of assigning fixed labels, these networks compute similarity between images, allowing comparisons against a reference database (Yoshihashi et al., 2019). By embedding images into a feature space where distances reflect species similarity, object re-identification provides a more flexible solution for handling novel timbers (Chen et al., 2017; Ghosh et al., 2023; Ye et al., 2021). Despite its promise, this approach remains underexplored in wood identification (Hwang and Sugiyama, 2021; Silva et al., 2022).

Despite ongoing advances in both identification keys and CV approaches for wood identification, these methods have largely developed in parallel, with no current system effectively integrating their complementary strengths. Both rely on macroscopic anatomical features, yet they extract and interpret these features in fundamentally different ways. Deep learning models autonomously learn complex visual representations through optimization, often identifying patterns that differ from those traditionally recognized by wood anatomists. This divergence is not a limitation but a potential advantage: expert-defined anatomical features may provide structured, interpretable information that complements the data-driven outputs of CNNs. Integrating these perspectives offers a promising path to improve both the accuracy and transparency of automated wood identification.

1.4 Structure of the PhD

This research bridges key gaps in wood identification by integrating traditional macroscopic anatomical assessment with state-of-the-art deep learning techniques. By doing so, it contributes to the development of more effective, scalable, and field-deployable solutions for combating illegal logging.

Chapter 2 – SmartWoodID: an image collection of large end-grain surfaces to support wood identification systems

To address the urgent need for rapid and accessible wood identification in the field—particularly in key countries like the DRC, where illegal logging poses a significant threat—we developed SmartWoodID, the largest reference database of sanded macroscopic cross-sectional images of local timber species, accompanied by descriptions of observable (expert-defined) wood anatomical features for wood identification. Chapter 2 describes the development of a comprehensive reference database to support rapid wood identification in the DRC. This effort leverages the extensive Tervuren wood collection, ensuring the inclusion of a vast number of DRC timber species, selecting species based on national and international timber trade relevance and standardized nomenclature using the World Checklist of Vascular Plants [202]. To account for natural variability within species, multiple specimens were digitized per species, prioritizing those with large end-grain surfaces to capture intra-individual variation. Specimens exhibiting natural anomalies (e.g., cracks, fungal damage, insect activity) were intentionally included for later evaluation of their impact on classification accuracy (chapter 4).

Chapter 3 – Evaluating Expert-Defined Anatomical Features for rapid identification of Congolese tree species

General Hypothesis: *Macroscopic anatomical features, as encoded in standardized expert-defined descriptors, retain measurable discriminatory power for taxonomic identification across taxonomic ranks (family, genus, species)—even in highly diverse tropical datasets.*

Despite their widespread use in field identification and identification keys, the diagnostic resolution of macroscopic features has not been systematically evaluated in species-rich contexts, such as the DRC. Chapter 3 addresses this gap by applying clustering and classification techniques to the SmartWoodID dataset. The goal is to quantify how well these 31 features distinguish among taxa at various taxonomic levels and to identify feature groupings that drive taxonomic separation. This chapter provides a baseline for comparison with more automated approaches.

Individual research questions:

- To what extent can *hierarchical agglomerative clustering* of the 31 expert-defined macroscopic feature codes delineate statistically distinct groups of DRC timber species?
- Which individual macroscopic features exhibit statistically significant associations with specific species clusters, and what is the strength of these associations?
- Are there recurrent groupings of macroscopic features that consistently co-occur within particular clusters, and can these be formalised into diagnostic feature sets?
- What is the classification accuracy at the family, genus, and species levels when species-level predictions are made, and their corresponding higher taxa are inferred and compared to the ground truth?
- How does the accuracy of the above hierarchical inference approach compare with that of models trained directly at each taxonomic rank (family, genus, species)?
- What is the minimum k in top- k predictions required to achieve $\geq 95\%$ correct identification at all taxonomic levels for the evaluated specimens?
- To what extent can species-level prediction accuracy be improved by constraining predictions using the output of a separately trained family-level classifier?
- Do the expert-defined macroscopic features annotated in SmartWoodID enable accurate classification of commercial timber species in other databases (e.g., InsideWood), and what are the measurable limits of this transferability?
- Can expert-defined macroscopic features reliably discriminate among morphologically similar, closely related species (for example African *Pterocarpus* spp.)?

Chapter 4 – Evaluating the effect of anomalous images on CV-based wood identification models

Hypothesis: *Physical anomalies common in real-world wood samples—such as cracks, discoloration, insect damage, and fungal decay— affect CNN-based classification accuracy by obscuring key diagnostic regions, and models trained on clean images may not generalize to damaged ones.*

This chapter investigates whether CNN performance degrades when confronted with anomalies that can be encountered on wood in the field and whether training on damaged or mixed-image sets improves resilience. Using Grad-CAM, model attention is visualized to determine whether CNNs rely on diagnostically meaningful regions or are misled by noise introduced by degradation. The findings provide information on the importance of this aspect regarding database construction for CV-based models by simulating the imperfect conditions typical of field and enforcement scenarios.

Individual research questions:

- What are the observed classification metrics—accuracy, precision, recall, and F1-score—of a pretrained, field-validated CNN in verifying the presence or absence of anomalies in macroscopic images of DRC timber specimens?

- Do Class Grad-CAM visualisations show consistent spatial correspondence between model-activated regions and visible anomalies in anomalous images, and can these correspondences explain specific false-positive and false-negative predictions?
- How does classification accuracy differ between models trained on (i) anomaly-free images only, (ii) anomalous images only, and (iii) a combined dataset, when evaluated on independent test specimens at both patch level and specimen level (using majority voting)? Do these differences vary systematically between anomalous and anomaly-free patches?
- How do classification performances under the three training scenarios vary across individual genera, and are certain taxa more susceptible to performance degradation caused by anomalies?
- For a genus-level model trained exclusively on anomaly-free images, which regions are activated according to Grad-CAM visualisations when classifying anomalous samples, and do these activations target non-diagnostic areas? Does incorporating anomalous images into the training dataset significantly improve both the alignment of attention with diagnostically relevant regions and overall classification performance?

Chapter 5 – Expert-defined vs. Deep Learning Approaches for Scalable and Generalizable Timber Identification

Hypothesis 1: *Deep learning models trained on raw images outperform classifiers trained on expert-defined anatomical features when both are applied to the same macroscopic specimens.*

In this chapter, classification models trained on 31 expert-coded anatomical features are directly compared with multiclass CNNs trained on the same Congolese species. This parallel analysis evaluates whether manually defined features can match or augment CNN performance and lays the foundation for integrating both approaches in Chapter 6.

Individual research questions:

- How do the classification metrics—accuracy, precision, recall, and F1-score—differ between (i) machine learning classifiers trained on 31 expert-coded anatomical features and (ii) CNNs trained directly on macroscopic images of the same specimens?
- Do differences between class-specific F1-scores vary across individual genera?

Hypothesis 2: Closed-world Comparison (SmartWoodID)

Open-world identification strategies, such as object re-identification via triplet learning and binary verification, provide improved classification flexibility compared to traditional closed-world multiclass CNN classifiers when applied to the same species used during training.

Individual research questions:

- How do performance metrics (accuracy, precision, recall, F1-score) compare between (i) closed-world CNN multiclass classification and (ii) open-world CNN methods, including binary verification and triplet-learning-based object re-identification?
- Do differences between class-specific F1-scores vary across individual genera?
- Which identification techniques demonstrates the smallest value of k in top- k predictions required to achieve $\geq 95\%$ correct identification?

These methods allow pairwise or similarity-based reasoning, better capturing intra-class variability and providing more robust predictions under real-world conditions where data imbalance and taxonomic convergence challenge multiclass architectures.

Hypothesis 3: Open-world Generalization (non-Congolese Timber Dataset)

Open-world identification strategy generalizes to unseen species.

By embedding novel samples into a learned feature space or comparing them to reference images, these approaches can offer scalable and adaptive solutions for timber identification across geographic regions, as demonstrated by performance on a separate dataset of non-Congolese timbers.

Individual research questions:

- How does the performance of binary verification—measured by accuracy, precision, recall, F1-score, the minimum k for $\geq 95\%$ correct identification, and the area under the top- k curve (AUC)—differ between genera included in training and novel genera not seen during training?
- How does classification performance vary when the number of possible candidate genera in the search space is systematically increased, thereby simulating progressively more challenging identification scenarios?

Chapter 6 – Combining expert-defined and CV-extracted macroscopic cross-sectional features for enhanced wood identification

Hypothesis: *Combining CNN-derived predictions with expert-defined anatomical data through a re-ranking or fusion model enhances the accuracy of automated wood identification systems.*

This integrative chapter tests whether a hybrid pipeline can leverage the strengths of both CV and human-curated anatomical knowledge. CNN outputs from Chapter 5 are re-ranked using a RF model trained on expert-defined features from Chapter 3. This hybrid framework aims to reduce false positives, improve species discrimination, and increase trust in model decisions—particularly for users in enforcement and forensic contexts.

Individual research questions:

- How do performance metrics—accuracy, precision, recall, and F1-score—change when re-ranking the top- k predicted genera from a binary verification CNN using a random forest (RF) model trained on expert-defined anatomical features? Performance will be assessed by visualizing metric distributions per genus for varying k values and aggregating them into boxplots to observe progression.
- How does re-ranking affect the classification performance of individual genera, as measured by changes in per-class precision, recall, and F1-scores?
- What is the optimal k value for which re-ranking consistently improves performance while minimizing adverse effects such as misranking correct predictions?
- Which factors explain genus-specific increases or decrease in performance after re-ranking? Potential factors include (i) differences in visual field coverage between CNN image features and expert-defined descriptors, (ii) reduction of morphological information during feature extraction, and (iii) taxonomic convergence or “look-alike” taxa within the top- k predictions.

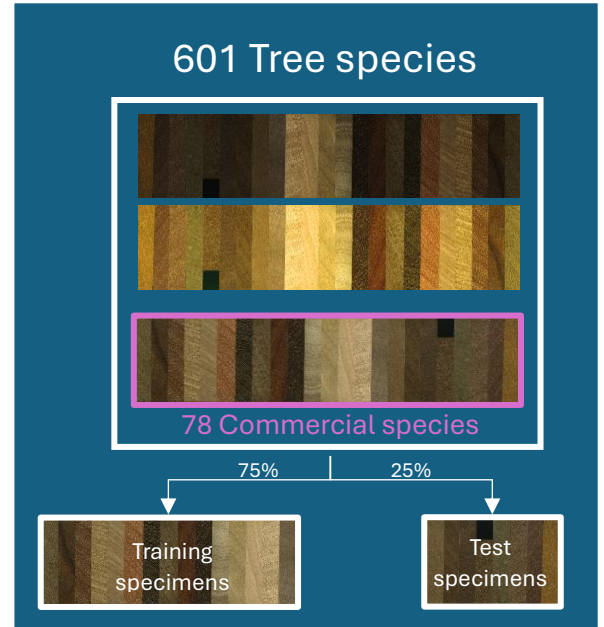
An overview of the data and methodologies used in each chapter and in function of each hypothesis is shown and described underneath.

SMARTWOODID



>1 specimen/species

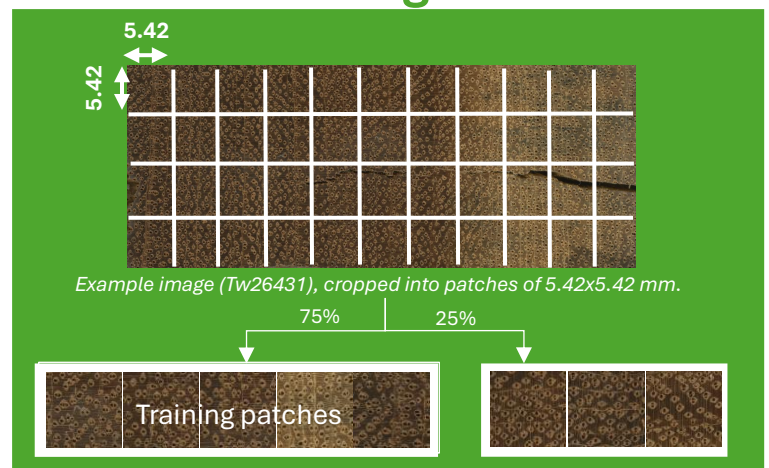
Dataset



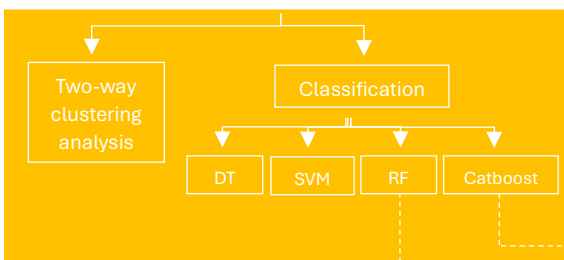
Annotations

Thirty-one annotated expert-defined anatomical features, coded as present, absent, variable or NA (could not be discriminated with certainty)

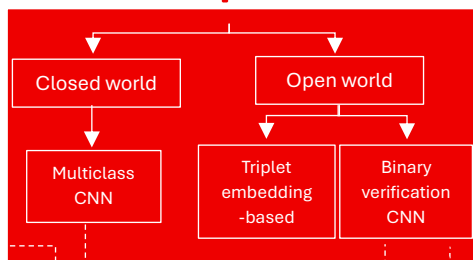
Images



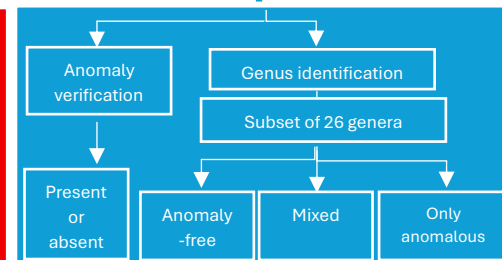
Chapter 3



Chapter 5



Chapter 4



Chapter 6

Re-ranking top k predicted genera by Binary verification CNN using expert-defined features and RF. This was evaluated only on subset of commercial timbers.

Separate image dataset of non-Congolese timbers

Hypotheses

Hypothesis ch. 3

Hypothesis ch. 5.1

Hypothesis ch. 6

Hypothesis ch. 5.2

Hypothesis ch. 5.3

Hypothesis ch. 4

1.5 Structure of the data in each chapter

This study is based on the SmartWoodID database, a digitized subcollection of the Tervuren xylarium housed at the RMCA, Belgium. SmartWoodID serves as a valuable resource for examining the relationship between macroscopic cross-sectional wood anatomy and the botanical diversity of Congolese tree taxa (De Blaere et al., 2023). The database contains high-resolution RGB scans of the macroscopic end-grain surfaces of 3,742 wood specimens, representing 954 species native to the DRC. Each specimen was prepared by scanning the cross-section at 2400 dpi using a flatbed scanner. This resolution allows for the visualization of macroscopic features essential for wood identification.

Each image is annotated with macroscopic IAWA features, which are observable at this resolution and assigned standardized feature numbers (NS, 1989; Ruffinatto et al., 2015). Features were recorded as Present (clearly visible), Variable (sporadically observed), Absent (below the threshold for Variable), or NA (undiscernible due to ambiguous visual cues or resolution limitations). Descriptions of growth rings were excluded, as these were often not discernible with sufficient certainty at the available resolution. Species and lower taxa are represented by multiple specimens, capturing both intra- and interspecific anatomical variation. This makes the database well-suited for studying wood identification using macroscopic anatomy. A complete overview of the database is provided in Chapter 2 and in De Blaere et al. (2023) (De Blaere et al., 2023), while Supplementary Materials Table 8.1 all unique specimen identifiers and metadata. To enable machine learning analysis, we selected only species represented by at least two specimens. This set comprises 2,296 digitized specimens across 601 species, 286 genera, and 64 families. Discriminatory power was mainly assessed by training and evaluating classification models on the specimens. Therefore, specimens were allocated random to training (75%) or test set (25%), while preserving distribution of species across both sets. Within both sets, a subset of 78 commercially important species was defined for targeted evaluation. Both the full set and the commercial subset are used throughout the analyses in this thesis unless specified otherwise.

Chapters 3 through 6 each build on this core dataset to evaluate different approaches to wood identification. Chapter 3 assesses the discriminatory power of expert-defined anatomical features as recorded in the SmartWoodID database. Chapter 4 to 5 focuses on training CNNs directly on macroscopic images for genus-level classification. Given the large surface area of the scanned images, training directly on full-resolution inputs is both computationally intensive and prone to overfitting due to the limited number of images. To mitigate this, we extracted non-overlapping patches of 512×512 pixels from each full image, corresponding to a physical area of 5.42×5.42 mm (see Figure 1.4). This patch size is sufficient to capture key diagnostic features while also reflecting natural variability between different regions of the same specimen.



Figure 1.4: An example of a typical image in the SmartWoodID collection (Tw26431), cropped into patches of 5.42×5.42 mm.

Patches were allocated random to training (75%) or validation set (25%), while preserving distribution of specimens across both sets, using different regions of the same specimens for training and validating. These patches allow for efficient model training while preserving macroscopic features relevant for identification and displaying anomalies (damage) on wood at an optimal scale for observation. Examples of different anomalies are presented in chapter 4 (Figure 4.1). For chapter 4 only, 26 genera were selected, to ensure a balanced distribution of damaged and undamaged specimens to ensure fair performance evaluation. Patches were manually inspected and labelled based on the presence or absence of visible damage. Furthermore, any full image yielding fewer than four usable patches was removed, as these would cover less than one square centimetre—below the typical area assessed by a human expert during wood anatomical identification. Chapter 5 extends the CNN approach to the full set of available species and also evaluates the open-world recognition potential on a separate image dataset of non-Congolese genera (for further details see section 5.3.9). Chapter 6 integrates the expert-defined features and CNN-derived image information into a tiered identification system that evaluates if the top genus predictions by a CNN can be refined by re-ranking the top k predicted genera using expert-defined features. This approach employs RF trained on coded anatomical features (chapter 3) and binary verification CNN (chapter 5). The dataset in chapter 6 is constrained to the 78 commercial species, which regards 56 genera.

All CNN models and associated training procedures were implemented using Python (v3.9.15) with TensorFlow (v2.6.0) and Scikit-learn (v1.3.0). Data analysis and visualizations were conducted using Pandas (v1.5.3) and Matplotlib (v3.7.1). Additional statistical analyses, including ecological metrics and heatmaps, were performed in R (v4.4.1) using the ‘vegan’ package (v2.6-8) and the ‘pheatmap’ package (v1.0.12). All computations were executed on a desktop workstation equipped with an Intel Core i9-11900F processor (8 cores, 16 threads, 2.5 GHz) and a single NVIDIA GeForce RTX 3080 Ti GPU.

Chapter 2: SmartWoodID—an image collection of large end-grain surfaces to support wood identification systems

Authors: Ruben De Blaere^a, Kévin Lievens^a, Dieter Van Hassel^a, Victor Deklerck^e, Tom De Mil^d, Wannes Hubau^{a,b}, Joris Van Acker^b, Nils Bourland^a, Jan Verwaeren^c, Jan Van den Bulcke^b, Hans Beeckman^a

a Service of Wood Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium, b UGent-Woodlab, Laboratory of Wood Technology, Department of Environment Ghent University, Coupure Links 653, 9000 Gent, Belgium, c UGent-KERMIT, research unit Knowledge-based, predictive and spatio-temporal modelling, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Gent, Belgium, d TERRA teaching and research center, Gembloux Agro-Bio Tech (Université de Liège), Passage des Déportés 2, 5030 Gembloux, Belgium, e Royal Botanic Gardens, Kew, Richmond, London, TW9 3A, UK

2.1 Abstract

Wood identification is a key step in enforcing laws and regulations against illegal timber trade. Effective wood identification tools, capable of distinguishing a large number of timbers, require a robust reference database. Such reference material is typically curated in botanical wood collections and consists of secondary xylem samples from lignified plants.

This chapter, adapted from De Blaere et al. (2023) (De Blaere et al., 2023), elucidates on the construction of SmartWoodID, a dataset derived from the Tervuren Wood Collection—one of the largest institutional wood collections worldwide and a major reference for DRC tree species with potential timber applications. SmartWoodID contains macroscopic RGB scans of cross-sectional surfaces enriched with expert-defined descriptions of 31 observable anatomical features. These annotated images support both interactive identification keys and CV models for AI-based wood identification.

Since the original publication (De Blaere et al., 2023), a full quality control assessment (Section 2.3.7) was completed to detect and exclude misclassified specimens, resulting in revised counts for specimens, botanical families, genera, and species. A first edition of the database consists of images of 954 taxa, with a focus on potential timber species from the DRC aiming to include at least four different specimens per species included.

2.2 Introduction

Illegal logging significantly impacts forests, posing a high risk of irreversible damage, particularly when exploiting populations of protected species. Thirty to ninety percent of traded tropical timber is estimated to have been harvested illegally (Hirschberger, 2008; Hoare, 2015; Magrath et al., 2009). Fast and accurate wood identification systems are key to properly enforce timber regulations (such as FLEGT, EUTR, U.S. Lacey Act, the Illegal Logging Prohibition Act in Australia) and CITES by verifying whether the traded species matches the species name on accompanying documents (Gasson et al., 2021; Piabuo et al., 2021).

The most commonly used and affordable method for wood identification is the wood anatomical assessment. It involves observing tissues and cells at different scales and planes to identify diagnostic features of the botanical taxon. Macroscopic cross-sectional wood anatomy provides an accessible range of diagnostic information, observable with a hand lens, for faster identification in the field (Koch et al., 2018; Ruffinatto et al., 2015) compared to other microscopic features, requiring a controlled laboratory environment with specialized equipment.

The macroscopic cross-sectional wood anatomy can be used for identification using different methods of feature extraction. Visual keys provide users with a list of exemplary images to compare with, relying on the expertise of the user with wood species. A secondary method aligns more closely with the traditional wood anatomical assessment, relying on standardized expert-defined features to ensure consistent predictions (Angyalossy et al., 2016; Committee, 2004; Gasson et al., 2011; NS, 1989; Ruffinatto et al., 2015; Wheeler, 2011). This second method can be applied in tandem with classification keys, allowing for the observation of anatomical features and provide a list of matching species. These classification keys are advantageous for their speed and flexibility, with some allowing for a specified number of feature mismatches or required presence/absence of certain features (LaPasha and Wheeler, 1987; Vander Mijnsbrugge and Beeckman, 1992). Keys can be accessed online with large reference material or offline, suitable for remote locations such as local lumber mills in the tropics.

CV-based wood identification is a third method for rapid identification in the field, that effectively automates the process of wood anatomical assessment. CV is a field of AI that trains computers to interpret and understand the visual world, relying on machine learning algorithms that use vast numbers of human-annotated reference images to distinguish timbers based on imagery (Bay et al., 2006; Hwang and Sugiyama, 2021; Lowe, 2004). CV-based wood identification tools have demonstrated their potential for real-world field deployment (e.g. evaluated in Ghana (Ravindran et al., 2019)), enabling non-expert field workers to perform timber tracking. This technique has the advantage of fast and easy application but faces risks due to the highly variable nature of wood, which exhibits inter- and intra-specific variability and anomalies like cracks, insect holes, and fungal damage that can hinder the recognition of wood features (Figure 2.1). This makes it challenging to train machine learning models for field-based wood identification.

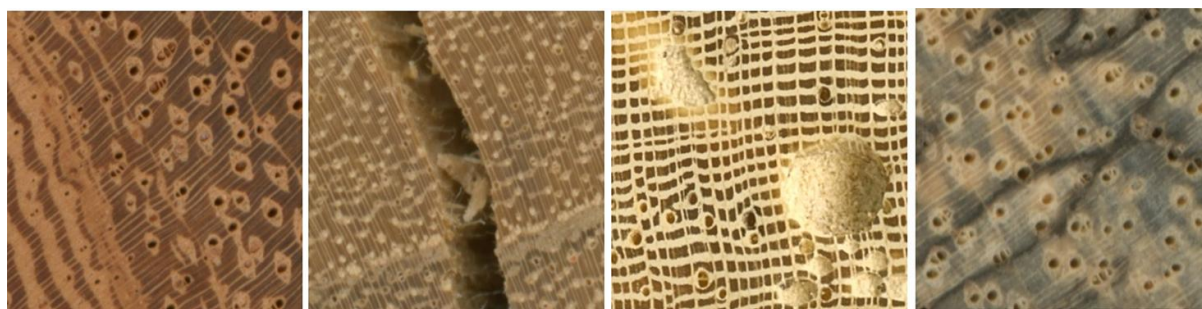


Figure 2.1: Examples of the intra-variability and anomalies that can be encountered on images of wood (end-grain surface). From left to right: (a) An example of the variability of wood anatomical features (such as axial parenchyma) on a single specimen. (b) An example of a possible anomaly on wood, a crack. (c) An example of a possible anomaly on wood, insect holes. (d) An example of a possible anomaly on wood, fungi damage. RubenDeBlaere@RMCA.

To build accurate identification tools, a database with typical species features and sufficient detail is needed, including information on multiple specimens to consider biological variation. Online databases,

such as macroHOLZdata (Richter et al., 2017) and the Atlas of Macroscopic Wood Identification (Ruffinatto and Crivellaro, 2019), provide macroscopic anatomical descriptions of wood, but they may not cover all intra-species variability of anatomical features that can occur in wood. The most complete online database for timber identification is InsideWood, a wood anatomy reference, research, and teaching tool, containing wood anatomical descriptions of wood based on the IAWA Lists of Microscopic Features for Hardwood and Softwood Identification accompanied by a collection of photomicrographs (NS, 1989; Wheeler, 2011; Wheeler et al., 2020). This database has a global scope and therefore incorporates timbers from all over the world, having over 9.400 wood anatomical descriptions of fossil and modern woody dicots, representing over 10.000 species and 200 plant families, accompanied by over 50.000 images of both microscopic and macroscopic features (Wheeler et al., 2020). Still, while this database serves a key purpose as a reference resource, this does not mean that its descriptions cover all intra-species variability of anatomical features that can occur in wood. Wood is variable, and requires descriptions of large areas on multiple specimens. Additionally, although this database is relatively large, certainly not all woody species are represented with a sufficient number of individuals. It is the paucity of large databases, that cover the variability of wood anatomical features, which is the main obstacle currently faced when building classification keys or machine-learning models (Hwang and Sugiyama, 2021; Ravindran et al., 2018). This paucity of large quality datasets stems from the difficulty of acquiring sufficient wood specimens that give a faithful representation of all species and their variability in a geographically delineated area.

We built the first edition of SmartWoodID, an image database of end-grain wood that includes macroscopic features and anomalies, such as cracks, fungi damage, and insect damage, and their variability along a radial gradient (i.e. a gradient from pith to bark). This information enables evaluation of the diagnostic power of macroscopic wood anatomy under visible light, comparing traditional expert-defined features with computer vision-extracted features for field-based wood identification. This first edition focuses on tree species from the DRC and serves as annotated training data for developing classification keys and AI for CV-based wood identification. SmartWoodID will be gradually extended with images of timbers from other continents in the coming years. The resulting database can also provide unique insights into the occurrence of characteristics, for example within families.

2.3 Material and methods

2.3.1 Delineating a set area for the first edition

Given the complex and variable structure of a wooden tissue and the large number of tree species worldwide, digitizing a large amount of wood specimens is a long-term process. The SmartWoodID database will therefore gradually be extended over the following years with images and annotated materials from wood species from all over the world. This will be done in several editions containing images and annotated materials from large geographically delineated regions, to ensure that the data are available and usable for research on entire biomes rather than adding species ad hoc.

The first edition of the database focuses on the tree species of Central Africa and more specifically the tree species from the DRC. The following definition of trees is used here: perennial woody seed plants with a single dominant stem that is self-supporting and undergoes secondary growth. The DRC was selected because the vast area of the country and the different forest biomes make the DRC rich in tree species and thereby representative of species richness for all countries in the Congo Basin and ensure that many timbers or potentially commercial tree taxa of tropical Africa are included. An overview of the procedure, explaining the material and methods, is shown as a flowchart in Figure 2.2.

SMARTWOOD ID

Delineating a set area for the first edition

A first edition of the SmartWoodID database focuses on the tree species of Central-Africa, and more specifically the tree species from the Democratic Republic of the Congo (DRC).



Species selection

The species were selected from the *Liste des essences forestières de la République Démocratique du Congo (2017)*, enumerating all timber producing species in the DRC.

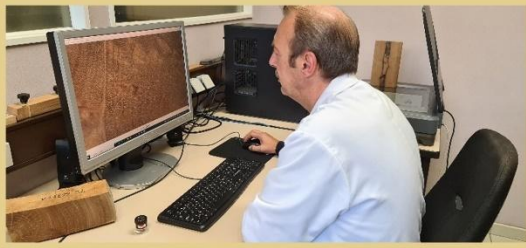
Collecting the specimens

Material of the Tervuren Wood Collection is used to create a database with high-resolution scans of the endgrain surface of 1190 species along with expert wood anatomical descriptions.



Scanning

The sanded surfaces of the samples are scanned on a flatbed scanner at high resolution (2400 dpi). A typical image (TIFF file of 80 MB) will cover the wood anatomical end-grain structure of a surface of about 7 cm long and 1-2 cm wide.



Sanding

The end-grain surfaces of the samples are sanded to ensure that the anatomical features are visible.

Annotating

The scans are visually analysed by experts in wood anatomy of the service of Wood Biology at the Royal museum for Central-Africa. The presence of 29 standardized macroscopic features is checked as well as the presence of anomalies (like fissures, traces of insect, fungi, etc.)

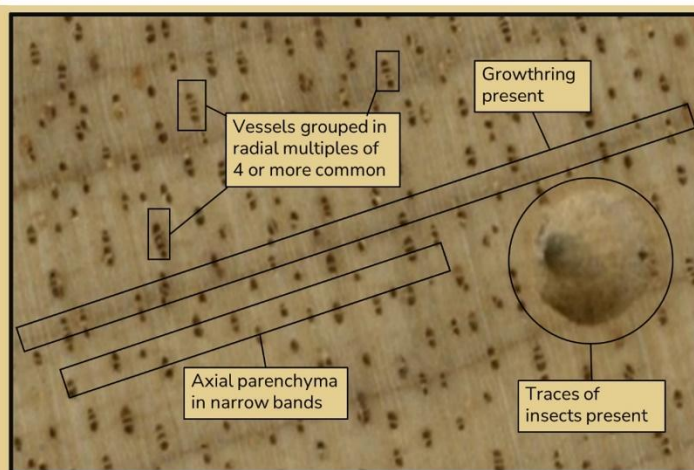


Figure 2.2: A flowchart showing the procedure of building the database.

2.3.2 Species selection

A list with accepted species names has been created according to the World Checklist of Vascular Plants (WCVP) (WCVP, 2022) and the African Plant Database (APD) (APD, 2012), providing information on accepted name status and synonymy. In this research, we will regard not only species but also accepted varieties and subspecies. All instances of species, varieties and sub-species shall be named ‘species and lower taxa’ from this point onwards to improve the ease of reading.

Two lists were used as a reference for all current and potential timbers in the DRC. The first is the list of the DRC forest administration *Direction Inventaire et Aménagement Forestiers (DIAF)*, summarizing all tree species and lower taxa present in DRC forests, along with an indication of their current economic value (DIAF, 2017). The second list is extracted from the RAINBIO database, from which all tree species and lower taxa were selected that occur in the DRC (Dauby et al., 2016). The accepted name status of species and lower taxa names was cross-referenced and harmonized with the WCVP (WCVP, 2022) as a reference, using a custom-developed Python script. The number of species and lower taxa for which no direct match to WCVP was found was checked against the APD (APD, 2012), a curated list of >205 456 names of African plants with their nomenclatural status being a product of a collaboration between the South African National Biodiversity Institute, the *Conservatoire et Jardin botaniques de la Ville de Genève, Tela Botanica* and the Missouri Botanical Garden. Taxa that did not match WCVP or APD were reviewed manually, and any misspellings or synonyms that had not been automatically detected were corrected. Having standardized taxonomic names, any records from species and lower taxa that did not meet our working definition of trees—perennial woody seed plants with a single dominant stem that is self-supporting and undergoes secondary growth—were manually removed from the database. This included removing all ferns, palms, lianas, strangler figs, bamboos, pandans, as well as a number of shrub species and lower taxa that rarely exceed 2 m in height and are generally multi-stemmed. Finally, the accepted name of each species and lower taxon was used to check their presence in the Tervuren Wood Collection. The list contains also introduced species and lower taxa.

The IUCN Red List of Threatened Species (IUCN, 2021) was used to add information on Red List Categories and population trends. This was done to give an overview of the threatened tree species and lower taxa in the DRC and give a perspective on the threatened nature of commercial timbers, which are provided by the indication of the economic value of those species and lower taxa.

Information on the occurrence of those species and lower taxa in different vegetation types was also added by combining the geographical occurrence data in the RAINBIO database (Dauby et al., 2016) with the geographical distribution of vegetation types in the Global Land Cover Map 2000 (GLC 2000 map) (“Global Land Cover 2000 database. European Commission, Joint Research Centre,” 2003). Twenty-seven different classes are used in the GLC 2000 map to classify African vegetation. These classes were combined into larger classes based on research by Fritz (Fritz et al., 2003) and consist of closed forests, edaphic forests, altitudinal forests, woodlands, shrub lands, savannahs, deserts, water bodies and urban areas. An overview of all classes is given in Table 2.1.

Table 2.1: Summary of Vegetation classes in the Global Land Cover map 2000, merged into broader classes used in this study.

Merged classes	Vegetation class (GLC 2000)
Altitudinal forest	Submontane forest (900 -1500 m)
	Montane forest (>1500 m)
Closed forest	Closed evergreen lowland forest
	Degraded evergreen lowland forest
	Mosaic Forest / Croplands
	Mosaic Forest / Savanna
Desert	Sandy desert and dunes
	Stony desert
	Bare rock

	Salt hardpans
Edaphic forest	Swamp forest
	Mangrove
	Swamp bushland and grassland
Savanna	Closed grassland
	Open grassland with sparse shrubs
	Open grassland
	Sparse grassland
	Croplands (>50%)
Shrubland	Deciduous shrubland with sparse trees
	Open deciduous shrubland
	Croplands with open woody vegetation
Urban areas	Irrigated croplands
	Tree crops
	Cities
Water bodies	Water bodies
Woodland	Closed deciduous forest
	Deciduous woodland

The GLC 2000 map of Africa (product two, version 5.0) was imported in QGIS 3.24.3 along with the occurrence data of trees in the RAINBIO database as point vector data and a third layer containing country borders (Esri, 2022). All layers were reprojected to the same coordinate reference system, EPSG:4326—WGS 84. Next, the class of every data point in the RAINBIO database was determined. The information on classes was then added to the SmartWoodID database by counting all occurrences of a species and lower taxon in the RAINBIO database and counting the occurrence of each class. Finally, the classes were merged into larger classes, in order to give an easier overview on the occurrence of tree species and lower taxa in vegetation classes.

2.3.3 Collecting the specimens

The Tervuren Wood Collection of the Royal Museum for Central Africa (RMCA, Belgium) was founded in 1898 to demonstrate the importance of African tropical timber for economic purposes. During the first half of the 20th century, the economic purpose has been gradually extended with a much broader scientific interest. Not only tropical species and lower taxa with commercial value but also any tropical African tree species and lower taxa that could be of interest in comparative wood anatomy or for the study of ethnographic objects were collected. From the middle of the 20th century and onwards, wood specimens from other continents were also incorporated in the collection (Beeckman, 2007, 2003; RMCA, 2019).

Today, the wood collection has become the Belgian scientific reference collection for wood, containing ca. 81000 specimens from 13533 species and lower taxa with accompanying microtome sections, ca. 20500 sets of thin sections in the three principal directions (Beeckman, 2007; Deklerck, 2019; RMCA, 2019). Most of the species and lower taxa are represented by multiple samples, each from a different specimen.

The Tervuren Wood Collection holds 26604 specimens of DRC-tree species and lower taxa, which encompasses 30% of the total collection, thereby offering the most complete collection of reference material for wood identification of >2000 woody species and lower taxa from the DRC (timber trees, small trees, shrubs, dwarf shrubs and lianas). Those aspects of the wood collection create the unique opportunity to provide the robust reference database needed for building classification keys and CV-based wood identification tools by valorising a vast collection of tree species and lower taxa with potential use as timber.

All DRC tree species and lower taxa, present in the Tervuren Wood Collection, are taken from the collection with at least four specimens per tree species and lower taxa. This ensures that variability in wood anatomical features between specimens of the same species and lower taxa is covered by the database. A typical wood collection sample is rarely intact because of the frequent presence of pin holes, traces of fungi attacks, cracks and other mechanical damage, making it difficult to produce clean polished surfaces that show the wood anatomical features without aforementioned anomalies. Specimens in the database that have such damage are not excluded from the database. They are included on purpose to ensure that the CV tools can learn to detect and ignore their presence. A lack of such damaged samples in the training data could cause the machine learning algorithm to explore such anomalies for recognizable and species and lower taxa defining characteristics.

2.3.4 Sanding

The end-grain surfaces of the samples are sanded before scanning to ensure that all features, necessary for determination, are visible. The samples are stacked together with clamping screws to facilitate the process. The parameters of the machinery, more specifically angles and distances between the table, the sanding surface, and the fulcrum, are set to be equal to ensure that every part of the surface is sanded at each grit. The samples are first sanded using a belt sander at one hundred grit to flatten the end-grain surface and subsequently using an eccentric sander. The end-grain of the samples is pressed against the belt sanding surface with the appropriate amount of force at 1-s intervals, to prevent scorch marks, which can hinder the visibility of anatomical features. Similarly, the end-grain surface of the samples is pressed against the eccentric sander while simultaneously performing lateral movements. Samples are sanded multiple times with gradually finer-grade sanding paper with each consecutive grit removing scratches from the previous grit and leaving shallower scratches. The eccentric sanding starts with a fine grade at 100 grit to remove all scratches of the belt sander and ends with an ultra-fine grade at 4000 grit at which point the end-grain surface is free of scratches and all macroscopically visible anatomical features are discernible with the naked eye or a $\times 10$ magnifying glass. At the end of sanding, a magnifying glass is used to check surface quality, and if necessary, both belt and eccentric sanding are repeated if any scratches are still present.

2.3.5 Scanning

The sanded end-grain surfaces are scanned to visualize all macroscopically visible anatomical features. The scanning is performed using an Epson Perfection V750 Pro scanner using the SilverFast Ai Studio Version 9 software package. The scanner is calibrated twice a day with a 10 $\times$ 15 cm reflective Fuji Advanced Colour Calibration Target to ensure consistent results. A resolution of 2400 dpi or ninety-five pixel/mm was used in order to find a balance between storage need and a required resolution for observing all macroscopically visible anatomical features. A bit depth of forty-eight bit was selected to maximize the quantitative information (RGB values) on the natural colour of the wood. A typical image (TIFF file of 80 MB) will cover the wood anatomical end-grain structure of a surface of ~ 7 cm long and 1–2cm wide. The digital images cover more variability compared to sections of the usual size and provide opportunities for building elaborate classification keys and for performing substantial data augmentation (i.e. increasing image variability) for Deep learning.

2.3.6 Annotating (updated after publication)

The resulting images are annotated by anatomical descriptions of the samples based on the list of macroscopic features (NS, 1989; Ruffinatto et al., 2015). Twenty-nine of those standardized features are visible on a typical high-resolution scan of the end-grain surface. These were summarized, along with the matching feature number for the microscopic IAWA features for hardwood identification (NS, 1989) as well as the matching macroscopic feature number (Ruffinatto and Crivellaro, 2019), in Table 2.2.

However, the feature definitions do not translate directly from one system to the other. InsideWood lacks features such as Solitary and radial multiples of 2–3 vessels, Ray visibility to the naked eye on the transverse surface. Vessel diameter classes are defined into different ranges, with InsideWood featuring $<50\mu\text{m}$, 50–100 μm , 100–200 μm , $>200\mu\text{m}$, and the macroscopic feature definitions using $<80\mu\text{m}$, 80–130 μm , and $>130\mu\text{m}$ to be more adept towards the macroscopic resolution. Vessel frequency classification for the macroscopic features pool InsideWood's finer resolution-based categories into a broader state (>20 vessels: 20–40, 40–100, >100). Similarly, vessel arrangements (radial and

diagonal/echelon) were also pooled into a single feature class for SmartWoodID, as SmartWoodID handles both as separate features while InsideWood treats them as a single feature. Each image is labelled with a single state for every anatomical feature. The feature states provide an overall assessment of the frequency at which the feature in question occurs on the specimen. The following states are used: Present/Variable/Absent/NA. Present, can be interpreted as the feature being clearly observable on the end-grain surface. Variable, can be interpreted as the feature being rare, occurring sporadically across the end-grain surface. The concept of Variable differs between SmartWoodID and InsideWood. In SmartWoodID, Variable reflects a low feature frequency across a larger specimen surface, while in InsideWood, Variable indicates feature instability across multiple specimens. This distinction necessitated cautious interpretation, given the lack of specimen-specific resolution and field-of-view details in InsideWood. Absent, can be interpreted as the feature being almost entirely unobservable on the end-grain surface, with too few examples (for example: only 1 or two vessel clusters among all vessels on an area of at least 1cm²) to count as Variable. Finally, NA, can be interpreted as the feature being unable to discern with certainty. The state "NA" is used to describe anatomical features when observations are hampered macroscopic resolution for features that are only a few cells wide, such as thin vasicentric parenchyma or banded parenchyma heights. The states may vary slightly depending on the human annotator as it is a qualitative assessment using the microscopic features as defined in the IAWA list of microscopic hardwood features. Anomalies due to biological or mechanical impact that do not have a diagnostic value are also coded using the same states because they can hamper the identification process by non-experts or automated expert systems. It should be noted that the damage must not be too dominant on the specimen. During the identification process and the process of deriving the anatomical description, we found that specimens for which the damaged area was over two-third, a proper identification was often hard to obtain.

Table 2.2: Summary of described macroscopic cross-sectional features in the SmartWoodID database, showing the structures and refined categories (property/character) along with the matching feature number(s) in the IAWA list for hardwood identification (NS, 1989) and the Macroscopic IAWA feature list defined by Ruffinato et al. (2019) (Ruffinato and Crivellaro, 2019).

Structure	Refined category	Feature	IAWA Description	Item	Macroscopic feature number IAWA
Growth rings	Growth rings	Growth rings distinct	1		1
		Growth ring boundaries indistinct or absent	2		1
Vessels	Porosity	Diffuse porous	5		3
		Semi-ring porous	4		4
		Ring porous	3		5
	Arrangement	Vessels in tangential bands	6		8
		Vessels in radial pattern	7		9
		Vessels in diagonal pattern (echelon)	7		10
		Vessels in dendritic pattern (flame-like)	8		11
	Grouping	Solitary and in radial multiples of 2-3 vessels	-		12
		Exclusively solitary (90% or more)	9		13
		Radial multiples of 4 or more common	10		14
		Clusters common	11		15
	Frequency	≤ 5 vessels per square mm	46		16
		6–20 vessels / square mm	47		17

	Vessel diameter/ pore visibility	> 20 vessels / square mm	48/49/50	18
		Small (not visible to the naked eye, less than 80µm)	-	19
		Medium (just visible to the naked eye, 80-130 µm)	-	20
		Large (commonly visible to the naked eye, larger than 130 µm)	-	21
Axial parenchyma	Distribution	Diffuse-in-aggregates	77	30
		Vasicentric	79	31
		Lozenge-aliform	81	32
		Winged-aliform	82	33
		Confluent	83	34
		Bands more than three cells wide	85	35
		Narrow bands or lines up to three cells wide	86	35
		Parenchyma in marginal or seemingly marginal bands	89	38
		Reticulate	87	39
		Scalariform	88	40
Rays	Visibility	Ray visibility to the naked eye on the transverse surface	-	43
	Width	Rays per millimetre (<= 4 / mm)	114	49
		Rays per millimetre (4-12 / mm)	115	49
Anomalies	visible damage	Rays per millimetre (>= 12 /mm)	116	49
		insect holes	-	-
		fungi	-	-
		Mechanical damage	-	-

The result is a list of 1700 tree species and lower taxa from the DRC, each with a description of the vegetation classes in which they grow, an indication of their commercial value and their threatened status in 2022 according to the IUCN Red List (IUCN, 2021), the CITES appendices (UNEP-WCMC (Comps.), 2022) and the European Union and Trade in Wild Fauna and Flora (Commission Regulation (EU), 2019). Of these 1700 species and lower taxa, 954 species and lower taxa are present in the Tervuren Wood Collection and are used to create images and annotations on the macroscopic anatomical features with at least four specimens available for all species and lower taxa, thereby resulting in 3742 surfaces to scan. The pursued number of four specimens per species and lower taxon was chosen to correspond to the available number of collection specimens in the Tervuren Wood Collection. In addition, it is common practice in wood anatomical assessments to base species and lower taxa descriptions on a relatively small number of specimens.

2.3.7 Quality control

A database with reliable reference material is the backbone of any application to identify a specimen in a taxonomy system. A first important aspect of reliability to address is the need for specimens to be correctly identified. If misidentified, it would cause the interpreter, being either a wood anatomist or a machine learning model, to focus on different distinguishing characteristics for said species and lower taxa, potentially resulting in misidentification. The Tervuren Wood Collection contains specimens that were

collected during field missions. During many field missions, herbarium material was also collected and stored in the collection of the Meise Botanic Garden (Abeelee et al., 2021). Specimens with reliable herbarium vouchers in the Meise Botanic Garden are primarily selected to ensure the reliability of the specimens. When specimens with herbarium material are not available, specimens from dependable collectors are chosen. Next, specimens are compared with descriptions in the InsideWood database to maximally avoid misidentification during annotation. Because the features checked during annotations are all macroscopic for hardwood identification listed by the IAWA and visible on the end-grain surface, the InsideWood database provides the perfect reference tool for checking the occurrence of IAWA features and the correct identification of wood collection samples. An unknown wood specimen can be a species and lower taxon not present in the database (Wheeler et al., 2020). The Tervuren Wood Collection contains multiple samples of most species and lower taxa, and for the image database that we present, four specimens of each species and lower taxon are selected. In order to ensure a good number of high-quality specimens, specimens with a large end-grain surface are preferred as they contain more information. Wide branches and stem disks are also included, if possible, as they have a large end-grain surface along with extra information, like pith and differences between heartwood and sapwood. Twigs and branches are only included if no other specimens are available because the smaller area of the end-grain surface consists of juvenile wood mainly and does not show the diagnostic features used in routine wood identification on the variability of anatomical features. Some of the different macroscopic features can also differ between different parts of the tree, for example, the size of vessels will be substantially larger in the stem compared to branches and especially twigs (Zimmennann and Potter, 1982).

2.3.8 Technical description of the database and functionalities

The specimen-based database with the collected observations is made accessible online by incorporating it in an IIIF environment for presenting and annotating content such as images and audio-visual files (IIIF, 2022a; McAulay, 2017). This framework was selected due to the potential it has for sharing data in a way that allows viewing, comparing, manipulating and annotating images in an environment that is easily accessible. The SmartWoodID database within the IIIF contains new high-resolution scans of wood and accompanying metadata such as geographical origin, accepted taxonomy according to the WCVP, descriptions of their anatomical features, the mean RGB values of intact wood and the density measurements.

The IIIF environment is implemented with the Image API and the Presentation API only, with plans to add the Content Search API in a later stage. The Image API defines how image servers deliver pixels to a viewer, and the Presentation API adds metadata and structures to these images, defining how they appear in IIIF-compliant viewers. This is done through an IIIF Manifest, a JSON file. These JSON files are generated with a custom Python script that fetches all relevant information from the database.

The Manifest file is presented by an IIIF-compliant viewer online. There are several (open source) viewers available, each with its own use case. Since IIIF is all about interoperability, the Manifest file can be reused potentially within different viewers. Mirador (IIIF, 2022b) was selected as the primary viewer as it is an open-source, highly configurable and extensible multi-window image viewing platform that allows researchers to view, zoom, rotate and compare image-based resources, making it both educative and useful for experts and wood enthusiast alike. The viewer is also not limited to view-able specimens but can view and compare any IIIF-enabled resource available, facilitating research across collections and institutions.

In a later stage, annotations will be added to display the macroscopic wood anatomical characteristics in order to visualize them for educative purposes and to potentially include them in identification applications such as classification keys and AI using object identification to recognize and quantify wood anatomical properties.

2.4 Results

2.4.1 Taxonomic coverage (updated after publication)

The database contains 954 tree species, and lower taxa present in the Tervuren Wood Collection encompassing 385 genera and 81 families. In total, 3,774 specimens were examined, of which 3,472 were retained after verification. Among these, 348 specimens had corresponding herbarium vouchers at Meise, with 335 confirmed as correctly identified—approximately 10% voucher coverage. The family with the largest number of species and lower taxa in the database is *Fabaceae*, covering 216 species and lower taxa and 22.7% of all DRC tree species and lower taxa. The fact that *Fabaceae* is the most diverse tree family is not surprising given that *Fabaceae* or *Leguminosae* is the third most diverse plant family after the (primarily herbaceous) families *Asteraceae* and *Orchidaceae* (54). The second and third most occurring families are the *Rubiaceae* and the *Annonaceae*, covering significantly less species and lower taxa with 47 and 43 species and lower taxa and 4.9% and 4.5% of all DRC tree species and lower taxa, respectively. The 12 most occurring families encompass 60% of the 954 species and lower taxa with not <30 species and lower taxa per family. *Diospyros* is the genus with the most species and lower taxa at 2.6% of all species and lower taxa closely followed by *Ficus* at 2.2% of all species and lower taxa.

2.4.2 Geographical coverage

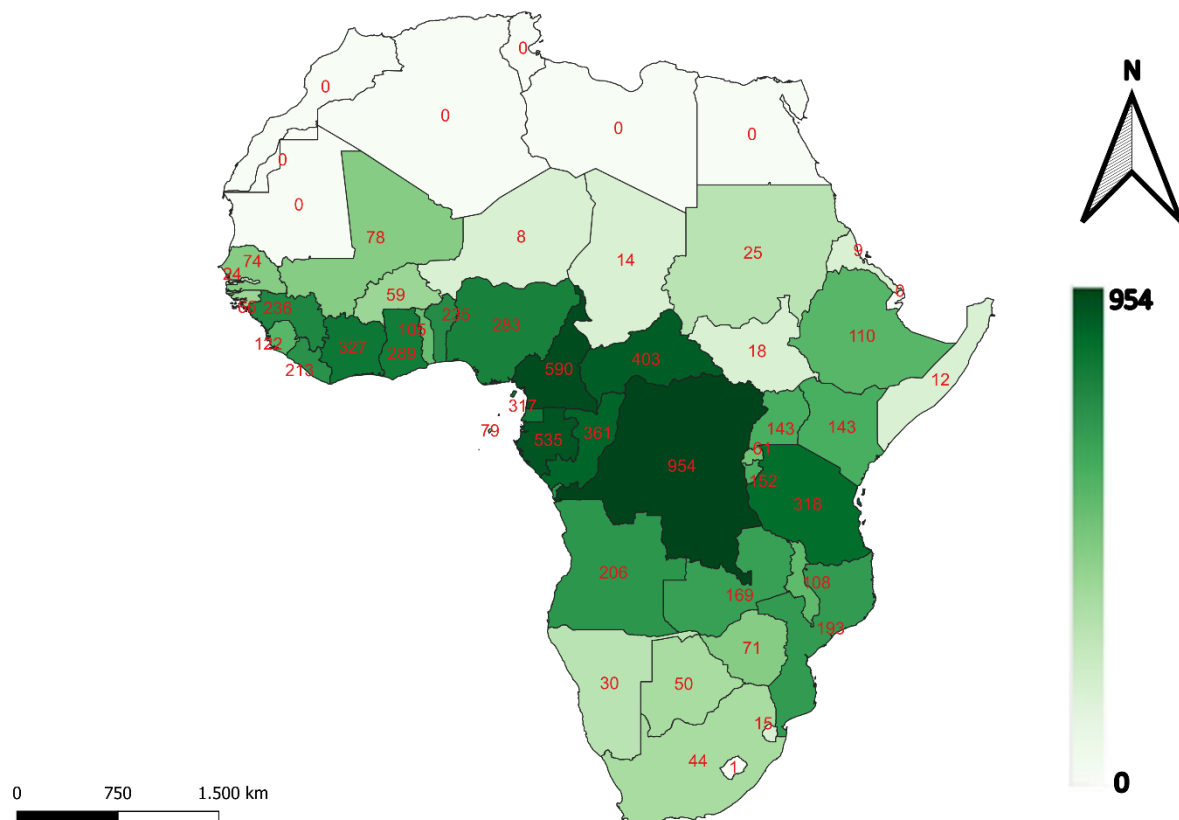


Figure 2.3: A map of the African continent, indicating the number (indicated in red) of DRC tree species present in each country. A darker colour represents a higher tree species richness of Congolese trees.

Figure 2.3 shows in which countries on the African continent the tree species and lower taxa from the SmartWoodID database are growing. The colour intensities represent the number of tropical tree species and lower taxa of the DRC present in that country according to the RAINBIO database, ranging from 0 to 954. There is a gradient moving away from the equator, as less DRC tree species and lower taxa occur further north or south of the continent, which is logical given the tropical boundaries. Given that the DRC is a large country covering a wide spectrum of phytogeographical regions, this obviously results in a large number of vegetation classes (Table 2.3) also present in neighbouring countries harbouring many of the same species and lower taxa. The DRC also covers the majority of the Guineo-Congolian regional centre of

endemism, one of the largest and most biodiverse regions of Central Africa, that encompasses both ‘Moist Central Africa’ and ‘Wet Central Africa’ (Fayolle et al., 2014; White, 1983). Those facts further support that the DRC is a relevant geographically delineated area to produce a robust reference database of images and wood anatomical descriptions for species identification.

Table 2.3: An overview of all general vegetation classes in the DRC defined by Fritz et al. (2003) (Fritz et al., 2003) with the total number of tree species (available in the SmartWoodID database) in each vegetation class and the percentage of threatened species present in each vegetation class.

General Vegetation Class	Number of tree species per class	Percentage of threatened species	Percentage of species at lower risk	Percentage of species with deficient data on threatened status
Closed forest	886	8	16,4	64,4
Altitudinal forest	302	7	17,5	68,2
Edaphic forest	323	2,8	11,5	71,5
Woodland	543	4,2	10,7	68,5
Shrubland	355	4,5	9,6	71,8
Savanna	304	3,9	9,2	73,4
Urban areas	178	3,9	6,2	73
Desert	0	0	0	0
Water bodies	306	3,3	13,7	69,9

2.4.3 Threatened status

Table 2.4: All tree species in the DRC that are appended to appendix II of Checklist of CITES Species (UNEP-WCMC (Comps.), 2022), along with their respective IUCN Red List Category, IUCN population trend and last year they were assessed (IUCN, 2021).

Species name	IUCN Red List Category	Population trend	Last year assessed by IUCN
<i>Azelia africana</i>	vulnerable	decreasing	2019
<i>Azelia bella</i>	least concern	stable	2019
<i>Azelia bipindensis</i>	vulnerable	unknown	1998
<i>Azelia pachyloba</i>	vulnerable	unknown	1998
<i>Azelia peturei</i>	vulnerable	decreasing	2019
<i>Azelia quanzensis</i>	least concern	decreasing	2019
<i>Alsophila camerooniana</i>	least concern	unknown	2016
<i>Dalbergia nitidula</i>	least concern	stable	2018
<i>Euphorbia abyssinica</i>	not evaluated	unknown	-
<i>Euphorbia ingens</i>	least concern	stable	2018
<i>Euphorbia teke</i>	not evaluated	unknown	-
<i>Guibourtia demeusei</i>	near threatened	decreasing	2020
<i>Khaya anthotheca</i>	vulnerable	unknown	1998
<i>Khaya grandifoliola</i>	vulnerable	unknown	1998
<i>Pericopsis elata</i>	endangered	decreasing	2020
<i>Prunus africana</i>	vulnerable	decreasing	2020
<i>Pterocarpus angolensis</i>	least concern	decreasing	2018

<i>Pterocarpus lucens</i>	least concern	stable	2010
<i>Pterocarpus rotundifolius</i>	least concern	stable	2018
<i>Pterocarpus soyauxii</i>	not evaluated	unknown	-
<i>Pterocarpus tessmannii</i>	near threatened	unknown	2020
<i>Pterocarpus tinctorius</i>	least concern	decreasing	2017

Only 23 of the 1700 DRC tree species and lower taxa are included in the Checklist of CITES (UNEP-WCMC (Comps.), 2022), which is shown in Table 2.4. All those tree species and lower taxa are listed in appendix II of the Checklist of CITES and annex B of the European Union and Trade in Wild Fauna and Flora regulation (Commission Regulation (EU), 2019) and are therefore considered species and lower taxa not necessarily threatened with extinction now, but they may become so unless trade is closely controlled. For each species and lower taxon, the IUCN Red List Category is presented with sometimes remarkable results as some species and lower taxa are of least concern according to the Red List, while the Checklist of CITES includes them. One reason for this might be the year of the last assessment by the IUCN. *Pterocarpus tinctorius* is a good example of this, as it was last assessed in 2017. It was appended to appendix II of the CITES at the Nineteenth meeting of the Conference of the Parties (CoP19) in Panama in 2022 because more recent assessment showed the heightened risk of extinction due to trade (CITES, 2022a). Outdated assessments by the IUCN are not the only reason that a species and lower taxon might be appended to CITES. *Afzelia bella*, for example, was last assessed in 2019, which showed that the population remains stable on a global scope and that the species is of least concern. It was however added at the CoP19 due to being a look-a-like species for threatened species such as *Afzelia africana*, *Afzelia bipindensis*, *Afzelia pachyloba* and *Afzelia quanzensis* (CITES, 2022b). Figure 2.4 shows how many of the listed species and lower taxa belong to each of the nine IUCN Red List Categories. Only 8 of the 10 categories are present in the list of DRC timber species and lower taxa as it contains no species and lower taxa that are extinct or extinct in the wild. Nine per cent of all listed species and lower taxa belong to one of the three threatened categories (vulnerable, endangered and critically endangered). The list contains 85 vulnerable species and lower taxa such as *Afzelia bipindensis*, *Baillonella toxisperma* and *Entandrophragma utile*. Thirty-six species and lower taxa are classified as endangered such as *Millettia laurentii*, *Pericopsis elata* and *Autranella congolensis*. Only three species are considered critically endangered, *Beilschmiedia donisii*, *Elaeophorbium drupifera* and *Warneckea superba*. Sixty per cent of all species and lower taxa are of lower risk, more specifically near threatened, least concern and conservation dependent. About 43 species and lower taxa are near threatened like *Milicia excelsa*, *Entandrophragma angolense* and *Dialium pentandrum*. Half of the listed species and lower taxa are of least concern to being threatened. Two species are conservation dependent, and six species belong to the category data deficient because there are little data about their distribution and/or abundance. The remaining 35% of the listed species and lower taxa have not been evaluated yet.

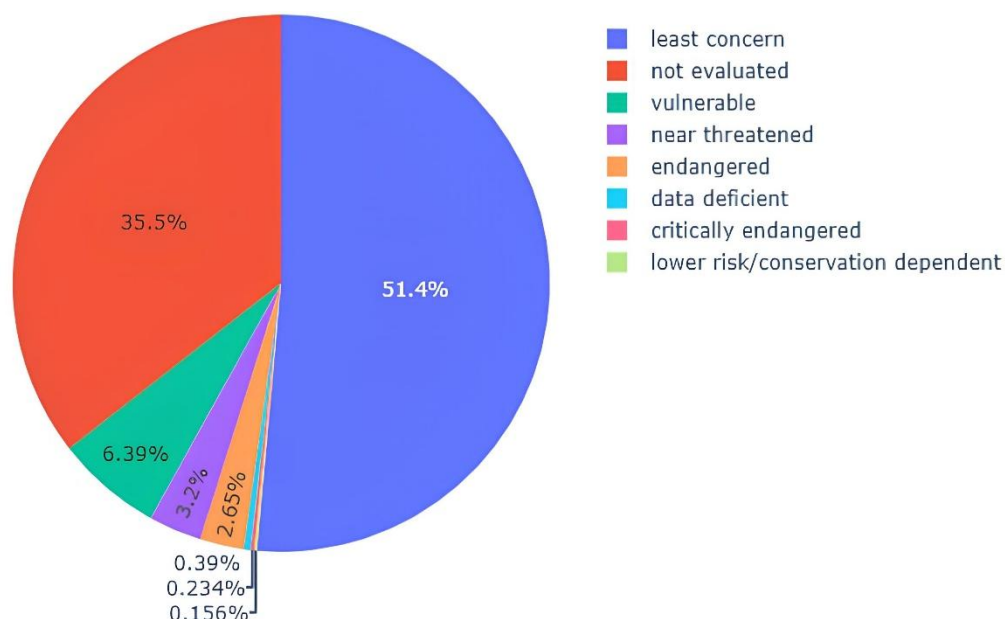


Figure 2.4: Pie chart showing the percentage of each category (according to the IUCN Red List) present in the list of trees in the DRC, capable of providing timber.

2.4.4 Economic value

The list of DIAF 2017, summarizing all tree species and lower taxa present in DRC forests, also indicates their current economic value (DIAF, 2017). The species and lower taxa are divided into four categories: I (commercially exploited species and lower taxa) consisting of 26 species and lower taxa, II (species and lower taxa with potential to be used commercially) consisting of 19 species and lower taxa, III (species and lower taxa with potential to be used commercially, but with few knowledge on their material properties) consisting of 42 species and lower taxa and IV (species and lower taxa with no known economic value) which is the majority of the database at 1613 species and lower taxa. Categories I, II and III are considered as economically important classes due to the use or potential use of these species and lower taxa. A study that analysed 31 logging concessions in the five Inter-national Tropical Timber organization member countries of the Congo Basin was able to determine the 35 timbers from tropical Africa which amount to 94.2% of the total timber volume produced annually in the Congo Basin (Pérez et al., 2005). Of those 35, only three species do not occur in the DRC according to the database. Those three species, *Distemonanthus benthamianus*, *Brachystegia cynometroides* and *Testulea gabonensis*, account for <3% of the total timber volume produced in tropical Africa. This shows that the DRC is rich in commercial species and lower taxa, although it is important to note that those species and lower taxa do not necessarily show the same abundance in the DRC compared to other countries. An example of this is *Aucoumea klaineana*, the most traded species in the Congo Basin, which is only sparsely present in the DRC because it mostly grows in West-Central-African countries such as Gabon (Pérez et al., 2005).

2.5 Discussion

2.5.1 The DRC and its representativeness

The choice of the DRC as a basis for developing the first version of the SmartWoodID database is affirmed by the large area of the DRC housing a large variety of vegetation types. This ensures that a large part of the species and lower taxa from the DRC are also present in neighbouring countries, making the database relevant in an international context. Illegal logging and fraudulent deliveries of timbers are not geographically limited to DRC. These malpractices with the same species and lower taxa happen in other African countries, and therefore, wood identification tools for all DRC timber-producing species can help combat illegal logging across borders.

Total species richness is however not the only parameter to select the DRC as the geographic area of choice for the database. The choice also depends on the type of species, their economic value and threatened status. Some species are protected by the CITES convention (UNEP-WCMC (Comps.), 2022) and The European Union and Trade in Wild Fauna and Flora (Commission Regulation (EU), 2019) while also being highly interesting for commercial use such as *Azelia bipindensis*, *Khaya anthotheca*, *Pterocarpus soyauxii* and *Pericopsis elata*. If such species are logged illegally, it can lead to severe population loss and even impact species that are not currently threatened or that are of least concern due to the damage to precious forest stands in search for valuable trees. A reference database for wood recognition should therefore contain the most prominent exploited timbers, as a wood identification tool will frequently encounter commercial and threatened timbers. The SmartWoodID database contains 32 of those 35 commercial timbers, showing that the DRC is host to almost all highly commercialized timbers from the Congo Basin (Pérez et al., 2005). The large number of commercial species and threatened species makes the SmartWoodID database and wood identification tools derived from it also usable in importing countries. This last aspect is particularly important as this is where regulations go in effect and where wood identification techniques must be applicable on a systematic basis.

2.5.2 Opportunities of the SmartWoodID database

An image database with information on wood anatomical features has clear goals to aid in identifying the botanical taxon of species, by serving as reference material for distinguishing them. In this regard, an image reference database with information on wood anatomy must be complete regarding its content. It should therefore not only aim at encompassing all species that logically can be encountered in trade but must also maximally cover all possible types of irregularities.

Containing all species is particularly important because the value of identification tools depends on the completeness of its reference data. Especially for a species-rich country like the DRC, databases should be as large as possible to reduce the risk that a tool is only developed for a small part of the flora and that a positive identification is not possible only because many species are not included in the database. Moreover, it is very unlikely that foreign species are being imported in the DRC, so the database should purely focus on the maximum of species present in the DRC.

Wood samples most often contain many irregularities visible on a wood surface due to its nature as a natural product that is subservient to the growing conditions of a tree or post-growth incidences such as mechanical damage or damage by insects or fungi. Those irregularities hamper a smooth identification process and are a main reason why expert knowledge is needed to distinguish between diagnostic characteristics and other features. This makes this information particularly relevant for anyone performing wood anatomical assessments in the field. Any tool used to identify wood with anatomy, such as classification keys or AI, must therefore take the irregularities into account. It is the inclusion of data on such irregularities that distinguish the image collection of the Tervuren Xylarium compared to other large image databases that contain high-resolution images of the end-grain surface, such as the database of the XyloTron system (Ravindran et al., 2020). Another difference between SmartWoodID and other databases is the large end-grain area scanned. A large end-grain surface contains a maximum of information on variably occurring macroscopic features. The amount of anatomical information is therefore higher and available to be used in research and development of identification tools. The variability of wood anatomical characteristics between specimens must also be considered as growth conditions and genetic traits can lead to varying wood anatomical features. Therefore, databases should contain information on several

specimens for each species in order to cope with the natural variability of wood between individual trees. The observed misidentifications, ~10% of the initial 3.774, also underscores the need for systematic cross-checking the taxonomy, since inconsistencies can propagate into downstream research on automated identification pipelines if left unresolved. This becomes even more crucial in absence of voucher-linked herbaceous material, as only ~10% of the specimens could be linked to herbaceous material in the Meise Herbarium.

The information on variably occurring macroscopic features and the recorded data on irregularities enrich the SmartWoodID database and ensure its robustness needed to create tools capable of aiding fieldworkers in accurate identification.

2.6 Conclusion

The SmartWoodID image database will offer new opportunities for developing identification systems based on recognition of diagnostic wood anatomical features. This database is unique since it covers a large number of African tree species and lower taxa of which the macroscopic structure is visualized and described. The Tervuren Wood Collection provides this thanks to its heritage of collecting reliable reference material over the span of more than a century. A total of 56% of all DRC tree species and lower taxa, listed in DIAF (2017) (DIAF, 2017), are currently available within the Tervuren Wood Collection. The first version of the SmartWoodID image database that is presented here consists of a set of 954 timber species and lower taxa present in the DRC forests (Dauby et al., 2016; DIAF, 2017). The database focuses on the macroscopic anatomical features that can be encountered on a high-resolution scan of end-grain wood surface. The database accounts for irregularities and natural variability, using multiple specimens with large end-grain surfaces. This makes it a robust reference database for research on wood in general and will allow the development of tools for aiding in law enforcement to combat illegal logging.

2.7 Data availability

The database and all its data will remain publicly available for a minimum of two years starting from the day of publication. **Database URL:** https://hdl.handle.net/20.500.12624/SmartWoodID_first_edition

2.8 Acknowledgements

This work was realized through the combined efforts of the staff at the service Wood biology of the RMCA in Tervuren, Belgium, and the staff at the UGent-Woodlab (Department of Environment, Faculty of Bioscience Engineering) of Ghent University. The collection and specimen were managed by Annelore Nackaerts, Daniel Wallenus, Eric Van Herreweghe, volunteer Richard Shutt and intern Guillaume Charles. The sanding protocols and work were performed by Stijn Willen, Toon Gheyle and Eric Van Herreweghe and were based on the work of Jean-Claude Cerre. The scanning and density measurements were mostly performed by Daniel Wallenus. The annotations were performed by Ir. Michael Monnoye.

Chapter 3: Evaluating expert-defined anatomical features for rapid identification of Congolese tree species

Authors: Ruben De Blaere^a, Hans Beeckman^a, Jan Van den Bulcke^b, Jan Verwaeren^c, Wannes Hubau^{a,b}

a Service of Wood Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium, b UGent-Woodlab, Laboratory of Wood Technology, Department of Environment Ghent University, Coupure Links 653, 9000 Gent, Belgium, c UGent-KERMIT, research unit Knowledge-based, predictive and spatio-temporal modelling, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Gent, Belgium

3.1 Abstract

Macroscopic wood anatomical analysis is a widely applied method for identifying wood taxa based on features visible to the naked eye or with a hand lens. This is essential for enforcing anti-illegal logging regulations and promoting sustainable timber trade. The cross section (or end-grain) is particularly relevant, as its accessibility make it valuable for field applications, requiring little equipment and limited sample preparation. However, the number of expert-defined diagnostic features observable at this scale is limited, posing challenges when distinguishing morphologically similar taxa within species-rich groups, especially in highly diverse tropical regions such as the Congo Basin. Despite the widespread use of this method, its performance has not been systematically quantified, hampering objective evaluation relative to microscopic or molecular identification techniques.

Therefore, this chapter assesses the diagnostic potential of expert-defined macroscopic cross-sectional anatomical features using the SmartWoodID database, the most complete image database of tree species that grow in the DRC. The evaluated features represent the most accessible anatomical characters for field-based identification. Classification performance was assessed at the species, genus, and family levels using a range of analytical approaches, including two-way hierarchical clustering, DTs, and advanced machine learning algorithms such as RF, CatBoost, and SVMs. Findings reveal that macroscopic features alone are insufficient for reliable species-level identification across the full taxonomic spectrum, with correct species predictions in fewer than 25% of cases. The clustering analysis showed that six was the optimal number of clusters that could be identified by clustering the tree species based on 31 accessible macroscopic cross-sectional features. Although accuracy improves at broader taxonomic levels, overlapping anatomical features still constrain resolution, with a maximum of approximately one-third of specimens correctly classified at the genus level, and around half at the family level. Nonetheless, when applied within narrow taxonomic scopes—such as specific genera or families—macroscopic features exhibit notable diagnostic value. For example, the study successfully distinguishes African *Pterocarpus* species, considered difficult to distinguish without laboratory-based methods according to the amendment of all African *Pterocarpus* species to CITES appendix II. These results underscore both the limitations and the potential of macroscopic wood anatomical analysis for trade monitoring and forensic wood identification. In field contexts, where speed, simplicity, and minimal training are critical, CV-based approaches offer promising solutions. By directly analysing macroscopic cross-sectional images, such methods can extract additional diagnostic patterns and enhance identification accuracy, supporting scalable and accessible wood identification systems.

This chapter is not submitted to a peer-reviewed journal on 19/08/2025.

3.2 Introduction

The ability to rapidly and accurately identify wood species is increasingly critical given the ambitious international regulations aimed at combating illegal logging (Gasson et al., 2021; Hirschberger, 2008; Hoare, 2015; Magrath et al., 2009; Piabuo et al., 2021; Van Brusselen et al., 2023). Effective identification methods must be both accurate and applicable in field conditions (e.g. forests, roadside, warehouse, container, harbour), where access to laboratory facilities is often limited (Gasson, 2011). Among the various techniques available, wood anatomical assessment remains one of the most reliable approaches for species identification. This method involves examining tissues and cells across different planes and scales to detect diagnostic features characteristic of specific taxa. The IAWA has standardized anatomical features, providing consistency and reliability across the World (Gasson et al., 2011; NS, 1989; Ruffinatto et al., 2015; Wheeler, 2011).

Identification keys include digital and offline formats and play a crucial role in simplifying and speeding up the identification process, especially given the immense diversity of tree species (Barefoot and Hankins, 1982; Gregory, 1980; Ilic, 1993; LaPasha and Wheeler, 1987). These keys are expert-driven and designed for practical use, guiding users through systematic evaluations of diagnostically valuable features by determining their presence, absence, or variation (LaPasha and Wheeler, 1987; Vander Mijnsbrugge and Beeckman, 1992). A broad array of 163 well-defined microscopic features were defined by IAWA and are today commonly used for precise identifications (NS, 1989). InsideWood is a widely recognized microscopic identification key using these IAWA features. It combines a comprehensive database of wood anatomical descriptions with a search engine that queries the descriptions. InsideWood is widely used for research, teaching, and reference purposes (NS, 1989; Wheeler, 2011; Wheeler et al., 2020). With a global scope, InsideWood hosts over 9,400 wood anatomical descriptions of both fossil and modern woody dicotyledons, covering more than 10,000 species and 200 plant families, accompanied by over 50,000 images (Wheeler et al., 2020).

However, comprehensive wood anatomical assessment is constrained by the need for specialized laboratory equipment and labour-intensive sample preparation techniques, such as microtomy, which increase both cost and processing time. Therefore, wood identification is most often performed using only macroscopic features that are quickly observable during *in situ* wood identification (e.g. in harbours) (Koch et al., 2018; Wheeler and Baas, 1998). A study by Ruffinatto et al. determined which IAWA features can be observed with the naked eye or with the aid of a simple hand lens. Using only these macroscopic features for identification significantly reduces the time and resources required for sample preparation and examination (Ruffinatto et al., 2015). Among the three anatomical planes, the cross section is the most readily exposed and provides up to 31 distinct macroscopic features that can support rapid identification (Ruffinatto and Crivellaro, 2019). These features enable wood identification to be performed outside of laboratory settings by individuals with foundational knowledge in wood anatomy.

Macroscopic identification keys, particularly dichotomous and multi-entry types, are commonly used in field contexts where time constraints necessitate immediate classification—often within minutes. This methodology is especially relevant in the context of timber trade enforcement, where rapid screening of large volumes of wood and wood-derived products is essential to detect fraudulent or illegally traded specimens, including those from protected species. Notable tools such as the *Atlas of Macroscopic Wood Identification* (Ruffinatto and Crivellaro, 2019) and *macroHOLZdata* (Richter et al., 2017) exemplify the application of macroscopic features in digital identification systems. The structural simplicity of dichotomous or multi-entry keys makes them popular, but also limits their capacity to fully leverage the power of anatomical features and often fails to accommodate intra-specific anatomical variability. Furthermore, the applied categorical feature states (present, variable, or absent) are inherently subjective, as the thresholds distinguishing these states are not strictly defined. For example, what qualifies as a 'present' feature can vary between observers, leading to inconsistent character scoring. In identification keys, this interpretive variability can significantly influence outcomes. When using a dichotomous key, an anatomist who codes a feature as 'present' may be guided to a different result than one who considers the same feature 'variable,' potentially excluding the correct taxon early in the process. While multi-entry keys

are somewhat more flexible, they can still produce misleading outcomes if the underlying database does not adequately capture the natural variation of features as they appear across different specimens.

More importantly, the restricted number of diagnostic features limits their discriminatory power in taxonomically diverse and morphologically convergent groups. The combination of observer bias (especially when using simple key-based systems) with the sheer scarcity of macroscopic features, might often lead to partial or full misidentifications. However, despite their widespread use, no systematic effort has been made to quantitatively assess the identification success of macroscopic features. This lack of empirical evaluation hampers objective comparisons between identification success of macroscopic features versus identification success of more advanced, yet resource-intensive laboratory-based methods such as microscopy.

Therefore, here we evaluate the ability of macroscopic anatomical features to identify wood specimens in the species-rich SmartWoodID database (De Blaere et al., 2023). To maximise identification success and valorise the full diagnostic potential of all discernible macroscopic features, we do not use simple dichotomic keys, but more sophisticated analytical techniques which can significantly increase accuracy and robustness (Boulesteix et al., 2012; Salman et al., 2024). We specifically use a statistical technique (two-way species and feature clustering analysis) and four machine learning techniques that were evaluated for their effectiveness in classifying wood specimens across the three taxonomic levels (species, genus, and family): DT, SVM, RF, and CatBoost. We assess the accuracy of each of these methods using Congolese species in the SmartWoodID database. In order to compare the accuracy with other online resources, the models were trained on a selection of commercial timbers and were used to assess the ability to predict InsideWood descriptions of the macroscopic cross-sectional IAWA features and to evaluate the capacity to distinguish high-demand timber species commonly traded in international markets.

3.3 Materials and Methods

3.3.1 Dataset Description

This chapter is based on the SmartWoodID database, serving as a valuable resource for examining the relationship between macroscopic cross-sectional wood anatomy and the botanical diversity of Congolese tree taxa (De Blaere et al., 2023). The following information in this section, re-uses the text in section 1.5 which provides an overview of all data used in each chapter.

The SmartWoodID database contains high-resolution RGB scans of the macroscopic end-grain surfaces of 3,742 wood specimens, representing 954 species native to the DRC. Each specimen was prepared by scanning the cross-section at 2400 dpi using a flatbed scanner. This resolution allows for the visualization of macroscopic features essential for wood identification.

Species and lower taxa are represented by multiple specimens, capturing both intra- and interspecific anatomical variation. This makes the database well-suited for studying wood identification using macroscopic anatomy. A complete overview of the database is provided in Chapter 2 and in De Blaere et al. (2023) (De Blaere et al., 2023), while Supplementary Materials Table 8.1 lists all unique specimen identifiers and metadata. To enable machine learning analysis, we selected only species represented by at least two specimens. This set comprises 2,296 digitized specimens across 601 species, 286 genera, and 64 families. Discriminatory power was mainly assessed by training and evaluating classification models on the specimens. Therefore, specimens were allocated random to training (75%) or test set (25%), while preserving distribution of species across both sets. Within both sets, a subset of 78 commercially important species was defined for targeted evaluation. An overview of the chapters, designated datasets used in each chapter and of which hypothesis they target is provided in section 1.4.

Macroscopic wood anatomical features were described for each end-grain image of the SmartWoodID specimens (see chapter 2, Table 2.2; and De Blaere et al. (2023) (De Blaere et al., 2023)). Each feature is assigned a Macroscopic IAWA feature number (Ruffinatto et al., 2015). We did not use descriptions on the presence of growth rings as the discernability at the used resolution was often not high enough to assess this feature with certainty. Each feature is annotated with one of four states: Present (clearly visible), Variable (sporadically observed), Absent (rarely observed, below the threshold for Variable), or NA (undiscernible due to resolution limits or ambiguous visual cues).

3.3.2 Macroscopic feature annotations

For clustering and classification, the macroscopic anatomical features needed to be encoded in a way that appropriately reflects their descriptive nature and variability. Grouping features into their overarching refined categories (see Table 2.2) and treating them as standard categorical variables was not feasible. For example, the feature "vessel frequency" cannot be simply categorized into fixed bins (e.g., 6–20 vessels / mm²), because the IAWA guidelines may define the same value range differently across species (Committee, 2004; NS, 1989). For example, on a specimen, a vessel frequency of ≤ 5 vessels / mm² might be considered "variable," 6–20 vessels / mm² as "present," and >20 vessels / mm² as "absent." This complexity means that each feature must be treated individually, and generic binning approaches cannot be applied. However, the 'variable' feature state complicates straightforward analysis further as the classification or clustering can no longer be regarded as a binary case (a feature is either present or absent). Treating all feature states as equal categorical levels would obscure important information—particularly the nuance provided by the "variable" state, which implies a range of occurrence rather than a binary presence or absence. To preserve this information while enabling quantitative analysis, each state was transformed into a numerical value representing the observed variability. Specifically, features were encoded per species as follows: "present" = 1, "variable" = 0.5, "absent" = 0, and "NA" (not assessable due to resolution or ambiguity) = 0. These values were used as inputs for clustering and classification models, allowing the analysis to incorporate the graded nature of anatomical trait expression.

To assess the identification potential of macroscopic wood anatomy in a commercial context, the data were adjusted to align SmartWoodID with InsideWood database standards, addressing differences in

feature definitions and resolutions to ensure data compatibility while maintaining the integrity of anatomical descriptions. An overview is presented in Chapter 2.3.6 (Table 2.2) and De Blaere et al. (2023) (De Blaere et al., 2023).

3.3.3 Clustering analysis

To identify intercorrelated macroscopic features and to visually and statistically explore whether species in SmartWoodID are significantly different from one another based on the macroscopic annotations, we performed a two-way hierarchical clustering analysis in R (version 4.4.1) (Borcard et al., 2011).

We then constructed two separate dissimilarity indices to independently assess variation among taxa and among anatomical features, using the ‘vegdist’ function of the ‘Vegan’ package (version 2.6-8) (Oksanen et al., 2025). For the clustering of taxa (i.e., botanical species), the Raup-Crick dissimilarity index was applied. This index is appropriate for binary presence-absence data and emphasizes stochastic differences in feature composition across taxa. For the clustering of anatomical features, the Bray-Curtis dissimilarity index was used, as it accounts for relative abundance and is commonly used in ecological studies. Each dissimilarity matrix was subjected to hierarchical agglomerative clustering using the Ward.D2 linkage method, which minimizes the total within-cluster variance and is effective for generating compact, interpretable clusters (Borcard et al., 2011). To determine the optimal number of clusters, we applied the Mantel test, which assesses the correlation between the original dissimilarity matrix and a binary matrix representing group membership at various cut-off levels (Borcard et al., 2011). For each value of k (number of clusters), a Mantel correlation coefficient was computed, and the k maximizing this correlation was selected as the optimal number of clusters for both dimensions. A two-way clustered heatmap was generated using the ‘pheatmap’ package (version 1.0.12), allowing for the simultaneous visualization of clusters across taxa (rows) and features (columns) (Kolde, 2018). This was accompanied by smaller figures showing the presence of each anatomical feature in each species cluster through boxplots, based on the species-specific values in the heatmap.

To evaluate the taxonomic consistency of anatomical clustering, we examined the distribution of species across clusters with respect to their genus. Genera represented by multiple species were identified, and we assessed whether their species were grouped within a single cluster or spread across multiple clusters. This allowed for the quantification of anatomical coherence at the genus level and the identification of genera with divergent or convergent anatomical patterns relative to their taxonomic classification.

3.3.4 Classification Models

3.3.4.1 Training and testing datasets

Four machine learning techniques were evaluated for their effectiveness in classifying wood specimens across the three taxonomic levels (species, genus, and family): DT, SVM, RF, and CatBoost. We split the selected image database in a training (75% of specimens) and a test (25% of specimens) set. To prevent information leakage and to minimize potential biases, we ensured a balanced distribution of species across the training and test sets. A summary of specimen distribution across training and test sets is presented in Table 10.1. This shows that each species has a minimum of 1 and a maximum of 17 specimens in the training dataset and a minimum of 1 and a maximum of 6 specimens in the test dataset. We used the same training and test sets to train each of the four models.

Table 3.1: Summary of specimen distribution between training and test datasets. This table reports the minimum, maximum and average number of specimens per taxonomic level (family, genus and species level) included in the training and test datasets.

	Training Specimens				Testing Specimens			
	Minimum	Maximum	Average	Standard deviation	Minimum	Maximum	Average	Standard deviation
Family	1	425	24	56	1	221	12	29
Genus	1	56	5	6	1	22	3	3
Species	1	17	3	2	1	6	1	1

3.3.4.2 Classification models

Traditional DTs function akin to identification keys. DTs recursively split datasets based on features that optimize criteria such as information gain or Gini impurity at each node, unlike identification keys, which rely on rules derived from the raw data to guide users through sequential feature evaluations without probabilistic modelling (Murthy et al., 1994; Quinlan, 1987). Due to their interpretability and ability to model non-linear relationships, DTs are well-suited for analysing wood anatomical datasets, which often include large numbers of discrete features (Vander Mijnsbrugge and Beeckman, 1992; Wheeler et al., 2020).

RF is an ensemble method that builds multiple DTs and combines their predictions to improve accuracy and reduce overfitting (Ho, 1998; Pes, 2021). Each tree is trained on a random subset of the data, and at each split, the algorithm evaluates a random subset of features, selecting the best feature from this subset to optimize the split. This promotes diversity among the trees in the ensemble. The probabilities of each class are then obtained by averaging the predictions of all trees. It provides a higher degree of robustness compared to a single DT and can effectively handle noisy data, which is common in wood anatomical datasets due to the inherent inter-specific variability of the features (Biau and Scornet, 2016; von Arx et al., 2016). Additionally, the built-in feature importance metrics can help highlight which anatomical features are most significant for classification, offering insights into the biological relevance of the features.

CatBoost, a gradient boosting algorithm specifically designed to handle categorical data effectively (Prokhorenkova et al., 2018). The algorithm works by sequentially training an ensemble of DTs, where each new tree attempts to correct the errors made by the previous trees (Ibrahim et al., 2020). Unlike traditional gradient boosting, which may require extensive preprocessing of categorical variables, CatBoost natively handles categorical features (Bentéjac et al., 2021; Wanga et al., n.d.). Its ability to generalize well on datasets with uneven representation makes it suitable for biological databases (Jumabek et al., 2021). Furthermore, CatBoost demonstrates faster training times and better performance compared to other boosting algorithms, such as XGBoost, making it highly suitable for large-scale wood anatomical datasets (Bentéjac et al., 2021; Mironov and Khuziev, 2022).

SVMs, which represent yet another powerful machine-learning technique for wood identification. SVMs are designed to identify the optimal hyperplane that best separates different classes by maximizing the margin between support vectors (Joachims, 2002). For non-linearly separable data, SVMs use kernel functions (e.g., radial basis functions or polynomial kernels) to map data into higher-dimensional spaces, enabling better class separation. In wood anatomical datasets, where subtle distinctions between species can be challenging, SVMs offer high precision in modelling complex relationships between features (Rosa da Silva et al., 2017).

3.3.4.3 Hyperparameter Optimization

Each model was optimized using a randomized grid search to enhance performance by sampling a predefined number of parameter combinations from a range of hyperparameter values. This strategy ensures a scalable and flexible search for the best hyperparameters, especially when dealing with large parameter spaces (such as for CatBoost). In total, 50 random hyperparameter settings were sampled from the parameter grid of each technique, consultable Supplementary materials: Table 8.2. During the randomized search, the models were iteratively trained, with training samples split for training and validation using an internal 4-fold cross-validation, to ensure robustness. To guarantee consistency and reproducibility across all experiments, the random number generators were seeded with a fixed value for all algorithms and folds. Class weights were used to address the inherent class imbalance. As Catboost is a gradient method, it is sensitive to overfitting with fixed iterations, as such the model was run for 1000 iterations with early stopping after 20 epochs if no improvement in loss. All data processing, classification, and evaluation tasks were executed using Python 3.9.15, Scikit-learn 1.3.0, Pandas 2.0.3, and Matplotlib 3.7.2.

3.3.4.4 Evaluating model performance across taxonomic levels

The performance of each classification model was evaluated across taxonomic levels using the species label with the highest predicted probability for every test specimen. Predicted species labels were mapped to their genus and family labels to provide insight into a species' model's ability to capture hierarchical taxonomic relationships. To evaluate the accuracy for each model, species predictions were visualized using horizontal stacked bar charts, illustrating accuracy at progressively higher taxonomic levels (e.g. species, genus, family) progression across taxonomic levels.

To complement this indirect mapping approach, the models were also trained to predict genus and family labels directly. The outcomes of these direct predictions were visualized using a Venn diagram, which showed the proportion of specimens correctly classified at each taxonomic level, as well as the extent of overlap and mismatch among them. This visualization enabled a comparison between hierarchical inference (from species predictions) and direct classification, helping to determine whether direct prediction at higher ranks could improve overall accuracy. Furthermore, it highlighted species that were frequently misclassified at certain levels, potentially revealing inconsistencies in their alignment with current botanical taxonomy or indicating taxonomic ambiguity within certain groups.

3.3.4.5 Optimal number of taxa for Reliable identification

To determine the optimal number of output classes needed for a reliable shortlist at each taxonomic level, predicted classes were ranked based on their assigned probabilities. Accuracy improvements from including the top k predictions were analysed using two approaches.

First, accuracy was assessed as a function of k , where the accuracy curve gradually approaches a maximum value of 1, representing perfect identification of all specimens. Because this function is cumulative, accuracy steadily increases as additional ranked classes (k) are considered. Additionally, the AUC was computed for each technique's cumulative curve in function of k , providing a single quantitative measure of overall classification performance across multiple k levels.

Second, accuracy was assessed as a function of probability cut-off-thresholds, where the accuracy curve gradually approaches a minimum value of 0, representing complete misidentification across all specimens. Because this function is reverse cumulative, accuracy steadily decreases as the cut-off threshold increases, leaving out more predicted classes. A 95% cumulative accuracy threshold was included for both functions to indicate the smallest k value (conversely the largest cut-off threshold) at which the model's top- k accuracy reaches 0.95, offering a practical benchmark for classification reliability.

3.3.4.6 Refining species prediction

Direct predictions at higher taxonomic levels (e.g. family level) could refine predictions at lower level (e.g. species). For example, certain species predictions could be ruled out by a secondary family prediction, eliminating species that do not belong to those families. The impact of using higher taxonomic-level predictions to refine lower taxonomic-level classifications was therefore assessed. Specifically, the top- k predicted families (derived from family-level models) were used to filter out predicted species class that are not a member of those botanical families, and the resulting improvement in accuracy was analysed. Initially, the value of k for filtering was determined based on the 95% average threshold value from the family-level model outputs, ensuring that the selected families were classified with a high degree of confidence. Next, individual families and genera were singled out for studying the contribution of anatomical features for refining genus and species predictions, respectively.

3.3.4.7 Assessing Recognition of InsideWood Descriptions for Commercial Species

The models were retrained using only commercial species to evaluate the models' effectiveness in distinguishing commercially significant timber species in a practical context, such as combatting illegal logging. This subset contains 537 specimens belonging to 76 commercial species, 56 genera and 22 families.

For this analysis, we tested evaluations conducted both on SmartWoodID and corresponding descriptions of the same species from InsideWood (Wheeler, 2011). The performance of each machine-learning technique was evaluated by selecting the class with the highest predicted probability for each test specimen. Species predictions were mapped to their corresponding genus and family labels to provide insight into a species' model's ability to capture hierarchical taxonomic relationships. By testing the models on InsideWood species descriptions, this analysis evaluates their ability to generalize beyond the SmartWoodID dataset, providing insight into the practical applicability of these techniques for real-world wood identification. These descriptions, compiled by multiple wood anatomists from diverse image types (resolution, field-of-view), serve as an external benchmark. The assessment determines whether models can classify commercial species based on macroscopic features described from different materials (e.g. thin sections), rather than SmartWoodID standardized scans.

3.4 Results

3.4.4 Clustering analysis

This analysis aimed to identify which anatomical features co-vary across timber species. Two-way clustering analysis was applied to provide clear visual representation of shared anatomical features and to highlight key diagnostic features relevant for wood identification. All features contributed equally to the distance metric; the method does not assign explicit weights or identify feature importance in a predictive sense. Instead, the decisive features for each cluster were interpreted post-hoc by examining co-occurrence patterns and concentrations within the heatmap. For example, cluster 5 is characterized by wide banded parenchyma, while cluster 3 is strongly associated with diffuse-in-aggregate parenchyma. Thus, while the clustering model itself does not indicate which feature group was most decisive, the visualizations allow us to infer which anatomical traits most strongly differentiate clusters. Results are visualized as a heatmap, where wood species and anatomical features are simultaneously ordered according to their hierarchical clustering. Figure 3.12345 depicts the relative abundance of 31 macroscopic anatomical features across 954 wood species. Rows represent species, and columns represent anatomical features, with clustering patterns visualized through dendrograms on the left (species) and top (features). The heatmap employs a Viridis colour gradient, with higher values indicating higher relative abundances of anatomical features. The species and features were clustered according to the mantel statistics, with clustered visualized on the heatmap. The heatmap reveals six broad species clusters and five feature clusters, reflecting both species-level differentiation and underlying structural relationships among wood anatomical features.

To facilitate interpretation of the heatmap, boxplots were generated for each macroscopic anatomical feature, showing how consistently that feature occurs across the six species clusters (Figure 3.2 to Figure 3.8). Species-specific heatmap scores of the heatmap in Figure 3.1 were used to construct boxplots per cluster, allowing visualization of whether a feature is consistently absent, consistently present, or variable within clusters. In total, 31 cross-sectional features were examined, grouped into figures according to the broader feature categories defined in Chapter 2 (Table 2.2).

