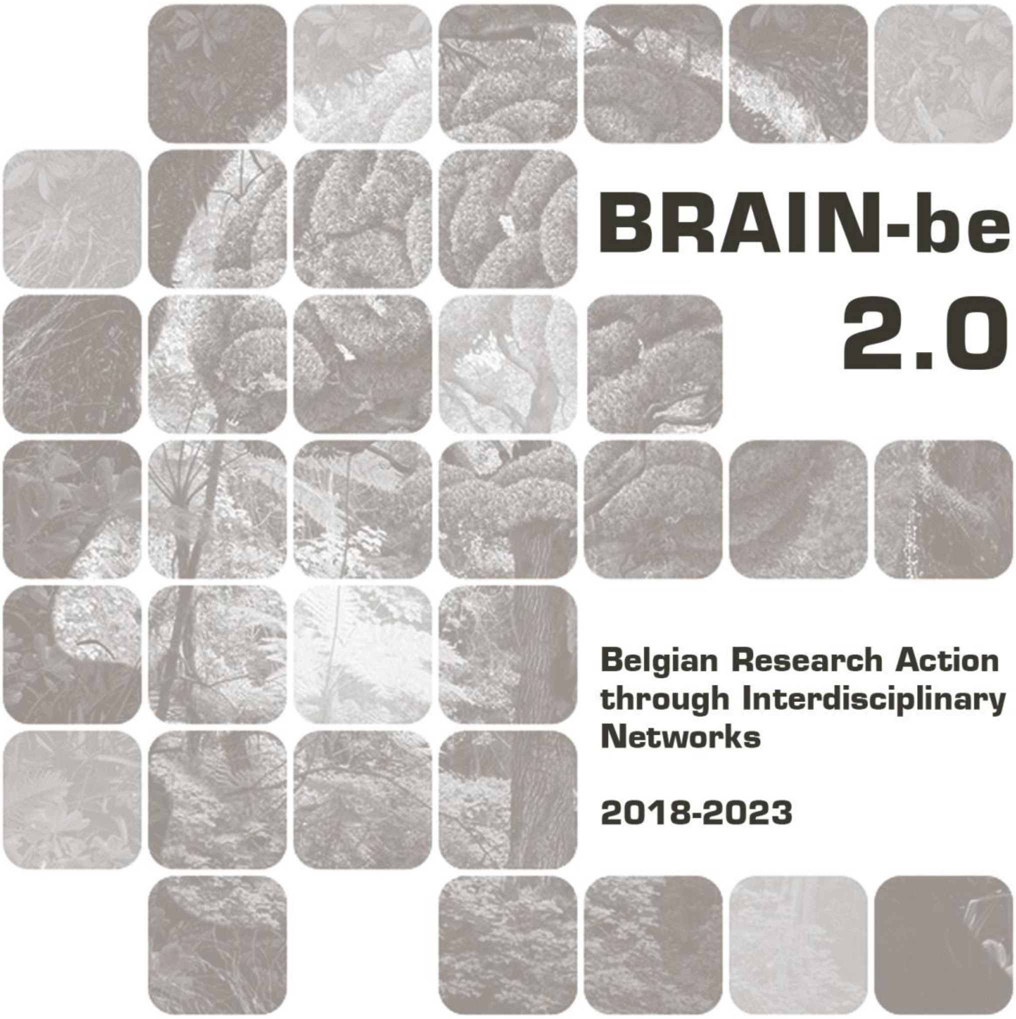

SmartWoodID

Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation

dr. ir. Ruben De Blaere (KMMA/RMAC) - Prof. dr. ir. Wannes Hubau (KMMA/RMAC, UGent) - dr. ir. Hans Beeckman (KMMA/RMAC)

Pillar 2: Heritage science

NETWORK PROJECT

SmartWoodID

Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation

Contract - B2 / 202 / P2 / SmartWoodID

FINAL REPORT

PROMOTORS: dr. ir. Hans Beeckman (KMMA/RMAC)
Prof. dr. ir. Tom De Mil (ULiège)
Prof. dr. ir. Wannes Hubau (KMMA/RMAC, UGent)
Prof. dr. ir. Jan Van den Bulcke (UGent)
Prof. dr. ir. Jan Verwaeren (UGent)