Some features—such as vessel porosity (ring, semi-ring, and diffuse) (Figure 3.2), ray visibility (Figure 3.3), and certain vessel arrangements (e.g., tangential bands, diagonal, and dendritic patterns) (figure Figure 3.4)—appear less informative for differentiating species clusters. These features display relatively uniform distribution across clusters, lacking pronounced cluster-specific concentrations. One exception is radial vessel arrangement, which tends to be more prevalent in the third cluster (Figure 3.4). Among species clusters, ray and vessel quantitative features are more informative. Clusters 1 and 2 predominantly exhibit moderate ray densities (4–12 rays/mm²), while cluster 5 includes species with lower ray densities, and clusters 3, 4, and 6 tend toward higher ray densities (see Figure 3.3). Vessel frequency shows complementary patterns: clusters 1, 4, and 6 primarily comprise species with moderate vessel densities (5–20 vessels/mm²); clusters 2 and 5 more frequently include species with low frequency (<5/mm²); and cluster 3 is characterized by high frequency (>20/mm²) (see Figure 3.7). These vessel frequency patterns inversely align with vessel diameter. For example, cluster 3 features mostly small vessels (<80 µm), while clusters 2 and 5 include more large vessels (>130 µm), and the remaining clusters exhibit intermediate diameters (Figure 3.6). These opposing patterns underscore a physiologically relevant trade-off: species with high vessel density tend to have smaller vessels, while those with low density often feature larger vessels. This aligns with known hydraulic strategies in trees, where high-conductivity vessels are fewer and larger (Carlquist and Hoekman, 1985).

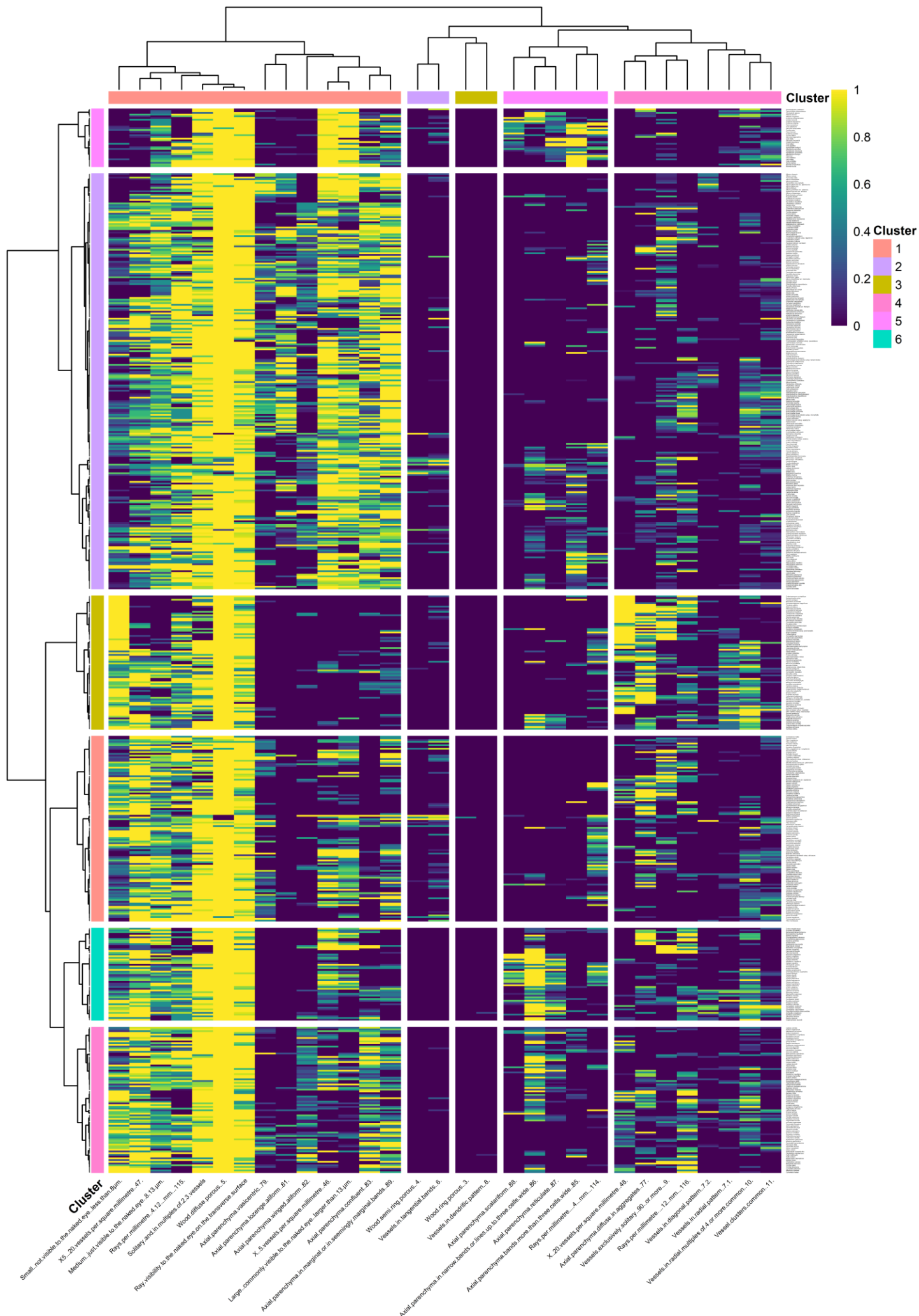


Figure 3.12345: Two-way clustered heatmap showing clustering of botanical Congolese tree species (rows) and macroscopic cross-sectional anatomical features (columns). The presence of features is highlighted using viridis colour gradient to ensure perceptual uniformity (Garnier et al., 2024). Broad clusters are visualized using distinct colours.

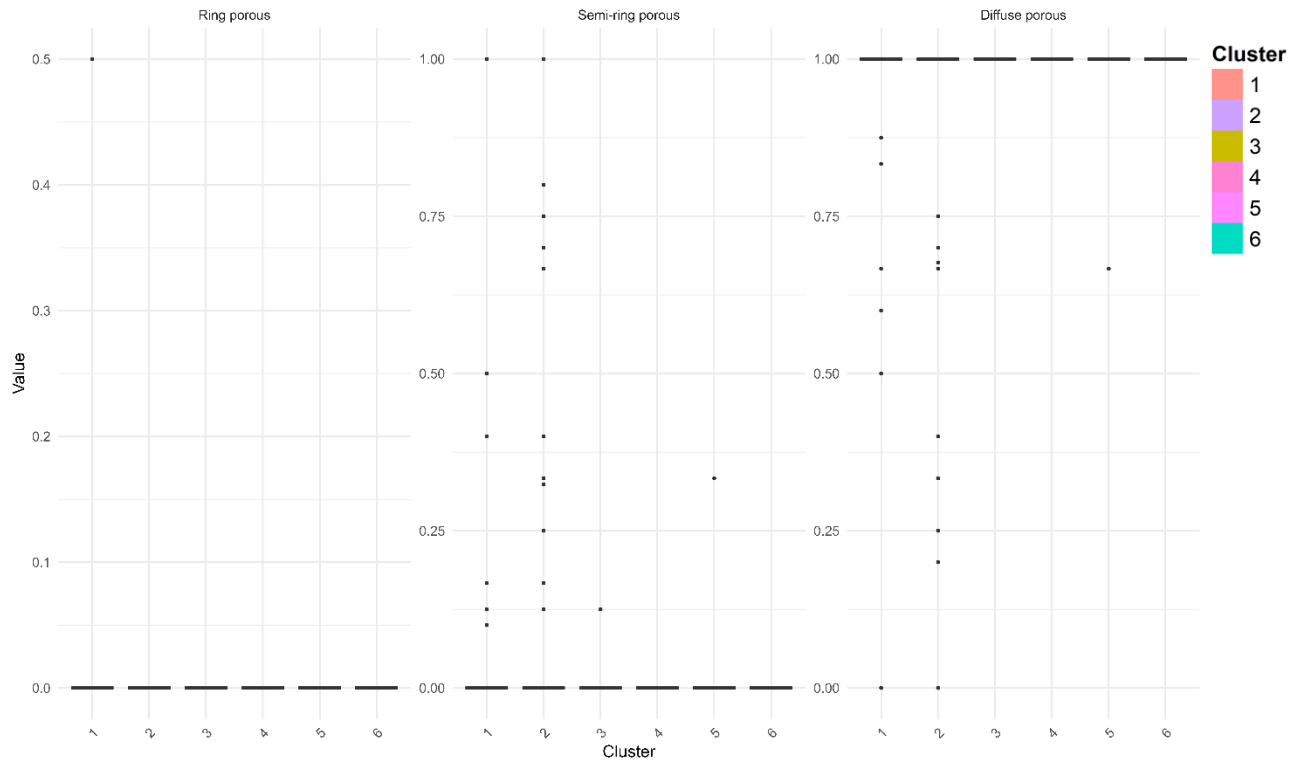


Figure 3.2: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Vessel porosity features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

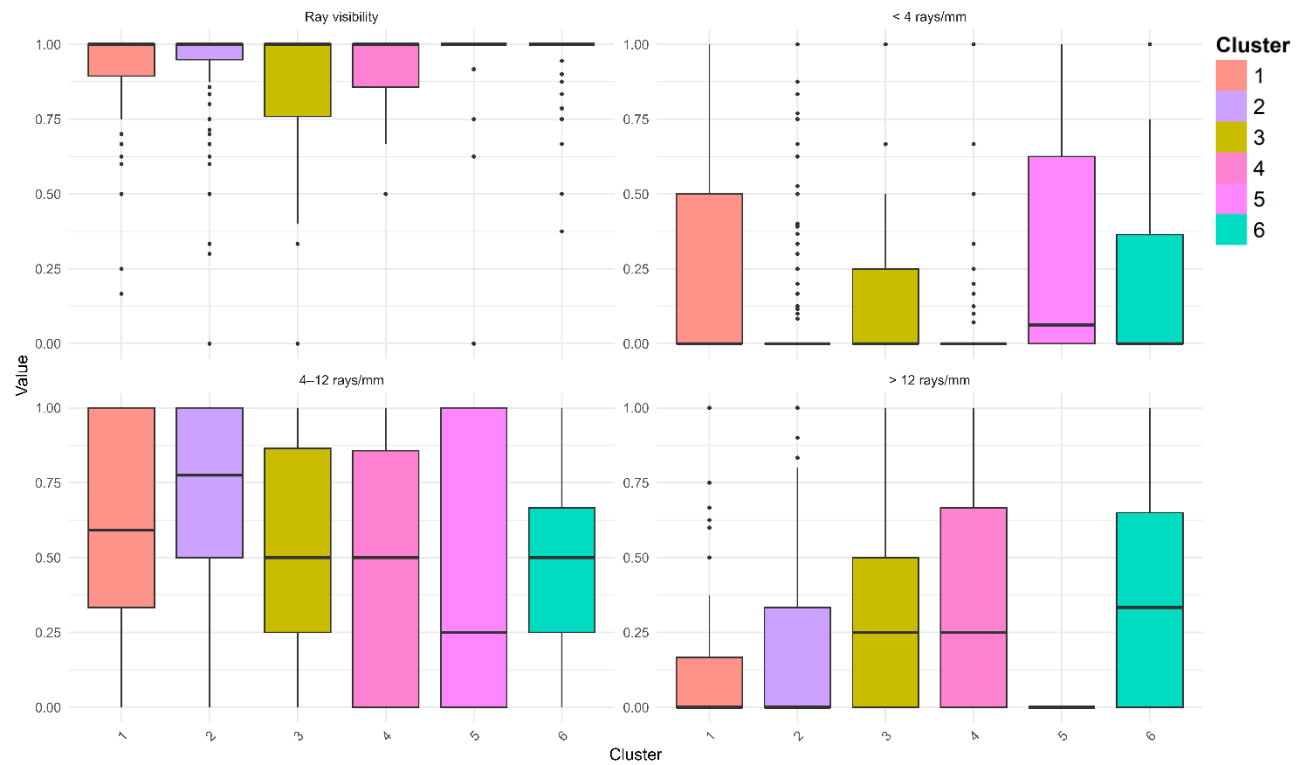


Figure 3.3: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Ray features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

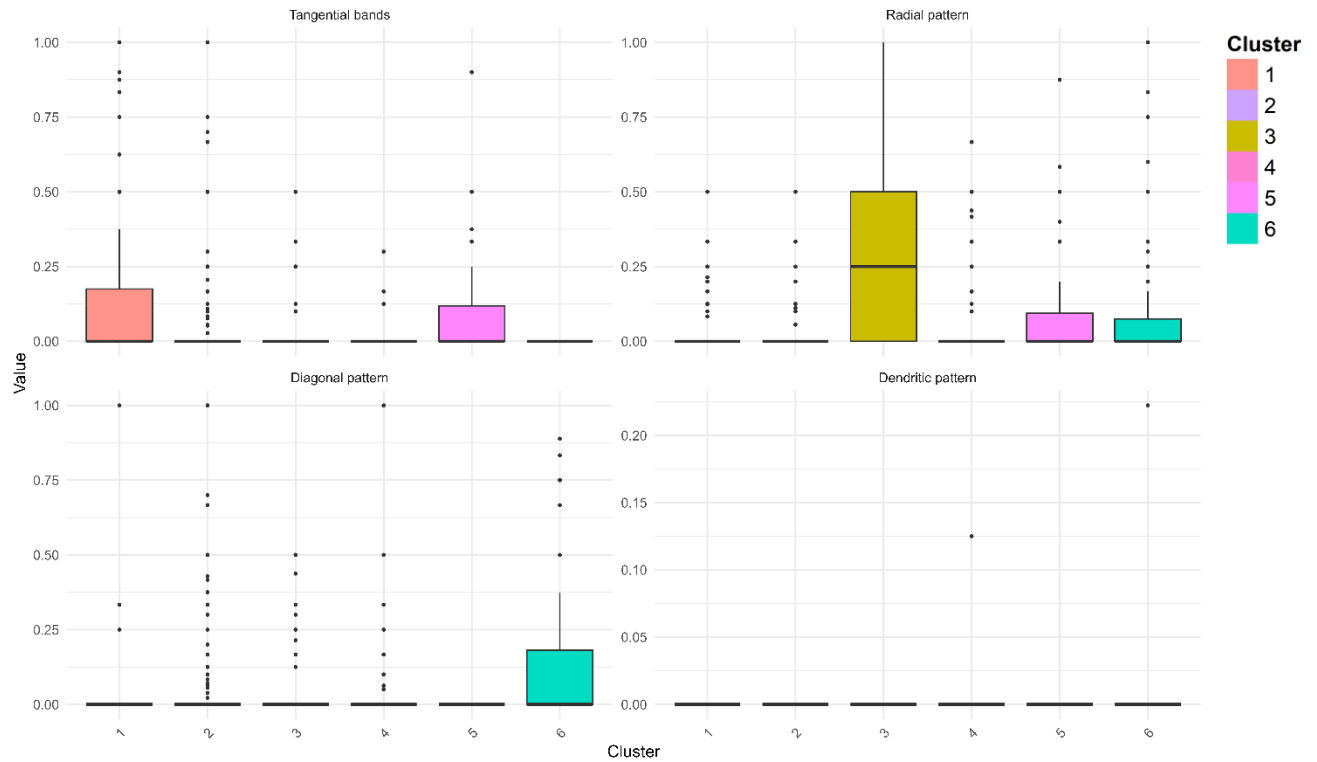


Figure 3.4: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Vessel arrangement features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

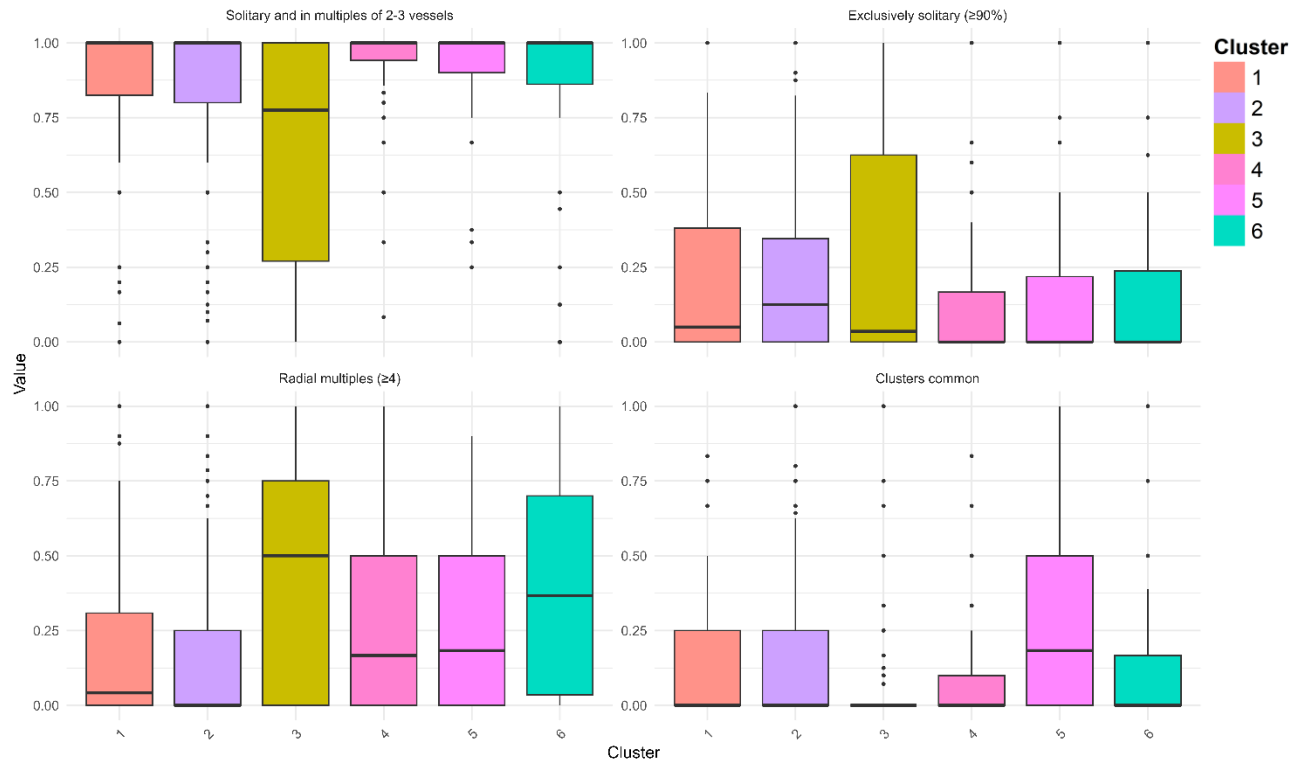


Figure 3.5: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Vessel grouping features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

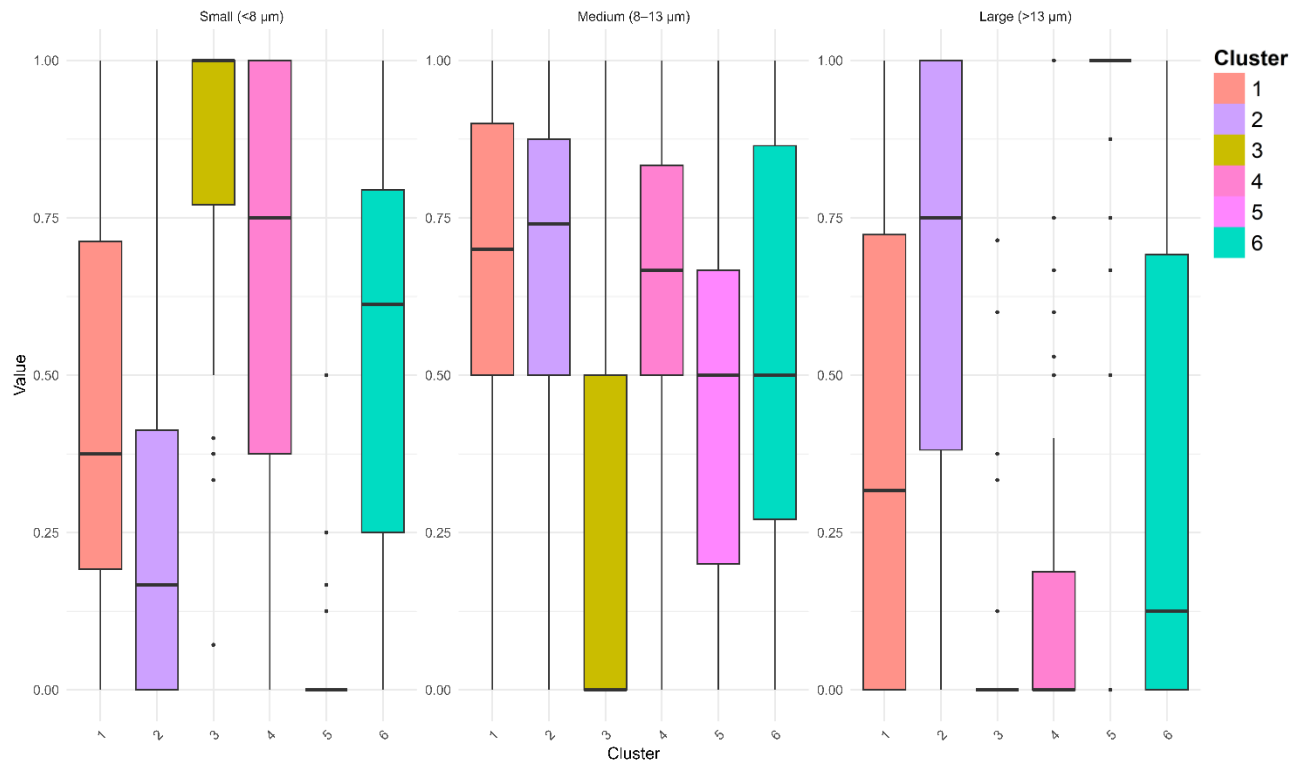


Figure 3.6: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Vessel diameter class features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

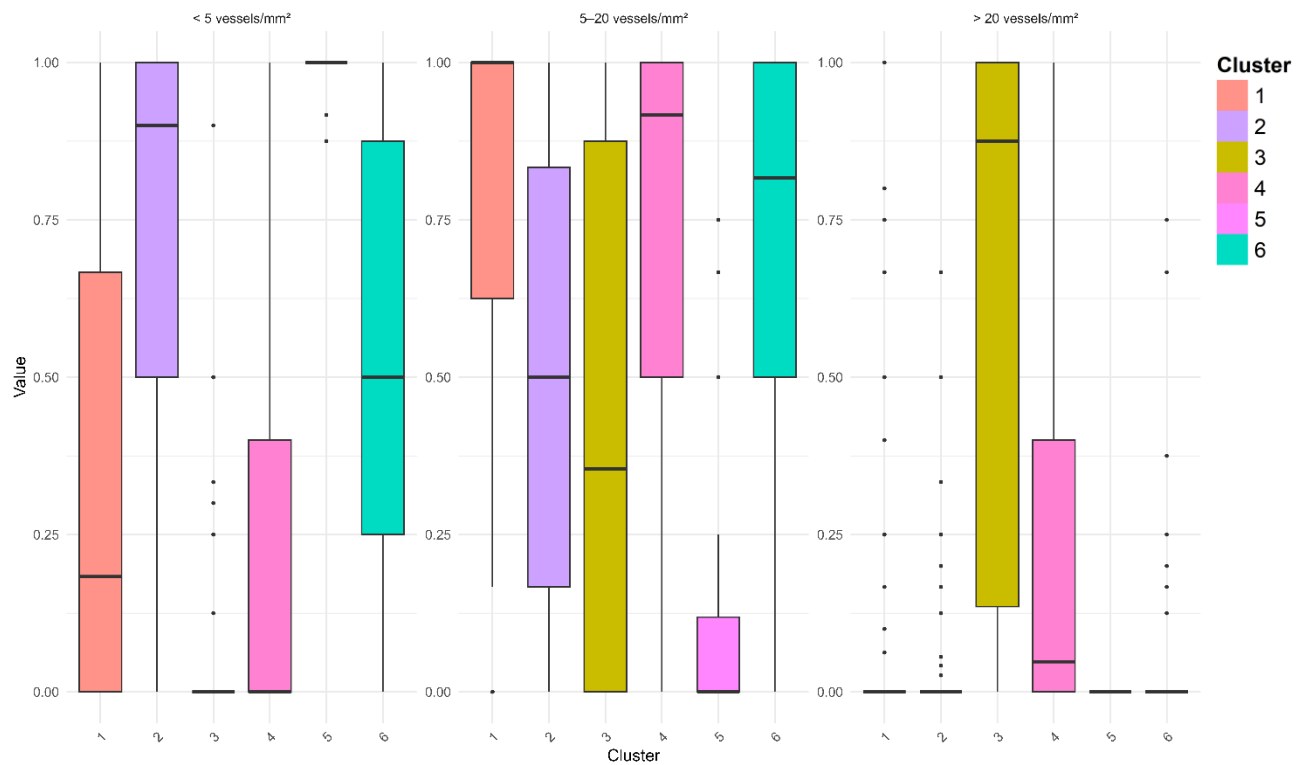


Figure 3.7: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Vessel frequency class features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

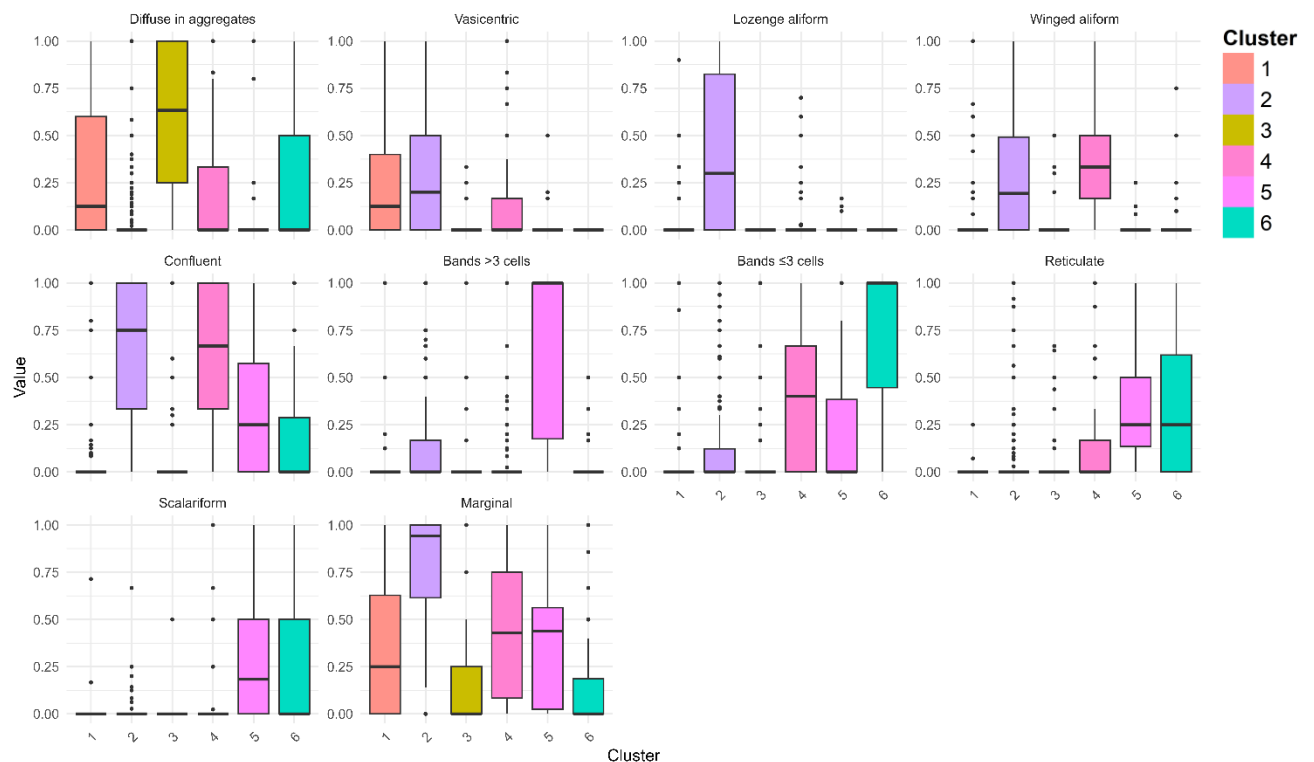


Figure 3.8: Boxplots showing distribution of macroscopic anatomical features across six species clusters. This figure shows Parenchyma features, with individual features printed above each subplot. Cluster numbers are shown on the x-axis and color-coded to the same standard as in Figure 3.1. The y-axis indicates feature prevalence in each cluster (0 = absent for all species, 1 = present for all species).

Axial parenchyma characteristics further distinguish species clusters (see Figure 3.8). Cluster 5 is defined by a predominance of wide banded parenchyma, while cluster 6 is more strongly associated with narrow bands. Both clusters feature reticulate parenchyma, but scalariform patterns are more frequent in cluster 5. Cluster 2 displays a range of paratracheal parenchyma types, including vasicentric and confluent forms, along with frequent marginal bands. Cluster 4 represents a transition from paratracheal to banded parenchyma, frequently featuring winged-aliform, confluent, and marginal bands, though less consistently than cluster 2. Cluster 3 is distinguished by its emphasis on apotracheal parenchyma—particularly diffuse-in-aggregates—which is largely mutually exclusive from the banded and paratracheal types found in other clusters. In cluster 1 axial parenchyma is less observed, though some species exhibit marginal bands, diffuse-in-aggregates, or vasicentric types.

Vessel grouping features are comparatively less diagnostic across clusters (Figure 3.5). Solitary vessels are broadly distributed but do not dominate any cluster. Radial multiples of four or more are largely absent from cluster 2 and somewhat more common in clusters 3, 4, 5, and 6. Vessel clusters are especially frequent in cluster 5, though not universally present.

These species-level observations intersect with five feature clusters that reflect co-occurrence patterns among anatomical features. Feature cluster 1 groups paratracheal parenchyma types such as confluent, vasicentric, lozenge-aliform, and winged-aliform, reflecting their structural continuum. These types often blend into one another morphologically and functionally, co-occurring in species with moderate to high vessel and ray densities. Feature cluster 2 links semi-ring porosity with vessels arranged in tangential bands—an association that aligns with the seasonal onset of earlywood vessels, often producing tangential alignments. Feature cluster 3 brings together ring-porous species and dendritic vessel arrangements. Both features are largely absent from Congolese species, and their co-location within this cluster likely reflects mutual absence. The lack of ring porosity can be attributed to the predominantly tropical and subtropical climates of the DRC, where seasonal growth rhythms are less pronounced.

Feature cluster 4 represents a structurally divergent group of features, including diffuse-in-aggregate parenchyma, exclusively solitary vessels, high vessel frequency, and vessel arrangements in radial or diagonal patterns. These features characterize anatomical strategies that are spatially and functionally distinct from the vessel-associated (paratracheal) parenchyma types. Diffuse-in-aggregate parenchyma is particularly removed from other axial types in the dendrogram, underscoring its apotracheal nature: it consists of loosely clustered cells not associated with vessels. This structural independence contrasts strongly with the tight vessel-parenchyma associations seen in paratracheal types. As such, apotracheal and paratracheal parenchyma types appear nearly mutually exclusive in their distribution, with only a few exceptions such as the genus *Afzelia bipindensis* (Fabaceae), *Balanites wilsoniana* (Zygophyllaceae), and species of *Beilschmiedia* (Lauraceae) exhibiting both forms. The integration of these patterns across species and features reinforces the idea that wood anatomical diversity in Congolese species reflects a continuum of trait syndromes, with certain parenchyma and vessel arrangements marking points of distinct structural divergence.

Feature cluster 5 groups narrow and wide banded parenchyma together with scalariform and reticulate patterns. This reflects established structural interdependencies: both reticulate and scalariform parenchyma require the presence of axial parenchyma bands and are shaped by their spacing relative to rays. Reticulate parenchyma forms regular, net-like networks where tangential bands align with ray spacing, whereas scalariform parenchyma consists of narrow, evenly spaced horizontal lines forming ladder-like patterns, with band intervals narrower than ray spacing. The co-occurrence of these patterns in species clusters 5 and 6 supports their structural and functional linkage. Observations from this dataset align with the IAWA list, in which reticulate parenchyma is common in families such as Annonaceae and Ebenaceae. In contrast, Chrysobalanaceae generally exhibits narrow banded parenchyma, with only limited occurrences of reticulate or scalariform patterns. Nonetheless, some species within *Maranthes* (Chrysobalanaceae) display reticulate parenchyma, notably *Maranthes gabunensis* (Tw25733) and *Maranthes glabra* (Tw103, Tw8219, Tw47797).

The clustering results provide insights into the degree of taxonomic coherence captured by the macroscopic cross-sectional anatomical patterns. Specifically, examining the taxonomic structure within each cluster reveals whether species belonging to the same genus or family tend to be grouped together or scattered across multiple clusters. Inconsistencies between taxonomy and clustering may reflect anatomical divergence, convergent features, or limitations in the clustering resolution. To assess this, a ring plot (Figure 3.9) was created to visualize the proportion of genera whose species were assigned to one or multiple clusters. The results show that approximately 75% of Congolese timber genera have all of their species assigned to a single cluster, suggesting strong anatomical consistency within these genera. Notably, several genera within the Fabaceae family—such as *Afzelia* and *Albizia*—are entirely assigned to cluster 2, indicating both taxonomic and anatomical cohesion. In contrast, about 20% of the genera exhibit inter-cluster species assignments. For example, species within the genus *Pterocarpus* (Fabaceae) are split across clusters: *P. angolensis*, *P. rotundifolius*, and *P. soyauxii* are assigned to cluster 2, whereas *P. tinctorius* is placed in cluster 4. A small number of genera—specifically nine—have species distributed across three clusters. For instance, the genus *Aningeria* includes *Aningeria adolfi-friederici* (cluster 3), *Aningeria altissima* (cluster 4), and *Aningeria pierrei* (cluster 6). The genus *Gambeya* shows the broadest dispersion, with species present in clusters 1, 3, 4, and 6.

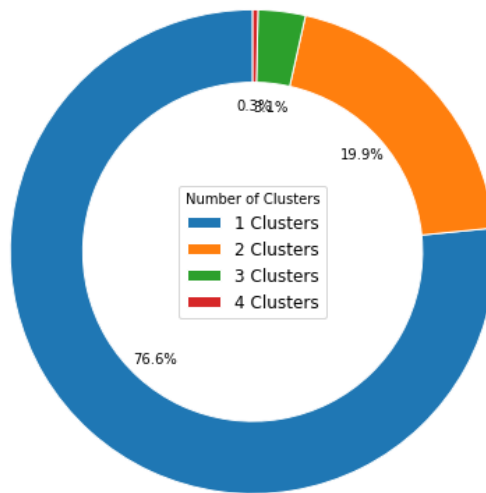


Figure 3.9: Distribution of Congolese timber genera assigned to one or more clusters (see Figure 3.12345). Genera with species assigned to a single cluster (Blue), assigned to two clusters (Orange), assigned to three clusters (Green), Assigned to four clusters (Red).

3.4.5 Classification models

3.4.5.1 Evaluating model performance across taxonomic levels

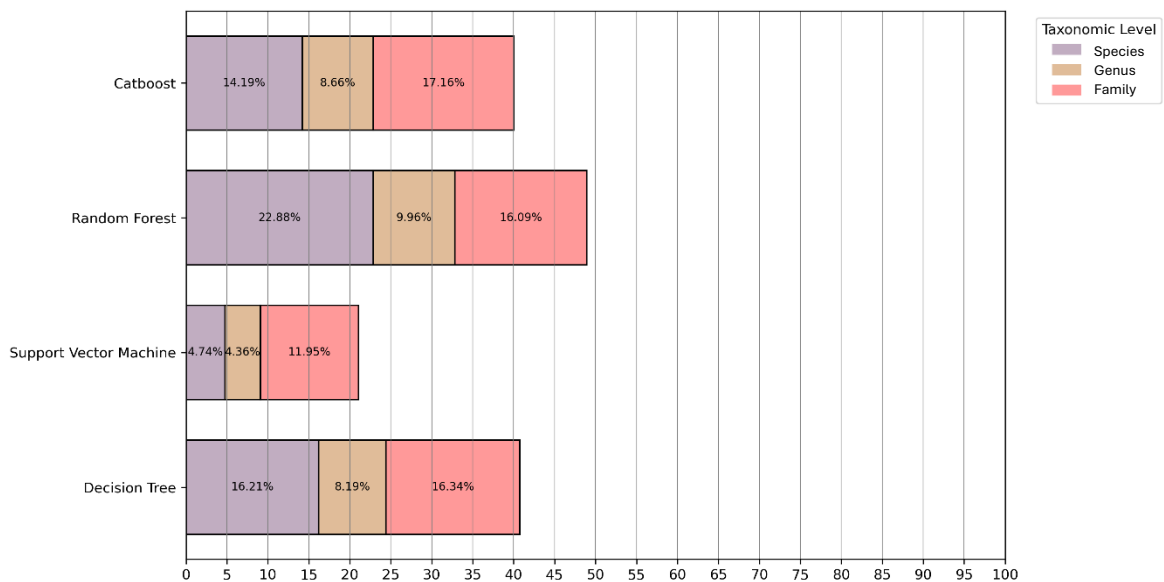


Figure 3.10: Stacked bar chart illustrating the accuracy of four different machine-learning classification techniques (named on the y-axis) across the full range of predicted Congolese tree species labels (in the SmartWoodID database); The stacked bar colours represent different taxonomic levels, with printed values representing the individual increase in accuracy gained at a higher taxonomic level.

The stacked bar chart (Figure 3.10) presents the cumulative classification accuracy achieved by each species classifier across progressively broader taxonomic levels. The purple bars, along with the printed values within them, represent the accuracy of species-level predictions. The yellow bars indicate the additional accuracy gained when species-level predictions are mapped to their corresponding genera,

resulting in a total genus-level accuracy of purple + yellow. Similarly, the pink bars reflect the further increase in accuracy when genus-level predictions are mapped to their respective families, yielding a total family-level accuracy of purple + yellow + pink. Overall, classification accuracy improves as taxonomic resolution broadens, though the magnitude of this improvement varies across methods, with RF exhibiting the most pronounced increase across taxonomic levels.

SVM demonstrated the weakest performance among classifiers, achieving the lowest species accuracy and showing modest gains at broader taxonomic levels, with genus-level accuracy increasing to 9.1% and family-level accuracy to a final 21.05%. Traditional DT classifiers outperformed SVM showing moderate results, with gains reaching 24.4% at the genus level and 40.74 % at the family level. Methods leveraging DTs in gradient frameworks or ensembles yielded nuanced results. CatBoost, a gradient-boosting method that refines DTs iteratively, showed comparable performance to DT and larger absolute family-level gains, culminating in 40.01% family-level accuracy. RF, combining multiple DTs as ensembles, effectively enhanced the performance of traditional DTs, achieving the highest species-level accuracy and demonstrated consistent improvements at higher taxonomic levels reaching the highest family accuracy (48.93%).

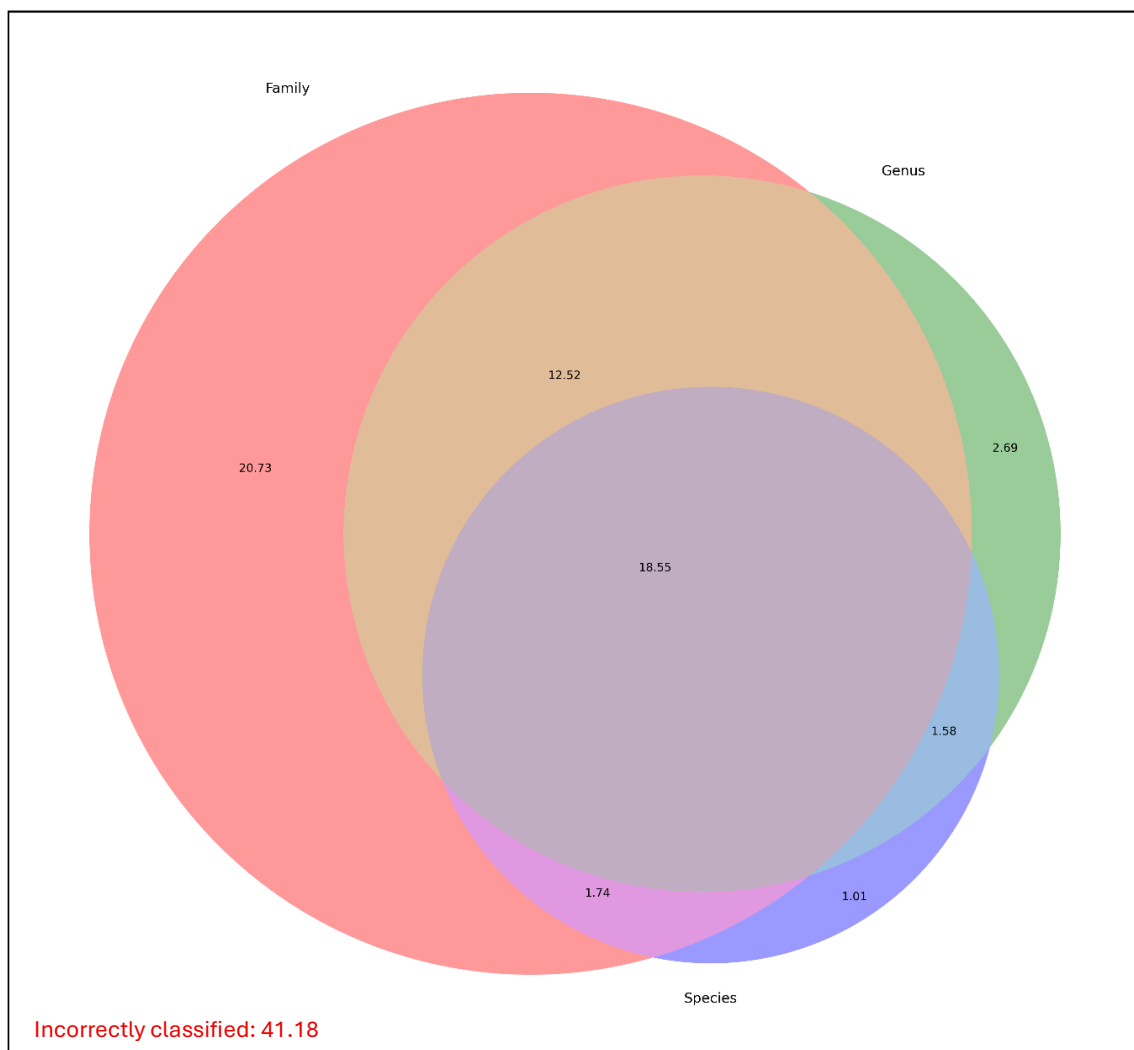


Figure 3.11: Venn diagram representing percentage of specimens correctly identified by three RF models (each trained directly at Species, Genus, and Family level); Intersected areas between models indicate percentage of specimens correctly identified by both techniques; Areas are proportionate to their printed percental value; The red printed value in the lower-left corner, represents the percentage of specimens misclassified across all taxonomic levels.

Machine learning models were also trained directly on genus and family labels. To evaluate the consistency of taxonomic predictions for the same specimens, an ensemble agreement analysis was conducted, focusing on the RF model due to its superior performance. The distribution of correctly classified

specimens across taxonomic levels was visualized using a Venn diagram (Figure 3.11), in which three overlapping circles represent the number of specimens accurately identified at the species, genus, and family levels. The relative area and numerical values within each section indicate the proportion of specimens correctly classified at one or more taxonomic levels. For example, 12.52% of specimens (highlighted in yellow) were correctly identified at the genus and family levels but misclassified at the species level. The white square enclosing the Venn diagram represents the total number of specimens misclassified at all taxonomic levels, with this subset highlighted in red in the lower-left corner of the figure.

To compare both methods, predicting species labels and mapping higher taxonomic labels according to known taxonomy, and predicting labels at all taxonomic labels directly, was compared. For RF, the accuracy for all specimens with correct species predictions, independent of misclassification on the higher taxonomic levels resulted in 22.88%, matching the accuracy in Figure 3.10. The majority of which also had correct genus and family predictions (18.55% of all specimens). By directly predicting the genus label, the genus accuracy rose with a percentile value of 15.21% (2.69% only at the genus level and 12.52% at genus and family level), while the genus predictions by mapping the species predictions only resulted in a rise of nearly 9.96%. A similar pattern could be observed for family predictions, where the mapped species level predictions resulted in an accuracy rise of 16.09% on family predictions, while the direct family level classification resulted in an accuracy rise of 20.73%. The other 4.33% having mismatches for both or individual higher taxonomic levels (each between 1-2%). Correct classification, solely at species level, occurred in 1.01% of cases. Mismatched predictions across taxonomic levels were observed for 144 different species, demonstrating that no species predominantly exhibited mismatched predictions, suggesting that classification inconsistencies cannot be attributed to specific species, genera, or families (see Supplementary materials Figure 8.2).

Despite the demonstrated advantages of training models specifically for different taxonomic levels, the overall accuracy of both approaches—species-derived predictions and direct family or genus classification—reveals the limitations of the macroscopic features. A significant proportion of specimens were entirely misidentified across all taxonomic levels, underscoring the challenge of achieving consistent accuracy. For instance, using the species predictions, the proportion of specimens misclassified at all levels was substantial: 59.26% for DT, 78.95% for SVM, 51.07% for RF, and 59.99% for CatBoost. Even with RF, the most effective technique, misclassification persisted for over half of the specimens. When applying the second approach of directly classifying family and genus levels using RF, the error rate improved but remained considerable, with 41.18% of specimens still misclassified across all levels.

3.4.5.2 *Optimal number of taxa for Reliable identification*

To study how the different machine-learning classification techniques rank predicted classes across taxonomic levels, classes were ranked in descending order based on their predicted probabilities. CDF plots were generated for each machine-learning technique and aggregated per taxonomic level to illustrate how accuracy improves as additional top-k predictions are considered. The progression of accuracy was functioned against two different parameters. First, the value of k which shows how accuracy cumulatively increased by considering k additional top predicted classes. Second, cut-off thresholds for the predicted probabilities which shows how sensitive the techniques are by applying increasing thresholds, leaving out classes with consecutively higher probabilities.

The CDFs against both parameters were visualized in different figures, where each taxonomic level is represented separately. At each level, the progression of top-k accuracy is depicted using aggregated coloured curves for each technique, obtained by averaging the values across the four validation folds. Additionally, a 95% cumulative threshold was computed for each taxonomic level and technique to identify the rank at which 95% of test specimens were correctly classified. These thresholds are indicated by vertical dotted lines corresponding to each technique. To assess the consistency of these thresholds across validation folds, a 95% confidence interval was calculated. This was done by determining the rank positions where the cumulative distribution curves exceeded 95%, computing the mean threshold rank, and deriving the standard error of mean (SEM). The SEM is visualized as semi-transparent shaded areas around each 95% rank threshold, with colours matching the respective techniques. The CDFs in function

of the value k were visualized in Figure 3.12. The CDFs in function of cut-off thresholds were visualized in Figure 3.13.

Table 3.2 provides further insight into model performance of the techniques across taxonomic levels summarizing three key metrics of the CDFs in function of number of classes (k): AUC, the 95% threshold rank, and the standard error of mean (SEM) around this threshold.

Table 3.2: Summary of metrics of the CDF plots for every Machine-learning technique across taxonomic levels: AUC, the 95% threshold rank, and standard error of mean (SEM) around this threshold.

Technique	DT			RF			CatBoost			SVM		
Metric	AUC	95% threshold rank	SEM	AUC	95% threshold rank	SEM	AUC	95% threshold rank	SEM	AUC	95% threshold rank	SEM
Family	0.703	62	0.80	0.936	21	6.28	0.908	22	3.98	0.942	17	1.30
Genus	0.880	262	2.05	0.902	197	24.94	0.930	92	17.11	0.630	279	2
Species	0.582	562	3.01	0.861	480	27.88	0.897	292	41.77	0.582	600	1

Across all models, accuracy improves as k increases (Figure 3.12) and declines as the cut-off probability threshold increases (Figure 3.13). The rank analysis (Figure 3.12) reveals that the DT classifier follows a predominantly linear growth pattern across taxonomic levels, with only modest accuracy gains as rank increases. This pattern suggests that DT extracts features that contribute minimally to class ranking, reflecting its basic and systematic nature. This is further reinforced by its low AUC values, as AUC values closer to 0.5 indicate a limited ability to effectively rank predicted classes (Table 3.2). The cut-off probability threshold analysis (Figure 3.13) further underscores this limitation—DT exhibits an immediate accuracy drop-off, suggesting that it assigns high confidence exclusively to its top predictions while neglecting lower-ranked classes. This indicates an overconfident classification approach, where all probability mass is concentrated on the highest-ranked class, resulting in poor ranking robustness.

In contrast, the other machine-learning techniques demonstrate greater ability to rank predicted classes. The rank analysis (Figure 3.12) shows that they follow a saturating growth pattern, characterized by rapid initial accuracy gains that gradually plateau as k increases. Among these, RF consistently ranks as one of the strongest performers, achieving high accuracy (Figure 3.12) and strong AUC values across taxonomic levels (Table 3.2). CatBoost also demonstrates strong performance, albeit with greater variability across taxonomic levels. While its top-1 accuracy is initially lower than RF's, its ranking ability improves as k increases, surpassing RF in ranking lower-probability classes (Figure 3.12). This is reflected in notably lower 95% threshold rank values, indicating that CatBoost requires fewer ranked predictions to classify 95% of test specimens correctly (Table 3.2). However, these advantages diminish at the family level, suggesting that CatBoost loses its edge when handling broader classification tasks with fewer, more general classes.

The cut-off threshold analysis (Figure 3.13) further highlights RF as the most resilient technique. RF exhibits the most gradual decline in accuracy across taxonomic levels, indicating that it effectively distributes probability across ranked predictions rather than overemphasizing the top prediction. In contrast, CatBoost's cut-off threshold curve (Figure 3.13) displays a sharp decrease in accuracy beyond the top predicted classes at all taxonomic levels. This suggests that CatBoost assigns a disproportionately large share of its probability mass to its top predictions, leaving little confidence distributed among lower-ranked classes.

The SVM exhibits the most pronounced performance gains at higher taxonomic levels. At the family level, its top- k accuracy curve (Figure 3.12) and AUC values closely align with those of RF, and it achieves a lower 95% threshold rank, indicating strong ranking performance (Table 3.2). However, as classification specificity increases, SVM's accuracy and AUC values decline, its 95% threshold rank increases, and its ranking ability for lower-probability classes deteriorates. A distinctive "tail effect" emerges, where accuracy stabilizes at high values of k before abruptly increasing at the final ranks. This effect becomes more pronounced at finer taxonomic levels, highlighting SVM's difficulty in ranking the least probable classes. The cut-off threshold analysis (Figure 3.13) confirms that SVM exhibits a more stable probability distribution across ranked predictions at higher taxonomic levels but also shows the greatest shift in probability distribution as taxonomic specificity increases.

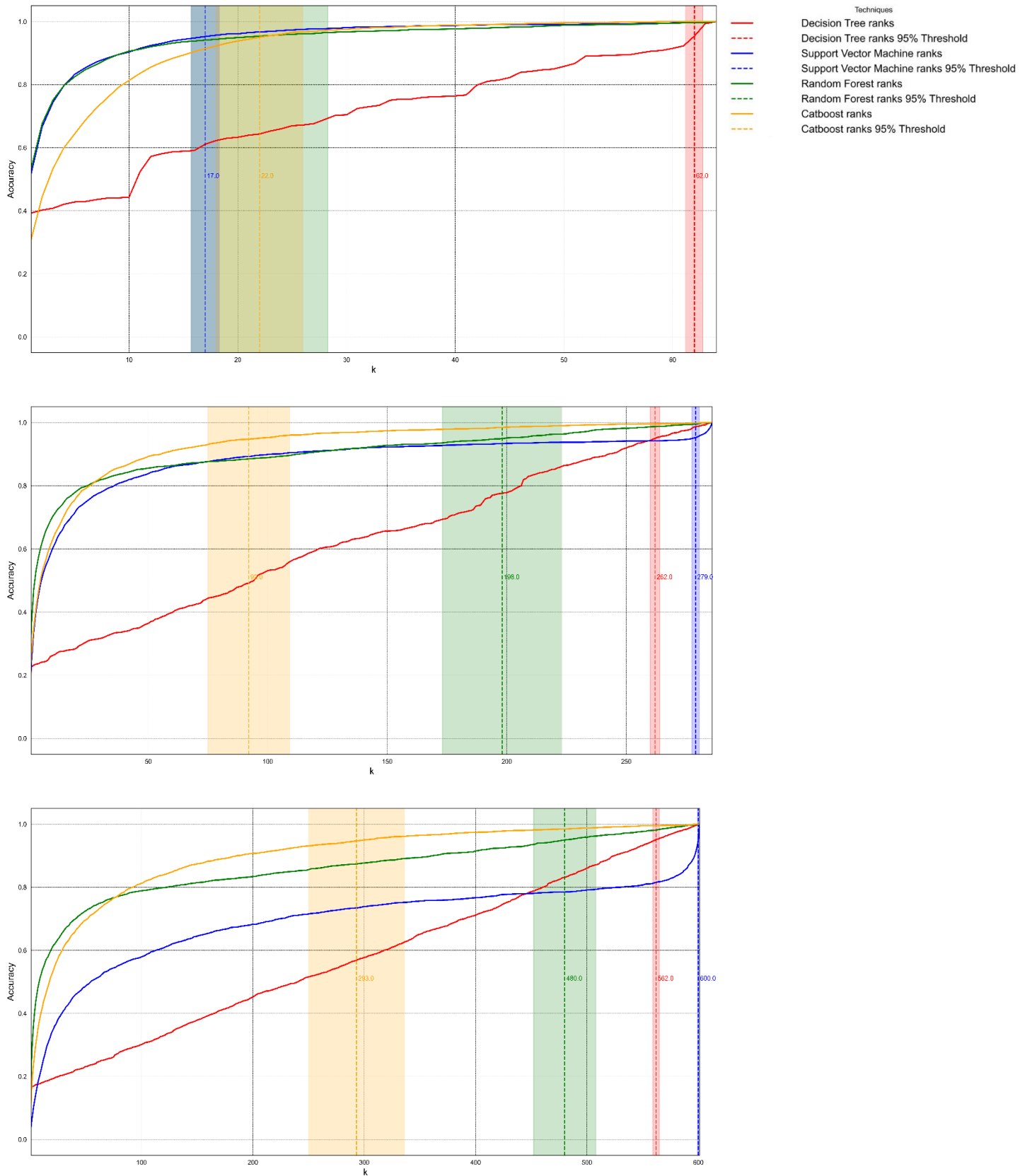


Figure 3.12: Cumulative distribution function plots, showing the increase in accuracy by considering k additional predicted classes for the four machine-learning techniques. The threshold, where each curve reaches 0.95 is indicated by a vertical dotted line with the standard error of mean around it, determined by aggregating the information across all 4 folds per technique. The top graph shows this for direct family predictions with x-axis range from 1 till the maximum number of classes (64 families), The central graph for direct genus predictions (286 genera), The bottom graph for direct species predictions (601 species).

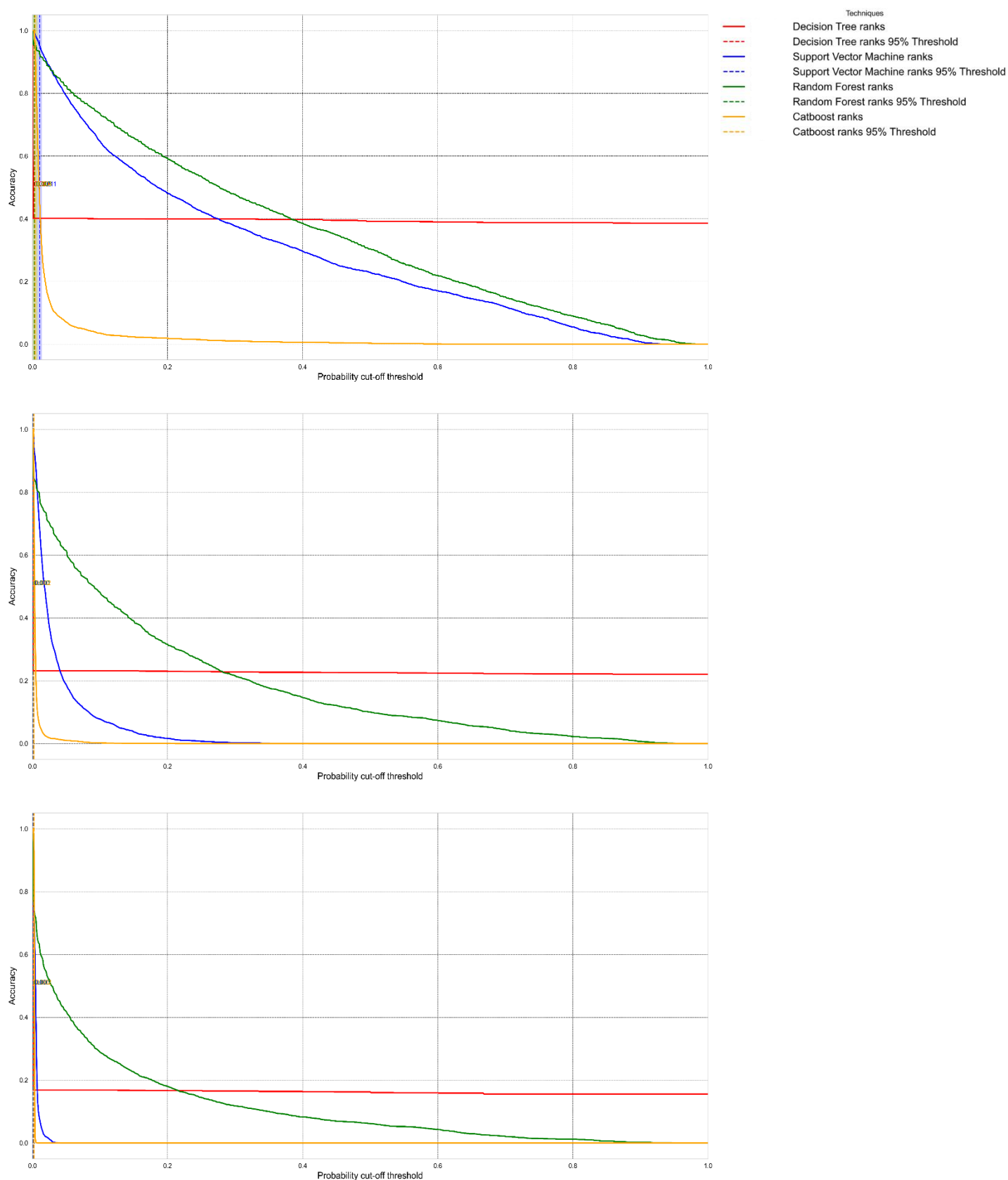


Figure 3.13: Cumulative distribution function plots, showing the decrease in accuracy by applying higher cut-off threshold values for leaving out predicted classes for the four machine-learning techniques. The threshold, where each curve reaches 0.95 is indicated by a vertical dotted line with the standard error of mean around it, determined by aggregating the information across all 4 folds per technique. The top graph shows this for direct family predictions with x-axis range from 0 till the maximum threshold value of 1, The central graph for direct species predictions, The bottom graph for direct genus predictions.

3.4.5.3 Refining species prediction

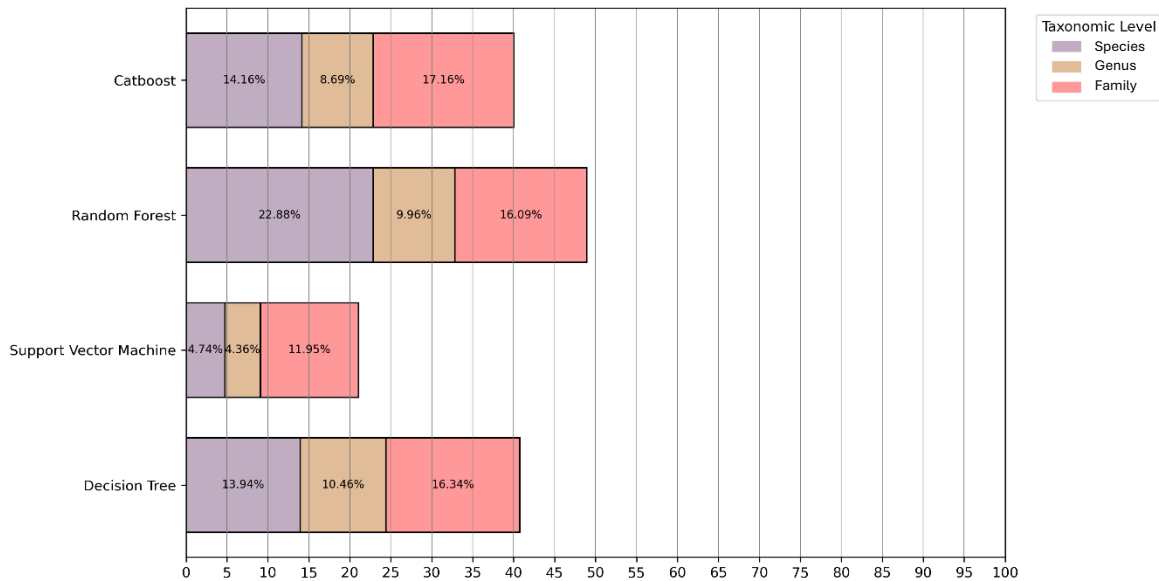


Figure 3.14: Stacked bar chart illustrates accuracy of four different machine-learning classification techniques (named on the y-axis) across the full range of predicted Congolese tree species labels (after filtering species predictions on the 25 top predicted families); The stacked bar colours represent different taxonomic levels, with printed values representing the individual increase in accuracy gained at a higher taxonomic level.

The rank analysis revealed that the top 20–25 directly predicted families allow the 95% accuracy threshold to be reached (3.4.5.2). These predicted family labels thereby also provide information for predictions at lower taxonomic levels. For example, predicted species labels (that do not belong to one of those families), can be ruled out, increasing accuracy. To assess this, species predictions were constrained to belong only to the top 25 families. Next, the same type of bar chart was produced as Figure 3.10, showing well the different machine learning techniques perform at genus and family levels, by mapping the filtered species predictions to their corresponding genus and family labels and re-evaluating performance. The new accuracy values were visualized in Figure 3.14, mirroring the representation of Figure 3.10. Marginal differences between both figures indicate that direct family prediction (using a reliable shortlist of 25 families) does not significantly enhance species predictions for any techniques. Only DT (decision tree) demonstrated different accuracy on species and genus levels, although the cumulative accuracy of the three taxonomic levels does not improve.

3.4.5.4 Assessing Recognition of InsideWood Descriptions for Commercial Species

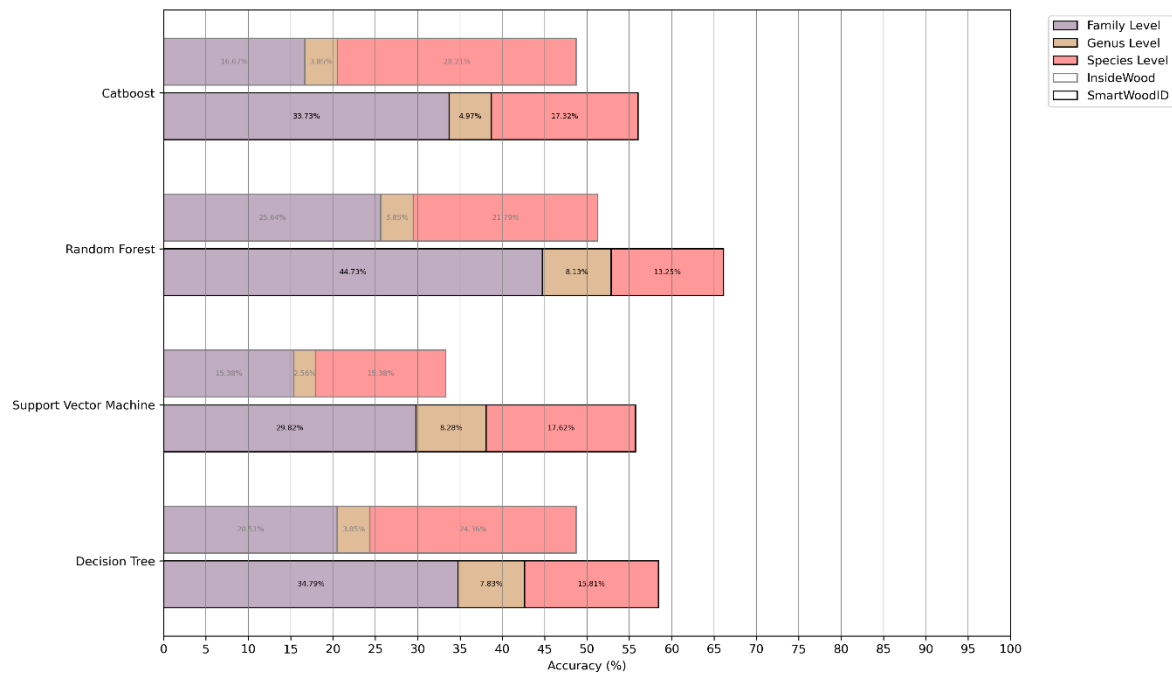


Figure 3.15: Stacked bar chart illustrates accuracy of four different machine-learning classification techniques (named on the y-axis) across Commercial Congolese tree species; The stacked bar colours represent different taxonomic levels, with printed values representing the individual increase in accuracy gained at a higher taxonomic level. The bars with grey borders show accuracy on the InsideWood descriptions, bars with black borders on SmartWoodID test specimens.

To evaluate the applicability of macroscopic wood anatomical descriptions for identifying Congolese timber in trade, species predictions were visualized using horizontal stacked bar charts. These charts illustrate accuracy progression across taxonomic levels, grouped by machine-learning technique. For each technique, two stacked bars represent results based on InsideWood descriptions (grey border) and SmartWoodID descriptions (black border) (see Figure 3.15).

Accuracy values in Figure 3.15 are notably higher than those in Figure 3.10, indicating that all techniques perform better on the commercial subset than on the full set of Congolese tree species. However, overall accuracy remains low, highlighting the limitations of macroscopic wood anatomical features for reliable field identification of commercial species.

A comparison between SmartWoodID and InsideWood descriptions reveals consistently lower accuracy across all taxonomic levels when using InsideWood data. This discrepancy suggests that models struggle with descriptions based on other material (thin sections, variations in resolution and field of view). Notably, the proportional accuracy gain from genus to family level is larger for InsideWood predictions, indicating that macroscopic wood anatomical descriptions from external sources may be more stable at broader taxonomic levels.

3.5 Discussion

The results indicate that macroscopic cross-sectional wood anatomical features do not enable accurate species predictions across most Congolese tree species. The two-way clustering analysis of macroscopic anatomical features across species corroborates this (section 3.4.4). Co-varying patterns across botanical species were observed on the heatmap. Namely, vessel arrangement, density, and diameter, were often grouped reflecting fundamental hydraulic trade-offs (Carlquist and Hoekman, 1985). Parenchyma types also clustered into meaningful groups, with banded, reticulate, and scalariform parenchyma forming one cluster, while lozenge-aliform, winged-aliform, and confluent parenchyma highlighted a structural continuum. In contrast, diffuse-in-aggregate parenchyma remained distinct, reinforcing the mutual exclusivity of species featuring either apotracheal or paratracheal parenchyma. Clustering the 601 species based on those accessible features in the field resulted in an optimal number of six broad clusters according to Mantel test. This underscores the limited ability to discern Congolese timber with macroscopic cross-sectional features.

Regarding the classification analysis, mapping species-level predictions to their corresponding genus and family labels, confirmed improvement at broader taxonomic levels (3.4.5). This reflects the hierarchical nature of taxonomy and the phylogenetic relationships present in wood anatomy (de Luna et al., 2018; Kobayashi et al., 2019). Training models directly at the genus and family levels further supports this trend. Additionally, mismatching classifications (for example: family and species correct, but genus incorrect) occur across a diverse range of species rather than being concentrated in specific taxa, confirming that errors stem from general limitations in macroscopic anatomical features rather than model biases. These findings suggest that models perform more reliably when applied to broader taxonomic categories, where reduced anatomical variability enhances classification performance. However, it is important to note this may partly result from the smaller number of classes at broader taxonomic levels, simplifying the classification problem for the machine-learning models and inherently raising performance (Li et al., 2015; Sanlı et al., 2020). The results on the commercial subset (3.4.5.4), elucidates this further as accuracy is higher across all machine-learning techniques and taxonomic levels, because of the strong reduction in look-a-like timbers. Despite these trends, overall accuracy remains low at higher taxonomic levels and across all machine-learning techniques.

The analysis of top-k rankings and probability cut-off thresholds reveals that all models struggle to rank predictions effectively, often assigning disproportionate confidence to top-ranked classes (3.4.5.2). Among them, RF demonstrates the most stable ranking ability and probability distribution. Several studies have demonstrated the potential of RF as a successful classification technique, suggesting that the dataset is likely the limiting factor explaining the poor performance (Biau and Scornet, 2016; Salman et al., 2024; J. Zhao et al., 2024). This highlights the necessity of considering multiple top-ranked predictions to mitigate misclassification risks in macroscopic timber identification, underscored by De Oliveira et al. (2019) noting the importance of considering classes beyond solely the best prediction (De Oliveira et al., 2019). The high number of taxa (families, genera, and species) required to achieve 95% accuracy, illustrates the limitations of macroscopic cross-sectional features for accurately identifying tree species across the full range of Congolese tree species. Refining species predictions through filtering on higher-level taxonomic predictions (e.g. top 25 families to attain 95% accuracy threshold) resulted only in marginal gains, confirming those limitations (section 3.4.5.3).

The low performance across taxonomic levels emphasizes the constraints of macroscopic cross-sectional features for differentiating a large and diverse set of timbers. To further elucidate, Figure 3.16 visualizes the anatomical features of different taxa, using *Millettia laurentii* (Fabaceae) and *Ficus bubu* (Moraceae) as a clear example. When examined side by side, these species are easily distinguishable based on differences in colour and vessel size. However, their formal macroscopic feature descriptions are nearly identical, with the only recorded distinction being *Millettia laurentii*'s tendency toward confluent axial parenchyma. Vessel diameter of both species is assigned to class ($>130\text{ }\mu\text{m}$), while *Millettia laurentii* exhibits noticeably larger vessel diameters within this broad range. This underscores that fine-scale differences such as colour

variation and quantitative features measurements are not adequately captured within the framework of the standardized macroscopic descriptors.

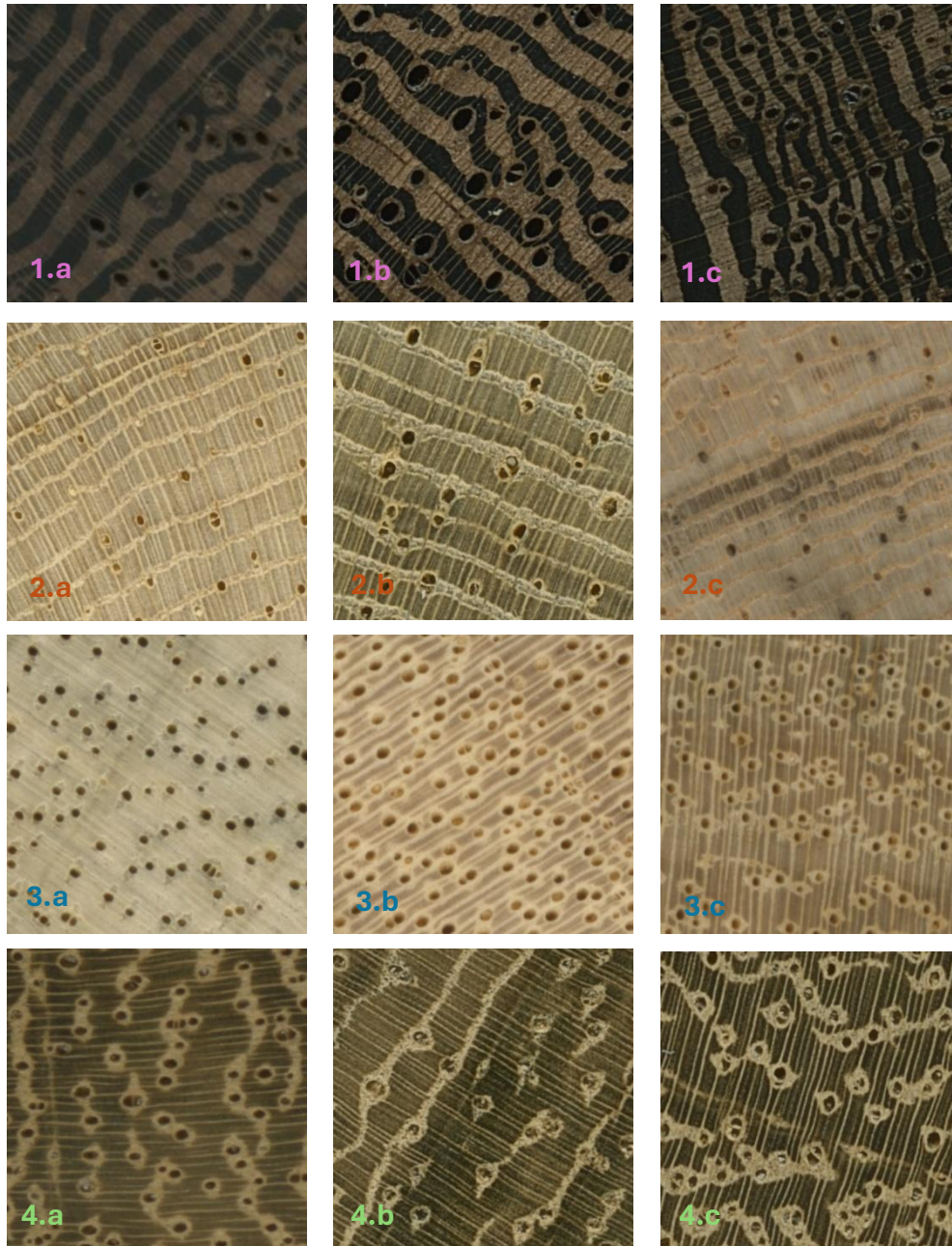


Figure 3.16: Macroscopic cross-sectional wood anatomy images (2400 dpi / 5.42x5.42mm) displaying differences between 1. *Millettia laurentii* (Fabaceae) (a: Tw3894, b: Tw425, c: Tw5227), 2. *Ficus bubu* (Moraceae) (a: Tw4824, b: Tw7602, c: Tw990), 3. *Antiaris toxicaria* (Moraceae) (a: Tw31032, b: Tw7338, c: Tw7568), and 4. *Milicia excelsa* (Moraceae) (a: Tw1121, b: Tw1462, c: Tw51741).

However, this does not imply that the macroscopic wood anatomical features contain no diagnostic value. Applying stricter hierarchical constraints on the classification analysis (e.g. within a single family or genus) revealed how feature combinations enable classification of species (undistinguishable across the full range of Congolese tree species). The clustering analysis shows this, as species within the same genus (e.g. *Aningeria*) were assigned to different broad clusters, indicating that in limited ranges, these features can still provide diagnostic value for identification. Visualizing the anatomy elucidates this further, as demonstrated for the Moraceae family (visualized on Figure 3.16). The genus of *Ficus* can be distinguished clearly from other Moraceae genera by differences in axial parenchyma expression: *Ficus* predominantly

exhibits axial parenchyma bands wider than three cells {85}, whereas other genera, including *Antiaris* and *Milicia*, primarily display confluent axial parenchyma {83}. Additionally, *Ficus* tends to have fewer than four rays per millimetre {114}, while other Moraceae genera typically range between four and twelve {115}. The results for *Milicia excelsa* and *Antiaris toxicaria* underscore the potential for distinguishing species within a predefined taxonomic range. While identification of these species failed when considering the entire Congolese timber dataset (see Table 8.3), their anatomical profiles remain distinct within Moraceae. The combination of vasicentric axial parenchyma {79}, confluent axial parenchyma {83}, and a vessel frequency of 5–20 vessels per mm² {47} differentiates *Milicia excelsa* and *Antiaris toxicaria* from other Moraceae species. Moreover, *Milicia excelsa* can be further distinguished from *Antiaris toxicaria* by the presence of axial parenchyma in marginal or seemingly marginal bands {89}, a feature predominantly found in *Milicia excelsa* but absent in *Antiaris toxicaria*.

The Lamiaceae species underscores this as vessel porosity enables clear distinction between species within this family. *Tectona grandis* displays clear ring-porous patterns on two specimens (Tw3805, Tw11055) and semi-ring-porous patterns on the others (Tw13935, Tw767). On the other hand, *Premna angolensis* and *Vitex* spp., display only limited semi-ring porosity, and *Gmelina arborea* consistently exhibits diffuse porosity. The Boraginaceae family, consisting only of the genus *Cordia* in Congolese tree species, mirrors this pattern, with *Cordia africana* being distinct by semi-ring porosity in three out of the four specimens (Tw2106, Tw6982, Tw7345). Similarly, in Bignoniaceae, semi-ring porosity is largely confined to the *Markhamia* genus, while other genera such as *Spathodea*, *Kigelia*, and *Stereospermum* exhibit diffuse porosity.

The genus *Pterocarpus* underscores this further still, indicating how individual species (*Pterocarpus tinctorius*, *Pterocarpus soyauxii*, *Pterocarpus rotundifolius*, and *Pterocarpus angolensis*) can be distinguished (see Figure 3.17). All species within this genus were listed in CITES Appendix II at CoP19, as existing literature indicated that reliable field identification was not possible without access to advanced scientific techniques (CITES, 2022a). To elaborate, previous studies reported *Pterocarpus* species as indistinguishable without laboratory-based techniques such as microscopic wood anatomy, DART-TOFMS, and fluorescence spectrometry (Liu et al., 2023; Price et al., 2021). *Pterocarpus rotundifolius* and *Pterocarpus angolensis* exhibit semi-ring porosity {4} and vessel frequency of 5–12 vessels per mm² {47}, whereas *Pterocarpus tinctorius* and *Pterocarpus soyauxii* have fewer than five {46}. *Pterocarpus tinctorius* and *Pterocarpus soyauxii* are more difficult to differentiate but display slight differences in vessel diameter distributions, varying between medium (80–130 µm) and large (>130 µm), with *Pterocarpus soyauxii* displaying a higher prevalence of large vessels (medium: variable / Large: present), while *Pterocarpus tinctorius* displays the opposite. The results suggest that readily observable features, such as macroscopic cross-sectional anatomy (Koch et al., 2018; Ruffinatto et al., 2015), facilitate species differentiation within visually similar, commercially important timbers, underscoring the potential for refining of international enforced protection lists for threatened timber species.

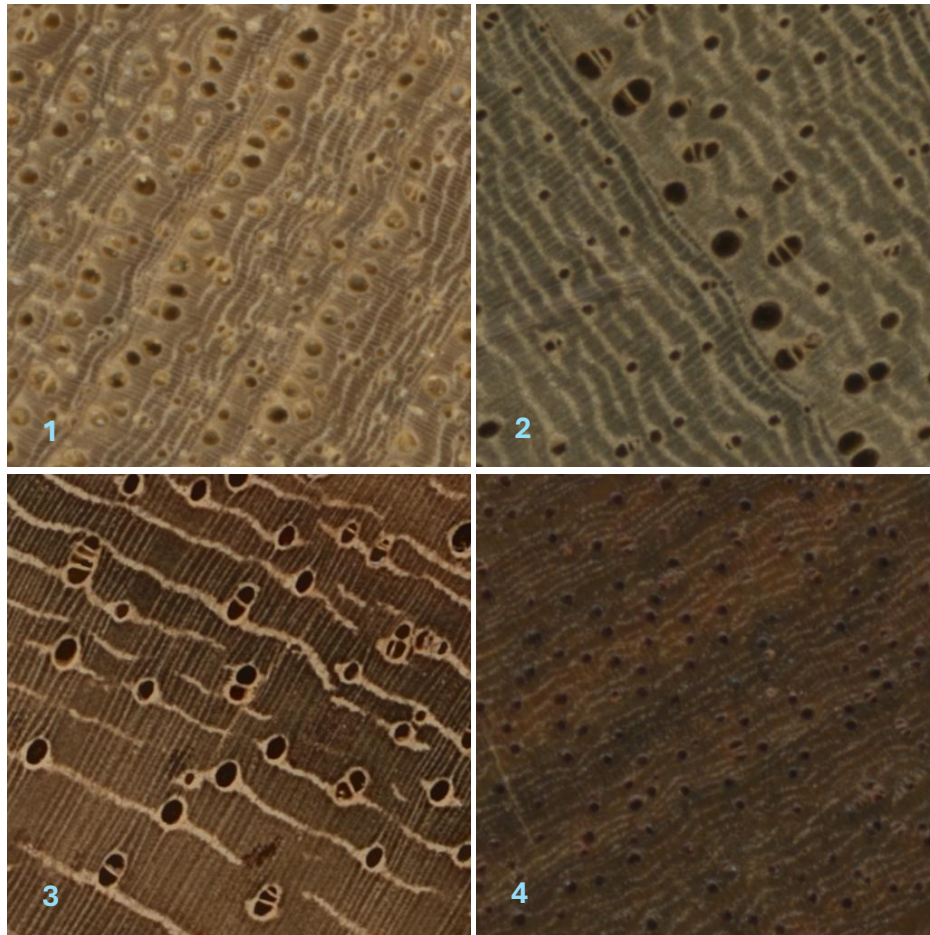


Figure 3.17: Macroscopic cross-sectional wood anatomy images (2400 dpi / 5.42x5.42mm) displaying differences between 1. *Pterocarpus rotundifolius* (Tw28123), 2. *Pterocarpus angolensis* (Tw768), 3. *Pterocarpus soyauxii* (Tw7654), 4. *Pterocarpus tinctorius* (Tw1010).

The ability of SmartWoodID-trained models to identify InsideWood species descriptions illustrates the advantages of specimen-based databases that incorporate large surface areas, such as SmartWoodID (De Blaere et al., 2023), providing more comprehensive insights into wood anatomical variation (De Blaere et al., 2023). Species accuracy is consistently lower on InsideWood descriptions across all models, and genus accuracy is also reduced, highlighting the difficulty of resolving fine taxonomic distinctions within a generalized dataset (section 3.4.5.4). However, accuracy at the family level improves significantly for InsideWood descriptions, suggesting that broader taxonomic classifications mitigate the effects of overgeneralization and feature variability. Lower accuracy on InsideWood descriptions can be attributed to the following factors: First, although wood anatomists follow standardized IAWA definitions, subtle subjective differences in interpretation can arise (Gasson et al., 2011; NS, 1989). Second is the structural difference between the datasets. InsideWood provides species-level descriptions compiled from multiple sources rather than direct observations of individual specimens, potentially leading to overgeneralization and a loss of specimen-specific detail (Wheeler et al., 2020). In contrast, SmartWoodID descriptions are specimen-based, maintaining a direct link between observed features and physical specimens, and explaining the observed higher accuracy (De Blaere et al., 2023). Third, field-of-view of specimens can significantly impact identification. Traditional anatomical descriptions often rely on thin sections covering only $\sim 1 \text{ mm}^2$, whereas SmartWoodID scans encompass a much larger area ($\sim 7 \text{ cm} \times 1\text{--}2 \text{ cm}$) (De Blaere et al., 2023). This broader perspective helps reveal structural variability and patterns that could be misinterpreted or missed entirely when focusing on a small region. For instance, a narrow field might suggest banded parenchyma, while a wider view could reveal an interrupted pattern indicative of confluent parenchyma (NS, 1989). The impact of these database differences was observed for images of *Pterocarpus* (see Figure 3.17), as key diagnostic features in the SmartWoodID database—such as the exclusively

solitary vessel arrangement ($\geq 90\%$) {9} distinguishing *Pterocarpus angolensis* within the genus *Pterocarpus*—were absent from the InsideWood database (Wheeler, 2011). Additionally, studying the anatomy of *Pterocarpus* (Figure 3.17) revealed previously unrecorded trends in vessel diameter distribution between *Pterocarpus soyauxii* and *Pterocarpus tinctorius*, which were both broadly categorized as $>200\ \mu\text{m}$ in InsideWood (Wheeler, 2011) despite notable differences in observed distribution patterns, corroborated in the results. These findings underscore the value of expanding anatomical databases through large-area imaging and multi-specimen analysis to enhance the accuracy of species differentiation.

However, the findings on differentiating *Pterocarpus tinctorius* and *Pterocarpus soyauxii* also underscore the need for enhanced feature extraction and alternative approaches to improve accuracy, as the broad categorization of macroscopic anatomical features (e.g. vessel diameter distributions varying between medium ($80\text{--}130\ \mu\text{m}$) and large ($>130\ \mu\text{m}$), with *Pterocarpus soyauxii* displaying a higher prevalence of large vessels (medium: variable / Large: present), while *Pterocarpus tinctorius* displays the opposite) and lack of fine-scale detail due to low resolution may limit classification between closely related taxa. Traditional methods, relying on expert-defined features, may miss critical information in raw anatomical images (Knauer et al., 2019). This stems in part not from the anatomy itself but from the subjective interpretation of most quantitative features within the IAWA framework. This is particularly evident for quantitative features such as vessel frequency, lumen diameter, and ray width, which are grouped into broad categorical states (NS, 1989). While these simplified categories (e.g., present, variable, absent) facilitate practical identification, more precise descriptors—such as means, ranges, and standard deviations—may better capture the inherent variability of wood samples (Beeckman and Yin, 2024; He et al., 2019; Van den Bulcke et al., 2025). Extracting such quantitative information presents additional challenges, as it requires precise measurements that account for the natural variability within and between species. This highlights the need for robust database construction methods and retrieval algorithms capable of effectively handling these quantitative features. Addressing this issue necessitates the preparation of large surface areas for analysis, digitization at a resolution high enough to enable accurate feature measurements (e.g., vessel diameter), and the development of advanced computational methods for automated feature extraction to enable measuring numerous individual features (e.g. vessel, rays). At 2400 dpi (or 10.58 micron), SmartWoodID captures visually interpretable anatomical features but may introduce ambiguities for finer structures (De Blaere et al., 2023). For example, axial vasicentric parenchyma — parenchyma cells forming a complete circular to oval sheath around a solitary vessel or vessel multiple (NS, 1989) — may appear indistinguishable from vessel walls when thin layers are present, causing potential misclassification. Similarly, banded parenchyma is classified based on band height in cell count (NS, 1989), but variations in cell size within a band can blur distinctions between wide- and narrow-banded parenchyma. Recent advances in automated digitization of wood anatomical surfaces, through robotics and neural network-based segmentation, provide a promising avenue for overcoming these limitations enabling fast and accurate digitization of large cross-sectional surfaces at higher resolution (2.25 micron), and automating measurement of large number of quantitative features (NS, 1989; Van den Bulcke et al., 2025).

It is important to recognize, however, that enhancing species-level identification accuracy through advanced imaging techniques and analysis of large wood surfaces remains difficult to implement under typical field conditions. In contexts where high volumes of timber must be rapidly assessed, the use of accessible tools—such as utility knives, sanding blocks, and low-cost imaging devices—does not support the acquisition of high-resolution data required for such advanced analyses. Nevertheless, research in CV has demonstrated that deep learning models applied directly to macroscopic cross-sectional images can effectively differentiate between closely related taxa across diverse timber types, including those commonly encountered in tropical regions (Ravindran et al., 2021, 2020, 2018). CNNs, in particular, show promise for extracting taxonomically informative features from end-grain images by leveraging subtle variations in colour, texture, and anatomical patterning, thereby enabling more accurate and scalable wood identification (Hwang and Sugiyama, 2021; Knauer et al., 2019).

3.6 Conclusion

This study highlights the inherent limitations of the 31 expert-defined, accessible macroscopic cross-sectional features for taxonomic identification across a diverse range of timber species, such as those found in the Congo Basin. While classification accuracy improves at higher taxonomic ranks, genus- and family-level predictions remain limited due to overlapping anatomical features among taxa. Nevertheless, macroscopic cross-sectional features retain diagnostic value when applied within narrower taxonomic scopes. The successful discrimination of *Pterocarpus* species—once considered indistinguishable without laboratory-based methods—demonstrates that readily observable anatomical features can enable species-level identification in the field. Further improvements in diagnostic accuracy can be achieved by incorporating high-resolution, large-area imaging and multi-specimen datasets. These approaches more effectively capture intra-specific anatomical variability than conventional single-sample methods and enable the extraction of quantitative anatomical information at a finer scale. Integrating such enhancements offers promising pathways to increase both taxonomic resolution and classification reliability. However, practical constraints in field environments—such as time pressures and limited equipment—necessitate alternative strategies for reliable in situ wood identification. Continued progress will depend on advancing CV-based identification systems, particularly CNNs, which can directly process macroscopic images to deliver accurate and rapid classification. These models offer strong potential for scalable, efficient, and field-ready timber verification applications.

3.7 Acknowledgements

This chapter was realized through the combined efforts of the staff at the service Wood biology of the RMCA in Tervuren, Belgium, and the staff at the UGent-Woodlab (Department of Environment, Faculty of Bioscience Engineering) of Ghent University. Describing wood anatomical features was aided by job students Michael Monnoye, Miro Cnops, Senne Suykerbuyk, and volunteers Tibo Deckers and Michèle Florquin.

3.8 Availability of code and metadata

Scripts and metadata can be requested by contacting the first author:

Contact: ruben.de.blaere@africamuseum.be

Chapter 4: Evaluating the effect of anomalous images on CV-based wood identification models

Authors: Ruben De Blaere^a, Kévin Lievens^a, Victor Deklerck^a, Tom De Mil^d, Wannes Hubau^{a,b}, Hans Beeckman^a, Jan Verwaeren^c, Jan Van den Bulcke^b

a Service of Wood Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium, b UGent-Woodlab, Laboratory of Wood Technology, Department of Environment Ghent University, Coupure Links 653, 9000 Gent, Belgium, c UGent-KERMIT, research unit Knowledge-based, predictive and spatio-temporal modelling, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Gent, Belgium, d TERRA teaching and research center, Gembloux Agro-Bio Tech (Université de Liège), Passage des Déportés 2, 5030 Gembloux, Belgium, e Meise Botanic Garden Nieuwelaan 38, 1860 Meise, Belgium

4.1 Abstract

Automating wood identification through CV offers improved objectivity, time-efficiency, and accuracy over traditional methods, using expert-defined features. Conventional wood anatomical assessments rely on intact mature tissue, avoiding damage (cracks, fungi deterioration, insect damage) and other anomalies (pith, bark, traumatic canals). The impact of using images from anomalous surfaces on automated identification remains underexplored in current research.

This chapter evaluates the performance of CNNs for classifying the presence of anomalies on images, and studies the impact of anomalies on genus identification by in- or excluding image of anomalous surfaces in the training data and assessing recall on the test data. The Xception network architecture was used to train the two types of classification models, on macroscopic cross-sectional images of 26 Congolese wood genera. The first model was trained for binary classification on the presence or absence of anomalies on >250.000 images of ~1000 Congolese tree species, demonstrating accuracy, precision, recall and f1-score of ~93% on 25.000 test images. This shows that CNNs can learn patterns to detect the presence of anomalies. The second model was trained and evaluated on a subset of those Congolese tree species, consisting of 26 timber genera with abundant different types of anomalies (cracks, fungi deterioration, insect damage, pith, bark, traumatic canals). Three different wood identification models were trained and evaluated on the images featuring a model trained only on all images (regardless of anomalies), a second model trained only on perfect (anomaly-free) images, and a third model trained only on images with anomalies. The three models were evaluated on different specimens and demonstrated macro-averaged recall scores of 88.4, 90.5%, and 79.1% for the respective models, showing that a model trained on images from intact end-grain wood/anomaly-free images performed best. Class (genus) specific recall scores demonstrated for the three models that model performance varies between genera. The class (genus) specific recall scores of *Millettia*, *Tessmannia*, *Celtis*, *Azelia*, *Beilschmiedia*, and *Vitex* are highest for the model a trained on all images (with and without anomalies). Conversely, the recall scores of *Cynometra* and *Microcos* were lower for this model. Grad-CAM analysis was performed to visualizes which regions on images were more activated for classification (wood identification), and revealed that the model focuses more on anomaly-free regions for wood identification, underscoring the importance of clear wood anatomy in training CNNs for wood identification.

This chapter was accepted for publication in the peer-reviewed Journal *Wood Science and Technology* (Springer) on 19/08/2025

4.2 Introduction

Wood identification has gained increasing interest due to rising concerns over sustainable sourcing, combating illegal logging, and ensuring product quality and authenticity (Hwang and Sugiyama, 2021; Silva et al., 2022). Traditional manual methods for wood identification can suffer from subjectivity, time inefficiency, and limited taxonomic accuracy. However, with the emergence and development of AI, there is a promising avenue for automatizing wood identification and providing additional insight in wood identification in a fast and accessible way (Andrade et al., 2020; Hwang and Sugiyama, 2021). Deep learning algorithms, and CNNs in particular, can learn complex patterns and features from large datasets, leading to more robust and accurate identification outcomes (Alzubaidi et al., 2021; Sarker, 2021; Taye, 2023).

CV-based wood identification, which emerged over the past decades, is a practise that uses images of wood as a source of diagnostic information (Hwang and Sugiyama, 2021; Silva et al., 2022). Often, researchers use macroscopic cross-sectional images to develop models that can distinguish timbers on multiple taxonomic levels (Hwang and Sugiyama, 2021; Silva et al., 2022). Such models have been achieving remarkable results in several studies in the last decade and often advocate for direct applicability in the field (Ravindran et al., 2021, 2020, 2019; Tang et al., 2018; Wiedenhoef, 2020).

Unlike many other CV domains, wood identification presents distinctive challenges. While advancements in CV have improved model robustness to variations in lighting, resolution, and image quality (e.g., blur) (Shorten and Khoshgoftaar, 2019), wood often exhibits physical anomalies that obscure key anatomical features and complicate classification (Goodell and Nielsen, 2023; Niemz et al., 2023a; Schmidt, 2006). As a biological material, wood is subject to degradation from disease, infestation, and physical stress. Insects and marine borers can damage wood, by removing wood material, and fungi and bacteria can cause discoloration and decay (Goodell and Nielsen, 2023; Schmidt, 2006). Furthermore, wood can crack, especially during drying (Niemz et al., 2023a). These anomalies can obscure diagnostic anatomical structures, thereby complicating identification.

In comparison, wood anatomists typically prefer anomaly-free mature heartwood tissue for assessment because identification protocols and image databases such as atlases traditionally avoid juvenile wood (Burley, 2004). The main reason why juvenile wood is not used, is because its anatomy is highly variable even within a specimen (e.g. different average vessel diameter compared to mature tissue), making them less ideal for taxonomic identification. Furthermore, commerce usually targets larger trees for use as timber, which emphasizes the need for using mature tissue for developing CV-based wood identification models, which will likely encounter the same mature wood in traded cargo. Sapwood is considered valuable for wood identification, as it retains some diagnostic anatomical features.

Aforementioned examples of anomalies are known to be often encountered in the field, but official numbers are seldom published. Similarly, there is little to no public record of anomalous surface area on collection material in wood collections. An example of a wood collection database that records the occurrence of anomalies on collection material is SmartWoodID (De Blaere et al., 2023). The overall presence of aforementioned anomalies on the SmartWoodID collection reveals that the majority of specimens portray anomalies on more than 50% of their surface area.

However, the impact of these anomalies on CV-based wood identification remains largely unexplored. In related topics, Ravindran et al. (2023) has demonstrated that suboptimal surface preparation by cutting or sanding can hamper CV-based wood identification. Furthermore, Owens et al. 2024 has systematically tested how CNN predictions are affected by digital perturbations mimicking real-world wood degradation. This underscores the need to study the impact of the aforementioned anomalies on CV-based wood identification.

Furthermore most other studies on CV-based wood identification have relied exclusively on anomaly-free specimens (Hwang and Sugiyama, 2021; Ravindran et al., 2021; Silva et al., 2022), overlooking the imperfections typically encountered in applied contexts. Ravindran et al. 2018, highlighted in their methodology that anomalies were annotated and excluded from training CNN for wood identification to rule-out influence caused by anomalies. To ensure reliable field deployment, it is therefore essential to

evaluate how such anomalies influence model predictions—and to develop mitigation strategies that improve classification resilience under realistic conditions.

While recent studies (Owens et al., 2024; Ravindran et al., 2023) have assessed how models trained on clean, anomaly-free wood images perform when evaluated on imperfect test images, they have largely overlooked the potential impact of incorporating such anomalous images directly into the training process. As a result, the influence of training data quality—particularly the inclusion or exclusion of anomalies—on model robustness and generalizability remains insufficiently understood.

This study addresses the aforementioned gaps by systematically evaluating how wood anomalies affect both the prediction performance of CNNs for wood identification under different images training datasets. Our investigation is structured around two central components. First, we trained a binary classification model to detect images with anomalies, exploring whether CNNs can reliably detect the diverse types of wood anomalies—such as cracks, fungal decay, insect damage, bark, pith, traumatic canals—and thereby enabling automated filtering of anomalous images data from training sets. This contributes to the field, as a means to automatically clean datasets according to the standards currently and widely used in CV-based wood identification, e.g. to avoid the use of anomalous images.

Second, we investigated the influence of anomaly inclusion in the training data on wood identification performance. Using the same CNN architecture, we trained three separate models on different image sets: (1) a balanced mix of anomaly-free and anomalous images, (2) only anomaly-free images, and (3) only anomalous images. All three models were trained to classify the same 26 Congolese timber genera using standardized macroscopic cross-sectional image patches. Each model was then evaluated on a test dataset composed of both anomaly-free and anomalous images, providing insight on how well each of those three training dataset configurations, perform on images independent test specimens with and without anomalies. Model performance was assessed using macro-averaged recall, allowing comparison across trained models and across individual classes (genera). By comparing these trained models and analysing classification behaviour, this study aims to clarify the role of anomalous image content in CNN-based wood identification.

To further enhance interpretability, we employed Grad-CAM, which visualizes the image regions most influential in the model's predictions. These visualizations are targeted to deliver insight into whether models trained on perfect, anomaly-free images rely on diagnostically relevant features (e.g. the clean anatomy) or are distracted by present anomalies.

4.3 Material and Methods

4.3.2 Dataset Description

This chapter is based on the SmartWoodID database, serving as a valuable resource for examining the relationship between macroscopic cross-sectional wood anatomy and the botanical diversity of Congolese tree taxa (De Blaere et al., 2023). The following information in this section, re-uses the text in section 1.5 which provides an overview of all data used in each chapter.

The database contains high-resolution RGB scans of the macroscopic end-grain surfaces of 3,742 wood specimens, representing 954 species native to the DRC. Each specimen was prepared by scanning the cross-section at 2400 dpi using a flatbed scanner. This resolution allows for the visualization of macroscopic features essential for wood identification.

Given the large size of the scanned surface, working directly with the full images is computationally demanding and may lead to overfitting due to the limited number of images. Therefore, full images are cropped into smaller patches that still contain the required diagnostic features and capture the inherent variability across different specimens. We opted for patches of 512x512 pixels, which corresponds to a surface area of 5.42x5.42 mm, providing sufficient detail to effectively distinguish damage such as insect holes, cracks, and fungal growth, while also preserving the necessary macroscopic features visible on the end-grain surface for accurate wood identification (see Figure 4.1). Figure 4.1 shows the different types of anomalies regarded, upper from left to right: a) Traumatic canals (*Millettia hockii* / Tw28835), b) insect damage (*Prioria gilbertii* / Tw3601), c) fungal discoloration (*Ficus vallis-choudae* / Tw44423). Lower from

left to right, it shows d) cracks (*Afzelia bipindensis* / Tw26431), e) pith (*Brachystegia spiciformis* / Tw2002), f) bark (*Diospyros bipindensis* / Tw35708).

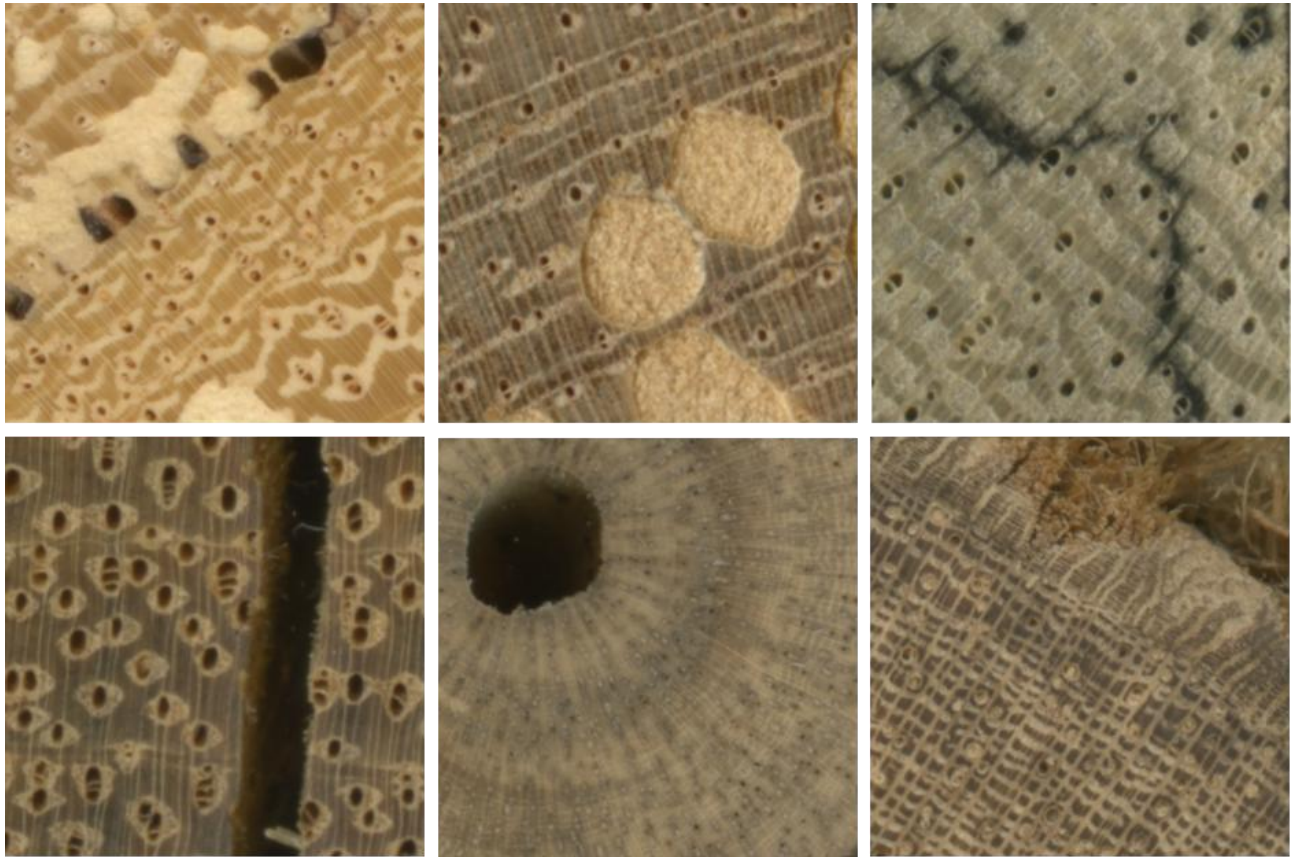


Figure 4.1: Anomalies on the end grain timber: a-c(upper/left-to-right) traumatic canals, insect damage, fungal discoloration, d-f(bottom/left-to-right) cracks, pith, bark.

The image patches were labelled on the presence of anomalies. Patches without visible wood anatomy were excluded from the training and validation data.

All full images with less than four patches were also left out, as their total endgrain surface area would accumulate to less than one square cm, a normal area size for anatomical assessment by humans.

4.3.3 Model architecture

Xception was used as a backbone due to its balance between required computational power and its satisfactory performance on classification of the ImageNet database [10] (a visualisation of the Xception architecture can be consulted in Supplementary materials Figure 8.3). The pre-trained Xception architecture (on the ImageNet database) was fine-tuned without freezing layers, using a batch size of eight and RMSprop as optimizer with a learning rate of $1e-4$. Class weighting was employed during training to address class imbalance. Early stopping was implemented to mitigate overfitting by tracking the validation loss progression during training and preserving the model with the lowest validation loss. Training started with one hundred epochs, with early stopping triggered if there was no improvement in validation loss for ten consecutive epochs with less than 0.1.

The following data augmentation techniques were applied to improve model robustness: random rotation (0-20 degrees), shearing (0-10 degrees), horizontal/vertical flipping, colour shifting (in a single, randomly selected, channel with -10-10 range) and random gaussian blurring (sigma range 0-4) (Owens et al., 2024). However, zooming and/or shifting the image along its height or width was avoided as that could inadvertently obscure anomalies. The CNN was implemented using TensorFlow (version 2.6.0, Python version 3.9.15). Pandas (version 1.5.3) and Matplotlib (version 3.7.1) were used for analysis and visualisation of the results.

4.3.4 Classification of image patches by presence or absence of anomalies

The CNN was then customised to classify image patches by presence or absence of anomalies. The presence of anomalies is essentially a binary image classification problem. Patches labelled as ‘containing anomalies’ were assigned to the positive class, patches labelled as anomaly-free to the negative class.

In summary, the entire SmartWoodID database contains 4740 images cropped to 255712 patches from which 44.76% are annotated as ‘anomalous’ and 55.24% as ‘anomaly-free’. This ensures a balanced representation to develop CNN for anomaly verification (De Blaere et al., 2023). All patches were randomly divided into training (60%), validation (30%), and test sets (10%).

The Xception CNN architecture was implemented as described in section 2.2. The classification head was replaced with a GlobalAveragePooling layer and a fully connected layer with Sigmoid activation. Binary cross-entropy was used as the loss function and binary accuracy as the evaluation metric. The performance was assessed using the Precision, Recall, Accuracy and by visually inspecting individual misclassifications.

Mislabeled patches were reported and corrected before training models on genus classification to ensure labeling to be correct.

4.3.5 Classification of image patches by genus

Classification is targeted at the genus level due to the high interspecific anatomical resemblance within most genera, and the limited number of specimens per species in the SmartWoodID database. Twenty-six botanical genera were selected for training the model architecture. As selection criterion, the number of specimens per genus was considered, prioritizing genera with a balanced representation of anomalous and anomaly-free specimens. For further comprehensive details regarding the species used in this study, see Table 8.4 in Supplementary materials.

The specimens were allocated exclusively into either the training/validation or the test set to ensure fair evaluation. The specimens were grouped by species, before allocation into training and test sets, to preserve a balanced representation of species, ensuring the model's ability to manage anatomical differences across all species in a genus. This resulted in two datasets with approximately 80% of the patches as material for training and 20% as test data. After the initial split, the image patches for training were grouped, by the presence of anomalies, and consecutively divided into training (80%) and validation (20%) using a stratified sampling approach. This preserved proportional representation of specimens and anomalies in both datasets, and aims to minimize class imbalance.

Due to the large variability in specimen size, the number of patches per specimen is very different, leading to a strong imbalance in the training set. Therefore, oversampling was implemented by randomly selecting additional patches from smaller specimens. Specifically, for each genus, the average number of patches was computed in the training data. Smaller specimens, with fewer patches than the genus-level mean, were then oversampled by randomly selecting additional patches until the number of patches matched the genus-level mean. Each new patch was evaluated by the binary anomaly classifier, ensuring balanced representation of anomalous and anomaly-free patches in the training data. For validation, patches were randomly sampled without replacement from specimens that were not oversampled, until the desired number of validation patches was obtained. This prevented data leakage to the validation data. For further comprehensive details regarding the specimen distribution in the training and test sets, see Table 8.5 in Supplementary materials.

The CNN architecture was implemented as described in section 2.2. The classification head was replaced with a global average pooling layer and a fully connected layer with SoftMax activation. Categorical cross-entropy was used as the loss function and accuracy as the evaluation metric on the training and validation data. Three training dataset configurations were investigated. The first model featured the entire training set with both anomaly-free and anomalous images. The second model and third model use exclusively anomaly-free patches and anomalous patches, respectively. The performance across models was evaluated by monitoring the evolution of accuracy and loss during training (for more information see Supplementary materials: 8.2) and comparing the recall scores on the test data. Recall with Macro

Averaging was selected as the evaluation metric on the test data to offer comprehensive assessment of CNN performance across and within the trained models. Recall measures the proportion of true positive predictions (correctly identified patches) out of all actual positive patches. Macro-Averaged Recall is a metric that computes the arithmetic mean of recall scores across different classes (here timber genera). For each model, the recall scores were used to compare, when discerning either anomaly-free patches or anomalous patches. To ascertain the model ability at correctly identifying the genus of entire specimens, the most frequently classified class was determined for every specimen. If the resulting majority vote differed from the true genus, the specimen was regarded as misidentified (False). The Boolean scores per specimen were then aggregated to Recall values per class and were further averaged across classes to obtain a "Specimen Recall score". In addition, differences in recall scores for individual classes were studied. By averaging the recall values for each class, Macro Averaging accounts for class imbalances and variations in dataset characteristics, ensuring a fair comparison of model performance and offering insights into the model's behaviour on individual classes in different training dataset configurations (three different models). Finally, Grad-CAM was used to elucidate patterns in the extracted features to classify anomalous timber images. Anomalous images were classified by the model trained only anomaly-free images. The generated heatmaps, represented as red-blue gradients, were overlaid onto those patches to pinpoint critical classification areas (Selvaraju et al., 2020; Wang and Zhang, 2023). The final convolutional layer's output was selected for Grad-CAM generation due to its ability to capture essential hierarchical features for the classification.

4.4 Results

4.4.1 Classification of patches by anomalies

The test set consisted of 32,493 image patches, comprising 16,217 anomaly-free patches and 16,276 anomalous patches. Each patch was classified by the model as either anomaly-free or anomalous, yielding four possible classification outcomes: true positive (anomaly-free patches correctly classified as anomaly-free), true negative (anomalous patches correctly classified as anomalous), false positive (anomalous patches incorrectly classified as anomaly-free), and false negative (anomaly-free patches incorrectly classified as anomalous), as illustrated in Figure 4.2. The resulting performance metrics derived from the confusion matrix are as follows: accuracy of 93.1%, precision of 91.7%, recall of 94.8%, and an F1-score of 93.2%.

At the centre of each quadrant of the confusion matrix, representative image patches are shown with overlaid Grad-CAM heatmaps. These visualizations, rendered in a blue-to-red colour gradient, indicate the regions within each image patch that contributed most strongly to the model's classification decision. In the true negative category, the model consistently focuses on visually prominent anomalous features such as cracks (Figure 4.2.s), fungal hyphae (Figure 4.2.t), insect boreholes (Figure 4.2.u), pith (Figure 4.2.x), bark (Figure 4.2.v), and traumatic canals (Figure 4.2.w). In contrast, the heatmaps corresponding to true positives display broader regions of activation, often spanning substantial portions of the patch (Figure 4.2a-f).

False negatives offer an opportunity to examine whether anomaly-free patches exhibited latent characteristics that may have led the model to misclassify them. Upon visual inspection, 270 of the 840 false negative patches were found to contain overlooked anomalies, suggesting mislabelling in the ground truth data. Heatmaps of these mislabelled patches, also shown in Figure 4.2, exhibit strong activation in areas corresponding to these hidden anomalies. For instance, Figure 4.2.g highlights insect damage, while Figure 4.2.h, i, and k show activation over traumatic canals, cracks, and bark tissue, respectively—indicating that the model detected the anomalies.

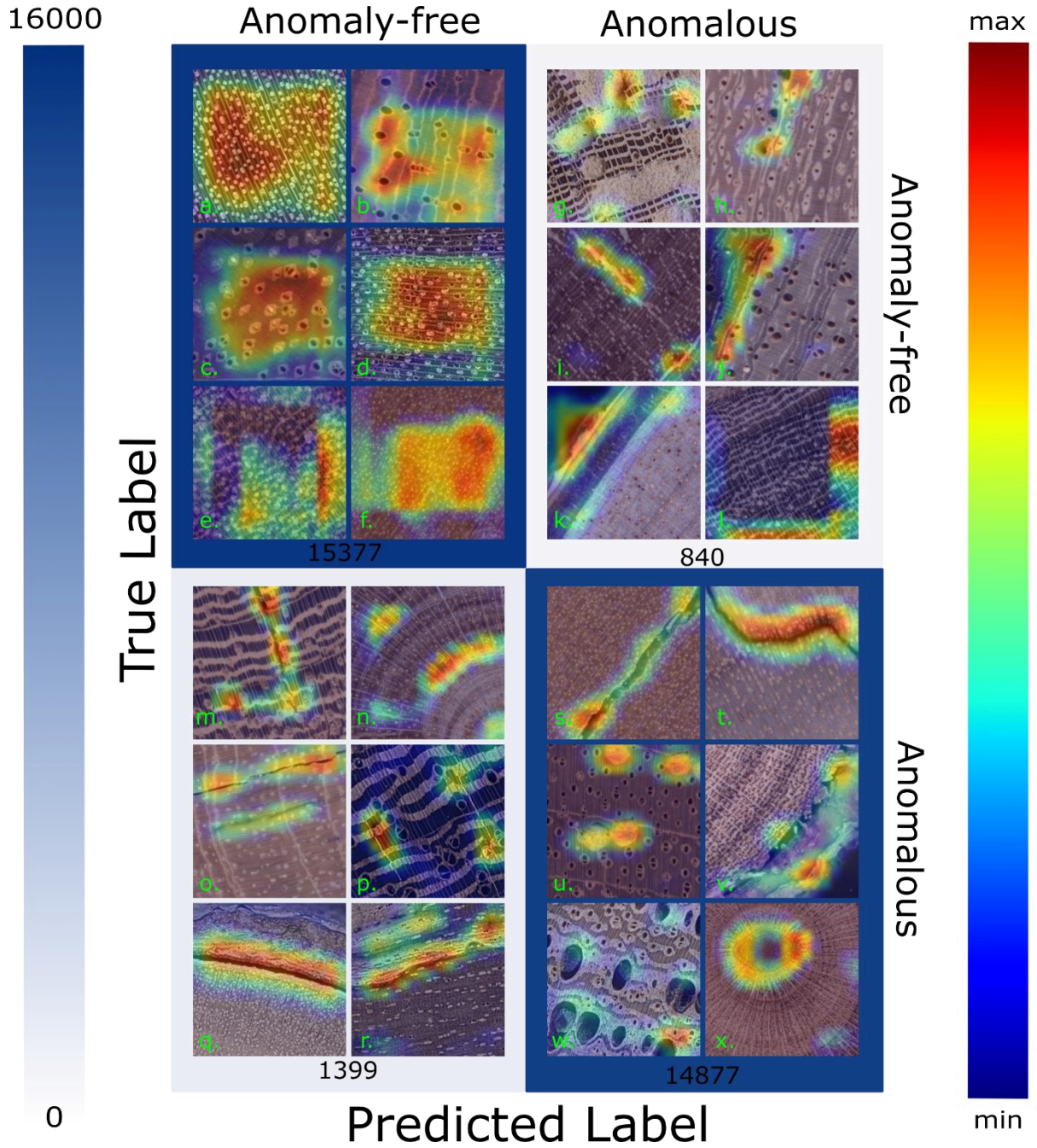


Figure 4.2: Confusion matrix showing the number of True positives (upper-left quadrant), False positives (lower-left quadrant), False negatives (upper-right quadrant), True positives (lower-right quadrant) of the binary anomaly detector model. The quadrants are coloured by a linear transparent blue colour gradient (shown on the left), in relation to the number of image patches (printed at the bottom of each quadrant). Each quadrant shows exemplary images overlayed with Grad-CAM to provide interpretation output in each quadrant. The level of activation of the Grad-CAM is shown on the right with a multi-colour gradient from blue to red. Pixel that have no colour overlay were not activated by the model during feature extraction. Letters (green) serve as unique identifiers for corroboration in text.

The remaining 570 false negative patches were confirmed as correctly labelled (i.e., truly anomaly-free). In these cases, the heatmaps generally lacked consistent activation patterns. Figure 4.2.j provides an example where the model appears to have responded to colour contrast between heartwood and sapwood, which may have been mistakenly interpreted as indicative of fungal discoloration. Figure 4.2.l exemplifies the majority of patches in this group, where no systematic visual features were consistently highlighted, suggesting uncertainty or over-sensitivity in the model's anomaly detection in these contexts. As framed in Section 2: Material and methods, the 270 mislabelled patches were relabelled to avoid

including this into the datasets for the genus classification models studied next. These mislabelled patches represent less than 1% of all test patches and thereby affirm that the labelling errors were rare. After updating the labels, this resulted in accuracy, precision, recall, and F1-score of 94.1%, 92.0%, 96.4%, and 94.2% respectively.

These observations highlight the importance of rigorous dataset validation. They also demonstrate the model's capacity to localize anomalies, even in the presence of possible label noise in the training data.

Further insights were gained from examining the false positive predictions. In these cases, the model identified anomalous features, but the corresponding patches were labeled as anomaly-free. Representative examples are presented in Figure 4.2. In Figure 4.2.m, o, and p, the model clearly activated on visible cracks. Figure 4.2.q shows the model responding to the boundary between wood and bark tissue, while Figure 4.2.r highlights areas affected by fungal degradation. Figure 4.2.n presents a patch where pith tissue was activated, in addition to other regions lacking visible anomalies.

4.4.2 Classification of patches by genus

4.4.2.1 Model performance on test data

Table 4.1 summarizes the macro-averaged recall scores for three training configurations (rows) evaluated on three filtered test sets (columns). For each configuration, recall is reported at the patch level (Recall (%)) and at the specimen level (Specimen Recall (%)), where the latter is based on majority voting across patches belonging to the same specimen.

Table 4.1: Overview of recall scores across training configuration (Rows represent different models trained with either both anomaly-free images and anomalous images, only anomaly-free images, and only anomalous images) and across filtered test data (columns represent the test patches, with the first column (from-the-left) showing the result on all patches, the centre column showing the recall based only on anomaly-free test patches, and the third column (from-the-left) showing the recall based only on anomalous test patches).

	All test patches (1)		Only anomaly-free test patches (2)		Only anomalous test patches (3)	
	Recall (%)	Specimen Recall (%)	Recall (%)	Specimen Recall (%)	Recall (%)	Specimen Recall (%)
Model trained on anomalous and anomaly-free patches (1)	86.8	92.7	88.4	92.9	80.0	91.9
Model trained only on anomaly-free patches (2)	87.0	95.8	90.5	95.6	77.1	95.1
Model trained only on anomalous patches (3)	78.4	89.1	79.1	90.3	73.2	88.1

The highest recall on anomaly-free test patches (column 2) was obtained by the model trained exclusively on anomaly-free training patches (row 2), exceeding the recall of the same model on anomalous patches (column 3) by 13.4 percentage points. This disparity indicates reduced generalization of the anomaly-free-trained model to unseen anomalous samples. Despite this, the same model (row 2) achieved the highest specimen-level recall overall (column 1) (95.8%), with minimal variance between specimen recall scores on anomaly-free (column 2) and anomalous (column 3) test subsets (95.6% and 95.1%, respectively), suggesting stable genus prediction of entire specimens despite variability at the patch level.

The model trained on both anomaly-free and anomalous patches (row 1) exhibited the most balanced recall between test subsets, with a difference of 8.4 percentage points between recall on anomaly-free (column 2) and anomalous test patches (column 3). Its recall on all patches (column 1) was comparable to the model trained only on anomaly-free patches (row 2), differing by only 0.2 percentage points. The specimen recall was higher across the test data (all columns) by ~3.0 percentage points if anomalous images were excluded from the training set (row 1 to row 2).

The model trained only on anomalous patches (row 3), yielded the lowest recall across all conditions. Compared to the model trained on all patches (row 1), overall recall (column 1) was 8.4 percentage points lower (78.4% vs. 86.8%), with reductions of 9.3 and 6.8 percentage points on anomaly-free (column 2) and anomalous (column 3) test subsets, respectively. Specimen recall was also lower to the other models (row 1 and 2), with an average of 89.2% and a higher variance (1.21%).

Across all three configurations, specimen recall consistently exceeded per-patch recall, with an average improvement of 10.1 percentage points and reduced variability across test conditions. This emphasizes the limited effect of adding or leaving out anomalous images of wood from the training data, on classifying test specimens based on the available surface (anomaly-free or anomalous).

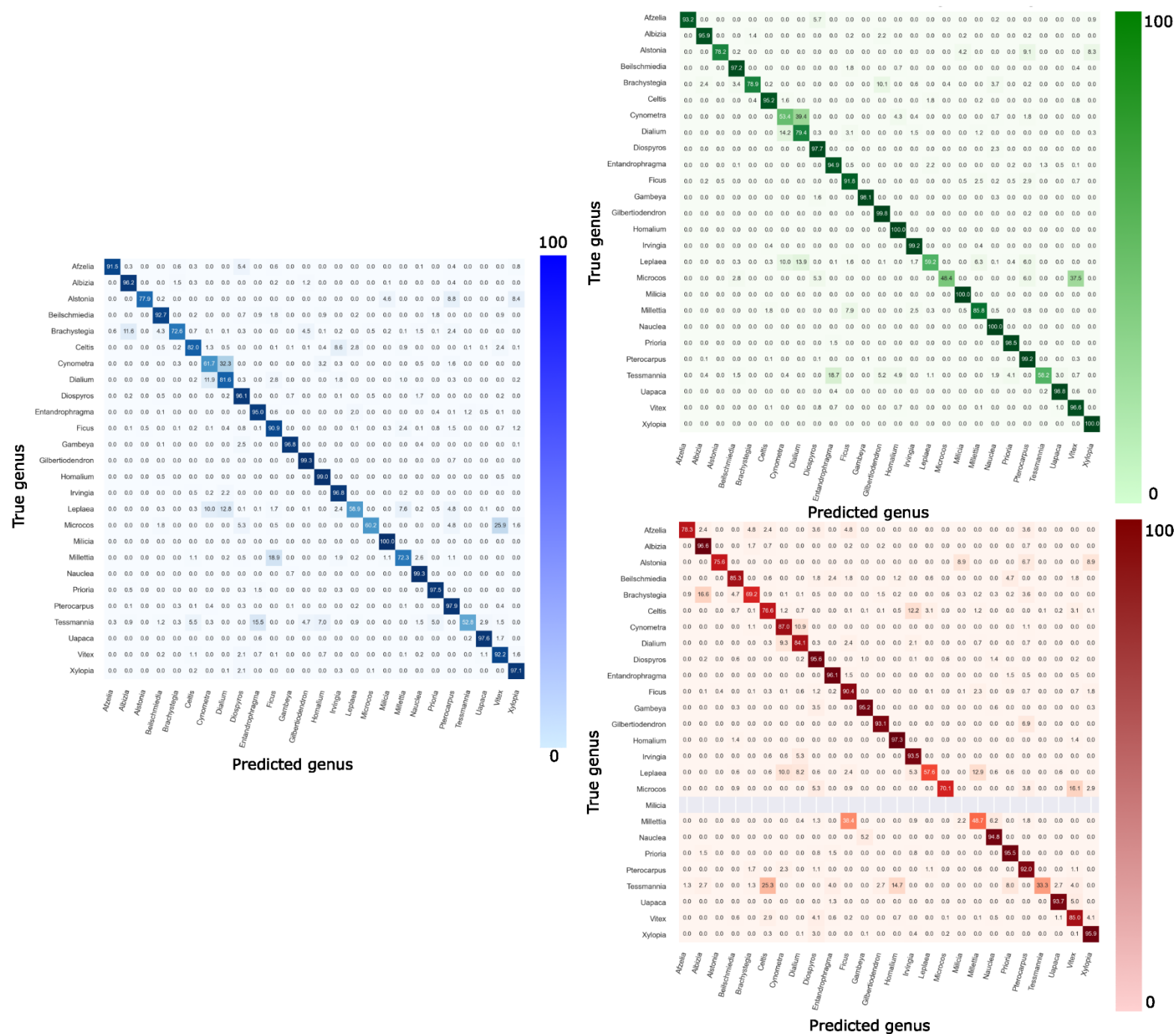


Figure 4.3: Confusion matrices showing the percentual recall scores for the 26 classes (genera) on the test patches, evaluated by the **model trained on both anomaly-free and anomalous images**. For each confusion matrix, the actual genus label is presented as rows, and the predicted label is presented as columns. The cells of the confusion matrix are coloured by a linear colour gradient shown on the right of every confusion matrix. The three matrices each represent the result on a different part of the test data. The blue confusion matrix (left) represents the results on all test patches. The green confusion matrix (upper-right) represents the results only on the anomaly-free test patches. The red confusion matrix (lower-right) represents the results only on the anomalous test patches.

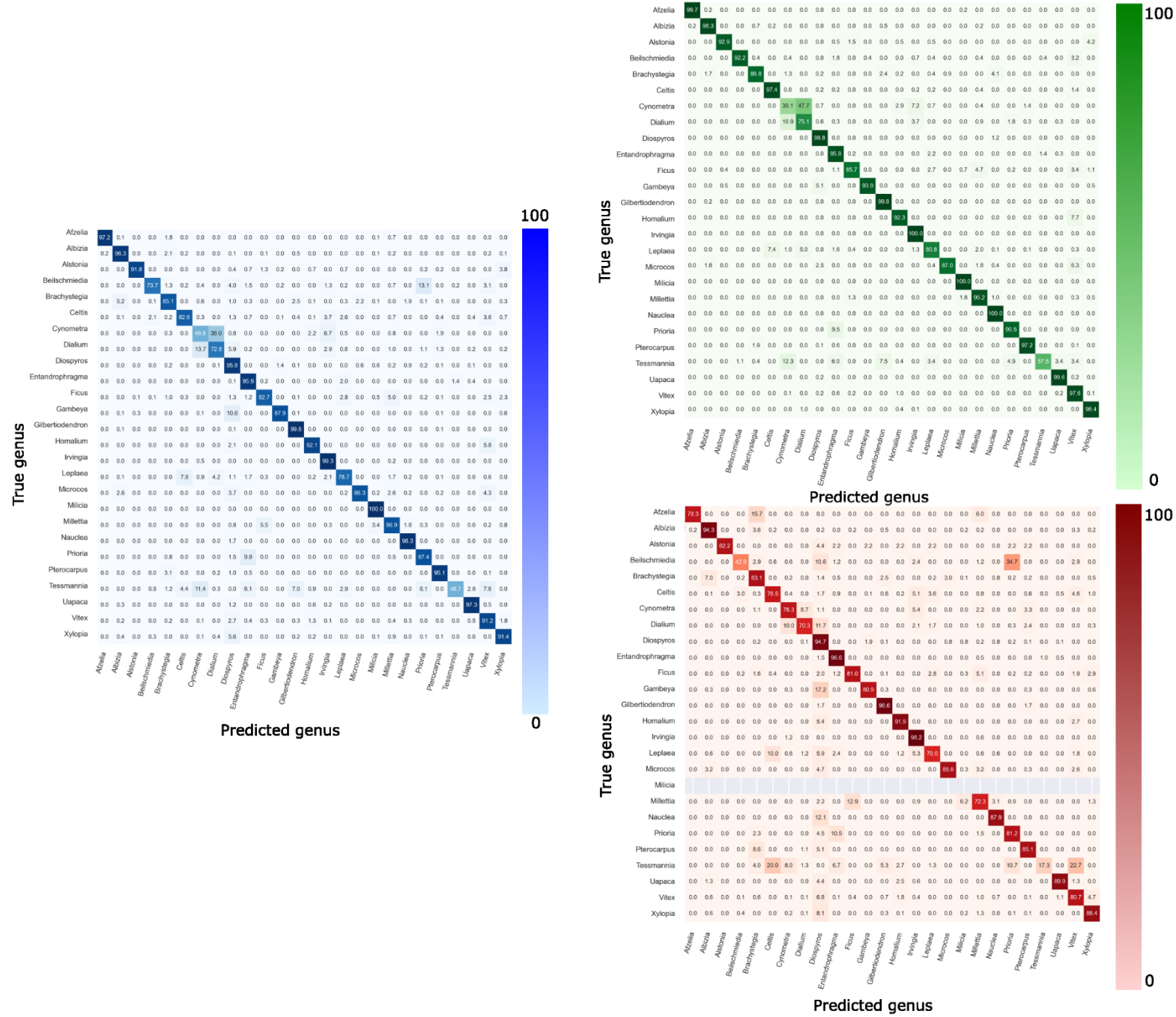


Figure 4.4: Confusion matrices showing the percentual recall scores for the 26 classes (genera) on the test patches, evaluated by the **model trained on only anomaly-free images**. For each confusion matrix, the actual genus label is presented as rows, and the predicted label is presented as columns. The cells of the confusion matrix are coloured by a linear colour gradient shown on the right of every confusion matrix. The three matrices each represent the result on a different part of the test data. The blue confusion matrix (left) represents the results on all test patches. The green confusion matrix (upper-right) represents the results only on the anomaly-free test patches. The red confusion matrix (lower-right) represents the results only on the anomalous test patches.

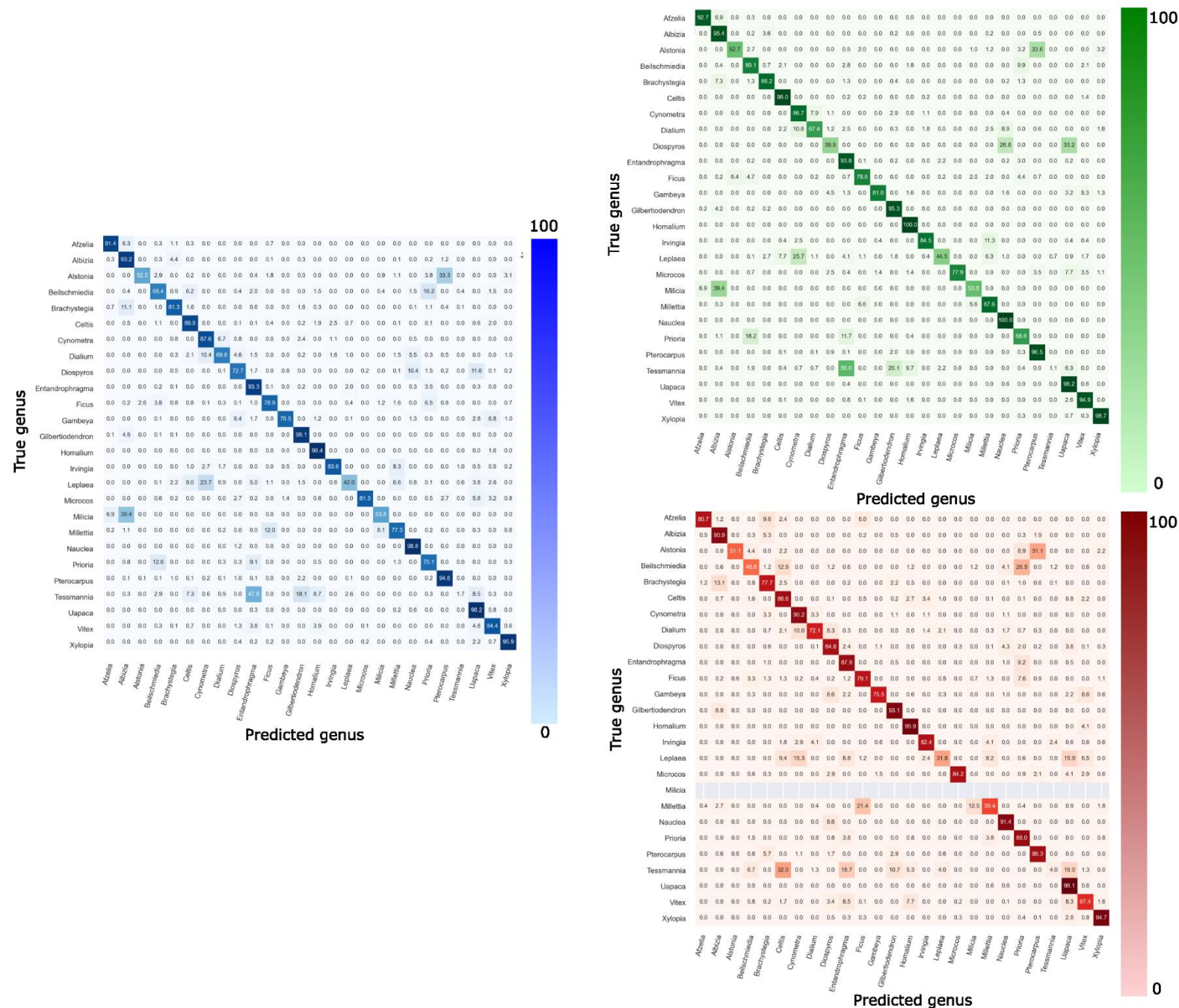


Figure 4.5: Confusion matrices showing the percentual recall scores for the 26 classes (genera) on the test patches, evaluated by the **model trained only on anomalous images**. For each confusion matrix, the actual genus label is presented as rows, and the predicted label is presented as columns. The cells of the confusion matrix are coloured by a linear colour gradient shown on the right of every confusion matrix. The three matrices each represent the result on a different part of the test data. The blue confusion matrix (left) represents the results on all test patches. The green confusion matrix (upper-right) represents the results only on the anomaly-free test patches. The red confusion matrix (lower-right) represents the results only on the anomalous test patches.

The confusion matrices illustrates the class (genus)-specific performance of wood image classification models in the three training configurations using confusion matrixes with recall scores: Figure 4.3: model trained on balanced dataset of anomaly-free and anomalous images; Figure 4.4: model trained only on anomaly-free images; and Figure 4.5: model trained only on anomalous images. The confusion matrices on all test patches (blue (left) matrices) revealed that the recall scores on most genera were well over 70% recall in Figure 4.3, except for *Cynometra* (61.7%) and *Tessmannia* (60.2%). Transitioning to Figure 4.4, the recall were higher for *Brachystegia* (12.5 percentages points), *Alstonia* (13.9 percentages points), *Millettia* (14.6 percentages points), *Leplaea* (19.8 percentages points), and *Microcos* (26.1 percentages points), while recall was observed lower for *Cynometra* (12.9 percentages points) and *Beilschmiedia* (19.0 percentages points). For the model only trained on anomalous patches (Figure 4.5), a decrease in recall was observed for *Ficus* (12.0 percentages points), *Dialium* (12.0 percentages points), *Irvingia* (13.2 percentages points), *Leplaea* (16.9 percentages points), *Gambeya* (18.3 percentages points), *Prioria* (22.4 percentages points), *Diospyros* (23.4 percentages points), *Beilschmiedia* (24.3 percentages points), *Alstonia* (25.4 percentages points), and *Tessmannia* (51.1 percentages points). However, recall improved for *Cynometra* (25.9 percentages points) and *Microcos* (21.1 percentages points).

For anomaly-free test patches (green (upper-right) matrices), the recall change between Figure 4.3 and Figure 4.4 was limited for most genera (<10%), except for improvements in *Alstonia* (14.7 percentages points), *Leplaea* (21.6 percentages points), and *Microcos* (38.6 percentages points). Limiting the dataset to only anomalous images in the training data (from Figure 4.3 to Figure 4.5) resulted in recall decreases for anomaly-free test patches (column 2) in *Diospyros* (57.8 percentages points), *Tessmannia* (57.1 percentages points), *Prioria* (29.9 percentages points), *Alstonia* (25.5 percentages points), *Gambeya* (17.1 percentages points), *Beilschmiedia* (17.1 percentages points), *Leplaea* (14.7 percentages points), *Irvingia* (14.7 percentages points), *Ficus* (13.2 percentages points), and *Dialium* (12.0 percentages points).

Decrease in recall on anomalous test patches (red (lower right) matrices) was observed for *Beilschmiedia* (from 85.3% Figure 4.3 to 42.9% on Figure 4.4), *Tessmannia* (16.0 percentages points), *Prioria* (15.3 percentages points), *Gambeya* (15.0 percentages points), and *Dialium* (13.8 percentages points). Conversely, improvements were observed for *Leplaea* (12.4 percentages points), *Brachystegia* (13.9 percentages points), *Microcos* (15.5 percentages points), and *Millettia* (23.7 percentages points). Transitioning from Figure 4.3 to Figure 4.5, recall on anomalous patches (red (lower-right) matrices) decreased further for *Beilschmiedia* (36.5%), *Tessmannia* (29.3%), *Leplaea* (25.8%), *Alstonia* (24.5%), *Gambeya* (19.7%), *Vitex* (17.6%), *Dialium* (12.0%), *Ficus* (11.3%), *Diospyros* (11.2%), and *Irvingia* (11.1%), while an increase (from 70.1% to 84.2%) was observed for *Microcos*. *Milicia* did not feature in any analysis of anomalous test data due to the absence of anomalous specimens.

The class-specific performance also differed within the three models between anomaly-free and anomalous patches. The effects are shown on the violin plots in Figure 4.6 and described underneath.

Each subplot represents a different trained model: (left/blue) model trained on both anomaly-free and anomalous patches, (middle/green) model trained only on anomaly-free patches, and (right/red) model trained only on anomalous patches. The y-axis depicts the difference in recall percentage between the two

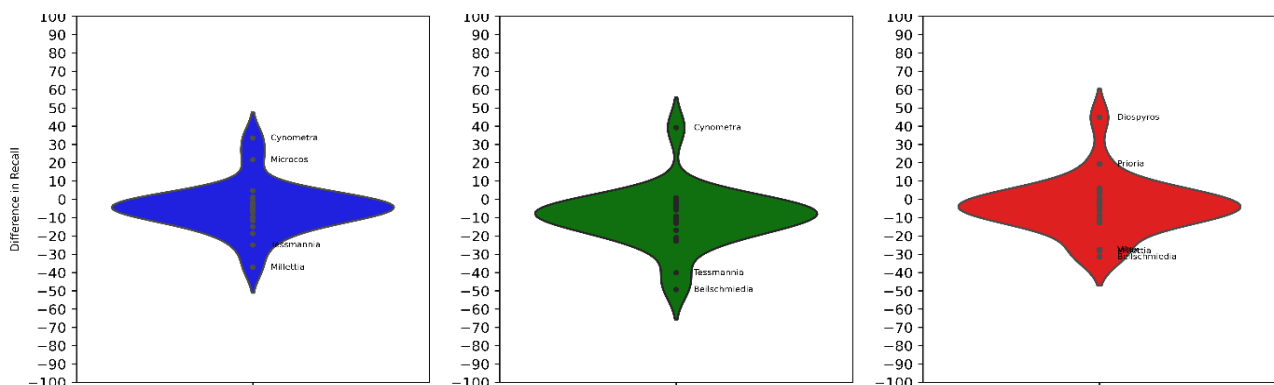


Figure 4.6: Violin plots showing recall variation between anomaly-free and anomalous patches across the three training configurations (left/blue) model trained on both anomaly-free and anomalous patches, (middle/green) model trained only on anomaly-free patches, and (right/red) model trained only on anomalous patches.

test image types, providing insights into how different training data impact the model's performance on classifying wood images in varying conditions.

For the (blue) model trained on both anomaly-free and anomalous patches, recall scores for anomaly-free test patches were generally higher than for anomalous patches in genera such as *Millettia* (85.8% vs. 48.7%), *Tessmannia* (58.2% vs. 33.3%), *Celtis* (95.2% vs. 76.6%), *Afzelia* (93.2% vs. 78.3%), *Beilschmiedia* (97.2% vs. 85.3%), and *Vitex* (96.6% vs. 85.0%). Conversely, higher recall scores were observed on anomalous patches for *Microcos* (48.4% vs. 70.1%) and *Cynometra* (53.4% vs. 87.0%).

Similar observations were made for the (green) model trained only on anomaly-free patches, with *Millettia* (95.2% vs. 72.3%), *Tessmannia* (57.5% vs. 17.3%), *Celtis* (97.4% vs. 76.5%), *Afzelia* (99.7% vs. 78.3%), *Beilschmiedia* (92.2% vs. 42.9%), *Vitex* (97.6% vs. 80.7%), *Gambeya* (93.9% vs. 80.9%), *Pterocarpus* (97.2% vs. 85.1%), *Nauclea* (100% vs. 87.9%), *Leplaea* (80.8% vs. 70.0%), and *Alstonia* (92.9% vs. 82.2%) all demonstrating improved identification on anomaly-free images. Once more, the reverse effect was observed for the recall value on *Cynometra* (39.1% vs. 78.3%), while *Microcos* showed smaller difference (87.0% vs. 85.6%).

For the (red) model trained only on anomalous patches, higher recall was observed on anomaly-free patches for *Beilschmiedia* (80.1% vs. 48.8%), *Millettia* (87.6% vs. 59.4%), *Afzelia* (92.7% vs. 80.7%), *Vitex* (94.9% vs. 67.4%), *Celtis* (98.0% vs. 86.6%), *Leplaea* (44.5% vs. 31.8%), *Pterocarpus* (96.5% vs. 86.3%), and *Brachystegia* (88.2% vs. 77.7%). There was no recall difference observed for *Microcos*, *Cynometra*, or *Tessmannia*, while patches of *Diospyros* (92.9% vs. 82.2%) and *Prioria* (92.9% vs. 82.2%) were more often predicted correctly when displaying anomalies.

4.4.2.2 Class activation maps

Figure 4.7 shows the Grad-CAMs derived from the neural network that classifies the wood at the genus level (for the model trained only on anomaly-free patches). The colour scale ranging from red to blue represents the relative intensity or importance of each pixel in the heatmap, relative to others within the same image. As such, although the same colormap is used, the specific colour intensities applied are unique to each image and cannot be directly compared across different images. These Grad-CAMs distinctly reveal that

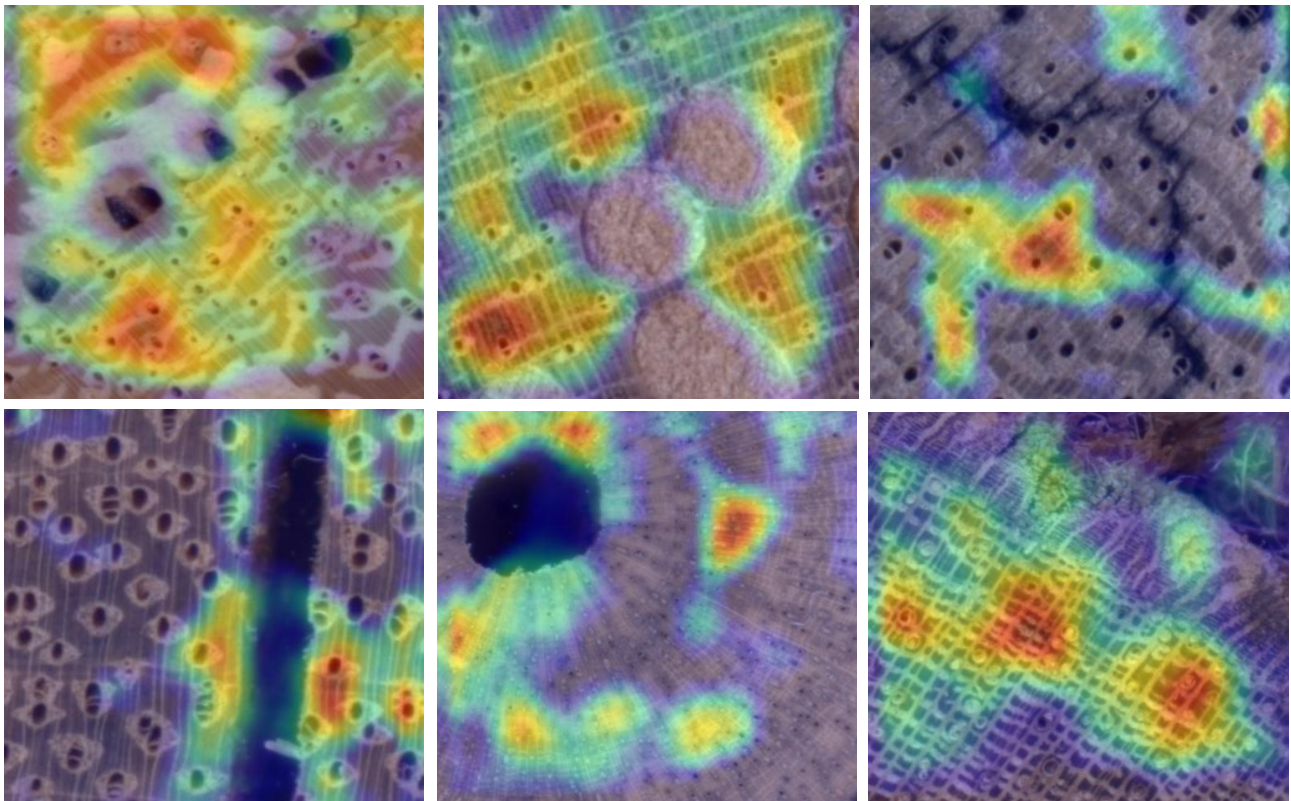


Figure 4.7: Grad-CAM heatmaps on image patches (originally shown in Figure) highlighting natural timber damage: a-c(upper/left-to-right) traumatic canals, insect damage, fungal discoloration, d-f(bottom/left-to-right) cracks, pith, bark.

areas exhibiting anomalies typically demonstrate comparatively lower activation levels. However, as illustrated in Figure 4.7, even anomaly-free wood anatomy regions do not consistently exhibit high activation levels. This behaviour can be attributed to the CNN's selective feature extraction process. The network focuses on specific image areas that are most indicative of the genus-level classification, rather than uniformly activating all clear anatomical structures. This selective activation ensures that the CNN prioritizes the most relevant and informative features for accurate classification.

4.5 Discussion

The Xception architecture has demonstrated good performance on classifying the labelled ImageNet database, (Chollet, 2017). This study suggests that the architecture also successfully extracts features from the wood anatomical image patches to distinguishing between anomaly-free patches and anomalous patches (featuring multiple types of anomalies), and identifying specific timber genera.

The binary anomaly verification model (section 3.1) can effectively discern between anomalous and anomaly-free patches and effectively localize regions on correctly classified test patches with anomalies. False negatives (Figure 4.2) demonstrate difficulty classification, which could be caused by particular wood anatomical patterns e.g. heartwood to sapwood transition possibly mistaken as fungal staining. Regarding the false positives, we could not provide definitive explanations. One possible contributing factor may be ambiguity in the labelling protocol. The binary classification of patches as either anomaly-free or anomalous was based on the presence or absence of any visually identifiable anomaly. However, the threshold for what constitutes sufficient anomaly to warrant labelling was not absolute and may have varied across annotations. For example, the presence of a minor crack—whether centrally located or at the edge of the patch—may or may not have warranted a label of "anomalous," depending on the annotator's subjective interpretation. Consequently, the model may have learned to detect such features but failed to align perfectly with the annotation heuristic, particularly when those features were subtle or limited in spatial extent. These findings underscore both the sensitivity of the model to complex visual patterns associated with anomalies, and the inherent challenges in curating high-quality labelled datasets for supervised learning in visual classification tasks. It is also important to note that the heatmaps on Figure 4.2 show that activation is more pronounced on anomalies, though not all pixels that make up an anomaly are perfectly segmented. This is a fundamental aspect also described in other previous literature in adjacent CV based classification domains (Jia and Shen, 2017; Liao et al., 2023; Yang et al., 2019). CNNs primarily activate regions it finds useful for classification, as demonstrated by class activation maps. These maps highlight specific areas contributing to predictions, indicating that not all image areas are utilized, but rather those deemed significant for the task at hand (Yang et al., 2019).

The model architecture could also successfully extract patterns to classify the 26 Congolese timbers onto genus level, as apparent by the obtained recall scores on the training, validation, and test data. Filtering the training data on the presence of anomalies, has a nuanced effect on the recall. The recall scores across classes, increasing marginally from 88.4% (trained on both anomaly-free and anomalous patches) to 90.5% (trained only on anomaly-free patches) and decreasing from 88.4% (trained on both anomaly-free and anomalous patches) to 79.1% (trained only on anomalous patches), on all test patches. This suggests a balanced trade-off in performance when removing anomalous patches, albeit with a high recall score for the model with lowest recall (e.g. trained only on anomalous patches: 79.1%).

The overall recall score on specimens is further testimony that anomalies neither significantly improves nor decreases the model's ability to classify them correctly. Additional findings on individual genera, however, suggest that while the overall performance change between training models was limited, individual genera exhibited varied responses to the presence of anomalies. Genera like *Leplaea*, *Alstonia*, *Brachystegia*, and *Millettia* demonstrated higher recall on the test data if the model was trained on anomaly-free images, whereas higher recall was observed for *Cynometra* when evaluated with models trained on anomalous patches. Other genera, including *Diospyros*, *Tessmannia*, and *Ficus*, were negatively affected by anomalies in training and did not improve when anomalous images were filtered out, indicating a preference for a balanced dataset. The peculiar case of *Microcos*, which showed higher recall regardless of anomalies for the model trained only on anomalous patches, suggests the need for further investigation. The performance of certain genera remained unaffected by altering the training data based on damage. The performance of *Azelia*, *Albizia*, *Entandrophragma*, *Pterocarpus*, *Xylopia*, *Uapaca*, *Gilbertiodendron*, *Nauclea*, *Vitex*, *Milicia* were generally unaffected by altering the training data based on anomalies. Possible correlations between these findings and genus-specific statistics of the training data (See Supplementary materials Table 8.5) were difficult to pinpoint as the applied sampling approach in this study, implies that the specimen balance across models was constant and could not have contributed to differences in genus-specific performances. Our findings on this are added in Supplementary materials 8.4 for further information. Our analysis, therefore, suggests that CNNs, like Xception, demonstrate higher recall in timber classification, when trained on anomaly-free images.

Overall, anomaly-free patches consistently yielded higher recall scores across all model evaluations than anomalous patches. Additionally, identical patterns can be observed concerning performance for the individual genera. In general, *Millettia*, *Tessmannia*, *Celtis*, *Azelaia*, *Beilschmiedia*, *Vitex* and other genera are more correctly identified (10-40% better using anomaly-free patches than anomalous patches). An exception to this is the genus *Cynometra*, for which the recall score was between 20-40% higher on anomalous test images in models trained with anomaly-free training images present (either only anomaly-free or both anomaly-free and anomalous), though not for the model trained only on anomalous images. Our analysis, therefore, suggests that CNNs, like Xception, demonstrate the overall highest proficiency in general timber classification, when testing anomaly-free image patches.

The Grad-CAM analysis adds further insight, revealing that the model prioritizes regions on patches without anomalies. This provides elucidation on the higher observed recall for models trained on anomaly-free patches. If the model predominantly focuses on anomaly-free regions for its classification, it follows logically that cleaner training images enhance its ability to capture diagnostic patterns, leading to improved discrimination between timbers. This also underscores the effectiveness of the model's convolutions in extracting diagnostically valuable patterns and applying it on anomalous wood specimen in the field. In addition, these findings affirm the reliability of employing Grad-CAMs as a way to provide valuable insights into the model's interpretability (Wang and Zhang, 2023) and its approach to handling anomalies on endgrain timber images.

Still, other research also emphasizes the importance of data augmentation to improve model robustness. Owens et al. (2024) conducted a review that explored the impact of various digital perturbations on test images to assess the robustness of a macroscopic CV wood identification model, to classify 24 Peruvian timbers. To compare for instance, cracks can be likened to the introduction of scratches in the study. On the other hand, fungi, and/or bacterial damage, leading to wood deterioration or discoloration, pose a slightly greater challenge for direct comparison with the studied perturbations. The combination of colour shifts in the blue channel (mimicking the effects of blue staining fungi), digital scratches (similar to hyphae on wood), and blurring (impairing the visibility of diagnostic features) can approximate these effects. In their study, the 24-class model performance showed sensitivity to blue channel shifts and blurring with the percentage of correctly classified images dropping along with the magnitude of the colour shift/blur. Digital scratches seemed to have only mild impact on model performance. Other anomalies like large cracks or insect holes are more difficult to link to but can be compared to other frequently used augmentation techniques such as random erasing, where random patches of pixels on images are replaced by either set or random values—an approach inspired by dropout regularization (Zhong et al., 2020). Similar parallels exist for images taken at the sides of specimens due to human error. The study by Owens et al. (2024) concludes by suggesting that the introduction of such digital perturbations on the training data could improve model robustness. Another study by Shorten and Khoshgoftaar (2019) conducted a review that explored various data augmentation techniques on CNN in general and their impacts on model performance. This study also suggests that applying such techniques or a combination can improve generalization on training images. This could have influenced the small differences between the recall scores of the different models, as the applied augmentation techniques during training (e.g. colour scaling and gaussian blurring (for more information see section 2.2)) could have improved the robustness of the models to better handle anomalous patches. However, Shorten and Khoshgoftaar (2019) also emphasizes the need for careful consideration in the design of data augmentation to align with the classification and not remove crucial diagnostic features. This is especially relevant for shallow datasets where excessive removal might lead to overfitting. This insight elucidates the observed tendency toward overfitting and the suboptimal evaluation of test specimens in models trained on image datasets lacking anomaly-free images in the training data.

Closer examination of Grad-CAMs revealed no discernible correlation between the type of wood anomalies and misclassification, although severely deteriorated images with little to no clear wood anatomy were prone to misidentification due to the absence of diagnostic features. This finding aligns with findings of Owens et al. (2024) that their model became progressively less robust as the area of medium-to-severe blur increases, suggesting that a CNN for macroscopic cross sectional wood identification requires intact regions of wood anatomy for correct classification. A peculiar pattern was also observed where severely deteriorated images were consistently misclassified as *Diospyros*, indicating that CNN classifiers might uniformly categorize such images without additional training classes. This suggests the need for including an unknown class to handle such anomalies. Supporting these findings, Ravindran et

al. (2023) demonstrated that the occlusion of wood anatomy negatively affects CNN performance, with multiple trained CNN architectures performing poorly on specimens sanded at low grits (80, 180) due to severe occlusion. Performance stabilized at higher grits (240 and onwards), underscoring the need for clear anatomical features for accurate classification, a requirement shared by both human observers and CNNs.

4.6 Conclusion

The results show that CNN, such as the applied Xception architecture, can successfully extract features for classifying image patches of sanded cross-sectional images to classify different timbers at the genus level, and distinguishing between anomaly-free and anomalous wood. The performance on the test data varied for individual genera, with some benefiting from training on anomaly-free images, while for other genera, like *Cynometra*, higher recall was observed for the model trained on anomalous images. However, the overall recall scores were higher for anomaly-free patches. Grad-CAM analysis revealed the model's preference for regions on patches showing unobscured wood anatomical tissue, underscoring the importance of clear wood anatomy in training CNNs for wood identification. This could enable CNNs to capture diagnostic patterns more effectively, which in turn would lead to better discrimination between timbers, even when applied to anomalous specimens in the field. The inclusion of anomalous patches had a limited impact, but subtly enhanced performance on anomalous patches. The findings therefore suggest that CNNs (like Xception) demonstrate the highest proficiency in timber classification when trained on anomaly-free images, making this approach highly effective for developing CV-based wood identification models for deployment in the field.

4.7 Acknowledgments

This work was realized through the combined efforts of the staff at the service of Wood biology at the RMCA in Tervuren, Belgium, and the staff at the UGent-Woodlab (Department of Environment, Faculty of Bioscience Engineering) of Ghent University. The collection and specimens were managed by Maaïke De Ridder (RMCA) and Daniel Wallenus (RMCA).

4.8 Availability of code and metadata

Scripts and metadata can be consulted on <https://github.com/rdblauer/Scripts-metadata-Evaluating-the-effect-of-anomalous-images-on-CV-based-wood-identification-models.git> :

Contact: ruben.de.blaere@africamuseum.be

Chapter 5: Expert-defined vs. deep learning approaches for scalable and generalizable timber identification

Authors: Ruben De Blaere^a, Wannes Hubau^{a,b}, Hans Beeckman^a, Jan Van den Bulcke^b, Jan Verwaeren^c

a Service of Wood Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium, b UGent-Woodlab, Laboratory of Wood Technology, Department of Environment Ghent University, Coupure Links 653, 9000 Gent, Belgium, c UGent-KERMIT, research unit Knowledge-based, predictive and spatio-temporal modelling, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Gent, Belgium

5.1 Abstract

CNNs show strong potential for automated timber identification, yet most prior studies have relied solely on closed-world multiclass classification, which assumes that all species encountered during deployment are present in the training data. This assumption limits applicability in real-world contexts, where previously unseen species are common—a significant challenge given the global diversity of traded timbers.

This chapter compares closed- and open-world genus-level identification methods using macroscopic cross-sectional wood anatomy. Four approaches were evaluated: (i) closed-world multiclass RF trained on expert-defined anatomical features (Chapter 3); (ii) closed-world multiclass CNN with CCE loss trained on 5.42 × 5.42 mm image patches (Chapter 4); (iii) open-world embedding-based CNN with triplet loss; and (iv) open-world binary verification CNN using composite images and BCE loss. All models were trained and tested on the SmartWoodID dataset, with performance assessed across all Congolese species and a subset of commercially important Congolese timbers.

For commercial Congolese timbers, the multiclass CNN achieved the highest metrics (accuracy = 0.86, precision = 0.87, recall = 0.85, F1 = 0.84), followed by binary verification (accuracy = 0.75, precision = 0.73, recall = 0.71, F1 = 0.69) and RF (accuracy = 0.58, precision = 0.56, recall = 0.54, F1 = 0.53). The embedding-based CNN performed substantially worse (accuracy = 0.28, precision = 0.21, recall = 0.24, F1 = 0.21), likely due to the sensitivity of the triplet loss function.

Top-*k* analysis confirmed the superior ranking ability of CNN-based models: for all Congolese species, 95% correct identification was reached by considering the top 16% of genera for the multiclass CNN and top 7% for commercial species, compared to 72% and 64%, respectively, for RF. Binary verification required slightly larger *k* values than the multiclass CNN but outperformed RF. In contrast, the embedding-based CNN required *k* values of 84% (all species) and 71% (commercial species) to reach the same accuracy threshold.

Open-world performance was further tested on an independent dataset of non-Congolese timbers. The binary verification approach successfully identified unseen genera, achieving 95% correct identification when considering the top 30% of genera. These findings demonstrate that CNN-based methods—particularly multiclass classification and binary verification—substantially outperform feature-based RF models. Moreover, binary verification offers flexibility for identifying previously unseen timbers, supporting forensic and enforcement applications by enabling rapid screening and flagging of potentially suspicious shipments.

Embedding-based methods offer a powerful framework for open-world recognition with strong advantages regarding computational efficiency. However, observed performance in this chapter was poor, suggesting the need for more robust training strategies, improved loss formulations, and potentially the integration of complementary data sources (e.g., microscopic anatomy, chemical signatures) to reach their full potential for scalable, generalizable timber identification.

This chapter is not submitted to a peer-reviewed journal on 19/08/2025.

5.2 Introduction

Illegal logging remains a significant driver of deforestation, biodiversity loss, and environmental degradation, requiring the development of rapid and reliable wood identification techniques to enforce regulations and international conventions (Dormontt et al., 2015; Tacconi, 2012; Thompson and Magrath, 2021). Recent advancements in wood forensics have leveraged deep learning techniques to classify timber species, with the analysis of macroscopic cross-sectional images of wood anatomy emerging as a popular approach (Silva et al., 2022; Wheeler and Baas, 1998). Cross-sectional images are particularly advantageous due to their distinct structural patterns and accessibility, enabling non-experts to locate the appropriate cross-section for evaluation with relative ease (Chen et al., 2022; Hwang and Sugiyama, 2021; Wu et al., 2021).

Traditional deep learning approaches, such as multiclass classification, have been widely applied to wood classification tasks (Silva et al., 2022). However, a multiclass approach faces inherent limitations when applied to biological data (McCarthy and Hayes, 1981; Yoshihashi et al., 2019). First, they operate under a closed-set assumption, meaning the model is trained only on a predefined set of classes and struggles with unknown species encountered during inference (Sünderhauf et al., 2018; Wilber et al., 2013). While opt-out mechanisms or "other" classes can partially mitigate this limitation, they remain imperfect solutions (Entezari and Saukh, 2020; Geifman and El-Yaniv, 2019). Second, these approaches generally rely on balanced class distributions, with minor imbalances managed using class weight dictionaries. This requirement places a burden on projects aiming to classify a broad range of timbers, as it demands extensive and rigorous data collection. Reliable specimens often originate from fieldwork guided by experienced botanists or from curated wood collections. While wood collections are advantageous due to their pre-processed samples with clear anatomical planes and reliable metadata, they typically offer limited replicates per species. This scarcity makes it challenging to ensure sufficient variability for robust model generalization (Ravindran et al., 2018; Silva et al., 2022).

To address the limitations of multiclass classification, deep learning research has increasingly focused on open-world recognition, with object re-identification emerging as a powerful open-world alternative for image classification (Geng et al., 2020; Scheirer et al., 2012; Yoshihashi et al., 2019). In contrast to multiclass models that produce discrete class labels, object re-identification networks are trained to compute image similarities, enabling the identification of the most corresponding reference image(s) (Ye et al., 2021). Object re-identification networks have achieved significant success in fields like facial recognition, relying on prominent methods such as binary verification and triplet learning (Schroff et al., 2015). Binary verification determines whether two given samples belong to the same class by learning an optimal decision threshold for similarity scores (Chen et al., 2017). In contrast, triplet learning maps images into an embedding space. In this space, each image is represented by an embedding vector. These vectors encode distinctive features, with smaller distances between vectors indicating greater similarity. This essentially creates a "digital fingerprint" for each class, enabling comparisons against a reference database of embeddings. A smaller distance between the embeddings of a sample and the embeddings of a reference sample indicates a stronger match (Ghosh et al., 2023; Ye et al., 2021). While these techniques enhance generalization, they also introduce new challenges, such as the need for hard example mining to maximize learning efficiency (Bai et al., 2018; Chen et al., 2017; Hermans et al., 2017). Additionally, balancing and generating appropriate sample pairs are crucial for effective training, requiring careful attention to loss function design and data preparation (Hermans et al., 2017). At present, applying object re-identification for wood identification is underexplored in literature (Hwang and Sugiyama, 2021; Silva et al., 2022).

This study provides a comprehensive evaluation of CNN-based classification strategies for timber identification, comparing traditional multiclass classification with object re-identification approaches, providing insights into the trade-offs between closed-world and open-world approaches for timber identification, and ultimately guiding the development of more robust forensic wood identification methods for field deployment. A high-performing CNN architecture was trained to classify wood at the genus level using the SmartWoodID dataset (De Blaere et al., 2023). Three distinct CNN-based

classification approaches were developed and evaluated: 1) traditional multiclass classification, where a CNN was trained as a conventional closed-world multiclass classifier; 2) object re-identification, where an embedding-CNN was trained, which subsequently allows to differentiate between specimens based on learned feature embeddings; 3) object re-identification, with binary verification using composite twin images, where each input pair was labelled as either belonging to the same genus or different genera.

To systematically assess and compare the performance of these models, we conducted a series of experiments using the same scenarios as Chapter 3. The first scenario features 601 Congolese species from SmartWoodID to validate the use of macroscopic anatomical features for classifying taxa within the Congo Basin's biodiverse ecosystem, thereby in supporting biodiversity research and conservation efforts. The second scenario features a subset of 78 Congolese commercial species to evaluate the models' capacity to classify high-demand timber species commonly traded in international markets, supporting legal compliance and sustainable timber trade practices.

To test the hypothesis that deep learning models trained on raw wood images outperform traditional classifiers based on expert-defined anatomical features, we conducted a series of comparative experiments using the SmartWoodID dataset. These experiments build on the scenarios presented in Chapter 3, enabling direct comparison of model performance across consistent taxonomic and ecological contexts. Next, we compared the traditional multiclass CNN with object re-identification approaches—triplet learning and binary verification—to test the hypothesis that open-world frameworks offer greater classification flexibility and better accommodate intra-species variability. Model evaluation was extended using top-k accuracy metrics, quantifying the likelihood that the correct class appears among the model's top predictions. Finally, to test generalization capacity beyond the training domain, object re-identification models were applied to a separate dataset of non-Congolese timber species digitized using the same protocol. This experiment examines the hypothesis that open-world strategies provide scalable, region-independent solutions to wood identification, even under increasing classification complexity and the presence of unseen taxa.

5.3 Material and methods

5.3.2 Dataset Description

This chapter is based on the SmartWoodID database, serving as a valuable resource for examining the relationship between macroscopic cross-sectional wood anatomy and the botanical diversity of Congolese tree taxa (De Blaere et al., 2023). The following information in this section, re-uses the text in section 1.5 which provides an overview of all data used in each chapter.

The database contains high-resolution RGB scans of the macroscopic end-grain surfaces of 3,742 wood specimens, representing 954 species native to the DRC. Each specimen was prepared by scanning the cross-section at 2400 dpi using a flatbed scanner. This resolution allows for the visualization of macroscopic features essential for wood identification.

Species and lower taxa are represented by multiple specimens, capturing both intra- and interspecific anatomical variation. This makes the database well-suited for studying wood identification using macroscopic anatomy. A complete overview of the database is provided in Chapter 2 and in De Blaere et al. (2023) (De Blaere et al., 2023), while Supplementary Materials Table 8.1 all unique specimen identifiers and metadata. To enable machine learning analysis, we selected only species represented by at least two specimens. This set comprises 2,296 digitized specimens across 601 species, 286 genera, and 64 families. Discriminatory power was mainly assessed by training and evaluating classification models on the specimens. Therefore, specimens were allocated random to training (75%) or test set (25%), while preserving distribution of species across both sets. Within both sets, a subset of 78 commercially important species was defined for targeted evaluation. An overview of the chapters, designated datasets used in each chapter and of which hypothesis they target is provided in section 1.4.

Given the large size of the scanned end-grain surfaces, directly processing full images is computationally expensive and may lead to overfitting due to the limited number of images. To address this, the full images were cropped into smaller, non-overlapping patches that retained diagnostic features while capturing anatomical variability. We selected patch dimensions of 512×512 pixels, corresponding to a physical area of approximately 5.42×5.42 mm, ensuring sufficient detail for anatomical analysis without obscuring critical features. Full images yielding fewer than four patches (i.e., total end-grain surface area below 1 cm²) were excluded to maintain consistency and relevance for anatomical assessment.

Patches were allocated random to training (75%) or validation set (25%), while preserving distribution of specimens across both sets, using different regions of the same specimens for training and validating. Furthermore, any full image yielding fewer than four usable patches was removed, as these would cover less than one square centimetre—below the typical area assessed by a human expert during wood anatomical identification.

5.3.3 CNN Backbone Architecture

The Xception architecture was chosen as the CNN backbone due to its balance between computational efficiency and strong classification performance on large-scale datasets such as ImageNet (Chollet, 2017). The model was pretrained on ImageNet and fine-tuned on the SmartWoodID patches using a batch size of 8 and RMSprop as the optimizer with a learning rate of 1×10^{-4} . Early stopping was implemented to prevent overfitting, with validation loss monitored during training. Training was capped at 100 epochs, with early stopping triggered after 10 consecutive epochs of no improvement in validation loss (<0.1).

5.3.4 Classification Techniques

Four different approaches were used to develop CNNs for wood genus identification, ranging from multiclass and multilabel classification to representation learning via embedding networks and binary classification. The output layers and loss functions were customized to suit the unique requirements of each classification task.

5.3.4.1 Multiclass Classification

The multiclass classification model aims to assign each image patch to one of the available classes. The CNN backbone was extended with a global average pooling layer and a classification head. The classification head consisted of a fully-connected layer with a SoftMax activation to output probabilities across all genera. The model was compiled using the categorical cross-entropy loss function (L_{CCE} : see Equation 1) quantifying the difference between the predicted class probabilities and the actual class labels across all samples. Specifically, for a total of N samples and C classes (in this case, genera), the loss is calculated based on the binary indicator $y_{i,j}$, which equals 1 if sample i belongs to class j , and 0 otherwise. The predicted probability for class j for sample i is denoted as $\hat{y}_{i,j}$.

$$L_{CCE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \log(\hat{y}_{i,j})$$

Equation 1: CCE Loss function

5.3.4.2 Object Re-Identification Network with Triplet Learning

To explore open-world recognition, we implemented an object re-identification network trained with triplet loss. This approach maps input image patches into a lower-dimensional embedding space where embeddings from the same genus are closer together, and those from different genera are further apart. This enables comparison of the test images to reference images based on the learned feature embeddings. A triplet loss function was employed to train the model to minimize the distance between embeddings of images belonging to the same genus (anchor-positive pairs) while maximizing the distance from images belonging to different genera (anchor-negative pairs). Figure 5.1 shows how triplet loss operates on the embedded versions of these image patches. Figure 5.2 shows an example of a triplet of image patches. The Xception backbone was trained to map the images to into fixed-length feature embeddings. The model outputs the concatenated embedding vectors of the anchor, positive, and negative image patches for triplet loss computation. The triplet loss function ($L_{triplet}$: see Equation 2) calculated the distance between anchor-positive pairs and anchor-negative pairs, optimizing the network to maximize inter-class separation while minimizing intra-class variation. For each triplet of samples—consisting of an anchor (x^a_i), a positive sample from the same class (x^p_i), and a negative sample from a different class (x^n_i)—the loss is computed based on the squared Euclidean distances ($\|\cdot\|_2^2$) between the embedded representations $f(x)$ of these samples. Formally, the loss ensures that the distance between the anchor and the positive embedding is smaller than the distance between the anchor and the negative embedding by at least a predefined margin α (e.g., $\alpha = 0.2$). This is achieved through the hinge function ($[.]$), which returns the value if positive, or zero otherwise, thus preventing negative loss values. The total number of triplets is denoted as N , and the loss encourages the condition:

The loss measured the margin between the positive and negative distances, penalizing configurations where the negative distance was not sufficiently larger than the positive distance. Key hyperparameters, such as the margin ($\alpha = 0.2$) and learning rate (1×10^{-5}), were fine-tuned for optimal performance.

$$L_{triplet} = \frac{1}{N} \sum_{i=1}^N \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]$$

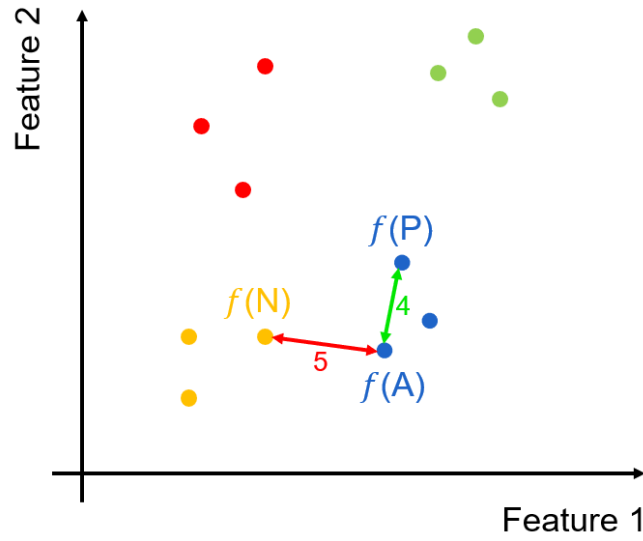


Figure 5.1: Visualisation of the concept of triplet loss optimization in an exemplary 2D feature space. The generated embeddings of the triplet model are converted into points in the feature space. The different colours (red, yellow, blue, green) represent different classes. Triplet learning aims to learn representations where the distance between anchor(A) and positive(P) is smaller than the difference between anchor(A) and negative(N).



Figure 5.2: Example of hard triplet consisting of an anchor (left: Afzelia; Tw11081), positive (centre: Afzelia; Tw47200), and negative image (right: Anthonotha; Tw95) for Object Re-Identification network training with triplet loss.

5.3.4.3 Object Re-Identification Network with Binary Verification

This approach serves as an intermediate method between an embedding network for object re-identification and a binary classification network. The added layers closely resemble those used in binary classification, where the extracted features are passed through a Global Average Pooling layer followed by a fully connected layer. The fully connected layer contains a single output neuron with a sigmoid activation function, designed to perform binary classification. The model was compiled using binary cross-entropy loss (L_{BCE} : see Equation 3) with one class, a suitable choice for binary tasks as it quantifies the error between the predicted probabilities and the true labels for each genus. For a total of N samples, each with a true label y_i (where 1 indicates the positive class and 0 the negative), and a predicted probability $\hat{y}_i$, the binary cross-entropy loss penalizes the model based on how far off each prediction is from its corresponding true label.

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

The images themselves encode the anchor-positive-negative relationship. For each anchor image, two composite images are created, each consisting of two halves: (anchor-positive) labelled as 1 and another (anchor-negative) labelled as 0 (an example is presented in Figure 5.3). This design allows the model to learn to distinguish between images from the same genus and those from different genera. Class balance is inherently maintained, as for each anchor, both composite images are always generated. This ensures the model does not inadvertently focus more on one class over the other, promoting equitable learning across both categories.

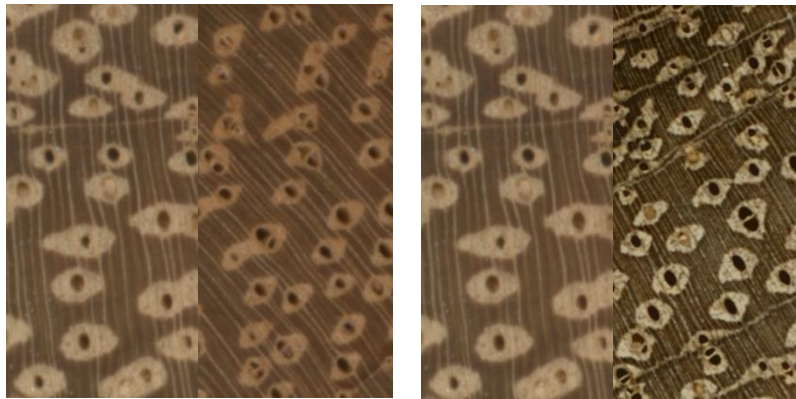


Figure 5.3: An example of two composite images representing on the left the anchor (Afzelia; Tw11081)-positive (Afzelia; Tw47200), and on the right anchor (Afzelia; Tw11081)-negative (Anthonotha; Tw95) image relationship for Object Re-Identification network training with Binary Cross Entropy.

As previously discussed in 4.5, restricting the training dataset to include only undamaged image patches is crucial for this method. This is because the patches are reduced to half their original area to form composite images. Including damaged patches could result in the removal of areas containing key diagnostic information during this reduction process.

5.3.5 Balancing the dataset

The dataset's inherent imbalance—arising from some genera containing more specimens and producing more image patches due to larger cross-sections—was addressed using a sampling strategy at the start of every epoch. To ensure fairness, each genus was limited to a predefined number of three specimens by randomly sampling unique specimen identifiers. For each specimen, a fixed number of four image patches was selected, with replacement applied for specimens with fewer available patches. By exposing the model to a diverse set of images in each epoch, this approach balanced representation across genera and allowed the model to capitalize on all data during the whole training process without biasing the model towards more prolific genera. The samples were also shuffled to prevent the memorization of image sequences, prior to being loaded into the CNN.

To ensure a robust evaluation and facilitate direct comparison with the machine learning classification techniques in Chapter 3, specimens are allocated into the same subsets (training and testing) as described in that chapter. An additional stratified split was applied on the training dataset, allocating 75% of the patches for actual training and 25% for periodic validation after each epoch. This approach guaranteed that every specimen was represented in both training and validation while maintaining distinct patches to prevent data leakage.

5.3.6 Data augmentation

Data augmentation involves artificially inflating the training dataset to prevent overfitting by warping available data (Shorten and Khoshgoftaar, 2019). This is achieved by applying transformations to the images to increase variability, improve model generalisation, and simulate the diversity of real-world conditions—such as changes in orientation, scale, lighting, or background—without the need for

additional manual data collection. All data augmentation techniques were applied uniformly across the experiments to maintain consistency in training conditions while adapting the CNN architecture to specific classification tasks. The augmentation was only applied to the training data to ensure that the validation and test datasets remained representative of the original data distribution, reflecting its ability to generalize. These techniques were applied using custom built data generators to leverage the images through the model.

To enhance the robustness and generalization of the CNN, an advanced data augmentation pipeline was implemented. The pipeline combines traditional augmentation techniques with custom filters and random erasing to introduce variability while preserving key diagnostic features of wood anatomy. Standard augmentations were applied, including random rotations up to 45°, horizontal and vertical shifts up to 20% of the image dimensions, random brightness scaling within a range of 0.5–1.0, shearing up to 30°, horizontal and vertical flipping, and normalizing the pixel intensities. To further increase variability and simulate realistic imperfections, the pipeline included custom filtering and random erasing. A randomly selected filter from the following pool was applied to each image; Gaussian blur: Random σ -values between 0.0 and 3.0; Median filtering: Random kernel sizes between 1×1 and 5×5; Minimum, maximum, and uniform filters: Kernel sizes randomly chosen between 1×1 and 5×5; Percentile filtering: Random percentiles between 1 and 99 with a 3×3 kernel; Colour channel manipulation: Random scaling of pixel intensities in red, green, or blue channels within a range of ± 20

Each filter was applied with a randomly selected mode (reflect or nearest) to further diversify the outputs. Next, random erasing was applied replacing random small rectangular regions of the image with random colour noise. The number and size of these erased regions were determined by a random sampling of the image area (0.1–1.0% per region).

This augmentation strategy not only ensured variability in training data but also introduced robustness against noise and potential artifacts in real-world applications, such as damage to wood surfaces or inconsistencies in imaging conditions. The range of each augmentation step was constrained to avoid excessive damage to diagnostic features. This was assessed by generating augmented patches and enhancing specific augmentation steps, verifying if the anatomical features were still recognizable through trial-and-error.

5.3.7 Hard mining for Object Re-Identification approaches

Training object re-identification networks with triplet learning effectively requires guiding the model towards learning from challenging examples (Hermans et al., 2017). This logic is also applicable to the binary verification approach, as binary classification problems are known to prioritize minimizing classification error through optimal threshold selection rather than optimizing ranking accuracy. Consequently, while binary verification may achieve low misclassification rates, it risks ranking errors in closely related genera, potentially leading to incorrect top-1 predictions (Chen et al., 2017). Challenging triplets or pairs, where the positive sample is visually similar to the anchor but still belongs to the same class, or where the negative sample is visually similar to the anchor but belongs to a different class, are critical for maximizing the network's ability to discern fine-grained differences (Ghosh et al., 2023; Hermans et al., 2017). These challenging examples force the network to focus on subtle and discriminative features rather than relying on easy-to-learn patterns, thereby enhancing its robustness and generalization capability.

To effectively guide the model toward learning from challenging triplets and pairs, three key aspects were addressed. First, a structured understanding of inter-class similarity was necessary. To facilitate this, domain knowledge of macroscopic cross-sectional anatomical features was integrated into the selection process for both object re-identification approaches. Integrating domain knowledge can improve the selection of informative training samples by ensuring that the model learns from biologically meaningful variations rather than relying solely on data-driven feature extraction (Guo and Lovell, 2024). Specifically, macroscopic features (Table 2.2) were numerically encoded to construct a similarity matrix for all specimens. Each feature was assigned a numerical value: present = 1, variable = 0.5, absent = 0, and NA =

0. Pairwise Euclidean distances were then computed between the average feature vectors of each specimen group, and these distances were transformed into similarity scores using a reciprocal transformation to emphasize smaller distances.

Second, the performance of object re-identification models benefits from diversity in triplet and pairwise combinations, as repeated exposure to the same combinations may lead to overfitting (Chen et al., 2023; Ying, 2019). To enhance diversity, new triplet and pair selections were generated at the start of each training epoch, preventing the model from memorizing specific relationships and improving generalization (Kumar et al., 2020).

Finally, while prioritizing the most challenging positive and negative examples, i.e., selecting within-genus specimens that are most dissimilar to the anchor and between-genus specimens that are most similar to the anchor, is crucial for fine-grained discrimination, exclusively sampling the hardest cases can hinder generalization (Schroff et al., 2015; Shi et al., 2016). Overexposure to outliers may cause the model to focus disproportionately on rare cases rather than learning broader inter-class relationships (Hermans et al., 2017). To mitigate this, the similarity scores were used as weights during sampling of positive and negative examples. This ensured that challenging pairs and triplets were more likely to be sampled while still allowing occasional exposure to easier combinations. By maintaining a balanced representation of difficult and diverse training examples, this approach enhanced the robustness of the object re-identification models to subtle anatomical variations.

5.3.8 Standardizing CNN output for model comparison

To ensure comparability across classification models, the output of all three CNN approaches is transformed to align with the output format of the RF model. This ensures a standardized probability range for each test specimen.

For the multiclass CNN, the model generates a probability distribution across all classes for each image patch. These probabilities are averaged per specimen across patches and subsequently normalized across classes to ensure consistency with the RF model. Two different models are trained for each set of timber genera: one for all Congolese tree species and one using only the subset of commercial timbers.

The evaluation of object re-identification models differs due to their reliance on a reference database for comparison. In this study, training and validation specimens serve as reference samples. To evaluate the object re-identification models' performance across the full timber dataset, encompassing all tree species, and the commercially relevant subset, non-commercial species were excluded from the reference dataset when assessing the subset. This approach is valid because object re-identification, unlike multiclass classification, does not necessitate training distinct models for varying class ranges.

For object re-identification using Triplet Learning, the CNN generates an embedding vector for each image patch. These vectors are aggregated at the specimen level by averaging. The resulting embeddings from the training and validation specimens are then used to train a separate RF model, following the methodology outlined in Chapter 3. To optimize classification performance, randomized grid search is employed, systematically sampling parameter combinations from a predefined hyperparameter grid (see Chapter 3, Table 8.2). Class weights are incorporated to mitigate class imbalance, and 4-fold cross-validation ensures robustness across different data splits. To maintain consistency and reproducibility, all random number generators were seeded with a fixed value. Consequently, the output of this RF model on the embedded vectors aligns with the output of the RF model trained on wood anatomical descriptions.

The binary verification approach requires generating a large set of composite image pairs, where each test image patch is matched against reference patches to produce a similarity score between 0 and 1, reflecting the likelihood of a genus match. Given the dataset's size—37,577 test patches, 18,136 validation patches, and 54,405 training patches—an exhaustive comparison using all training and validation patches as references would result in 2,725,873,157 image pairs. With an estimated computation time of 0.05 seconds per image for the applied workstation (single standard GPU (NVIDIA GeForce RTX 3080 Ti)), this approach would require approximately >1500 days, rendering it computationally infeasible.

To address this, a sampling strategy was implemented; Thirty percent of the patches were retained for each test specimen. Four reference patches were sampled per reference specimen. And finally, one reference specimen per genus was selected from the training/validation set. This down-sampling resulted in 1,144 reference patches and 12,525 test patches, leading to 14,329,362 comparisons and a significantly reduced computation time of approximately 9 days. After generating similarity scores, the results were aggregated in two stages : first, per reference and test specimen by averaging the output scores; second, per reference genus yielding a probability distribution (0–1) for each test specimen across all genera; and finally normalization, ensuring that the output probabilities of this model align with those of the multiclass CNN and the RF model, allowing for direct comparison.

5.3.9 Model evaluation

The performance of all classification models was evaluated using Accuracy, Precision, Recall, and F1-Score, calculated from standardized test predictions for both the full set of Congolese timbers and the commercially relevant subset (Definitions of those metrics are provided in supplementary materials section 8.2.). The experiments include model comparisons on African timber genera and the generalization of the object re-identification approach to a distinct group of different timber genera for which the model was not trained. Model comparisons were conducted in three distinct scenarios:

1. Comparison of multiclass Classification Models – The performance of the multiclass CNN classifier was compared with that of a RF model, previously trained on manually extracted wood anatomical features (see Chapter 3). This analysis assessed the relative effectiveness of deep learning-based feature extraction versus human-engineered anatomical descriptors for wood classification.
2. Comparison of CNN-Based Models – The traditional multiclass CNN was evaluated against both object re-identification models to examine the differences between direct class prediction and similarity-based classification.
3. Top-*k* Accuracy Analysis – To better understand the reliability of each model, we extended the evaluation beyond top-1 accuracy by assessing the classification accuracy of all four models (RF/multiclass-CNN/object re-identification-CNN (Triplet learning)/object re-identification-CNN (binary verification)), when considering the top-*k* predictions. This analysis offers deeper insight into the reliability of model predictions, particularly in highly biodiverse settings where multiple timber species may exhibit similar macroscopic characteristics.

To systematically compare classification behaviour across techniques, four complementary analyses were conducted on the first two scenarios:

1. Performance Metrics Summary – Accuracy, Precision, Recall, and F1-Score were computed across classes. Macro-averaging was applied to Precision, Recall, and F1-Score to ensure that each class contributed equally to the overall evaluation, mitigating the impact of class imbalances.
2. Matrix Comparison – The relationship between model predictions was assessed by constructing matrices. These matrices captured the number of specimens correctly classified by both models, only one, or neither, providing insight into classification overlap and discrepancies.
3. Genus-Level Performance Analysis – To assess classification performance across taxa, F1-scores were calculated per genus for each model. These values were plotted in scatter plots, with one model's performance on the x-axis and the other on the y-axis. The F1-score was selected as the primary evaluation metric due to its balanced consideration of precision and recall, ensuring that both false positives and false negatives were accounted for.

To compare classification performance beyond the top-ranked prediction between all models (scenario 3), we evaluated each model's ability to correctly identify the true class within the top-*k* predicted classes. CDF plots were generated to show classification accuracy improvement with increase of *k*. A 95% cumulative accuracy threshold was included to indicate the smallest *k* value at which the model's top-*k*

accuracy surpasses 95%, offering a practical benchmark for classification reliability. Additionally, the AUC was computed for each model's CDF curve, providing a single quantitative measure of overall classification performance across multiple k levels.