AUTHORS: dr. ir. Ruben De Blaere (KMMA/RMAC)
Prof. dr. ir. Wannes Hubau (KMMA/RMAC, UGent)
dr. ir. Hans Beeckman (KMMA/RMAC)

Published in 2025 by the Belgian Science Policy Office
WTCIII
Simon Bolivarlaan 30 bus 7
Boulevard Simon Bolivar 30 bte 7
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11
http://www.belspo.be
http://www.belspo.be/brain-be

Contact person: Georges JAMART
Tel: +32 2 238 36 90

R. De Blaere, W. Hubau, H. Beeckman. ***Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation***. Final Report. Brussels: Belgian Science Policy Office 2021 – 59 p. (BRAIN-be 2.0 - (Belgian Research Action through Interdisciplinary Networks))

TABLE OF CONTENTS

ABSTRACT

Context

A substantial part of the timber trade is still illegal, and illegal logging is the most profitable biodiversity crime. UN Environment estimates that illegal logging and the associated timber trade counts up to US$50 to $152 billion per year. Illegal logging involves a high risk of irreversible damage to ecosystems associated with the exploitation of highly sought after, sometimes protected, species. Timber regulations are already active (CITES, FLEGT, EUTR), but implementation and enforcement are a challenge. Currently, Belgium has the negative connotation of being the 'hub of illegal timber trade'. 27.5% of the total EU28 imports of primary tropical timber products are imported via Belgium (mainly via the port of Antwerp). Wood identification is a key process in the enforcement that needs to check whether the shipment corresponds with the products mentioned on the accompanying documents. For this reason, there is a growing demand for timber identification tools that can be applied by law enforcement officers.

The Tervuren xylarium is the Belgian governmental collection of wood samples. It is an internationally renowned part of the federal scientific heritage, housed by the Royal Museum for Central Africa and comprises reference material of 13 000 different botanical species. One of the growing actual functions of the collection is supporting forensic research through verification of a species' identity. The most common technique of timber identification is a wood anatomical assessment. Machine Learning methods are likely to be able to assist the wood identification process for non-specialists. Wood species have indeed characteristic features at different microscopical magnifications. However, some of those features are highly variable, which hampers the development of classical dichotomy identification keys that can be used by non-specialists. Moreover, many features seen on wood surfaces are to be understood as artifacts (fissures, traces of mechanical damage, fungi and insect attacks) and are not always easy to distinguish from diagnostic characteristics for the untrained eye. The Tervuren xylarium offers the most complete assemblage of reference material for the development of new wood identification approaches.

The project aimed at automating part of the wood identification process by applying artificial intelligence techniques for the analysis of wood anatomical images of timber species of the Democratic Republic of the Congo. The tree flora of Central Africa comprises 3013 species, 27 of these belong to the class 1 commercial timber species of the DRC and are actually intensively logged and traded, 20 to class 2 (have potentially a big commercial value), 44 to class 3 (are considered to be promoted) and 879 to class 4 (commercial value is not yet known). The project used xylarium samples of all the species of the four classes and took advantage of the power of modern deep learning approaches. The project relied on expert wood anatomical descriptions which served as annotated training data to develop the software. The project was unique because of the large number of African species, the application of deep learning and a database of standardized descriptions that are available online. In a first work package expert annotations of microscopic and mesoscopic images of transverse surfaces of 1000 Congolese wood species were made. In Work package 2 we developed an image processing pipeline for semi-automated annotation of microscopic and mesoscopic wood sections. In Work package 3 we focused on the production of a user-friendly interface.

Objectives

The SmartwoodID project aimed at improving the process of analysis of timber species from the Democratic Republic of the Congo (DRC), a large species-rich country where identification is a daunting task, leading to regular law breaches. The general ambition was to explore the possibilities for automating identification of timber species by applying machine learning, using large datasets of macroscopic features and high-resolution optical scans of endgrain surfaces.