To study the generalization capabilities of the object re-identification approach, another image dataset was created of non-Congolese timber genera, through the same methodology as the SmartWoodID database. The end-grain surfaces of the specimens were also sanded to a fine grit of 4000 and digitized as high-resolution RGB images using a flatbed scanner at 2400 dpi. Each species or lower taxon in the database is represented by multiple specimens from different trees, providing coverage of both intra- and inter-species anatomical variability. This diversity makes the dataset a robust resource for applying to the trained object re-identification models in this study. A summary of the taxonomic range of the included timbers is provided in supplementary materials Table 8.8. The dataset covers 58 families, 135 genera, and 234 species. The "Training and Testing Specimens" section provides statistical measures of specimen distribution per class. Specifically, the minimum indicates the number of specimens in the smallest class, the maximum represents the number in the largest class, the average reflects the mean number of specimens across all classes, and the standard deviation quantifies the variability in specimen distribution.

Table 5.1: Summary of the taxonomic range and specimen distribution in the datasets of non-Congolese timbers (for full list see Table 8.8). The "Training and Testing Specimens" section provides statistical measures of specimen distribution per class, including the minimum, maximum, average, and standard deviation.

	Training Specimens				Testing Specimens			
	Minimum	Maximum	Average	Standard deviation	Minimum	Maximum	Average	Standard deviation
Family	1	88	18	23	1	39	8	10
Genus	1	83	8	10	1	28	3	4
Species	1	83	5	7	1	28	2	2

As stated earlier, a reference database is required to assess an object re-identification model. Therefore, the dataset was partitioned into a reference set (75%) and a test set (25%), ensuring that images from the same specimen were allocated exclusively to one set to prevent data leakage.

To evaluate the model's generalization performance, a simulation framework was employed in which N classes were randomly sampled and the evaluation repeated m times per value of N . As N increased, performance trends at higher class counts were examined. To maintain computational efficiency while ensuring statistical reliability, the number of repetitions m was reduced for larger N . All simulations were conducted using predefined random seeds to ensure reproducibility. Full simulation details, including values of N and m , are provided in Supplementary materials Table 8.9.

For this experiment the binary verification model, using composite images, was evaluated by predicting a similarity score for paired images, composed of reference and test specimens. The model's output scores were processed using the same aggregation approach applied to African timbers, converting them into a probabilistic distribution across genera for each specimen. For each class size, key performance metrics were computed, including accuracy, macro-averaged precision, macro-averaged recall, macro-averaged F1-score, and the area under the CDF of ranked genus predictions. Performance trends across different class sizes were visualized using metric curves, with confidence intervals estimated via the standard error of multiple simulation results.

5.4 Results

5.4.2 Comparison of multiclass classification models

Two subsets of the dataset were used: all Congolese tree species (Blue) and the commercially traded Congolese timber species (Green). Table 5.2 presents the performance metrics for multiclass Classification (RF model trained on the described macroscopic cross-sectional anatomical features VS CNN trained on the macroscopic cross sectional image patches).

Table 5.2: A summary of the general metrics across classes (genera) for both sets of timbers and for both multiclass classification techniques (RF and CNN).

Subset	All Congolese tree species		Commercial Congolese Timbers	
Techniques	RF	CNN	RF	CNN
Accuracy	0,36	0,63	0,58	0,86
Precision	0,25	0,48	0,56	0,87
Recall	0,28	0,54	0,54	0,85
F1-Score	0,25	0,49	0,53	0,84

Across both subsets, the CNN consistently outperforms the RF model in terms of accuracy, precision, recall, and F1-score. The performance metrics of the subset of commercial timbers are significantly higher than the full set of Congolese tree species.

All tested specimens were accounted in matrices to study the correlation between both techniques. The matrices indicate the number of specimens correctly or incorrectly classified by both or either technique, with a colour gradient indicating a higher number of specimens (see Figure 5.4).

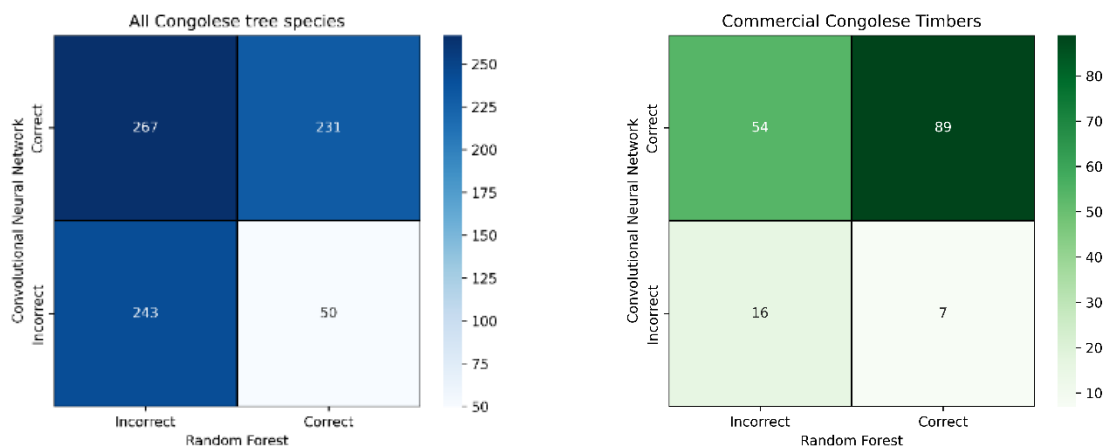


Figure 5.4: Matrices showing the number of specimens classified correctly by both, either, or none of the multiclass classification techniques (for both sets of timbers). The number of each case is indicated by the intensity of the colour gradient.

The matrices reveal a strong positive correlation between the predictions of both classifiers. When RF correctly classifies a sample, CNN is also correct in 82.2% of those cases for all Congolese tree genera and 92.7% for the commercial subset. Likewise, when CNN is correct, RF is also correct 46.4% of those cases in the full dataset and 62.2% in the commercial subset. Notably, the correlation between the two classifiers is stronger within the commercial subset (higher conditional probabilities). The asymmetry in the matrices further reveals that CNN correctly classifies a substantial number of test specimens where RF fails, but the reverse is less common. For the commercially important subset, both classifiers improve in performance. However, the CNN approach still has fewer misclassifications and a stronger ability to differentiate closely related genera.

Individual genera may exhibit varying classification performance. To account for these differences, F1-scores were computed for each genus across both classification techniques. These values were visualized in a scatter plot, with the x-axis representing RF performance and the y-axis representing CNN performance (see Figure 5.5). To better illustrate the distribution of genera according to classification performance, the scatter plot is divided into four quadrants. Genera positioned in the upper and right quadrants achieved an F1-score ≥ 0.5 for the respective models. To facilitate interpretation, the number of genera in each quadrant is displayed at its centre. These counts provide a balanced representation of model performance across taxa compared to the matrices (Figure 5.4), which is inherently skewed by class distribution and absolute specimen counts.

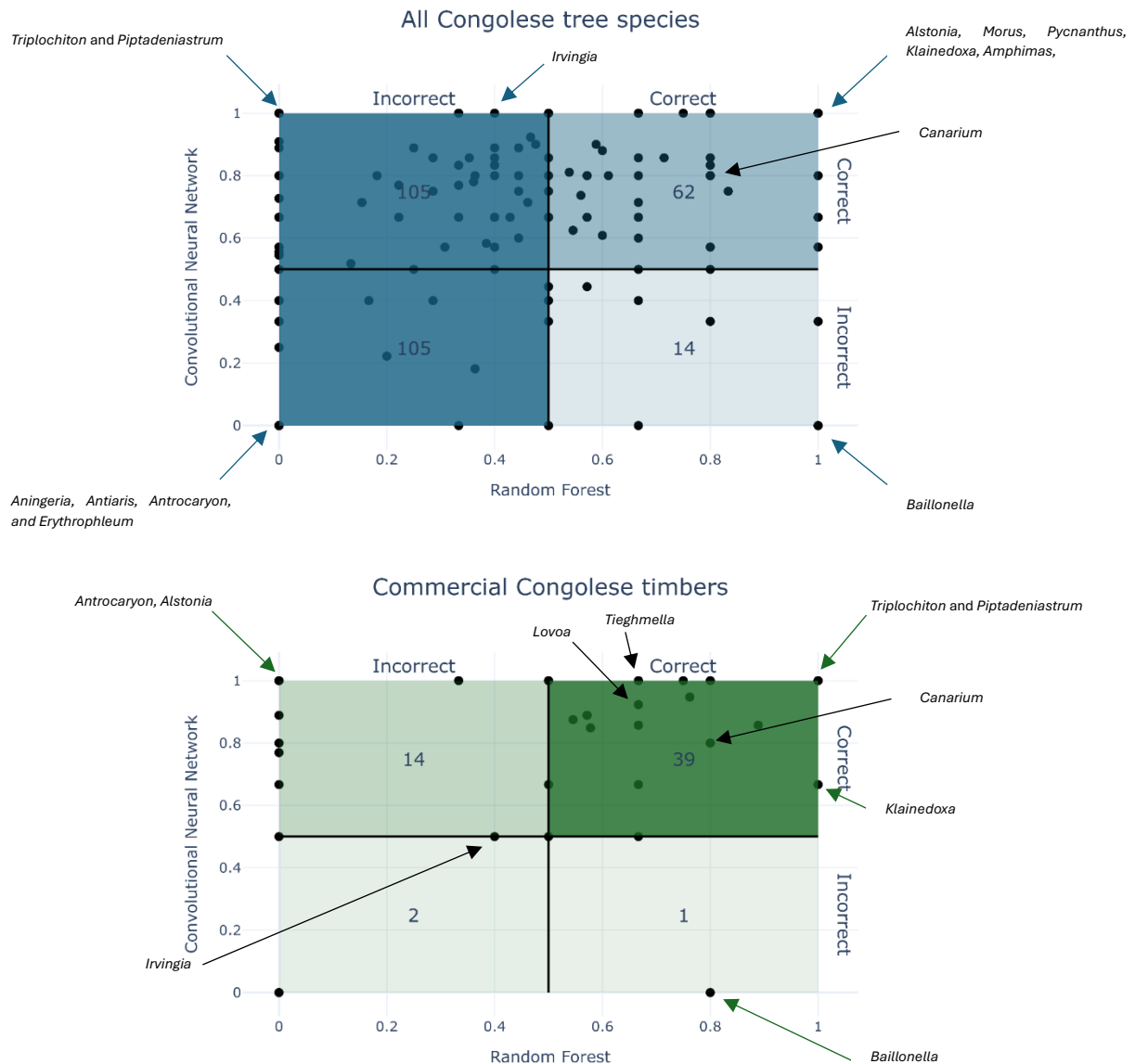


Figure 5.5: Scatterplots visualizing the F1-scores of individual genera with the Random forest model F1-score on the x-axis and the multiclass CNN model F1-score on the y-axis. The plots are divided into four quadrants of equal size to indicate performance. The number of genera in each quadrant is printed centrally in each quadrant.

Genera such as *Triplochiton* and *Piptadeniastrum* exhibit an F1-score of 1 for the CNN in the full dataset of Congolese tree genera, while the RF model fails to recognize them (F1-score = 0). This suggests that macroscopic cross-sectional IAWA features alone are insufficient for distinguishing these genera among all Congolese tree genera, whereas the CNN's extracted features enable recognition. Similarly, genera like *Alstonia*, *Morus*, *Pycnanthus*, *Klainedoxa*, *Amphimas*, and *Canarium* achieve F1-scores of 1 with both techniques, indicating their clear distinction from all other genera in the dataset. On the other hand, certain

genera, including *Aningeria*, *Antiaris*, *Antrocaryon*, and *Erythrophleum*, are poorly recognized by both methods, highlighting their morphological similarity to other genera based on macroscopic features. A distinct group of genera, such as *Baillonella*, display a clear contrast in recognition performance between the two techniques, with CNN failing to extract distinguishing features while RF succeeds.

When models are trained exclusively on the commercial subset, shifts in performance are expected, reflecting improvement in F1-scores due to a reduced complexity (e.g. taxonomic scope). Notably, *Triplochiton* and *Piptadeniastrum* transition from the upper left quadrant (misclassified by RF) to the upper right quadrant (perfect recognition by both models), suggesting that the presence of non-commercial lookalike genera in the full dataset previously impeded their classification. Additionally, *Antrocaryon* moves from the lower left quadrant to the upper left, indicating that while macroscopic features and RF do not enable accurate classification, the CNN successfully differentiates it within the commercial subset. Certain genera show improved recognition by both techniques in the commercial subset, including *Maesopsis*, *Ongokea*, *Tieghemella*, and *Lovoa*. However, some genera, such as *Pycnanthus*, *Pericopsis*, *Klainedoxa*, *Canarium*, *Baillonella*, *Amphimas*, and *Aningeria*, do not experience enhanced recognition despite the reduced taxonomic scope. Some commercial genera demonstrate a decrease in F1-score when restricting training to only commercial genera. For instance, the RF model's performance significantly drops for *Alstonia*, *Morus*, *Millettia*, *Pachylobus*, *Gambeya*, and *Tessmannia*. A similar but less pronounced effect is observed with the CNN, where performance decreases only for *Irvingia*, *Albizia*, and *Beilschmiedia*.

5.4.3 Comparison of CNN-based approaches

Table 5.3 presents the performance metrics for three different CNN classification approaches—multiclass classification, Object Re-Identification (trained with either triplet learning or binary verification)—evaluated on two subsets of the dataset: all tree genera found in the DRC (Blue) and the commercially traded timber species (Green).

Table 5.3: A summary of the general metrics across classes (genera) for both sets of timbers and for all CNNs techniques (multiclass classification, Object re-identification (Triplet learning), and Object re-identification (binary verification)).

Subset	All Congolese tree species			Commercial Congolese Timbers		
Technique	Multi-class	Object re-identification		Multi-class	Object re-identification	
		Triplet learning	Binary verification		Triplet learning	Binary verification
Accuracy	0.63	0.13	0.46	0.86	0.28	0.75
Precision	0.48	0.07	0.32	0.87	0.21	0.73
Recall	0.54	0.08	0.34	0.85	0.24	0.71
F1-Score	0.49	0.07	0.30	0.84	0.21	0.69

Across both subsets, the traditional multiclass classification approach consistently outperforms the object re-identification techniques in terms of accuracy, precision, recall, and F1-score. The performance gap is particularly pronounced for object re-identification trained with Triplet learning, which has low values for the metrics. However, the binary verification approach demonstrates considerably better performance than triplet loss, achieving scores much closer to those of the traditional multiclass CNN. The performance metrics of the subset of commercial timbers are significantly higher than the full set of Congolese tree species.

All tested specimens were accounted in matrices to study the correlation between the multiclass CNN on the one hand and the object re-identification techniques on the other. The matrices indicate the number of specimens correctly or incorrectly classified by both or either technique, with a colour gradient indicating a higher number of specimens (see Figure 5.6). The relationship between multiclass - object re-identification (Triplet learning) is shown in the upper matrices, The relationship between multiclass - binary verification is shown in the lower matrices.

The matrices for the multiclass vs. object re-identification (Triplet learning) models demonstrate a strong skew in favour of the multiclass CNN, indicating that misclassifications tend to occur in either both models or that only the multiclass CNN can correctly identify. Specifically, when the multiclass CNN correctly classifies a specimen, the object re-identification CNN (Triplet learning) is also correct in only 18.5% of cases for the full set of Congolese tree genera and 30.1% for the commercial subset. On the other hand, when the Triplet Loss CNN correctly classifies a specimen, the multiclass CNN is also correct in 89.3% of cases for the full dataset and 93.5% for the commercial subset. The binary verification CNN shows a stronger alignment with the multiclass CNN. When the multiclass CNN correctly classifies a specimen, the binary verification CNN is also correct in 60.0% of cases for the full dataset and 82.5% for the commercial subset. In contrast, when the binary verification CNN correctly classifies a specimen, the multiclass CNN is also correct in 82.8% of cases for the full dataset and 95.2% for the commercial subset.

Similar with the comparison of multiclass classification models, F1-scores were computed for each genus across both classification techniques to address differences between individual genera. These values were visualized as scatter plots using the same principles (quadrant counts), with the x-axis representing the multiclass CNN performance and the y-axis representing either object re-identification performance (See Figure 5.7).

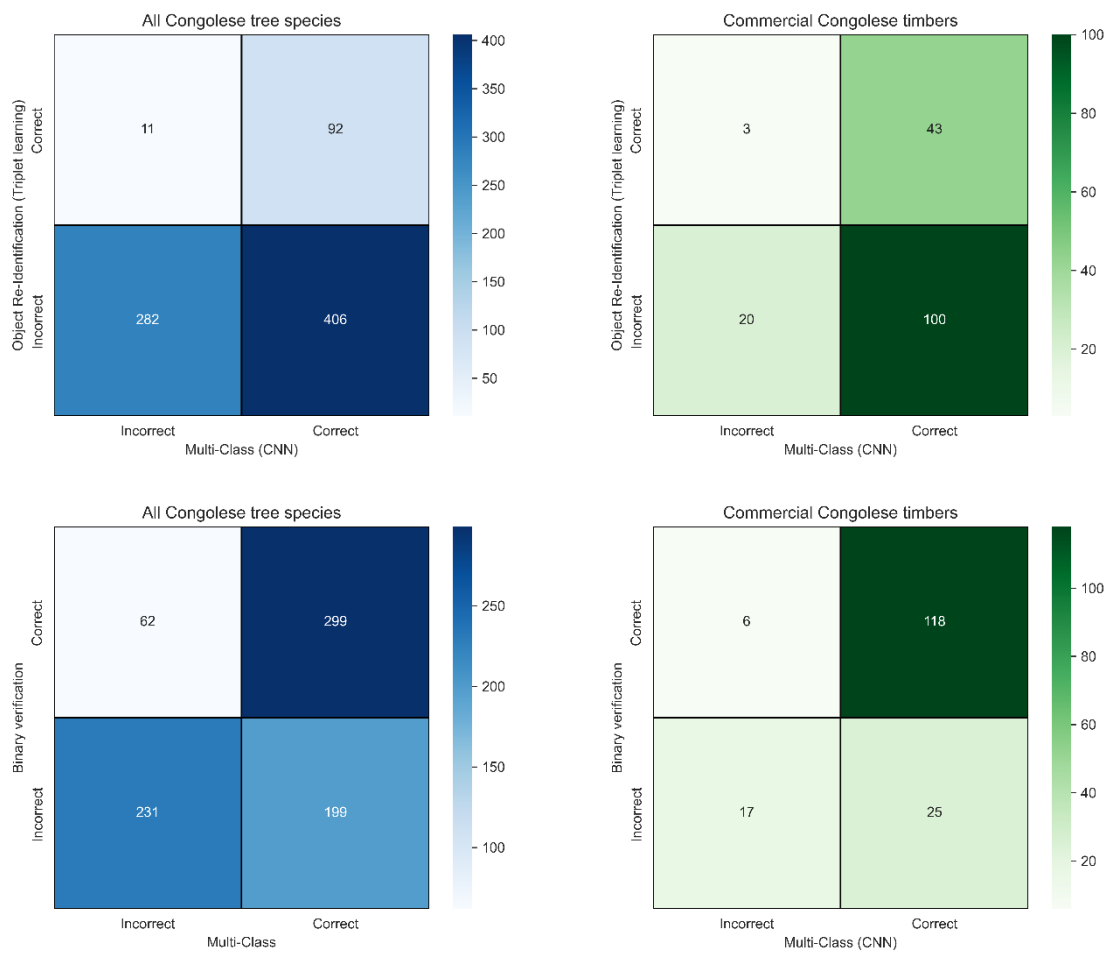


Figure 5.6: Matrices showing the number of specimens classified correctly by both, either, or none of the multiclass classification techniques (for both sets of timbers). The number of each case is indicated by the intensity of the colour gradient. The top matrices show the correlation between the multiclass CNN on the x-axis and the Object reidentification model (Triplet learning) on the y-axis; The bottom matrices show the correlation between the multiclass CNN on the x-axis and the Binary verification model on the y-axis.

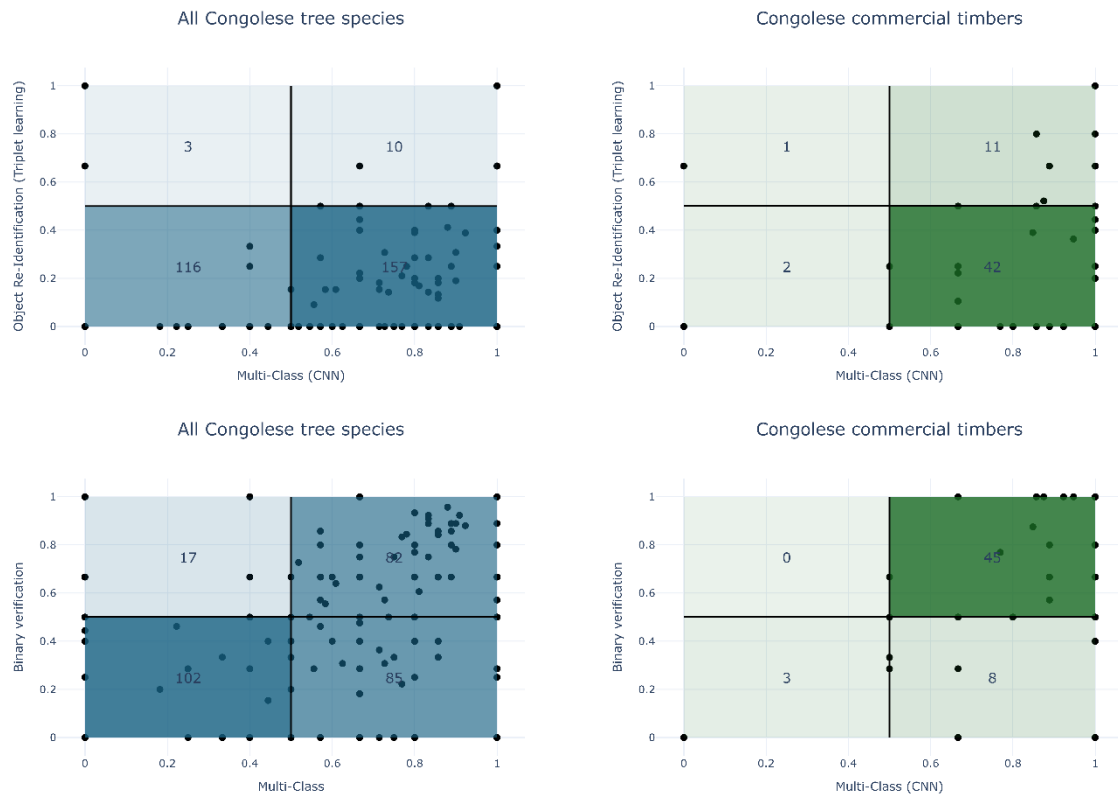


Figure 5.7: Scatterplots visualizing the F1-scores of individual genera with the multiclass CNN model F1-score on the x-axis and the object reidentification model F1-score on the y-axis. The top plots visualize the correlation with the triplet loss object reidentification model, the bottom plots with the Binary Cross Entropy loss object reidentification model. The blue plots on the left show the results for the full set of Congolese tree species, the green plots on the right for the subset of commercial timbers. The plots are divided into four quadrants of equal size to indicate performance. The number of genera in each quadrant is given in the centre of each quadrant.

Figure 5.7 shows that specific genera have markedly higher associated F1-scores for the multiclass CNN compared to the object re-identification model trained with triplet learning (top graphs), as the majority of the points (genera) are in the lowest quadrants indicating either a low F1-score according to both models or only high score for the multiclass model (bottom right quadrant). This is different for the relation between the multiclass model and the binary verification model (bottom graphs), where for all Congolese tree species (bottom left graph), the genera are more evenly distributed across quadrants. For the commercial subset, this shifts to nearly all genera having high F1-scores with both approaches.

5.4.4 Top- k classification accuracy analysis

The classification performance of the different models, assessed beyond the top-ranked prediction, is visualized in Figure 5.8 using CDF plots. The plots demonstrate the increase in accuracy as the number of considered classes (k) expands. The coloured values on the y-axis represent the accuracy for $k=1$, which corresponds to the accuracy (see Table 5.2 and Table 5.3). A key reference point in the plot is the 95% cumulative accuracy threshold, marked by a vertical dotted line, which represents the minimum k value at which each model achieves at least 95% top k accuracy. Furthermore, the AUC values are displayed for each model, providing an aggregated metric that encapsulates overall classification performance across varying k levels.

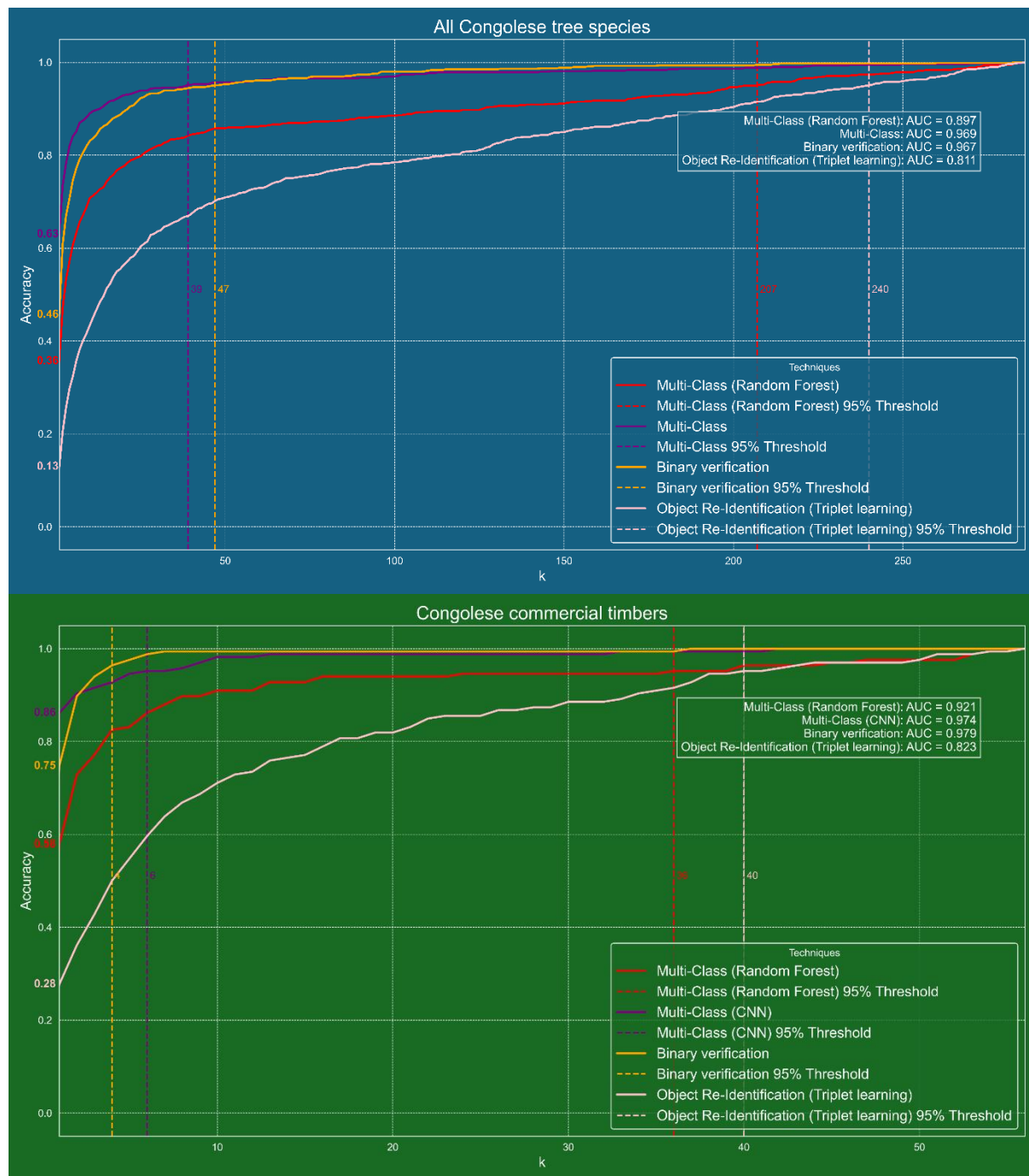


Figure 5.8: Cumulative distribution function plots, showing the progression of top k accuracy with increasing numbers of k ranked predictions for multiclass models (Random Forest, CNN (Categorical Cross Entropy (CCE)) and Object Reidentification models (Triplet loss and Binary Cross Entropy loss). The vertical dotted lines represent the value k where the top k accuracy exceeds 0.95. The coloured values on the y-axis represent the top-1 accuracy (normal accuracy). The text box on the left displays the Area Under Curve for the CDF of each model. The top plot (blue) shows the results for all Congolese tree species, the bottom plot (green) shows this for the subset of commercial timbers.

Top-1 accuracy serves as a reference point for evaluating model performance, providing insight into initial classification effectiveness. Across all models, accuracy improves with increasing values of k , following a characteristic saturating pattern—where initial gains are pronounced but progressively diminish as k increases.

Across both the full dataset and the commercial timber subset, the RF model consistently underperforms compared to the multiclass CNN and the object re-identification CNN trained for binary verification. RF exhibits lower top-1 accuracy, a slower rate of accuracy improvement, and lower AUC values, indicating weaker overall classification performance. Notably, RF requires a substantially higher k value to exceed 95% accuracy, suggesting that its predictions stabilize at a slower rate compared to the CNN-based approaches.

When tested on the full dataset, the multiclass CNN outperforms the binary verification CNN, achieving higher top- k accuracy across $k = 1$ to 40. However, a different pattern emerges for the commercial timber subset. While the multiclass CNN initially surpasses the binary verification CNN at $k = 1$, this trend reverses beyond $k = 2$, where the binary verification model achieves superior accuracy. By $k = 10$, both models reach similar top- k performance, with accuracy approaching 1. These findings are further contextualized by the 95% cumulative accuracy threshold. For the full dataset, the multiclass CNN maintains superior classification performance. However, for the commercial subset, the binary verification CNN demonstrates stronger performance at lower k values. In contrast, the object re-identification CNN trained with Triplet Loss demonstrates the weakest performance across all k values. It exhibits the lowest top-1 accuracy, the slowest rate of accuracy improvement, and the lowest AUC values. Furthermore, it requires a substantially higher k value to exceed 95% accuracy, indicating a slower stabilization of predictions.

5.4.5 Generalization to unseen taxa

Open-world recognition through the binary verification model, trained on Congolese timber genera, was assessed on unseen non-Congolese timber genera through the implementation of a structured simulation framework. The evaluation involved partitioning non-Congolese timber specimens into a test set and a reference set to facilitate genus similarity assessment. A subset of N genera was sampled, and all corresponding specimens from both sets were selected. Composite images were generated by pairing each test specimen with each reference specimen, enabling the binary verification model to produce similarity scores that indicate the likelihood of genus correspondence between the two images. These similarity scores were then transformed into a normalized probability vector of length N , where each probability represented the likelihood of the test specimen belonging to a particular genus in the sample. This transformation followed the methodology detailed in Section 5.3.8 of the Materials and Methods.

To assess the model's performance under increasing classification difficulty, the evaluation was repeated across seven values of N , representing the number of genera sampled (see Table 8.2). As N increased, the classification task became more challenging due to a larger verification set. For each value of N , the sampling process was repeated m times (e.g., $N = 10$ with $m = 50$ repetitions) to ensure result robustness. In each repetition, key performance metrics were calculated, including accuracy, macro-averaged precision, recall, and F1-score. Additionally, top- k accuracy was analysed by computing the area under the cumulative distribution of ranked genus predictions and identifying the smallest k required to achieve 95% accuracy. Results are summarized in Figure 5.9, which includes six subplots—one for each metric—showing performance trends as a function of N . Each data point reflects the average metric across repetitions, with error bars indicating the standard error of the mean (SEM), and numerical values of mean $\pm$ SEM displayed above.

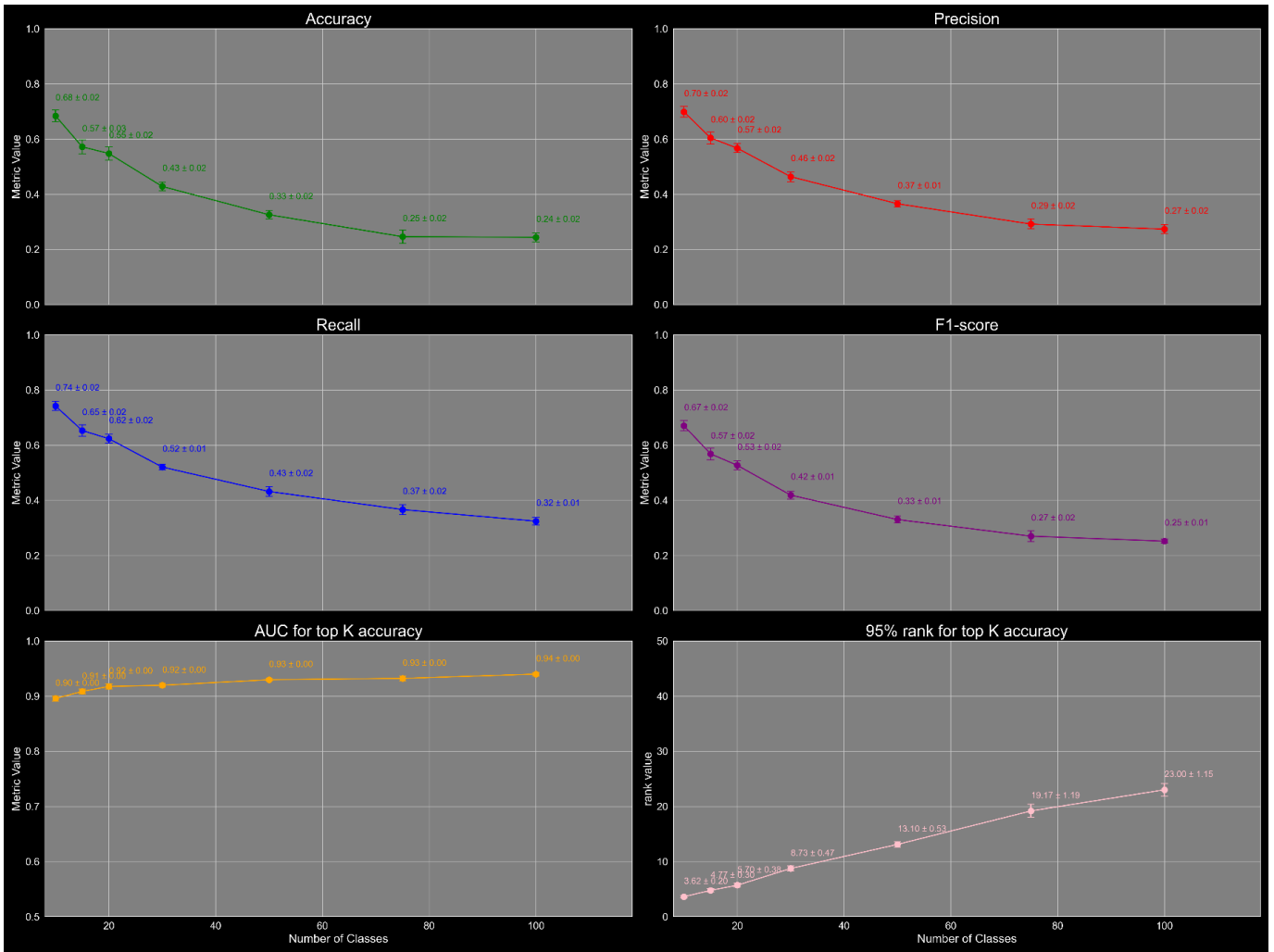


Figure 5.9: Performance evaluation of a binary verification model on unseen non-Congolese timbers, showing accuracy (top left, green), precision (top right, red), recall (centre left, blue), F1-score (centre right, purple), AUC for top K accuracy (bottom left, yellow), and 95% rank for top K accuracy (bottom right, pink) across different numbers of genera (x-axis). Metrics for each number of genera are visualized as points with metric value (y-value) printed on the right of each point. Each point demonstrates error bars representing the 95% confidence interval around every metric value, based on the metric values of each simulation for that point.

The accuracy, precision, recall, and F1-score, in Figure 5.9 curves exhibit matching decreasing patterns across all tested values of numbers of genera (N). For each N, the average values of these metrics are close to one another, and their associated standard errors of the mean are consistently between 0.01 and 0.02, indicating a high degree of stability in the model's predictions across all X simulation runs for each N. As the number of considered genera (N) increases, the accuracy, precision, recall, and F1-score decrease, as the curves stabilize. The AUC of the top-k CDF remains relatively stable across all values of N, suggesting that the model consistently maintains its ability to rank the correct genus among the top predictions across numbers of considered genera (N). AUC values closer to 1 indicate ideal model ranking performance and values near 0.5 suggest random guessing. The observed values between 0.90-0.94 suggest that the model is generally successful in prioritizing the correct genus in its ranked predictions. The 95% rank for top-k accuracy further supports this interpretation. As N increases, a larger number of top predicted classes must be considered to capture 95% of the correct predictions. For example, when classifying among 100 unknown genera, the correct genus falls within the top 23 predictions in 95% of cases. However, the proportion of genera remains stable across different values of N, with the average proportion and standard mean of error being 0.29 ± 0.02 , indicating that approximately the top 30% predicted genera must be considered to be correct in 95% of the cases.

5.5 Discussion

The results indicate that, despite methodological differences, all models capture similar underlying patterns, as evidenced by the positive correlation between their predictions in the matrices. This is further demonstrated by the fact that the probability of a model with a correct classification, given that another model is correct, consistently exceeds the model's overall accuracy—an observation that holds for all models. This suggests that models tend to classify the same specimens correctly more frequently than expected by chance. Correlations between human and CNN-based classification have been explored in other research domains, where studies have examined that both extract similar patterns and achieve comparable accuracies (Thoidis et al., 2023). Despite this shared capacity for pattern recognition, CNNs consistently outperform the RF model, likely due to their ability to directly extract discriminative features from images rather than relying on predefined expert-defined features. This is evident in the skewed matrices, where CNNs correctly classify test specimens that RF misclassifies more frequently than the reverse. This aligns with the expectation that CNNs, by learning image representations, capture finer-grained details that RF models cannot access. These findings are corroborated by literature, such as the study by Knauer et al. (2019), which demonstrated that CNNs surpassed machine-learning approaches like RF and SVMs in tree species identification based on hyperspectral image data (Knauer et al., 2019). The CNN model in that study was able to intrinsically integrate spectral and spatial information, achieving superior performance levels comparable to RF and SVM models that relied on handcrafted features, reinforcing the broader advantage of CNNs in feature extraction.

In commercial timber classification, the good performance metrics are primarily attributed to the reduced taxonomic scope, which simplifies the classification task and generally leads to higher accuracy (Ali et al., 2024). However, this trend does not apply uniformly across all genera. Certain genera exhibit lower F1-scores when trained exclusively on a restricted dataset of commercial timbers. This phenomenon could be attributed to CNNs trained on the full dataset learning more generalized, texture-based features, whereas models trained on commercial genera only can be overfit to superficial differences, thereby reducing their capacity for generalization.

The top-*k* accuracy findings further underscore the effectiveness of deep learning in timber genus identification across both ecological and commercial contexts. In high-diversity environments (such as the broad range of tree species in the Congo basin), CNN-based models achieve a 95% probability of correct genus classification when considering the top 40–50 ranked predictions of the 286 genera (approximately top 16% of all genera), whereas for commercial timber classification, this confidence threshold is reached with only the top 4–6 predicted genera for the 56 commercial genera (approximately top 7% of all genera). These results quantify the practical utility of CNN-based classification pipelines for large-scale timber identification with only the macroscopic cross-sectional wood anatomy. They also emphasize the importance of considering not only the timber class with this highest probability. Research by De Oliveira et al. (2019) confirms that CNNs trained for wood classification can frequently assign the correct timber label as the second or third most probable prediction, highlighting the importance of considering predictions beyond the top-ranked classification (De Oliveira et al., 2019). However, these outcomes do not represent the upper limit of what can be achieved using macroscopic cross-sectional data. Increasing the number of specimens in the training dataset has the potential to further enhance model performance. This is supported by findings from a study published as part of conference proceedings at the 26th IUFRO world congress (Stockholm, Sweden 2024), demonstrating that multiclass classification of the same Congolese commercial genera, using five-fold cross-validation, achieved stable accuracies of 0.95 across folds (De Blaere et al., 2024). Their dataset, which comprised 1,519 specimens digitized similarly as the SmartWoodID specimens, featured a broader range of cross-sections, contributing to improved classification robustness.

Object re-identification offers expanded opportunities for automated wood identification beyond traditional multiclass classification. The binary verification approach achieved performance comparable to the multiclass CNN. This method presents a viable and accessible alternative for developing wood identification models, with the added advantage of successfully verifying timbers it had never encountered

during training. Although performance on the non-Congolese timber genera decreases when verifying to a higher number of considered classes, the proportion of considered classes to reach 95% probability remains stable at approximately the top 30% of predicted genera. This indicates that the identification probability of unseen timbers is lower, compared to the Congolese timber genera the model was trained on. Object re-identification offers additional advantages for wood identification, particularly in terms of scalability and adaptability. Unlike traditional multiclass classification models, which require training on a fixed set of timbers and must be retrained whenever new timbers are introduced, an object re-identification model can operate flexibly across diverse datasets. This eliminates the need for maintaining numerous separate multiclass models, reducing the burden of data collection and enhancing deployment efficiency. Additionally, running multiple independent classification models can lead to inconsistencies, reduced interpretability, and challenges in maintaining model stability. In contrast, a re-identification approach allows a single model to generalize across different timber datasets, facilitating seamless updates and broader applicability. Forensic research also benefits from object re-identification, as it not only classifies specimens but also retrieves the most closely matching reference samples. This enhances interpretability by providing direct comparisons to vouchered specimens, strengthening the reliability of forensic conclusions. Furthermore, a single re-identification model can integrate digitized specimens from multiple institutional collections worldwide, allowing forensic cases to be examined against a far more extensive and diverse set of reference materials. This level of global collaboration significantly improves identification accuracy, as the model can match with the most well-documented and vouchered specimens across various collections, rather than relying on a single institutional dataset. A key limitation of the binary verification approach is its reduced ability to rank multiple class predictions effectively, as it prioritizes minimizing classification error through optimal threshold selection rather than optimizing ranking accuracy (Chen et al., 2017). Unlike multiclass classification, which assigns probabilities to all possible classes and facilitates ranking based on confidence scores, binary verification generates only a similarity score for each pairwise comparison, necessitating post-classification processing to derive a ranked list. Consequently, the binary verification model is more prone to misranking closely related timber genera, increasing the likelihood of incorrect top-1 predictions. This limitation is reflected in the lower top-1 accuracy observed in the top-k accuracy analysis for both the full set of Congolese tree species and the commercial subset. Furthermore, the 95% cumulative threshold analysis confirms this discrepancy, as the binary verification approach requires eight additional predicted classes (ranks k) to reach this threshold compared to the multiclass model for the full Congolese dataset. However, results also indicate that reducing the taxonomic scope can mitigate this issue. For the commercial subset, the binary verification model achieves the 95% cumulative threshold with fewer additional ranks than the multiclass CNN. The assessment of performance on unseen non-Congolese timbers underscores this, as a lower number of possible genera resulted in higher accuracy, precision, recall, and F1 score. Chen et al. (2017) suggests that integrating a ranking-based loss function could enhance the model's ability to distinguish between highly similar genera while preserving classification performance (Chen et al., 2017). Performance could also potentially be improved by changing the backbone architecture, loss functions or hyperparameters (e.g. batch size, learning rate, ...). Another key limitation of the binary verification approach is its reliance on generating large numbers of image pairs between test and reference specimens, which significantly increases computational demands (Chen et al., 2017; Hermans et al., 2017). We note that this could have influenced the results by limiting the scope of the reference database and the number of available patches per test specimen to accommodate evaluation in this study. Still the results provide clear indication of this techniques' potential for field application and for further accuracy improvement. In addition, while the computational demands pose challenges for large-scale applications—where processing times may become prohibitively long (see 0)—it remains a practical solution for forensic casework. For example, comparing a single unknown specimen (1×1 cm or four 0.5×0.5 cm patches) against a reference set of 600 commercial genera would take approximately six minutes using the computational resources available in this study. Processing time could be further reduced by optimizing hardware, using a smaller reference set, or implementing a multi-step approach. One such approach could involve an initial stage that narrows down the verification to a limited set of reference specimens per genus. This refined selection could then be verified against multiple specimens and multiple image patches per specimen, ensuring a reliable identification while maintaining efficiency. The binary verification approach also provides promising

perspectives for ruling out certain timbers. Wood verification is a crucial process in tiered approaches to highlight suspicious cargo for further forensic in-depth analysis. The binary verification methodology aligns with the desired output for wood verification, producing a single independent score representing possible matches to a selection of timbers. For example, the binary verification model can reference expertise samples to a limited number of internationally protected timbers, providing a direct similarity score and enabling a fast and accurate screening of cargo.

In contrast, the embedding-based object re-identification approach using triplet loss provides less doubt in ranking classes and is more computationally efficient (Chen et al., 2017; Hermans et al., 2017; Schroff et al., 2015). By transforming reference images into embedding vectors, the reference database can be stored in tabular form, requiring minimal memory while enabling rapid predictions using machine-learning methods such as RF or K Nearest Neighbours with cosine similarity. However, in this study, embedding-based object re-identification consistently underperformed across all evaluation metrics—including accuracy, precision, recall, F1-score, matrices, scatterplots, and top-k accuracy—indicating poor generalisation. This contrasts with the performance of the binary verification approach, despite both methods utilizing the same backbone CNN architecture, augmentation strategies, dataset balancing techniques, and an identical mining strategy that selected hard anchor-positive-negative examples based on macroscopic wood anatomical descriptions to optimize learning for look-alike genera. Existing literature suggests that embedding-based performance can be improved by incorporating more advanced mining strategies, such as online learning, where the model dynamically selects the most challenging triplet combinations based on its own performance rather than relying on predefined macroscopic wood anatomical features (Hermans et al., 2017; Schroff et al., 2015). This approach enhances training efficiency by prioritizing samples where the model struggles most. While these techniques have been successfully implemented in frameworks like PyTorch, their integration within TensorFlow proved challenging in this study. Further improvements may be achievable using more advanced loss functions, such as histogram loss, which can optimize the distribution of embeddings in feature space more effectively (Ustinova and Lempitsky, 2016). Additionally, integrating classification-reconstruction learning approaches, such as the CROSR framework, could enhance open-world classification performance (Yoshihashi et al., 2019).

Despite the suboptimal performance observed in this study, embedding-based object re-identification remains a promising avenue for advancing CV-based wood identification, given its success in other domains (Bai et al., 2018; Chen et al., 2023; Ghosh et al., 2023; Hermans et al., 2017; Kumar et al., 2020; Schroff et al., 2015; Shen et al., n.d.; Ye et al., 2024). Beyond its computational efficiency, this approach holds significant potential for integration with other diagnostic information within a unified framework. The embeddings generated by the CNN function as numerical fingerprints of macroscopic wood anatomy, encoding diagnostic features as fixed-length vectors. This data format aligns well with other established methods, such as microscopic wood anatomical assessment and DART-TOFMS. Microscopic wood anatomy, the most traditional wood identification approach, encodes diagnostic traits from all three anatomical planes into a structured vector based on qualitative observations (Committee, 2004; NS, 1989; Wheeler, 2011). DART-TOFMS generates chemical fingerprints by ionizing wood compounds and analysing their mass-to-charge ratios (Deklerck, 2022, 2019; Deklerck et al., 2020). Since all three techniques produce structured numerical representations of wood characteristics, they offer the potential for integration into machine-learning or deep learning models. By combining macroscopic, microscopic, and chemical data, future models could substantially enhance diagnostic accuracy, potentially enabling fine-grained identification at the species, subspecies, or populations. This is also suggested by Knauer et al. (2019) demonstrating that combining different models for tree identification can yield improved results (Knauer et al., 2019).

5.6 Conclusion

This study demonstrates the potential of deep learning for automated timber genus identification. The results highlight that while all classification models capture similar underlying patterns, CNNs outperform RF models on IAWA anatomical features due to their ability to extract discriminative image features without relying on predefined descriptors. Performance trends across different taxonomic scopes emphasize the importance of training data diversity, as models trained on broader datasets exhibit greater generalization capabilities compared to those trained exclusively on commercial timbers.

Beyond standard multiclass classification, object re-identification approaches provide valuable alternatives, particularly in forensic contexts where identification may be less critical than ruling out certain timbers. The binary verification approach demonstrates strong performance in this regard, though effectiveness is constrained by ranking limitations and computational demands. Embedding-based re-identification, while computationally efficient, underperforms in this study, suggesting that improved mining strategies and loss functions could enhance its reliability. Additionally, object re-identification produces information on similarity to specific reference specimens, rather than producing a direct prediction of classes (e.g. genera in this study), providing valuable information for forensic researchers.

Future research should address the limitations of embedding-based models using the current state-of-the-art to offer a powerful approach for automated wood identification. These models should be explored in hybrid studies that integrate multiple diagnostic data modalities—macroscopic images, microscopic wood anatomy, and chemical fingerprinting—to improve classification accuracy and enable identification at sharper taxonomic resolution. As global efforts to combat illegal logging and enforce sustainable trade regulations intensify, advancing AI-driven timber identification will be essential for strengthening forensic capabilities and ensuring responsible resource management.

5.7 Acknowledgements

This work was realized through the combined efforts of the staff at the service of Wood biology at the RMCA in Tervuren, Belgium, and the staff at the UGent-Woodlab (Department of Environment, Faculty of Bioscience Engineering) of Ghent University. The collection and specimens were managed by Maaïke De Ridder (RMCA) and Daniel Wallenus (RMCA).

5.8 Availability of code and metadata

Scripts and metadata can be requested by contacting the first author:

Contact: ruben.de.blaere@africamuseum.be

Chapter 6: Combining expert-defined and CV-extracted macroscopic cross-sectional features for enhanced wood identification

Authors: Ruben De Blaere^a, Wannes Hubau^{a,b}, Hans Beeckman^a, Jan Verwaeren^c, Jan Van den Bulcke^b

a Service of Wood Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium, b UGent-Woodlab, Laboratory of Wood Technology, Department of Environment Ghent University, Coupure Links 653, 9000 Gent, Belgium, c UGent-KERMIT, research unit Knowledge-based, predictive and spatio-temporal modelling, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Gent, Belgium

6.1 Abstract

As reported in Chapter 3, using expert-defined anatomical features for classifying commercial timbers results in limited accuracy, with genus-level identification reaching only around 53%. However, these features were effective for distinguishing closely related taxa, such as different species of *Pterocarpus*. In contrast, Chapter 5 demonstrated that CNNs perform well across broad taxonomic ranges (e.g. all Congolese tree species and commercial species). The binary verification approach, in particular, showed strong generalization with high accuracy, precision, recall, and F1-score on both trained (Congolese) and untrained timbers—including the correct genus in 95% of the specimens when considering the top 7% of predicted genera for Congolese timbers and up to 30% for non-Congolese timbers. Still, CNNs do not necessarily rely on the diagnostic patterns used by wood anatomists for classification. Therefore, expert-defined anatomical features may offer complementary information that can help refine or validate the top predicted taxa generated by a CNN.

This chapter evaluates a re-ranking framework where expert-defined features are used in RF models to refine CNN genus predictions. This chapter explores the tiering of computer-extracted CNN and expert-defined macroscopic cross-sectional wood anatomy to improve the accuracy, precision, and sensitivity (recall) of genus identification. Results show that while CNNs extract diagnostic patterns, targeted re-ranking—especially within the top three to five predictions—yields modest but consistent performance gains. Beyond this range, refinement leads to general performance becoming lower than without re-ranking, particularly in genera with overlapping or ambiguous anatomical traits. Moreover, re-ranking is not always beneficial: around one-third of the 56 evaluated genera showed decreased performance, most notably in sensitivity (recall). This decline in recall is especially concerning in regulatory applications, where speed, accessibility, and high sensitivity are essential—such as verifying timber species noted in trade documentation and identifying protected taxa. Misclassifying CITES-listed species, such as within the genus *Khaya*, would be problematic. These findings underscore the value of integrating CNNs with expert knowledge but emphasize the need for targeted, genus-specific refining to enhance diagnostic accuracy without compromising the reliability or practicality of field-based timber verification systems.

This chapter is not submitted to a peer-reviewed journal on 19/08/2025.

6.2 Introduction

Macroscopic cross-sectional wood anatomy has been studied thoroughly for wood identification in field settings, due to the ease with which diagnostic features can be observed (Ravindran et al., 2021; Ruffinatto et al., 2015). Identification is typically performed by examining the endgrain surface of a cut or sanded specimen, using the naked eye or a hand lens to assess anatomical features such as vessel arrangement, ray width, and parenchyma patterns (Koch et al., 2018; Naturalis Biodiversity center, 2025; Ruffinatto et al., 2015). These observations are then entered with identification keys, which are available online for different geographic or taxonomic groups of timbers (Malaysian Timber Council, 2018; Richter et al., 2017; Ruffinatto and Crivellaro, 2019). In CV-based wood identification, an image of the endgrain surface is captured and processed directly by an AI model (Hwang and Sugiyama, 2021; Silva et al., 2022). One example is the Xylotron, which uses a purpose-built imaging device (the ‘xyloscope’) to standardize image acquisition, ensuring compatibility with the data used to train the model (Ravindran et al., 2020, 2019). Other tools, such as the Xylorix smartphone app, rely on integrated smartphone cameras, requiring careful image augmentation to minimize decrease in performance due to differences in camera hardware and settings (Artemov, 2025; Tang and Tay, 2019).

In Chapter 3 of this dissertation the classification performance of expert-defined anatomical features was investigated, while in Chapter 4 and 5 features were extracted automatically through CV (specifically CNNs). Expert-defined macroscopic features yielded limited accuracy—achieving a maximum of 52.86% accuracy across Congolese commercial timber genera using a RF classifier. However, within narrow taxonomic groups, such as species within the genus *Pterocarpus*, these features proved more effective. By contrast, CNNs demonstrated significantly higher classification performance across broader taxonomic ranges. In particular, the binary verification approach studied in Chapter 5 produced high accuracy, precision, recall, and F1-score not only on timbers from the training set (Congolese species) but also on unseen timbers from other regions. Nevertheless, for accurate genus-level classification, it was necessary to consider the top 7% of predicted genera for Congolese timbers and up to 30% for non-Congolese specimens to include the correct genus in 95% of cases. For practical applications, this translates to reviewing the top 4–6 predictions for the 56 Congolese commercial genera (see section 5.4.4).

These findings highlight the potentially complementary strengths of both approaches. While expert-defined features are effective at distinguishing among similar-looking timbers within a narrow range, CNNs are better suited for broader classification tasks. The integration of both methods leverages their complementary strengths, enhancing both the accuracy and precision of timber identification. Combining methodologies that rely on different diagnostic patterns has proven to increase the performance of wood identification (Dormontt et al., 2015; Knauer et al., 2019). Chapter 4 further supports this integration, showing that CNNs activate regions on images with clear anatomy for classification, as visualized through gradient-weighted class activation maps. However, the features used by CNNs do not align directly with traditional anatomical descriptors, suggesting that expert-defined features could add complementary value.

This study investigates whether integrating expert-defined macroscopic wood anatomical features can improve genus-level timber identification by refining the predictions of a CNN binary verification model. Using specimens of Congolese commercial timber from the SmartWoodID database, we evaluate a tiered re-ranking approach that applies anatomical features to the CNN’s top-k predicted genera (from top-2 to top-10). The aim is to assess the extent to which expert knowledge can enhance classification accuracy, precision, recall, and F1-score—both overall and at the genus level—for the forensic identification of traded timber products.

6.3 Material and methods

6.3.1 Dataset Description

This chapter is based on the SmartWoodID database, serving as a valuable resource for examining the relationship between macroscopic cross-sectional wood anatomy and the botanical diversity of Congolese tree taxa (De Blaere et al., 2023). The following information in this section, re-uses the text in section 1.5 which provides an overview of all data used in each chapter.

The database contains high-resolution RGB scans of the macroscopic end-grain surfaces of 3,742 wood specimens, representing 954 species native to the DRC. Each specimen was prepared by scanning the cross-section at 2400 dpi using a flatbed scanner. This resolution allows for the visualization of macroscopic features essential for wood identification.

Species and lower taxa are represented by multiple specimens, capturing both intra- and interspecific anatomical variation. This makes the database well-suited for studying wood identification using macroscopic anatomy. A complete overview of the database is provided in Chapter 2 and in De Blaere et al. (2023) (De Blaere et al., 2023), while Supplementary Materials Table 8.1 all unique specimen identifiers and metadata. To enable machine learning analysis, we selected only species represented by at least two specimens. This set comprises 2,296 digitized specimens across 601 species, 286 genera, and 64 families. Discriminatory power was mainly assessed by training and evaluating classification models on the specimens. Therefore, specimens were allocated random to training (75%) or test set (25%), while preserving distribution of species across both sets. Within both sets, a subset of 78 commercially important species was defined for targeted evaluation. An overview of the chapters, designated datasets used in each chapter and of which hypothesis they target is provided in section 1.4.

Macroscopic wood anatomical features were described for each end-grain image of the SmartWoodID specimens (see chapter 2, Table 2.2; and De Blaere et al. (2023) (De Blaere et al., 2023)). Each feature is assigned a Macroscopic IAWA feature number (Ruffinatto et al., 2015). We did not use descriptions on the presence of growth rings as the discernability at the used resolution was often not high enough to assess this feature with certainty. Each feature is annotated with one of four states: Present (clearly visible), Variable (sporadically observed), Absent (rarely observed, below the threshold for Variable), or NA (undiscernible due to resolution limits or ambiguous visual cues).

Given the large size of the scanned end-grain surfaces, directly processing full images is computationally expensive and may lead to overfitting due to the limited number of images. To address this, the full images were cropped into smaller, non-overlapping patches that retained diagnostic features while capturing anatomical variability. We selected patch dimensions of 512×512 pixels, corresponding to a physical area of approximately 5.42×5.42 mm, ensuring sufficient detail for anatomical analysis without obscuring critical features. Full images yielding fewer than four patches (i.e., total end-grain surface area below 1 cm²) were excluded to maintain consistency and relevance for anatomical assessment.

Patches were allocated random to training (75%) or validation set (25%), while preserving distribution of specimens across both sets, using different regions of the same specimens for training and validating. Furthermore, any full image yielding fewer than four usable patches was removed, as these would cover less than one square centimetre—below the typical area assessed by a human expert during wood anatomical identification.

This chapter only applied images and annotations of the subset of commercial timbers. First, the binary verification CNN (chapter 5) was applied to each test specimen to generate a normalized vector of class probabilities across 56 commercial timber genera. For further details regarding the image data, data pre-processing, CNN architecture, training procedures, model selection, and hyperparameter optimization, see section 5.3.

The second component of the tiered approach involves re-ranking the CNN's top predictions—genera with the highest assigned probability—by using expert-defined macroscopic anatomical features to refine the

initial ranking. These features are described in detail in section 2.3.6. For each test specimen, a custom machine-learning classifier was trained using the anatomical annotations of training specimens belonging to the top- k predicted genera. The re-ranking process was structured incrementally: starting with the top two predicted genera per specimen, then increasing to the top three, and so on up to the top ten. In Chapter 5 it was shown that for the commercial genera subset, the correct genus is present within the top four predictions in 95% of test cases and within the top ten in approximately 99%, justifying the choice to restrict re-ranking to a range of k from two to ten. RF was selected as the classification technique based on its consistently strong performance across taxonomic levels. The RF technique was then applied to the same test specimen to produce a refined probability distribution over the top- k genera. These RF models were trained using the optimal hyperparameters of the hyperparameter grid (see Table 8.2), which consisted of 200 DTs, a maximum DT depth of 20 nodes, a minimum of two samples to split and a minimum of 1 sample per leaf. Class weighting was applied to prevent overfitting to genera with more training specimens.

Overall, this integration framework provides a structured means to assess the added value of expert-defined anatomical features in refining CNN-based genus predictions within a commercially relevant timber dataset.

A flowchart of the tiered approach is shown in Figure 6.1.

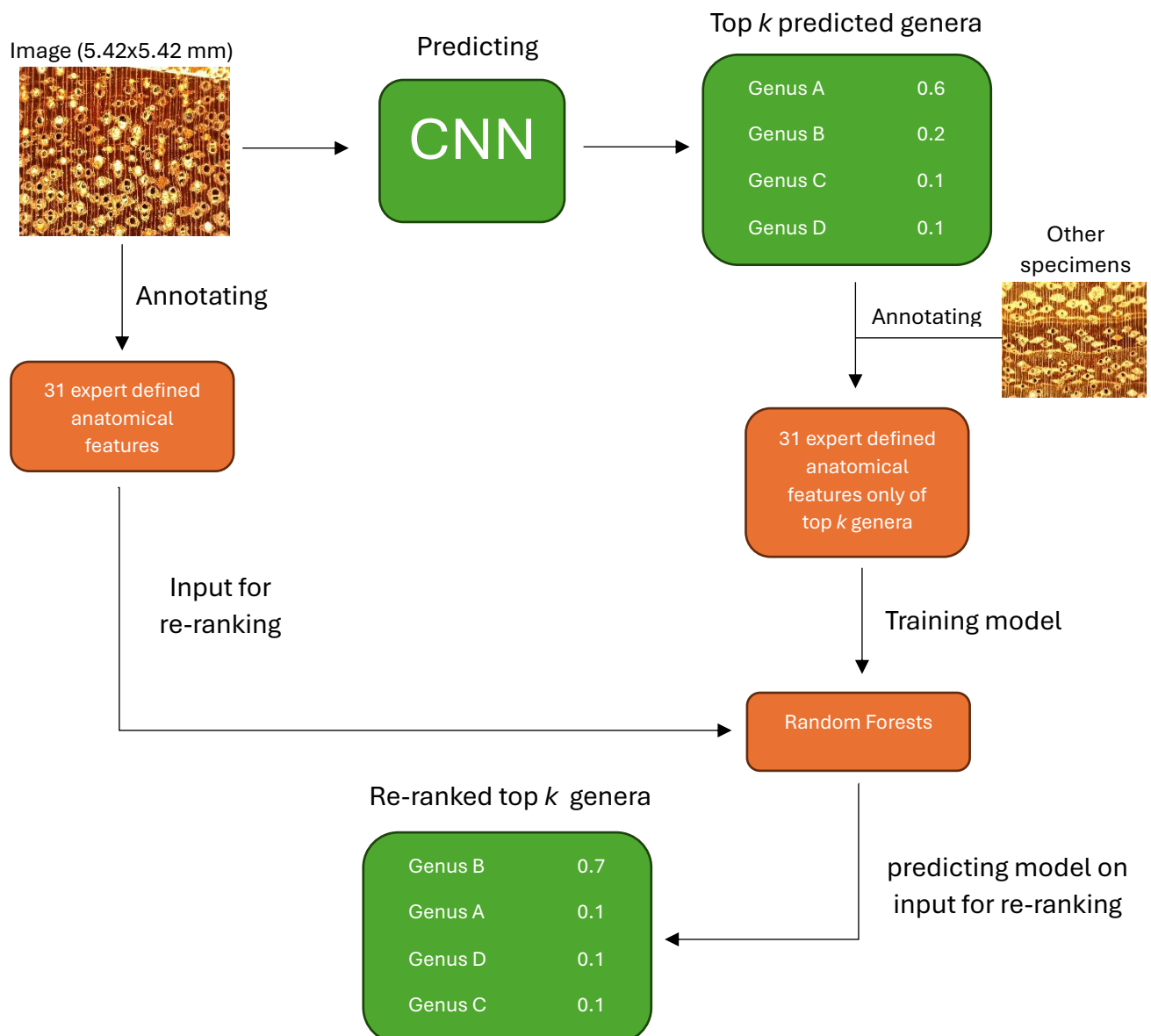


Figure 6.1: Flowchart showing the re-ranking approach

6.4 Results

The overall performance trends of the tiered approach are presented in Figure 6.2, which plots accuracy (green), recall (blue), precision (red), and F1-score (purple) as a function of the number of top predicted genera included for re-ranking. Definitions of those metrics are provided in supplementary materials section 8.2. The first x-value corresponds to the metric values obtained from the binary verification approach without any re-ranking. Subsequent points reflect metric values when only the top n predicted genera (e.g., top 2, top 3, etc.) were re-ranked using a RF classifier trained on macroscopic wood anatomical features. In the recall, precision, and F1-score plots, each x-value also includes a boxplot representing the distribution of class-specific (i.e., genus-level) scores, with the grey bar representing the interquartile range, the points being the macro-average and the coloured line representing the median.

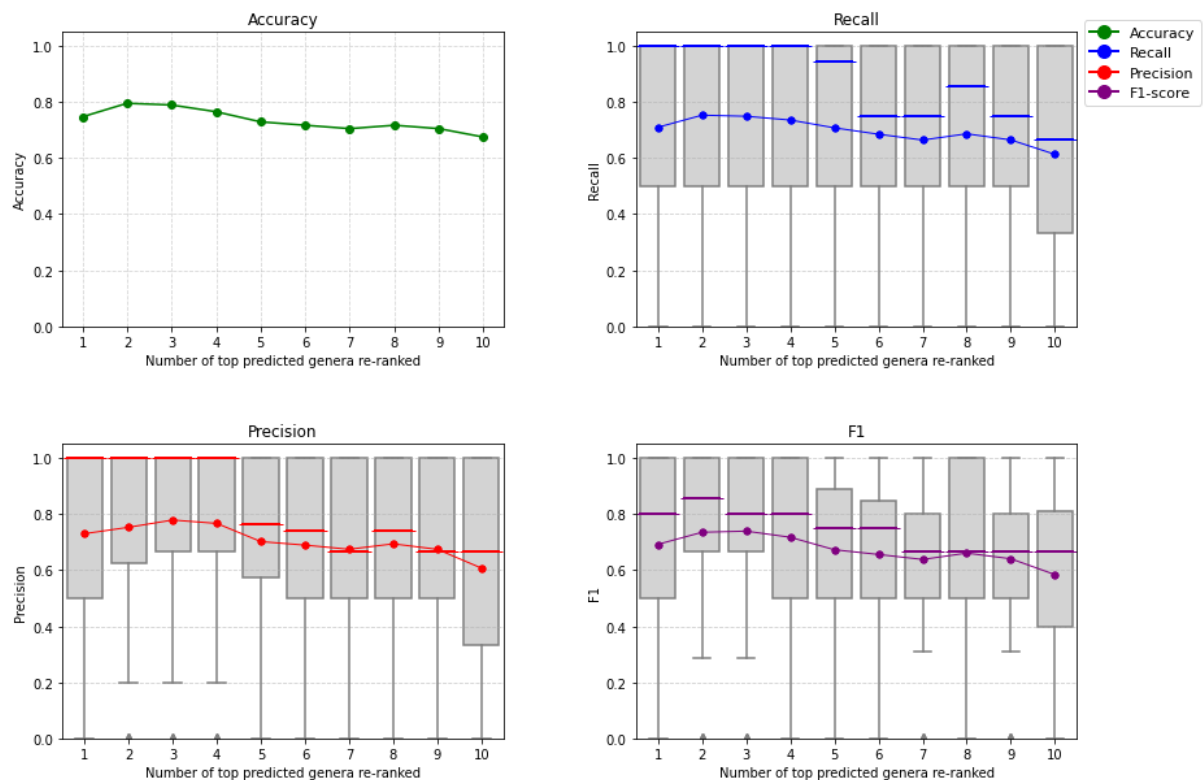


Figure 6.2: Performance metrics plotted against the number of top predicted genera included in the random forest re-ranking step using macroscopic wood anatomical features. Accuracy (green), recall (blue), precision (red), and F1-score (purple) are shown as macro-averaged values across all genera. For recall, precision, and F1-score, boxplots display the distribution of class-specific values at each level of re-ranking. The first point represents performance without re-ranking (CNN-only output), while subsequent points reflect increasing numbers of top predicted genera included in the re-ranking process.

Figure 6.2 demonstrates that refining CNN genus predictions with expert-defined anatomical features results in moderate but meaningful improvements in classification performance, particularly within the top-4 predicted genera. Across all considered top- k levels, the averaged metrics (points) remain relatively stable, with overall means and standard errors of mean (0.741 ± 0.012), Recall (0.700 ± 0.015), Precision (0.711 ± 0.018), and F1-score (0.677 ± 0.017), suggesting only minor changes in performance. The median values of the recall and precision remain 1 for re-ranking in the top four genera, suggesting that most genera are well classified regardless of re-ranking. In fact, including more top predicted classes allows median recall and median precision to decrease, indicating that most genera perform worse, likely due to misclassifications with the macroscopic features. The median values of the F1 score elucidates the discrepancy between genera with likely different genera having high precision and high recall. The highest median F1 score is observed when re-ranking the top 2 predicted classes.

The points (macro-averaged values) and interquartile ranges (boxes on the boxplots) provide further detail on re-ranking within the top 4 predicted genera by the CNN. All metrics experience an increase in average value by re-ranking the top predicted genera, with the highest values ranging in re-ranking within the top 4 predicted genera by the CNN. Accuracy peaks when the top two predicted genera are re-ranked using expert-defined anatomical features. Extending the re-ranking to the top three or four predictions results in only marginal changes, after which accuracy gradually decreases. Macro-averaged recall improves with re-ranking up to the top three genera. The stability of the interquartile range suggests that little variability is introduced in the recall by re-ranking. Precision follows a similar trend, with the highest values observed at ranks four and five. The interquartile range decreases most notably at rank four, from $Q1 = 0.78$ to $Q3 = 1.00$, indicating greater consistency and less genus-specific variability in precision at this point. A narrow $Q1$ – $Q3$ range in this context implies that the refinement process yields uniformly higher precision across most genera. F1-score, which balances recall and precision, is also highest for re-ranking up to the top four predictions. The median values of F1 are consistently lower than those of recall and precision, starting at 0.8 for the CNN output, peaking at 0.85 for re-ranking the top three predictions, and returning to 0.8 through rank four before declining after rank 5. This discrepancy suggests that individual genera can exhibit uneven gains in recall versus precision. The interquartile range for F1 decreases most at rank three/four, indicating a convergence of genus-specific F1 scores toward higher values and greater uniformity.

To evaluate the impact of re-ranking for each genus individually across different re-ranking depths, recall, precision, and F1-score were assessed and summarized in Table 6.1, Table 6.2, and Table 6.3. These tables use a Viridis colour gradient—ranging from yellow (high) to green (intermediate) to blue (low)—to facilitate visual interpretation of performance changes from no re-ranking (column 1) to re-ranking the top ten predicted classes. Across these tables, five distinct patterns emerge. The first represents a clear improvement in metric values due to re-ranking, where the values in column 1 are lower than in subsequent columns. The second pattern is a decrease in performance, where values decrease relative to the initial CNN output. The third pattern indicates no substantial change across all re-ranking depths. The fourth captures genera that show a temporary increase in performance, which then returns to the original (non-re-ranked) value. Conversely, the fifth scenario involves an initial decrease followed by a return to the original value. Fourteen genera increase (pattern 1) and sixteen decrease (pattern 2) in recall with re-ranking, fourteen increase and twenty-two decrease in precision, and sixteen increase and twenty-five decrease in F1-score. Most remaining genera fall into pattern three, with little to no change observed: 18 for recall, 10 for precision, and 7 for F1-score. Only a small number of genera belong to patterns four and five. For recall, four genera increase and stabilize, and four other genera show a temporary decrease; for precision, three genera increase then return to baseline, while seven show the opposite; and for F1-score, five genera increase then return to baseline, while three show the opposite.

Importantly, the effect of re-ranking is not consistent across genera. Certain genera, such as *Antiaris*, *Berlinia*, *Copaifera*, *Gilbertiodendron*, *Khaya*, *Leplaea*, *Milicia*, *Terminalia*, and *Zanthoxylum*, show a decrease in both precision and recall following re-ranking. In contrast, genera including *Aningeria*, *Austranella*, *Baillonella*, *Canarium*, *Celtis*, *Klainedoxa*, *Pachylobus*, *Staudtia*, and *Tessmannia* show concurrent improvements in both metrics. Some genera, such as *Alstonia*, *Morus*, *Ricinodendron*, and *Triplochiton* were already classified correctly by the CNN and thus remained unaffected by re-ranking. Others, like *Antrocaryon* and *Beilschmiedia*, were consistently misclassified and did not benefit from refinement. In several cases, re-ranking decreased or increased recall and increased or decreased precision respectively. For example, genera such as *Entandrophragma*, *Guibourtia*, *Irvingia*, and *Turraeanthus* show an increase in one metric and a decrease in the other. Additional genera showed stability in one metric with either a decrease or increase in the other, further illustrating the lack of a uniform effect of re-ranking. Despite general improvements when re-ranking is limited to the top four predictions, the results clearly demonstrate that refinement using macroscopic anatomical features does not benefit all genera equally. Approximately 30% of the genera suffer notable decreases in recall and 40% in precision due to re-ranking. Genera such as *Antiaris*, *Gilbertiodendron*, *Zanthoxylum*, *Nauclea*, *Khaya*, and *Irvingia* consistently decrease in recall across all re-ranking depths. For *Guibourtia*, *Lovoa*, and *Terminalia*, recall performance decreases notably when re-ranking beyond the top four CNN predictions.

Table 6.1: Genus-specific recall values across different numbers of top predicted genera included in the re-ranking step using macroscopic wood anatomical features. Column "1" represents recall values from the original binary verification model without re-ranking (CNN-only output). Columns "2" through "10" represent increasing numbers of top predicted genera (e.g., top 2, top 3, etc.) included in the re-ranking step. Recall values range from 0 to 1, with a colour gradient applied to highlight trends (yellow = high recall, green = intermediate, blue = low recall).

Genus	Number of top predicted genera reshuffled									
	1	2	3	4	5	6	7	8	9	10
<i>Afzelia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Albizia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Alstonia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Amphimas</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Aningeria</i>	0,00	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,00	0,00
<i>Antiaris</i>	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Antrocaryon</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Autranella</i>	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
<i>Baillonella</i>	0,00	1,00	0,00	0,50	1,00	1,00	1,00	1,00	1,00	1,00
<i>Beilschmiedia</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Berlinia</i>	1,00	0,33	1,00	1,00	1,00	1,00	0,67	0,33	0,33	0,33
<i>Bobgunnia</i>	1,00	0,50	1,00	0,50	0,50	1,00	0,50	0,50	1,00	0,50
<i>Brachystegia</i>	0,50	0,50	1,00	1,00	0,50	0,50	0,50	0,50	0,50	0,50
<i>Canarium</i>	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Celtis</i>	0,00	0,33	1,00	0,33	0,33	0,33	0,33	0,67	0,33	0,67
<i>Copaifera</i>	1,00	0,50	1,00	0,50	0,50	0,50	0,50	0,50	0,50	0,50
<i>Cynometra</i>	0,67	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Daniellia</i>	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	0,00
<i>Diospyros</i>	0,50	0,00	0,50	0,50	0,00	0,00	0,00	0,00	0,50	0,50
<i>Entandrophragma</i>	0,82	0,94	0,82	0,88	0,88	0,82	0,88	0,88	0,88	0,82
<i>Erythrophleum</i>	0,33	0,00	0,67	0,67	0,33	0,00	0,00	0,00	0,00	0,00
<i>Gambeya</i>	0,33	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,33
<i>Gilbertiodendron</i>	1,00	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
<i>Guibourtia</i>	1,00	0,75	1,00	1,00	0,75	0,75	0,75	0,75	0,75	0,75
<i>Holoptelea</i>	0,50	0,50	1,00	1,00	1,00	1,00	0,50	0,50	0,50	0,50
<i>Irvingia</i>	0,67	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33
<i>Julbernardia</i>	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Khaya</i>	0,83	0,33	0,50	0,50	0,50	0,50	0,33	0,33	0,33	0,33
<i>Klainedoxa</i>	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Lepalea</i>	1,00	0,89	1,00	0,89	0,89	0,89	0,89	0,89	0,89	0,89
<i>Lophira</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Lovoa</i>	1,00	0,67	1,00	1,00	0,83	0,67	0,83	0,67	0,83	0,83
<i>Maesopsis</i>	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Milicia</i>	1,00	0,67	1,00	1,00	1,00	0,67	0,67	0,67	0,67	0,67
<i>Millettia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Mitragyna</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Morus</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Nauclea</i>	1,00	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
<i>Nesogordonia</i>	0,50	0,75	0,75	0,50	0,75	0,75	0,75	0,75	0,50	0,75
<i>Ongokea</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Pachylobus</i>	0,00	1,00	0,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Pericopsis</i>	1,00	0,50	1,00	1,00	1,00	1,00	1,00	0,50	0,50	0,50
<i>Petersianthus</i>	1,00	0,00	0,50	0,50	0,50	0,00	0,00	0,50	0,50	0,50
<i>Piptadeniastrum</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Prioria</i>	1,00	0,50	0,75	0,75	0,75	0,75	0,75	0,75	1,00	0,75
<i>Pterocarpus</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Pterygota</i>	1,00	0,33	1,00	0,67	0,33	0,33	0,33	0,33	0,33	0,33
<i>Pycnanthus</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Ricinodendron</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Staudtia</i>	0,00	0,00	0,00	1,00	1,00	0,50	0,50	0,50	0,50	0,50
<i>Terminalia</i>	0,67	0,00	0,67	0,67	0,33	0,33	0,33	0,33	0,00	0,00
<i>Tessmannia</i>	0,50	0,75	1,00	1,00	1,00	0,50	0,75	0,75	1,00	1,00
<i>Tieghemella</i>	0,25	0,75	0,50	0,75	1,00	1,00	0,75	0,75	0,75	0,75
<i>Triplochiton</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Turraeanthus</i>	0,33	1,00	1,00	1,00	0,67	1,00	1,00	1,00	1,00	1,00
<i>Zanthoxylum</i>	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Table 6.2: Genus-specific precision values across different numbers of top predicted genera included in the re-ranking step using macroscopic wood anatomical features. Column "1" represents precision values from the original binary verification model without re-ranking (CNN-only output). Columns "2" through "10" represent increasing numbers of top predicted genera (e.g., top 2, top 3, etc.) included in the re-ranking step. Precision values range from 0 to 1, with a colour gradient applied to highlight trends (yellow = high precision, green = intermediate, blue = low precision).

Genus	Number of top predicted genera reshuffled									
	1	2	3	4	5	6	7	8	9	10
<i>Afzelia</i>	1,00	0,67	1,00	1,00	1,00	0,75	0,67	0,67	0,67	0,67
<i>Albizia</i>	1,00	0,67	1,00	1,00	1,00	1,00	1,00	0,67	1,00	0,67
<i>Alstonia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Amphimas</i>	1,00	0,67	1,00	1,00	0,67	0,67	0,67	1,00	1,00	1,00
<i>Aningeria</i>	0,00	0,33	0,20	0,20	0,20	0,33	0,33	0,33	0,00	0,00
<i>Antiaris</i>	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Antrocaryon</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Autranella</i>	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
<i>Baillonella</i>	0,00	1,00	0,00	1,00	1,00	0,67	1,00	1,00	1,00	1,00
<i>Beilschmiedia</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Berlinia</i>	1,00	0,50	0,75	0,75	0,50	0,50	0,50	0,50	0,50	0,50
<i>Bobgunnia</i>	0,67	0,33	1,00	1,00	1,00	1,00	1,00	1,00	0,67	0,50
<i>Brachystegia</i>	0,50	1,00	0,50	1,00	1,00	1,00	1,00	0,50	0,50	0,50
<i>Canarium</i>	0,50	0,67	0,67	0,67	0,67	0,67	0,67	0,50	0,67	0,67
<i>Celtis</i>	0,00	0,50	1,00	1,00	1,00	1,00	1,00	0,67	0,50	0,67
<i>Copaifera</i>	1,00	0,33	0,67	0,50	0,50	0,33	0,33	0,50	0,50	0,33
<i>Cynometra</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Daniellia</i>	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,00	0,33	0,00
<i>Diospyros</i>	0,20	0,00	0,50	0,50	0,00	0,00	0,00	0,00	0,50	1,00
<i>Entandrophragma</i>	0,93	0,80	0,82	0,79	0,79	0,78	0,75	0,75	0,75	0,78
<i>Erythrophleum</i>	1,00	0,00	1,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00
<i>Gambeya</i>	1,00	1,00	0,00	0,50	1,00	1,00	1,00	1,00	1,00	1,00
<i>Gilbertiodendron</i>	1,00	0,50	1,00	1,00	1,00	1,00	0,50	0,50	0,50	0,50
<i>Guibourtia</i>	0,67	0,75	1,00	1,00	1,00	0,75	0,75	0,75	0,75	0,75
<i>Holoptelea</i>	0,20	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Irvingia</i>	0,67	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Julbernardia</i>	0,33	0,00	1,00	0,50	0,50	0,50	0,50	0,50	1,00	0,50
<i>Khaya</i>	0,71	0,29	0,50	0,50	0,50	0,38	0,29	0,29	0,29	0,29
<i>Klainedoxa</i>	0,50	1,00	0,50	0,67	0,67	0,67	0,67	0,67	1,00	0,67
<i>Lepalea</i>	1,00	0,80	1,00	0,89	0,80	0,80	0,80	0,80	0,89	0,73
<i>Lophira</i>	1,00	1,00	1,00	0,50	0,67	0,67	0,67	0,67	0,50	0,67
<i>Lovoa</i>	1,00	0,67	1,00	1,00	1,00	0,80	1,00	0,80	1,00	1,00
<i>Maesopsis</i>	1,00	1,00	1,00	1,00	1,00	0,67	1,00	1,00	1,00	1,00
<i>Milicia</i>	1,00	0,50	1,00	1,00	0,75	0,67	0,67	0,67	0,67	0,67
<i>Millettia</i>	1,00	0,67	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Mitragyna</i>	1,00	0,80	1,00	0,75	0,75	0,80	0,80	0,80	0,50	0,80
<i>Morus</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Nauclea</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Nesogordonia</i>	1,00	0,50	1,00	1,00	0,75	0,75	0,75	0,75	0,67	0,50
<i>Ongokea</i>	1,00	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
<i>Pachylobus</i>	0,00	1,00	0,00	0,00	0,50	1,00	1,00	0,50	1,00	1,00
<i>Pericopsis</i>	1,00	0,50	0,67	1,00	1,00	0,67	0,67	1,00	1,00	1,00
<i>Petersianthus</i>	1,00	0,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00	1,00
<i>Piptadeniastrum</i>	0,50	0,67	1,00	1,00	1,00	1,00	0,67	1,00	0,67	0,67
<i>Prioria</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Pterocarpus</i>	1,00	0,67	1,00	0,80	0,73	0,80	0,73	0,67	0,73	0,67
<i>Pterygota</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Pycnanthus</i>	1,00	0,50	1,00	1,00	1,00	1,00	0,50	0,50	0,50	0,50
<i>Ricinodendron</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Staudtia</i>	0,00	0,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Terminalia</i>	0,67	0,00	0,50	0,50	0,33	0,33	0,25	0,50	0,00	0,00
<i>Tessmannia</i>	0,67	1,00	0,80	0,67	1,00	1,00	1,00	1,00	1,00	1,00
<i>Tieghemella</i>	1,00	1,00	1,00	0,75	1,00	0,80	0,75	0,80	0,75	0,75
<i>Triplochiton</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Turraeanthus</i>	1,00	0,50	0,75	0,75	0,33	0,43	0,80	0,80	0,80	0,80
<i>Zanthoxylum</i>	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Table 6.3: Genus-specific F1-scores across different numbers of top predicted genera included in the re-ranking step using macroscopic wood anatomical features. Column "1" represents F1-scores from the original binary verification model without re-ranking (CNN-only output). Columns "2" through "10" represent increasing numbers of top predicted genera (e.g., top 2, top 3, etc.) included in the re-ranking step. F1-scores range from 0 to 1, with a colour gradient applied to highlight trends (yellow = high F1-score, green = intermediate, blue = low F1-score).

Genus	Number of top predicted genera reshuffled									
	1	2	3	4	5	6	7	8	9	10
<i>Afzelia</i>	1,00	0,80	1,00	1,00	1,00	0,86	0,80	0,80	0,75	0,80
<i>Albizia</i>	1,00	0,80	1,00	1,00	1,00	1,00	1,00	0,80	1,00	0,80
<i>Alstonia</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Amphimas</i>	1,00	0,80	1,00	1,00	0,80	0,80	0,80	1,00	1,00	1,00
<i>Aningeria</i>	0,00	0,40	0,29	0,29	0,29	0,40	0,40	0,40	0,00	0,00
<i>Antiaris</i>	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Antrocaryon</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Autranella</i>	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
<i>Baillonella</i>	0,00	1,00	0,00	0,67	1,00	0,80	1,00	1,00	1,00	1,00
<i>Beilschmiedia</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>Berlinia</i>	1,00	0,40	0,86	0,86	0,67	0,67	0,67	0,40	0,40	0,40
<i>Bobgunnia</i>	0,80	0,40	1,00	0,67	0,67	1,00	0,67	0,67	0,80	0,50
<i>Brachystegia</i>	0,50	0,67	0,67	1,00	0,67	0,67	0,67	0,50	0,50	0,50
<i>Canarium</i>	0,50	0,80	0,80	0,80	0,80	0,80	0,80	0,67	0,80	0,80
<i>Celtis</i>	0,00	0,40	1,00	0,50	0,50	0,50	0,50	0,67	0,40	0,67
<i>Copaifera</i>	1,00	0,40	0,80	0,50	0,50	0,40	0,40	0,50	0,50	0,40
<i>Cynometra</i>	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Daniellia</i>	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,00	0,50	0,00
<i>Diospyros</i>	0,29	0,00	0,50	0,50	0,00	0,00	0,00	0,00	0,50	0,67
<i>Entandrophragma</i>	0,87	0,86	0,82	0,83	0,83	0,80	0,81	0,81	0,81	0,80
<i>Erythrophleum</i>	0,50	0,00	0,80	0,80	0,50	0,00	0,00	0,00	0,00	0,00
<i>Gambeya</i>	0,50	0,50	0,00	0,40	0,50	0,50	0,50	0,50	0,50	0,50
<i>Gilbertiodendron</i>	1,00	0,50	0,67	0,67	0,67	0,67	0,50	0,50	0,50	0,50
<i>Guibourtia</i>	0,80	0,75	1,00	1,00	0,86	0,75	0,75	0,75	0,75	0,75
<i>Holoptelea</i>	0,29	0,67	1,00	1,00	1,00	1,00	0,67	0,67	0,67	0,67
<i>Irvingia</i>	0,67	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
<i>Julbernardia</i>	0,50	0,00	1,00	0,67	0,67	0,67	0,67	0,67	1,00	0,67
<i>Khaya</i>	0,77	0,31	0,50	0,50	0,50	0,43	0,31	0,31	0,31	0,31
<i>Klainedoxa</i>	0,50	1,00	0,67	0,80	0,80	0,80	0,80	0,80	1,00	0,80
<i>Leplaea</i>	1,00	0,84	1,00	0,89	0,84	0,84	0,84	0,84	0,89	0,80
<i>Lophira</i>	1,00	1,00	1,00	0,67	0,80	0,80	0,80	0,80	0,67	0,80
<i>Lovoa</i>	1,00	0,67	1,00	1,00	0,91	0,73	0,91	0,73	0,91	0,91
<i>Maesopsis</i>	0,67	1,00	1,00	1,00	1,00	0,80	1,00	1,00	1,00	1,00
<i>Milicia</i>	1,00	0,67	1,00	1,00	0,86	0,67	0,67	0,67	0,67	0,67
<i>Millettia</i>	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Mitragyna</i>	1,00	0,75	1,00	0,86	0,86	0,75	0,75	0,75	0,67	0,75
<i>Morus</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Nauclea</i>	1,00	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
<i>Nesogordonia</i>	0,67	0,67	0,86	0,67	0,75	0,75	0,75	0,75	0,67	0,67
<i>Ongokea</i>	1,00	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80
<i>Pachylobus</i>	0,00	1,00	0,00	0,00	0,67	1,00	1,00	0,67	1,00	1,00
<i>Pericopsis</i>	1,00	0,50	0,80	1,00	1,00	0,80	0,80	0,67	0,67	0,67
<i>Petersianthus</i>	1,00	0,00	0,67	0,67	0,67	0,00	0,00	0,67	0,67	0,67
<i>Piptadeniastrum</i>	0,67	0,80	1,00	1,00	1,00	1,00	0,80	1,00	0,80	0,80
<i>Prioria</i>	1,00	0,67	0,86	0,86	0,86	0,86	0,86	0,86	1,00	0,86
<i>Pterocarpus</i>	1,00	0,80	1,00	0,89	0,84	0,89	0,84	0,80	0,84	0,80
<i>Pterygota</i>	1,00	0,50	1,00	0,80	0,50	0,50	0,50	0,50	0,50	0,50
<i>Pycnanthus</i>	1,00	0,67	1,00	1,00	1,00	1,00	0,67	0,67	0,67	0,67
<i>Ricinodendron</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Staudtia</i>	0,00	0,00	0,00	1,00	1,00	0,67	0,67	0,67	0,67	0,67
<i>Terminalia</i>	0,67	0,00	0,57	0,57	0,33	0,33	0,29	0,40	0,00	0,00
<i>Tessmannia</i>	0,57	0,86	0,89	0,80	1,00	0,67	0,86	0,86	1,00	1,00
<i>Tieghemella</i>	0,40	0,86	0,67	0,75	1,00	0,89	0,75	0,67	0,75	0,75
<i>Triplochiton</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
<i>Turraeanthus</i>	0,50	0,67	0,86	0,86	0,44	0,60	0,75	0,75	0,75	0,75
<i>Zanthoxylum</i>	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

6.5 Discussion

Previous research has emphasized the value of combining multiple techniques to enhance wood identification. Dormontt et al. (2015) proposed that integrating several methods is essential for improving diagnostic accuracy and narrowing the taxonomic resolution in forensic wood investigations (Dormontt et al., 2015). Research by Knauer et al. (2019), also demonstrated that refining CNN predictions with RF classifiers (of different size and operating in different feature spaces) for tree species identification using hyperspectral image information can significantly improve tree species identification (Knauer et al., 2019). These studies all indicate that leveraging diverse diagnostic features—such as microscopic anatomical characteristics, wood chemical composition, or DNA—can offer complementary information, thereby improving both the accuracy and the taxonomic specificity of wood identification systems.

The results of our approach on the accessible macroscopic cross-section, which combines CV with expert-defined anatomical features, highlight the nuanced effects of re-ranking CNN genus predictions using expert-defined macroscopic wood anatomical features on the endgrain. While the integration of expert-defined features through RF models can lead to modest improvements in performance (accuracy, recall, and precision), the benefits are not uniform across all genera. Median values of recall and precision remain stable up to the top 4 re-ranked predictions, indicating that most genera are well classified by the CNN output alone. This is underscored by the analysis on individual genera, with no change in recall observed for 18 out of the 56 genera and no change in precision observed for 10 out of 56 genera. This suggests that the CNN approach already captures valuable taxonomic signals from the images, and in many cases, re-ranking may offer limited or no added benefit.

This limited improvement in performance following re-ranking may be attributed to the hard mining strategy employed during training of the binary verification CNN (for additional details see section 5.3.7). In object re-identification tasks—such as binary verification—model robustness is enhanced by selectively training on the most challenging specimen pairs rather than on easy (highly dissimilar) examples (Ghosh et al., 2023; Hermans et al., 2017). For this model, hard negatives were systematically generated based on the same expert-defined macroscopic anatomical features, prioritizing specimen pairs that shared similar standardized anatomical patterns in endgrain sections. As a result, the binary verification CNN was implicitly encouraged to learn discriminative features that extend beyond those defined by traditional anatomical descriptors. This training strategy likely enabled the CNN to extract subtle diagnostic signals from the wood images that are not easily captured by predefined macroscopic features alone, thereby explaining the limited gains observed through subsequent re-ranking using RF models.

Nonetheless, improvements can be observed when re-ranking is applied judiciously. Across various metrics, re-ranking within the top two to four CNN predictions resulted in increased performance (averaged across genera). Specifically, the top 3 yields the most consistent gains across accuracy, recall, precision, and F1-score, with the narrowest interquartile range in precision and F1-score occurring at this number of genera. Beyond the top four predictions, classification performance begins to decrease, indicating that incorporating macroscopic features for lower-ranked genera may introduce noise. This is likely due to the RF model assigning proportionally larger probabilities to incorrect genera suggesting that re-ranking with lower-ranked genera introduces noise, and thereby misguiding the re-ranking process. This underscores the limitations of macroscopic wood anatomical features observed in chapter 3 and provides perspective on the limited range where these anatomical features can provide compelling value at refining CNN-based predictions.

The analysis on individual genera further supports this interpretation. While recall improvement is observed in 25% of the genera (for example for *Aningeria*, *Canarium*, and *Tessmannia*) across all three metrics (recall, precision, and F1-score) due to re-ranking, 29% of genera shows a decrease. This divergence highlights that re-ranking with expert-defined anatomical features is not uniformly advantageous and may be taxon-dependent. A key explanation could lie in differences in field-of-view. Whereas the CNN operates on fixed-size cropped patches, the expert-defined features were annotated across entire scans, reflecting a broader context. Certain diagnostic traits—such as banded or confluent

parenchyma, or irregular vessel groupings—are not uniformly distributed across the wood surface and may only become apparent at larger scales or through multi-scale assessment. Consequently, the CNN may miss features that experts can reliably recognize by zooming in and out or scanning across surfaces, enhancing the synergy between the two approaches. This underscores the importance of future research on how FOV, image resolution, and multi-scale anatomical integration affect CV-based wood identification. Developing strategies that explicitly incorporate scale-adaptive representations could help bridge the gap between patch-based cv-based models and anatomical assessment.

The diverging results on the individual genera also underscores that re-ranking is not a universally beneficial strategy, and application in forensic cases should be targeted at specific cases (genera) and not be performed blindly. This awareness raises the difficulty of applying such approaches in the field. Moreover, the results show that recall and precision do not always move in the same direction. Several genera experience increased precision but reduced recall, or vice versa. This asymmetry carries important implications in the context of timber verification for detecting illegal logging of protected timbers. From a regulatory screening standpoint, a decrease in recall is more problematic than in precision. Failing to identify a true positive (e.g. a protected timber species) is a more critical error than generating a false positive. This trade-off is particularly evident for high-priority genera such as *Khaya*, listed under CITES Appendix II (CITES, 2022c; UNEP-WCMC (Comps.), 2022). Applying the re-ranking approach on specimens of threatened timber genera resulted in an immediate and consistent decrease in recall. This outcome poses a tangible risk in real-world applications, as it may lead to under-detection of protected species, potentially facilitating illegal trade. Therefore, caution must be exercised when applying refinement methods that rely on macroscopic cross-sectional features. Additional diagnostic information, such as microscopic anatomical features or visual assessment using visual keys (Leggett and Kirchoff, 2011), may provide the necessary detail to improve identification accuracy without compromising recall. Incorporating such features, albeit easy to observe in field conditions (e.g. warehouse, forest, roadside), could help capture taxonomically relevant traits, enabling more reliable refinement of CNN predictions in forensic timber identification in field conditions.

An additional factor influencing the effectiveness of re-ranking could be the presence of taxonomic look-alikes. Many timber genera share macroscopic in cross-sectional features as evident from the observed clusters in section 3.1.6 such as similar vessel arrangements or parenchyma patterns. When such genera co-occur in the top k CNN predictions, re-ranking has the potential to refine the ordering by leveraging complementary anatomical descriptors. However, in cases where the CNN has already implicitly learned to discriminate these subtle traits—partly due to the hard negative mining strategy employed during training—the added value of re-ranking diminishes.

The critical challenge then lies in defining what constitutes a look-alike. Both CV-based extracted features and expert-defined features are derived from the same underlying visual information, but the effective field-of-view may shift the balance. A CNN restricted to cropped patches may fail to capture diagnostic traits visible only at larger scales, while expert annotations operate across the full cross-section, potentially recognizing features outside the CNN's receptive field. This raises the question of whether look-alike relationships should be determined primarily by CNN similarity metrics, anatomical descriptors, or a hybrid strategy. Multi-scale approaches offer a promising way forward, as they could align the CNN's representational scope more closely with expert assessment, enabling more robust differentiation of visually similar genera and a better understanding of when re-ranking adds diagnostic value.

6.6 Conclusion

This study demonstrates that while integrating expert-defined macroscopic anatomical features can yield moderate improvements in CNN-based genus predictions, these benefits are highly dependent on both the specific genera and the depth of re-ranking applied. The overall results affirm that CNN models alone already encode substantial taxonomic information, likely due to their training on challenging diagnostic comparisons that extend beyond traditional anatomical descriptors. Crucially, re-ranking within the top two to five CNN predictions offers the most consistent performance gains across accuracy, precision, and recall—especially at the top three threshold. Beyond this range, performance diminishes due to misclassifications introduced by overemphasizing weak or misleading anatomical features. This finding underscores the limited but strategic utility of macroscopic cross-sectional wood anatomy for refining identifications. From a field application perspective, particularly in the context of frontline timber verification, the implications are twofold. First, the CNN model offers a rapid and accessible method for genus-level identification that already performs well in most cases. Second, refinement methods such as re-ranking must be applied selectively, as indiscriminate use—especially on protected taxa—can reduce recall, increasing the risk of overlooking high-priority timbers such as *Khaya*. Future research on field implementations should aim for using different diagnostic information that provides complementary value for refining wood identifications.

6.7 Acknowledgements

This work was realized through the combined efforts of the staff at the service of Wood biology at the RMCA in Tervuren, Belgium, and the staff at the UGent-Woodlab (Department of Environment, Faculty of Bioscience Engineering) of Ghent University.

6.8 Availability of code and metadata

Scripts and metadata can be requested by contacting the first author:

Contact: ruben.de.blaere@africamuseum.be

7 General discussion

7.1 Evaluating expert-defined macroscopic cross-sectional features for identification of Congolese tree species

The need for rapid and accessible wood identification remains pressing, particularly in the context of regulating timber trade and enforcing regulations to halt illegal logging (Gasson et al., 2021; Hirschberger, 2008; Hoare, 2015; Magrath et al., 2009; Piabuo et al., 2021; Van Brusselen et al., 2023). Given the volume of timber shipments and the limited movement space often encountered in warehouses and containers, speed and practicality are essential (Dormontt et al., 2015; Tacconi, 2012; Thompson and Magrath, 2021). Therefore, field officers currently rely primarily on manual macroscopic assessment using simple, low-cost tools such as a sharp cutter, a 10× or 15× hand lens, and a field guide (Koch et al., 2018; Wheeler and Baas, 1998). Typically, cross-sectional wood anatomy is exposed, and the specimen is compared against standardized anatomical descriptors or reference images of expected timber species (Leggett and Kirchoff, 2011; Tardif and Conciatori, 2015; Wheeler and Baas, 1998). The objective of this study was to assess the feasibility of authenticating the botanical identity of timbers in species-rich contexts, where many morphologically similar taxa coexist, using the most accessible diagnostic information in the field: macroscopic cross-sectional wood anatomy. Chapter 3 assessed the diagnostic potential of expert-defined macroscopic features using the SmartWoodID dataset (hypothesis 2). While practical for field use, these features had not been systematically evaluated. However, as demonstrated in Chapter 3, the discriminatory power of those accessible and frequently used expert-defined anatomical features is limited when applied across the diverse range of timbers in the DRC. Two-way clustering of 601 species suggested an optimal partitioning into six groups based on Mantel test (Borcard et al., 2011), indicating a broad anatomical overlap. Predictive modelling (classification) using these features achieved approximately 50% accuracy at the genus level when restricted to 56 Congolese commercial genera. Chapter 5 underscored these findings, showing that macroscopic cross-sectional features also yielded low recall and precision, approximately 50% for commercial genera and around 30% when considering all Congolese species (hypothesis 4.1). Moreover, rank-based accuracy analysis confirmed the limited discriminatory capacity of macroscopic features: to achieve a 95% probability of including the correct genus among the top predictions, one would have to consider 36 out of 56 commercial genera. These results indicate that, in species-rich contexts such as the DRC, macroscopic anatomical assessment of the cross-section alone is insufficiently reliable for wood identification across a broad range of timbers. This emphasizes the importance of incorporating more anatomical features available across all three principal anatomical planes—cross-section, radial, and tangential—and at multiple levels of magnification, to improve diagnostic accuracy and enable finer taxonomic resolution in wood identification workflows.

The low performance of classification based on expert-defined features in this study is likely due to the limited number of anatomical features observable on the cross-section at macroscopic resolution. A full anatomical assessment incorporates up to 163 anatomical features observed on all three sections of (cross-section, radial section, tangential section) (Gasson, 2011; NS, 1989; Wheeler, 2011) rather than the 31 cross-sectional features used in this study. These microscopic features remain to this day the standard method for taxonomic identification (Koch et al., 2015; Richter and Dallwitz, 2000; Wheeler, 2011). As such, it is essential to study additional microscopic features for identifying Congolese timbers through wood anatomy. The limited accessibility to these features currently poses challenges regarding the feasibility in field conditions, underscoring the need for further research developing methods for rapid visualization.

Low performance is also influenced by the current interpretation of the IAWA framework (Gasson et al., 2011; NS, 1989). Categorical states—such as “present,” “variable,” or “absent”—fail to capture the full spectrum of quantitative variation. For example, vessel diameter is typically recorded in broad ranges, obscuring important differences within those classes (small (<80 µm), medium (80–130 µm) and large (>130 µm)) (NS, 1989). An illustrative example is the genus *Pterocarpus*, specifically the challenge of

distinguishing *P. tinctorius* from *P. soyauxii* (see chapter 3: Figure 3.17). *P. soyauxii* tends to exhibit predominantly large vessels (small (<80 µm) is absent, medium (80–130 µm) is variable and large (>130 µm) is present), whereas *P. tinctorius* shows a predominance of medium-sized vessels (small (<80 µm) is absent, medium (80–130 µm) is present and large (>130 µm) is variable). This scrutinized form of coding highlights the limitations of the conventional binning strategies, which can mask subtle yet diagnostically useful anatomical patterns. Improving taxonomic resolution may thus depend on extracting finer-grained data, such as mean and standard deviation of vessel sizes or ray width. This also underscores that higher resolution is crucial in that regard as it enables more precise measurements. However, acquiring such data requires digitization of wood surfaces at high resolution, along with algorithms capable of automated feature extraction. These automated procedures are becoming more accessible as time progresses with recent advances showing opportunity for automating digitization at high resolution (Van den Bulcke et al., 2025). Advances in wood anatomy digitization—through high-resolution scanning, robotics, and deep learning—offer promising avenues for addressing these limitations (Van den Bulcke et al., 2025). Systems operating at ~2.25 µm resolution can now automatically segment and measure thousands of vessels, rays, and parenchyma structures (Van den Bulcke et al., 2025). While these techniques are impractical under field conditions, they hold tremendous potential for curating high-quality reference databases and developing downstream classification tools that outperform conventional feature-based methods (Van den Bulcke et al., 2025).

7.2 The importance of colour and subtle patterns for identification

In field settings, expert-defined anatomical assessments are frequently complemented by expert visual comparison, where law enforcement officers evaluate the macroscopic cross-sectional appearance of wood specimens against reference images (Leggett and Kirchoff, 2011). This step allows the inclusion of colour and texture, beyond the scope of the expert-defined anatomical features (Committee, 2004; Gasson, 2011; NS, 1989; Wheeler, 2011). Unlike the codified expert-defined features, these visual estimations are based on tacit experience and offer potentially valuable, albeit subjective, discriminatory patterns (Kirchoff et al., 2011). Although this informal visual component was not explicitly assessed in the present study, it remains a plausible contributor to improved identification accuracy and warrants further empirical investigation. This was studied in Chapter 5 by training and evaluating CV models that can learn directly from these subtle, non-standardized patterns (hypothesis 4.1). This chapter explored traditional ‘closed-set’ multiclass classification and ‘open-set’ recognition techniques for identification of the same Congolese tree species studied in chapter 3. The performance of CV-based models in chapter 5 indirectly highlights the diagnostic value of these subtle, non-standardized patterns. By training CNNs directly on macroscopic colour images of cross-sections, the models preserved key visual information—achieving approximately 0.85 in precision, recall, and accuracy for identifying commercial Congolese timber genera (Chapter 5). Rank-based evaluation further showed that the correct genus appeared within the top six predictions in over 95% of cases across 56 genera.

These results highlight that raw image data, retaining subtle differences in colour and texture, carries significantly more discriminatory information than expert-defined anatomical features alone (Knauer et al., 2019). This reinforces the notion that such subtle patterns—routinely leveraged by human experts—can be systematically captured through CV, making it a robust and scalable approach for field-based wood identification.

7.3 Refining CV-Based wood identification using expert-defined features

However, as previously noted, field-based wood identification draws on both standardized anatomical features and holistic visual impressions from full cross-sectional images. This highlights the importance of evaluating these two information sources both independently and in combination to determine whether their integration enhances identification accuracy. Chapter 5 demonstrated that CV-based techniques on macroscopic cross-sectional images do not provide full proof identification as the top 4-6 genera must be

considered among all 56 commercial genera in order not to exclude the correct answer in 95% of the test specimens. Chapter 3 demonstrated that expert-defined features can facilitate reliable differentiation between narrow ranges of taxa (such as within a single family or genus). For instance, within Moraceae, *Ficus* species show wide axial parenchyma bands and fewer rays per millimetre than genera like *Antiaris* and *Milicia*. Similarly, vessel porosity effectively distinguishes Lamiaceae genera: *Tectona* displays ring-porous wood, while *Vitex* and *Premna* show diffuse or semi-ring porosity. Another notable example are all African *Pterocarpus* species, which were amended to CITES Appendix II due to challenges in field identification (CITES, 2022a). Contrary to prior claims that macroscopic features were insufficient (Liu et al., 2023; Price et al., 2021), our study revealed clear differences between species in this genus. For example, *P. rotundifolius* and *P. angolensis* exhibit semi-ring porosity and moderate vessel frequency compared to *P. tinctorius* and *P. soyauxii*.

Chapter 6 attempted to address whether the expert-defined features provide complementary information for refining CV-based classifications or whether they provide little added value or can even result in misidentifications (hypothesis 5). This was evaluated by re-ranking the top k CNN predictions using tailored RF models trained on expert-defined features for distinguishing only the top k genera (with k ranging from 2 to 10). This showed that re-ranking in the top 4 predicted genera improves performance (accuracy, recall, precision) in general across genera, though the effect is small, the averaged metrics (points) remain relatively stable, with overall means and standard errors of mean (0.741 ± 0.012), Recall (0.700 ± 0.015), Precision (0.711 ± 0.018), and F1-score (0.677 ± 0.017), suggesting only minor changes in performance and additionally the added recall and precision is only for specific genera. Approximately 30% of the genera suffer notable decreases in recall and 40% in precision due to re-ranking. Genera such as *Antiaris*, *Gilbertiodendron*, *Zanthoxylum*, *Nauclea*, *Khaya*, and *Irvingia* consistently decrease in recall across all re-ranking depths. For *Guibourtia*, *Lovoa*, and *Terminalia*, recall performance decreases notably when re-ranking beyond the top four CNN predictions.

This underscores that while integrating expert-defined macroscopic anatomical features can yield moderate improvements in CNN-based genus predictions, these benefits are highly dependent on both the specific genera and the depth of re-ranking applied. The overall results affirm that CNN models alone already encode substantial taxonomic information, likely due to their training on challenging diagnostic comparisons that extend beyond traditional anatomical descriptors. This finding underscores the limited but strategic use of expert-defined macroscopic cross-sectional wood anatomy for refining identifications. Refinement methods such as re-ranking must be applied with caution, as blind use—especially on protected taxa—can reduce recall, increasing the risk of overlooking high-priority timbers such as *Khaya* (CITES, 2022c).

7.4 Reflecting on future research for wood identification

Collectively, our findings—on hypotheses 2, 4.1, and 5—highlight the importance of periodically and critically reassessing established methodologies in wood identification, particularly in light of evolving demands in forensic and regulatory contexts. Macroscopic cross-sectional wood anatomy remains a valuable and accessible diagnostic tool for rapid field-based screening. However, our results indicate that expert-defined features alone offer limited discriminatory power when applied to Congolese tree species. This suggests that further research is needed to systematically evaluate the diagnostic value of macroscopic features across a broader taxonomic spectrum, that represents all traded timber species. This effort should be coupled with parallel research into complementary techniques—improving their field accessibility and exploring opportunities for integrated identification strategies to refine taxonomic specificity. Furthermore, greater attention should be given to understanding how visual similarity—both in terms of intra- and interspecific variation—influences the reliability of visual comparisons with reference material. It is essential that regulatory agencies and enforcement bodies are aware of the intrinsic limitations of traditional approaches. This awareness is not to diminish the contributions of expert-based methodologies, but rather to foster a more adaptive, evidence-driven strategy that integrates conventional practices with complementary technologies to preserve the operational advantages of in-field expertise while bolstering accuracy and consistency.

In this context, CV -based approaches using CNNs have emerged as promising tools for large-scale, automated wood identification. Collectively, the findings of this study —on hypotheses 4.1, 4.2, 4.3— affirm the potential of CV as a viable strategy for enabling scalable and reliable wood identification in field conditions (Andrade et al., 2020; Hwang and Sugiyama, 2021; Ravindran et al., 2021, 2020, 2019; Silva et al., 2022; Tang et al., 2018; Wiedenhoef, 2020). The findings in chapter 4—on hypotheses 3—provide further key insight into the robustness of CV for wood identification. However, deploying CV-based models in operational contexts (e.g. border inspections, roadside checks, or timber warehouses) not only depends on quantified performance in controlled experiments. It is essential to consider the differences between both contexts when constructing databases and developing models (Ravindran and Wiedenhoef, 2022; Spiecker et al., 2000; Tardif and Conciatori, 2015). For database construction, these differences are related to factors such as specimen selection, processing methods, and image acquisition techniques.

7.4.1 Database construction

A critical factor in the development of reliable CV-based classifiers is the extent to which the training data reflect the intra-class anatomical variability that may be encountered during deployment (Halevy et al., 2009; Kala et al., 2022). This includes anatomical variation among specimens of the same species and within individual trees (e.g., across growth rings or along the trunk axis). The analysis in Chapter 3, which focused on predicting expert-defined features from wood anatomy, extends to the full images and sheds light on the importance of capturing the variability of the wood anatomy. The SmartWoodID dataset was constructed as a specimen-database with this in mind, using a minimum of four specimens per species and linking each anatomical description to specific, digitized specimens (De Blaere et al., 2023). In contrast, InsideWood aggregates species-level descriptions compiled from disparate sources (Wheeler, 2011; Wheeler et al., 2020)(Wheeler et al., 2020). This difference in design was reflected in model performance: models trained on SmartWoodID consistently achieved higher accuracy when evaluated on SmartWoodID specimen descriptions than when tested on generalized InsideWood descriptions. Further empirical support for the importance of specimen-based data comes from a small study published in the conference proceedings of the 26th IUFRO World congress (Stockholm, Sweden 2024), evaluating CV-based classification using an expanded dataset with an increased number of specimens per Congolese wood genus (De Blaere et al., 2024). Results showed that, using identical model architectures, training parameters, and image augmentation protocols as applied in this study, and increasing the number of digitized specimens per genus led to a performance increase from 0.85 (observed in Chapter 5) to 0.94. Because both studies used the same target genera and similar imaging protocols, these findings show that covering intra-species variability is a key driver for building robust CV-based wood identification models. This underscores the importance of collaboration efforts between institutions for building comprehensive reference databases.

Beyond the number of specimens, the anatomical surface area captured per specimen also influences model performance. Traditional anatomical descriptions often rely on thin sections covering only $\sim 1 \text{ mm}^2$, whereas SmartWoodID scans encompass a much larger area ($\sim 7 \text{ cm} \times 1\text{--}2 \text{ cm}$) (De Blaere et al., 2023), offering broader visual coverage of structural variation within a single piece of wood, potentially revealing diagnostically relevant features that might be absent from smaller samples. For instance, *Pterocarpus angolensis*, which in SmartWoodID has $\geq 90\%$ solitary vessels—a diagnostic trait within the genus—was not represented different in InsideWood (Wheeler, 2011). In addition, this study identified notable interspecific differences in vessel diameter distributions between *P. soyauxii* and *P. tinctorius*, despite both being described as “ $>200 \mu\text{m}$ ” in InsideWood (Wheeler, 2011). We note that this comparison is limited by the lack of information on how many specimens were used in the InsideWood descriptions; nonetheless, it underscores the importance of using multiple specimens to reliably capture interspecific variation in quantitative features. While these examples concern expert-defined features, they affect CV-based wood identification model’s ability to learn such distinctions from raw imagery indirectly. This emphasizes the need for further research into optimal thresholds for either the number of specimens or the anatomical surface area required for effective training. Identifying the point of diminishing returns in model accuracy with respect to increased anatomical representation could enable more efficient dataset construction strategies, reducing unnecessary labour while preserving classification performance (Szyk, 2020).

Variation in anatomical features across a pith to bark gradient, at different heights of a tree or in specific organs (branches) were not explicitly quantified in this study. However, these aspects could potentially be measured in subsequent studies and could help guide the field forward to improving database construction for robust model development.

A further critical consideration in developing robust CV-based wood identification models lies in how datasets are partitioned for training and evaluation of models. This is essential to ensure realistic performance assessment: when images from the same specimen appear in both training and test sets, models risk exploiting specimen-specific artefacts rather than learning generalizable anatomical signals (Ravindran and Wiedenhoef, 2022). Enforcing splits at the specimen level, where distinct specimens are used for training, validation, and testing, provides a stricter but more ecologically valid test of model robustness, reflecting real-world deployment where unknown specimens will be encountered.

Beyond the anatomy, wood can however also show different anomalies—such as cracks (Niemz et al., 2023a), insect damage (Goodell and Nielsen, 2023; Schmidt, 2006), and fungal deterioration (Goodell and Nielsen, 2023; Schmidt, 2006)—that are the result of the nature of wood being a natural material. Although this factor is rarely addressed in the CV wood identification literature, it is highly relevant for field applications, where perfect specimens are rarely available. This was investigated in the context of model performance in chapter 4, evaluating the impact of wood damage on CV-based identification using CNNs (hypothesis 3). This was studied by training a CNN architecture to classify 26 timber genera under three training conditions: (1) mixed sets containing both damage-free and damaged patches, (2) only damage-free patches, and (3) only damaged patches. Each model was evaluated on a balanced test set including both types of wood, and macro-average recall was used to assess overall and class-specific performance. Grad-CAM visualizations were applied to enhance interpretability by highlighting regions used for classification. Results show that models trained exclusively on damage-free image patches showed the highest recall (90.5%), followed closely by mixed-condition patches (88.4%) and, to a lesser extent, damaged-only patches (79.1%). Grad-CAM visualizations revealed that the CNN focused on intact anatomical regions and largely ignored areas affected by anomalies. Our research sheds light on this gap in literature affirming that intact wood anatomical surfaces provide the most important information for CV-based wood identification models. However, it also suggests that while damage-free specimens should be prioritized during database construction, the inclusion of imperfect but anatomically interpretable samples does not substantially compromise predictive performance—so long as diagnostically relevant features remain discernible.

The method of specimen preparation significantly influences the quality of anatomical information available for wood identification (Spiecker et al., 2000). In this study, all cross-sectional surfaces were prepared using a standardized sanding protocol involving a sequence of fine-grit abrasives to ensure smooth, high-quality surfaces optimized for observing the wood anatomy (De Blaere et al., 2023). While this method is ideal for generating high-quality reference images (Van den Bulcke et al., 2025), quality is more difficult to achieve in the field due to limitations in equipment, time, and operator expertise (Spiecker et al., 2000). Cutting with a sharp blade represents an alternative for field use, though it may introduce surface irregularities such as scratches or warp surfaces which in turn can affect magnification (Ravindran et al., 2023). Relevant work by Ravindran et al. (2023) demonstrated that CV models trained on high-quality sanded specimens maintained strong performance when evaluated on images of cut surfaces or on those sanded with intermediate grits (P240 or finer) (Ravindran et al., 2023). This suggests that reduced preparation quality does not impair accurate classification, which underscores the use of high-quality surface preparation for image database construction (Ravindran et al., 2023).

Image acquisition plays a central role in any CV-based model. Two stable parameters—field-of-view and resolution—directly affect diagnostic patterns on images (Chen and Guan, 2019; Gorodissky et al., 2018; Miyata et al., 2020). In CV-based classification workflows, it is standard practice to resize input images to fixed values, ensuring uniform input dimensions for model training (Chollet and Chollet, 2021; Talebi and Milanfar, 2021). However, this approach was not applied in the present study due to the significant variation in scan dimensions across specimens, which presents a concrete risk of losing diagnostically relevant

anatomical detail through downscaling. The largest cross-sectional scan in the dataset measured approximately 287.6 mm × 215.1 mm, while the smallest had a vertical dimension of 11 mm × 11 mm, highlighting the variability in specimen dimensions. General resizing across such a range would result in uneven loss of image fidelity and potentially compromise anatomical interpretability. To preserve image detail and maintain a consistent input structure for the CNN, all scans were instead cut into non-overlapping patches of equal size. Specifically, images were divided into 512 × 512-pixel patches, corresponding to a physical area of approximately 5.42 × 5.42 mm. This patch size was selected to align with the scale typically used in anatomical assessments of cross-sectional thin sections, ensuring sufficient context for recognizing wood anatomical features while enabling systematic analysis of anomalies, as discussed in Chapter 4. Importantly, overlap between patches was avoided to prevent data leakage between the training and validation sets (Ravindran and Wiedenhoef, 2022). Chapters 4 and 5 demonstrate that this patch-based approach supports effective CNN training, with the chosen field-of-view proving adequate to capture diagnostically relevant patterns for wood identification. Nevertheless, previous studies have shown that smaller patches can also yield competitive results. For instance, Ravindran et al. (2021) used patch sizes of 512 × 192 pixel, corresponding to 1.59 mm × 0.60 mm, and achieved 97% accuracy for a 24-class model to distinguish Peruvian timbers (Ravindran et al., 2021). This was achieved by cutting Xylotron images of 2048x2048 pixels (6.35x6.35mm) into those smaller patches. This suggests that field-deployed system like the Xylotron, which uses the full image for direct input into the neural network for wood identification, might not require that field-of-view. Furthermore, different resolutions can also yield high accuracy as the resolution of 2400 dpi in this study proved effective, while the study by Ravindran et al. 2021 used a resolution of 8192 dpi and other studies such as Ravindran et al. (2018) used a resolution of 4096 dpi and achieved 87.4% accuracy for classifying ten Neotropical Meliaceae timbers species (Ravindran et al., 2021, 2018).

Findings from Chapter 6 further underscore the importance of field-of-view in wood identification. While the overall benefit of re-ranking CV-based predictions was limited, certain genera were better identified when expert-defined features were incorporated. This could be attributed to the difference in scale at which features were extracted. Expert-defined features were assessed across the entire scan, with continuous zooming in and out to evaluate anatomical characteristics, whereas CNN-based features are derived from fixed patches of restricted size. Some diagnostic traits, such as banded or confluent parenchyma, may only become visible at larger spatial scales. Others may be rare, either due to natural variability along the pith-to-bark gradient or stem height, or because they are excluded during the patch-cutting process. Patch-based CNNs are effective in capturing local diagnostic signals and make computational processing tractable, but they inherently lack the broader spatial context that experts routinely exploit when scanning whole cross-sections. This highlights the need for multi-scale strategies in CV-based wood identification. Models capable of dynamically “zooming in and out” across resolutions would be better able to capture both localized and large-scale features, thereby narrowing the gap between automated approaches and expert practice. These complementary findings suggest that the optimal patch size and minimum effective resolution remains undetermined. Future research should therefore explore the relationship between field-of-view, anatomical feature representation, and model performance. Such investigations could help determine the smallest viable patch size that balances diagnostic fidelity with computational efficiency. Additionally, alternative strategies such as multi-scale or zoom-adaptive approaches may further enhance model generalization and robustness (Talebi and Milanfar, 2021). These approaches would enable CNNs to capitalize on larger available surface areas when present, while still functioning reliably under conditions where only limited tissue is accessible (e.g., small specimens or partial samples).

Beyond field-of-view and resolution, field-based image acquisition can be influenced in the field by lighting conditions, device-specific colour profiles, magnification differences, and sensor artifacts—that can cause “distribution shifts” (Owens et al., 2024). This phenomenon, widely documented in the CV literature, arises when the statistical properties of test-time data diverge from those observed during training, leading to degraded model performance (Alomar et al., 2023; Liu and Mirzsoleiman, 2022; Liu et al., 2020; Nanni et al., 2021). Such shifts are particularly relevant in mobile deployments, where variability in camera

hardware and user behaviour is difficult to control (Liu et al., 2020). Two primary strategies exist to address this challenge. First, hardware-based standardization, as implemented in systems like the XyloTron (Ravindran et al., 2020) or XyloPhone (Wiedenhoeft, 2020), can minimize acquisition variability by using fixed optics, controlled lighting, and predefined imaging protocols. The disadvantage of this is the need for developing robust and affordable tools to standardize image acquisition the field. Second, model robustness can be improved through data augmentation—a widely used technique in CV applications (Shorten and Khoshgoftaar, 2019). By simulating realistic perturbations during training (e.g., rotation, brightness shifts, random erasing), data augmentation exposes the model to a broader distribution of inputs, improving its ability to generalize (Alomar et al., 2023; Shorten and Khoshgoftaar, 2019). Regularization techniques such as dropout and weight decay further reduce the risk of overfitting to idealized training conditions (Srivastava et al., 2014; Xie et al., 2022). These techniques are important for mobile applications that depend on integrated cameras which each have different resolution, field-of-view due to magnification, distance to the surfaces etc. (e.g. Xylorix (Tang and Tay, 2019)) (Liu et al., 2020). Evidence supporting these strategies comes from Owens et al. (2024), who evaluated model performance under controlled image perturbations that mimic field-based acquisition artifacts (Owens et al., 2024). Their study found that CV-based wood identification models were relatively robust to superficial scratches and mild colour distortions, but more sensitive to significant blur and magnification changes (Owens et al., 2024). These results emphasize the importance of tailoring augmentation protocols to expected deployment conditions—particularly when developing smartphone-based applications where user-induced variability is pronounced (Liu et al., 2020). Furthermore, generalization for smart-phone models can be improved by training on images captured with different devices, as it incorporates variability in device-specific colour profiles and resolution (Biney and Sellahewa, 2013; Liu et al., 2020). Further research is needed to systematically evaluate how cross-device training impacts model performance and to identify strategies that best simulate field conditions, ultimately supporting the development of more resilient and deployable identification tools.

The applicability of CV-based identification to traded wood products extends beyond roundwood, and raises important questions about how diagnostic features are preserved across different levels of processing, and how reference databases should be designed to reflect these realities. Alternative, non-anatomical techniques face own challenges regarding identification of engineered wood products: for instance, glue can interfere with chemical analyses, and drying processes may reduce the amount of extractable DNA (Jiao et al., 2020; Michael Höltsken et al., 2012). While anatomical features generally remain accessible, their visibility is also affected by processing. As wood is converted into veneers or engineered products (e.g., plywood, MDF, particleboard, fibreboard), the field-of-view available for assessing anatomical surfaces becomes progressively smaller. Features that require larger continuous surfaces for assessment—such as vessel distribution patterns or banded parenchyma—are especially affected. Reference databases therefore need to account for the limitations of macroscopic anatomy to remain broadly applicable. Addressing this challenge calls for two complementary strategies. First, expanding analysis beyond the cross-sectional plane to include tangential and radial orientations may improve performance, since these surfaces are often present in blocks, veneers, or particles, and can also be obtained from splinters. Splinters are particularly valuable, as they can be collected from virtually any wood product and are widely used in forensic research. Importantly, splinters typically expose slightly larger tangential and radial surfaces compared to cross-sections, providing a richer anatomical basis for identification. Second, refining CV models to operate reliably on smaller fields of view is a promising avenue for applications to veneers and thin plywood layers. Composite materials such as fibreboard, however, requires other approaches (Helmling et al., 2018; Nieradzik et al., 2024). In these products, anatomical integrity is reduced to dispersed fragments of cells, cell walls, or isolated vessel elements. Here, CV approaches may need to be paired with segmentation or other vision-based preprocessing techniques in order to recover meaningful diagnostic signals (Helmling et al., 2018). Taken together, these prospects highlight a broader trajectory for future work: adapting CV to the practical realities of traded wood forms and engineered products, thereby enhancing its utility for enforcement and customs inspection in complex product streams.

Another critical dimension of database construction is ensuring the taxonomic reliability of reference collections (Deklerck, 2019). Misidentified specimens are present in xylaria, and careless inclusion into reference databases can influence downstream identification applications and forensic casework. CV,

when trained on highly curated collections that have been cross validated by experts and supported by ancillary methods such as herbarium vouchers or other identification techniques, offers a promising avenue for verifying the integrity of other collections. A concrete example comes from the evaluation of the INERA–Yangambi xylarium, where a CNN trained on the SmartWoodID dataset was applied to 193 specimens (presented at the 26th IUFRO Congress in Stockholm, Sweden) (De Blaere et al., 2024). By aggregating patch-level predictions at the specimen level through majority voting, the model classified 70.5% of specimens correctly to the genus level, while 57 were flagged as potentially misidentified. Such CV-based models could serve as a first-pass screening step for digitisation projects, flagging specimens whose metadata is incongruent with visual predictions and therefore warranting re-examination with complementary methods (e.g., microscopic anatomy, chemical profiling, or genetic testing). In this way, CV models trained on verified collection material may not only accelerate the detection of misidentified samples but also support the long-term goal of improving reference dataset integrity, which is fundamental to advancing automated wood identification.

These findings underscore the importance of diligent database construction for the development of robust CV-based wood identification models. While certain best practices can already be recommended (e.g., use of multiple specimens, larger scan areas, and high-quality preparation), many aspects—such as optimal resolution, acceptable surface quality, and augmentation strategies and adding information from other anatomical planes (tangential, radial)—require further empirical study. Furthermore, it underscores the importance of collaboration across wood collections due to the highly diverse pool of timber species in commerce (Chudnoff, 1984; Council and Organization, 2012; Mark et al., 2014; Richter and Dallwitz, 2000; tropicaux, 1979) and the tailored nature of wood collections focussing predominantly on delineated regions rather than covering all species world-wide (Ravindran et al., 2018; Silva et al., 2022). Leveraging this individual expertise is essential for building robust datasets to cover timbers across the world with enough specimens per species and sufficient variation in wood anatomy within trees. Collaboration also streamlines progress by reducing duplicated efforts across institutions and maximizing the utility of existing resources. In addition, different institutes are likely to have digitized their own collections through different imaging techniques, by using available resources based on expertise. Training models on such data can enhance model robustness by including device-specific variability, eliminating the need for individual institutes to invest in multiple imaging setups or repeatedly digitize their collections. For example, digitizing the 3,742 specimens of the Tervuren collection took over two years of dedicated work—using multiple imaging techniques would have significantly extended this timeline. Coordinated efforts save time and labour while broadening research opportunities across institutions. However, it is important to note that such collaboration must be carefully managed; if different institutes employ inconsistent imaging protocols or focus on distinct timber groups, it could introduce biases and compromise model generalizability.

7.4.2 Model development

Beyond the construction of robust reference databases, the design of the model itself plays a critical role in the performance, scalability, and interpretability of CV-based wood identification systems. In the broader CV domain, model architecture—including factors such as layer depth, filter size, residual connections, and hyperparameters—significantly influences a network’s capacity to generalize to unseen data (Alzubaidi et al., 2021; Bentéjac et al., 2021; Chollet and Chollet, 2021). In this study, we adopted a single architecture (Xception) with consistent hyperparameters across all models to isolate and evaluate the effects of classification strategy (Chollet, 2017). However, as CV technologies continue to evolve, future research should explore more resource-efficient architectures that can maximize predictive power while minimizing computational demands—especially for mobile deployment where processing capabilities are limited.

CNN architectures remain well suited for wood identification, as their hierarchical receptive fields allow features to be aggregated from local structures (e.g., vessels, parenchyma, rays) to global anatomical patterns across the entire image area (Taye, 2023), aligning closely with the multi-scale nature of wood anatomy. At the same time, transformer-based architectures, which leverage global self-attention, have demonstrated strong classification performance on large datasets in other domains (Gufran et al., 2023)

and may offer complementary advantages once sufficient training data are available. Hybrid approaches that combine CNNs and transformers are also promising, as they can balance local feature extraction with global context integration. Future work should therefore benchmark multiple architectural designs of neural networks systematically to determine the most effective balance between predictive performance, efficiency, and robustness for wood identification tasks.

7.4.2.1 *Classification strategy*

The findings presented in Chapter 5 highlight that the classification strategy is a key consideration in the development of wood identification models for forensic and regulatory applications. While multiclass classification has been the predominant classification strategy in studies on CV-based wood identification (Silva et al., 2022) and serves as the classification strategy in current deployed tools (e.g. Xylophone (Wiedenhoeft, 2020), Xylotron (Ravindran et al., 2020), Xylorix (Tang and Tay, 2019)), its underlying functioning makes it inherently ill-suited for the biological complexity and real-world variability that characterize the wood identification in timber trade (McCarthy and Hayes, 1981; Yoshihashi et al., 2019). Multiclass models operate under a closed-set assumption: they are trained to assign inputs to one of a fixed set of predefined classes and assume that all possible categories are represented in the training data (Sünderhauf et al., 2018; Wilber et al., 2013). This poses a limitation for wood identification, which must contend with vast diversity of timber species and fragmented coverage in existing collections, which often include only a few specimens per species (Chudnoff, 1984; Council and Organization, 2012; Mark et al., 2014; Richter and Dallwitz, 2000; tropicaux, 1979). Expanding taxonomic coverage is a slow and resource-intensive process, requiring either fieldwork by experienced botanists or collaboration across wood collections, both of which are constrained by logistics, funding, and expertise (Ravindran et al., 2018; Silva et al., 2022). Even where digitization efforts are underway, institutions face limitations in terms of imaging capacity, metadata quality, and specimen availability (Gasson et al., 2021). As a result, reference datasets grow incrementally, and models must be regularly updated to reflect the expanding taxonomic scope. Here, multiclass models exhibit a critical rigidity with architectures fixed to the classes included during training. Introducing new taxa necessitates retraining of the model—a process that is not only time-consuming but also costly in terms of computational and environmental resources. While using transfer learning can minimize these costs, they remain necessary to expand the pool of timbers. This poses a scalability bottleneck for global deployment. Moreover, multiclass models perform poorly in open-set conditions, where inputs may include species that are absent in training data. Opt-out mechanisms, such as probability thresholds or “unknown” classes, offer partial mitigation, but these strategies are imperfect solutions (Entezari and Saukh, 2020; Geifman and El-Yaniv, 2019). A workaround is to train region-specific models tailored to the local timber trade. While pragmatic, this leads to the proliferation of models trained on different taxa, using varying architectures, data sources, and augmentation strategies. This undermines consistency and comparability across institutions, complicates interpretation, and limits the potential for global standardization. In the absence of shared benchmarks, protocols, or certification frameworks for field-ready identification models, such fragmentation reduces transparency and poses challenges for regulatory harmonization (Ravindran and Wiedenhoeft, 2022).

Object re-identification offers a more flexible classification strategy for wood identification by decoupling recognition from a fixed set of categories (Geng et al., 2020; Scheirer et al., 2012; Yoshihashi et al., 2019). In contrast to multiclass models that produce discrete class labels, object re-identification networks are trained to compute image similarities, enabling the identification of the most corresponding reference image(s) (Ye et al., 2021). In this study, we demonstrated that object re-identification is a promising classification strategy for wood identification in the field. We tested two approaches, verified in literature in different fields (facial (Chen et al., 2017; Hermans et al., 2017; Schroff et al., 2015; Shi et al., 2016; Ye et al., 2021) and vehicle recognition (Bai et al., 2018; He et al., 2020; Kumar et al., 2020; Shen et al., n.d.; Tang et al., 2019)): binary verification (Chen et al., 2017; Ye et al., 2021) and triplet learning (Bai et al., 2018; Chen et al., 2023; Ghosh et al., 2023; Guo and Lovell, 2024). This provided the following insights for performance on the same Congolese timbers compared to the traditional closed-world multiclass technique (hypothesis 4.2); and for performance of binary verification for identifying non-Congolese timbers (hypothesis 4.3).

Our results demonstrated that the technique of binary verification represents a viable classification strategy for wood identification in open-world contexts hypothesis. Rather than assigning an input to a predefined class, binary verification compares an unknown specimen to reference images and computes a similarity score per image (Chen et al., 2017; Ye et al., 2021). This pairwise structure allows models to operate independently of the number or identity of classes seen during training, enabling generalization to taxa not present in the training data. Our results show that binary verification achieves performance metrics comparable to those of traditional multiclass CNNs, particularly when evaluated on the same set of Congolese commercial timbers. The binary verification model exhibited only slightly lower performance (accuracy: 0.75; F1-score: 0.69) than the multiclass model (accuracy: 0.86; F1-score: 0.84). Notably, when applied to unseen (non-Congolese) timbers excluded from the training data, the binary model correctly identified the genus in 95% of cases when the top 30% of candidate genera were considered. This ability to maintain reliable performance on previously unseen taxa is particularly important for real-world deployment, where inspected samples may originate from diverse geographic and taxonomic sources. Binary verification offers additional advantages in transparency and interpretability. Unlike multiclass models, which provide only class probabilities with no direct linkage to specific specimens, binary verification identifies similarity to individual reference images. This enables traceable predictions, where the basis of classification can be examined through direct visual comparison or review of specimen metadata. In forensic and regulatory settings, such transparency is crucial for ensuring accountability, enabling expert corroboration, and supporting documentation standards such as chain-of-custody records. Finally, the design of binary verification models facilitates international collaboration and dataset integration. Because the model evaluates similarity at the image level, it can operate across digitized reference databases compiled from multiple institutions, without requiring retraining for each new dataset or region. This enables a more harmonized approach to timber identification, supporting broader taxonomic coverage while avoiding the fragmentation that results from maintaining multiple region-specific multiclass models.

Beyond technical performance, classification strategy also carries important semantic implications for how model outputs are interpreted and used in practice. In the context of timber trade monitoring and law enforcement, field-based wood identification serves as an initial screening step rather than a definitive taxonomic determination (Ravindran et al., 2021). At this stage, the primary objective is not to determine the botanical taxon of traded timber, but to assess whether the declared taxon on accompanying documentation plausibly matches. This process supports the early detection of potential non-compliance, including the misdeclaration of protected species, and informs decisions about whether further identification is warranted (Ravindran et al., 2021). In such workflows, models need not provide full species resolution but must instead support reliable verification—determining whether a specimen is consistent with or significantly deviates from known reference examples. Binary verification is well suited to this task. Its design aligns with the operational logic of preliminary screening by producing a similarity score that quantifies the degree of match between a query image and a reference set of specimens (Chen et al., 2017). This structure enables rapid flagging of potentially suspicious cases, such as when a specimen shows high similarity to a CITES-listed genus but is declared as a different, non-protected timber.

Practical implementations of verification-like strategies have been applied in field-deployable systems, such as the Xylorix platform (Tang and Tay, 2019), which can output similarity scores for individual timbers via its API. While the underlying methodology has not been fully detailed in the literature and remains subject of speculation, it is plausible that such systems rely on conventional approaches. One conventional possibility is the use of dedicated binary classifiers trained for individual timbers (e.g., “*Afzelia*” vs. “not *Afzelia*”). Although this produces intuitive outputs between 0–1, the interpretability of the resulting scores depends heavily on how the “other” class is constructed, which may vary widely in taxonomic coverage. Moreover, this strategy requires developing and maintaining a large library of binary models—potentially one per commercial timber species. Such an approach incurs substantial computational and logistical costs, as each model must be trained, validated, and periodically retrained when new reference data become available. This creates scalability challenges for practical deployment in forensic and regulatory contexts. An alternative is to derive verification-like outputs from multiclass

models, by evaluating whether the declared species is assigned the highest posterior probability among the included classes. While this avoids the need for hundreds of dedicated binary models, it remains constrained by the closed-set assumption of multiclass classification. Probabilities are conditioned only on the taxa represented in training, which limits their reliability when encountering species outside the reference set and makes it difficult to capture genuine uncertainty. Together, these strategies illustrate different ways in which verification can be operationalized, but they also highlight the advantages of the binary verification framework developed in this dissertation. Unlike multiclass probabilities or species-specific binary classifiers, verification networks directly quantify the similarity between query and reference images and can be extended to individual reference specimens. This design is both scalable and interpretable, making it particularly well-suited to forensic and regulatory applications where robustness to open-set conditions and transparent decision-making are critical.

However, binary verification also has inherent limitations. Its reliance on exhaustive pairwise comparisons result in significantly increased computational load—particularly when working with large reference sets or performing high-resolution patch-based analysis (Chen et al., 2017; Hermans et al., 2017). Although strategies such as hierarchical narrowing or limiting comparisons to genus-level representatives can reduce this, they add complexity to the model pipeline (Chen et al., 2017; Hermans et al., 2017; Schroff et al., 2015). Furthermore, binary verification models are optimized to distinguish between matches and non-matches based on a threshold, rather than to distribute probability across a class hierarchy (Chen et al., 2017; Ye et al., 2021). Identification therefore implies transforming the model output of similarity scores to an output that enables ranking predicted genera to establish the most likely taxon (Chen et al., 2017). The manner in which predictions are aggregated across image patches to obtain a specimen-level decision is therefore critical. Naïve averaging of patch-level probabilities (applied in Chapter 5) offers a straightforward means of aggregation, as it preserves the full distribution of confidence scores across classes and enables ranking beyond the top prediction. However, it is inherently sensitive to variation, as a single patch can disproportionately distort the aggregate. This may explain why top-k analysis (section 5.4.4) revealed that more top-ranked genera needed to be considered to include the correct genus in 95% of the specimens compared to multiclass CNN. Still, we note that for Congolese commercial timbers this was limited with the accuracy matching multiclass classification by considering the top two predicted genera, representing still a strong and accurate result for field identification across 56 commercial genera and underscoring it as a valid technique. Majority voting, as applied in Chapter 4, provides a robust alternative to occasional misclassified patches, since the dominant signal across patches determines the final decision. Its stability is exemplified in the results presented in Table 4.1 (Section 4.4.2.1) showing that specimen-aggregated recall across all training scenarios exceeded recall values calculated at the patch level, indicating that majority voting can secure high overall performance. Nonetheless, this approach can also become unstable when the number of patches per specimen is small or when class votes are closely divided. Weighted aggregation techniques represent a pragmatic middle ground. By amplifying the influence of diagnostically strong patches while reducing the effect of uncertain ones, such methods increase resilience to local noise without sacrificing the richness of probability distributions. Future research should systematically investigate how different aggregation strategies affect both class-level performance and specimen-specific outcomes, as methodological choices at this stage may have a decisive impact on the reliability of wood identification results.

We also investigated a second re-identification strategy: embedding-based object re-identification using triplet learning. This method learns a discriminative feature space where similar images are close together and dissimilar ones are far apart (Hermans et al., 2017), effectively transforming an image into a digital fingerprint (embedding vector) that enables rapid classification using machine-learning algorithms such as nearest-neighbour classification with cosine similarity, RF (Salman et al., 2024), or gradient-weighted boosting algorithms such as XGBoost (Bentéjac et al., 2021). Despite its conceptual advantages and strong performance in other CV domains metrics (Bai et al., 2018; Chen et al., 2023; Ghosh et al., 2023; Hermans et al., 2017; Kumar et al., 2020; Schroff et al., 2015; Shen et al., n.d.; Ye et al., 2024), our implementation of triplet learning consistently underperformed (see sections 5.4.3 and 5.4.4). This may be attributed to the hard mining methodology implemented in this study, which worked well for binary verification but may not

suffice for triplet learning. These findings align with prior research showing that embedding-based methods are highly sensitive to the quality of the mining strategy and loss function design (Bai et al., 2018; Chen et al., 2017; Hermans et al., 2017; Schroff et al., 2015). More advanced approaches—such as online hard example mining (Hermans et al., 2017; Schroff et al., 2015), histogram loss (Ustinova and Lempitsky, 2016), or hybrid models like CROSR (Yoshihashi et al., 2019)—may better structure the embedding space and improve generalization.

7.4.2.2 *Integrating different wood identification techniques*

Integrating wood identification techniques is key in forensic research on wood identification, as combining multiple sources of diagnostic information strengthens the reliability of taxonomic assignments in the timber trade (Schmitz et al., 2020). Each method contributes unique perspectives on wood identity—whether anatomical, chemical, genetic, or isotopic—and together they provide a stronger evidentiary basis than any single technique alone. However, effective integration requires careful consideration, as each method relies on different measurement principles, produces different data structures, and is subject to different limitations.

An effective strategy for improving accuracy would be refining predictions through a serial integration of techniques. In such a workflow, CV-based models could serve as an initial screening stage: narrowing candidates to a top-*k* set and then directing which auxiliary method is most suitable for definitive resolution. This could be applied with any CV-based models, including the output of other tools such as Xylophone (Wiedenhoeft, 2020), Xylotron (Ravindran et al., 2020), Xylorix (Tang and Tay, 2019). Unlocking such potential requires close collaboration with taxonomic experts to establish which techniques are most effective for specific species or species groups—an endeavour that is central to advancing wood identification practices. A particularly promising pathway is to leverage anatomical information as a bridge between macroscopic and microscopic analysis. Databases such as InsideWood provide an invaluable foundation for this integration by codifying the expert tradition of anatomical wood identification (Wheeler, 2011). By linking macroscopic CV predictions with feature importance scores derived from microscopic descriptors, systems could flag the traits most diagnostic for distinguishing among taxa in the top-*k* set. For instance, if predictions converge on closely related *Pterocarpus* species, the model could highlight which microscopic features merit closer inspection, or suggest when complementary analyses—chemical, genetic, or otherwise—are required. In this way, CV classifiers could evolve from stand-alone predictive engines into interactive diagnostic tools, guiding forensic workflows by prioritizing critical features, pointing to confirmatory tests, and ultimately enabling more reliable taxonomic resolution.

However, prediction refinement need not be restricted to serial approaches. Combining different techniques into a joint prediction framework can also be highly powerful. The wood identification techniques (as discussed in section 1.2) (Schmitz et al., 2020) follow a common structure: feature extraction followed by classification based on modality-specific reference data. Across techniques, the extracted features typically take the form of sequential or vector-based data. For instance, anatomical assessments rely on expert interpretation, codified into binary or categorical descriptors indicating the presence of microscopic anatomical features (Committee, 2004; Koch et al., 2018; NS, 1989; Wheeler, 2011). DART-TOFMS produces chemical fingerprints by ionizing low-molecular-weight compounds through thermal desorption and measuring their mass-to-charge ratios via TOFMS, generating numerical spectral profiles (Cody et al., 2005; Deklerck, 2022, 2019; Deklerck et al., 2020; Price et al., 2022). Similarly, NIRS captures the absorbance of light in the 800–2500 nm range by high-molecular-weight compounds such as cellulose, lignin, and extractives, yielding vectors of wavelength-specific intensity values (Deklerck, 2019; Lowe et al., 2016; Tsuchikawa et al., 2003; Tsuchikawa and Kobori, 2015). DNA-based approaches—including barcoding (Jiao et al., 2020, 2019) and fingerprinting (Lowe et al., 2010; Thünen Institute of Forest Genetics, 2015)—encode taxonomic or individual identity as nucleotide sequences, while stable isotope analysis provides geographic origin data through continuous variables reflecting local environmental isotope ratios (Camin et al., 2017; Dormontt et al., 2015; Horacek et al., 2009; Kagawa and Leavitt, 2010; Lin et al., 2024). Traditional CV-based methods (multiclass) differ from these approaches by merging feature extraction and classification into a single, end-to-end process (Alzubaidi et al., 2021; Chollet and

Chollet, 2021; Taye, 2023; Wu, 2017). While effective for image-based predictions, this architecture complicates integration with other sequential data streams.

Integration of anatomical information with other sequential data can be approached at two main stages of the identification chain: *a priori* (before feature extraction) or *a posteriori* (after classification). *A priori* integration involves transforming input data from different modalities into a common representational format, such as converting sequential data into images suitable for CNN training. For instance, Jahanbanifard et al. (2023) demonstrated this by converting DART-TOFMS spectra into image-like visual plots for CNN classification (Jahanbanifard et al., 2023). While this method allows multiple modalities to be represented in a unified way, the transformation is indirect, potentially lossy, and methodologically complex.

A posteriori integration, in contrast, combines the outputs of independent models after prediction. This approach typically aggregates posterior probabilities from different techniques to reach a consensus decision. Its main strength lies in flexibility: each method can operate within its native domain without modification, while integration occurs only at the decision-making stage. A *posteriori* integration would also enable leveraging the prediction of different multiclass CV-based models embedded in deployed tools such as Xylophone (Wiedenhoeft, 2020), Xylotron (Ravindran et al., 2020), Xylorix (Tang and Tay, 2019). Provided that full class probabilities are accessible through their APIs, such integration could facilitate cross-referencing of predictions against multiple deployed tools, thereby enhancing the robustness of automated wood identification. Yet, the probabilistic outputs of multiclass models must be interpreted with care, since those tools are conditioned on closed sets of classes (timbers): probabilities reflect the relative likelihood of a taxon given the training set, rather than an absolute measure of similarity. However, *a posteriori* integration treats the features of each technique as separate and therefore cannot leverage new patterns in features across modalities, limiting their potential to exploit the full diagnostic value of combined data.

Recent advances in classification strategies have opened the way for a third approach that blends advantages of both. Rather than transforming input data prior to modelling or combining only posterior probabilities, images can be transformed into structured, sequential data via embedding-based object re-identification (Schroff et al., 2015). By extracting fixed-length vectors from raw anatomical images, embeddings map samples into a multidimensional space where anatomical similarity corresponds to spatial proximity (Ghosh et al., 2023). Unlike conventional CNN outputs, these embeddings are structured numerical representations that can be directly compared or integrated with other vectorized data from chemical, genetic, or isotopic sources. This alignment facilitates deeper integration across diagnostic methods, enhancing species- and potentially population-level identification. This underscores that embedding-based object re-identification represents a promising pathway toward fully interoperable, multimodal wood identification systems. As reference databases grow and techniques for embedding alignment advance, this strategy offers scalable, field-adaptable, and forensically reliable tools for timber identification and trade enforcement.

7.4.2.3 Context-sensitive predictions in wood identification

Another crucial consideration for wood identification is *a posteriori* processing in general. This can serve not only to integrate different techniques but also to refine model outputs after prediction, bringing them closer to the ground truth. Such refinement is particularly important because wood, as a biological material, exhibits substantial variability and complexity, while classification is embedded within a hierarchical taxonomic system. Even when robust modelling techniques are applied, raw outputs may not fully align with practical identification needs—especially in forensic and conservation contexts where the consequences of error are uneven across timbers.

A promising avenue lies in incorporating information on anomalies. As shown in Section 4.4.2.1 (Table 4.1), recall was consistently lower for anomalous test patches compared to anomaly-free ones, confirming that predictions are more reliable when clear anatomical structures are present. Aggregation at the specimen level using majority voting reduced this disparity, as reflected in higher recall scores, though specimens

with limited usable area may remain vulnerable. *A posteriori* weighting of patch contributions could further mitigate such effects by prioritizing anomaly-free regions. This could be achieved through fixed scoring rules or probabilistic weighting based on the outputs of a binary anomaly verification model. More advanced approaches would involve segmenting anomalies (e.g., cracks, insect damage, fungal staining, pith, bark) and quantifying their extent to generate continuous quality indicators. Grad-CAM visualizations may also be leveraged to estimate anomaly distributions directly, which could inform aggregation strategies or even be integrated into model architectures via transfer learning. Collectively, these strategies underscore the need for further research into output standardization methods that explicitly incorporate patch quality in order to generate more reliable specimen-level predictions.

Beyond basic *a posteriori* processing—such as weighting predictions according to the quality of individual images or surfaces to reduce the impact of anomalies—more advanced approaches can account for relationships both between classes, and between classes and their broader context. Bayesian cost modelling offers a principled framework for such adjustments, enabling the integration of uncertainty estimates and contextual information into the decision-making process (Lampinen and Vehtari, 2001). Unlike likelihood-based methods, which treat all classes equally during training and prediction, posterior approaches enable rebalancing outcomes according to their real-world significance (Jackson et al., 2010; Mediavilla-Relaño et al., 2023). These frameworks are relevant for handling look-alike *taxa*—cases where visually similar woods carry different regulatory or economic significance. Here, posterior adjustments could be applied, ensuring that closely resembling species are either flagged for additional scrutiny or classified conservatively at a higher taxonomic rank when certainty is insufficient. Furthermore, *a posteriori* processing is also critical in law enforcement, where errors are not equally consequential: confusing two common species may be acceptable, but failing to detect a CITES-listed taxon carries ecological and legal repercussions. Posterior adjustments can address this asymmetry by, for example, recalibrating predictions with priors reflecting species frequencies in trade, geographic origin, or shipment documentation. Cost-sensitive decision frameworks can also assign higher penalties to false negatives on protected species, aligning decisions with enforcement priorities rather than purely statistical likelihoods. Adaptive thresholds may further increase recall for high-risk *taxa* by lowering the probability required for positive identification, thereby reducing the likelihood of missing endangered species. Importantly, such methods must be validated carefully using both precision and recall, to avoid unfair identifications that could, for instance, wrongly implicate businesses based on overly cautious thresholds. Automated approaches such as CV therefore should not be applied in isolation for legal determinations but rather in tandem with expert assessment, with their primary utility lying in high-throughput screening rather than courtroom evidence.

Although not implemented in the present study, these strategies point to a clear direction for future development. By integrating anomaly weighting, contextual priors, and cost-sensitive adjustments, wood identification systems can evolve from tools of visual similarity to robust decision-support frameworks that explicitly account for biological variability, taxonomic complexity, and regulatory stakes—better aligning automated outputs with the priorities of forensic enforcement and conservation biology.

In light of these considerations, it is also critical to address a further limitation of current approaches. Wood, as a biological material, is embedded within a nested taxonomic hierarchy, yet most models produce predictions at only a single level of classification. Ideally, outputs should distribute probabilities across taxonomic levels—for example, estimating confidence that a specimen belongs to Meliaceae (family), *Khaya* (genus), or *Khaya anthotheca* (species). To date, this has not been systematically explored in the wood identification literature, where models typically target a single taxonomic rank or rely on trade groupings that do not align strictly with taxonomy (e.g., Meranti wood encompassing multiple *Shorea* spp., or genera such as *Lophira* and *Milicia* that are represented in trade by a single species). In other domains, hierarchical classification has been advanced through strategies such as hierarchical loss functions (La Grassa et al., 2021; Yan et al., 2015) and contrastive learning (Kokilepersaud et al., 2024). Applying such approaches to wood identification would not only improve predictive reliability but also generate biologically faithful representations of uncertainty, allowing practitioners to determine the taxonomic level at which conclusions can be drawn with confidence. Future work should therefore move beyond single-

rank predictions to explore hierarchical architectures that provide multi-level outputs, integrate taxonomic priors, and calibrate uncertainty across ranks, enabling decisions that are both more robust and more aligned with biological reality.

Circling back to the *a posteriori* approaches, applying cost-sensitive decision frameworks to look-alike timbers requires clear definition of what constitutes a look-alike. In essence, a timber is a look-alike when distinguishing features are so subtle that consistent separation is difficult. In rapid, field-based identification using macroscopic cross-sections, this depends directly on the scope of information available. As shown in section 3.4.4, expert-defined anatomical descriptors can cluster taxa with broadly similar features, but these descriptors are reduced abstractions of wood anatomy and thus provide only coarse groupings. CNN-extracted features (Chapter 5), by contrast, capture finer-grained patterns that allow separation of taxa that experts could not distinguish with reduced descriptors. This is reinforced in Chapter 6, where the limited added value of re-ranking with expert features shows that CNNs already encode much of the same diagnostic information. Binary verification models trained with hard example mining (Section 5.3.7) further built on this, as they were explicitly challenged to distinguish between anatomically similar specimens (according to the coded expert-defined anatomical features). Together, these results suggest that CNN-derived representations form a stronger foundation for defining look-alikes than expert-defined features alone.

At the same time, the results in Chapter 6 show that CNNs do not capture all relevant information. Some genera benefited from re-ranking with expert-defined traits, indicating that diagnostic signals visible only at larger field-of-view scales remain underutilized. Whereas CNNs work with fixed-size cropped patches, experts evaluate features across entire cross-sections, often scanning between magnifications to detect rare or large-scale patterns (e.g., banded parenchyma or vessel groupings). This gap highlights the need for multi-scale approaches that allow CNNs to integrate information across resolutions and spatial contexts, aligning them more closely with expert assessment and enabling a clearer operational definition of look-alikes. Geographic variation is crucial to consider species that appear distinct in one region may be virtually indistinguishable in another due to convergent traits or local adaptations. Future frameworks for defining look-alikes in CV-based wood identification should therefore combine CNN feature hierarchies with expert descriptors, multi-scale anatomical representations, while covering a broad taxonomic scope beyond the boundaries of a single country (as was the case for this study on the DRC). Such integration would provide a more reliable definition of look-alikes, which is essential for cost-sensitive decision-making in forensic and regulatory applications.

Funding

BELSPO through the BRAIN-be 2.0 (Belgian Research Action through Interdisciplinary Networks PHASE 2–2018-2023) (B2/202/P2/SmartWoodID); the UGent Special Research Fund through the BOF Starting Grant JVdB (BOFSTG2018000701).

Bibliography

- (Rome), F. and A.O. of the U.N.F.D., 2010. Global forest resources assessment 2010: Main report. Food and Agriculture Organization of the United Nations.
- Abeebe, S. Vanden, Beeckman, H., De Mil, T., De Troyer, C., Deklerck, V., Engledow, H., Hubau, W., Stoffelen, P., Janssens, S.B., 2021. When xylarium and herbarium meet: linking Tervuren xylarium wood samples with their herbarium specimens at Meise Botanic Garden. *Biodivers. Data J.* 9.
- Aleman, J.C., Blarquez, O., Gourlet-Fleury, S., Bremond, L., Favier, C., 2017. Tree cover in Central Africa: determinants and sensitivity under contrasted scenarios of global change. *Sci. Rep.* 7, 41393.
- Alexander, K., 2014. The Lacey Act: protecting the environment by restricting trade. Congressional Research Service Washington, DC, USA.
- Ali, A.K., Abdullah, A.M., Raheem, S.F., 2024. Impact the Classes' Number on the Convolutional Neural Networks Performance for Image Classification. *Int. J. Adv. Sci. Comput. Eng.* 6, 64–69.
- Alomar, K., Aysel, H.I., Cai, X., 2023. Data augmentation in classification and segmentation: A survey and new strategies. *J. Imaging* 9, 46.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8, 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Amari, S., 1993. Backpropagation and stochastic gradient descent method. *Neurocomputing* 5, 185–196.
- Andrade, B., Basso, V., Latorraca, J.V., 2020. Machine vision for field-level wood identification. *IAWA J.* 41, 1–18. <https://doi.org/10.1163/22941932-bja10001>
- Angyalossy, V., Pace, M.R., Evert, R.F., Marcati, C.R., Oskolski, A.A., Terrazas, T., Kotina, E., Lens, F.P., Mazzoni-Viveiros, S.C., Angeles, G., 2016. IAWA list of microscopic bark features. *IAWA J.* 37, 517–615.
- APD, 2012. African Plant Database (version 3.4.0) [WWW Document]. *Conserv. Jard. Bot. la V. Genève South African Natl. Biodivers. Institute, Pretoria*. URL <http://africanplantdatabase.ch> (accessed 2.22.22).
- Artemov, M., 2025. Wood Identifier: Tree & Plant.
- Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A., 2017. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* 34, 26–38.
- Arzac, A., López-Cepero, J.M., Babushkina, E.A., Gomez, S., 2018. Applying methods of hard tissues preparation for wood anatomy: Imaging polished samples embedded in polymethylmethacrylate. *Dendrochronologia* 51, 76–81.
- Augustino, S., Hall, J.B., 2008. Population status of *Pterocarpus tinctorius*: A medicinal plant species in Urumwa forest reserve, Tanzania. *Tanzania J. For. Nat. Conserv.* 78.
- Ayed, S.H., 2006. Council Regulation 2173/2005/EC of 20 December 2005 Establishing a Flegt Licence Scheme for Imports of Timber in the European Community. *Environ. Law Rev.* 8, 218–224.
- Baas, P., Werker, E., 1981. A New Record of Vestured Pits in Cistaceae. *IAWA J.* 2, 41–42. <https://doi.org/https://doi.org/10.1163/22941932-90000393>
- Bai, Y., Lou, Y., Gao, F., Wang, S., Wu, Y., Duan, L.-Y., 2018. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* 20, 2385–2399.
- Banerjee, C., Mukherjee, T., Pasiliao Jr, E., 2019. An empirical study on generalizations of the ReLU activation function, in: *Proceedings of the 2019 ACM Southeast Conference*. pp. 164–167.
- Barefoot, A.C., Hankins, F.W., 1982. Identification of modern and tertiary woods. Oxford University Press.

- Barima, Y.S.S., Djibu, J.P., Alongo, S., Ndayishimiye, J., Bomolo, O., Kumba, S., Iyongo, L., Bamba, I., Mama, A., Toyi, M., 2011. Deforestation in Central and West Africa: landscape dynamics, anthropogenic effects and ecological consequences. 9781617287749.
- Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features, in: European Conference on Computer Vision. Springer, pp. 404–417.
- Beech, E., Rivers, M., Oldfield, S., Smith, P.P., 2017. GlobalTreeSearch: The first complete global database of tree species and country distributions. *J. Sustain. For.* 36, 454–489.
- Beeckman, H., 2007. Collections of the RMCA - Wood. Royal Museum for Central Africa.
- Beeckman, H., 2003. A xylarium for the sustainable management of biodiversity: the wood collection of the Royal Museum for Central Africa, Tervuren, Belgium. *Bull. l'APAD*.
- Beeckman, H., Jolivet-Blanc, C., Boeschoten, L., Braga, J.W.B., Cabezas, J.A., Chaix, G., Crameri, S., Degen, B., Deklerck, V., Dormontt, E., 2020. Overview of current practices in data analysis for wood identification: a guide for the different timber tracking methods. Schmitz, N., ed. GTTN, Glob. Timber Track. Netw. GTTN Secr. Eur. For. Inst. Thunen Institute. 143 p.
- Beeckman, H., Yin, Y., 2024. BOOK OF ABSTRACTS: FORESTS & SOCIETY TOWARDS 2050, in: 26th IUFRO world Congress (Ed.), T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification: Digitizing a Xylarium beyond Shared Meta-Data: The Importance of Quantitative Wood Anatomy, Images and Chemistry. 26th IUFRO world Congress, Stockholm, Sweden, p. 3524.
- Bejani, M.M., Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks. *Artif. Intell. Rev.* 54, 6391–6438.
- Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 54, 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25, 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Biney, A.G., Sellahewa, H., 2013. Analysis of smartphone model identification using digital images, in: 2013 IEEE International Conference on Image Processing. IEEE, pp. 4487–4491.
- Borcard, D., Gillet, F., Legendre, P., 2011. Numerical ecology with R. Springer.
- Boulesteix, A., Janitza, S., Kruppa, J., König, I.R., 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 2, 493–507.
- Brack, D., Gray, K., Hayman, G., 2002. Controlling the international trade in illegally logged timber and wood products. Royal Institute of International Affairs London.
- Brazier, J.D., Franklin, G.L., 1961. Identification of hardwoods: a microscope key.
- Bremananth, R., Nithya, B., Saipriya, R., 2009. Wood species recognition using GLCM and correlation, in: 2009 International Conference on Advances in Recent Technologies in Communication and Computing. IEEE, pp. 615–619.
- Burley, J., 2004. Encyclopedia of forest sciences. Academic Press.
- Butterfield, B., 2012. Three-dimensional structure of wood: an ultrastructural approach. Springer Science & Business Media.
- Cabala Kaleba, S., Useni Sikuzani, Y., Sambieni, K.R., Bogaert, J., Munyemba Kankumbi, F., 2017. Dynamique des écosystèmes forestiers de l'Arc Cuprifère Katangais en République Démocratique du Congo. I. Causes, transformations spatiales et ampleur. *Tropicultura* 35.
- Camin, F., Boner, M., Bontempo, L., Fauhl-Hassek, C., Kelly, S.D., Riedl, J., Rossmann, A., 2017. Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends Food*

- Sci. Technol. 61, 176–187.
- Cao, J., Su, Z., Yu, L., Chang, D., Li, X., Ma, Z., 2018. Softmax cross entropy loss with unbiased decision boundary for image classification, in: 2018 Chinese Automation Congress (CAC). IEEE, pp. 2028–2032.
- Carlquist, S., Hoekman, D.A., 1985. Ecological wood anatomy of the woody southern Californian flora. *Iawa J.* 6, 319–347.
- Chen, J., Chen, H., Jiang, X., Gu, B., Li, W., Gong, T., Zheng, F., 2023. On the stability and generalization of triplet learning, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 7033–7041.
- Chen, S., Awano, T., Yoshinaga, A., Sugiyama, J., 2022. Flexural behavior of wood in the transverse direction investigated using novel computer vision and machine learning approach. *Holzforschung* 76, 875–885.
- Chen, W., Chen, X., Zhang, J., Huang, K., 2017. A multi-task deep network for person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence.
- Chen, Y., Guan, W., 2019. Disease recognition from images having a large field of view.
- Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1251–1258.
- Chollet, Francois, Chollet, François, 2021. Deep learning with Python. Simon and Schuster.
- Chudnoff, M., 1984. Tropical timbers of the world. US Department of Agriculture, Forest Service.
- CITES, 2022a. CONSIDERATION OF PROPOSALS FOR AMENDMENT OF APPENDICES I AND II - Proposal 50, in: Nineteenth Meeting of the Conference of the Parties. CONVENTION ON INTERNATIONAL TRADE IN ENDANGERED SPECIES OF WILD FAUNA AND FLORA, Panama City (Panama), pp. 1–23.
- CITES, 2022b. CONSIDERATION OF PROPOSALS FOR AMENDMENT OF APPENDICES I AND II - Proposal 46, in: Nineteenth Meeting of the Conference of the Parties. CONVENTION ON INTERNATIONAL TRADE IN ENDANGERED SPECIES OF WILD FAUNA AND FLORA, Panama City (Republic of Panama), pp. 1–23.
- CITES, 2022c. CONSIDERATION OF PROPOSALS FOR AMENDMENT OF APPENDICES I AND II - Proposal 51, in: Nineteenth Meeting of the Conference of the Parties. CONVENTION ON INTERNATIONAL TRADE IN ENDANGERED SPECIES OF WILD FAUNA AND FLORA, Panama City (Republic of Panama).
- CITES, 2019. CONSIDERATION OF PROPOSALS FOR AMENDMENT OF APPENDICES I AND II - Proposal 54, in: Eighteenth Meeting of the Conference of the Parties. CONVENTION ON INTERNATIONAL TRADE IN ENDANGERED SPECIES OF WILD FAUNA AND FLORA, Colombo (Sri Lanka), pp. 1–12.
- Cody, R.B., Laramée, J.A., Durst, H.D., 2005. Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal. Chem.* 77, 2297–2302.
- Commission Regulation (EU), 2019. Commission Regulation (EU) 2019/2117 of 29 November 2019 amending Council Regulation (EC) No 338/97 on the protection of species of wild fauna and flora by regulating trade therein [WWW Document].
- Committee, I.A. of W.A., 2004. IAWA list of microscopic features for softwood identification. *IAWA J.* 25, 1–70.
- Council, I.T.T., Organization, I.T.T., 2012. Annual review and assessment of the world timber situation.
- Cunningham, P., Cord, M., Delany, S.J., 2008. Supervised learning, in: Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval. Springer, pp. 21–49.
- Dargie, G.C., Lewis, S.L., Lawson, I.T., Mitchard, E.T.A., Page, S.E., Bocko, Y.E., Ifo, S.A., 2017. Age, extent and carbon storage of the central Congo Basin peatland complex. *Nature* 542, 86–90.
- Dauby, G., Zaiss, R., Blach-Overgaard, A., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield,

- J., Droissart, V., Duarte, M.C., 2016. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* 1.
- De Blaere, R., Laurent, F., Angoboy Ilondea, B., Lievens, K., Van den Bulcke, J., Verwaeren, J., Beeckman, H., 2024. BOOK OF ABSTRACTS: FORESTS & SOCIETY TOWARDS 2050, in: 26th IUFRO world Congress (Ed.), T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification: Valorizing Xylaria Using Computer Vision-Based Wood Identification: A Case Study on the INERA-Yangambi Xylarium. 26th IUFRO world Congress, Stockholm, Sweden, p. 3536.
- De Blaere, R., Lievens, K., Van Hassel, D., Deklerck, V., De Mil, T., Hubau, W., Van Acker, J., Bourland, N., Verwaeren, J., Van den Bulcke, J., Beeckman, H., 2023. SmartWoodID—an image collection of large end-grain surfaces to support wood identification systems. *Database* 2023, baad034. <https://doi.org/10.1093/database/baad034>
- de Luna, B.N., Freitas, M. de F., Barros, C.F., 2018. Systematic and phylogenetic implications of the wood anatomy of six Neotropical genera of Primulaceae. *Plant Syst. Evol.* 304, 775–791.
- De Oliveira, W., de Paula Filho, P.L., Martins, J.G., 2019. Software for forest species recognition based on digital images of wood. *Floresta* 49.
- DeFries, R., Hansen, A., Turner, B.L., Reid, R., Liu, J., 2007. Land use change around protected areas: management to balance human needs and ecological function. *Ecol. Appl.* 17, 1031–1038.
- Deklerck, V., 2022. Timber origin verification using mass spectrometry: challenges, opportunities, and way forward. *Forensic Sci. Int. Anim. Environ.* 100057.
- Deklerck, V., 2019. National treasure: valorisation of the Federal Xylarium in Belgium for timber identification and wood technology.
- Deklerck, V., Lancaster, C.A., Van Acker, J., Espinoza, E.O., Van den Bulcke, J., Beeckman, H., 2020. Chemical fingerprinting of wood sampled along a pith-to-bark gradient for individual comparison and provenance identification. *Forests* 11, 107.
- Denk, T., Grimm, G.W., Manos, P.S., Deng, M., Hipp, A.L., 2017. An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. *Oaks Physiol. Ecol. Explor. Funct. Divers. genus Quercus L.* 13–38.
- DIAF, 2017. Guide opérationnel-Liste des essences de la République Démocratique du Congo, Série: Généralités –N°2, Version révisée.
- Dierickx, S., Genbrugge, S., Beeckman, H., Hubau, W., Kibleur, P., Van den Bulcke, J., 2024. Non-destructive wood identification using X-ray μ CT scanning: which resolution do we need? *Plant Methods* 20, 98.
- Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., Koch, G., 2015. Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biol. Conserv.* 191, 790–798.
- Downes, G.M., Drew, D.M., 2008. Climate and growth influences on wood formation and utilisation. *South. For. a J. For. Sci.* 70, 155–167.
- Dubey, S.R., Singh, S.K., Chaudhuri, B.B., 2022. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* 503, 92–108.
- EC, 2003. Communication from the Commission to the Council and the European Parliament. Forest Law Enforcement, Governance and Trade (FLEGT). Proposal for an EU Action Plan. Brussels, 21.05. 2003. COM (2003) 251 final.
- Entezari, R., Saukh, O., 2020. Class-dependent pruning of deep neural networks, in: 2020 IEEE Second Workshop on Machine Learning on Edge in Sensor Systems (SenSys-ML). IEEE, pp. 13–18.
- Esri, 2022. World Countries (Generalized).

- European Commission, 2019. FLEGT REGULATION - FLEGT Voluntary Partnership Agreements (VPAs) [WWW Document]. URL <http://ec.europa.eu/environment/forests/flegt.htm> (accessed 8.18.25).
- FAO, T., 2022. The state of the world's forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies. Rome, FAO.
- Fayolle, A., Swaine, M.D., Bastin, J., Bourland, N., Comiskey, J.A., Dauby, G., Doucet, J., Gillet, J., Gourlet-Fleury, S., Hardy, O.J., 2014. Patterns of tree species composition across tropical African forests. *J. Biogeogr.* 41, 2320–2331.
- Fritz, S., Bartholomé, E., Belward, A., Hartley, A., Stibig, H.J., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., 2003. Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version).
- Fukuta, S., Nomura, M., Ikeda, T., Yoshizawa, M., Yamasaki, M., Sasaki, Y., 2016. UV laser machining of wood. *Eur. J. wood wood Prod.* 74, 261–267.
- Gabris, J.M., 2025. Zero Tolerance for Environmental Crimes. *Ecol. Integr. Int. Law Peace, Public Heal. Glob. Secur.*
- Garnier, S., Ross, N., Rudis, R., Camargo, P.A., Sciaiini, M., Scherer, C., 2024. viridis(Lite) - Colorblind-Friendly Color Maps for R. <https://doi.org/doi:10.5281/zenodo.4679423>
- Gasson, P., 2011. How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. *IAWA J.* 32, 137–154.
- Gasson, P., Baas, P., Wheeler, E., 2011. Wood anatomy of CITES-listed tree species. *IAWA J.* 32, 155–198.
- Gasson, P.E., Lancaster, C.A., Young, R., Redstone, S., Miles-Bunch, I.A., Rees, G., Guillery, R.P., Parker-Forney, M., Lebow, E.T., 2021. WorldForestID: Addressing the need for standardized wood reference collections to support authentication analysis technologies; a way forward for checking the origin and identity of traded timber. *Plants, People, Planet* 3, 130–141.
- Geifman, Y., El-Yaniv, R., 2019. Selectivenet: A deep neural network with an integrated reject option, in: International Conference on Machine Learning. PMLR, pp. 2151–2159.
- Geng, C., Huang, S., Chen, S., 2020. Recent advances in open set recognition: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3614–3631.
- Ghosh, A., Shanmugalingam, K., Lin, W.-Y., 2023. Relation preserving triplet mining for stabilising the triplet loss in re-identification systems, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4840–4849.
- Gibbs, C., Boratto, R., 2017. Environmental crime, in: Oxford Research Encyclopedia of Criminology and Criminal Justice.
- Global Land Cover 2000 database. European Commission, Joint Research Centre, 2003.
- Goodell, B., Nielsen, G., 2023. Wood Biodeterioration, in: Springer Handbook of Wood Science and Technology. Springer, pp. 139–177.
- Gorodissky, H., Harari, D., Ullman, S., 2018. Large Field and High Resolution: Detecting Needle in Haystack, *Journal of Vision*. <https://doi.org/10.1167/18.10.517>
- Government of Canada, 2025. Wild Animal and Plant Protection and Regulation of International and Interprovincial Trade Act. [WWW Document]. URL <https://lois-laws.justice.gc.ca/eng/acts/W-8.5/page-1.html#h-468887> (accessed 8.18.25).
- Gregory, M., 1980. Wood identification: an annotated bibliography. *Iawa J.* 1, 3–41.
- Gufran, D., Tiku, S., Pasricha, S., 2023. VITAL: Vision transformer neural networks for accurate smartphone heterogeneity resilient indoor localization, in: 2023 60th ACM/IEEE Design Automation Conference (DAC). IEEE, pp. 1–6.

- Guo, K., Lovell, B.C., 2024. Domain-aware triplet loss in domain generalization. *Comput. Vis. Image Underst.* 243, 103979.
- Guo, X., Deng, M., Hu, Y., Wang, Y., Ye, T., 2021. Morphology, mechanism and kerf variation during CO₂ laser cutting pine wood. *J. Manuf. Process.* 68, 13–22.
- Haji, S.H., Abdulazeez, A.M., 2021. Comparison of optimization techniques based on gradient descent algorithm: A review. *PalArch's J. Archaeol. Egypt/Egyptology* 18, 2715–2743.
- Halevy, A., Norvig, P., Pereira, F., 2009. The unreasonable effectiveness of data. *IEEE Intell. Syst.* 24, 8–12.
- Hanin, B., 2018. Which neural net architectures give rise to exploding and vanishing gradients? *Adv. Neural Inf. Process. Syst.* 31.
- Hao, W., Yizhou, W., Yaqin, L., Zhili, S., 2020. The role of activation function in CNN, in: 2020 2nd International Conference on Information Technology and Computer Application (ITCA). IEEE, pp. 429–432.
- He, S., Luo, H., Chen, W., Zhang, M., Zhang, Y., Wang, F., Li, H., Jiang, W., 2020. Multi-domain learning and identity mining for vehicle re-identification, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 582–583.
- He, T., Marco, J., Soares, R., Yin, Y., Wiedenhoef, A.C., 2019. Machine learning models with quantitative wood anatomy data can discriminate between *Swietenia macrophylla* and *Swietenia mahagoni*. *Forests* 11, 36.
- Helmling, S., Olbrich, A., Heinz, I., Koch, G., 2018. Atlas of vessel elements: identification of Asian timbers. *Iawa J.* 39, 249–352.
- Hermans, A., Beyer, L., Leibe, B., 2017. In defense of the triplet loss for person re-identification. *arXiv Prepr. arXiv1703.07737*.
- Hirschberger, P., 2008. Illegal wood for the European market: an analysis of the EU import and export of illegal wood and related products. WWF-Germany.
- Ho, T.K., 1998. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 832–844. <https://doi.org/10.1109/34.709601>
- Hoare, A., 2015. Tackling illegal logging and the related trade. *What Prog. where next* 79.
- Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.* 6, 107–116.
- Hong, Z., Wu, Z., Zhao, K., Yang, Z., Zhang, N., Guo, J., Tembrock, L.R., Xu, D., 2020. Comparative analyses of five complete chloroplast genomes from the genus *Pterocarpus* (Fabaceae). *Int. J. Mol. Sci.* 21, 3758.
- Horacek, M., Jakusch, M., Krehan, H., 2009. Control of origin of larch wood: discrimination between European (Austrian) and Siberian origin by stable isotope analysis. *Rapid Commun. Mass Spectrom. An Int. J. Devoted to Rapid Dissem. Up-to-the-Minute Res. Mass Spectrom.* 23, 3688–3692.
- Hourticq, J., Carole Megevand, C., 2013. Dynamiques de déforestation dans le bassin du Congo Réconcilier la croissance économique et la protection de la forêt.
- Hubau, W., Lewis, S.L., Phillips, O.L., Affum-Baffoe, K., Beeckman, H., Cuní-Sanchez, A., Daniels, A.K., Ewango, C.E.N., Fauset, S., Mukinzi, J.M., 2020. Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* 579, 80–87.
- Hunt, E.B., 2014. Artificial intelligence. Academic Press.
- Hwang, S.-W., Sugiyama, J., 2021. Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review. *Plant Methods* 17, 1–21.
- Ibrahim, A.A., Ridwan, R.L., Muhammed, M.M., Abdulaziz, R.O., Saheed, G.A., 2020. Comparison of the

- CatBoost classifier with other machine learning methods. *Int. J. Adv. Comput. Sci. Appl.* 11.
- Ifo, S.A., Moutsambote, J.-M., Koubouana, F., Yoka, J., Ndzai, S.F., Bouetou-Kadilamio, L.N.O., Mampouya, H., Jourdain, C., Bocko, Y., Mantota, A.B., 2016. Tree species diversity, richness, and similarity in intact and degraded forest in the tropical rainforest of the Congo Basin: case of the forest of Likouala in the Republic of Congo. *Int. J. For. Res.* 2016, 7593681.
- IIIF, 2022a. International Image Interoperability Framework [WWW Document]. URL <https://iiif.io/> (accessed 9.6.22).
- IIIF, 2022b. Mirador [WWW Document]. URL <https://projectmirador.org/> (accessed 9.6.22).
- Ilic, J., 1993. Computer aided wood identification using CSIROID. *IAWA J.* 14, 333–340.
- Inatimi, S.A., 2023. The need to conserve and protect forest resources: African perspective, in: *Sustainable Utilization and Conservation of Africa's Biological Resources and Environment*. Springer, pp. 203–233.
- ITTO, 2021. Biennial review and assessment of the world timber situation 2019–2020.
- IUCN, 2021. The IUCN Red List of Threatened Species. Version 2021-3 [WWW Document]. URL <https://www.iucnredlist.org/search?dl=true&permalink=d037e92e-6c37-4d09-be01-42bfa8f9dd7f> (accessed 2.22.22).
- Jackson, C.H., Sharples, L.D., Thompson, S.G., 2010. Structural and parameter uncertainty in Bayesian cost-effectiveness models. *J. R. Stat. Soc. Ser. C Appl. Stat.* 59, 233–253.
- Jacobs, R.L., Baker, B.W., 2018. The species dilemma and its potential impact on enforcing wildlife trade laws. *Evol. Anthropol. Issues, News, Rev.* 27, 261–266.
- Jahanbanifard, M., Price, E., González, B.A., Raggi, L.A., Javanmardi, S., Lens, F., Gravendeel, B., Espinoza, E., Verbeek, F.J., 2023. A novel method to analyse DART TOFMS spectra based on Convolutional Neural Networks: A case study on methanol extracts of wool fibres from endangered camelids. *Int. J. Mass Spectrom.* 489, 117050.
- James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., 2023. Unsupervised learning, in: *An Introduction to Statistical Learning: With Applications in Python*. Springer, pp. 503–556.
- Jansen, S., Baas, P., Smets, E., 2001. Vestured pits: their occurrence and systematic importance in eudicots. *Taxon* 50, 135–167. <https://doi.org/https://doi.org/10.2307/1224516>
- JANSEN, S., KITIN, P., DE PAUW, H., IDRIS, M., BEECKMAN, H., SMETS, E., 1998. PREPARATION OF WOOD SPECIMENS FOR TRANSMITTED LIGHT MICROSCOPY AND SCANNING ELECTRON MICROSCOPY. *Belgian J. Bot.* 131, 41–49.
- Jansen, S., Smets, E., 1998. Vestured pits in some woody Gentianaceae. *IAWA J.* 19, 35–42.
- Jansen, S., Smets, E., Nordenstam, B., ElGhazaly, G., Kassas, M., 2000. Morphology, distribution and systematic importance of vestures in the Gentianales, in: *PLANT SYSTEMATICS FOR THE 21ST CENTURY*. PORTLAND PRESS LTD, pp. 277–296.
- Jia, X., Shen, L., 2017. Skin Lesion Classification using Class Activation Map. <https://doi.org/10.48550/arXiv.1703.01053>
- Jian-Wei, L., Hui-Dan, Z., Xiong-Lin, L., Jun, X., 2020. Research progress on batch normalization of deep learning and its related algorithms. *Acta Autom. Sin.* 46, 1090–1120.
- Jiang, Y., Li, X., Luo, H., Yin, S., Kaynak, O., 2022. Quo vadis artificial intelligence? *Discov. Artif. Intell.* 2, 4.
- Jiao, L., Lu, Y., He, T., Guo, J., Yin, Y., 2020. DNA barcoding for wood identification: Global review of the last decade and future perspective. *IAWA J.* 41, 620–643.
- Jiao, L., Lu, Y., He, T., Li, J., Yin, Y., 2019. A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus*

- species. *Planta* 250, 95–104.
- Jiao, L., Yin, Y., Xiao, F., Sun, Q., Song, K., Jiang, X., 2012. Comparative analysis of two DNA extraction protocols from fresh and dried wood of *Cunninghamia lanceolata* (Taxodiaceae). *Iawa J.* 33, 441–456.
- Joachims, T., 2002. Support Vector Machines BT - Learning to Classify Text Using Support Vector Machines, in: Joachims, T. (Ed.), . Springer US, Boston, MA, pp. 35–44. https://doi.org/10.1007/978-1-4615-0907-3_3
- Johann Heinrich von Thünen-Institut, 2025a. Thünen Centre of Competence on the Origin of Timber [WWW Document]. URL <https://www.thuenen.de/en/thuenen-institute/compound-structures/thuenen-kompetenzzentrum-holzherkuenfte> (accessed 4.24.25).
- Johann Heinrich von Thünen-Institut, 2025b. The scientific wood collection (Xylothek) of the Thünen Institute in Hamburg [WWW Document]. URL <https://www.thuenen.de/en/thuenen-institute/infrastructure/the-thuenen-centre-of-competence-on-the-origin-of-timber/the-scientific-wood-collection-xylothek-1> (accessed 4.24.25).
- 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.
- Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. *J. big data* 6, 1–54.
- Jonsson, R., Giurca, A., Masiero, M., Pepke, E., Pettenella, D., Prestemon, J.P., Winkel, G., 2015. Assessment of the EU timber regulation and FLEGT action plan: From Sci. to policy.
- Jumabek, A., Yang, S., Noh, Y., 2021. CatBoost-based network intrusion detection on imbalanced CIC-IDS-2018 dataset. *한국통신학회논문지* 46, 2191–2197.
- Kabulu Djibu, J.P., Bamba, I., Munyemba Kankumbi, F., Defourny, P., Vancutsem, C., Nyembwe, N.S., Ngongo, M.L., Bogaert, J., 2008. Analyse de la structure spatiale des forêts au Katanga. *Ann. la Fac. des Sci. Agron. l'Université Lubumbashi* 1.
- Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996. Reinforcement learning: A survey. *J. Artif. Intell. Res.* 4, 237–285.
- Kagawa, A., Leavitt, S.W., 2010. Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin of timber. *J. Wood Sci.* 56, 175–183.
- Kala, A., McCollum, E.D., Elhilali, M., 2022. Implications of clinical variability on computer-aided lung auscultation classification, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 4421–4425.
- Kazadi, S., Kaoru, F., 1996. Interannual and long-term climate variability over the Zaire River Basin during the last 30 years. *J. Geophys. Res. Atmos.* 101, 21351–21360.
- Kirchoff, B., Remington, D., Fu, L., Sadri, F., 2008. A new type of image-based key, in: 2008 International Conference on BioMedical Engineering and Informatics. IEEE, pp. 825–829.
- Kirchoff, B.K., Leggett, R., Her, V., Moua, C., Morrison, J., Poole, C., 2011. Principles of visual key construction—with a visual identification key to the Fagaceae of the southeastern United States. *AoB Plants* 2011, plr005.
- Knauer, U., von Rekowski, C.S., Stecklina, M., Krokotsch, T., Pham Minh, T., Hauffe, V., Kilias, D., Ehrhardt, I., Sagischewski, H., Chmara, S., 2019. Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers. *Remote Sens.* 11, 2788.
- Kobayashi, K., Kegasa, T., Hwang, S.-W., Sugiyama, J., 2019. Anatomical features of Fagaceae wood statistically extracted by computer vision approaches: Some relationships with evolution. *PLoS One* 14, e0220762.
- Koch, G., Haag, V., Heinz, I., Richter, H.-G., Schmitt, U., 2015. Control of internationally traded timber-the role of macroscopic and microscopic wood identification against illegal logging. *J. Forensic Res* 6,

1000317.

- Koch, G., Heinz, I., Schmitt, U., Richter, H.-G., 2018. Wood anatomy-the role of macroscopic and microscopic wood identification against illegal logging, in: 8 TH HARDWOOD CONFERENCE - NEW ASPECTS OF HARDWOOD UTILIZATION - FROM SCIENCE TO TECHNOLOGY. 8 TH HARDWOOD CONFERENCE - NEW ASPECTS OF HARDWOOD UTILIZATION - FROM SCIENCE TO TECHNOLOGY, Sopron, Austria, pp. 10–11.
- Koch, G., Richter, H.-G., Schmitt, U., 2011. Design and application of CITESwoodID computer-aided identification and description of CITES-protected timbers. *IAWA J.* 32, 213–220.
- Kokilepersaud, K., Yarici, Y., Prabhushankar, M., AlRegib, G., 2024. Taxes are all you need: Integration of taxonomical hierarchy relationships into the contrastive loss, in: 2024 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 270–276.
- Kolde, R., 2018. pheatmap: Pretty Heatmaps. R package.
- Köthke, M., Lippe, M., Elsasser, P., 2023. Comparing the former EUTR and upcoming EUDR: Some implications for private sector and authorities. *For. Policy Econ.* 157, 103079.
- Kumar, R., Weill, E., Aghdasi, F., Sriram, P., 2020. A strong and efficient baseline for vehicle re-identification using deep triplet embedding. *J. Artif. Intell. Soft Comput. Res.* 10, 27–45.
- La Grassa, R., Gallo, I., Landro, N., 2021. Learn class hierarchy using convolutional neural networks. *Appl. Intell.* 51, 6622–6632.
- Lampinen, J., Vehtari, A., 2001. Bayesian approach for neural networks—review and case studies. *Neural networks* 14, 257–274.
- LaPasha, C.A., Wheeler, E.A., 1987. A microcomputer based system for computer-aided wood identification. *IAWA J.* 8, 347–354.
- Lawson, S., 2014. Illegal logging in the Democratic Republic of the Congo. *Energy, Environ. Resour. EER* 2014.
- Leggett, R., Kirchoff, B.K., 2011. Image use in field guides and identification keys: review and recommendations. *AoB Plants* 2011, plr004.
- Lewis, S.L., Edwards, D.P., Galbraith, D., 2015. Increasing human dominance of tropical forests. *Science* (80-.). 349, 827–832.
- Li, L., Wu, Y., Ye, M., 2015. Experimental comparisons of multi-class classifiers. *Informatica* 39.
- Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. neural networks Learn. Syst.* 33, 6999–7019.
- Liao, Y., Gao, Y., Zhang, W., 2023. Feature activation map: Visual explanation of deep learning models for image classification. *arXiv Prepr. arXiv2307.05017*.
- Lin, C., Jiao, L., Lu, Y., Yin, Y., 2024. BOOK OF ABSTRACTS: FORESTS & SOCIETY TOWARDS 2050, in: 26th IUFRO world Congress (Ed.), T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification: Origin Traceability of Santalum Wood Using Stable Isotope and Mineral Elements. 26th IUFRO world Congress, Stockholm, Sweden, p. 3527.
- Lirëza, L., Koçi, G., 2023. Environmental crimes: Their nature, scope, and problems in identification. *Interdiscip. J. Res. Dev.* 10, 237.
- Liu, B., Chen, Q., Tang, L., Zhu, L., Zou, X., Li, B., Fan, W., Fu, Y., Lu, Y., 2023. Screening of potential chemical marker with interspecific differences in *Pterocarpus* wood and a spatially-resolved approach to visualize the distribution of the characteristic markers. *Front. Plant Sci.* 14, 1133848.
- Liu, M., Chen, L., Du, X., Jin, L., Shang, M., 2021. Activated gradients for deep neural networks. *IEEE Trans. Neural Networks Learn. Syst.* 34, 2156–2168.