The project had three research objectives:

- **Research objective 1:** Constructing a database of flat-bed scans from end-grain surfaces and macroscopic features of timber species from the Democratic Republic of the Congo.
- **Research objective 2:** Development of illustrated classification keys for wood identification of Congo.
- **Research objective 3:** Developing of an illustrated classification key assisted by machine learning.

Conclusions

The SmartWoodID image database offers 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 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. 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.

The study into identification keys (WP2) 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.

The results on WP3 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. 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. This 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 identification keys on the 31 macroscopic cross-sectional IAWA 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 CNN, object re-identification CNN 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.

The results on an integrated key using CV and the 31 expert-defined features demonstrate 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 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.

Keywords

Wood Anatomy
Wood Identification
Macroscopic wood anatomical assessment
Cross-section
Computer Vision-based wood identification
Object re-identification
Gradient-weighted class activation mapping

1. INTRODUCTION

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 Evironmental 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.

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; 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, 2019, 2022). 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). 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.

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.

Reference list

(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.

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.

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.

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.

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.

CITES, 2022. 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, 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.

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.

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.

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.

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.

Gabris, J.M., 2025. Zero Tolerance for Environmental Crimes. Ecol. Integr. Int. Law Peace, Public Heal. Glob. Secur.

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.

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.

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).

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.

Hoare, A., 2015. Tackling illegal logging and the related trade. What Prog. where next 79.

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 (Fabacaeae). Int. J. Mol. Sci. 21, 3758.

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.

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.

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.

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.

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.

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.

Lawson, S., 2014. Illegal logging in the Democratic Republic of the Congo. Energy, Environ. Resour. EER 2014.

Lewis, S.L., Edwards, D.P., Galbraith, D., 2015. Increasing human dominance of tropical forests. Science (80-.). 349, 827–832.

Lirëza, L., Koçi, G., 2023. Environmental crimes: Their nature, scope, and problems in identification. Interdiscip. J. Res. Dev. 10, 237.

Lowe, A.J., Dormontt, E.E., Bowie, M.J., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoeft, A., Yin, Y., 2016. Opportunities for improved transparency in the timber trade through scientific verification. Bioscience 66, 990–998.

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.

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.

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.

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. ajoutee, GIZ, MEDD, DFS, GFA.

Nellemann, C., INTERPOL Evironmental 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

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.

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., Altstatt, 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.

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.

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.

Stokstad, E., 2014. The empty forest.

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.

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.

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.

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.

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.

White, R., 2018. Transnational environmental crime: Toward an eco-global criminology. Willan.

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.

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).

2. STATE OF THE ART AND OBJECTIVES

State of current knowledge at national and international level

A total of 5 billion m^3 roundwood is harvested annually from the world's forests. A substantial part of timber harvest is illegal, and illegal logging is the most profitable natural resource crime (May, 2017). The associated illegal timber trade represents 50 to 152 billion USD per year, equivalent to 10-30 % of the total (legal + illegal) global timber trade (UNEP & INTERPOL, 2012). Illegal logging involves a high risk of irreversible damage to ecosystems, associated with the exploitation of highly sought after, sometimes protected, species. This is especially pertinent for tropical species, as it is estimated that 30-90% of the tropical timber volume is harvested illegally (Deklerck *et al.*, 2020; Vlam *et al.*, 2018; Hoare, 2015; Hirschberger, 2008). Timber regulations are already active (CITES, FLEGT, EUTR), but implementation and enforcement are a challenge. Currently, Belgium has the negative connotation of being the 'illegal timber trade hub' of Europe. 27.5% of the total European (28 countries) imports of primary tropical timber products are shipped via Belgium (mainly via the port of Antwerp). Wood identification is a key process in enforcement. Law officers need this to check whether the shipment corresponds with the products and species mentioned in the accompanying documents. Hence, there is a growing demand for robust and routine timber identification tools that are easily accessible for law enforcement officers. This societal context primarily determined the actual themes of research groups involved in wood identification.

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.

Various approaches to extract stable diagnostic patterns are being studied (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, 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, enable 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).