- Liu, T.Y., Mirzasoleiman, B., 2022. Data-efficient augmentation for training neural networks. *Adv. Neural Inf. Process. Syst.* 35, 5124–5136.
- Liu, Z., Lian, T., Farrell, J., Wandell, B.A., 2020. Neural network generalization: The impact of camera parameters. *IEEE Access* 8, 10443–10454.
- Lowe, A.J., Dormontt, E.E., Bowie, M.J., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoef, A., Yin, Y., 2016. Opportunities for improved transparency in the timber trade through scientific verification. *Bioscience* 66, 990–998.
- Lowe, A.J., Wong, K.N., Tiong, Y.S., Iyerh, S., Chew, F.T., 2010. A DNA Method to Verify the Integrity of Timber Supply Chains; Confirming the Legal Sourcing of Merbau Timber From Logging Concession to Sawmill. *Silvae Genet.* 59, 263.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110.
- Lu, Y., Jiao, L., Yin, Y., 2024. BOOK OF ABSTRACTS: FORESTS & SOCIETY TOWARDS 2050, in: 26th IUFRO world Congress (Ed.), T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification: New Advances in DNA-Based Methods for Modern and Archaeological Woods Identification in China. 26th IUFRO world Congress, Stockholm, Sweden, p. 3526.
- Magrath, W., Younger, P., Phan, H., 2009. An INTERPOL Perspective on Law Enforcement in Illegal Logging. *Lyon INTERPOL Gen. Secr.*
- Malaisse, F., 1997. *Se nourrir en foret claire africaine: approche écologique et nutritionnelle.*
- Malaysian Timber Council, 2018. MTC Wood Wizard [WWW Document]. URL <https://appadvice.com/app/mtc-wood-wizard/1373564891> (accessed 9.14.22).
- Mark, J., Newton, A., Oldfield, S., Rivers, M., 2014. A Working List of Commercial Timber Tree Species.
- Mayaux, P., Pekel, J.-F., Desclée, B., Donnay, F., Lupi, A., Achard, F., Clerici, M., Bodart, C., Brink, A., Nasi, R., 2013. State and evolution of the African rainforests between 1990 and 2010. *Philos. Trans. R. Soc. B Biol. Sci.* 368, 20120300.
- Mboma, R., Chevillotte, H., Doumenge, C., PP, L.I.I., 2022. Lack of Coherence Between Commercial and Scientific Names for Timber Species: Implications for the Sustainable Exploitation of Andoung in Gabon.
- McAulay, L., 2017. Benefits of the International Image Interoperability Framework (IIIF) Featuring Medieval Palimpsest Manuscripts.
- McCarthy, J., Hayes, P.J., 1981. Some philosophical problems from the standpoint of artificial intelligence, in: *Readings in Artificial Intelligence*. Elsevier, pp. 431–450.
- Mediavilla-Relaño, J., Lázaro, M., Figueiras-Vidal, A.R., 2023. Imbalance example-dependent cost classification: A Bayesian based method. *Expert Syst. Appl.* 213, 118909.
- Michael Höltnen, A., Schröder, H., Wischnewski, N., Degen, B., Magel, E., Fladung, M., 2012. Development of DNA-based methods to identify CITES-protected timber species: a case study in the Meliaceae family.
- Mironov, A., Khuziev, I., 2022. Optimization of Oblivious Decision Tree Ensembles Evaluation for CPU. *arXiv Prepr. arXiv2211.00391*.
- Mishler, B.D., 2000. Deep phylogenetic relationships among “plants” and their implications for classification. *Taxon* 49, 661–683.
- Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. *Carbon Balance Manag.* 12, 1–22.

- Miyata, T., Yanagawa, M., Hata, A., Honda, O., Yoshida, Y., Kikuchi, N., Tsubamoto, M., Tsukagoshi, S., Uranishi, A., Tomiyama, N., 2020. Influence of field of view size on image quality: ultra-high-resolution CT vs. conventional high-resolution CT. *Eur. Radiol.* 30, 3324–3333.
- Mulvagh, L., 2006. The impact of commercial logging and forest policy on indigenous peoples in the Democratic Republic of Congo. For. Peoples Program.
- Münkner, C.A., Bouquet, M., Muakana, R., 2015. Elaboration du schéma directeur d’approvisionnement durable en bois-énergie pour la ville de Lubumbashi (Katanga). Program. Biodivers. ForSts Proj. Filiere Bois/Chaînes Val. ajoutée, GIZ, MEDD, DFS, GFA.
- Murthy, S.K., Kasif, S., Salzberg, S., 1994. A system for induction of oblique decision trees. *J. Artif. Intell. Res.* 2, 1–32.
- Nanni, L., Paci, M., Brahnam, S., Lumini, A., 2021. Comparison of different image data augmentation approaches. *J. imaging* 7, 254.
- Naturalis Biodiversity center, 2025. The power of comparative wood anatomy [WWW Document]. URL <https://www.naturalis.nl/en/science/power-of-comparative-wood-anatomy> (accessed 3.26.25).
- Nellemann, C., INTERPOL Environmental Crime Programme (eds), 2012. Green Carbon, Black Trade: Illegal Logging, Tax Fraud and Laundering in the Worlds Tropical Forests. A Rapid Response Assessment. United Nations Environment Programme, GRID_Arendal. <https://doi.org/978-82-7701-102-8>
- Niemz, P., Sonderegger, W., Keplinger, T., Jiang, J., Lu, J., 2023a. Physical Properties of Wood and Wood-Based Materials, in: Springer Handbook of Wood Science and Technology. Springer, pp. 281–353.
- Niemz, P., Teischinger, A., Sandberg, D., 2023b. Springer handbook of wood science and technology. Springer.
- Nieradzik, L., Sieburg-Rockel, J., Helmling, S., Keuper, J., Weibel, T., Olbrich, A., Stephani, H., 2024. Automating wood species detection and classification in microscopic images of fibrous materials with deep learning. *Microsc. Microanal.* 30, 508–520.
- NS, I.B., 1989. IAWA list of microscopic features for hardwood identification.
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F., Weedon, J., Borman, T., 2025. vegan: Community Ecology Package Ordination methods, diversity analysis and other functions for community and vegetation ecologists. <https://doi.org/10.32614/CRAN.package.vegan>
- Owens, F.C., Ravindran, P., Costa, A., Chavesta, M., Montenegro, R., Shmulsky, R., Wiedenhoeft, A.C., 2024. Robustness of a macroscopic computer-vision wood identification model to digital perturbations of test images. *IAWA J.* 1–16. <https://doi.org/https://doi.org/10.1163/22941932-bja10167>
- Parliament, E., 2023. Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and r. *J. Eur. Union* 150, 206–247.
- Partnership., C.B.F., 2005. The forests of the Congo Basin: a preliminary assessment.
- Pérez, M.R., de Blas, D.E., Nasi, R., Sayer, J.A., Sassen, M., Angoué, C., Gami, N., Ndoye, O., Ngono, G., Nguinguiri, J.-C., 2005. Logging in the Congo Basin: a multi-country characterization of timber companies. *For. Ecol. Manage.* 214, 221–236.
- Pes, B., 2021. Learning from high-dimensional and class-imbalanced datasets using random forests. *Information* 12, 286.

- Piabuo, S.M., Minang, P.A., Tieguhong, C.J., Foundjem-Tita, D., Nghobuoche, F., 2021. Illegal logging, governance effectiveness and carbon dioxide emission in the timber-producing countries of Congo Basin and Asia. *Environ. Dev. Sustain.* 23, 14176–14196.
- Potapov, P. V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstadt, A., Mane, L., Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. *Remote Sens. Environ.* 122, 106–116.
- Prechelt, L., 2002. Early stopping-but when?, in: *Neural Networks: Tricks of the Trade*. Springer, pp. 55–69.
- Price, E.R., McClure, P.J., Huffman, A.N., Voin, D., Espinoza, E.O., 2022. Reliability of wood identification using DART-TOFMS and the ForeST© database: A validation study. *Forensic Sci. Int. Anim. Environ.* 2, 100045. <https://doi.org/https://doi.org/10.1016/j.fsiae.2022.100045>
- Price, E.R., Miles-Bunch, I.A., Gasson, P.E., Lancaster, C.A., 2021. Pterocarpus wood identification by independent and complementary analysis of DART-TOFMS, microscopic anatomy, and fluorescence spectrometry. *IAWA J.* 42, 397–418.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* 31.
- Quinlan, J.R., 1987. DECISION TREES AS PROBABILISTIC CLASSIFIERS, in: Langley, P.B.T.-P. of the F.I.W. on M.L. (Ed.), . Morgan Kaufmann, pp. 31–37. <https://doi.org/https://doi.org/10.1016/B978-0-934613-41-5.50007-6>
- Ravindran, P., Costa, A., Soares, R., Wiedenhoeft, A.C., 2018. Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14, 1–10.
- Ravindran, P., Ebanyenle, E., Ebeheakey, A.A., Abban, K.B., Lambog, O., Soares, R., Costa, A., Wiedenhoeft, A.C., 2019. Image based identification of Ghanaian timbers using the XyloTron: Opportunities, risks and challenges. *arXiv Prepr. arXiv1912.00296*.
- Ravindran, P., Owens, F.C., Costa, A., Rodrigues, B.P., Chavesta, M., Montenegro, R., Shmulsky, R., Wiedenhoeft, A.C., 2023. EVALUATION OF TEST SPECIMEN SURFACE PREPARATION ON MACROSCOPIC COMPUTER VISION WOOD IDENTIFICATION. *WOOD FIBER Sci.* 55, 176–202. <https://doi.org/10.22382/wfs-2023-15>
- Ravindran, P., Owens, F.C., Wade, A.C., Vega, P., Montenegro, R., Shmulsky, R., Wiedenhoeft, A.C., 2021. Field-deployable computer vision wood identification of Peruvian timbers. *Front. Plant Sci.* 12, 647515.
- Ravindran, P., Thompson, B.J., Soares, R.K., Wiedenhoeft, A.C., 2020. The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. *Front. Plant Sci.* 11, 1015.
- Ravindran, P., Wiedenhoeft, A.C., 2022. Caveat emptor: on the need for baseline quality standards in computer vision wood identification. *Forests* 13, 632.
- Réjou-Méchain, M., Mortier, F., Bastin, J.-F., Cornu, G., Barbier, N., Bayol, N., Bénédet, F., Bry, X., Dauby, G., Deblauwe, V., 2021. Unveiling African rainforest composition and vulnerability to global change. *Nature* 593, 90–94.
- Richter, H.G., Dallwitz, M., 2000. Commercial timbers: descriptions, illustrations, identification, and information retrieval. English, French, Ger. Spanish. Version 4th May.
- Richter, H.G., Oelker, M., Koch, G., 2017. macroHOLZdata: descriptions, illustrations, identification, and information retrieval. English Ger. Version.
- RMCA, 2019. Tervuren Xylarium Wood Database.
- Rosa da Silva, N., De Ridder, M., Baetens, J.M., Van den Bulcke, J., Rousseau, M., Martinez Bruno, O., Beeckman, H., Van Acker, J., De Baets, B., 2017. Automated classification of wood transverse cross-section micro-imagery from 77 commercial Central-African timber species. *Ann. For. Sci.* 74, 30. <https://doi.org/10.1007/s13595-017-0619-0>

- Rosa da Silva, N., Deklerck, V., Baetens, J.M., Van den Bulcke, J., De Ridder, M., Rousseau, M., Bruno, O.M., Beeckman, H., Van Acker, J., De Baets, B., 2022. Improved wood species identification based on multi-view imagery of the three anatomical planes. *Plant Methods* 18, 79.
- Ruffinatto, F., Castro, G., Cremonini, C., Crivellaro, A., Zanuttini, R., 2019. A new atlas and macroscopic wood identification software package for Italian timber species. *IAWA J.* 41, 393–411.
- 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.
- Ruffinatto, F., Crivellaro, A., Wiedenhoef, A.C., 2015. Review of macroscopic features for hardwood and softwood identification and a proposal for a new character list. *IAWA J.* 36, 208–241.
- Salman, H.A., Kalakech, A., Steiti, A., 2024. Random Forest Algorithm Overview. *Babylonian J. Mach. Learn.* 2024, 69–79.
- Sanlı, T., Sıcakyüz, Ç., Yüregir, O.H., 2020. Comparison of the accuracy of classification algorithms on three data-sets in data mining: Example of 20 classes. *Int. J. Eng. Sci. Technol.* 12, 81–89.
- Santos, C.F.G. Dos, Papa, J.P., 2022. Avoiding overfitting: A survey on regularization methods for convolutional neural networks. *ACM Comput. Surv.* 54, 1–25.
- Sarker, I.H., 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput. Sci.* 2, 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boulton, T.E., 2012. Toward open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1757–1772.
- Schmidt, O., 2006. Wood and tree fungi. Springer.
- Schmitz, N., Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J.W.B., Cabezas, J.-A., Chaix, G., Crameri, S., Degen, B., Deklerck, V., 2020. Overview of current practices in data analysis for wood identification-A guide for the different timber tracking methods.
- Schmitz, N., Beeckman, H., Cabezas Martínez, J.A., Cervera, M.T., Espinoza, E., Fernández-Golfín, J.I., Gasson, P., Hermanson, J.C., Arteaga, M.J., Koch, G., 2019. The Timber Tracking Tool Infogram. Overview of wood identification methods' capacity. *Inst. Ciencias For.*
- Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 815–823.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* 128, 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Shapiro, A.C., Grantham, H.S., Aguilar-Amuchastegui, N., Murray, N.J., Gond, V., Bonfils, D., Rickenbach, O., 2021. Forest condition in the Congo Basin for the assessment of ecosystem conservation status. *Ecol. Indic.* 122, 107268.
- Shen, F., Du, X., Zhang, L., Shu, X., Tang, J., n.d. Triplet contrastive representation learning for unsupervised vehicle re-identification. *ACM Trans. Multimed. Comput. Commun. Appl.*
- Shi, H., Yang, Y., Zhu, X., Liao, S., Lei, Z., Zheng, W., Li, S.Z., 2016. Embedding deep metric for person re-identification: A study against large variations, in: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* 14. Springer, pp. 732–748.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. big data* 6, 1–48.
- Silva, J.L., Bordalo, R., Pissarra, J., de Palacios, P., 2022. Computer Vision-Based Wood Identification: A Review. *Forests*. <https://doi.org/10.3390/f13122041>

- Souza, D.V., Santos, J.X., Vieira, H.C., Naide, T.L., Nisgoski, S., Oliveira, L.E.S., 2020. An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. *Wood Sci. Technol.* 54, 1065–1090.
- Spiecker, H., Schinker, M.G., Hansen, J., Park, Y.-I., Ebding, T., Döll, W., 2000. Cell structure in tree rings: novel methods for preparation and image analysis of large cross sections. *IAWA J.* 21, 361–373.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Stackpole, D.J., Vaillancourt, R.E., Alves, A., Rodrigues, J., Potts, B.M., 2011. Genetic variation in the chemical components of *Eucalyptus globulus* wood. *G3 Genes| Genomes| Genet.* 1, 151–159.
- Stokstad, E., 2014. The empty forest.
- Sugiarto, B., Prakasa, E., Wardoyo, R., Damayanti, R., Dewi, L.M., Pardede, H.F., Rianto, Y., 2017. Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier, in: 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, pp. 337–341.
- Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., Upcroft, B., Abbeel, P., Burgard, W., Milford, M., 2018. The limits and potentials of deep learning for robotics. *Int. J. Rob. Res.* 37, 405–420.
- Sutton, R.S., Barto, A.G., 1999. Reinforcement learning. *J. Cogn. Neurosci.* 11, 126–134.
- Szyc, K., 2020. Determining the minimal number of images required to effectively train convolutional neural networks, in: Theory and Applications of Dependable Computer Systems: Proceedings of the Fifteenth International Conference on Dependability of Computer Systems DepCoS-RELCOMEX, June 29–July 3, 2020, Brunów, Poland 15. Springer, pp. 652–661.
- Tacconi, L., 2012. *Illegal Logging: Law Enforcement, Livelihoods and the Timber Trade*. Routledge.
- Tacconi, L., Cerutti, P.O., Leipold, S., Rodrigues, R.J., Savaresi, A., To, P.X., Weng, X., 2016. Defining illegal forest activities and illegal logging.
- Talebi, H., Milanfar, P., 2021. Learning to resize images for computer vision tasks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 497–506.
- Tang, C., Zhu, Q., Wu, W., Huang, W., Hong, C., Niu, X., 2020. PLANET: improved convolutional neural networks with image enhancement for image classification. *Math. Probl. Eng.* 2020, 1245924.
- Tang, X.J., Tay, Y.H., 2019. O-12: Xylorix: An AI-as-a-service platform for wood identification, in: Program & Abstract of the IAWA-IUFRO International Symposium on Challenges and Opportunities for Updating Wood Identification. IAWA/IUFRO/CAF. pp. 31–32.
- Tang, X.J., Tay, Y.H., Siam, N.A., 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. pp. 37–43.
- Tang, Z., Naphade, M., Birchfield, S., Tremblay, J., Hodge, W., Kumar, R., Wang, S., Yang, X., 2019. Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 211–220.
- Tardif, J.C., Conciatori, F., 2015. Microscopic examination of wood: Sample preparation and techniques for light microscopy. *Plant Microtech. Protoc.* 373–415.
- Taye, M.M., 2023. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*. <https://doi.org/10.3390/computers12050091>
- Tegegne, Y.T., Cramm, M., Van Brusselen, J., 2018. Sustainable forest management, FLEGT, and REDD+: exploring interlinkages to strengthen forest policy coherence. *Sustainability* 10, 4841.

- Thoidis, I., Gaultier, C., Goehring, T., 2023. Perceptual analysis of speaker embeddings for voice discrimination between machine and human listening, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1–5.
- Thompson, S.T., Magrath, W.B., 2021. Preventing illegal logging. *For. Policy Econ.* 128, 102479.
- Thünen Institute of Forest Genetics, 2015. Development and implementation of a species identification and timber tracking system with DNA fingerprints and isotopes in Africa (ITTO project PD 620/11 Rev.1 (M)). Hamburg, Germany.
- tropicaux, A. technique internationale des bois, 1979. Nomenclature générale des bois tropicaux. Centre technique forestier tropical.
- Tsuchikawa, S., Inoue, K., Noma, J., Hayashi, K., 2003. Application of near-infrared spectroscopy to wood discrimination. *J. Wood Sci.* 49, 29–35.
- Tsuchikawa, S., Kobori, H., 2015. A review of recent application of near infrared spectroscopy to wood science and technology. *J. Wood Sci.* 61, 213–220.
- UNEP-WCMC (Comps.), 2022. The Checklist of CITES Species Website [WWW Document]. CITES Secr. Geneva, Switzerland. Compil. by UNEP-WCMC, Cambridge, UK. URL <http://checklist.cites.org> (accessed 8.12.22).
- Union, E., 2010. Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market. *Off. J. L* 295, 23–34.
- University of Wisconsin–Madison, 2025. USDA Forest Service Xylarium [WWW Document]. URL <https://herbarium.wisc.edu/collections/wood/> (accessed 3.26.25).
- Ustinova, E., Lempitsky, V., 2016. Learning deep embeddings with histogram loss. *Adv. Neural Inf. Process. Syst.* 29.
- Van Brusselen, J., Cramm, M., Tegegne, Y.T., 2023. Wood identification services in support of legal supply chains: A market study. *Sustain. Futur.* 6, 100128.
- Van den Bulcke, J., Boone, M., Van Acker, J., Stevens, M., Van Hoorebeke, L., 2009. X-ray tomography as a tool for detailed anatomical analysis. *Ann. For. Sci.* 66, 1–12.
- Van den Bulcke, J., Verschuren, L., De Blaere, R., Vansuyt, S., Dekegeleer, M., Kibleur, P., Pieters, O., De Mil, T., Hubau, W., Beeckman, H., 2025. Enabling high-throughput quantitative wood anatomy through a dedicated pipeline. *Plant Methods* 21, 11.
- van Uhm, D., 2024. Green Criminology and Biodiversity Loss: Crimes and Harms against Flora and Fauna, in: *Oxford Research Encyclopedia of Criminology and Criminal Justice*.
- Vander Mijnsbrugge, K., Beeckman, H., 1992. Knowledge modelling for a wood identification system. *Silva Gandav.* 57.
- von Arx, G., Crivellaro, A., Prendin, A.L., Čufar, K., Carrer, M., 2016. Quantitative Wood Anatomy—Practical Guidelines. *Front. Plant Sci.* 7.
- Wang, S.H., Zhang, Y.D., 2023. Grad-CAM: Understanding AI Models. *C. Mater. Contin.* 76, 1321–1324. <https://doi.org/10.32604/cmc.2023.041419>
- Wanga, H., Pengb, L., Li Changc, Z.L., Guoa, Y., Lid, Q., Yanga, X., n.d. CatBoost-Based Framework for Intelligent Prediction and Reaction Condition Analysis of Coupling Reaction.
- Wani, M., Bhat, F., Afzal, S., Khan, A., 2020. Basics of Supervised Deep Learning. pp. 13–29. https://doi.org/10.1007/978-981-13-6794-6_2
- Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J., Dokken, D.J., 2000. Land use, land use change, and forestry.

- WCVP, 2022. The World Checklist of Vascular Plants, Version 8.0 [WWW Document]. R. Bot. Gard. Kew. URL <http://wcvp.science.kew.org/> (accessed 3.24.22).
- Weerth, C., 2024. Citeswoodid—a capable app based identification tool for law enforcement officers around the globe (cites listed timbers).
- Wheeler, E.A., 2011. Inside Wood—A web resource for hardwood anatomy. *Iawa J.* 32, 199–211.
- Wheeler, E.A., Baas, P., 1998. Wood identification—a review. *Iawa J.* 19, 241–264.
- Wheeler, E.A., Gasson, P.E., Baas, P., 2020. Using the InsideWood web site: potentials and pitfalls. *Iawa J.* 41, 412–462.
- White, F., 1983. The vegetation of Africa.
- White, R., 2018. Transnational environmental crime: Toward an eco-global criminology. Willan.
- Wiedenhoeft, A.C., 2020. The XyloPhone: toward democratizing access to high-quality macroscopic imaging for wood and other substrates. *Iawa J.* 41, 699–719.
- Wiedenhoeft, A.C., 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.
- Wilber, M.J., Scheirer, W.J., Leitner, P., Heflin, B., Zott, J., Reinke, D., Delaney, D.K., Boulton, T.E., 2013. Animal recognition in the mojave desert: Vision tools for field biologists, in: 2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE, pp. 206–213.
- Wiley, E.O., Lieberman, B.S., 2011. Phylogenetics: theory and practice of phylogenetic systematics. John Wiley & Sons.
- Winston, P.H., 1984. Artificial intelligence. Addison-Wesley Longman Publishing Co., Inc.
- Wodzicki, T.J., 2001. Natural factors affecting wood structure. *Wood Sci. Technol.* 35, 5–26.
- World Resources Institute, 2024. Illegal Logging Prohibition Amendment (Strengthening Measures to Prevent Illegal Timber Trade) Act 2024 [WWW Document]. URL <https://forestpolicy.org/policy-law/australia-illegal-logging-prohibition-act> (accessed 8.18.25).
- Wu, F., Gazo, R., Haviarova, E., Benes, B., 2021. Wood identification based on longitudinal section images by using deep learning. *Wood Sci. Technol.* 55, 553–563.
- Wu, J., 2017. Introduction to convolutional neural networks. *Natl. Key Lab Nov. Softw. Technol. Nanjing Univ. China* 5, 495.
- Xie, X., Xie, M., Moshayedi, A.J., Noori Skandari, M.H., 2022. A hybrid improved neural networks algorithm based on L2 and dropout regularization. *Math. Probl. Eng.* 2022, 8220453.
- Yan, Z., Jagadeesh, V., DeCoste, D., Di, W., Piramuthu, R., 2015. HD-CNN: Hierarchical deep convolutional neural network for image classification, in: International Conference on Computer Vision (ICCV). Citeseer, pp. 435–443.
- Yang, R., Xu, X., Xu, Z., Ding, C., Pu, F., 2019. A Class Activation Mapping Guided Adversarial Training Method for Land-Use Classification and Object Detection, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 9474–9477. <https://doi.org/10.1109/IGARSS.2019.8897938>
- Yang, Y., Ferguson, D.K., Liu, B., Mao, K.-S., Gao, L.-M., Zhang, S.-Z., Wan, T., Rushforth, K., Zhang, Z.-X., 2022. Recent advances on phylogenomics of gymnosperms and a new classification. *Plant Divers.* 44, 340–350.
- Ye, M., Chen, S., Li, C., Zheng, W.-S., Crandall, D., Du, B., 2024. Transformer for object re-identification: A survey. *Int. J. Comput. Vis.* 1–31.

- Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.H., 2021. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 2872–2893.
- Ying, X., 2019. An overview of overfitting and its solutions, in: *Journal of Physics: Conference Series*. IOP Publishing, p. 22022.
- Yoshihashi, R., Shao, W., Kawakami, R., You, S., Iida, M., Naemura, T., 2019. Classification-reconstruction learning for open-set recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4016–4025.
- Zhao, J., Lee, C.-D., Chen, G., Zhang, J., 2024. Research on the Prediction Application of Multiple Classification Datasets Based on Random Forest Model, in: *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*. IEEE, pp. 156–161.
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., Parmar, M., 2024. A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* 57, 99.
- Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y., 2020. Random erasing data augmentation, in: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 13001–13008.
- Zhu, X.J., 2005. Semi-supervised learning literature survey.
- Zimmennann, M.H., Potter, D., 1982. Vessel-length distribution in branches, stem and roots of *Acer rubrum* L. *IAWA Bull. new Ser.* 3, 103–109.

Chapter 8: Supplementary materials

8.1 CNNs for image classification

AI refers to the ability of computer systems to perform tasks that typically require human intelligence, such as pattern recognition, decision-making, problem-solving, and data analysis (Hunt, 2014; Winston, 1984). Over the past decades, AI has evolved to self-learning models that autonomously adapt (Chollet and Chollet, 2021). Deep learning, a subset of machine learning, has revolutionized AI applications by employing artificial neural networks. These models are computational models inspired by the structure and function of the human brain, consisting of interconnected layers of artificial neurons that process input data and learn patterns through iterative adjustments. Each neuron receives an input, applies a mathematical transformation, and passes the result to the next layer. Artificial neural networks can process diverse data types, including time-series data, text, images, and video material, allowing them to perform increasingly complex tasks such as image recognition, speech processing, and natural language understanding (Jiang et al., 2022). AI has been widely adopted across various fields, including environmental science, where it is used for biodiversity monitoring, deforestation detection, and climate

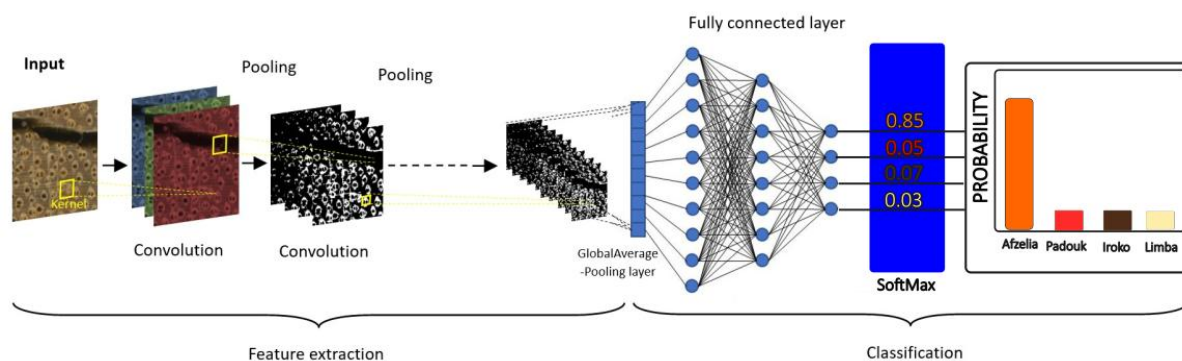


Figure 8.1: Schematic overview of a multiclass CNN for wood identification, showing feature extraction through convolutional and pooling layers, followed by classification into timber genera via fully connected layers with SoftMax activation for probabilistic outputs.

pattern prediction (Knauer et al., 2019).

CV, a subfield of AI, enables computers to interpret and analyse image data, extracting meaningful patterns that mimic human perception (Bay et al., 2006; Hwang and Sugiyama, 2021; Lowe, 2004). When combined with deep learning, CV has given rise to powerful approaches capable of automatically extracting complex visual features, enabling robust pattern recognition and classification across diverse image datasets (Chollet and Chollet, 2021). Among deep learning models, CNNs are particularly effective for learning from image data (Li et al., 2021; Tang et al., 2020; Wu, 2017; X. Zhao et al., 2024). Unlike traditional image analysis approaches that require manual feature selection, CNNs automate feature extraction by recognizing spatial hierarchies within an image. CNNs detect patterns at multiple levels of abstraction, beginning with basic structures such as edges and progressing to more complex textures, shapes, and high-level representations. CNNs consist of multiple layer types, each serving a specific role in processing visual information (Alzubaidi et al., 2021; Chollet and Chollet, 2021; Wu, 2017). An overview of the inner workings of a CNN for wood identification is presented in Figure 8.1, illustrating the sequence of convolutional and pooling layers used to extract hierarchical features from cross-sectional images, followed by fully connected layers that perform classification by outputting class probabilities across all candidate genera. Convolutional layers apply small, learnable filters (kernels) to an image, detecting relevant features such as edges, textures, and structural patterns. Digital images consist of a grid of pixels, each representing intensity values for specific wavelengths of electromagnetic (EM) radiation. Grayscale images contain a single intensity value per pixel, while multichannel images capture multiple wavelength ranges. Standard colour images (RGB) use three channels—red (620–780 nm), green (490–570 nm), and blue (440–490 nm)—encoding visible light to align digital data with human visual perception. Hyperspectral

images capture reflectance across numerous narrow spectral bands (beyond the visible spectrum), stacking channels to provide detailed spectral signatures for material identification. The kernels of the convolutional layer slide across the input while preserving spatial relationships between pixels, enabling CNNs to learn localized patterns. A single convolutional layer applies multiple filters to the input image, increasing the number of feature representations (channels), while maintaining the original spatial dimensions. The number of filters determines the number of output channels; For example, a standard colour image (three channels) passed through a convolutional layer with 64 filters, will output a feature map with 64 feature representations. As the network deepens, later convolutional layers typically apply more filters, progressively extracting more abstract representations. Pooling layers down-sample the spatial dimensions of a feature map, to improve computational efficiency and reduce sensitivity to small transformations. Max pooling, the most common approach, selects the highest pixel value in a kernel, ensuring that small variations in an image—such as changes in lighting, orientation, or slight distortions—do not drastically affect the model’s ability to recognize objects (Chollet and Chollet, 2021). Global Average Pooling computes the average value of each feature map across its entire spatial dimensions, unlike max pooling, which selects the highest value (Chollet and Chollet, 2021). This results in a single value per feature map, significantly reducing the number of parameters while preserving essential information. It is frequently used in literature to produce a compact representation, providing a clear link between feature extraction and subsequent classification. By stacking multiple convolutional and pooling layers in succession, CNNs enable hierarchical feature learning, where shallow layers detect low-level structures, and deeper layers extract increasingly abstract and complex representations of the image (Wu, 2017).

After convolution and pooling operations, the extracted features are passed to fully connected layers, which integrate and refine the learned representations before transforming them into the desired output format (Chollet and Chollet, 2021; Taye, 2023). The classification task is facilitated by assigning labels to images according to a predetermined class distribution. CNNs facilitate classification by converting the image into a vector of class probabilities, where each value corresponds to the likelihood of the input belonging to a specific class. The model allocates higher probability values to the classes that are more likely to represent the image, based on its learned features. Training models to classify based on labelled examples is known as supervised learning (Cunningham et al., 2008; Sarker, 2021; Wani et al., 2020). In contrast, models aiming to clusters within the data, without relying on labelling, is called unsupervised learning (James et al., 2023; Sarker, 2021; Shen et al., n.d.). Both can also be combined into semi-supervised learning using a small amount of labelled data with a larger set of unlabelled data to improve model performance. Semi-supervised learning is a good approach if labelling data is expensive and/or time-consuming (Zhu, 2005). Reinforcement learning involves an agent learning through interactions with its environment, receiving feedback in the form of rewards or penalties to maximize long-term goals, suiting well to interactive applications such as robotics or game play (Arulkumaran et al., 2017; Kaelbling et al., 1996; Sutton and Barto, 1999).

To achieve optimal performance, CNNs rely on an iterative optimization process that adjusts the kernel weights of the convolutional and fully connected layers based on the discrepancy between the model’s predictions and the actual labels (Alzubaidi et al., 2021). This is achieved by repeatedly passing the data through the network and updating the kernel weights at intermediate steps. A complete pass through the entire dataset is called an epoch (Chollet and Chollet, 2021). During each epoch, the dataset is typically divided into smaller batches, with weight updates occurring after processing each batch. This mini-batch approach allows the model to refine its parameters gradually, preventing large, unstable updates (Jian-Wei et al., 2020). The entire process is referred to as training a model, where images pass through the network to learn meaningful feature representations by minimizing the discrepancy between the model’s predictions and the actual labels (Chollet and Chollet, 2021). This discrepancy is quantified by a loss function, with cross-entropy loss commonly used for classification tasks (Cao et al., 2018). Performance is optimized by minimizing loss through gradient descent (Haji and Abdulazeez, 2021). The gradient of the loss function measures how the loss changes with respect to each kernel weight. By computing the gradient, the model determines the direction in which each weight should be adjusted to decrease the loss. To determine the influence of each weight on the loss, the error signal from the fully-connected layer is

propagated backwards through the network (Amari, 1993). The weights are updated accordingly by an optimization algorithm (e.g., Stochastic Gradient Descent (SGD), Adam) by taking a step in the descending direction of the gradient, scaled by a learning rate—a hyperparameter that controls the step size (Alzubaidi et al., 2021; Haji and Abdulazeez, 2021). If the learning rate is too large, the model may overshoot the optimal solution; if it is too small, convergence may be slow (Alzubaidi et al., 2021; Chollet and Chollet, 2021). Optimization is performed iteratively over multiple batches and epochs, gradually refining the model's weights to minimize loss and improve classification. To enhance the learning process, CNNs incorporate activation functions, which introduce non-linearity and allow the network to learn complex relationships beyond simple linear transformations (Dubey et al., 2022; Hao et al., 2020). One of the most widely used activation functions is the ReLU, which replaces negative values with zero (Banerjee et al., 2019). This is an effective countermeasure for phenomena where the gradient vanishes or explodes (Liu et al., 2021). When gradients approach values close to zero (and 'vanish'), weight updates become negligible, slowing down or even stopping learning. Alternatively, they can become excessively large (or 'explode'), causing unstable weight updates that lead to drastic fluctuations in model parameters (Chollet and Chollet, 2021; Hanin, 2018; Hochreiter, 1998). This instability can prevent convergence, making training ineffective.

While optimization sounds straightforward, ensuring that a model generalizes well to real-world data requires careful consideration. One of the main challenges in deep learning is overfitting, where a model becomes overly specialized in recognizing patterns from the training data but fails to perform well on new, unseen examples (Bejani and Ghatee, 2021; Ying, 2019). A crucial strategy to mitigate overfitting is validation, where a separate validation dataset, distinct from the training set, is passed through the network during, without altering the weights (Chollet and Chollet, 2021). Loss is calculated on those images to evaluate the model's performance throughout training. This helps detect overfitting by monitoring how well the model generalizes to unseen data rather than just memorizing training patterns. If the model performs significantly better on the training data than on the validation data, it indicates that it has learned to fit the training data too closely rather than capturing generalizable features. To counteract this, training can be stopped early by monitoring the validation performance and halting training if iterations no longer minimize loss, calculated on the validation data (Prechelt, 2002). Regularization techniques can be embedded in the model to further prevent overfitting (Santos and Papa, 2022). Dropout is a widely used regularization method that randomly deactivates a subset of neurons during training, forcing the network to rely on multiple feature representations rather than depending too heavily on specific patterns (Srivastava et al., 2014). Another regularization method, L2 regularization (also known as weight decay), discourages excessively large weight values by adding a penalty term to the loss function (Xie et al., 2022). Weight decay constraint helps prevent the model from assigning too much importance to individual features, promoting smoother and more stable learning (Santos and Papa, 2022). To improve model generalization further, the images can be transformed (e.g. rotation, flipping, cropping, brightness adjustments, noise addition) to artificially expand the variation in the training data (Shorten and Khoshgoftaar, 2019). These augmentations introduce variations in lighting, orientation, and scale, helping the model learn to recognize objects despite these changes. As a result, data augmentation enhances the model's ability to generalize to real-world images, where such variations naturally occur (Shorten and Khoshgoftaar, 2019). A crucial factor in achieving generalization is maintaining a balanced dataset. An imbalanced dataset can lead to biased predictions, where the model disproportionately favours classes with more training samples (Chollet and Chollet, 2021; Johnson and Khoshgoftaar, 2019). This issue is particularly relevant in wood identification, as certain species may be underrepresented in wood collections. To mitigate this, techniques such as class weighting and sampling can be employed. Class weighting increases the importance of underrepresented classes during training, ensuring they contribute proportionally to optimization (Johnson and Khoshgoftaar, 2019). Alternatively, oversampling augments the number of samples in minority classes, while undersampling reduces the prevalence of dominant classes, thereby improving model balance.

After training, models are typically evaluated on a separate dataset containing distinct images (Alzubaidi et al., 2021). While the validation dataset can be repurposed for this evaluation, it is recommended to use an entirely separate test dataset. This prevents potential data leakage, which can occur if overfitting

mitigation strategies, such as early stopping, were applied. If training is halted based on the progression of validation loss, the model is effectively optimized for that specific dataset, potentially leading to an overestimation of its performance on truly unseen data (Prechelt, 2002).

Table 8.1: Full overview of all SmartWoodID specimens (split across training and test sets) for training and evaluating wood classification models, with a mention of the commercial status.

Family	Number of Training Specimen par Family	Number of Testing Specimen par Family	Genus	Number of Training Specimen par Genus	Number of Testing Specimen par Genus	Species	Training Specimen ID Tervuren xylarium	Testing Specimen ID Tervuren xylarium	Commercial Status
Achariaceae	12	6	Caloncoba Gilg	5	2	Caloncoba glauca (P.Beauv.) Gilg	Tw984, Tw2022, Tw7180	Tw928	Not Commercial
						Caloncoba welwitschii (Oliv.) Gilg	Tw63887, Tw61053	Tw59461	Not Commercial
			Lepisanthes Blume	2	1	Lepisanthes senegalensis (Poir.) Leenh.	Tw60200, Tw60172	Tw32771	Not Commercial
			Lindackeria C.Presl	2	2	Lindackeria bukobensis Gilg	Tw59391	Tw21926	Not Commercial
						Lindackeria dentata (Oliv.) Gilg	Tw61173	Tw146	Not Commercial
			Scottellia Oliv.	3	1	Scottellia klaineana Pierre	Tw8313, Tw1907, Tw1311	Tw1919	Not Commercial
Anacardiaceae	37	21	Antrocaryon Pierre	4	3	Antrocaryon micraster A.Chev. & Guillaumin	Tw53856	Tw29940	Not Commercial
						Antrocaryon nannanii De Wild.	Tw7682, Tw7534, Tw7567	Tw7119, Tw1406	Commercial
			Clausena Burm.f.	1	1	Clausena anisata (Willd.) Hook.fil.	Tw28798	Tw23531	Not Commercial
			Ganophyllum Blume	3	2	Ganophyllum giganteum (A.Chev.) Hauman	Tw9885, Tw5325, Tw3644	Tw1696, Tw1536	Not Commercial
			Lannea A.Rich.	3	1	Lannea welwitschii (Hiern) Engl.	Tw1265, Tw1464, Tw33	Tw1574	Not Commercial
			Mangifera L.	1	1	Mangifera indica L.	Tw457	Tw19558	Not Commercial
			Myrsine L.	2	1	Myrsine melanophloeos (L.) R.Br.	Tw21937, Tw18728	Tw13215	Not Commercial
			Panda Pierre	3	1	Panda oleosa Pierre	Tw1865, Tw5241, Tw947	Tw7143	Not Commercial
			Pseudospondias Engl.	3	2	Pseudospondias longifolia Engl.	Tw61416	Tw60636	Not Commercial
						Pseudospondias microcarpa (A.Rich.) Engl.	Tw1278, Tw43692	Tw125	Not Commercial
			Santiria Blume	1	1	Santiria trimera (Oliv.) Aubrév.	Tw10196	Tw10152	Not Commercial
			Sorindeia Thouars	5	3	Sorindeia africana (Engl.) Van der Veken	Tw9012, Tw8716	Tw192	Not Commercial
						Sorindeia juglandifolia	Tw61418, Tw61456, Tw41112	Tw32374, Tw32768	Not Commercial

						(A.Rich.) Planch. ex Oliv.			
			Spondias L.	3	1	Spondias dulcis Parkinson	Tw31049, Tw72037, Tw71044	Tw71775	Not Commercial
			Trichoscypha Hook.f.	5	3	Trichoscypha acuminata Engl.	Tw7333, Tw182, Tw1906	Tw7326	Not Commercial
						Trichoscypha lucens Oliv.	Tw8481	Tw21117	Not Commercial
						Trichoscypha oddonii De Wild.	Tw43719	Tw112	Not Commercial
			Xymalos Baill.	3	1	Xymalos monospora (Harv.) Baill.	Tw8483, Tw13214, Tw17193	Tw4333	Not Commercial
Anisophylleaceae	2	1	Anisophyllea R.Br. ex Sabine	2	1	Anisophyllea boehmii Engl.	Tw24226, Tw28526	Tw20594	Not Commercial
Annonaceae	52	29	Annickia Setten & Maas	3	2	Annickia affinis (Exell) Versteegh & Sosef	Tw4820, Tw10808	Tw10268	Not Commercial
						Annickia lebrunii (Robyns & Ghesq.) Setten & Maas	Tw2479	Tw2009	Not Commercial
			Anonidium Engl. & Diels	4	2	Anonidium mannii (Oliv.) Engl. & Diels	Tw619, Tw1892, Tw362, Tw2510	Tw1191, Tw1858	Not Commercial
			Brieya De Wild.	1	1	Brieya fasciculata De Wild.	Tw7624	Tw1261	Not Commercial
			Cleistopholis Pierre ex Engl.	4	3	Cleistopholis glauca Pierre ex Engl. & Diels	Tw8030	Tw7408	Not Commercial
						Cleistopholis patens (Benth.) Engl. & Diels	Tw357, Tw1820, Tw1888	Tw1795, Tw1407	Not Commercial
			Duguetia A.St.-Hil.	2	1	Duguetia staudtii (Engl. & Diels) Chatrou	Tw1244, Tw970	Tw10344	Not Commercial
			Greenwayodendron Verdc.	4	2	Greenwayodendron suaveolens (Engl. & Diels) Verdc.	Tw205, Tw404, Tw880, Tw552	Tw199, Tw150	Not Commercial
			Hexalobus A.DC.	4	3	Hexalobus crispiflorus A.Rich.	Tw10812, Tw1469, Tw14809	Tw10144, Tw10811	Not Commercial
						Hexalobus monopetalus (A.Rich.) Engl. & Diels	Tw41433	Tw30055	Not Commercial
			Isolona Engl.	3	2	Isolona congolana (De Wild. & T.Durand) Engl. & Diels	Tw8290, Tw8243	Tw3603	Not Commercial
						Isolona hexaloba Pierre ex Engl. & Diels	Tw8312	Tw63477	Not Commercial

			Monodora Dunal	1	1	Monodora angolensis Welw.	Tw7238	Tw2219	Not Commercial
			Platymitra Boerl.	2	1	Platymitra arborea (Blanco) Kessler	Tw87, Tw8267	Tw1279	Not Commercial
			Xylophia L.	24	11	Xylophia aethiopica (Dunal) A.Rich.	Tw356, Tw3683	Tw170	Not Commercial
						Xylophia aurantiiodora De Wild. & T.Durand	Tw611	Tw599	Not Commercial
						Xylophia cupularis Mildbr.	Tw4822, Tw6958	Tw33463	Not Commercial
						Xylophia flamignii Boutique	Tw7749, Tw2481, Tw2480	Tw2482	Not Commercial
						Xylophia gilbertii Boutique	Tw7581, Tw1915	Tw1894	Not Commercial
						Xylophia hypolampra Mildbr. & Diels	Tw57, Tw1451	Tw1421	Not Commercial
						Xylophia katangensis De Wild.	Tw21176, Tw3604	Tw1973	Not Commercial
						Xylophia phloiodora Mildbr.	Tw1944, Tw3687, Tw1955	Tw32755	Not Commercial
						Xylophia rubescens Oliv.	Tw2484, Tw791	Tw10831	Not Commercial
						Xylophia staudtii Engl. & Diels	Tw50, Tw10359, Tw10832	Tw4867	Not Commercial
						Xylophia wilwerthii De Wild.	Tw58862, Tw58861	Tw32601	Not Commercial
Apocynaceae	31	16	Alstonia R.Br.	6	3	Alstonia boonei De Wild.	Tw225, Tw1640, Tw575	Tw1168, Tw1470	Commercial
						Alstonia congensis Engl.	Tw294, Tw475, Tw145	Tw317	Not Commercial
			Diplorhynchus Welw. ex Ficalho & Hiern	3	1	Diplorhynchus condylocarpon (Müll.Arg.) Pichon	Tw20543, Tw28193, Tw24217	Tw24382	Not Commercial
			Funtumia Stapf	2	1	Funtumia elastica (Preuss) Stapf	Tw616, Tw487	Tw17972	Not Commercial
			Holarrhena R.Br.	2	1	Holarrhena floribunda (G.Don) T.Durand & Schinz	Tw5160, Tw439	Tw272	Not Commercial
			Hunteria Roxb.	2	1	Hunteria umbellata (K.Schum.) Hallier fil.	Tw14478, Tw23851	Tw114	Not Commercial
			Picralima Pierre	1	1	Picralima nitida (Stapf) T.Durand & H.Durand	Tw32461	Tw10214	Not Commercial
			Pleiocarpa Benth.	1	1	Pleiocarpa pycnantha (K.Schum.) Stapf	Tw33076	Tw29762	Not Commercial

			Rauvolfia L.	5	2	Rauvolfia caffra Sond.	Tw24035, Tw25770	Tw18706	Not Commercial
						Rauvolfia vomitoria Afzel.	Tw1248, Tw21824, Tw20976	Tw21822	Not Commercial
			Tabernaemontana Plum. ex L.	6	3	Tabernaemontana crassa Benth.	Tw52, Tw26564, Tw21945, Tw26563	Tw143, Tw10351	Not Commercial
						Tabernaemontana pachysiphon Stapf	Tw39181, Tw28855	Tw28178	Not Commercial
			Voacanga Thouars	3	2	Voacanga africana Stapf ex Scott Elliot	Tw753, Tw28921	Tw21823	Not Commercial
						Voacanga thouarsii Roem. & Schult.	Tw8544	Tw8539	Not Commercial
Araliaceae	4	3	Cussonia Thunb.	2	2	Cussonia arborea Hochst. ex A.Rich.	Tw57545	Tw31821	Not Commercial
						Cussonia spicata Thunb.	Tw25917	Tw17154	Not Commercial
			Polyscias J.R.Forst. & G.Forst.	2	1	Polyscias fulva (Hiern) Harms	Tw875, Tw749	Tw7291	Not Commercial
Asteraceae	4	2	Brenandendron H.Rob.	3	1	Brenandendron donianum (DC.) H.Rob.	Tw181, Tw17507, Tw28136	Tw26577	Not Commercial
			Gymnanthemum Cass.	1	1	Gymnanthemum amygdalinum (Delile) Sch.Bip. ex Walp.	Tw51060	Tw50919	Not Commercial
Bignoniaceae	16	10	Fernandoa Welw. ex Seem.	1	1	Fernandoa adolfi-friderici (Gilg & Mildbr.) Heine	Tw7107	Tw1217	Not Commercial
			Kigelia DC.	3	1	Kigelia africana (Lam.) Benth.	Tw23847, Tw17478, Tw27684	Tw24304	Not Commercial
			Markhamia Seem.	8	5	Markhamia lutea (Benth.) K.Schum.	Tw22717, Tw8182, Tw25680	Tw1868, Tw1866	Not Commercial
						Markhamia obtusifolia (Baker) Sprague	Tw334	Tw1972	Not Commercial
						Markhamia tomentosa (Benth.) K.Schum. ex Engl.	Tw41, Tw8565, Tw28509	Tw8392	Not Commercial
						Markhamia zanzibarica (Bojer ex DC.) K.Schum.	Tw23618	Tw18709	Not Commercial
						Spathodea Beauverd	2	1	Spathodea campanulata Beauverd
			Stereospermum Cham.	2	2	Stereospermum harmsianum K.Schum.	Tw327	Tw24354	Not Commercial
						Stereospermum kunthianum Cham.	Tw11296	Tw11295	Not Commercial
Boraginaceae	11	5	Cordia L.	11	5	Cordia africana Lam.	Tw7345, Tw2010, Tw2106	Tw6982	Not Commercial

						Cordia millenii Baker	Tw11445, Tw33112, Tw14028	Tw17784	Not Commercial
						Cordia monoica Roxb.	Tw28614, Tw28677	Tw28481	Not Commercial
						Cordia myxa L.	Tw60259	Tw30079	Not Commercial
						Cordia platythyrsa Baker	Tw1259, Tw22655	Tw1242	Not Commercial
Burseraceae	20	12	Aucoumea Pierre	3	2	Aucoumea klaineana Pierre	Tw25782, Tw18752, Tw57884	Tw14988, Tw12938	Not Commercial
			Canarium L.	3	2	Canarium schweinfurthii Engl.	Tw255, Tw128, Tw364	Tw1164, Tw1195	Commercial
			Commiphora Jacq.	1	1	Commiphora mollis (Oliv.) Engl.	Tw28183	Tw23062	Not Commercial
			Pachylobus G.Don	13	7	Pachylobus buettneri (Engl.) Guillaumin	Tw10770, Tw22559, Tw13196	Tw18754	Commercial
						Pachylobus edulis G.Don	Tw2358, Tw2064	Tw1547	Not Commercial
						Pachylobus igaganga (Aubrév. & Pellegr.) Byng & Christenh.	Tw57881, Tw18755, Tw43722	Tw51795	Not Commercial
						Pachylobus normandii (Aubrév. & Pellegr.) Byng & Christenh.	Tw57880	Tw18756	Not Commercial
						Pachylobus osika Guillaumin	Tw9760	Tw8021	Not Commercial
						Pachylobus pubescens Vermoesen	Tw982, Tw1516, Tw267	Tw130, Tw1433	Not Commercial
Calophyllaceae	4	3	Endodesmia Benth.	1	1	Endodesmia calophylloides Benth.	Tw30676	Tw29814	Not Commercial
			Mammea L.	3	2	Mammea africana G.Don	Tw452, Tw472, Tw467	Tw382, Tw421	Not Commercial
Cannabaceae	21	11	Celtis L.	15	7	Celtis gomphophylla Baker	Tw64995, Tw64993, Tw7603, Tw6969, Tw64991, Tw3605	Tw26920, Tw64992	Commercial
						Celtis latifolia (Blume) Planch.	Tw71106	Tw71087	Not Commercial
						Celtis mildbraedii Engl.	Tw1560, Tw3437, Tw1561	Tw2135	Not Commercial
						Celtis philippensis Blanco	Tw573	Tw31998	Not Commercial
						Celtis tessmannii Rendle	Tw62768	Tw62746	Commercial

						Celtis zenkeri Engl.	Tw1515, Tw2008, Tw1509	Tw1553	Not Commercial
			Morus L.	3	2	Morus mesozygia Stapf	Tw1471, Tw17738, Tw1472	Tw1238, Tw1249	Commercial
			Trema Lour.	3	2	Trema orientale (L.) Blume	Tw57698, Tw59409, Tw59650	Tw28460, Tw23439	Not Commercial
Celastraceae	4	2	Apodytes E.Mey. ex Bernh.	1	1	Apodytes dimidiata E.Mey. ex Arn.	Tw21931	Tw19979	Not Commercial
			Cassine L.	3	1	Cassine peragua L.	Tw21954, Tw29739, Tw39086	Tw29740	Not Commercial
Chrysobalanaceae	15	10	Magnistipula Engl.	1	1	Magnistipula butayi De Wild.	Tw31487	Tw14119	Not Commercial
			Maranthes Blume	7	4	Maranthes chrysophylla (Oliv.) Prance ex F.White	Tw26573	Tw25712	Not Commercial
						Maranthes gabunensis (Engl.) Prance	Tw7177	Tw25733	Not Commercial
						Maranthes glabra (Oliv.) Prance	Tw8219, Tw47797, Tw2395	Tw597	Not Commercial
						Maranthes kerstingii (Engl.) Prance ex F.White	Tw22962, Tw44434	Tw10383	Not Commercial
			Parinari Aubl.	7	5	Parinari congensis Didr.	Tw6999	Tw3488	Not Commercial
						Parinari curatellifolia Planch. ex Benth.	Tw23504, Tw3889, Tw23510	Tw19261, Tw21126	Not Commercial
						Parinari excelsa Sabine	Tw8098, Tw33991, Tw736	Tw1270, Tw1136	Not Commercial
Clusiaceae	30	14	Allanblackia Oliv.	8	3	Allanblackia floribunda Oliv.	Tw358, Tw53, Tw1835	Tw395	Not Commercial
						Allanblackia kisonghi Vermoesen	Tw273, Tw8574, Tw2402	Tw8314	Not Commercial
						Allanblackia parviflora A.Chev.	Tw23010, Tw22983	Tw10380	Not Commercial
			Garcinia L.	14	6	Garcinia chromocarpa Engl.	Tw63602, Tw63601	Tw63598	Not Commercial
						Garcinia epunctata Stapf	Tw167, Tw155, Tw2003	Tw178	Not Commercial
						Garcinia huillensis Welw.	Tw20553, Tw28476, Tw10219	Tw21754	Not Commercial
						Garcinia ovalifolia Oliv.	Tw33362	Tw10181	Not Commercial
						Garcinia punctata Oliv.	Tw7469, Tw1139, Tw32224	Tw3643	Not Commercial
						Garcinia smeathmannii (Planch. & Triana) Oliv.	Tw2210, Tw21821	Tw10210	Not Commercial

			Lebrunia Staner	2	1	Lebrunia busbaie Staner	Tw1645, Tw2193	Tw1067	Not Commercial
			Pentadesma Sabine	5	3	Pentadesma butyracea Sabine	Tw7627, Tw142, Tw171	Tw10339, Tw11616	Not Commercial
						Pentadesma grandifolia Baker fil.	Tw7134, Tw7234	Tw553	Not Commercial
			Symphonia L.f.	1	1	Symphonia globulifera L.fil.	Tw54142	Tw25413	Not Commercial
Combretaceae	40	23	Combretum Loeft.	10	6	Combretum adenogonium Steud. ex A.Rich.	Tw60236	Tw29011	Not Commercial
						Combretum collinum Fresen.	Tw28266, Tw28709	Tw28207	Not Commercial
						Combretum collinum subsp. elgonense (Exell) Okafor	Tw23672, Tw23674	Tw20545	Not Commercial
						Combretum lokafor Liben	Tw33108	Tw32744	Not Commercial
						Combretum molle R.Br. ex G.Don	Tw28503	Tw26054	Not Commercial
						Combretum zeyheri Sond.	Tw24029, Tw24349, Tw28225	Tw28209	Not Commercial
			Terminalia L.	30	17	Terminalia anisoptera (Welw. ex M.A.Lawson) Gere & Boatwr.	Tw28181, Tw28172	Tw24370	Not Commercial
						Terminalia brachystemma Welw. ex Hiern	Tw28537	Tw24340	Not Commercial
						Terminalia catappa L.	Tw25689, Tw57696, Tw59834	Tw59490	Not Commercial
						Terminalia hylandron (Mildbr.) Gere & Boatwr.	Tw8530, Tw5156, Tw6957	Tw1104, Tw30030	Not Commercial
						Terminalia ivorensis A.Chev.	Tw47938, Tw20807, Tw40978, Tw26420, Tw17974	Tw14029, Tw11064	Not Commercial
						Terminalia leiocarpa (DC.) Baill.	Tw31659, Tw30912, Tw41424	Tw30077, Tw30693	Not Commercial
						Terminalia macroptera Guill. & Perr.	Tw49440	Tw41423	Not Commercial
						Terminalia mollis M.A.Lawson	Tw11366, Tw11372, Tw340	Tw11300, Tw11264	Not Commercial
						Terminalia sericea Burch. ex DC.	Tw28590, Tw24039	Tw17177	Not Commercial

						Terminalia superba Engl. & Diels	Tw1430, Tw57060, Tw1354, Tw57207, Tw55833, Tw88	Tw1353, Tw1318, Tw1409	Commercial
						Terminalia welwitschii Gere & Boatwr.	Tw7741	Tw28134	Not Commercial
Daphniphyllaceae	2	1	Plagiostyles Pierre	2	1	Plagiostyles africana (Müll.Arg.) Prain	Tw2491, Tw2490	Tw2403	Not Commercial
Dipterocarpaceae	5	4	Marquesia Gilg	3	2	Marquesia macroura Gilg	Tw341, Tw9816, Tw24249	Tw20548, Tw22727	Not Commercial
			Monotes A.DC.	2	2	Monotes hypoleucus var. angolensis (De Wild.) Meerts	Tw28221	Tw28139	Not Commercial
						Monotes katangensis (De Wild.) De Wild.	Tw307	Tw23912	Not Commercial
Ebenaceae	13	7	Diospyros L.	13	7	Diospyros batocana Hiern	Tw20550, Tw56593, Tw28273	Tw29615	Not Commercial
						Diospyros crassiflora Hiern	Tw523, Tw788, Tw57052, Tw7651	Tw14483, Tw33268	Commercial
						Diospyros dendo Welw. ex Hiern	Tw8077	Tw22664	Not Commercial
						Diospyros ferrea (Willd.) Bakh.	Tw22862, Tw1802	Tw1263	Not Commercial
						Diospyros iturensis (Gürke) Letouzey & F.White	Tw33548	Tw32832	Not Commercial
						Diospyros mespiliformis Hochst. ex A. DC.	Tw14485, Tw17165	Tw10293	Not Commercial
Erythroxylaceae	3	2	Erythroxylum Browne	3	2	Erythroxylum mannii Oliv.	Tw29886, Tw816, Tw30886	Tw22563, Tw22564	Not Commercial
Euphorbiaceae	48	27	Cavacoa J.Léonard	1	1	Cavacoa quintasii (Pax & K.Hoffm.) J.Léonard	Tw33564	Tw10142	Not Commercial
			Croton L.	8	6	Croton haumanianus J.Léonard	Tw9759	Tw33372	Not Commercial
						Croton macrostachyus Hochst. ex Delile	Tw39143	Tw28714	Not Commercial
						Croton mayumbensis J.Léonard	Tw4318, Tw1929	Tw168	Not Commercial
						Croton megalocarpus Hutch.	Tw26071	Tw18849	Not Commercial
						Croton mubango Müll.Arg.	Tw8084	Tw7128	Not Commercial

						Croton sylvaticus Hochst.	Tw7714, Tw7921	Tw27710	Not Commercial
			Dichostemma Pierre	2	1	Dichostemma glaucescens Pierre	Tw2391, Tw1286	Tw1155	Not Commercial
			Discoglyprena Prain	2	1	Discoglyprena caloneura (Pax) Prain	Tw624, Tw6967	Tw1224	Not Commercial
			Grossera Pax	3	2	Grossera macrantha Pax	Tw7329, Tw8216, Tw7227	Tw10143, Tw30664	Not Commercial
			Hevea Aubl.	3	2	Hevea brasiliensis (Willd. ex A.Juss.) Müll.Arg.	Tw61294, Tw61295, Tw50322	Tw26141, Tw19370	Not Commercial
			Klaineanthus Pierre ex Prain	3	1	Klaineanthus gabonae Pierre ex Prain	Tw8543, Tw11439, Tw8434	Tw8446	Not Commercial
			Macaranga Thouars	8	3	Macaranga kilimandscharica Pax	Tw19426, Tw4916, Tw24166	Tw2563	Not Commercial
						Macaranga monandra Müll.Arg.	Tw10224, Tw1223, Tw2651	Tw2649	Not Commercial
						Macaranga spinosa Müll.Arg.	Tw7122, Tw8163	Tw1149	Not Commercial
			Maprounea Aubl.	2	1	Maprounea africana Müll.Arg.	Tw32701, Tw28261	Tw24351	Not Commercial
			Neoboutonia Müll.Arg.	3	1	Neoboutonia macrocalyx Pax	Tw21942, Tw31481, Tw17221	Tw24189	Not Commercial
			Ricinodendron Müll.Arg.	4	2	Ricinodendron heudelotii (Baill.) Heckel	Tw43839, Tw41100	Tw12959	Commercial
						Ricinodendron heudelotii subsp. africanum (Müll.Arg.) J.Léonard	Tw2471, Tw122	Tw1119	Not Commercial
			Schinziphyton Hutch. ex Radcl.Sm.	3	2	Schinziphyton rautanenii (Schinz) Radcl.-Sm.	Tw11329, Tw335, Tw11333	Tw11292, Tw11273	Not Commercial
			Sclerocroton Hochst.	2	1	Sclerocroton cornutus (Pax) Kruijt & Roebers	Tw8228, Tw2249	Tw10218	Not Commercial
			Shirakiopsis Esser	1	1	Shirakiopsis elliptica (Hochst.) Esser	Tw907	Tw2060	Not Commercial
			Tetrorchidium Poepp.	3	2	Tetrorchidium didymostemon (Baill.) Pax & K.Hoffm.	Tw2189, Tw2049, Tw967	Tw1909, Tw164	Not Commercial
Fabaceae	425	221	Afzelia Sm.	28	11	Afzelia africana Sm. ex Pers.	Tw14035, Tw5024, Tw53917,	Tw5023, Tw11081	Not Commercial

							Tw53506, Tw47200, Tw53948		
						Afzelia bella Harms	Tw375, Tw10452, Tw42694	Tw3899	Not Commercial
						Afzelia bipindensis Harms	Tw26431, Tw1567, Tw3898, Tw53508, Tw14072, Tw886, Tw30643, Tw11236	Tw13492, Tw2417, Tw29571	Commercial
						Afzelia pachyloba Harms	Tw14458, Tw23635, Tw2100, Tw57088, Tw62, Tw65751, Tw57069	Tw62357, Tw20811, Tw51747	Commercial
						Afzelia quanzensis Welw.	Tw322, Tw28248, Tw308, Tw342	Tw28143, Tw1022	Not Commercial
			Albizia Durazz.	38	21	Albizia adianthifolia (Schumach.) W.Wight	Tw2034, Tw3678, Tw4962, Tw24343	Tw1927, Tw11235	Not Commercial
						Albizia adianthifolia var. intermedia (De Wild. & T.Durand) Villiers	Tw5343	Tw518	Not Commercial
						Albizia altissima Hook.f.	Tw985, Tw8552, Tw3948	Tw30094, Tw3582	Not Commercial
						Albizia antunesiana Harms	Tw29622, Tw29058, Tw28848	Tw28191, Tw24213	Not Commercial
						Albizia chinensis (Osbeck) Merr.	Tw25206, Tw10957, Tw50580	Tw41850	Not Commercial
						Albizia coriaria Welw. ex Oliv.	Tw1214, Tw20703, Tw7131	Tw6962	Not Commercial
						Albizia ferruginea (Guill. & Perr.) Benth.	Tw724, Tw1544, Tw609, Tw842	Tw1450, Tw1172	Commercial
						Albizia glaberrima (Schumach. & Thonn.) Benth.	Tw29572, Tw29594	Tw25101	Not Commercial
						Albizia glaberrima var. glabrescens (Oliv.) Brenan	Tw8122	Tw4764	Not Commercial
						Albizia gummifera (J.F.Gmel.) C.A.Sm.	Tw690, Tw528	Tw1141	Not Commercial
						Albizia gummifera var. ealaensis (De Wild.) Brenan	Tw2164, Tw926, Tw1111	Tw262	Not Commercial
						Albizia laurentii De Wild.	Tw20615, Tw3607	Tw1670	Not Commercial
						Albizia lebbeck (L.) Benth.	Tw30609	Tw10908	Not Commercial
						Albizia schimperiana Oliv.	Tw43691	Tw43690	Not Commercial
						Albizia versicolor Welw. ex Oliv.	Tw6998, Tw7584, Tw3586	Tw2014, Tw2389	Not Commercial

						Albizia zygia (DC.) J.F.Macbr.	Tw3625, Tw657	Tw20707	Not Commercial
			Amphimas Pierre ex Harms	5	2	Amphimas ferrugineus Pierre ex Pellegr.	Tw32663, Tw2036, Tw1289	Tw32495	Commercial
						Amphimas pterocarpoides Harms	Tw394, Tw3694	Tw1404	Commercial
			Annea Mackinder & Wieringa	2	1	Annea laxiflora (Benth.) Mackinder & Wieringa	Tw28601, Tw32418	Tw28470	Not Commercial
			Anthonotha P.Beauv.	4	3	Anthonotha brieyi (De Wild.) J.Léonard	Tw19374, Tw95	Tw1206	Not Commercial
						Anthonotha fragrans (Baker f.) Exell & Hillc.	Tw3600	Tw1056	Not Commercial
						Anthonotha pynaertii (De Wild.) Exell & Hillc.	Tw7539	Tw33471	Not Commercial
			Aphanocalyx Oliv.	4	2	Aphanocalyx cynometroides Oliv.	Tw62474, Tw7639	Tw62443	Not Commercial
						Aphanocalyx microphyllus (Harms) Wieringa	Tw7145, Tw4949	Tw2050	Not Commercial
			Baikiaea Benth.	5	3	Baikiaea insignis Benth.	Tw24394, Tw24393, Tw25687	Tw21538, Tw10287	Not Commercial
						Baikiaea robynsii Ghesq. ex Laing	Tw40341, Tw39908	Tw33344	Not Commercial
			Baphia Afzel. ex G.Lodd.	10	6	Baphia bequaertii De Wild.	Tw28161, Tw24225, Tw24350	Tw23671, Tw23670	Not Commercial
						Baphia dewevrei De Wild.	Tw49182	Tw39910	Not Commercial
						Baphia massaiensis Taub.	Tw28666	Tw28567	Not Commercial
						Baphia nitida G.Lodd.	Tw29890, Tw14446, Tw14497	Tw28779	Not Commercial
						Baphia pubescens Hook.f.	Tw32737, Tw30012	Tw29855	Not Commercial
			Berlinia Sol. ex Hook.f. & Benth.	10	5	Berlinia bracteosa Benth.	Tw10786, Tw789, Tw113	Tw57428	Commercial
						Berlinia confusa Hoyle	Tw24418, Tw12967, Tw32021	Tw30016	Not Commercial
						Berlinia congolensis (Baker f.) Keay	Tw32626	Tw1438	Not Commercial
						Berlinia grandiflora (Vahl) Hutch. & Dalziel	Tw8439, Tw3492, Tw49435	Tw22674, Tw32753	Commercial

			Bobgunnia J.H.Kirkbr. & Wiersema	6	3	Bobgunnia fistuloides (Harms) J.H.Kirkbr. & Wiersema	Tw22594, Tw18782, Tw51757, Tw25708	Tw13500, Tw11245	Commercial
						Bobgunnia madagascariensis (Desv.) J.H.Kirkbr. & Wiersema	Tw304, Tw28165	Tw20586	Not Commercial
			Brachystegia Benth.	27	14	Brachystegia boehmii Taub.	Tw61780, Tw44583, Tw19267	Tw57541	Not Commercial
						Brachystegia bussei Harms	Tw61765, Tw61764	Tw19268	Not Commercial
						Brachystegia laurentii (De Wild.) Louis ex J.Léonard	Tw30675, Tw2540, Tw1562, Tw2414	Tw1363, Tw1030	Commercial
						Brachystegia longifolia Benth.	Tw28190, Tw11379, Tw329	Tw11278, Tw11288	Not Commercial
						Brachystegia manga De Wild.	Tw56370, Tw19270, Tw61812	Tw61785	Not Commercial
						Brachystegia spiciformis Benth.	Tw3887, Tw320, Tw2002, Tw345	Tw11303, Tw11326	Not Commercial
						Brachystegia tamarindoides subsp. microphylla (Harms) Chikuni	Tw19274, Tw28416	Tw19271	Not Commercial
						Brachystegia tamarindoides subsp. tamarindoides	Tw28607	Tw1577	Not Commercial
						Brachystegia taxifolia Harms	Tw3592, Tw24234, Tw330	Tw11269, Tw11302	Not Commercial
						Brachystegia utilis Hutch. & Burt Davy	Tw2370, Tw20562	Tw1373	Not Commercial
			Burkea Benth.	3	1	Burkea africana Hook.	Tw19275, Tw29610, Tw17181	Tw20563	Not Commercial
			Cassia L.	5	3	Cassia mannii Oliv.	Tw3598, Tw40369	Tw264	Not Commercial
						Cassia sieberiana DC.	Tw53923, Tw51672, Tw49425	Tw10885, Tw41426	Not Commercial
			Copaifera L.	4	3	Copaifera mildbraedii Harms	Tw8430, Tw7472, Tw8299	Tw7109, Tw5228	Commercial
						Copaifera religiosa J.Léonard	Tw48331	Tw26874	Not Commercial
			Craibia Harms & Dunn	5	3	Craibia affinis (De Wild.) De Wild.	Tw973	Tw4198	Not Commercial
						Craibia grandiflora (Micheli) Baker f.	Tw5068, Tw7353	Tw2360	Not Commercial
						Craibia lujae De Wild.	Tw7750, Tw8298	Tw7125	Not Commercial

			Crudia Schreb.	3	2	Crudia harmsiana De Wild.	Tw8410, Tw39900	Tw39899	Not Commercial
						Crudia laurentii De Wild.	Tw39884	Tw3954	Not Commercial
			Cryptosepalum Benth.	1	1	Cryptosepalum exfoliatum De Wild.	Tw28259	Tw28160	Not Commercial
			Cylicodiscus Harms	3	2	Cylicodiscus gabunensis Harms	Tw29781, Tw25735, Tw18881	Tw18765, Tw14033	Not Commercial
			Cynometra L.	13	7	Cynometra alexandri C.H.Wright	Tw3489, Tw1992, Tw1747	Tw2394	Commercial
						Cynometra hankei Harms	Tw1668, Tw2152, Tw589	Tw1662, Tw1158	Commercial
						Cynometra lujae De Wild.	Tw7161, Tw4765	Tw1440	Not Commercial
						Cynometra mannii Oliv.	Tw25692, Tw24420	Tw24419	Not Commercial
						Cynometra sessiliflora Harms	Tw8063, Tw3490, Tw7136	Tw27064, Tw1529	Not Commercial
			Dalbergia L.f.	1	1	Dalbergia boehmii Taub.	Tw24384	Tw21191	Not Commercial
			Daniellia Benn.	13	6	Daniellia alsteeniana P.A.Duvign.	Tw20565, Tw21750, Tw778	Tw7182	Not Commercial
						Daniellia klainei Pierre ex A.Chev.	Tw51768, Tw4838, Tw48454	Tw41218, Tw10752	Not Commercial
						Daniellia oliveri (Rolfe) Hutch. & Dalziel	Tw14038, Tw877, Tw20501	Tw3676	Not Commercial
						Daniellia pynaertii De Wild.	Tw51745, Tw901	Tw23020	Commercial
						Daniellia soyauxii (Harms) Rolfe	Tw5352, Tw8395	Tw18766	Not Commercial
			Dialium L.	22	9	Dialium englerianum Henriq.	Tw26915, Tw8414, Tw10201	Tw44562	Not Commercial
						Dialium excelsum Louis ex Steyaert	Tw7638, Tw3612, Tw7342	Tw1107, Tw2429	Not Commercial
						Dialium pachyphyllum Harms	Tw61981, Tw51772, Tw852, Tw6745, Tw57704	Tw100, Tw4826	Not Commercial
						Dialium pentandrum Louis ex Steyaert	Tw21873, Tw13337, Tw7580	Tw3685	Not Commercial
						Dialium polyanthum Harms	Tw696, Tw2474, Tw2473	Tw2478	Not Commercial
						Dialium tessmannii Harms	Tw61162, Tw32636	Tw32248	Not Commercial
						Dialium zenkeri Harms	Tw10163, Tw7743, Tw5018	Tw7708	Not Commercial

			Dichrostachys (A.DC.) Wight & Arn.	4	2	Dichrostachys cinerea (L.) Wight & Arn.	Tw28572, Tw29050, Tw26730, Tw3957	Tw23518, Tw22650	Not Commercial
			Entada Adans.	1	1	Entada abyssinica Steud. ex A.Rich.	Tw3593	Tw21194	Not Commercial
			Erythrina L.	6	4	Erythrina abyssinica Lam.	Tw755	Tw20580	Not Commercial
						Erythrina droogmansiana De Wild. & T.Durand	Tw8441, Tw5022, Tw26829	Tw5161	Not Commercial
						Erythrina excelsa Baker	Tw5255	Tw24380	Not Commercial
						Erythrina orophila Ghesq.	Tw5078	Tw1068	Not Commercial
			Erythrophleum Afzel. ex G.Don	9	5	Erythrophleum africanum (Benth.) Harms	Tw312, Tw29619, Tw3588	Tw11356, Tw18720	Not Commercial
						Erythrophleum suaveolens (Guill. & Perr.) Brenan	Tw321, Tw853, Tw931, Tw96, Tw34933, Tw864	Tw131, Tw1187, Tw32962	Commercial
			Faidherbia A.Chev.	3	1	Faidherbia albida (Delile) A.Chev.	Tw44661, Tw28184, Tw29593	Tw30347	Not Commercial
			Fillaeopsis Harms	3	2	Fillaeopsis discophora Harms	Tw531, Tw981, Tw7685	Tw44, Tw2134	Not Commercial
			Gilbertiodendron J.Léonard	13	8	Gilbertiodendron dewevrei (De Wild.) J.Léonard	Tw568, Tw549, Tw5215	Tw2018, Tw332	Commercial
						Gilbertiodendron grandiflorum (De Wild.) J.Léonard	Tw5175, Tw5094	Tw105	Not Commercial
						Gilbertiodendron grandistipulatum (De Wild.) J.Léonard	Tw5097, Tw5093, Tw5092, Tw5095	Tw1252, Tw42	Not Commercial
						Gilbertiodendron mayombense (Pellegr.) J.Léonard	Tw26259	Tw19376	Not Commercial
						Gilbertiodendron ogoouense (Pellegr.) J.Léonard	Tw5138, Tw398, Tw3441	Tw3439, Tw3440	Not Commercial
			Gilletiodendron Vermoesen	5	3	Gilletiodendron kisantuense (Vermoesen ex De Wild.) J.Léonard	Tw60050, Tw60051	Tw29479	Not Commercial
						Gilletiodendron mildbraedii (Harms) Vermoesen	Tw3693, Tw1975, Tw2390	Tw1169, Tw1106	Not Commercial
			Guibourtia Benn.	10	5	Guibourtia arnoldiana (De	Tw1622, Tw1414, Tw1507	Tw1508	Commercial

						Wild. & T.Durand J.Léonard			
						Guibourtia coleosperma (Benth.) J.Léonard	Tw13508, Tw4203, Tw22751	Tw29612	Not Commercial
						Guibourtia demeusei (Harms) J.Léonard	Tw32757	Tw31	Commercial
						Guibourtia ehie (A.Chev.) J.Léonard	Tw22580, Tw18884, Tw25777	Tw13426, Tw12955	Commercial
			Hylodendron Taub.	3	1	Hylodendron gabunense Taub.	Tw1275, Tw1422, Tw3636	Tw25757	Not Commercial
			Hymenostegia Harms	3	2	Hymenostegia mundungu (Pellegr.) J.Léonard	Tw7256, Tw8062, Tw7737	Tw7162, Tw10141	Not Commercial
			Intsia Thouars	3	2	Intsia bijuga var. bijuga	Tw27153, Tw22581, Tw26099	Tw11652, Tw22514	Not Commercial
			Isoberlinia Craib & Stapf	6	3	Isoberlinia angolensis (Welw. ex Benth.) Hoyle & Brenan	Tw5257, Tw28173	Tw20568	Not Commercial
						Isoberlinia doka Craib & Stapf	Tw48547, Tw49428, Tw18192	Tw48548	Not Commercial
						Isoberlinia tomentosa (Harms) Craib & Stapf	Tw49427	Tw3879	Not Commercial
			Julbernardia Pellegr.	12	7	Julbernardia briei (De Wild.) Troupin	Tw30657	Tw1441	Not Commercial
						Julbernardia globiflora (Benth.) Troupin	Tw19280, Tw19279, Tw763	Tw19281	Not Commercial
						Julbernardia paniculata (Benth.) Troupin	Tw24348, Tw3891, Tw27515	Tw24347, Tw20570	Not Commercial
						Julbernardia pellegriniana Troupin	Tw29484	Tw18779	Commercial
			Lonchocarpus Kunth	1	1	Julbernardia seretii (De Wild.) Troupin	Tw5326, Tw2285, Tw3958, Tw2281	Tw1539, Tw1611	Not Commercial
						Lonchocarpus sericeus (Poir.) Kunth ex DC.	Tw28204	Tw11240	Not Commercial
			Millettia Wight & Arn.	15	7	Millettia drastica Welw. ex Baker	Tw891, Tw21965	Tw1198	Not Commercial
						Millettia dura Dunn	Tw4320, Tw905	Tw1213	Not Commercial
						Millettia eetveldeana (Micheli) Hauman	Tw4328, Tw8255	Tw2043	Not Commercial

						Millettia laurentii De Wild.	Tw532, Tw59847, Tw3894, Tw5227, Tw425	Tw1988, Tw281	Commercial
						Millettia stuhlmannii Taub.	Tw25676	Tw20822	Not Commercial
						Millettia versicolor Welw. ex Baker	Tw20679, Tw18, Tw5	Tw28	Not Commercial
			Newtonia Baill.	12	5	Newtonia aubrevillei (Pellegr.) Keay	Tw10414, Tw808, Tw10403	Tw3602	Not Commercial
						Newtonia buchananii (Baker) G.C.C.Gilbert & Boutique	Tw8095, Tw20708, Tw18879	Tw2202	Not Commercial
						Newtonia glandulifera (Pellegr.) G.C.C.Gilbert & Boutique	Tw8393, Tw7587, Tw1448	Tw7753	Not Commercial
						Newtonia leucocarpa (Harms) G.C.C.Gilbert & Boutique	Tw7340, Tw64766, Tw4821	Tw25765, Tw2133	Not Commercial
			Pachyelasma Harms	3	1	Pachyelasma tessmannii (Harms) Harms	Tw3959, Tw1246, Tw1163	Tw29478	Not Commercial
			Paramacrolobium J.Léonard	3	1	Paramacrolobium coeruleum (Taub.) J.Léonard	Tw697, Tw7699, Tw1129	Tw729	Not Commercial
			Parkia R.Br.	3	1	Parkia bicolor A.Chev.	Tw43409, Tw38609, Tw64617	Tw57203	Not Commercial
			Peltophorum (Vogel) Benth.	1	1	Peltophorum africanum Sond.	Tw26830	Tw20571	Not Commercial
			Pentaclethra Benth.	9	4	Pentaclethra eetveldeana De Wild. & T.Durand	Tw43406, Tw2524, Tw2527, Tw2525, Tw2522, Tw2528	Tw152, Tw1369, Tw2523	Not Commercial
						Pentaclethra macrophylla Benth.	Tw1458, Tw5205, Tw1720	Tw3946	Not Commercial
			Pericopsis Thwaites	9	4	Pericopsis angolensis (Baker) Meeuwen	Tw21116, Tw3888, Tw28843	Tw20572, Tw18888	Not Commercial
						Pericopsis elata (Harms) Meeuwen	Tw7648, Tw248, Tw1861, Tw1237, Tw3613, Tw1520	Tw1100, Tw17985	Commercial
			Piliostigma Hochst.	1	1	Piliostigma thonningii (Schumach.) Milne-Redh.	Tw28157	Tw20556	Not Commercial

			Piptadeniastrum Brenan	3	2	Piptadeniastrum africanum (Hook.f.) Brenan	Tw25103, Tw957, Tw960	Tw1533, Tw1985	Commercial
			Platysepalum Welw. ex Baker	2	2	Platysepalum chevalieri Harms	Tw9174	Tw5069	Not Commercial
						Platysepalum violaceum Welw. ex Baker	Tw4330	Tw33039	Not Commercial
			Prioria Griseb.	9	6	Prioria balsamifera (Vermoesen) Breteler	Tw51767, Tw33490, Tw3616	Tw1334, Tw1417	Commercial
						Prioria buchholzii (Harms) Breteler	Tw435, Tw29630	Tw1269	Not Commercial
						Prioria mannii (Baill.) Breteler	Tw7680	Tw11241	Not Commercial
						Prioria oxyphylla (Harms) Breteler	Tw60038, Tw60049, Tw60033	Tw60031, Tw1109	Commercial
			Pterocarpus Jacq.	38	14	Pterocarpus angolensis DC.	Tw344, Tw18724, Tw11369, Tw17187, Tw10286, Tw11636, Tw19284, Tw8261, Tw768, Tw1013, Tw11266, Tw11390	Tw8325, Tw11635, Tw11338, Tw19369, Tw20583	Not Commercial
						Pterocarpus rotundifolius (Sond.) Druce	Tw28199, Tw28123	Tw23063	Not Commercial
						Pterocarpus soyauxii Taub.	Tw29805, Tw1826, Tw955, Tw7654, Tw29088, Tw51777, Tw3760, Tw236, Tw392, Tw7671, Tw10773, Tw1131	Tw463, Tw836, Tw718, Tw346	Commercial
						Pterocarpus tinctorius Welw.	Tw1010, Tw1322, Tw13506, Tw26, Tw762, Tw1021, Tw20585, Tw11420, Tw11327, Tw30, Tw19386, Tw313	Tw41191, Tw1426, Tw29542, Tw24162	Commercial
			Senegalia Raf.	1	1	Senegalia senegal (L.) Britton	Tw41460	Tw23499	Not Commercial
			Senna Mill.	3	1	Senna siamea (Lam.) H.S.Irwin & Barneby	Tw951, Tw11106, Tw1032	Tw710	Not Commercial
			Tamarindus Tourn. ex L.	3	2	Tamarindus indica L.	Tw19625, Tw18301, Tw5163	Tw17418, Tw11018	Not Commercial

			Tessmannia Harms	8	6	Tessmannia africana Harms	Tw56967	Tw1272	Commercial
						Tessmannia anomala (Micheli) Harms	Tw2362, Tw3953	Tw14835	Commercial
						Tessmannia anomala var. flamignii J.Léonard	Tw2279	Tw2276	Not Commercial
						Tessmannia lescrauwaetii (De Wild.) Harms	Tw7173, Tw5295, Tw7228	Tw5211, Tw5231	Commercial
						Tessmannia yangambiensis Louis ex J.Léonard	Tw7199	Tw1189	Not Commercial
			Tetraberlinia (Harms) Hauman	3	2	Tetraberlinia bifoliolata (Harms) Hauman	Tw43754, Tw42963, Tw44694	Tw24415, Tw25691	Not Commercial
			Tetrapleura Benth.	3	2	Tetrapleura tetraptera (Schumach. & Thonn.) Taub.	Tw56, Tw61105, Tw53891	Tw120, Tw2197	Not Commercial
			Vachellia Wight & Arn.	2	2	Vachellia abyssinica (Hochst. ex Benth.) Kyal. & Boatwr.	Tw39177	Tw28151	Not Commercial
						Vachellia seyal (Delile) P.J.H.Hurter	Tw30353	Tw30352	Not Commercial
Gentianaceae	7	4	Anthocleista Afzel. ex R.Br.	7	4	Anthocleista grandiflora Gilg	Tw17189	Tw1185	Not Commercial
						Anthocleista nobilis G.Don	Tw3071, Tw3069, Tw3070	Tw25768, Tw3066	Not Commercial
						Anthocleista schweinfurthii Gilg	Tw1183, Tw29297, Tw1229	Tw23926	Not Commercial
Huaceae	1	1	Hua Pierre ex De Wild.	1	1	Hua gaboni Pierre ex De Wild.	Tw7995	Tw32243	Not Commercial
Hypericaceae	3	1	Harungana Lam.	3	1	Harungana madagascariensis Poir.	Tw871, Tw2500, Tw2564	Tw819	Not Commercial
Irvingiaceae	18	11	Desbordesia Pierre ex Tiegh.	3	2	Desbordesia glaucescens (Engl.) Tiegh.	Tw22741, Tw25705, Tw18758	Tw10793, Tw1208	Not Commercial
			Irvingia Hook.f.	12	7	Irvingia gabonensis (Aubry-Lecomte ex O'Rorke) Baill.	Tw8166, Tw5238, Tw9749	Tw1647, Tw260	Commercial
						Irvingia grandifolia (Engl.) Engl.	Tw2485, Tw137, Tw1117	Tw1415	Commercial
						Irvingia robur Mildbr.	Tw7980, Tw8235, Tw2487	Tw10145, Tw2486	Not Commercial

						Irvingia smithii Hook.fil.	Tw9783	Tw13339	Not Commercial
						Irvingia tenuinucleata Tiegh.	Tw7142, Tw9781	Tw3049	Not Commercial
			Klainedoxa Pierre ex Engl.	3	2	Klainedoxa gabonensis Pierre	Tw8279, Tw5332, Tw6954	Tw1435, Tw2398	Commercial
Ixonanthaceae	2	1	Phyllocosmus Klotzsch	2	1	Phyllocosmus africanus (Hook.fil.) Klotzsch	Tw2216, Tw9770	Tw1120	Not Commercial
Kirkiaceae	2	1	Kirkia Oliv.	2	1	Kirkia acuminata Oliv.	Tw28189, Tw24054	Tw18997	Not Commercial
Lamiaceae	31	14	Gmelina L.	9	3	Gmelina arborea Roxb. ex Sm.	Tw25291, Tw57692, Tw18090, Tw14064, Tw31785, Tw14451, Tw29514, Tw44885, Tw60284	Tw29515, Tw14280, Tw53935	Not Commercial
			Premna L.	3	1	Premna angolensis Gürke	Tw39150, Tw28127, Tw20731	Tw28477	Not Commercial
			Tectona L.f.	3	1	Tectona grandis L.f.	Tw767, Tw11055, Tw13935	Tw3805	Not Commercial
			Vitex L.	16	9	Vitex congolensis De Wild. & T.Durand	Tw20613, Tw7237	Tw10222	Not Commercial
						Vitex congolensis var. congolensis	Tw2039, Tw1791, Tw7575	Tw1563, Tw1114	Not Commercial
						Vitex doniana Sweet	Tw18592, Tw1245, Tw302	Tw1000, Tw10800	Not Commercial
						Vitex ferruginea Schumacher & Thonn.	Tw28639, Tw61051, Tw61425	Tw61373	Not Commercial
						Vitex madiensis Oliv.	Tw2033, Tw780	Tw56974	Not Commercial
						Vitex madiensis subsp. milanjiensis (Britten) F.White	Tw1825, Tw754	Tw28453	Not Commercial
						Vitex mombassae Vatke	Tw8527	Tw1146	Not Commercial
Lauraceae	16	8	Beilschmiedia Nees	13	7	Beilschmiedia congolana Robyns & R.Wilczek	Tw5210, Tw33511	Tw32863	Commercial
						Beilschmiedia corbisieri (Robyns) Robyns & R.Wilczek	Tw839, Tw543, Tw719	Tw1255, Tw513	Not Commercial
						Beilschmiedia louisii Robyns & R.Wilczek	Tw4334	Tw1099	Not Commercial

						Beilschmiedia mannii (Meisn.) Robyns & R.Wilczek	Tw44011, Tw10372, Tw12941	Tw30093	Not Commercial
						Beilschmiedia oblongifolia Robyns & R.Wilczek	Tw7457, Tw1607, Tw5082	Tw5083	Not Commercial
						Beilschmiedia ugandensis Rendle	Tw45063	Tw4201	Not Commercial
			Persea Mill.	3	1	Persea americana Mill.	Tw24643, Tw19159, Tw19160	Tw22468	Not Commercial
Lecythidaceae	3	2	Petersianthus Merr.	3	2	Petersianthus macrocarpus (P.Beauv.) Liben	Tw989, Tw815, Tw889	Tw282, Tw693	Commercial
Loganiaceae	7	4	Brenania Keay	2	1	Brenania brieyi (De Wild.) E.M.A.Petit	Tw66, Tw30902	Tw30897	Not Commercial
			Strychnos L.	5	3	Strychnos cocculoides Baker	Tw28270	Tw24361	Not Commercial
						Strychnos innocua Delile	Tw24360	Tw2399	Not Commercial
						Strychnos spinosa Lam.	Tw56901, Tw24357, Tw56904	Tw56902	Not Commercial
Malvaceae	46	29	Adansonia L.	1	1	Adansonia digitata L.	Tw53835	Tw25915	Not Commercial
			Ceiba Mill.	3	2	Ceiba pentandra (L.) Gaertn.	Tw622, Tw8159, Tw8138	Tw50225, Tw251	Not Commercial
			Cola Schott & Endl.	9	6	Cola ballayi Cornu ex Tschirch & O.Oesterle	Tw8486, Tw2217, Tw62026	Tw6953	Not Commercial
						Cola cordifolia (Cav.) R.Br.	Tw41336	Tw30698	Not Commercial
						Cola gigantea A.Chev.	Tw56987	Tw22813	Not Commercial
						Cola lateritia K.Schum.	Tw9175, Tw8224	Tw10338	Not Commercial
						Cola nitida (Vent.) Schott & Endl.	Tw57419	Tw57182	Not Commercial
						Cola welwitschii Exell & Mendonça	Tw29586	Tw28133	Not Commercial
			Desplatsia Bocq.	3	1	Desplatsia suberica Bocq.	Tw7313, Tw185, Tw1860	Tw6975	Not Commercial
			Dombeya Cav.	4	2	Dombeya rotundifolia (Hochst.) Planch.	Tw26043, Tw24381, Tw44530	Tw26110	Not Commercial
						Dombeya torrida (J.F.Gmel.) Bamps	Tw24195	Tw19993	Not Commercial
			Duboscia Bocq.	3	1	Duboscia viridiflora (K.Schum.) Mildbr.	Tw29808, Tw30889, Tw62418	Tw43837	Not Commercial
			Grewia L.	1	1	Grewia louisii R.Wilczek	Tw8225	Tw7544	Not Commercial
			Heritiera Aiton	1	1	Heritiera littoralis Dryand. ex Aiton	Tw13932	Tw11632	Not Commercial

			Microcos L.	4	3	Microcos coriacea (Mast.) Burret	Tw2645	Tw2644	Not Commercial
						Microcos pinnatifida (Mast.) Burret	Tw2647, Tw2643, Tw7253	Tw1897, Tw187	Not Commercial
			Nesogordonia Baill.	6	4	Nesogordonia kabingaensis (K.Schum.) Capuron	Tw64765, Tw3634, Tw266	Tw129, Tw1512	Commercial
						Nesogordonia papaverifera (A.Chev.) Capuron	Tw14117, Tw14118, Tw14060	Tw11446, Tw13942	Commercial
			Pterygota Schott & Endl.	5	3	Pterygota bequaertii De Wild.	Tw25719, Tw2434	Tw1138	Commercial
						Pterygota macrocarpa K.Schum.	Tw2721, Tw14057, Tw21879	Tw11082, Tw11443	Commercial
			Sterculia L.	4	3	Sterculia quinqueloba (Garcke) K.Schum.	Tw732	Tw2247	Not Commercial
						Sterculia tragacantha Lindl.	Tw7693, Tw5088, Tw6961	Tw28250, Tw1475	Not Commercial
			Triplochiton K.Schum.	2	1	Triplochiton scleroxylon K.Schum.	Tw879, Tw68416	Tw65071	Commercial
Melastomataceae	1	1	Dichaetanthera Endl.	1	1	Dichaetanthera corymbosa (Cogn.) Jacq.-Fél.	Tw33391	Tw24202	Not Commercial
Meliaceae	152	62	Carapa Aubl.	6	2	Carapa procera DC.	Tw7506, Tw22440, Tw257, Tw8101, Tw909, Tw7104	Tw67, Tw1603	Not Commercial
			Ekebergia Sparrm.	5	3	Ekebergia benguelensis Welw. ex C.DC.	Tw28179, Tw9815	Tw24219	Not Commercial
						Ekebergia capensis Sparrm.	Tw9785, Tw738, Tw9800	Tw20915, Tw23849	Not Commercial
			Entandrophragma C.DC.	56	22	Entandrophragma angolense (Welw.) C.DC.	Tw243, Tw554, Tw530, Tw133, Tw713, Tw1550, Tw5354, Tw2536, Tw1212, Tw1543, Tw384, Tw378, Tw8484, Tw1982	Tw1432, Tw3630, Tw1118, Tw7468, Tw43	Commercial
						Entandrophragma candollei Harms	Tw5036, Tw5035, Tw1653, Tw9805, Tw32292, Tw566, Tw5032, Tw1535, Tw350, Tw5031, Tw20824	Tw9755, Tw14044, Tw11747, Tw51782	Commercial

						Entandrophragma cylindricum (Sprague) Sprague	Tw866, Tw14076, Tw941, Tw1324, Tw1043, Tw912, Tw11069, Tw527, Tw11465	Tw10476, Tw233, Tw867	Commercial
						Entandrophragma delevoyi De Wild.	Tw11394, Tw310, Tw20588	Tw26881	Not Commercial
						Entandrophragma excelsum (Dawe & Sprague) Sprague	Tw1232, Tw14125, Tw750, Tw8096	Tw1057, Tw1059	Not Commercial
						Entandrophragma palustre Staner	Tw733, Tw862, Tw904	Tw605, Tw1160	Not Commercial
						Entandrophragma utile (Dawe & Sprague) Sprague	Tw8192, Tw1534, Tw134, Tw2161, Tw20826, Tw914, Tw18783, Tw1315, Tw3632, Tw14046, Tw1102, Tw1329	Tw1410, Tw245, Tw9801, Tw13391, Tw32368	Commercial
			Khaya A.Juss.	14	6	Khaya anthotheca (Welw.) C.DC.	Tw8381, Tw26913, Tw9787, Tw209, Tw44570, Tw9814, Tw8382, Tw558	Tw215, Tw45, Tw838	Commercial
						Khaya grandifoliola C.DC.	Tw39363, Tw7355, Tw7351, Tw18904, Tw43840, Tw50773	Tw18193, Tw22771, Tw14507	Commercial
			Leplaea Vermoesen	30	10	Leplaea cedrata (A.Chev.) E.J.M.Koenen & J.J.de Wilde	Tw20827, Tw32811, Tw18784, Tw43852, Tw50299, Tw26497, Tw17991, Tw1305, Tw85, Tw660, Tw7455, Tw1655, Tw14047, Tw8322, Tw29916	Tw406, Tw12958, Tw14077, Tw295, Tw7484	Commercial
						Leplaea laurentii (De Wild.) E.J.M.Koenen & J.J.de Wilde	Tw373, Tw338, Tw621	Tw603	Not Commercial
						Leplaea thompsonii (Sprague & Hutch.) E.J.M.Koenen & J.J.de Wilde	Tw1190, Tw7262, Tw26919, Tw32715, Tw5338, Tw7157, Tw46414, Tw9799, Tw26905, Tw2030, Tw1110, Tw1388	Tw9752, Tw3233, Tw5339, Tw5145	Commercial
			Lovoa Harms	17	6	Lovoa trichilioides Harms	Tw18786, Tw32236, Tw29794, Tw449, Tw7622, Tw20604, Tw11065, Tw593,	Tw14049, Tw51784, Tw26496, Tw10779, Tw39119, Tw9798	Commercial

							Tw17993, Tw12932, Tw7358, Tw533, Tw20829, Tw25775, Tw29906, Tw1465, Tw613		
			Trichilia P. Browne	17	10	Trichilia dregeana Sond.	Tw39121	Tw22775	Not Commercial
						Trichilia emetica subsp. emetica	Tw39147	Tw13289	Not Commercial
						Trichilia gilgiana Harms	Tw29632	Tw1911	Not Commercial
						Trichilia gillettii De Wild.	Tw33536, Tw7540, Tw33154	Tw34956	Not Commercial
						Trichilia monadelpha (Thonn.) J. De Wild.	Tw22969, Tw33357, Tw2428	Tw29964	Not Commercial
						Trichilia prieuriana A. Juss.	Tw1896, Tw104, Tw29868	Tw2055	Not Commercial
						Trichilia rubescens Oliv.	Tw2876, Tw3284, Tw86	Tw21947, Tw2874	Not Commercial
						Trichilia tessmannii Harms	Tw29883	Tw2044	Not Commercial
						Trichilia welwitschii C. DC.	Tw7322	Tw7178	Not Commercial
			Turraeanthus Baill.	7	3	Turraeanthus africanus (Welw. ex C. DC.) Pellegr.	Tw20830, Tw54557, Tw1635, Tw2410, Tw858, Tw29968, Tw26490	Tw17994, Tw2413, Tw64413	Commercial
Moraceae	44	22	Antiaris Lesch.	7	4	Antiaris toxicaria (Pers.) Lesch.	Tw31032, Tw7338, Tw7568	Tw25, Tw11457	Commercial
						Antiaris toxicaria subsp. welwitschii (Engl.) C. C. Berg	Tw2266	Tw2112	Not Commercial
						Antiaris toxicaria var. africana Scott Elliot ex A. Chev.	Tw14512, Tw17995, Tw53862	Tw18911	Not Commercial
			Artocarpus J.R. Forst. & G. Forst.	1	1	Artocarpus altilis (Parkinson) Fosberg	Tw18579	Tw11626	Not Commercial
			Bosqueiopsis De Wild. & T. Durand	2	1	Bosqueiopsis gillettii De Wild. & T. Durand	Tw8244, Tw7218	Tw7130	Not Commercial
			Ficus L.	18	10	Ficus bubu Warb.	Tw4824, Tw990, Tw7602	Tw2020, Tw173	Not Commercial
						Ficus demeusei Warb.	Tw7740	Tw28805	Not Commercial
						Ficus elastica Roxb.	Tw48761	Tw44660	Not Commercial
						Ficus lutea Vahl	Tw11628, Tw47922, Tw53945	Tw53828	Not Commercial

						Ficus mucoso Welw. ex Ficaltho	Tw1807, Tw1923	Tw1411	Not Commercial
						Ficus sur Forssk.	Tw28210	Tw2111	Not Commercial
						Ficus sycomorus L.	Tw26530	Tw13325	Not Commercial
						Ficus thonningii Blume	Tw24368, Tw7529, Tw24026	Tw33888	Not Commercial
						Ficus vogeliana (Miq.) Miq.	Tw30668, Tw7942, Tw3609	Tw6956	Not Commercial
			Milicia Sim	9	3	Milicia excelsa (Welw.) C.C.Berg	Tw54822, Tw58, Tw51741, Tw1462, Tw256, Tw785, Tw235, Tw1463, Tw1121	Tw54832, Tw1416, Tw391	Commercial
			Treculia Decne.	2	1	Treculia africana Decne. ex Trécul	Tw53901, Tw57422	Tw3608	Not Commercial
Myristicaceae	15	6	Trilepisium Thouars	5	2	Trilepisium madagascariense DC.	Tw40, Tw32949, Tw992, Tw59682, Tw291	Tw25108, Tw20715	Not Commercial
			Coelocaryon Warb.	6	2	Coelocaryon botryoides Vermoesen	Tw1707, Tw948, Tw615	Tw8405	Not Commercial
						Coelocaryon preussii Warb.	Tw1644, Tw94, Tw135	Tw1849	Not Commercial
			Pycnanthus Warb.	3	1	Pycnanthus angolensis (Welw.) Exell	Tw17999, Tw14052, Tw10336	Tw1437	Commercial
			Staudtia Warb.	6	3	Staudtia kamerunensis Warb.	Tw64739, Tw61419, Tw61366	Tw61457	Not Commercial
Myrtaceae	12	7	Syzygium Gaertn.	12	7	Staudtia kamerunensis var. gabonensis (Warb.) Fouilloy	Tw51752, Tw57429, Tw59844	Tw473, Tw328	Commercial
						Syzygium aromaticum (L.) Merr. & L.M.Perry	Tw8185	Tw1034	Not Commercial
						Syzygium cordatum Hochst.	Tw27770, Tw28873, Tw766	Tw17196, Tw23906	Not Commercial
						Syzygium guineense (Willd.) DC.	Tw59423, Tw57682	Tw21948	Not Commercial
						Syzygium owariense (P.Beauv.) Benth.	Tw39115, Tw30941	Tw24353	Not Commercial
						Syzygium parvifolium (Engl.) Mildbr.	Tw31491	Tw26066	Not Commercial
						Syzygium staudtii (Engl.) Mildbr.	Tw176, Tw33557, Tw33058	Tw33517	Not Commercial

Ochnaceae	6	3	Lophira Banks ex C.F.Gaertn.	6	3	Lophira alata ex Banks C.F.Gaertn.	Tw11463, Tw374, Tw11060, Tw731	Tw10767, Tw10345	Commercial
						Lophira lanceolata Tiegh. ex Keay	Tw30064, Tw887	Tw27540	Not Commercial
Olacaceae	29	13	Coula Baill.	3	2	Coula edulis Baill.	Tw7536, Tw1940, Tw25713	Tw10758, Tw10346	Not Commercial
			Diogoia Exell ex Mendonça	1	1	Diogoia zenkeri (Engl.) Exell & Mendonça	Tw726	Tw3686	Not Commercial
			Heisteria Jacq.	3	1	Heisteria parvifolia Sm.	Tw1798, Tw29791, Tw10139	Tw2407	Not Commercial
			Okoubaka Pellegr. & Normand	3	1	Okoubaka aubrevillei Pellegr. & Normand	Tw6577, Tw21883, Tw8498	Tw7135	Not Commercial
			Ongokea Pierre	6	2	Ongokea gore (Hua) Pierre	Tw247, Tw824, Tw407, Tw56966, Tw470, Tw206	Tw197, Tw389	Commercial
			Strombosia Blume	9	3	Strombosia grandifolia Hook.fil. ex Benth.	Tw1291, Tw1456, Tw97	Tw2015	Not Commercial
						Strombosia pustulata var. pustulata	Tw1258, Tw1899, Tw10335	Tw18794	Not Commercial
						Strombosia scheffleri Engl.	Tw7504, Tw19983, Tw20718	Tw7270	Not Commercial
			Strombosiosis Engl.	3	2	Strombosiosis tetrandra Engl.	Tw1902, Tw1987, Tw1713	Tw10162, Tw1122	Not Commercial
			Ximenia L.	1	1	Ximenia americana L.	Tw28424	Tw28372	Not Commercial
Oleaceae	6	3	Olea L.	6	3	Olea capensis subsp. macrocarpa (C.H.Wright) I. Verd.	Tw19991, Tw23492, Tw18196	Tw23490	Not Commercial
						Olea europaea subsp. cuspidata (Wall. & G.Don) Cif.	Tw19995	Tw19984	Not Commercial
						Olea welwitschii (Knobl.) Gilg & G.Schellenb.	Tw39082, Tw39094	Tw28211	Not Commercial
Pandaceae	1	1	Microdesmis Planch.	1	1	Microdesmis kasaiensis J.Léonard	Tw8081	Tw7316	Not Commercial
Passifloraceae	3	1	Barteria Hook.f.	3	1	Barteria nigritana Hook.fil.	Tw7209, Tw7701, Tw8094	Tw7746	Not Commercial
Penaeaceae	2	1	Psydrax Gaertn.	2	1	Psydrax palma (K.Schum.) Bridson	Tw722, Tw355	Tw10158	Not Commercial
Peraceae	2	1	Chaetocarpus Thwaites	2	1	Chaetocarpus africanus Pax	Tw8074, Tw8043	Tw7924	Not Commercial
Phyllanthaceae	57	32	Antidesma L.	5	3	Antidesma laciniatum Müll.Arg.	Tw62507, Tw62516	Tw32891	Not Commercial

						Antidesma membranaceum Müll.Arg.	Tw7332, Tw23522, Tw13377	Tw10464, Tw10205	Not Commercial
			Bridelia Willd.	10	5	Bridelia atroviridis Müll.Arg.	Tw829, Tw1856, Tw2077	Tw7719	Not Commercial
						Bridelia brideliifolia (Pax) Fedde	Tw44655	Tw21953	Not Commercial
						Bridelia ferruginea Benth.	Tw39022, Tw28437, Tw28463	Tw10207, Tw22732	Not Commercial
						Bridelia micrantha (Hochst.) Baill.	Tw24387, Tw17168, Tw8221	Tw28138	Not Commercial
			Cleistanthus Hook.f. ex Planch.	7	4	Cleistanthus caudatus Pax	Tw8449, Tw62504, Tw62495, Tw62477, Tw9776	Tw50653, Tw62435	Not Commercial
						Cleistanthus inundatus J.Léonard	Tw40348	Tw33366	Not Commercial
						Cleistanthus polystachyus Hook.f. ex Planch.	Tw8099	Tw1604	Not Commercial
			Hymenocardia Wall. ex Lindl.	5	3	Hymenocardia acida Tul.	Tw7203, Tw23624	Tw10217	Not Commercial
						Hymenocardia ulmoides Oliv.	Tw7220, Tw28202, Tw8078	Tw1006, Tw10184	Not Commercial
			Maesobotrya Benth.	1	1	Maesobotrya staudtii (Pax) Hutch.	Tw39892	Tw184	Not Commercial
			Margaritaria L.f.	1	1	Margaritaria discoidea (Baill.) G.L.Webster	Tw59600	Tw1444	Not Commercial
			Phyllanthus L.	3	1	Phyllanthus physocarpus Müll.Arg.	Tw9778, Tw1127, Tw10198	Tw33620	Not Commercial
			Pseudolachnostyli s Pax	3	1	Pseudolachnostyli s maprouneifolia Pax	Tw24221, Tw28175, Tw28232	Tw28180	Not Commercial
			Uapaca Baill.	22	13	Uapaca guineensis Müll.Arg.	Tw59820, Tw61424, Tw61372	Tw380, Tw59813	Not Commercial
						Uapaca heudelotii Baill.	Tw8247, Tw53895, Tw9768	Tw26721, Tw195	Not Commercial
						Uapaca kirkiana Müll.Arg.	Tw1684, Tw24371	Tw1596	Not Commercial
						Uapaca mole Pax	Tw368, Tw79, Tw30884, Tw8172	Tw1423, Tw196	Not Commercial
						Uapaca nitida Müll.Arg.	Tw772, Tw13501, Tw13514	Tw306	Not Commercial
						Uapaca pilosa Hutch.	Tw676, Tw23602	Tw23600	Not Commercial
						Uapaca robynsii De Wild.	Tw23917	Tw23913	Not Commercial

						Uapaca sansibarica Pax	Tw24359, Tw44707	Tw24305	Not Commercial
						Uapaca togoensis Pax	Tw49444	Tw30053	Not Commercial
						Uapaca vanhouttei De Wild.	Tw954	Tw63	Not Commercial
Proteaceae	3	2	Faurea Harv.	3	2	Faurea rochetiana (A.Rich.) Chiov. ex Pic.Serm.	Tw751	Tw23977	Not Commercial
						Faurea saligna Harv.	Tw324, Tw7497	Tw24169	Not Commercial
Putranjivaceae	7	3	Drypetes Vahl	7	3	Drypetes angustifolia Pax & K.Hoffm.	Tw62412, Tw62577	Tw62406	Not Commercial
						Drypetes gerrardii Hutch.	Tw33866, Tw23517	Tw23503	Not Commercial
						Drypetes gossweileri S.Moore	Tw1126, Tw1302, Tw1219	Tw1284	Not Commercial
Rhamnaceae	4	3	Maesopsis Engl.	3	2	Maesopsis eminii Engl.	Tw1192, Tw28149, Tw1393	Tw1103, Tw1170	Commercial
			Ziziphus Mill.	1	1	Ziziphus abyssinica Hochst. ex A.Rich.	Tw60205	Tw20593	Not Commercial
Rhizophoraceae	11	7	Anopyxis Pierre ex Engl.	3	2	Anopyxis klaineana (Pierre) Engl.	Tw3103, Tw711, Tw3109	Tw2511, Tw2515	Not Commercial
			Cassipourea Aubl.	5	3	Cassipourea congoensis R.Br. ex DC.	Tw24194, Tw39872	Tw24024	Not Commercial
						Cassipourea gummiflua Tul.	Tw31482	Tw31022	Not Commercial
						Cassipourea malosana (Baker) Alston	Tw26100, Tw30363	Tw19990	Not Commercial
			Rhizophora L.	3	2	Rhizophora racemosa G.Mey.	Tw347, Tw25567, Tw7	Tw14054, Tw22177	Not Commercial
Rosaceae	4	2	Hagenia J.F.Gmel.	2	1	Hagenia abyssinica (Bruce) J.F.Gmel.	Tw740, Tw21932	Tw21621	Not Commercial
			Prunus L.	2	1	Prunus africana (Hook.fil.) Kalkman	Tw18972, Tw3583	Tw17207	Not Commercial
Rubiaceae	62	33	Aidia Lour.	1	1	Aidia ochroleuca (K.Schum.) E.M.A.Petit	Tw7441	Tw1304	Not Commercial
			Aulacocalyx Hook.f.	2	1	Aulacocalyx jasmiflora Hook.f.	Tw33037, Tw32945	Tw23009	Not Commercial
			Coffea L.	2	1	Coffea liberica W.Bull	Tw25823, Tw28578	Tw25652	Not Commercial
			Corynanthe Welw.	6	4	Corynanthe macroceras K.Schum.	Tw289, Tw7516, Tw35920	Tw254, Tw10150	Not Commercial

					Corynanthe paniculata Welw.	Tw977, Tw288, Tw983,	Tw1442, Tw154	Not Commercial	
			Craterispermum Benth.	2	2	Craterispermum cerinanthum Hiern	Tw62542	Tw41594	Not Commercial
						Craterispermum laurinum (Poir.) Benth.	Tw38625	Tw30111	Not Commercial
			Crossopteryx Fenzl	3	1	Crossopteryx febrifuga (Afzel. ex G.Don) Benth.	Tw8079, Tw1959, Tw28443	Tw7293	Not Commercial
			Gardenia J.Ellis	4	2	Gardenia imperialis K.Schum.	Tw24241, Tw36120, Tw40359	Tw39101	Not Commercial
						Gardenia ternifolia subsp. jovis- tonantis (Welw.) Verdc.	Tw28429	Tw28237	Not Commercial
			Heinsia DC.	3	1	Heinsia crinita (Wennberg) G.Taylor	Tw59364, Tw59360, Tw59359	Tw59362	Not Commercial
			Massularia (K.Schu m.) Hoyle	2	1	Massularia acuminata (G.Don) Bullock ex Hoyle	Tw40329, Tw8294	Tw33202	Not Commercial
			Mitragyna Korth.	12	5	Mitragyna ledermannii (K.Krause) Ridsdale	Tw29825, Tw108, Tw14055	Tw14082	Not Commercial
						Mitragyna rubrostipulata (K.Schum.) Havil.	Tw26094, Tw31476	Tw25679	Not Commercial
						Mitragyna stipulosa (DC.) Kuntze	Tw1161, Tw546, Tw276, Tw438, Tw241, Tw32, Tw349	Tw293, Tw12, Tw520	Commercial
			Morinda L.	5	4	Morinda chrysorrhiza (Thonn.) DC.	Tw1267	Tw102	Not Commercial
						Morinda citrifolia L.	Tw691	Tw55585	Not Commercial
						Morinda lucida Benth.	Tw846, Tw635, Tw728	Tw491, Tw608	Not Commercial
			Nauclea L.	10	4	Nauclea diderrichii (De Wild.) Merr.	Tw5209, Tw32869, Tw263, Tw17, Tw32256, Tw383	Tw1144, Tw315	Commercial
						Nauclea latifolia Sm.	Tw12947	Tw10363	Not Commercial
						Nauclea pobeguinii (Hua ex Pobég.) Merr.	Tw1962, Tw30041, Tw712	Tw34	Not Commercial
			Oxyanthus DC.	1	1	Oxyanthus tubiflorus (Andrews) DC.	Tw894	Tw32546	Not Commercial

			Psychotria L.	1	1	Psychotria dermatophylla (K.Schum.) E.M.A.Petit	Tw966	Tw821	Not Commercial
			Rothmannia Thunb.	4	2	Rothmannia longiflora Salisb.	Tw59623	Tw33023	Not Commercial
						Rothmannia lujae (De Wild.) Keay	Tw109, Tw8036, Tw1671	Tw41588	Not Commercial
			Schumanniphyton Harms	2	1	Schumanniphyton magnificum (K.Schum.) Harms	Tw40372, Tw68428	Tw40364	Not Commercial
			Tricalysia A.Rich. ex DC.	2	1	Tricalysia pallens Hiern	Tw30246, Tw32306	Tw28555	Not Commercial
Rutaceae	15	7	Citrus L.	4	2	Citrus x aurantium L.	Tw25991, Tw3782	Tw23577	Not Commercial
						Citrus x limon (L.) Osbeck	Tw60223, Tw42211	Tw1950	Not Commercial
			Vepris Comm. ex A.Juss.	1	1	Vepris louisii G.C.C.Gilbert	Tw7525	Tw2418	Not Commercial
			Zanthoxylum L.	10	4	Zanthoxylum chalybeum Engl.	Tw760, Tw319	Tw21933	Not Commercial
						Zanthoxylum gillettii (De Wild.) P.G.Waterman	Tw9820, Tw60730, Tw61094	Tw61374	Commercial
						Zanthoxylum heitzii (Aubrév. & Pellegr.) P.G.Waterman	Tw98, Tw18798, Tw25749	Tw61993	Not Commercial
						Zanthoxylum lemairei (De Wild.) P.G.Waterman	Tw43831, Tw29298	Tw13227	Not Commercial
Salicaceae	12	5	Homalium Jacq.	12	5	Homalium abdessammadii Asch. & Schweinf.	Tw2046, Tw757	Tw1874	Not Commercial
						Homalium africanum (Hook.fil.) Benth.	Tw2052, Tw6960, Tw92	Tw709	Not Commercial
						Homalium letestui Pellegr.	Tw26607, Tw30383	Tw18858	Not Commercial
						Homalium longistylum Mast.	Tw1171, Tw698, Tw4955	Tw69	Not Commercial
						Homalium stipulaceum Welw. ex Mast.	Tw7971, Tw32277	Tw10147	Not Commercial
Sapindaceae	21	13	Allophylus L.	2	2	Allophylus africanus P.Beauv.	Tw28154	Tw1461	Not Commercial
						Allophylus dummeri Baker fil.	Tw34780	Tw33534	Not Commercial
			Blighia K.D.Koenig	3	1	Blighia welwitschii (Hiern) Radlk.	Tw1424, Tw1116, Tw1669	Tw1652	Not Commercial
			Chytranthus Hook.f.	3	2	Chytranthus carneus Radlk.	Tw8437	Tw34807	Not Commercial

						Chytranthus setosus Radlk.	Tw19379, Tw1859	Tw10421	Not Commercial
			Haplocoelum Radlk.	1	1	Haplocoelum intermedium Hauman	Tw8231	Tw174	Not Commercial
			Lecaniodiscus Planch. ex Benth.	2	1	Lecaniodiscus cupanioides Planch.	Tw8433, Tw8308	Tw10213	Not Commercial
			Majidea J.Kirk ex Oliv.	1	1	Majidea fosteri (Sprague) Radlk.	Tw8369	Tw8134	Not Commercial
			Pancovia Willd.	4	2	Pancovia floribunda Pellegr.	Tw827	Tw7956	Not Commercial
						Pancovia laurentii (De Wild.) Gilg ex De Wild.	Tw2931, Tw1220, Tw2934	Tw2933	Not Commercial
			Sapindus L.	2	1	Sapindus rarak DC.	Tw44659, Tw20973	Tw10229	Not Commercial
			Zanha Hiern	3	2	Zanha golungensis Hiern	Tw32011, Tw39959, Tw29876	Tw2080, Tw10454	Not Commercial
Sapotaceae	76	42	Aningeria Aubrév. & Pellegr.	7	3	Aningeria adolfi-friederici (Engl.) Robyns & Gilbert	Tw2555	Tw2554	Not Commercial
						Aningeria altissima (A.Chev.) Aubrév. & Pellegr.	Tw13234, Tw22809, Tw51779	Tw5147	Commercial
						Aningeria pierrei (A.Chev.) Aubrév. & Pellegr.	Tw8049, Tw29934, Tw14523	Tw64760	Commercial
			Autranella A.Chev.	6	3	Autranella congolensis (De Wild.) A.Chev.	Tw57417, Tw633, Tw51736, Tw299, Tw515, Tw430	Tw32272, Tw1175, Tw13496	Commercial
			Baillonella Pierre	4	2	Baillonella toxisperma Pierre	Tw1675, Tw51781, Tw46, Tw59843	Tw1666, Tw1673	Commercial
			Brevia Aubrév. & Pellegr.	1	1	Brevia sericea Aubrév. & Pellegr.	Tw30679	Tw29904	Not Commercial
			Donella Pierre ex Baill.	1	1	Donella pruniformis (Engl.) Pierre ex Engl.	Tw7300	Tw1457	Not Commercial
			Englerophytum K.Krause	3	2	Englerophytum laurentii (De Wild.) L.Gaut.	Tw959, Tw614	Tw33333	Not Commercial
						Englerophytum magalismsontanum (Sond.) T.D.Penn.	Tw26912	Tw26104	Not Commercial
			Gambeya Pierre	19	10	Gambeya africana (A.DC.) Pierre	Tw51786, Tw61383, Tw33010, Tw61335, Tw61110	Tw1132, Tw1571	Commercial

					Gambeya albida (G.Don) Aubrév. & Pellegr.	Tw25112, Tw18990	Tw18989	Not Commercial
					Gambeya beguei (Aubrév. & Pellegr.) Aubrév. & Pellegr.	Tw31499	Tw1705	Not Commercial
					Gambeya lacourtiana (De Wild.) Aubrév. & Pellegr.	Tw13, Tw1674, Tw971	Tw57424	Commercial
					Gambeya lungi (De Wild.) Aubrév. & Pellegr.	Tw820, Tw996	Tw1303	Not Commercial
					Gambeya perpulchra (Mildbr. ex Hutch. & Dalziel) Aubrév. & Pellegr.	Tw57667, Tw32713, Tw53867	Tw29938, Tw20726	Not Commercial
					Gambeya subnuda (Baker) Pierre	Tw1853, Tw2678, Tw2679	Tw1130, Tw101	Not Commercial
			Letestua Lecomte	1	Letestua durissima (A.Chev.) Lecomte	Tw13884	Tw13326	Not Commercial
			Malacantha Pierre	1	Malacantha alnifolia (Baker) Pierre	Tw59528	Tw53936	Not Commercial
			Manilkara Adans.	4	Manilkara moehria (Baker) Dubard	Tw61732	Tw41091	Not Commercial
					Manilkara obovata (Sabine & G.Don) J.H.Hemsl.	Tw30896, Tw21560, Tw28583	Tw13231, Tw20727	Not Commercial
			Mimusops L.	2	Mimusops zeyheri Sond.	Tw759, Tw26522	Tw1991	Not Commercial
			Neolemonniera Heine	1	Neolemonniera clatandriifolia (A.Chev.) Heine	Tw937	Tw8560	Not Commercial
			Omphalocarpum P.Beauv.	3	Omphalocarpum briei De Wild.	Tw93, Tw5538	Tw2206	Not Commercial
					Omphalocarpum lecomteanum Pierre ex Engl.	Tw63032	Tw34779	Not Commercial
			Synsepalum (A.DC.) Daniell	9	Synsepalum afzelii (Engl.) T.D.Penn.	Tw29980, Tw7230	Tw10176	Not Commercial
					Synsepalum brevipes (Baker) T.D.Penn.	Tw8533	Tw30074	Not Commercial
					Synsepalum revolutum (Baker) T.D.Penn.	Tw7722	Tw13946	Not Commercial
					Synsepalum stipulatum (Radlk.) Engl.	Tw8391, Tw8424	Tw7327	Not Commercial

						Synsepalum subcordatum De Wild.	Tw1150, Tw62740, Tw62838	Tw62806	Not Commercial
			Tieghemella Pierre	9	4	Tieghemella africana Pierre	Tw25779, Tw20837, Tw18800	Tw10761, Tw15000	Commercial
						Tieghemella heckelii (A.Chev.) Pierre ex Dubard	Tw26510, Tw18005, Tw31670, Tw55117, Tw14526, Tw21571	Tw19951, Tw12949	Commercial
			Tridesmostemon Engl.	3	1	Tridesmostemon omphalocarpoides Engl.	Tw67108, Tw29894, Tw5237	Tw57057	Not Commercial
			Vitellaria C.F.Gaertn.	2	1	Vitellaria paradoxa C.F.Gaertn.	Tw7620, Tw26914	Tw18988	Not Commercial
Simaroubaceae	5	3	Odyndea Engl.	2	1	Odyndea gabunensis (Pierre) Engl.	Tw62535, Tw8029	Tw62438	Not Commercial
			Pierreodendron Engl.	3	2	Pierreodendron africanum (Hook.fil.) Little	Tw7725, Tw7531, Tw8055	Tw5288, Tw25760	Not Commercial
Sladeniaceae	1	1	Ficalhoa Hiern	1	1	Ficalhoa laurifolia Hiern	Tw24199	Tw22510	Not Commercial
Stilbaceae	1	1	Nuxia Comm. ex Lam.	1	1	Nuxia congesta R.Br.	Tw26542	Tw13211	Not Commercial
Ulmaceae	3	2	Holoptelea Planch.	3	2	Holoptelea grandis (Hutch.) Mildbr.	Tw969, Tw1836, Tw1876	Tw1419, Tw1698	Commercial
Urticaceae	9	5	Musanga R.Br.	3	2	Musanga cecropioides R.Br. ex Tedlie	Tw20, Tw1847, Tw1898	Tw1180, Tw1468	Not Commercial
			Myrianthus P.Beauv.	6	3	Myrianthus arboreus P.Beauv.	Tw7566, Tw36, Tw8220	Tw82	Not Commercial
						Myrianthus holstii Engl.	Tw2261, Tw872	Tw17234	Not Commercial
						Myrianthus preussii Engl.	Tw33243	Tw28794	Not Commercial
Violaceae	1	1	Rinorea Aubl.	1	1	Rinorea welwitschii (Oliv.) Kuntze	Tw34805	Tw32909	Not Commercial
Vochysiaceae	3	1	Erismadelphus Mildbr.	3	1	Erismadelphus exsul Mildbr.	Tw725, Tw25756, Tw30910	Tw33352	Not Commercial
Zygophyllaceae	2	1	Balanites Delile	2	1	Balanites wilsoniana Dawe & Sprague	Tw8404, Tw9779	Tw8287	Not Commercial

Table 8.2: Hyperparameters grid for finding the optimal parameter settings across all machine-learning classifiers in Chapter 3.

Classifier	Hyperparameter	Full Description	Functionality	Values
DT	max_depth	Maximum depth of the tree, controlling how deep the tree can grow before splitting stops.	Unlimited depth ('None') raises complexity, risking overfitting, while limited depth can lead to improved generalization	None, 10, 20, 30
	min_samples_split	Minimum number of samples required to split an internal node.	Larger values prevent the tree from growing too complex.	2, 5, 10
	min_samples_leaf	Minimum number of samples required to be at a leaf node.	Larger values simplify the model by preventing overly small and noisy splits, improving generalization	1, 2, 4
SVM	C	Regularization parameter that controls the trade-off between a smooth decision boundary and classifying training points correctly.	Lower values increase regularization (smoother boundary), while higher values prioritize fitting the training data	0.1, 1, 10, 100
	kernel	Specifies the kernel type used in the algorithm (linear, radial basis function (RBF), polynomial, or sigmoid).	Specifies the function used to map data into higher-dimensional space, allowing the model to capture either linear or non-linear relationships in the data.	linear, radial basis function (RBF), polynomial, or sigmoid
	gamma	Kernel coefficient for non-linear kernels (RBF, poly, sigmoid)	This controls the influence of individual training examples on the decision boundary in non-linear kernel methods.	scale', 'auto'
RF	n_estimators	The number of trees in the RF.	More trees can improve accuracy but may increase computational cost.	50, 100, 200
	max_depth	Maximum depth of each tree in the forest, limiting how deep the tree can grow.	Unlimited depth ('None') raises complexity, risking overfitting, while limited depth can lead to improved generalization	None, 10, 20
	min_samples_split	Minimum number of samples required to split an internal node.	Larger values prevent the tree from growing too complex.	2, 5, 10
	min_samples_leaf	Minimum number of samples required at a leaf node.	Larger values simplify the model by preventing overly small and noisy splits, improving generalization	1, 2, 4
Catboost	learning_rate	Step size used to update weights, controlling how quickly the model learns.	This parameter governs the model's learning rate. Lower values slow down learning but enhance accuracy, while higher values accelerate learning at the cost of accuracy.	0.01, 0.05, 0.1
	depth	Depth of each tree, determining the complexity of the model.	Increasing tree depth enhances pattern recognition but risks overfitting, while shallow trees generalize well but may overlook complex relationships.	3, 4, 6
	l2_leaf_reg	Coefficient for L2 regularization	Leaf weights are regularized to prevent overfitting by penalizing large leaf values, encouraging simpler models with better generalization.	1, 3, 5
	border_count	The number of splits for numerical features.	Higher values increase the model's granularity in handling numerical data, potentially improving accuracy but at a higher computational cost.	32, 64, 128

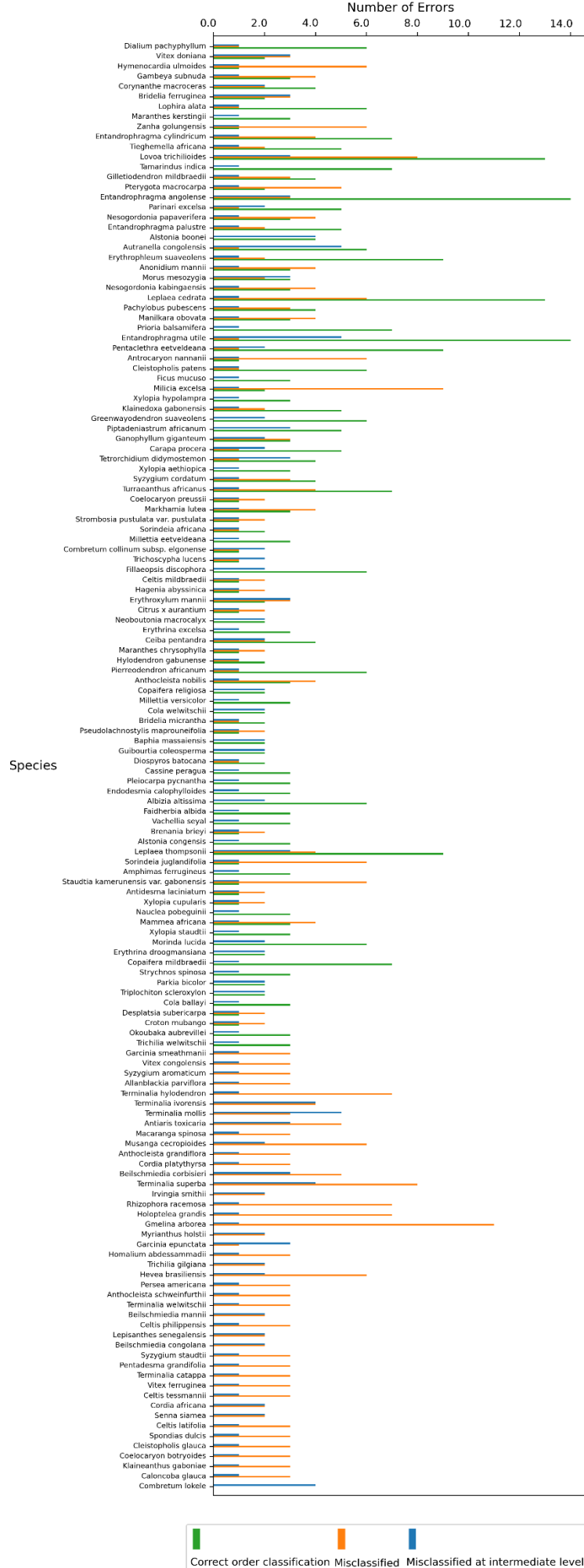


Figure 8.2: Horizontal grouped bar chart showing classification outcomes for 144 species with observed taxonomic mismatches. Each species is represented by three bars indicating the number of specimens that fall into one of the following categories: Blue (mismatched): Specimens with partial classification errors—at least one taxonomic level is correct, but others are incorrect (e.g., Family and Species correct, Genus wrong; or Genus correct, Family and Species wrong); Yellow (fully misclassified): Specimens that were misclassified at all three taxonomic levels—Family, Genus, and Species; Green (correct): Specimens correctly classified in a hierarchical order—Family must be correct, with Genus and/or Species also correct (e.g., Family correct but Genus and Species wrong; Family and Genus correct, Species wrong; or all three levels correct). This visualization helps identify patterns of classification accuracy and inconsistency across taxonomic levels.

Table 8.3: Summary of accuracy par species across taxonomic levels, showing for each species what percentages of their specimens was correctly identified by the RF model on Species, Genus, or Family level and what percentage of each species was misidentified.

Species	Correct Species	Genus	Correct Genus	Family	Correct Family	Completely Wrong
<i>Caloncoba glauca</i>	25	<i>Caloncoba</i>	25	Achariaceae	25	75
<i>Caloncoba welwitschii</i>	0	<i>Caloncoba</i>	0	Achariaceae	0	100
<i>Lepisanthes senegalensis</i>	50	<i>Lepisanthes</i>	50	Achariaceae	50	50
<i>Lindackeria bukobensis</i>	0	<i>Lindackeria</i>	0	Achariaceae	0	100
<i>Lindackeria dentata</i>	0	<i>Lindackeria</i>	0	Achariaceae	0	100
<i>Scottellia klaineana</i>	0	<i>Scottellia</i>	0	Achariaceae	0	100
<i>Antrocaryon micraster</i>	0	<i>Antrocaryon</i>	0	Anacardiaceae	25	75
<i>Antrocaryon nannanii</i>	12.5	<i>Antrocaryon</i>	25	Anacardiaceae	25	75
<i>Clausena anisata</i>	0	<i>Clausena</i>	0	Anacardiaceae	0	100
<i>Ganophyllum giganteum</i>	62.5	<i>Ganophyllum</i>	62.5	Anacardiaceae	62.5	37.5
<i>Lannea welwitschii</i>	0	<i>Lannea</i>	0	Anacardiaceae	0	100
<i>Mangifera indica</i>	0	<i>Mangifera</i>	0	Anacardiaceae	0	100
<i>Myrsine melanophloeos</i>	100	<i>Myrsine</i>	100	Anacardiaceae	100	0
<i>Panda oleosa</i>	0	<i>Panda</i>	0	Anacardiaceae	0	100
<i>Pseudospondias longifolia</i>	0	<i>Pseudospondias</i>	0	Anacardiaceae	75	25
<i>Pseudospondias microcarpa</i>	0	<i>Pseudospondias</i>	0	Anacardiaceae	0	100
<i>Santiria trimera</i>	0	<i>Santiria</i>	0	Anacardiaceae	0	100
<i>Sorindeia africana</i>	0	<i>Sorindeia</i>	75	Anacardiaceae	75	25
<i>Sorindeia juglandifolia</i>	12.5	<i>Sorindeia</i>	25	Anacardiaceae	25	75
<i>Spondias dulcis</i>	25	<i>Spondias</i>	25	Anacardiaceae	25	75
<i>Trichoscypha acuminata</i>	50	<i>Trichoscypha</i>	50	Anacardiaceae	50	50
<i>Trichoscypha lucens</i>	75	<i>Trichoscypha</i>	75	Anacardiaceae	75	25
<i>Trichoscypha oddonii</i>	25	<i>Trichoscypha</i>	75	Anacardiaceae	75	25
<i>Xymalos monospora</i>	0	<i>Xymalos</i>	0	Anacardiaceae	0	100
<i>Anisophyllea boehmii</i>	0	<i>Anisophyllea</i>	0	Anisophylleaceae	0	100
<i>Annickia affinis</i>	0	<i>Annickia</i>	0	Annonaceae	75	25
<i>Annickia lebrunii</i>	0	<i>Annickia</i>	0	Annonaceae	75	25
<i>Anonidium mannii</i>	12.5	<i>Anonidium</i>	12.5	Annonaceae	50	50
<i>Brieya fasciculata</i>	0	<i>Brieya</i>	0	Annonaceae	75	25
<i>Cleistopholis glauca</i>	0	<i>Cleistopholis</i>	25	Annonaceae	25	75
<i>Cleistopholis patens</i>	87.5	<i>Cleistopholis</i>	87.5	Annonaceae	87.5	12.5
<i>Duguetia staudtii</i>	25	<i>Duguetia</i>	25	Annonaceae	75	25
<i>Greenwayodendron suaveolens</i>	100	<i>Greenwayodendron</i>	100	Annonaceae	100	0
<i>Hexalobus crispiflorus</i>	62.5	<i>Hexalobus</i>	62.5	Annonaceae	100	0
<i>Hexalobus monopetalus</i>	0	<i>Hexalobus</i>	0	Annonaceae	25	75
<i>Isolona congolana</i>	50	<i>Isolona</i>	50	Annonaceae	100	0
<i>Isolona hexaloba</i>	0	<i>Isolona</i>	75	Annonaceae	100	0
<i>Monodora angolensis</i>	0	<i>Monodora</i>	0	Annonaceae	100	0
<i>Platymitra arborea</i>	75	<i>Platymitra</i>	75	Annonaceae	100	0

<i>Xylopia aethiopica</i>	25	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia aurantioidora</i>	100	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia cupularis</i>	0	<i>Xylopia</i>	50	Annonaceae	50	50
<i>Xylopia flamignii</i>	75	<i>Xylopia</i>	75	Annonaceae	100	0
<i>Xylopia gilbertii</i>	0	<i>Xylopia</i>	75	Annonaceae	75	25
<i>Xylopia hypolampira</i>	25	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia katangensis</i>	0	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia phloiodora</i>	0	<i>Xylopia</i>	75	Annonaceae	75	25
<i>Xylopia rubescens</i>	0	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia staudtii</i>	0	<i>Xylopia</i>	100	Annonaceae	100	0
<i>Xylopia wilwerthii</i>	0	<i>Xylopia</i>	25	Annonaceae	25	75
<i>Alstonia boonei</i>	12.5	<i>Alstonia</i>	100	Apocynaceae	100	0
<i>Alstonia congensis</i>	100	<i>Alstonia</i>	100	Apocynaceae	100	0
<i>Diplorhynchus condylocarpon</i>	0	<i>Diplorhynchus</i>	0	Apocynaceae	50	50
<i>Funtumia elastica</i>	0	<i>Funtumia</i>	0	Apocynaceae	0	100
<i>Holarrhena floribunda</i>	25	<i>Holarrhena</i>	25	Apocynaceae	25	75
<i>Hunteria umbellata</i>	50	<i>Hunteria</i>	50	Apocynaceae	50	50
<i>Picralima nitida</i>	0	<i>Picralima</i>	0	Apocynaceae	0	100
<i>Pleiocarpa pycnantha</i>	100	<i>Pleiocarpa</i>	100	Apocynaceae	100	0
<i>Rauvolfia caffra</i>	0	<i>Rauvolfia</i>	25	Apocynaceae	50	50
<i>Rauvolfia vomitoria</i>	0	<i>Rauvolfia</i>	0	Apocynaceae	75	25
<i>Tabernaemontana crassa</i>	25	<i>Tabernaemontana</i>	50	Apocynaceae	50	50
<i>Tabernaemontana pachysiphon</i>	0	<i>Tabernaemontana</i>	25	Apocynaceae	25	75
<i>Voacanga africana</i>	0	<i>Voacanga</i>	0	Apocynaceae	0	100
<i>Voacanga thouarsii</i>	0	<i>Voacanga</i>	0	Apocynaceae	0	100
<i>Cussonia arborea</i>	0	<i>Cussonia</i>	0	Araliaceae	0	100
<i>Cussonia spicata</i>	0	<i>Cussonia</i>	0	Araliaceae	0	100
<i>Polyscias fulva</i>	0	<i>Polyscias</i>	0	Araliaceae	0	100
<i>Brenandendron donianum</i>	0	<i>Brenandendron</i>	0	Asteraceae	0	100
<i>Gymnanthemum amygdalinum</i>	0	<i>Gymnanthemum</i>	0	Asteraceae	0	100
<i>Fernandoa adolfi-friderici</i>	0	<i>Fernandoa</i>	0	Bignoniaceae	25	75
<i>Kigelia africana</i>	0	<i>Kigelia</i>	0	Bignoniaceae	0	100
<i>Markhamia lutea</i>	37.5	<i>Markhamia</i>	50	Bignoniaceae	50	50
<i>Markhamia obtusifolia</i>	0	<i>Markhamia</i>	0	Bignoniaceae	0	100
<i>Markhamia tomentosa</i>	0	<i>Markhamia</i>	75	Bignoniaceae	75	25
<i>Markhamia zanzibarica</i>	0	<i>Markhamia</i>	25	Bignoniaceae	25	75
<i>Spathodea campanulata</i>	0	<i>Spathodea</i>	0	Bignoniaceae	0	100
<i>Stereospermum harmsianum</i>	0	<i>Stereospermum</i>	0	Bignoniaceae	0	100
<i>Stereospermum kunthianum</i>	0	<i>Stereospermum</i>	0	Bignoniaceae	0	100
<i>Cordia africana</i>	25	<i>Cordia</i>	50	Boraginaceae	50	50
<i>Cordia millenii</i>	0	<i>Cordia</i>	0	Boraginaceae	0	100
<i>Cordia monoica</i>	0	<i>Cordia</i>	0	Boraginaceae	0	100
<i>Cordia myxa</i>	0	<i>Cordia</i>	0	Boraginaceae	0	100

<i>Cordia platythyrsa</i>	25	<i>Cordia</i>	25	Boraginaceae	25	75
<i>Aucoumea klaineana</i>	0	<i>Aucoumea</i>	0	Burseraceae	37.5	62.5
<i>Canarium schweinfurthii</i>	62.5	<i>Canarium</i>	75	Burseraceae	87.5	12.5
<i>Commiphora mollis</i>	0	<i>Commiphora</i>	0	Burseraceae	75	25
<i>Pachylobus buettneri</i>	75	<i>Pachylobus</i>	75	Burseraceae	75	25
<i>Pachylobus edulis</i>	0	<i>Pachylobus</i>	0	Burseraceae	75	25
<i>Pachylobus igaganga</i>	0	<i>Pachylobus</i>	25	Burseraceae	25	75
<i>Pachylobus normandii</i>	0	<i>Pachylobus</i>	0	Burseraceae	0	100
<i>Pachylobus osika</i>	0	<i>Pachylobus</i>	0	Burseraceae	0	100
<i>Pachylobus pubescens</i>	0	<i>Pachylobus</i>	62.5	Burseraceae	62.5	37.5
<i>Endodesmia calophylloides</i>	100	<i>Endodesmia</i>	100	Calophyllaceae	100	0
<i>Mammea africana</i>	50	<i>Mammea</i>	50	Calophyllaceae	50	50
<i>Celtis gomphophylla</i>	50	<i>Celtis</i>	50	Cannabaceae	50	50
<i>Celtis latifolia</i>	0	<i>Celtis</i>	25	Cannabaceae	25	75
<i>Celtis mildbraedii</i>	25	<i>Celtis</i>	50	Cannabaceae	50	50
<i>Celtis philippensis</i>	0	<i>Celtis</i>	25	Cannabaceae	25	75
<i>Celtis tessmannii</i>	25	<i>Celtis</i>	25	Cannabaceae	25	75
<i>Celtis zenkeri</i>	100	<i>Celtis</i>	100	Cannabaceae	100	0
<i>Morus mesozygia</i>	75	<i>Morus</i>	75	Cannabaceae	75	25
<i>Trema orientale</i>	0	<i>Trema</i>	0	Cannabaceae	0	100
<i>Apodytes dimidiata</i>	0	<i>Apodytes</i>	0	Celastraceae	0	100
<i>Cassine peragua</i>	100	<i>Cassine</i>	100	Celastraceae	100	0
<i>Magnistipula butayi</i>	0	<i>Magnistipula</i>	0	Chrysobalanaceae	75	25
<i>Maranthes chrysophylla</i>	0	<i>Maranthes</i>	25	Chrysobalanaceae	50	50
<i>Maranthes gabunensis</i>	0	<i>Maranthes</i>	0	Chrysobalanaceae	0	100
<i>Maranthes glabra</i>	0	<i>Maranthes</i>	0	Chrysobalanaceae	0	100
<i>Maranthes kerstingii</i>	75	<i>Maranthes</i>	100	Chrysobalanaceae	100	0
<i>Parinari congensis</i>	0	<i>Parinari</i>	100	Chrysobalanaceae	100	0
<i>Parinari curatellifolia</i>	37.5	<i>Parinari</i>	75	Chrysobalanaceae	75	25
<i>Parinari excelsa</i>	0	<i>Parinari</i>	62.5	Chrysobalanaceae	87.5	12.5
<i>Allanblackia floribunda</i>	0	<i>Allanblackia</i>	0	Clusiaceae	0	100
<i>Allanblackia kisonghi</i>	0	<i>Allanblackia</i>	0	Clusiaceae	25	75
<i>Allanblackia parviflora</i>	0	<i>Allanblackia</i>	25	Clusiaceae	25	75
<i>Garcinia chromocarpa</i>	0	<i>Garcinia</i>	0	Clusiaceae	0	100
<i>Garcinia epunctata</i>	75	<i>Garcinia</i>	75	Clusiaceae	75	25
<i>Garcinia huillensis</i>	0	<i>Garcinia</i>	0	Clusiaceae	0	100
<i>Garcinia ovalifolia</i>	0	<i>Garcinia</i>	0	Clusiaceae	0	100
<i>Garcinia punctata</i>	0	<i>Garcinia</i>	0	Clusiaceae	0	100
<i>Garcinia smeathmanii</i>	0	<i>Garcinia</i>	25	Clusiaceae	25	75
<i>Lebrunia busbaie</i>	0	<i>Lebrunia</i>	0	Clusiaceae	0	100
<i>Pentadesma butyracea</i>	0	<i>Pentadesma</i>	0	Clusiaceae	0	100
<i>Pentadesma grandifolia</i>	0	<i>Pentadesma</i>	25	Clusiaceae	25	75
<i>Symphonia globulifera</i>	0	<i>Symphonia</i>	0	Clusiaceae	0	100

<i>Combretum adenogonium</i>	0	<i>Combretum</i>	25	Combretaceae	25	75
<i>Combretum collinum</i>	0	<i>Combretum</i>	0	Combretaceae	0	100
<i>Combretum collinum subsp. elgonense</i>	0	<i>Combretum</i>	75	Combretaceae	75	25
<i>Combretum lokele</i>	0	<i>Combretum</i>	100	Combretaceae	100	0
<i>Combretum molle</i>	0	<i>Combretum</i>	75	Combretaceae	75	25
<i>Combretum zeyheri</i>	100	<i>Combretum</i>	100	Combretaceae	100	0
<i>Terminalia anisoptera</i>	0	<i>Terminalia</i>	25	Combretaceae	25	75
<i>Terminalia brachystemma</i>	0	<i>Terminalia</i>	0	Combretaceae	0	100
<i>Terminalia catappa</i>	0	<i>Terminalia</i>	25	Combretaceae	25	75
<i>Terminalia hylodendron</i>	0	<i>Terminalia</i>	12.5	Combretaceae	12.5	87.5
<i>Terminalia ivorensis</i>	25	<i>Terminalia</i>	50	Combretaceae	50	50
<i>Terminalia leiocarpa</i>	0	<i>Terminalia</i>	0	Combretaceae	0	100
<i>Terminalia macroptera</i>	0	<i>Terminalia</i>	0	Combretaceae	0	100
<i>Terminalia mollis</i>	25	<i>Terminalia</i>	62.5	Combretaceae	62.5	37.5
<i>Terminalia sericea</i>	0	<i>Terminalia</i>	0	Combretaceae	0	100
<i>Terminalia superba</i>	33.33333333	<i>Terminalia</i>	33.33333333	Combretaceae	33.33333333	66.66666667
<i>Terminalia welwitschii</i>	0	<i>Terminalia</i>	25	Combretaceae	25	75
<i>Plagiostyles africana</i>	0	<i>Plagiostyles</i>	0	Daphniphyllaceae	0	100
<i>Marquesia macroura</i>	100	<i>Marquesia</i>	100	Dipterocarpaceae	100	0
<i>Monotes hypoleucus var. angolensis</i>	0	<i>Monotes</i>	0	Dipterocarpaceae	25	75
<i>Monotes katangensis</i>	0	<i>Monotes</i>	0	Dipterocarpaceae	0	100
<i>Diospyros batocana</i>	25	<i>Diospyros</i>	50	Ebenaceae	75	25
<i>Diospyros crassiflora</i>	0	<i>Diospyros</i>	0	Ebenaceae	0	100
<i>Diospyros dendo</i>	0	<i>Diospyros</i>	25	Ebenaceae	25	75
<i>Diospyros ferrea</i>	0	<i>Diospyros</i>	0	Ebenaceae	0	100
<i>Diospyros iturensis</i>	0	<i>Diospyros</i>	0	Ebenaceae	0	100
<i>Diospyros mespiliformis</i>	0	<i>Diospyros</i>	0	Ebenaceae	0	100
<i>Erythroxylum mannii</i>	62.5	<i>Erythroxylum</i>	62.5	Erythroxylaceae	62.5	37.5
<i>Cavacoa quintasii</i>	0	<i>Cavacoa</i>	0	Euphorbiaceae	0	100
<i>Croton haumanianus</i>	0	<i>Croton</i>	25	Euphorbiaceae	25	75
<i>Croton macrostachyus</i>	0	<i>Croton</i>	0	Euphorbiaceae	0	100
<i>Croton mayumbensis</i>	0	<i>Croton</i>	0	Euphorbiaceae	0	100
<i>Croton megalocarpus</i>	0	<i>Croton</i>	0	Euphorbiaceae	0	100
<i>Croton mubango</i>	0	<i>Croton</i>	50	Euphorbiaceae	50	50
<i>Croton sylvaticus</i>	0	<i>Croton</i>	0	Euphorbiaceae	25	75
<i>Dichostemma glaucescens</i>	0	<i>Dichostemma</i>	0	Euphorbiaceae	0	100
<i>Discoglypsemna caloneura</i>	0	<i>Discoglypsemna</i>	0	Euphorbiaceae	75	25
<i>Grossera macrantha</i>	0	<i>Grossera</i>	0	Euphorbiaceae	37.5	62.5
<i>Hevea brasiliensis</i>	25	<i>Hevea</i>	25	Euphorbiaceae	25	75
<i>Klaineanthus gaboniae</i>	0	<i>Klaineanthus</i>	25	Euphorbiaceae	25	75
<i>Macaranga kilimandscharica</i>	0	<i>Macaranga</i>	0	Euphorbiaceae	0	100
<i>Macaranga monandra</i>	0	<i>Macaranga</i>	25	Euphorbiaceae	75	25
<i>Macaranga spinosa</i>	0	<i>Macaranga</i>	25	Euphorbiaceae	25	75

<i>Maprounea africana</i>	75	<i>Maprounea</i>	75	Euphorbiaceae	75	25
<i>Neoboutonia macrocalyx</i>	75	<i>Neoboutonia</i>	75	Euphorbiaceae	100	0
<i>Ricinodendron heudelotii</i>	100	<i>Ricinodendron</i>	100	Euphorbiaceae	100	0
<i>Ricinodendron heudelotii</i> subsp. <i>africanum</i>	25	<i>Ricinodendron</i>	25	Euphorbiaceae	75	25
<i>Schinziophyton rautanenii</i>	87.5	<i>Schinziophyton</i>	87.5	Euphorbiaceae	100	0
<i>Sclerocroton cornutus</i>	0	<i>Sclerocroton</i>	0	Euphorbiaceae	50	50
<i>Shirakiopsis elliptica</i>	0	<i>Shirakiopsis</i>	0	Euphorbiaceae	75	25
<i>Tetrorchidium didymostemon</i>	50	<i>Tetrorchidium</i>	62.5	Euphorbiaceae	87.5	12.5
<i>Afzelia africana</i>	100	<i>Afzelia</i>	100	Fabaceae	100	0
<i>Afzelia bella</i>	50	<i>Afzelia</i>	100	Fabaceae	100	0
<i>Afzelia bipindensis</i>	58.33333333	<i>Afzelia</i>	91.66666667	Fabaceae	100	0
<i>Afzelia pachyloba</i>	25	<i>Afzelia</i>	66.66666667	Fabaceae	100	0
<i>Afzelia quanzensis</i>	37.5	<i>Afzelia</i>	87.5	Fabaceae	100	0
<i>Albizia adianthifolia</i>	12.5	<i>Albizia</i>	87.5	Fabaceae	100	0
<i>Albizia adianthifolia</i> var. <i>intermedia</i>	0	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia altissima</i>	62.5	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia antunesiana</i>	0	<i>Albizia</i>	50	Fabaceae	87.5	12.5
<i>Albizia chinensis</i>	0	<i>Albizia</i>	100	Fabaceae	100	0
<i>Albizia coriaria</i>	0	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia ferruginea</i>	62.5	<i>Albizia</i>	100	Fabaceae	100	0
<i>Albizia glaberrima</i>	75	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia glaberrima</i> var. <i>glabrescens</i>	0	<i>Albizia</i>	25	Fabaceae	100	0
<i>Albizia gummifera</i>	25	<i>Albizia</i>	100	Fabaceae	100	0
<i>Albizia gummifera</i> var. <i>ealaensis</i>	0	<i>Albizia</i>	0	Fabaceae	100	0
<i>Albizia laurentii</i>	25	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia lebbeck</i>	0	<i>Albizia</i>	75	Fabaceae	100	0
<i>Albizia schimperiana</i>	100	<i>Albizia</i>	100	Fabaceae	100	0
<i>Albizia versicolor</i>	25	<i>Albizia</i>	37.5	Fabaceae	100	0
<i>Albizia zygia</i>	0	<i>Albizia</i>	50	Fabaceae	100	0
<i>Amphimas ferrugineus</i>	75	<i>Amphimas</i>	100	Fabaceae	100	0
<i>Amphimas pterocarpoides</i>	100	<i>Amphimas</i>	100	Fabaceae	100	0
<i>Annea laxiflora</i>	100	<i>Annea</i>	100	Fabaceae	100	0
<i>Anthonothea brieyi</i>	0	<i>Anthonothea</i>	0	Fabaceae	100	0
<i>Anthonothea fragrans</i>	0	<i>Anthonothea</i>	0	Fabaceae	100	0
<i>Anthonothea pyraertii</i>	0	<i>Anthonothea</i>	0	Fabaceae	100	0
<i>Aphanocalyx cynometroides</i>	0	<i>Aphanocalyx</i>	0	Fabaceae	100	0
<i>Aphanocalyx microphyllus</i>	0	<i>Aphanocalyx</i>	0	Fabaceae	25	75
<i>Baikiaea insignis</i>	12.5	<i>Baikiaea</i>	12.5	Fabaceae	75	25
<i>Baikiaea robynsii</i>	0	<i>Baikiaea</i>	0	Fabaceae	50	50
<i>Baphia bequaertii</i>	0	<i>Baphia</i>	0	Fabaceae	87.5	12.5
<i>Baphia dewevrei</i>	0	<i>Baphia</i>	0	Fabaceae	0	100
<i>Baphia massaiensis</i>	50	<i>Baphia</i>	75	Fabaceae	100	0
<i>Baphia nitida</i>	0	<i>Baphia</i>	0	Fabaceae	100	0

<i>Baphia pubescens</i>	0	<i>Baphia</i>	0	Fabaceae	100	0
<i>Berlinia bracteosa</i>	25	<i>Berlinia</i>	50	Fabaceae	100	0
<i>Berlinia confusa</i>	50	<i>Berlinia</i>	50	Fabaceae	100	0
<i>Berlinia congolensis</i>	0	<i>Berlinia</i>	25	Fabaceae	100	0
<i>Berlinia grandiflora</i>	0	<i>Berlinia</i>	62.5	Fabaceae	100	0
<i>Bobgunnia fistuloides</i>	50	<i>Bobgunnia</i>	62.5	Fabaceae	87.5	12.5
<i>Bobgunnia madagascariensis</i>	0	<i>Bobgunnia</i>	25	Fabaceae	50	50
<i>Brachystegia boehmii</i>	0	<i>Brachystegia</i>	25	Fabaceae	100	0
<i>Brachystegia bussei</i>	0	<i>Brachystegia</i>	75	Fabaceae	100	0
<i>Brachystegia laurentii</i>	62.5	<i>Brachystegia</i>	62.5	Fabaceae	100	0
<i>Brachystegia longifolia</i>	12.5	<i>Brachystegia</i>	87.5	Fabaceae	100	0
<i>Brachystegia manga</i>	0	<i>Brachystegia</i>	75	Fabaceae	100	0
<i>Brachystegia spiciformis</i>	25	<i>Brachystegia</i>	75	Fabaceae	100	0
<i>Brachystegia tamarindoides</i> subsp. <i>microphylla</i>	0	<i>Brachystegia</i>	100	Fabaceae	100	0
<i>Brachystegia tamarindoides</i> subsp. <i>tamarindoides</i>	0	<i>Brachystegia</i>	100	Fabaceae	100	0
<i>Brachystegia taxifolia</i>	25	<i>Brachystegia</i>	50	Fabaceae	100	0
<i>Brachystegia utilis</i>	0	<i>Brachystegia</i>	25	Fabaceae	100	0
<i>Burkea africana</i>	25	<i>Burkea</i>	25	Fabaceae	100	0
<i>Cassia mannii</i>	0	<i>Cassia</i>	0	Fabaceae	100	0
<i>Cassia sieberiana</i>	0	<i>Cassia</i>	0	Fabaceae	100	0
<i>Copaifera mildbraedii</i>	50	<i>Copaifera</i>	50	Fabaceae	100	0
<i>Copaifera religiosa</i>	75	<i>Copaifera</i>	75	Fabaceae	100	0
<i>Craibia affinis</i>	0	<i>Craibia</i>	0	Fabaceae	100	0
<i>Craibia grandiflora</i>	0	<i>Craibia</i>	0	Fabaceae	100	0
<i>Craibia lujae</i>	0	<i>Craibia</i>	0	Fabaceae	25	75
<i>Crudia harmsiana</i>	0	<i>Crudia</i>	0	Fabaceae	75	25
<i>Crudia laurentii</i>	0	<i>Crudia</i>	25	Fabaceae	100	0
<i>Cryptosepalum exfoliatum</i> subsp. <i>pseudotaxus</i>	100	<i>Cryptosepalum</i>	100	Fabaceae	100	0
<i>Cylicodiscus gabunensis</i>	0	<i>Cylicodiscus</i>	12.5	Fabaceae	100	0
<i>Cynometra alexandri</i>	50	<i>Cynometra</i>	50	Fabaceae	75	25
<i>Cynometra hankei</i>	50	<i>Cynometra</i>	75	Fabaceae	100	0
<i>Cynometra lujae</i>	100	<i>Cynometra</i>	100	Fabaceae	100	0
<i>Cynometra mannii</i>	75	<i>Cynometra</i>	100	Fabaceae	100	0
<i>Cynometra sessiliflora</i>	0	<i>Cynometra</i>	25	Fabaceae	62.5	37.5
<i>Dalbergia boehmii</i>	0	<i>Dalbergia</i>	0	Fabaceae	75	25
<i>Daniellia alsteeniana</i>	0	<i>Daniellia</i>	50	Fabaceae	100	0
<i>Daniellia klainei</i>	0	<i>Daniellia</i>	25	Fabaceae	100	0
<i>Daniellia oliveri</i>	0	<i>Daniellia</i>	0	Fabaceae	100	0
<i>Daniellia pynaertii</i>	25	<i>Daniellia</i>	25	Fabaceae	75	25
<i>Daniellia soyauxii</i>	75	<i>Daniellia</i>	100	Fabaceae	100	0
<i>Dialium englerianum</i>	0	<i>Dialium</i>	0	Fabaceae	75	25
<i>Dialium excelsum</i>	87.5	<i>Dialium</i>	87.5	Fabaceae	100	0
<i>Dialium pachyphyllum</i>	25	<i>Dialium</i>	50	Fabaceae	87.5	12.5

<i>Dialium pentandrum</i>	0	<i>Dialium</i>	75	Fabaceae	75	25
<i>Dialium polyanthum</i>	0	<i>Dialium</i>	0	Fabaceae	25	75
<i>Dialium tessmannii</i>	0	<i>Dialium</i>	25	Fabaceae	50	50
<i>Dialium zenkeri</i>	0	<i>Dialium</i>	0	Fabaceae	25	75
<i>Dichrostachys cinerea</i>	50	<i>Dichrostachys</i>	50	Fabaceae	87.5	12.5
<i>Entada abyssinica</i>	0	<i>Entada</i>	0	Fabaceae	25	75
<i>Erythrina abyssinica</i>	75	<i>Erythrina</i>	75	Fabaceae	75	25
<i>Erythrina droogmansiana</i>	50	<i>Erythrina</i>	75	Fabaceae	100	0
<i>Erythrina excelsa</i>	100	<i>Erythrina</i>	100	Fabaceae	100	0
<i>Erythrina orophila</i>	0	<i>Erythrina</i>	0	Fabaceae	25	75
<i>Erythrophleum africanum</i>	37.5	<i>Erythrophleum</i>	37.5	Fabaceae	100	0
<i>Erythrophleum suaveolens</i>	16.66666667	<i>Erythrophleum</i>	16.66666667	Fabaceae	83.33333333	16.66666667
<i>Faidherbia albida</i>	100	<i>Faidherbia</i>	100	Fabaceae	100	0
<i>Fillaeopsis discophora</i>	25	<i>Fillaeopsis</i>	37.5	Fabaceae	100	0
<i>Gilbertiodendron dewevrei</i>	50	<i>Gilbertiodendron</i>	50	Fabaceae	100	0
<i>Gilbertiodendron grandiflorum</i>	25	<i>Gilbertiodendron</i>	75	Fabaceae	100	0
<i>Gilbertiodendron grandistipulatum</i>	25	<i>Gilbertiodendron</i>	75	Fabaceae	100	0
<i>Gilbertiodendron mayombense</i>	0	<i>Gilbertiodendron</i>	0	Fabaceae	25	75
<i>Gilbertiodendron ogoouense</i>	62.5	<i>Gilbertiodendron</i>	75	Fabaceae	100	0
<i>Gilletiodendron kisantuense</i>	75	<i>Gilletiodendron</i>	75	Fabaceae	75	25
<i>Gilletiodendron mildbraedii</i>	12.5	<i>Gilletiodendron</i>	12.5	Fabaceae	62.5	37.5
<i>Guibourtia arnoldiana</i>	75	<i>Guibourtia</i>	75	Fabaceae	100	0
<i>Guibourtia coleosperma</i>	75	<i>Guibourtia</i>	75	Fabaceae	100	0
<i>Guibourtia demeusei</i>	75	<i>Guibourtia</i>	100	Fabaceae	100	0
<i>Guibourtia ehie</i>	0	<i>Guibourtia</i>	0	Fabaceae	100	0
<i>Hylodendron gabunense</i>	50	<i>Hylodendron</i>	50	Fabaceae	75	25
<i>Hymenostegia mundungu</i>	25	<i>Hymenostegia</i>	50	Fabaceae	75	25
<i>Intsia bijuga</i> var. <i>bijuga</i>	12.5	<i>Intsia</i>	12.5	Fabaceae	100	0
<i>Isobertlinia angolensis</i>	50	<i>Isobertlinia</i>	50	Fabaceae	100	0
<i>Isobertlinia doka</i>	0	<i>Isobertlinia</i>	50	Fabaceae	100	0
<i>Isobertlinia tomentosa</i>	0	<i>Isobertlinia</i>	25	Fabaceae	100	0
<i>Julbernardia brieyi</i>	0	<i>Julbernardia</i>	25	Fabaceae	100	0
<i>Julbernardia globiflora</i>	0	<i>Julbernardia</i>	25	Fabaceae	100	0
<i>Julbernardia paniculata</i>	0	<i>Julbernardia</i>	62.5	Fabaceae	100	0
<i>Julbernardia pellegriniana</i>	100	<i>Julbernardia</i>	100	Fabaceae	100	0
<i>Julbernardia seretii</i>	0	<i>Julbernardia</i>	25	Fabaceae	37.5	62.5
<i>Lonchocarpus sericeus</i>	0	<i>Lonchocarpus</i>	0	Fabaceae	100	0
<i>Millettia drastica</i>	0	<i>Millettia</i>	25	Fabaceae	50	50
<i>Millettia dura</i>	25	<i>Millettia</i>	75	Fabaceae	75	25
<i>Millettia eetveldeana</i>	100	<i>Millettia</i>	100	Fabaceae	100	0
<i>Millettia laurentii</i>	87.5	<i>Millettia</i>	87.5	Fabaceae	87.5	12.5
<i>Millettia stuhlmannii</i>	0	<i>Millettia</i>	50	Fabaceae	100	0
<i>Millettia versicolor</i>	75	<i>Millettia</i>	100	Fabaceae	100	0

<i>Newtonia aubrevillei</i>	0	<i>Newtonia</i>	0	Fabaceae	100	0
<i>Newtonia buchananii</i>	0	<i>Newtonia</i>	25	Fabaceae	75	25
<i>Newtonia glandulifera</i>	0	<i>Newtonia</i>	0	Fabaceae	0	100
<i>Newtonia leucocarpa</i>	0	<i>Newtonia</i>	0	Fabaceae	25	75
<i>Pachyelasma tessmannii</i>	0	<i>Pachyelasma</i>	0	Fabaceae	50	50
<i>Paramacrolobium coeruleum</i>	0	<i>Paramacrolobium</i>	0	Fabaceae	75	25
<i>Parkia bicolor</i>	50	<i>Parkia</i>	75	Fabaceae	100	0
<i>Peltophorum africanum</i>	0	<i>Peltophorum</i>	0	Fabaceae	100	0
<i>Pentaclethra eetveldeana</i>	75	<i>Pentaclethra</i>	75	Fabaceae	91.66666667	8.333333333
<i>Pentaclethra macrophylla</i>	0	<i>Pentaclethra</i>	0	Fabaceae	100	0
<i>Pericopsis angolensis</i>	0	<i>Pericopsis</i>	25	Fabaceae	37.5	62.5
<i>Pericopsis elata</i>	62.5	<i>Pericopsis</i>	75	Fabaceae	75	25
<i>Piliostigma thonningii</i>	0	<i>Piliostigma</i>	0	Fabaceae	100	0
<i>Piptadeniastrum africanum</i>	37.5	<i>Piptadeniastrum</i>	37.5	Fabaceae	100	0
<i>Platysepalum chevalieri</i>	0	<i>Platysepalum</i>	50	Fabaceae	75	25
<i>Platysepalum violaceum</i>	0	<i>Platysepalum</i>	0	Fabaceae	75	25
<i>Prioria balsamifera</i>	12.5	<i>Prioria</i>	25	Fabaceae	100	0
<i>Prioria buchholzii</i>	0	<i>Prioria</i>	0	Fabaceae	50	50
<i>Prioria mannii</i>	0	<i>Prioria</i>	0	Fabaceae	100	0
<i>Prioria oxyphylla</i>	37.5	<i>Prioria</i>	37.5	Fabaceae	100	0
<i>Pterocarpus angolensis</i>	75	<i>Pterocarpus</i>	100	Fabaceae	100	0
<i>Pterocarpus rotundifolius</i>	50	<i>Pterocarpus</i>	75	Fabaceae	100	0
<i>Pterocarpus soyauxii</i>	93.75	<i>Pterocarpus</i>	93.75	Fabaceae	100	0
<i>Pterocarpus tinctorius</i>	31.25	<i>Pterocarpus</i>	87.5	Fabaceae	100	0
<i>Senegalia senegal</i>	0	<i>Senegalia</i>	25	Fabaceae	100	0
<i>Senna siamea</i>	50	<i>Senna</i>	50	Fabaceae	50	50
<i>Tamarindus indica</i>	12.5	<i>Tamarindus</i>	12.5	Fabaceae	100	0
<i>Tessmannia africana</i>	25	<i>Tessmannia</i>	75	Fabaceae	100	0
<i>Tessmannia anomala</i>	0	<i>Tessmannia</i>	0	Fabaceae	50	50
<i>Tessmannia anomala var. flamignii</i>	0	<i>Tessmannia</i>	0	Fabaceae	25	75
<i>Tessmannia lescrauwaetii</i>	62.5	<i>Tessmannia</i>	62.5	Fabaceae	75	25
<i>Tessmannia yangambiensis</i>	0	<i>Tessmannia</i>	0	Fabaceae	100	0
<i>Tetraberlinia bifoliolata</i>	37.5	<i>Tetraberlinia</i>	37.5	Fabaceae	100	0
<i>Tetrapleura tetraptera</i>	0	<i>Tetrapleura</i>	0	Fabaceae	100	0
<i>Vachellia abyssinica</i>	0	<i>Vachellia</i>	0	Fabaceae	100	0
<i>Vachellia seyal</i>	100	<i>Vachellia</i>	100	Fabaceae	100	0
<i>Anthocleista grandiflora</i>	0	<i>Anthocleista</i>	25	Gentianaceae	25	75
<i>Anthocleista nobilis</i>	37.5	<i>Anthocleista</i>	50	Gentianaceae	50	50
<i>Anthocleista schweinfurthii</i>	25	<i>Anthocleista</i>	25	Gentianaceae	25	75
<i>Hua gaboni</i>	0	<i>Hua</i>	0	Huaceae	0	100
<i>Harungana madagascariensis</i>	0	<i>Harungana</i>	0	Hypericaceae	0	100
<i>Desbordesia glaucescens</i>	25	<i>Desbordesia</i>	25	Irvingiaceae	25	75
<i>Irvingia gabonensis</i>	25	<i>Irvingia</i>	25	Irvingiaceae	25	75

<i>Irvingia grandifolia</i>	0	<i>Irvingia</i>	25	Irvingiaceae	25	75
<i>Irvingia robur</i>	0	<i>Irvingia</i>	25	Irvingiaceae	25	75
<i>Irvingia smithii</i>	0	<i>Irvingia</i>	50	Irvingiaceae	50	50
<i>Irvingia tenuinucleata</i>	0	<i>Irvingia</i>	0	Irvingiaceae	50	50
<i>Klainedoxa gabonensis</i>	75	<i>Klainedoxa</i>	75	Irvingiaceae	75	25
<i>Phyllocosmus africanus</i>	0	<i>Phyllocosmus</i>	0	Ixonanthaceae	0	100
<i>Kirkia acuminata</i>	0	<i>Kirkia</i>	0	Kirkiaceae	0	100
<i>Gmelina arborea</i>	8.333333333	<i>Gmelina</i>	8.333333333	Lamiaceae	8.333333333	91.66666667
<i>Premna angolensis</i>	0	<i>Premna</i>	0	Lamiaceae	50	50
<i>Tectona grandis</i>	0	<i>Tectona</i>	0	Lamiaceae	50	50
<i>Vitex congolensis</i>	0	<i>Vitex</i>	25	Lamiaceae	25	75
<i>Vitex congolensis</i> var. <i>congolensis</i>	37.5	<i>Vitex</i>	50	Lamiaceae	75	25
<i>Vitex doniana</i>	37.5	<i>Vitex</i>	62.5	Lamiaceae	62.5	37.5
<i>Vitex ferruginea</i>	0	<i>Vitex</i>	25	Lamiaceae	25	75
<i>Vitex madiensis</i>	0	<i>Vitex</i>	0	Lamiaceae	33.33333333	66.66666667
<i>Vitex madiensis</i> subsp. <i>milanjensis</i>	0	<i>Vitex</i>	60	Lamiaceae	60	40
<i>Vitex mombassae</i>	0	<i>Vitex</i>	75	Lamiaceae	75	25
<i>Beilschmiedia congolana</i>	0	<i>Beilschmiedia</i>	50	Lauraceae	50	50
<i>Beilschmiedia corbisieri</i>	37.5	<i>Beilschmiedia</i>	37.5	Lauraceae	37.5	62.5
<i>Beilschmiedia louisii</i>	0	<i>Beilschmiedia</i>	0	Lauraceae	0	100
<i>Beilschmiedia mannii</i>	0	<i>Beilschmiedia</i>	50	Lauraceae	50	50
<i>Beilschmiedia oblongifolia</i>	0	<i>Beilschmiedia</i>	0	Lauraceae	0	100
<i>Beilschmiedia ugandensis</i>	0	<i>Beilschmiedia</i>	0	Lauraceae	0	100
<i>Persea americana</i>	0	<i>Persea</i>	25	Lauraceae	25	75
<i>Petersianthus macrocarpus</i>	0	<i>Petersianthus</i>	0	Lecythidaceae	0	100
<i>Brenania brieii</i>	50	<i>Brenania</i>	50	Loganiaceae	50	50
<i>Strychnos cocculoides</i>	100	<i>Strychnos</i>	100	Loganiaceae	100	0
<i>Strychnos innocua</i>	0	<i>Strychnos</i>	0	Loganiaceae	0	100
<i>Strychnos spinosa</i>	75	<i>Strychnos</i>	75	Loganiaceae	100	0
<i>Adansonia digitata</i>	0	<i>Adansonia</i>	0	Malvaceae	100	0
<i>Ceiba pentandra</i>	25	<i>Ceiba</i>	37.5	Malvaceae	75	25
<i>Cola ballayi</i>	75	<i>Cola</i>	100	Malvaceae	100	0
<i>Cola cordifolia</i>	0	<i>Cola</i>	0	Malvaceae	0	100
<i>Cola gigantea</i>	0	<i>Cola</i>	75	Malvaceae	75	25
<i>Cola lateritia</i>	0	<i>Cola</i>	0	Malvaceae	0	100
<i>Cola nitida</i>	100	<i>Cola</i>	100	Malvaceae	100	0
<i>Cola welwitschii</i>	75	<i>Cola</i>	75	Malvaceae	100	0
<i>Desplatsia subericarpa</i>	0	<i>Desplatsia</i>	25	Malvaceae	50	50
<i>Dombeya rotundifolia</i>	25	<i>Dombeya</i>	25	Malvaceae	25	75
<i>Dombeya torrida</i>	0	<i>Dombeya</i>	0	Malvaceae	25	75
<i>Duboscia viridiflora</i>	75	<i>Duboscia</i>	75	Malvaceae	75	25
<i>Grewia louisii</i>	0	<i>Grewia</i>	0	Malvaceae	75	25
<i>Heritiera littoralis</i>	0	<i>Heritiera</i>	0	Malvaceae	25	75

<i>Microcos coriacea</i>	0	<i>Microcos</i>	0	Malvaceae	25	75
<i>Microcos pinnatifida</i>	0	<i>Microcos</i>	0	Malvaceae	37.5	62.5
<i>Nesogordonia kabingaensis</i>	25	<i>Nesogordonia</i>	25	Malvaceae	50	50
<i>Nesogordonia papaverifera</i>	50	<i>Nesogordonia</i>	50	Malvaceae	50	50
<i>Pterygota bequaertii</i>	100	<i>Pterygota</i>	100	Malvaceae	100	0
<i>Pterygota macrocarpa</i>	37.5	<i>Pterygota</i>	37.5	Malvaceae	37.5	62.5
<i>Sterculia quinqueloba</i>	0	<i>Sterculia</i>	0	Malvaceae	0	100
<i>Sterculia tragacantha</i>	25	<i>Sterculia</i>	37.5	Malvaceae	62.5	37.5
<i>Triplochiton scleroxylon</i>	50	<i>Triplochiton</i>	50	Malvaceae	100	0
<i>Dichaetanthera corymbosa</i>	0	<i>Dichaetanthera</i>	0	Melastomataceae	0	100
<i>Carapa procera</i>	12.5	<i>Carapa</i>	37.5	Meliaceae	87.5	12.5
<i>Ekebergia benguelensis</i>	50	<i>Ekebergia</i>	50	Meliaceae	100	0
<i>Ekebergia capensis</i>	0	<i>Ekebergia</i>	0	Meliaceae	37.5	62.5
<i>Entandrophragma angolense</i>	40	<i>Entandrophragma</i>	80	Meliaceae	85	15
<i>Entandrophragma candollei</i>	0	<i>Entandrophragma</i>	37.5	Meliaceae	56.25	43.75
<i>Entandrophragma cylindricum</i>	16.66666667	<i>Entandrophragma</i>	50	Meliaceae	66.66666667	33.33333333
<i>Entandrophragma delevoyi</i>	0	<i>Entandrophragma</i>	25	Meliaceae	25	75
<i>Entandrophragma excelsum</i>	12.5	<i>Entandrophragma</i>	25	Meliaceae	50	50
<i>Entandrophragma palustre</i>	25	<i>Entandrophragma</i>	75	Meliaceae	75	25
<i>Entandrophragma utile</i>	30	<i>Entandrophragma</i>	70	Meliaceae	95	5
<i>Khaya anthotheca</i>	8.333333333	<i>Khaya</i>	8.333333333	Meliaceae	50	50
<i>Khaya grandifoliola</i>	0	<i>Khaya</i>	16.66666667	Meliaceae	41.66666667	58.33333333
<i>Leplaea cedrata</i>	50	<i>Leplaea</i>	55	Meliaceae	70	30
<i>Leplaea laurentii</i>	75	<i>Leplaea</i>	75	Meliaceae	75	25
<i>Leplaea thompsonii</i>	12.5	<i>Leplaea</i>	50	Meliaceae	75	25
<i>Lovoa trichilioides</i>	45.83333333	<i>Lovoa</i>	45.83333333	Meliaceae	66.66666667	33.33333333
<i>Trichilia dregeana</i>	0	<i>Trichilia</i>	0	Meliaceae	75	25
<i>Trichilia emetica</i> subsp. <i>emetica</i>	0	<i>Trichilia</i>	0	Meliaceae	0	100
<i>Trichilia gilgiana</i>	0	<i>Trichilia</i>	50	Meliaceae	50	50
<i>Trichilia gillettii</i>	0	<i>Trichilia</i>	25	Meliaceae	25	75
<i>Trichilia monadelpha</i>	0	<i>Trichilia</i>	0	Meliaceae	0	100
<i>Trichilia prieuriana</i>	0	<i>Trichilia</i>	0	Meliaceae	0	100
<i>Trichilia rubescens</i>	0	<i>Trichilia</i>	0	Meliaceae	0	100
<i>Trichilia tessmannii</i>	0	<i>Trichilia</i>	0	Meliaceae	0	100
<i>Trichilia welwitschii</i>	0	<i>Trichilia</i>	25	Meliaceae	100	0
<i>Turraeanthus africanus</i>	33.33333333	<i>Turraeanthus</i>	33.33333333	Meliaceae	66.66666667	33.33333333
<i>Antiaris toxicaria</i>	37.5	<i>Antiaris</i>	37.5	Moraceae	37.5	62.5
<i>Antiaris toxicaria</i> subsp. <i>welwitschii</i>	0	<i>Antiaris</i>	0	Moraceae	0	100
<i>Antiaris toxicaria</i> var. <i>africana</i>	0	<i>Antiaris</i>	0	Moraceae	0	100
<i>Artocarpus altilis</i>	0	<i>Artocarpus</i>	0	Moraceae	0	100
<i>Bosqueiopsis gillettii</i>	0	<i>Bosqueiopsis</i>	0	Moraceae	0	100
<i>Ficus bubu</i>	0	<i>Ficus</i>	25	Moraceae	25	75
<i>Ficus demeusei</i>	0	<i>Ficus</i>	100	Moraceae	100	0

<i>Ficus elastica</i>	0	<i>Ficus</i>	100	Moraceae	100	0
<i>Ficus lutea</i>	0	<i>Ficus</i>	50	Moraceae	50	50
<i>Ficus mucoso</i>	0	<i>Ficus</i>	75	Moraceae	100	0
<i>Ficus sur</i>	0	<i>Ficus</i>	100	Moraceae	100	0
<i>Ficus sycomorus</i>	0	<i>Ficus</i>	100	Moraceae	100	0
<i>Ficus thonningii</i>	75	<i>Ficus</i>	75	Moraceae	75	25
<i>Ficus vogeliana</i>	0	<i>Ficus</i>	100	Moraceae	100	0
<i>Milicia excelsa</i>	25	<i>Milicia</i>	25	Moraceae	25	75
<i>Treculia africana</i>	0	<i>Treculia</i>	0	Moraceae	0	100
<i>Trilepisium madagascariense</i>	0	<i>Trilepisium</i>	0	Moraceae	0	100
<i>Coelocaryon botryoides</i>	25	<i>Coelocaryon</i>	25	Myristicaceae	25	75
<i>Coelocaryon preussii</i>	0	<i>Coelocaryon</i>	25	Myristicaceae	50	50
<i>Pycnanthus angolensis</i>	100	<i>Pycnanthus</i>	100	Myristicaceae	100	0
<i>Staudtia kamerunensis</i>	0	<i>Staudtia</i>	0	Myristicaceae	0	100
<i>Staudtia kamerunensis</i> var. <i>gabonensis</i>	0	<i>Staudtia</i>	25	Myristicaceae	25	75
<i>Syzygium aromaticum</i>	0	<i>Syzygium</i>	25	Myrtaceae	25	75
<i>Syzygium cordatum</i>	0	<i>Syzygium</i>	62.5	Myrtaceae	62.5	37.5
<i>Syzygium guineense</i>	0	<i>Syzygium</i>	25	Myrtaceae	50	50
<i>Syzygium owariense</i>	0	<i>Syzygium</i>	25	Myrtaceae	25	75
<i>Syzygium parvifolium</i>	0	<i>Syzygium</i>	75	Myrtaceae	75	25
<i>Syzygium staudtii</i>	25	<i>Syzygium</i>	25	Myrtaceae	25	75
<i>Lophira alata</i>	87.5	<i>Lophira</i>	87.5	Ochnaceae	87.5	12.5
<i>Lophira lanceolata</i>	0	<i>Lophira</i>	0	Ochnaceae	0	100
<i>Coula edulis</i>	0	<i>Coula</i>	0	Olacaceae	12.5	87.5
<i>Diogoia zenkeri</i>	0	<i>Diogoia</i>	0	Olacaceae	25	75
<i>Heisteria parvifolia</i>	75	<i>Heisteria</i>	75	Olacaceae	100	0
<i>Okoubaka aubrevillei</i>	75	<i>Okoubaka</i>	75	Olacaceae	100	0
<i>Ongokea gore</i>	50	<i>Ongokea</i>	50	Olacaceae	87.5	12.5
<i>Strombosia grandifolia</i>	25	<i>Strombosia</i>	25	Olacaceae	25	75
<i>Strombosia pustulata</i> var. <i>pustulata</i>	0	<i>Strombosia</i>	50	Olacaceae	50	50
<i>Strombosia scheffleri</i>	0	<i>Strombosia</i>	25	Olacaceae	100	0
<i>Strombosiaopsis tetrandra</i>	12.5	<i>Strombosiaopsis</i>	12.5	Olacaceae	50	50
<i>Ximenia americana</i>	0	<i>Ximenia</i>	0	Olacaceae	0	100
<i>Olea capensis</i> subsp. <i>macrocarpa</i>	75	<i>Olea</i>	75	Oleaceae	75	25
<i>Olea europaea</i> subsp. <i>cuspidata</i>	0	<i>Olea</i>	75	Oleaceae	75	25
<i>Olea welwitschii</i>	0	<i>Olea</i>	25	Oleaceae	25	75
<i>Microdesmis kasaiensis</i>	0	<i>Microdesmis</i>	0	Pandaceae	0	100
<i>Barteria nigritana</i>	50	<i>Barteria</i>	50	Passifloraceae	50	50
<i>Psydrax palma</i>	0	<i>Psydrax</i>	0	Penaeaceae	0	100
<i>Chaetocarpus africanus</i>	50	<i>Chaetocarpus</i>	50	Peraceae	75	25
<i>Antidesma laciniatum</i>	0	<i>Antidesma</i>	25	Phyllanthaceae	50	50
<i>Antidesma membranaceum</i>	0	<i>Antidesma</i>	12.5	Phyllanthaceae	50	50
<i>Bridelia atroviridis</i>	25	<i>Bridelia</i>	75	Phyllanthaceae	75	25

<i>Bridelia brideliifolia</i>	0	<i>Bridelia</i>	0	Phyllanthaceae	75	25
<i>Bridelia ferruginea</i>	25	<i>Bridelia</i>	50	Phyllanthaceae	62.5	37.5
<i>Bridelia micrantha</i>	0	<i>Bridelia</i>	50	Phyllanthaceae	75	25
<i>Cleistanthus caudatus</i>	0	<i>Cleistanthus</i>	0	Phyllanthaceae	0	100
<i>Cleistanthus inundatus</i>	0	<i>Cleistanthus</i>	0	Phyllanthaceae	0	100
<i>Cleistanthus polystachyus</i>	0	<i>Cleistanthus</i>	0	Phyllanthaceae	0	100
<i>Hymenocardia acida</i>	50	<i>Hymenocardia</i>	50	Phyllanthaceae	50	50
<i>Hymenocardia ulmoides</i>	12.5	<i>Hymenocardia</i>	25	Phyllanthaceae	25	75
<i>Maesobotrya staudtii</i>	0	<i>Maesobotrya</i>	0	Phyllanthaceae	100	0
<i>Margaritaria discoidea</i>	0	<i>Margaritaria</i>	0	Phyllanthaceae	0	100
<i>Phyllanthus physocarpus</i>	0	<i>Phyllanthus</i>	0	Phyllanthaceae	25	75
<i>Pseudolachnostylis maprouneifolia</i>	50	<i>Pseudolachnostylis</i>	50	Phyllanthaceae	50	50
<i>Uapaca guineensis</i>	0	<i>Uapaca</i>	37.5	Phyllanthaceae	37.5	62.5
<i>Uapaca heudelotii</i>	12.5	<i>Uapaca</i>	50	Phyllanthaceae	50	50
<i>Uapaca kirkiana</i>	0	<i>Uapaca</i>	50	Phyllanthaceae	50	50
<i>Uapaca mole</i>	0	<i>Uapaca</i>	50	Phyllanthaceae	62.5	37.5
<i>Uapaca nitida</i>	25	<i>Uapaca</i>	75	Phyllanthaceae	75	25
<i>Uapaca pilosa</i>	0	<i>Uapaca</i>	75	Phyllanthaceae	100	0
<i>Uapaca robynsii</i>	0	<i>Uapaca</i>	75	Phyllanthaceae	75	25
<i>Uapaca sansibarica</i>	50	<i>Uapaca</i>	75	Phyllanthaceae	75	25
<i>Uapaca togoensis</i>	0	<i>Uapaca</i>	100	Phyllanthaceae	100	0
<i>Uapaca vanhouttei</i>	0	<i>Uapaca</i>	50	Phyllanthaceae	50	50
<i>Faurea rochetiana</i>	0	<i>Faurea</i>	25	Proteaceae	25	75
<i>Faurea saligna</i>	0	<i>Faurea</i>	0	Proteaceae	0	100
<i>Drypetes angustifolia</i>	100	<i>Drypetes</i>	100	Putranjivaceae	100	0
<i>Drypetes gerrardii</i>	0	<i>Drypetes</i>	0	Putranjivaceae	0	100
<i>Drypetes gosswelleri</i>	0	<i>Drypetes</i>	50	Putranjivaceae	50	50
<i>Maesopsis eminii</i>	0	<i>Maesopsis</i>	0	Rhamnaceae	0	100
<i>Ziziphus abyssinica</i>	0	<i>Ziziphus</i>	0	Rhamnaceae	0	100
<i>Anopyxis klaineana</i>	25	<i>Anopyxis</i>	25	Rhizophoraceae	37.5	62.5
<i>Cassipourea congoensis</i>	75	<i>Cassipourea</i>	75	Rhizophoraceae	75	25
<i>Cassipourea gummiflua</i>	0	<i>Cassipourea</i>	0	Rhizophoraceae	0	100
<i>Cassipourea malosana</i>	0	<i>Cassipourea</i>	0	Rhizophoraceae	0	100
<i>Rhizophora racemosa</i>	12.5	<i>Rhizophora</i>	12.5	Rhizophoraceae	12.5	87.5
<i>Hagenia abyssinica</i>	50	<i>Hagenia</i>	50	Rosaceae	50	50
<i>Prunus africana</i>	0	<i>Prunus</i>	0	Rosaceae	0	100
<i>Aidia ochroleuca</i>	0	<i>Aidia</i>	0	Rubiaceae	25	75
<i>Aulacocalyx jasminiflora</i>	0	<i>Aulacocalyx</i>	0	Rubiaceae	25	75
<i>Coffea liberica</i>	0	<i>Coffea</i>	0	Rubiaceae	75	25
<i>Corynanthe macroceras</i>	50	<i>Corynanthe</i>	50	Rubiaceae	75	25
<i>Corynanthe paniculata</i>	0	<i>Corynanthe</i>	12.5	Rubiaceae	50	50
<i>Craterispermum cerinanthum</i>	0	<i>Craterispermum</i>	0	Rubiaceae	50	50
<i>Craterispermum laurinum</i>	25	<i>Craterispermum</i>	25	Rubiaceae	75	25

<i>Crossopteryx febrifuga</i>	75	<i>Crossopteryx</i>	75	Rubiaceae	100	0
<i>Gardenia imperialis</i>	50	<i>Gardenia</i>	50	Rubiaceae	100	0
<i>Gardenia ternifolia</i> subsp. <i>jovis-tonantis</i>	0	<i>Gardenia</i>	0	Rubiaceae	75	25
<i>Heinsia crinita</i>	100	<i>Heinsia</i>	100	Rubiaceae	100	0
<i>Massularia acuminata</i>	50	<i>Massularia</i>	50	Rubiaceae	75	25
<i>Mitragyna ledermannii</i>	0	<i>Mitragyna</i>	25	Rubiaceae	50	50
<i>Mitragyna rubrostipulata</i>	0	<i>Mitragyna</i>	0	Rubiaceae	25	75
<i>Mitragyna stipulosa</i>	75	<i>Mitragyna</i>	75	Rubiaceae	83.33333333	16.66666667
<i>Morinda chrysorrhiza</i>	0	<i>Morinda</i>	0	Rubiaceae	0	100
<i>Morinda citrifolia</i>	0	<i>Morinda</i>	25	Rubiaceae	25	75
<i>Morinda lucida</i>	100	<i>Morinda</i>	100	Rubiaceae	100	0
<i>Nauclea diderrichii</i>	62.5	<i>Nauclea</i>	62.5	Rubiaceae	62.5	37.5
<i>Nauclea latifolia</i>	0	<i>Nauclea</i>	25	Rubiaceae	50	50
<i>Nauclea pobeguinii</i>	25	<i>Nauclea</i>	25	Rubiaceae	100	0
<i>Oxyanthus tubiflorus</i>	0	<i>Oxyanthus</i>	0	Rubiaceae	75	25
<i>Psychotria dermatophylla</i>	0	<i>Psychotria</i>	0	Rubiaceae	0	100
<i>Rothmannia longiflora</i>	0	<i>Rothmannia</i>	0	Rubiaceae	75	25
<i>Rothmannia lujae</i>	0	<i>Rothmannia</i>	0	Rubiaceae	25	75
<i>Schumanniphyton magnificum</i>	100	<i>Schumanniphyton</i>	100	Rubiaceae	100	0
<i>Tricalysia pallens</i>	75	<i>Tricalysia</i>	75	Rubiaceae	100	0
<i>Citrus x aurantium</i>	25	<i>Citrus</i>	50	Rutaceae	50	50
<i>Citrus x limon</i>	75	<i>Citrus</i>	75	Rutaceae	75	25
<i>Vepris louisii</i>	0	<i>Vepris</i>	0	Rutaceae	0	100
<i>Zanthoxylum chalybeum</i>	0	<i>Zanthoxylum</i>	0	Rutaceae	0	100
<i>Zanthoxylum gillettii</i>	0	<i>Zanthoxylum</i>	25	Rutaceae	25	75
<i>Zanthoxylum heitzii</i>	0	<i>Zanthoxylum</i>	0	Rutaceae	0	100
<i>Zanthoxylum lemairei</i>	0	<i>Zanthoxylum</i>	25	Rutaceae	25	75
<i>Homalium abdessammadii</i>	0	<i>Homalium</i>	25	Salicaceae	25	75
<i>Homalium africanum</i>	0	<i>Homalium</i>	25	Salicaceae	25	75
<i>Homalium letestui</i>	0	<i>Homalium</i>	50	Salicaceae	50	50
<i>Homalium longistylum</i>	0	<i>Homalium</i>	75	Salicaceae	75	25
<i>Homalium stipulaceum</i>	0	<i>Homalium</i>	50	Salicaceae	50	50
<i>Allophylus africanus</i>	0	<i>Allophylus</i>	0	Sapindaceae	0	100
<i>Allophylus dummeri</i>	0	<i>Allophylus</i>	0	Sapindaceae	0	100
<i>Blighia welwitschii</i>	0	<i>Blighia</i>	0	Sapindaceae	0	100
<i>Chytranthus carneus</i>	50	<i>Chytranthus</i>	50	Sapindaceae	50	50
<i>Chytranthus setosus</i>	0	<i>Chytranthus</i>	0	Sapindaceae	0	100
<i>Haplocoelum intermedium</i>	0	<i>Haplocoelum</i>	0	Sapindaceae	0	100
<i>Lecaniodiscus cupanioides</i>	0	<i>Lecaniodiscus</i>	0	Sapindaceae	0	100
<i>Majidea fosteri</i>	0	<i>Majidea</i>	0	Sapindaceae	0	100
<i>Pancovia floribunda</i>	100	<i>Pancovia</i>	100	Sapindaceae	100	0
<i>Pancovia laurentii</i>	75	<i>Pancovia</i>	75	Sapindaceae	75	25
<i>Sapindus rarak</i>	0	<i>Sapindus</i>	0	Sapindaceae	0	100

<i>Zanha golungensis</i>	12.5	<i>Zanha</i>	25	Sapindaceae	25	75
<i>Aningeria adolfi-friederici</i>	0	<i>Aningeria</i>	0	Sapotaceae	25	75
<i>Aningeria altissima</i>	0	<i>Aningeria</i>	0	Sapotaceae	75	25
<i>Aningeria pierrei</i>	0	<i>Aningeria</i>	0	Sapotaceae	75	25
<i>Autranella congolensis</i>	91.66666667	<i>Autranella</i>	91.66666667	Sapotaceae	91.66666667	8.333333333
<i>Baillonella toxisperma</i>	50	<i>Baillonella</i>	50	Sapotaceae	100	0
<i>Brevia sericea</i>	0	<i>Brevia</i>	0	Sapotaceae	75	25
<i>Donella pruniformis</i>	0	<i>Donella</i>	0	Sapotaceae	75	25
<i>Englerophytum laurentii</i>	0	<i>Englerophytum</i>	0	Sapotaceae	100	0
<i>Englerophytum magalismontanum</i>	0	<i>Englerophytum</i>	0	Sapotaceae	0	100
<i>Gambeya africana</i>	0	<i>Gambeya</i>	37.5	Sapotaceae	100	0
<i>Gambeya albida</i>	0	<i>Gambeya</i>	25	Sapotaceae	50	50
<i>Gambeya beguei</i>	0	<i>Gambeya</i>	50	Sapotaceae	75	25
<i>Gambeya lacourtiana</i>	0	<i>Gambeya</i>	75	Sapotaceae	75	25
<i>Gambeya lungi</i>	0	<i>Gambeya</i>	50	Sapotaceae	75	25
<i>Gambeya perpulchra</i>	0	<i>Gambeya</i>	37.5	Sapotaceae	50	50
<i>Gambeya subnuda</i>	0	<i>Gambeya</i>	25	Sapotaceae	50	50
<i>Letestua durissima</i>	0	<i>Letestua</i>	0	Sapotaceae	100	0
<i>Malacantha alnifolia</i>	0	<i>Malacantha</i>	0	Sapotaceae	100	0
<i>Manilkara mochisia</i>	0	<i>Manilkara</i>	0	Sapotaceae	75	25
<i>Manilkara obovata</i>	37.5	<i>Manilkara</i>	37.5	Sapotaceae	50	50
<i>Mimusops zeyheri</i>	0	<i>Mimusops</i>	0	Sapotaceae	75	25
<i>Neolemonniera clitandrifolia</i>	0	<i>Neolemonniera</i>	0	Sapotaceae	75	25
<i>Omphalocarpum briei</i>	0	<i>Omphalocarpum</i>	0	Sapotaceae	25	75
<i>Omphalocarpum lecomteanum</i>	0	<i>Omphalocarpum</i>	0	Sapotaceae	0	100
<i>Synsepalum afzelii</i>	0	<i>Synsepalum</i>	0	Sapotaceae	50	50
<i>Synsepalum brevipes</i>	0	<i>Synsepalum</i>	75	Sapotaceae	100	0
<i>Synsepalum revolutum</i>	0	<i>Synsepalum</i>	25	Sapotaceae	100	0
<i>Synsepalum stipulatum</i>	0	<i>Synsepalum</i>	0	Sapotaceae	75	25
<i>Synsepalum subcordatum</i>	75	<i>Synsepalum</i>	75	Sapotaceae	75	25
<i>Tieghemella africana</i>	0	<i>Tieghemella</i>	37.5	Sapotaceae	75	25
<i>Tieghemella heckelii</i>	37.5	<i>Tieghemella</i>	50	Sapotaceae	62.5	37.5
<i>Tridesmostemon omphalocarpoides</i>	0	<i>Tridesmostemon</i>	0	Sapotaceae	75	25
<i>Vitellaria paradoxa</i>	25	<i>Vitellaria</i>	50	Sapotaceae	100	0
<i>Odyndea gabunensis</i>	0	<i>Odyndea</i>	0	Simaroubaceae	0	100
<i>Pierreodendron africanum</i>	87.5	<i>Pierreodendron</i>	87.5	Simaroubaceae	87.5	12.5
<i>Ficalhoa laurifolia</i>	0	<i>Ficalhoa</i>	0	Sladeniaceae	0	100
<i>Nuxia congesta</i>	0	<i>Nuxia</i>	0	Stilbaceae	0	100
<i>Holoptelea grandis</i>	12.5	<i>Holoptelea</i>	12.5	Ulmaceae	12.5	87.5
<i>Musanga cecropioides</i>	25	<i>Musanga</i>	25	Urticaceae	25	75
<i>Myrianthus arboreus</i>	25	<i>Myrianthus</i>	25	Urticaceae	25	75
<i>Myrianthus holstii</i>	50	<i>Myrianthus</i>	50	Urticaceae	50	50
<i>Myrianthus preussii</i>	0	<i>Myrianthus</i>	0	Urticaceae	0	100

<i>Rinorea welwitschii</i>	100	<i>Rinorea</i>	100	Violaceae	100	0
<i>Erismadelphus exsul</i>	0	<i>Erismadelphus</i>	0	Vochysiaceae	0	100
<i>Balanites wilsoniana</i>	50	<i>Balanites</i>	50	Zygophyllaceae	50	50

Table 8.4: Summary of botanical genera and species used for training and testing the CNN on timber genus classification. The first column gives the botanical genus, the second column presents the full taxon of all species included in this study, the final column shows the number of specimens of each species.