Several alternative techniques have been developed in recent decades for identifying 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.

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 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.

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 Höltken et al., 2012). Especially due to heat treatments frequently applied for drying, which degrades DNA (Jiao et al., 2020; Michael Höltken 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, involves 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.

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.

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.

The project and research objectives

The project focuses on the DRC, which comprises more than 3,000 woody species and represents ~5% of the estimated world's tropical tree flora (Sosef *et al*. 2017). 27 of these species belong to the class I commercial timber species of the DRC (*Ministère de l'environnement et de développement durable*, 2017) and are intensively logged and traded; 20 species are in class II (with potentially a large commercial value), 44 are in class III (are considered to be promoted) and 879 are in class IV (sufficiently large trees to produce timber, but timber value is poorly known). As such, the DRC harbors 970 species that are naturally big enough to produce timber. The Tervuren xylarium offers the most complete collection of reference material for the development of wood classification and identification approaches for Congolese species, comprising more than 2000 woody species from the DRC (timber trees, small trees, shrubs, dwarf shrubs and lianas).

The project aims to use the macroscopic features seen on an end-grain (transversal) surface. Correct orientation of a sample is crucial in an assessment of wood anatomical structures. Finding the correctly oriented tangential and radial surfaces is often difficult, especially for enforcement officers who are not trained in the wood anatomical principles. An end-grain surface is much more straightforward to localize, and many of the wood anatomical features observed on this surface are macroscopically observable (i.e. without the need of a microscope). These macroscopic features are also recognizable by computer vision and are expected to allow a relatively fine classification (Ruffinato & Crivellaro, 2019). Screening of macroscopic features on end-grain surfaces is a much faster process than a complete microscopic description and allows constructing large databases in a relatively short time span. Furthermore, the focus on macroscopic end-grain features permits collecting the input data with commonly available equipment such as a flatbed scanner or a smartphone. Existing wood identification and classification tools typically do not take full advantage of the power of modern machine learning techniques. The construction of a large database with expert annotations will provide extensive training data to develop machine learning techniques. The value of illustrated identification keys depends on the completeness of the database. Especially for a species-rich country like the Democratic Republic of the Congo, 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 a maximum of species present in the DRC. A database will be constructed with flatbed scans of large end-grain surfaces and description of macroscopic features. The RGB values of the images, as a proxy of the natural color of the timbers, will be stored as well. Databases and atlases are usually based on a selection of artefact-free images. In contrast, a shipment inspection often encounters images that contain traces of damage inherent to the material. To account for these, a chapter of the database will, for the studied specimens, also contain information on non-anatomical phenomena like cracks, mechanical damage and traces of fungi and insect attacks. Machine learning will be trained to recognize such irregularities.

Research objective 1: Establishing a database of flat-bed scans from end-grain surfaces and text-based feature descriptions

Wood anatomical assessment is the most used, cheapest and most generally applicable method for wood identification. It is possible to identify material that is as diverse as plywood, charcoal and solid wood. We will make use of scientific reference material from the Tervuren xylarium to establish a database on end grain wood anatomical features of nearly 1,000 timber species present in the DRC forests. There are on average five specimens available for a Central African species present in the xylarium resulting in an estimated 5,000 surfaces to be scanned.

End-grain surfaces will be used because non-specialists normally do not encounter difficulties to find this orientation. This way the approach mimics a practical setting where enforcement officers are expected to do a first screening of timber that is being shipped. At the same time, an examination of the end-grain surface is a first logical step in a routine expert analysis of an unknown wood sample in a process aiming at identification of the botanical species.