Botanical Genus	Full taxon	Number of specimens
<i>Afzelia</i> Sm.	<i>Afzelia africana</i> Sm. ex Pers.	8
	<i>Afzelia bella</i> Harms	4
	<i>Afzelia bipindensis</i> Harms	11
	<i>Afzelia pachyloba</i> Harms	10
	<i>Afzelia peturei</i> De Wild.	1
	<i>Afzelia quanzensis</i> Welw.	7
<i>Albizia</i> Durazz.	<i>Albizia</i> spp. Durazz.	3
	<i>Albizia adianthifolia</i> (Schumach.) W.Wight	6
	<i>Albizia adianthifolia</i> var. <i>intermedia</i> (De Wild. & T.Durand) Villiers	2
	<i>Albizia altissima</i> Hook.f.	5
	<i>Albizia antunesiana</i> Harms	5
	<i>Albizia chinensis</i> (Osbeck) Merr.	6
	<i>Albizia coriaria</i> Welw. ex Oliv.	5
	<i>Albizia ferruginea</i> (Guill. & Perr.) Benth.	6
	<i>Albizia glaberrima</i> (Schumach. & Thonn.) Benth.	3

	<i>Albizia glaberrima</i> var. <i>glabrescens</i> (Oliv.) Brenan	2
	<i>Albizia grandibracteata</i> Taub.	2
	<i>Albizia gummifera</i> (J.F.Gmel.) C.A.Sm.	4
	<i>Albizia gummifera</i> var. <i>ealaensis</i> (De Wild.) Brenan	6
	<i>Albizia laurentii</i> De Wild.	4
	<i>Albizia lebbeck</i> (L.) Benth.	2
	<i>Albizia schimperiana</i> Oliv.	2
	<i>Albizia versicolor</i> Welw. ex Oliv.	5
	<i>Albizia zygia</i> (DC.) J.F.Macbr.	4
<i>Alstonia</i> R.Br.	<i>Alstonia boonei</i> De Wild.	5
	<i>Alstonia congensis</i> Engl.	4
<i>Beilschmiedia</i> Nees	<i>Beilschmiedia</i> spp. Nees	4
	<i>Beilschmiedia alata</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia congolana</i> Robyns & R.Wilczek	4
	<i>Beilschmiedia corbisieri</i> (Robyns) Robyns & R.Wilczek	5
	<i>Beilschmiedia donisii</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia gaboonensis</i> (Meisn.) Benth. & Hook.fil.	1
	<i>Beilschmiedia gilbertii</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia louisii</i> Robyns & R.Wilczek	2
	<i>Beilschmiedia mannii</i> (Meisn.) Robyns & R.Wilczek	5
	<i>Beilschmiedia mannoides</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia mayumbensis</i> Robyns & R.Wilczek	1

	<i>Beilschmiedia michelsonii</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia oblongifolia</i> Robyns & R.Wilczek	4
	<i>Beilschmiedia schmitzii</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia ugandensis</i> Rendle	2
	<i>Beilschmiedia variabilis</i> Robyns & R.Wilczek	1
	<i>Beilschmiedia zenkeri</i> Engl.	1
<i>Brachystegia</i> Benth.	<i>Brachystegia</i> spp. Benth.	3
	<i>Brachystegia angustistipulata</i> De Wild.	1
	<i>Brachystegia boehmii</i> Taub.	5
	<i>Brachystegia bussei</i> Harms	3
	<i>Brachystegia floribunda</i> Benth.	3
	<i>Brachystegia gossweileri</i> Hutch. & Burt Davy	1
	<i>Brachystegia laurentii</i> (De Wild.) Louis ex J.Léonard	6
	<i>Brachystegia longifolia</i> Benth.	5
	<i>Brachystegia manga</i> De Wild.	5
	<i>Brachystegia spiciformis</i> Benth.	6
	<i>Brachystegia stipulata</i> De Wild.	1
	<i>Brachystegia tamarindoides</i> subsp. <i>microphylla</i> (Harms) Chikuni	3
	<i>Brachystegia tamarindoides</i> subsp. <i>tamarindoides</i>	4
	<i>Brachystegia taxifolia</i> Harms	5
	<i>Brachystegia utilis</i> Hutch. & Burt Davy	3
<i>Celtis</i> L.	<i>Celtis</i> spp. L.	1

	<i>Celtis adolfi-friderici</i> Engl.	4
	<i>Celtis africana</i> Burm.fil.	3
	<i>Celtis gomphophylla</i> Baker	8
	<i>Celtis latifolia</i> (Blume) Planch.	3
	<i>Celtis mildbraedii</i> Engl.	5
	<i>Celtis philippensis</i> Blanco	5
	<i>Celtis tessmannii</i> Rendle	3
	<i>Celtis zenkeri</i> Engl.	4
<i>Cynometra</i> L.	<i>Cynometra alexandri</i> C.H.Wright	4
	<i>Cynometra hankei</i> Harms	5
	<i>Cynometra lujae</i> De Wild.	3
	<i>Cynometra mannii</i> Oliv.	3
	<i>Cynometra michelsonii</i> J.Léonard	1
	<i>Cynometra schlechteri</i> Harms	1
	<i>Cynometra sessiliflora</i> Harms	5
<i>Dialium</i> L.	<i>Dialium</i> spp. L.	3
	<i>Dialium angolense</i> Welw. ex Oliv.	1
	<i>Dialium englerianum</i> Henriq.	4
	<i>Dialium excelsum</i> Louis ex Steyaert	5
	<i>Dialium guineense</i> Willd.	1
	<i>Dialium kasaiense</i> Louis ex Steyaert	2
	<i>Dialium pachyphyllum</i> Harms	8

	<i>Dialium pachyphyllum</i> Harms	8
	<i>Dialium pentandrum</i> Louis ex Steyaert	5
	<i>Dialium polyanthum</i> Harms	4
	<i>Dialium tessmannii</i> Harms	3
	<i>Dialium zenkeri</i> Harms	4
<i>Diospyros</i> L.	<i>Diospyros</i> spp. L.	5
	<i>Diospyros alboflavescens</i> (Gürke) F.White	1
	<i>Diospyros batocana</i> Hiern	4
	<i>Diospyros bipindensis</i> Gürke	3
	<i>Diospyros boala</i> De Wild.	2
	<i>Diospyros canaliculata</i> De Wild.	3
	<i>Diospyros chrysocarpa</i> F.White	2
	<i>Diospyros conocarpa</i> Gürke	1
	<i>Diospyros crassiflora</i> Hiern	6
	<i>Diospyros dendo</i> Welw. ex Hiern	3
	<i>Diospyros ferrea</i> (Willd.) Bakh.	5
	<i>Diospyros gillettii</i> De Wild.	1
	<i>Diospyros grex</i> F.White	1
	<i>Diospyros heterotricha</i> (Welw. ex Hiern) F.White	3
	<i>Diospyros hoyleana</i> F.White	3
	<i>Diospyros iturensis</i> (Gürke) Letouzey & F.White	4
	<i>Diospyros mespiliformis</i> Hochst. ex A.DC.	3

	<i>Diospyros monbuttensis</i> Gürke	1
	<i>Diospyros physocalycina</i> Gürke	1
	<i>Diospyros piscatoria</i> Gürke	1
	<i>Diospyros polystemon</i> Gürke	1
	<i>Diospyros pseudomespilus</i> Mildbr.	4
	<i>Diospyros pseudomespilus subsp. undabunda</i> (Hiern ex Greves) F.White	1
	<i>Diospyros vermoesenii</i> De Wild.	1
	<i>Diospyros viridicans</i> Hiern	2
	<i>Diospyros zenkeri</i> (Gürke) F.White	2
<i>Entandrophragma</i> C.DC.	<i>Entandrophragma</i> spp. C.DC.	6
	<i>Entandrophragma angolense</i> (Welw.) C.DC.	22
	<i>Entandrophragma candollei</i> Harms	17
	<i>Entandrophragma cylindricum</i> (Sprague) Sprague	14
	<i>Entandrophragma delevoyi</i> De Wild.	4
	<i>Entandrophragma excelsum</i> (Dawe & Sprague) Sprague	6
	<i>Entandrophragma palustre</i> Staner	5
	<i>Entandrophragma utile</i> (Dawe & Sprague) Sprague	19
<i>Ficus</i> Tourn. ex L.	<i>Ficus</i> spp. Tourn. ex L.	1
	<i>Ficus ampelos</i> Burm.fil.	1
	<i>Ficus ardisioides</i> Warb.	1
	<i>Ficus bubu</i> Warb.	5
	<i>Ficus craterostoma</i> Warb. ex Mildbr. & Burret	1

	<i>Ficus demeusei</i> Warb.	2
	<i>Ficus dicranostyla</i> Mildbr.	2
	<i>Ficus elastica</i> Roxb.	2
	<i>Ficus glumosa</i> Delile	2
	<i>Ficus ingens</i> (Miq.) Miq.	1
	<i>Ficus lutea</i> Vahl	4
	<i>Ficus mucoso</i> Welw. ex Ficalho	4
	<i>Ficus recurvata</i> De Wild.	1
	<i>Ficus sarmentosa</i> var. <i>sarmentosa</i>	1
	<i>Ficus sur</i> Forssk.	4
	<i>Ficus sycomorus</i> L.	4
	<i>Ficus thonningii</i> Blume	4
	<i>Ficus trichopoda</i> Baker	1
	<i>Ficus vallis-choudae</i> Delile	2
	<i>Ficus variifolia</i> Warb.	3
	<i>Ficus vogeliana</i> (Miq.) Miq.	4
	<i>Ficus wildemaniana</i> Warb. ex De Wild. & T.Durand	2
Gambeya Pierre	<i>Gambeya africana</i> (A.DC.) Pierre	9
	<i>Gambeya albida</i> (G.Don) Aubrév. & Pellegr.	3
	<i>Gambeya beguei</i> (Aubrév. & Pellegr.) Aubrév. & Pellegr.	2
	<i>Gambeya gorungosana</i> (Engl.) Liben	2
	<i>Gambeya lacourtiana</i> (De Wild.) Aubrév. & Pellegr.	4

	<i>Gambeya lungi</i> (De Wild.) Aubrév. & Pellegr.	3
	<i>Gambeya perpulchra</i> (Mildbr. ex Hutch. & Dalziel) Aubrév. & Pellegr.	5
	<i>Gambeya subnuda</i> (Baker) Pierre	5
<i>Gilbertiodendron</i> J.Léonard	<i>Gilbertiodendron</i> J.Léonard	1
	<i>Gilbertiodendron dewevrei</i> (De Wild.) J.Léonard	5
	<i>Gilbertiodendron grandiflorum</i> (De Wild.) J.Léonard	3
	<i>Gilbertiodendron grandistipulatum</i> (De Wild.) J.Léonard	6
	<i>Gilbertiodendron mayombense</i> (Pellegr.) J.Léonard	3
	<i>Gilbertiodendron ogoouense</i> (Pellegr.) J.Léonard	5
<i>Homalium</i> Jacq.	<i>Homalium</i> spp. Jacq.	1
	<i>Homalium abdessammadii</i> Asch. & Schweinf.	3
	<i>Homalium africanum</i> (Hook.f.) Benth.	4
	<i>Homalium letestui</i> Pellegr.	5
	<i>Homalium longistylum</i> Mast.	5
	<i>Homalium stipulaceum</i> Welw. ex Mast.	4
<i>Irvingia</i> Hook.f.	<i>Irvingia gabonensis</i> (Aubry-Lecomte ex O'Rorke) Baill.	5
	<i>Irvingia grandifolia</i> (Engl.) Engl.	4
	<i>Irvingia robur</i> Mildbr.	6
	<i>Irvingia smithii</i> Hook.fil.	2
	<i>Irvingia tenuinucleata</i> Tiegh.	3
<i>Leplaea</i> Vermeesen	<i>Leplaea cedrata</i> (A.Chev.) E.J.M.Koenen & J.J.de Wilde	20
	<i>Leplaea laurentii</i> (De Wild.) E.J.M.Koenen & J.J.de Wilde	4

	<i>Leplaea mayombensis</i> (Pellegr.) Staner	1
	<i>Leplaea thompsonii</i> (Sprague & Hutch.) E.J.M.Koenen & J.J.de Wilde	17
<i>Microcos</i> L.	<i>Microcos coriacea</i> (Mast.) Burret	2
	<i>Microcos mildbraedii</i> Burret	1
	<i>Microcos oligoneura</i> Burret	1
	<i>Microcos pinnatifida</i> (Mast.) Burret	5
<i>Milicia</i> Sim	<i>Milicia excelsa</i> (Welw.) C.C.Berg	13
<i>Millettia</i> Wight & Arn.	<i>Millettia</i> Wight & Arn.	2
	<i>Millettia drastica</i> Welw. ex Baker	3
	<i>Millettia dura</i> Dunn	3
	<i>Millettia eetveldeana</i> (Micheli) Hauman	3
	<i>Millettia hockii</i> De Wild.	1
	<i>Millettia hylobia</i> Louis ex Hauman	2
	<i>Millettia laurentii</i> De Wild.	7
	<i>Millettia limbutuensis</i> De Wild.	1
	<i>Millettia macroura</i> Harms	1
	<i>Millettia stuhlmannii</i> Taub.	3
	<i>Millettia versicolor</i> Welw. ex Baker	4
<i>Nauclea</i> L.	<i>Nauclea</i> spp. L.	3
	<i>Nauclea diderrichii</i> (De Wild.) Merr.	10
	<i>Nauclea latifolia</i> Sm.	3
	<i>Nauclea pobeguinii</i> (Hua ex Pobég.) Merr.	5

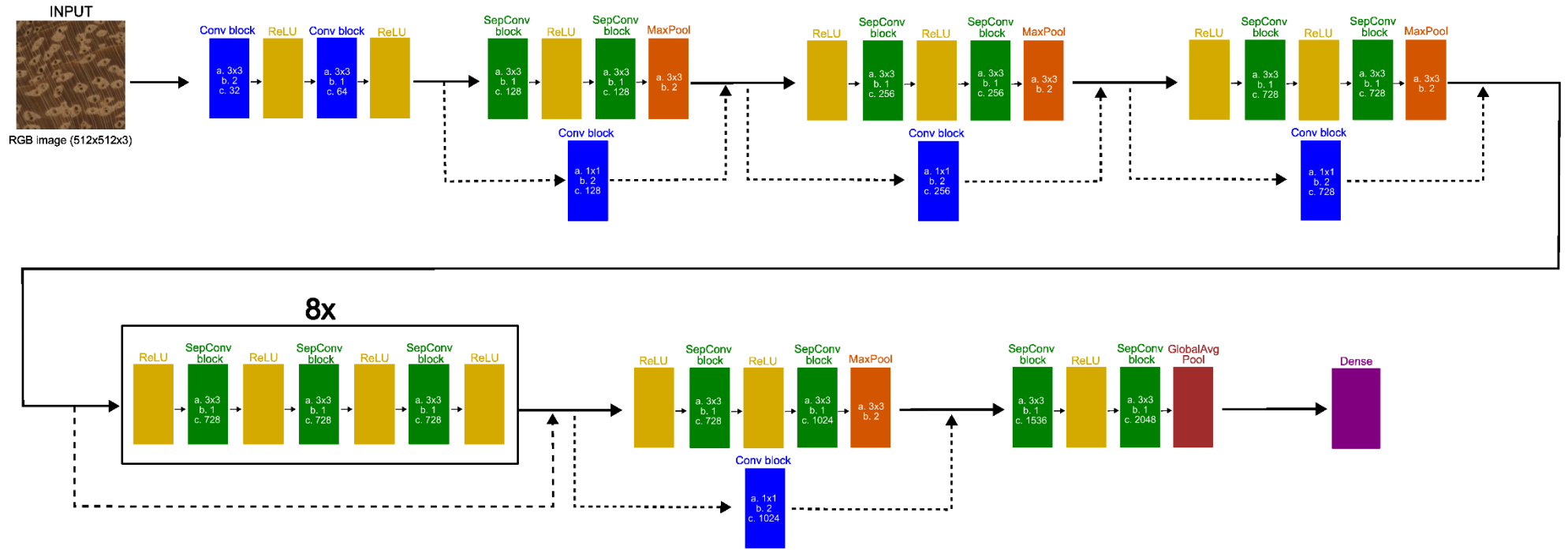
	<i>Nauclea vanderguchtii</i> (De Wild.) E.M.A.Petit	1
<i>Prioria</i> Griseb.	<i>Prioria</i> spp. Griseb.	3
	<i>Prioria balsamifera</i> (Vermoesen) Breteler	7
	<i>Prioria buchholzii</i> (Harms) Breteler	3
	<i>Prioria gilbertii</i> (J.Léonard) Breteler	1
	<i>Prioria mannii</i> (Baill.) Breteler	2
	<i>Prioria oxyphylla</i> (Harms) Breteler	5
<i>Pterocarpus</i> Jacq.	<i>Pterocarpus</i> spp. Jacq.	7
	<i>Pterocarpus angolensis</i> DC.	17
	<i>Pterocarpus gillettii</i> De Wild.	1
	<i>Pterocarpus lucens</i> Lepr. ex Guill. & Perr.	1
	<i>Pterocarpus rotundifolius</i> (Sond.) Druce	3
	<i>Pterocarpus soyauxii</i> Taub.	16
	<i>Pterocarpus tessmannii</i> Harms	1
	<i>Pterocarpus tinctorius</i> Welw.	18
<i>Tessmannia</i> Harms	<i>Tessmannia africana</i> Harms	2
	<i>Tessmannia anomala</i> (Micheli) Harms	3
	<i>Tessmannia anomala</i> var. <i>flamignii</i> J.Léonard	2
	<i>Tessmannia copallifera</i> J.Léonard	2
	<i>Tessmannia dewildemaniana</i> Harms	1
	<i>Tessmannia lescrauwaetii</i> (De Wild.) Harms	5
	<i>Tessmannia yangambiensis</i> Louis ex J.Léonard	4

<i>Uapaca</i> Baill.	<i>Uapaca</i> spp. Baill.	6
	<i>Uapaca guineensis</i> Müll.Arg.	5
	<i>Uapaca heudelotii</i> Baill.	5
	<i>Uapaca kirkiana</i> Müll.Arg.	3
	<i>Uapaca kirkiana</i> var. <i>gossweileri</i> (Hutch.) Meerts	1
	<i>Uapaca mole</i> Pax	6
	<i>Uapaca nitida</i> Müll.Arg.	4
	<i>Uapaca pilosa</i> Hutch.	3
	<i>Uapaca pynaertii</i> De Wild.	1
	<i>Uapaca robynsii</i> De Wild.	2
	<i>Uapaca sansibarica</i> Pax	4
	<i>Uapaca togoensis</i> Pax	2
	<i>Uapaca vanhouttei</i> De Wild.	2
<i>Vitex</i> L.	<i>Vitex</i> spp. L.	2
	<i>Vitex congolensis</i> De Wild. & T.Durand	10
	<i>Vitex cuspidata</i> Hiern	1
	<i>Vitex doniana</i> Sweet	5
	<i>Vitex ferruginea</i> Schumach. & Thonn.	7
	<i>Vitex fischeri</i> Gürke	1
	<i>Vitex grandifolia</i> Gürke	1
	<i>Vitex madiensis</i> Oliv.	6
	<i>Vitex madiensis</i> subsp. <i>milanjiensis</i> (Britten) F.White	2

	<i>Vitex micrantha</i> Gürke	2
	<i>Vitex mombassae</i> Vatke	2
	<i>Vitex rubroaurantiaca</i> De Wild.	1
<i>Xylopia</i> L.	<i>Xylopia</i> spp. L.	
	<i>Xylopia acutiflora</i> (Dunal) A.Rich.	4
	<i>Xylopia aethiopica</i> (Dunal) A.Rich.	3
	<i>Xylopia aurantiiodora</i> De Wild. & T.Durand	2
	<i>Xylopia cupularis</i> Mildbr.	6
	<i>Xylopia flamignii</i> Boutique	4
	<i>Xylopia gilbertii</i> Boutique	4
	<i>Xylopia hypolampra</i> Mildbr. & Diels	5
	<i>Xylopia katangensis</i> De Wild.	3
	<i>Xylopia longipetala</i> De Wild. & T.Durand	1
	<i>Xylopia odoratissima</i> Welw. ex Oliv.	1
	<i>Xylopia phloiodora</i> Mildbr.	5
	<i>Xylopia rubescens</i> Oliv.	5
	<i>Xylopia staudtii</i> Engl. & Diels	4
	<i>Xylopia toussaintii</i> Boutique	2
	<i>Xylopia villosa</i> Chipp	3
	<i>Xylopia wilwerthii</i> De Wild.	5

Table 8.5: The table shows quantified characteristics of the dataset for individual classes after division specimens into training and testing. Quantified characteristics are portrayed by gradient colour scales, portraying positive deviation of the average value.

Botanical Genus	Total number of specimens in Training data	Total number of image patches in Training data	Average number of training patches across specimens per class (genus)	Total number of specimens in Test data	Total number of image patches in Test data	Average number of test patches across specimens per class (genus)
<i>Entandrophragma</i> C.DC.	65	5698	88	27	2283	85
<i>Pterocarpus</i> Jacq.	49	2305	47	13	341	26
<i>Albizia</i> Durazz.	45	3695	82	26	1085	42
<i>Diospyros</i> L.	31	1564	50	32	1121	35
<i>Uapaca</i> Baill.	31	1659	54	15	969	65
<i>Xylopia</i> L.	31	3848	124	29	2235	77
<i>Brachystegia</i> Benth.	30	3976	133	23	1240	54
<i>Azela</i> Sm.	27	1869	69	14	716	51
<i>Dialium</i> L.	26	2529	97	14	508	36
<i>Leplaea</i> Vermoesen	26	1846	71	16	812	51
<i>Beilschmiedia</i> Nees	24	1120	47	10	422	42
<i>Gambeya</i> Pierre	24	2912	121	9	684	76
<i>Vitex</i> L.	23	1724	75	29	2235	77
<i>Ficus</i> Tourn. ex L.	21	1973	94	31	1504	49
<i>Millettia</i> Wight & Arn.	21	1653	79	6	404	67
<i>Celtis</i> L.	20	1021	51	15	613	41
<i>Cynometra</i> L.	18	1188	66	4	365	91
<i>Gilbertiodendron</i> J. Léonard	18	1961	109	4	668	167
<i>Nauclea</i> L.	16	1309	82	5	396	79
<i>Prioria</i> Griseb.	16	1994	125	14	964	69
<i>Homalium</i> Jacq.	14	2083	149	8	189	24
<i>Irvingia</i> Hook.f.	13	2792	215	7	348	50
<i>Milicia</i> Sim	11	1265	115	9	614	68
<i>Tessmannia</i> Harms	5	698	140	13	563	43
<i>Alstonia</i> R.Br.	3	875	292	6	333	56
<i>Microcos</i> Burm. ex L.	3	1577	526	6	432	72



Legend

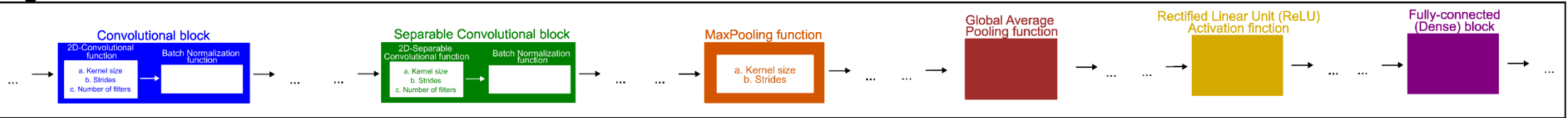


Figure 8.3: The Xception architecture, implemented in this study. Layers are sequenced from top-left to bottom-right. The legend underneath the architecture highlights the detailed structure of each layer (or "block") and the structure of the parameters (if applicable). The parameters (numerical values) for the layers in the figure are structured according to the listed parameters in the legend. Dashed lines represent residual (or skip) connections, which enables a bypass across layers and was implemented to prevent the vanishing gradient problem. The "8x" above the boxed section represents that the boxed sequence is repeated 8x sequentially.

8.2 Definitions of classification performance metrics: Accuracy, Precision, Recall, and F1-score

In classification problems, the performance of a model is commonly evaluated using a confusion matrix, which summarizes predictions relative to ground-truth labels. For a binary classification problem, the outcomes are defined as:

True Positives (TP): correctly predicted positive samples

True Negatives (TN): correctly predicted negative samples

False Positives (FP): negative samples incorrectly predicted as positive

False Negatives (FN): positive samples incorrectly predicted as negative

8.2.1 Accuracy

Accuracy measures the overall proportion of correctly classified instances among all predictions:

$$Accuracy = \frac{TP + TN + FP + FN}{TP + TN}$$

While accuracy is intuitive, it can be misleading in imbalanced datasets (e.g., many more negative than positive cases), as it does not distinguish between error types.

8.2.2 Precision

Precision (also called positive predictive value) quantifies the reliability of positive predictions by measuring the proportion of predicted positives that are actually correct:

$$Precision = \frac{TP + FP}{TP}$$

High precision indicates that when the model predicts a positive class, it is usually correct.

8.2.3 Recall

Recall (also called sensitivity or true positive rate) measures the ability of the model to identify actual positives:

$$Recall = \frac{TP + FN}{TP}$$

High recall indicates that the model successfully detects most of the positive cases.

8.2.4 F1-Score

Precision and recall often trade off against each other: increasing recall may lower precision, and vice versa. To balance these metrics, the F1-score is frequently used, defined as the harmonic mean of precision and recall:

$$F1score = \frac{2 \times (Precision + Recall)}{Precision \times Recall}$$

This provides a single measure that rewards models with both high precision and high recall.

8.3 Model performance during training

The Xception architecture achieved high training performance across all three models, consistently surpassing 99% accuracy with training loss remaining below 0.5. Within five epochs, accuracy exceeded 95% and loss fell below 0.2, after which training stabilized (Figure 8.4). Validation accuracy closely tracked training accuracy in models with the training data containing anomaly(damage)-free patches. However, the inclusion of anomalous patches in the model trained with both types of patches, introduced greater variability in validation loss during the first 30 epochs and a temporary decline in validation accuracy compared to the model with only anomaly-free patches. In the model with only anomalous patches, validation accuracy dropped to 87.9% with a loss of 0.82.

The reduced variability and faster convergence observed from the model with all patches, and the model with only anomaly-free patches suggest that excluding anomalous patches facilitates a more stable training process. Conversely, omitting anomaly-free patches led to a performance decline, with validation accuracy decreasing from 98.6% (model trained on both) to 87.9%. This indicates that a slight negative effect of anomalous patches on training stability and generalization, although the consistently high validation accuracies across all models suggest that the model's capacity to classify individual patches effectively is not compromised.

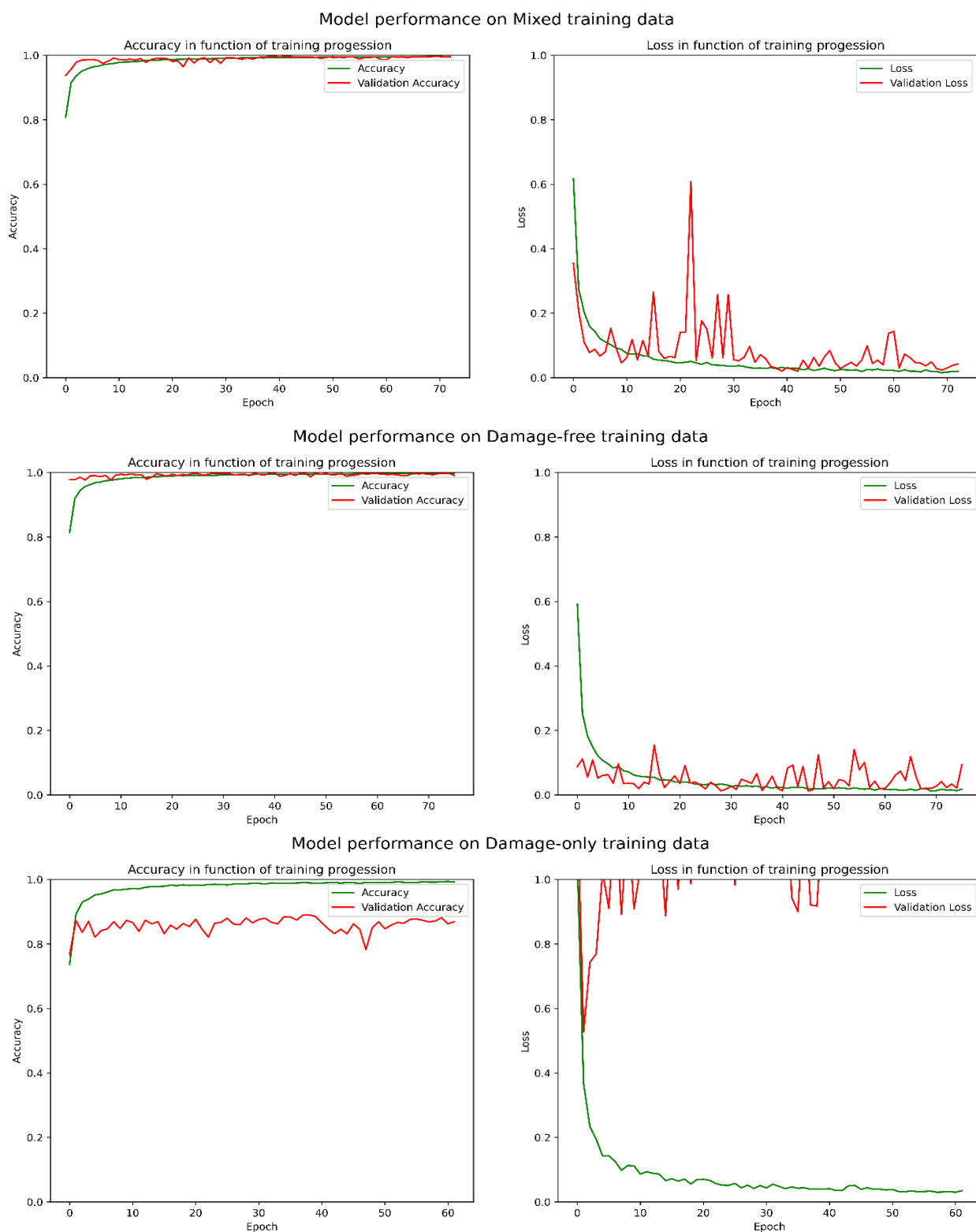


Figure 8.4: Training progress of accuracy and loss for the 3 models: Upper: model 1 (trained on all patches), Middle: model 2 (trained only on anomaly(damage)-free patches, Lower: model 3 (trained only on anomalous patches).

8.4 Correlations between class-specific changes in recall

8.4.1 Across trained models and model-specific training metrics

Pearson correlation coefficients revealed weak links between training data metrics and recall differences, except for stronger correlations with the average number of training patches per genus between models 1 and 2 (0.55 to 0.67). This suggests larger endgrain surface areas improve performance on anomaly-free specimens.

Table 10.6: Pearson correlation coefficients between the class-specific metrics of the training data and the performance differences between the three models on the (filtered) test data. Model 1 was trained on all images, model 2 was trained only on anomaly-free images, model 3 was trained only on anomalous images

	Total number of specimens in Training data	Total number of image patches in Training data	Average number of image patches across specimens per class (genus)
Performance difference on all test images (model 1 vs 2)	-0,15	0,01	0,55
Performance difference on all test images (model 1 vs 3)	0,16	0,20	0,14
Performance difference on anomaly-free test images (model 1 vs 2)	-0,21	-0,06	0,67
Performance difference on anomaly-free test images (model 1 vs 3)	0,10	0,20	0,20
Performance difference on anomalous test images (model 1 vs 2)	-0,05	0,15	0,36
Performance difference on anomalous test images (model 1 vs 3)	0,07	0,16	0,16

8.4.2 Within trained models (between anomaly-free and anomalous test data) and model-specific training metrics

Pearson correlations between class-specific performance differences and training data metrics were weak (<30%), except for a moderate positive correlation with the average number of image patches per genus in the model trained on all patches (see Table 2 and Supplementary material 1). This indicates that the training data structure does not significantly correlate with recall differences.

Table 10.7: Pearson correlation coefficients between the class-specific metrics of the training data and the performance differences between the anomaly-free and anomalous test patches on the training models.

	Total number of specimens in Training data	Total number of image patches in Training data	Average number of image patches across specimens per class (genus)
Performance difference on anomaly-free and anomalous test images (model trained on all patches)	-0,006	0,115	0,352
Performance difference on anomaly-free and anomalous test images (model trained only on anomaly-free patches)	0,101	0,262	0,137
Performance difference on anomaly-free and anomalous test images (model trained only on anomalous patches)	-0,089	-0,027	0,180

Table 8.8: Full overview of all specimens (split across reference and test sets) of non-Congolese timbers used in chapter 5 on the generalisation of object re-identification beyond the learned taxonomic scope of African timbers.

Family	Number of Training Specimen par Family	Number of Testing Specimen par Family	Genus	Number of Training Specimen par Genus	Number of Testing Specimen par Genus	Species	Training Specimen ID Tervuren xylarium	Testing Specimen ID Tervuren xylarium
Achariaceae	18	7	Erythrospermum Lam.	8	3	Erythrospermum candidum Becc.	Tw74955, Tw74581, Tw80935, Tw73711, Tw74744, Tw73826, Tw72021, Tw73975	Tw74576, Tw73380, Tw73899
			Pangium Reinw.	5	2	Pangium edule Reinw.	Tw74592, Tw72760, Tw73531, Tw71974, Tw73773	Tw71113, Tw71060
			Trichadenia Thwaites	5	2	Trichadenia philippinensis Merr.	Tw74201, Tw72936, Tw73402, Tw74745, Tw74640	Tw72756, Tw72221
Anacardiaceae	82	33	Buchanania Spreng.	27	10	Buchanania arborescens Blume	Tw74789, Tw71377, Tw74311, Tw74905	Tw71294, Tw71289
						Buchanania macrocarpa Merr. ex Setch.	Tw72131, Tw72630, Tw73652, Tw71702, Tw73172, Tw73296, Tw72935a, Tw71496, Tw73848, Tw72794, Tw74714, Tw73400, Tw72761, Tw74292, Tw71824, Tw74492, Tw74037, Tw73213, Tw74883a, Tw71226, Tw74883b, Tw72284, Tw74468	Tw73184, Tw72246, Tw72244, Tw72943, Tw71115, Tw74865, Tw74012, Tw71688
			Campnosperma Thwaites	19	8	Campnosperma brevipetiolatum Volkens	Tw71390, Tw72779, Tw72422, Tw71922, Tw72988, Tw73406, Tw71283, Tw71766, Tw73775, Tw74983, Tw71150, Tw74464, Tw74482, Tw74615, Tw72981, Tw72806	Tw74922, Tw72965, Tw74800, Tw73734, Tw71045, Tw71284
						Campnosperma coriaceum (Jack) Hallier fil.	Tw73003	Tw72168
						Campnosperma montanum Lauterb.	Tw74914, Tw73058	Tw72871
			Dracontomelon Blume	16	6	Dracontomelon dao (Blanco) Merr. & Rolfe	Tw71997, Tw72408, Tw73645, Tw73502, Tw73803, Tw73314, Tw72602, Tw73201, Tw73858, Tw72675, Tw74550, Tw73552, Tw72922, Tw74899, Tw71823, Tw73916	Tw73339, Tw74785, Tw72384, Tw74828, Tw73672, Tw71435

			Gluta L.	1	1	Gluta papuana Ding Hou	Tw72489	Tw72181
			Koordersiodendron Engl. ex Koord.	6	3	Koordersiodendron pinnatum (Blanco) Merr.	Tw74258, Tw74187, Tw74593, Tw72855, Tw71707, Tw74943	Tw71437, Tw72480, Tw71245
			Pentaspadon Hook.f.	6	2	Pentaspadon motleyi Hook.fil.	Tw72087, Tw72641, Tw71472, Tw71506, Tw72331, Tw72968	Tw71402, Tw72065
			Semecarpus L.f.	7	3	Semecarpus aruensis Engl.	Tw74892	Tw72588
Annonaceae	26	11	Cananga (Dunal) Hook.f. & Thomson	20	7	Cananga odorata (Lam.) Hook.f. & Thomson	Tw73981, Tw74260, Tw73295, Tw72071, Tw73781, Tw74569, Tw72968	Tw73468, Tw71287
							Tw74185, Tw72653, Tw74695, Tw71779, Tw71695, Tw72283, Tw71387, Tw73083, Tw73474, Tw71339, Tw72000, Tw74502, Tw73150, Tw74275, Tw73489, Tw74938, Tw73320, Tw73494, Tw72944, Tw74716	Tw73530, Tw72685, Tw73075, Tw73144, Tw71660, Tw72196, Tw71190
			Drepananthus Maingay ex Hook.f. & Thomson	3	2	Drepananthus petiolatus (Diels) Survesw. & R.M.K.Saunders	Tw73908, Tw73514	Tw71931
			Miliusa Lesch. ex A.DC.	3	2	Drepananthus polycarpus (C.T.White & W.D.Francis) Survesw. & R.M.K. & Miliusa koolsii (Kosterm.) J.Sinclair	Tw74452	Tw72734
Apocynaceae	18	9	Cerbera L.	6	3	Cerbera floribunda K.Schum.	Tw74484, Tw74885, Tw74966, Tw74656, Tw74367, Tw72768	Tw72418, Tw71454, Tw73783
			Ochrosia Juss.	6	3	Ochrosia ficifolia (S.Moore) Markgr.	Tw74528, Tw72709, Tw74698, Tw72757	Tw72573, Tw72632
			Wrightia R.Br.	6	3	Ochrosia glomerata (Blume) F.Muell.	Tw74064, Tw72464	Tw72204
						Wrightia laevis Hook.fil.	Tw74755, Tw72515, Tw74838, Tw74405, Tw73796, Tw72380	Tw72462, Tw72377, Tw72295
Aquifoliaceae	2	1	Ilex L.L.	2	1	Ilex cymosa Blume	Tw71170, Tw71235	Tw71074
Bignoniaceae	3	1	Dolichandrone (Fenzl) Seem.	3	1	Dolichandrone spathacea (L.fil.) K.Schum.	Tw72180, Tw74595, Tw73022	Tw74274
Burseraceae	34	13	Garuga Roxb.	1	1	Garuga floribunda Decne.	Tw74288	Tw73661

			Haplolobus H.J.Lam	33	12	Haplolobus acuminatus (K.Schum.) H.J.Lam	Tw73821, Tw73968, Tw74912, Tw73145	Tw72939, Tw72706
						Haplolobus floribundus (K.Schum.) H.J.Lam	Tw73596, Tw71674, Tw72738, Tw74529, Tw71296, Tw71556, Tw73164, Tw73573, Tw71676, Tw73693, Tw73478, Tw73699, Tw73115, Tw73352, Tw71677, Tw73316, Tw74834, Tw74812, Tw73301, Tw72989, Tw74632, Tw73073, Tw72780, Tw72990, Tw74149, Tw73591, Tw73729	Tw71684, Tw71585, Tw73732, Tw72695, Tw74897, Tw73961, Tw71169, Tw73357, Tw74148
						Haplolobus leifolius (Lauterb.) H.J.Lam	Tw71964, Tw72502	Tw71921
Calophyllaceae	10	5	Calophyllum L.	10	5	Calophyllum euryphyllum Lauterb.	Tw73653	Tw71330
						Calophyllum peekelii Lauterb.	Tw71587	Tw71174
						Calophyllum soulattri Burm.fil.	Tw74760, Tw74741	Tw72044
						Calophyllum suberosum P.F.Stevens	Tw72978, Tw72172, Tw73013	Tw72980
						Calophyllum vexans P.F.Stevens	Tw71225, Tw71987, Tw71465	Tw71968
Cannabaceae	22	8	Gironniera Gaudich.	22	8	Gironniera hirta Ridl.	Tw74113, Tw74151	Tw74111
						Gironniera subaequalis Planch.	Tw71604, Tw74433, Tw73212, Tw71253, Tw74807, Tw72764, Tw71259, Tw73227, Tw74307, Tw74749, Tw71732, Tw73395, Tw72723, Tw75002a, Tw75002b, Tw72908, Tw71541, Tw73386, Tw72920, Tw73087	Tw72128, Tw71529, Tw72917, Tw72929, Tw72777, Tw73874, Tw71096
Cardiopteridaceae	3	2	Citronella D.Don	2	1	Citronella suaveolens (Blume) R.A.Howard	Tw73049, Tw74473	Tw71929
			Gonocaryum Miq.	1	1	Gonocaryum litorale (Blume) Sleumer	Tw72504	Tw72473
Celastraceae	8	3	Lophopetalum Wight ex Arn.	8	3	Lophopetalum ledermannii (Loes.) Ding Hou	Tw72876, Tw72804	Tw72796
						Lophopetalum torricellense Loes.	Tw72640, Tw74921, Tw74574, Tw74848, Tw74996, Tw74862	Tw72589, Tw74795

Centroplacaceae	2	1	Bhesa Buch.-Ham. ex Arn.	2	1	Bhesa archboldiana (Merr. & L.M.Perry) Ding Hou	Tw74353, Tw74896	Tw73401
Chrysobalanaceae	15	5	Atuna Raf.	6	2	Atuna excelsa (Jack) Kosterm.	Tw72306, Tw72309, Tw72600, Tw74406	Tw71507, Tw72501
			Parastemon A.DC.	9	3	Parastemon versteeghii Merr. & Perry	Tw74377, Tw73598, Tw73987, Tw73822, Tw72627	Tw73867, Tw73606, Tw72485
Cornaceae	10	4	Alangium Lam.	10	4	Alangium javanicum (Blume) Wangerin	Tw74070, Tw74023, Tw73941, Tw74950, Tw74305, Tw73976, Tw73717, Tw73812, Tw73817, Tw74303	Tw72079, Tw74306, Tw74671, Tw72110
Ctenolophonaceae	2	1	Ctenolophon Oliv.	2	1	Ctenolophon parvifolius Oliv.	Tw74223, Tw72365	Tw72256
Cunoniaceae	3	1	Schizomeria D.Don	3	1	Schizomeria serrata (Hochr.) Hochr.	Tw72716, Tw73578, Tw74768	Tw74162
Dilleniaceae	11	5	Dillenia L.	11	5	Dillenia castaneifolia (Miq.) Martelli ex T.Durand & B.D.Jacks.	Tw73650, Tw73845	Tw72047
						Dillenia papuana Martelli	Tw71152, Tw71765, Tw74777, Tw72900, Tw74446, Tw73050, Tw72175, Tw73167	Tw72405, Tw72910, Tw71719
						Dillenia pteropoda (Miq.) Hoogland	Tw72506	Tw71468
Dipterocarpaceae	44	17	Anisoptera Korth.	14	5	Anisoptera thurifera (Blanco) Blume	Tw73869, Tw74814, Tw73000, Tw72074, Tw73683, Tw74222, Tw74366, Tw73007, Tw71923, Tw72866, Tw72795, Tw73617, Tw73002, Tw73872	Tw73188, Tw72169, Tw74401, Tw71565, Tw73687
			Hopea Roxb.	9	5	Hopea iriana Slooten	Tw73686, Tw72954, Tw74673	Tw72946, Tw72495
						Hopea novoguineensis Slooten	Tw72816, Tw74934	Tw72491
						Hopea papuana Diels	Tw73712, Tw71992	Tw71991
						Hopea scabra P.S.Ashton	Tw73580, Tw71971	Tw71962
			Vatica L.	21	7	Vatica rassak (Korth.) Blume	Tw74364, Tw74968, Tw71307, Tw72826, Tw72153, Tw71509, Tw72537, Tw72498, Tw73629, Tw71734, Tw72314, Tw74967, Tw73720, Tw71501,	Tw72237, Tw72130, Tw71727, Tw73601, Tw72313, Tw71265, Tw71434

							Tw71547, Tw71425, Tw74627, Tw72864	Tw74283, Tw72299, Tw71731,	
Elaeocarpaceae	54	23	Elaeocarpus L.	35	14	Elaeocarpus altisectus Schltr.	Tw74347, Tw73042	Tw72577,	Tw73620
						Elaeocarpus dolichostylus Schltr.	Tw71623, Tw74453, Tw74120, Tw74876, Tw71633, Tw72775	Tw72953, Tw73925, Tw74121, Tw73922,	Tw74618, Tw74775, Tw71338, Tw71368
						Elaeocarpus miegei Weibel	Tw71181, Tw73792	Tw71228,	Tw73024
						Elaeocarpus nouhuysii Koord.	Tw71184, Tw72397, Tw71641, Tw73139, Tw71773, Tw74873	Tw71656, Tw73378, Tw73756, Tw72316, Tw73291,	Tw71077, Tw73439, Tw71111, Tw74509
						Elaeocarpus sepikanus Schltr.	Tw74400, Tw71630	Tw71620,	Tw73749
						Elaeocarpus sphaericus (Gaertn.) Ettingsh.	Tw73888, Tw73777, Tw72671, Tw74319		Tw71761, Tw71808
						Elaeocarpus undulatus Warb.	Tw73820		Tw72271
			Sloanea L.	19	9	Sloanea aberrans (Brandis) A.C.Sm.	Tw74621, Tw74954, Tw74371		Tw72496, Tw73258
						Sloanea brachystyla (Schltr.) A.C.Sm.	Tw73030		Tw73025
						Sloanea forbesii F.Muell.	Tw73486, Tw72317		Tw71104
						Sloanea paradisearum F.Muell.	Tw74075, Tw74459		Tw73244
						Sloanea pullei ex O.C.Schmidt A.C.Sm.	Tw72941, Tw73102, Tw72740, Tw72844, Tw73142, Tw72762, Tw73414, Tw73735, Tw72420, Tw74375, Tw74988		Tw71628, Tw74441, Tw71581, Tw73737
Euphorbiaceae	21	12	Aleurites J.R.Forst. & G.Forst.	1	1	Aleurites moluccanus (L.) Willd.	Tw74977		Tw71476
			Balakata Esser	1	1	Balakata luzonica (Vidal) Esser	Tw73338		Tw72027
			Endospermum Benth.	11	5	Endospermum medullosum L.S.Sm.	Tw74117, Tw72650, Tw74198, Tw74704, Tw74128, Tw74706		Tw72310, Tw72201, Tw73814
						Endospermum moluccanum (Teijsm. & Binn.) Kurz	Tw74864, Tw71762, Tw72870, Tw73260, Tw72460		Tw71129, Tw71127

			Excoecaria L.	1	1	Excoecaria myrionura Airy Shaw	Tw74361	Tw74157
			Mallotus Lour.	6	3	Mallotus philippensis (Lam.) Müll.Arg.	Tw73521, Tw73554	Tw72466
						Mallotus pleiogynus Pax & K.Hoffm.	Tw72727, Tw74228, Tw74963, Tw74277	Tw72385, Tw71989
			Spathiostemon Blume	1	1	Spathiostemon javensis Blume	Tw73719	Tw73669
Fabaceae	23	12	Adenanthera L.	2	1	Adenanthera novoguineensis Baker f.	Tw72477, Tw74435	Tw72112
			Archidendron F.Muell.	1	1	Archidendron molle (K.Schum.) de Wit	Tw73664	Tw73648
			Falcataria (I.C.Nielsen) Barneby & J.W.Grimes	3	2	Falcataria falcata (L.) Greuter & R.Rankin	Tw74474, Tw74322, Tw74566	Tw73240, Tw73483
			Inocarpus J.R.Forst. & G.Forst.	11	5	Inocarpus fagifer ex (Parkinson F.A.Zorn) Fosberg	Tw74940, Tw72179, Tw72452, Tw71549, Tw71768, Tw72404, Tw71391, Tw72960, Tw74483, Tw72356	Tw71050, Tw74756, Tw74894, Tw71379
						Inocarpus papuanus Kosterm.	Tw73193	Tw73179
			Pongamia Adans.	3	2	Pongamia pinnata (L.) Pierre	Tw73856, Tw73533, Tw73722	Tw72789, Tw71994
			Serianthes Benth.	3	1	Serianthes minahassae (Koord.) Merr. & L.M.Perry	Tw74895, Tw71503, Tw72483	Tw72585
Fagaceae	7	3	Castanopsis (D.Don) Spach	2	1	Castanopsis acuminatissima (Blume) A.DC.	Tw73451, Tw73477	Tw73252
			Lithocarpus Blume	5	2	Lithocarpus celebicus (Miq.) Rehder	Tw73742, Tw72919	Tw71171
						Lithocarpus rufovillosus (Markgr.) Rehder	Tw74373, Tw72676, Tw73436	Tw73450
Gentianaceae	3	2	Fagraea Thunb.	3	2	Fagraea gracilipes A.Gray	Tw72950, Tw73626, Tw73105	Tw72812, Tw72690
Hernandiaceae	15	5	Hernandia L.	15	5	Hernandia ovigera L.	Tw74471, Tw74910, Tw73843, Tw72805, Tw71705, Tw71572, Tw73463, Tw73107, Tw73498, Tw74450, Tw74549, Tw73276, Tw71976, Tw74973, Tw71480	Tw71689, Tw73780, Tw73399, Tw71196, Tw74869
Himantandraceae	6	2	Galbulimima F.M.Bailey	6	2	Galbulimima belgraveana (F.Muell.) Sprague	Tw74080, Tw73225, Tw74156, Tw73415, Tw74540, Tw74372	Tw73613, Tw73113

Juglandaceae	8	3	Engelhardia Lesch. ex Blume	8	3	Engelhardia rigida Blume	Tw74065, Tw74047, Tw74399, Tw73445, Tw74412, Tw73738	Tw73111, Tw73746
						Engelhardia spicata Lesch. ex Blume	Tw74819, Tw74669	Tw74628
Lamiaceae	7	3	Teijsmanniodendron Koord.	7	3	Teijsmanniodendron bogoriense Koord.	Tw72104, Tw73873, Tw73416, Tw73413, Tw74815, Tw72752, Tw72721	Tw73403, Tw74428, Tw72587
Lauraceae	68	35	Alseodaphne Nees	5	2	Alseodaphne archboldiana (C.K.Allen) Kosterm.	Tw71687, Tw74281, Tw72951, Tw71746, Tw73248	Tw71430, Tw71222
			Cinnamomum Schaeff.	2	1	Cinnamomum grandiflorum Kosterm.	Tw72520, Tw72976	Tw71925
			Cryptocarya R.Br.	9	5	Cryptocarya alleniana C.T.White	Tw73430, Tw73119	Tw72923
						Cryptocarya brevipes C.K.Allen	Tw74112	Tw74106
						Cryptocarya diversifolia Blume	Tw73710	Tw72503
						Cryptocarya murrayi F.Muell.	Tw74316, Tw73834, Tw73562	Tw74123
						Cryptocarya verrucosa Teschner	Tw71334, Tw71306	Tw71292
			Endiandra R.Br.	7	4	Endiandra brassii C.K.Allen	Tw74650, Tw72693	Tw71449
						Endiandra forbesii Gamble	Tw74004, Tw74804	Tw71234
						Endiandra latifolia Kosterm.	Tw74489, Tw74526	Tw71643
						Endiandra ledermannii Teschner	Tw74849	Tw72896
			Litsea Lam.	41	20	Litsea calophyllantha K.Schum.	Tw74684	Tw72755
						Litsea elliptica Blume	Tw74176	Tw74132
						Litsea firma (Blume) Hook.fil.	Tw71215, Tw74357, Tw74699, Tw74362, Tw74060, Tw72912, Tw73104, Tw72417, Tw72884	Tw72146, Tw74136, Tw74015, Tw74079
						Litsea glutinosa (Lour.) C.B.Rob.	Tw74878	Tw74500
						Litsea grandis (Wall. ex Nees) Hook.fil.	Tw74061	Tw74025
						Litsea irianensis Kosterm.	Tw74118, Tw73541, Tw73311	Tw72620, Tw71117
						Litsea ledermannii Teschner	Tw73411, Tw74919, Tw72867, Tw73464, Tw73421, Tw74806	Tw73097, Tw72666, Tw72724

						Litsea timoriana Span.	Tw74125, Tw72629, Tw74532, Tw74284, Tw74248, Tw73308, Tw71081, Tw72655, Tw73884, Tw73524	Tw2885, Tw74124, Tw71632, Tw74246, Tw72341, Tw74559, Tw74636, Tw72129, Tw72320, Tw73524	Tw72750, Tw73495, Tw71071, Tw72343	Tw74245, Tw73264, Tw71737,
			Nothaphoebe Blume	3	2	Nothaphoebe archboldiana C.K.Allen	Tw74608		Tw73763	
						Nothaphoebe elata (Kosterm.) Kosterm.	Tw71146, Tw73085		Tw71080	
			Phoebe Nees	1	1	Phoebe forbesii Gamble	Tw74083		Tw73190	
Lecythidaceae	9	4	Barringtonia J.R.Forst. & G.Forst.	7	3	Barringtonia lauterbachii R.Knuth	Tw73971, Tw74675, Tw74739, Tw71112	Tw73804, Tw73844, Tw80930,	Tw74903, Tw73960	Tw71158,
			Planchonia Blume	2	1	Planchonia papuana R.Knuth	Tw73949, Tw73937		Tw72214	
Loganiaceae	3	2	Neuburgia Blume	3	2	Neuburgia corynocarpa (A.Gray) Leenh.	Tw74253, Tw74272	Tw74270,	Tw71075, Tw74166	
Lythraceae	4	3	Duabanga Buch.-Ham.	1	1	Duabanga moluccana Blume	Tw74945		Tw74419	
			Lagerstroemia L.	2	1	Lagerstroemia celebica Blume	Tw72901, Tw72624		Tw72584	
			Sonneratia L.f.	1	1	Sonneratia caseolaris (L.) Engl.	Tw72555		Tw72359	
Magnoliaceae	4	2	Magnolia Plum. ex L.	4	2	Magnolia tsiampacca (L.) Figlar & Noot.	Tw74710, Tw74898, Tw73427	Tw74158, Tw73427	Tw72711, Tw72613	
Malvaceae	16	8	Colona Cav.	9	3	Colona scabra (Sm.) Burret	Tw72433, Tw72773, Tw74722, Tw72430, Tw74729	Tw72162, Tw74328, Tw73082, Tw74510,	Tw74649, Tw73151, Tw72415	
			Kleinhovia L.	3	2	Kleinhovia hospita L.	Tw73526, Tw73784	Tw72516,	Tw72265, Tw71340	
			Trichospermum Blume	4	3	Trichospermum pleiostigma (F.Muell.) Kosterm.	Tw73458, Tw74479	Tw73488,	Tw73440, Tw73455	
						Trichospermum tripyxis (Schum.) Kosterm.	Tw74638		Tw71475	
Melastomataceae	8	4	Astronia Blume	5	2	Astronia hollrungii Cogn.	Tw74490, Tw74591, Tw74451	Tw73204, Tw73435,	Tw72751, Tw72959	

			Astronidium A.Gray	2	1	Astronidium biakense J.F.Maxwell	Tw74030, Tw74035	Tw74029
			Pternandra Jack	1	1	Pternandra tuberculata (Korth.) M.P.Nayar	Tw74674	Tw71532
Meliaceae	86	39	Aglaia Lour.	22	11	Aglaia flavida Merr. & L.M.Perry	Tw72018, Tw72992	Tw71085
						Aglaia lawii (Wight) C.J.Saldanha	Tw73299, Tw74332, Tw72670, Tw73284, Tw73304	Tw72481, Tw72070
						Aglaia lepiorrhachis Harms	Tw74942	Tw73948
						Aglaia oligophylla Miq.	Tw74712	Tw74421
						Aglaia sexipetala Griff.	Tw73662	Tw73532
						Aglaia silvestris (M.Roem.) Merr.	Tw72785, Tw73833, Tw74203, Tw72014, Tw73921, Tw74205, Tw74742, Tw74668	Tw73919, Tw74066, Tw72357
						Aglaia spectabilis (Miq.) S.S.Jain & Bennet	Tw71322	Tw71304
						Aglaia subminutiflora C.DC.	Tw73286, Tw71328, Tw73564	Tw73329
			Aphanamixis Pierre	1	1	Aphanamixis polystachya (Wall.) R.Parker	Tw73384	Tw72342
			Azadirachta A.Juss.	3	2	Azadirachta excelsa (Jack) Jacobs	Tw73211, Tw72425, Tw72429	Tw71785, Tw72290
			Chisocheton Blume	20	9	Chisocheton ceramicus Miq.	Tw73503, Tw74350, Tw73067, Tw72414, Tw71063, Tw74243, Tw72272, Tw74570	Tw73426, Tw72066, Tw72592
						Chisocheton cumingianus (C.DC.) Harms	Tw73808	Tw73441
						Chisocheton lasiocarpum (Miq.) Valeton	Tw71998, Tw71333, Tw72543, Tw74438, Tw74182, Tw73232, Tw74449, Tw73665	Tw73368, Tw73064, Tw71508
						Chisocheton longistipitatus (F.M.Bailey) L.S.Sm.	Tw71800	Tw71203
						Chisocheton stellatus P.F.Stevens	Tw73846, Tw73663	Tw71189
			Dysoxylum Blume	11	5	Dysoxylum excelsum Blume	Tw74478, Tw73835, Tw73496, Tw73404, Tw74560	Tw72963, Tw73210
						Dysoxylum mollissimum Blume	Tw73482, Tw73283, Tw73432	Tw73434

						Dysoxylum oppositifolium F.Muell.	Tw72997, Tw73034	Tw74542,	Tw71502, Tw72421
			Pseudocarapa Hemsl.	6	2	Pseudocarapa inopinata Harms	Tw74096, Tw74647, Tw72742, Tw74388	Tw72798, Tw74014,	Tw72722, Tw72934
			Sandoricum Cav.	4	2	Sandoricum koetjape (Burm.fil.) Merr.	Tw74734, Tw74282, Tw74928	Tw74290,	Tw74271, Tw74231
			Toona (Endl.) M.Roem.	5	2	Toona sureni (Blume) Merr.	Tw74979, Tw73794, Tw73500	Tw73733, Tw73472,	Tw73268, Tw73467
			Vavaea Benth.	12	4	Vavaea amicorum Benth.	Tw72673, Tw74172, Tw72661, Tw74285, Tw74291, Tw72488, Tw73743	Tw74981, Tw72134, Tw71970, Tw72667,	Tw74192, Tw74590, Tw74167, Tw73631
			Xylocarpus J.Koenig	2	1	Xylocarpus granatum J.Koenig	Tw74603, Tw74730		Tw71414
Metteniusaceae	4	2	Platea Blume	4	2	Platea excelsa Blume	Tw71616, Tw73758, Tw72776, Tw71721		Tw71308, Tw71124
Moraceae	18	7	Parartocarpus Baill.	4	2	Parartocarpus venenosa (Zoll. & Moritzi) Becc.	Tw74078, Tw74582, Tw75001, Tw74696		Tw72080, Tw71937
Moraceae	18	7	Paratrophis Blume	8	3	Paratrophis glabra (Merr.) Steenis	Tw73906, Tw74661		Tw73237
						Paratrophis philippinensis (Bureau) Fern.-Vill.	Tw73340, Tw72603, Tw72928, Tw74890, Tw72260, Tw73180		Tw72714, Tw71069
			Prainea King ex Hook.f.	6	2	Prainea limpató (Miq.) Beumée ex K.Heyne	Tw73923, Tw74140, Tw74978, Tw71580, Tw74831, Tw74635		Tw74031, Tw71559
Myristicaceae	50	22	Endocomia W.J.de Wilde	3	1	Endocomia macrocoma (Miq.) W.J.de Wilde	Tw74678, Tw72697, Tw73818		Tw74491
			Horsfieldia Willd.	16	7	Horsfieldia irya (Gaertn.) Warb.	Tw74425		Tw71094
						Horsfieldia pachyrachis W.J.de Wilde	Tw74497		Tw71367
						Horsfieldia sylvestris (Houtt.) Warb.	Tw73681, Tw71806, Tw71639, Tw73839, Tw71219, Tw71137, Tw72862, Tw74867, Tw72771, Tw71794, Tw71791, Tw71792, Tw71756, Tw71653		Tw74855, Tw71078, Tw71793, Tw71946, Tw71432
			Myristica Gronov.	31	14	Myristica fatua Houtt.	Tw73057, Tw71221, Tw74374, Tw74280, Tw74926, Tw74547, Tw71535		Tw71504, Tw74753, Tw74226

						Myristica garciniifolia Warb.	Tw72546	Tw71271
						Myristica globosa Warb.	Tw73709, Tw73569	Tw73341
						Myristica hollrungii Warb.	Tw72167, Tw72040, Tw73037, Tw74503, Tw72388, Tw71610, Tw72238, Tw71375, Tw72006, Tw72681	Tw71042, Tw73106, Tw73133, Tw71299
						Myristica inutilis Rich. ex A.Gray	Tw74682, Tw72219, Tw72045	Tw73895
						Myristica mediovibex W.J.de Wilde	Tw74393, Tw73882	Tw73592
						Myristica scripta W.J.de Wilde	Tw74131	Tw74119
						Myristica sulcata Warb.	Tw73649, Tw74469, Tw73614, Tw73830	Tw73156, Tw72710
Myrtaceae	10	5	Eucalyptopsis C.T.White	1	1	Eucalyptopsis papuana C.T.White	Tw72559	Tw72367
			Rhodamnia Jack	4	2	Rhodamnia pachyloba A.J.Scott	Tw73374, Tw72707, Tw74499	Tw74434
			Rhodamnia reticulata A.J.Scott				Tw74663	Tw72726
			Tristaniopsis Brongn. & Gris	5	2	Tristaniopsis ferruginea (C.T.White) Paul G.Wilson & J.T.Waterh.	Tw73701, Tw73692, Tw74705, Tw72879, Tw74210	Tw72857, Tw72143
Nyssaceae	3	1	Mastixia Blume	3	1	Mastixia kaniensis Melch.	Tw72835, Tw74700, Tw73682	Tw73885
Oleaceae	2	1	Chionanthus Royen	2	1	Chionanthus ramiflorus Roxb.	Tw74936, Tw74302	Tw72125
Pentaphylacaceae	3	1	Ternstroemia Mutis ex L.f.	3	1	Ternstroemia merrilliana Kobuski	Tw72248, Tw74667, Tw71517	Tw72561
Phyllanthaceae	11	5	Baccaurea Lour.	3	1	Baccaurea nanihua Merr.	Tw74073, Tw74155, Tw72895	Tw74126
			Bischofia Blume	4	2	Bischofia javanica Blume	Tw74565, Tw73547, Tw74515, Tw73520	Tw72637, Tw72007
			Glochidion J.R.Forst. & G.Forst.	4	2	Glochidion lucidum Blume	Tw74995	Tw71533
						Glochidion stenophyllum Airy Shaw	Tw72072, Tw72595, Tw74152	Tw73689
Podocarpaceae	4	2	Nageia Gaertn.	3	1	Nageia wallichiana (Presl) Kuntze	Tw72873, Tw72536, Tw74626	Tw72916
			Podocarpus L'Hér. ex Pers.	1	1	Podocarpus rumphii Blume	Tw74331	Tw74251
Polygalaceae	4	2	Xanthophyllum Roxb.	4	2	Xanthophyllum papuanum Whitmore ex Meijden	Tw71351, Tw72754, Tw71070	Tw72656

						Xanthophyllum suberosum C.T.White	Tw75005	Tw74224
Rhamnaceae	7	4	Alphitonia Reissek ex Endl.	7	4	Alphitonia excelsa (Fenzl) Benth.	Tw73480, Tw74900, Tw73457, Tw73456	Tw73424, Tw72509
						Alphitonia macrocarpa Mansfield	Tw74676, Tw74539	Tw74708, Tw71635, Tw73372
Rhizophoraceae	12	6	Bruguiera Lam.	2	1	Bruguiera gymnorhiza (L.) Lam.	Tw74604, Tw74686	Tw72156
			Carallia Roxb. ex R.Br.	9	4	Carallia brachiata (Lour.) Merr.	Tw72401, Tw74351, Tw72875, Tw73787, Tw73523, Tw73076, Tw73091, Tw74792, Tw74623	Tw74074, Tw74114, Tw72401a, Tw73725
			Gynotroches Blume	1	1	Gynotroches axillaris Blume	Tw74929	Tw74798
Rubiaceae	31	14	Adina Salisb.	2	1	Adina multifolia Havil.	Tw72275, Tw73166	Tw72251
			Mastixiodendron Melch.	11	4	Mastixiodendron pachyclados (K.Schum.) Melch.	Tw72036, Tw71555, Tw74641, Tw71701, Tw71557, Tw71936	Tw71709, Tw72200, Tw71825, Tw71941, Tw73768, Tw71255, Tw73956, Tw71199
			Neolamarckia Bosser	7	3	Neolamarckia cadamba (Roxb.) Bosser	Tw74980, Tw73305, Tw74552, Tw74769	Tw74178, Tw71927, Tw73141, Tw72098, Tw74964
			Neonauclea Merr.	9	5	Neonauclea brassii S.Moore	Tw74447, Tw74925, Tw73565, Tw74141	Tw73108, Tw73247, Tw72493, Tw74501, Tw72662, Tw73257
						Neonauclea hagenii (Lauterb. & K.Schum.) Merr.	Tw73449	Tw72969
						Neonauclea lanceolata (Blume) Merr.	Tw74818	Tw74229
			Timonius Rumph. ex DC.	2	1	Timonius timon (Spreng.) Merr.	Tw72004, Tw73779	Tw71043
Rutaceae	34	16	Flindersia R.Br.	14	6	Flindersia amboinensis Poir.	Tw74356, Tw73487, Tw71131, Tw72399, Tw72609, Tw72427, Tw73255, Tw74553, Tw74597	Tw74536, Tw72383, Tw73361
						Flindersia pimenteliana F.Muell.	Tw74368, Tw74095, Tw74003, Tw74021	Tw73605, Tw73597
						Flindersia schottiana F.Muell.	Tw73563	Tw73259
			Halfordia F.Muell.	1	1	Halfordia kendack (Montrouz.) Guillaumin	Tw73608	Tw72792

			Melicope J.R.Forst. & G.Forst.	19	9	Melicope bonwickii (F.Muell.) T.G.Hartley	Tw74408, Tw72362	Tw73558,	Tw71782, Tw72279
						Melicope elleryana (F.Muell.) T.G.Hartley	Tw72538, Tw73173, Tw72434, Tw74530, Tw73940, Tw71750, Tw73954, Tw74466	Tw73223, Tw72578, Tw71777, Tw73074, Tw73280, Tw73066, Tw72850,	Tw74537, Tw73136, Tw71153, Tw73407, Tw74735, Tw72165
						Melicope exuta T.G.Hartley	Tw72853		Tw72277
Sabiaceae	1	1	Meliosma Blume	1	1	Meliosma pinnata (Roxb.) Maxim.	Tw74217		Tw72135
Salicaceae	4	2	Casearia Jacq.	1	1	Casearia grewiaefolia Vent.	Tw74586		Tw74567
			Itoa Hemsl.	3	1	Itoa stapfii (Koord.) Sleumer	Tw73275, Tw73490	Tw73285,	Tw73471
Sapindaceae	88	32	Arytera Blume	1	1	Arytera litoralis Blume	Tw74461		Tw74352
			Dimocarpus Lour.	1	1	Dimocarpus longan Lour.	Tw71738		Tw71673
			Pometia J.R.Forst. & G.Forst.	83	28	Pometia pinnata J.R.Forst. & G.Forst.	Tw73343, Tw72012, Tw71554, Tw73121, Tw73392a, Tw74998b, Tw74858, Tw74596, Tw72243, Tw71324, Tw72808, Tw71148, Tw71944, Tw72625, Tw71242, Tw73473, Tw72913, Tw71389, Tw72240, Tw73881a, Tw72807, Tw74511, Tw73159, Tw73397, Tw74780, Tw72608, Tw71041, Tw71563, Tw73392, Tw71144,	Tw72540, Tw73807, Tw71661, Tw72409, Tw73881, Tw73518, Tw71319, Tw71729, Tw71224, Tw71257, Tw73730, Tw71478, Tw73408, Tw71954, Tw71650, Tw72888, Tw72455, Tw73609, Tw73219, Tw74583, Tw74989, Tw74889, Tw74426, Tw72581, Tw74631, Tw74998a, Tw72325, Tw72529a, Tw71314,	Tw71461, Tw72743, Tw71950, Tw73630, Tw71120, Tw71960, Tw72529, Tw74998c, Tw72234, Tw71386, Tw72524a, Tw73604, Tw71363, Tw72043a, Tw71380, Tw72443, Tw74516, Tw74170, Tw72599, Tw71151, Tw71820, Tw73325, Tw72530, Tw71126, Tw73375, Tw72541, Tw72534, Tw71065

							Tw72043, Tw74177, Tw72787, Tw72970, Tw71913, Tw72601, Tw73220, Tw73890, Tw72118, Tw71358, Tw71419, Tw73398, Tw73375a	Tw71492, Tw72431, Tw74263, Tw71477, Tw72228, Tw73176, Tw71400, Tw72524, Tw71605, Tw71975, Tw72582, Tw73375a	
			Tristiropsis Radlk.	3	2	Tristiropsis acutangula Radlk.	Tw72225, Tw72458	Tw73932,	Tw72223, Tw71947
Sapotaceae	86	38	Burckella Pierre	14	5	Burckella macropoda (K.Krause) H.J.Lam	Tw74329, Tw74326, Tw74200, Tw74737	Tw73632, Tw73053,	Tw72512, Tw73953
						Burckella polymera P.Royen	Tw73695, Tw71498, Tw74961, Tw71481, Tw71603	Tw74924, Tw71448, Tw74754,	Tw71474, Tw72352, Tw71594
			Madhuca Buch.-Ham. ex J.F.Gmel.	4	3	Madhuca boerlageana (Burck) Baehni	Tw74068		Tw73607
						Madhuca leucodermis (K.Krause) H.J.Lam	Tw73221, Tw73405	Tw74160,	Tw71258, Tw71669
			Palaquium Blanco	40	17	Palaquium amboinense Burck	Tw72042a, Tw74278, Tw74204, Tw74841, Tw71967, Tw71963, Tw72972, Tw74327, Tw73123a, Tw71300	Tw71809, Tw74606, Tw74175, Tw72973, Tw71297, Tw71982, Tw71955, Tw71816, Tw71300	Tw71366, Tw72042, Tw71164, Tw73123, Tw74657, Tw73891
						Palaquium galactoxylum (F.Muell.) H.J.Lam	Tw73603, Tw73006	Tw72991,	Tw73008
						Palaquium lobbianum Burck	Tw74051, Tw71320, Tw74917, Tw73101, Tw71369, Tw71374	Tw74032, Tw74770, Tw72786, Tw72937, Tw73793,	Tw74888, Tw71309, Tw71200, Tw74707
						Palaquium obtusifolium Burck	Tw74485, Tw73540	Tw73770,	Tw73307, Tw72725
						Palaquium ridleyi King & Gamble	Tw74209, Tw74359	Tw74358,	Tw73599, Tw72797
						Palaquium supfianum Schltr.	Tw72924		Tw71583

						Palaquium warburgianum Schltr. ex K.Krause	Tw71571	Tw71263
			Planchonella Pierre	25	11	Planchonella chartacea (F.Muell. ex Benth.) H.J.Lam	Tw74391, Tw74213	Tw73658
						Planchonella keyensis H.J.Lam	Tw72413, Tw72926, Tw72435	Tw71431, Tw71047
						Planchonella menait (Vink) Swenson	Tw74665, Tw73684, Tw74087, Tw74028	Tw72574, Tw72025
						Planchonella myrsinodendron (F.Muell.) Swenson, Bartish & Munzinger	Tw74313, Tw74935	Tw72475
						Planchonella obovata (R.Br.) Pierre	Tw71485, Tw73390, Tw71346	Tw72241
						Planchonella pomifera (Pierre ex Baill.) Dubard	Tw72749	Tw71089
						Planchonella torricellensis (K.Schum.) H.J.Lam	Tw73538, Tw74517, Tw73566, Tw73358	Tw71813, Tw72889
						Planchonella xylocarpa (C.T.White) Swenson, Bartish & Munzinger	Tw71810, Tw74607, Tw71811	Tw71812
			Sarcosperma Hook.f.	3	2	Sarcosperma paniculatum (King) Stapf & King	Tw74548, Tw73460, Tw74816	Tw73250, Tw73130
Simaroubaceae	7	3	Ailanthus Desf.	7	3	Ailanthus integrifolia Lam. ex Steud.	Tw73309, Tw73336, Tw71084, Tw74971	Tw73549, Tw71345, Tw73627, Tw71232, Tw73318
Stemonuraceae	6	2	Stemonurus Blume	6	2	Stemonurus monticola (G.Schellenb.) Sleumer	Tw74382, Tw73716, Tw74594, Tw73014	Tw80936, Tw74212, Tw73005
Styracaceae	3	1	Bruinsmia Boerl. & Koord.	3	1	Bruinsmia styracoides Boerl. & Koord.	Tw72860, Tw74709, Tw74216	Tw74677
Tetramelaceae	15	5	Octomeles Miq.	12	4	Octomeles sumatranum Miq.	Tw72782, Tw71342, Tw72784, Tw72360, Tw72133, Tw73556, Tw72335, Tw74783	Tw72085, Tw73791, Tw71948, Tw73149, Tw72837, Tw73942, Tw72407
			Tetrameles R.Br.	3	1	Tetrameles nudiflora R.Br.	Tw72160, Tw73815, Tw72449	Tw73148
Theaceae	6	3	Polyspora Sweet	6	3	Polyspora amboinensis (Miq.)	Tw74153, Tw74150	Tw74670, Tw73747, Tw73690

						Orel, Peter G.Wilson, Curry & Luu		
						Polyspora papuana (Kobuski) Orel, Peter G.Wilson, Curry & Luu	Tw74402, Tw72492, Tw72231	Tw73570

Table 8.9: Overview of the applied class counts and simulations for the experiment in Chapter 5 on generalisation of object re-identification beyond the learned taxonomic scope.

Number of Classes	Number of Simulations per Class count
10	50
15	30
20	20
30	15
50	10
75	6
100	3

Curriculum Vitae

Ir. RUBEN DE BLAERE

Windekekouter 35, 9860 Oosterzele, Belgium

+32 470 45 51 99 | ruben.de.blaere@telenet.be



PROFESSIONAL PROFILE

PhD researcher with a strong background in bio-engineering sciences and specialization in rapid and accessible wood identification of African timbers. Experienced in computer vision, deep learning, and botanical database management. Passionate about contributing to environmental and biodiversity-related challenges through scientific research, education, and applied technology. Currently working in collaboration between the Royal Museum for Central Africa (RMCA) and UGent-Woodlab on AI-driven wood identification solutions.

WORK EXPERIENCE

PhD Researcher in Fast and Scalable Wood Identification systems
Royal Museum for Central Africa (RMCA) – Department of Wood Biology
July 2021 – Present

- Leading researcher in the SmartWood ID project (BRAIN-be 2.0, Belspo-funded)
- Development of a computer vision-based tool for the identification of Congolese timber using deep learning techniques
- Responsible for digitizing and managing the image database of the renowned Tervuren wood collection
- Authoring scientific publications and presenting research at international conferences
- Providing lectures and workshops on wood anatomy and rapid field identification methods
- Supervising volunteers in the management of botanical collections

Project Assistant – Soil Investigations

AMIBO bv

April 2021 – June 2021

- Conducted and drafted exploratory soil investigations
- Assisted in fieldwork and environmental data collection

Junior Consultant

AGO Jobs

January 2021

- Conducted outreach to candidates regarding interest in administrative positions

Intern – Nature Management

Ecobois nv

August 2019 – September 2019

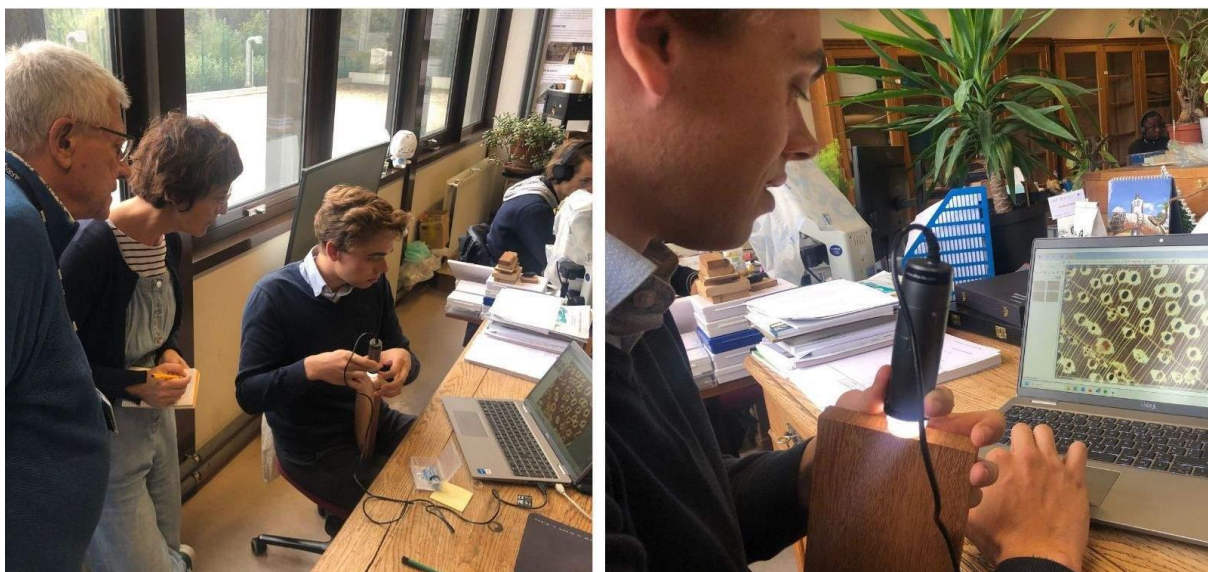
- Drafted a comprehensive nature management plan for a 15-hectare nature domain

EDUCATION

Ghent University (UGent)

Master of Science in Bioscience Engineering: Forest and Nature Management

Graduation Year: 2020



SKILLS & EXPERTISE

- Proficiency in wood anatomy and macroscopic cross-sectional wood identification
- Expert in advanced sanding techniques for unveiling wood anatomy for scientific research
- Programming advanced robotics (Universal Robots, Linux CNC)
- Computer vision and deep learning for material identification
- Scientific research & academic writing
- Advanced programming in python, R, MATLAB
- Botanical database development and management
- Fieldwork techniques in environmental and nature studies
- Supervising and coaching volunteers in collection management
- Public speaking: presenting research, teaching, and giving demonstrations

- Workshop and educational activity development
- Environmental impact assessments and field investigations

LANGUAGES

- Dutch: Native
- English: Fluent
- French: Basic proficiency

Publications:

- First author:
 - R. De Blaere *et al.*, “SmartWoodID—an image collection of large end-grain surfaces to support wood identification systems,” *Database*, vol. 2023, p. baad034, Jan. 2023, doi: 10.1093/database/baad034.
 - R. De Blaere *et al.*, “Evaluating the effect of anomalous images on CV-based wood identification models”, *Wood Science and Technology*; “Accepted for publication”; 19/08/2025
- Co-author
 - J. Van den Bulcke *et al.*, “Enabling high-throughput quantitative wood anatomy through a dedicated pipeline,” *Plant Methods*, vol. 21, no. 1, p. 11, 2025.

Conference proceedings:

- First author:
 - **GAPSYM14: Africa and its ecologies: people and nature in the age of climate change** (2021): Poster (~45 min) — *SmartWoodID: Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation*—Ghent, Belgium (online); https://www.africapplatform.ugent.be/sites/default/files/GAPSYM14_abstract%20book-final_1.pdf
 - **AMBITUS** (2022): Presentation (~15 min)— *Wood identification techniques for combatting illegal logging: Importance, issues, techniques, applications on wood anatomical assessment and the importance of wood collections*—Paris, France; <https://www.facebook.com/ambituseuropa/posts/closing-meeting-of-the-ambitus-project-on-2-and-3-june-in-paris-the-members-of-t/543063587316223/>
 - **Joint 6th Annual Meeting on Plant Ecology and Evolution & COBECORE meeting** (2022): Presentation (~15 min) — *Why digitizing wood collections and how?*—and workshop (~90 min) — *Why digitizing wood collections and how?*—Meise, Belgium; <https://sites.google.com/plantentuinmeise.be/ampee6/>

- **26th IUFRO world congress** (T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification) (2024): Short oral + poster (~5 min))— *Valorizing xylaria using computer vision-based wood identification: a case study on the INERA-Yangambi xylarium*—Stockholm, Sweden; <https://iufro2024.com/book-of-abstracts/> ; p. 3536.
- Co-author:
 - **26th IUFRO world congress** (T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification) (2024): Short oral + poster (~5 min))— *The two-fold role of wood anatomy on a local Congolese timber market: a tool for species level identification and a direct link to the adjacent*— Stockholm, Sweden; <https://iufro2024.com/book-of-abstracts/>
 - **26th IUFRO world congress** (T5.16 IAWA-IUFRO Symposium: Advancing Methods and Applications of Wood Identification) (2024): presentation (~10 min))— *In search of species level identifications for African CITES timber species: potential of combined methods and importance for provenancing*— Stockholm, Sweden; <https://iufro2024.com/book-of-abstracts/>