A rotating machine with gradually finer grain sanding paper mounted on a fixed stage will be used to sand the transverse surfaces of the samples. This will produce perfect flat surfaces that will be scanned at 1,200 dpi, which is a compromise between storing capacity and resolution. A typical image (tiff file of 50 MB) will cover the wood anatomical end-grain structure of a surface of 7 cm long and 1-2 cm large (Fig. 1). The digital images cover more variability compared to sections of the usual size and provide opportunities for substantial data augmentation (i.e. increasing image variability) for deep learning by rotation and cropping. The images also contain quantitative information (RGB values) on the natural color of the wood, which can be approximated by making use of a scanner color calibration tile and the procedure described in León *et al*. (2006). Artefacts like fissures, traces of insect, fungi or mechanical damages are present as well on many samples and will be included in a specimen-based chapter of the database. The resulting images with overlaid grids (Fig. 3) will be used to make text descriptions of macroscopic features (Table 1)

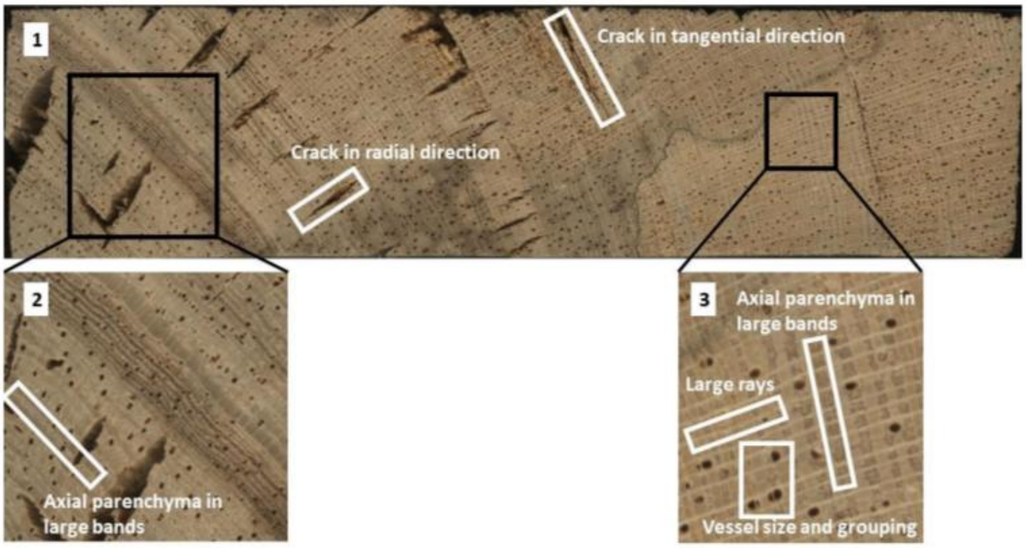

Figure 2: Image (2 cm x 7 cm; scanned at 1200 dpi, inset: digital magnifications) of end-grain surface of Ficus variegata *– Tw71055 . The image shows several large cracks in the tangential and radial direction and traces of fungi development. Macroscopic wood anatomical features can be recognized visually: vessel features (size of vessel elements, grouping), axial parenchyma in large bands, large rays. There are some darker bands within the ground tissue due to thicker walled fibers than the light-colored wood. This is a natural tissue variability potentially hampering a wood anatomical assessment.*

Research objective 2: building Illustrated keys

Based on the database illustrated keys for identification will be developed, at this stage without application of machine learning techniques. This first type of keys is sometimes called a Computer Aided Identification (CAI) system, giving access to necessary resources to relate observations of morpho-anatomical features with names of taxa and information related to these taxa. The Xper² (Ung *et al*., 2010, infosyslab.fr) and Xper³ (Vignes-Lebbe *et al*., 2015) systems are user friendly platforms that manage descriptive data and facilitate identifications through interactive keys. They are constructed around four modules: (1) a module for standardized descriptions of an unknown specimen, (2) an editor with the features of the taxa, including the related images, (3) management tools (e.g. to deal with uncertainties) and (4) a module for creation of interactive keys. Dichotomous keys have been the main tool for practical identification of biological material. Computers have allowed to develop multi-access systems including images and other sources of information. Xper² is a free access software that does not require special computer skills. It is a platform dedicated to taxonomic descriptions and computer-aided identification. It is quite comparable with other existing software like those developed on the DELTA format. It includes an editor to edit taxonomic standardized descriptions and several functionalities to identify specimens, to compare and compute morphological dissimilarities and to import or export data from other formats. Xper² allows developing offline systems, which is still appealing in conditions such as harbors and regions with limited internet connections. Xper³ is a web platform for descriptive data management and interactive identification. Data sharing and interactive identification can be done online.

Table 1: 29 different macroscopic features (Ruffinatto *et al.*, 2015) that can be seen on a high-resolution scan of an end-grain surface.

Structure	Property	Character	Character states	Macroscopic features #
Growth rings	Growth rings	Growth rings distinct	Present / absent / variable	1
Vessels	Porosity	Diffuse porous	Present / absent / variable	3
		Semi-ring porous	Present / absent / variable	4
		Ring porous	Present / absent / variable	5
	Arrangement	Vessels in tangential bands	Present / absent / variable	8
		Vessels in radial pattern	Present / absent / variable	9
		Vessels in diagonal pattern (echelon)	Present / absent / variable	10
		Vessels in dendritic pattern (flame-like)	Present / absent / variable	11
	Grouping	Solitary and in radial multiples of 2-3 vessels	Present / absent / variable	12
		Exclusively solitary (90% or more)	Present / absent / variable	13
		Radial multiples of 4 or more common	Present / absent / variable / NA	14
		Clusters common	Present / absent / variable / NA	15
	Frequency	≤ 5 vessels per square mm	Present / absent / variable	16
		6–20 vessels / square mm	Present / absent / variable	17
		> 20 vessels / square mm	Present / absent / variable	18
	Vessel diameter/ pore visibility	Small (not visible to the naked eye, less than 80µm)	Present / absent / variable	19
		Medium (just visible to the naked eye, 80-130 µm)	Present / absent / variable	20
		Large (commonly visible to the naked eye, larger than 130 µm)	Present / absent / variable	21
Axial parenchyma	Distribution	Diffuse-in-aggregates	Present / absent / variable	30
		Vasicentric	Present / absent / variable	31
		Lozenge-aliform	Present / absent / variable / unilateral	32
		Winged-aliform	Present / absent / variable / unilateral	33
		Confluent	Present / absent / variable / unilateral	34
		Banded	Majority wide / majority narrow / variable / absent	35
		Parenchyma in marginal or seemingly marginal bands	Present / absent / variable	38
		Parenchyma reticulate	Present / absent / variable	39
		Parenchyma scalariform	Present / absent / variable	40
Rays	Width	Ray visibility to the naked eye on the transverse surface	Rays not visible / all rays visible / only larger rays visible	43
	Rays per mm	Rays per mm	≤ 4 mm / 5–12 mm / > 12 mm / NA	49

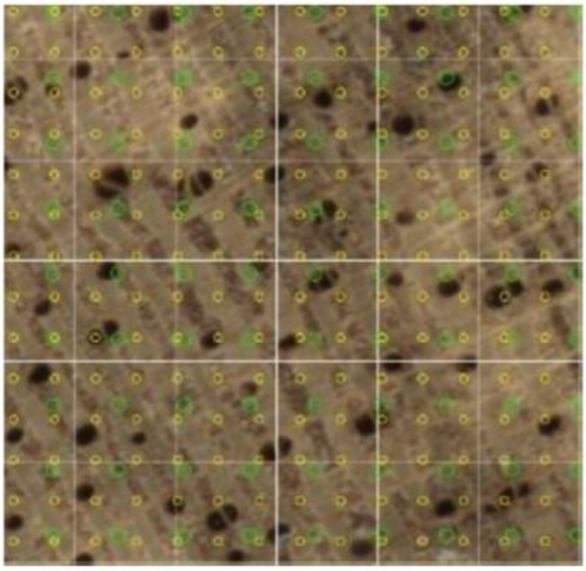

Figure 3: Scanned end-grain surface of Ficus variegata *with grids allowing a visual assessment of macroscopic features. Vessels (black "pores") are counted within the grid of whites 1 mm² squares, their size is compared to the grids with small (diameter of 80 µm) and bigger (diameter of 130 µm) circles, delimiting the size classes as defined by Ruffinatto et al. (2015).*

The database with images and annotations will be implemented in the Xper² system to develop an offline classification system and in Xper³ for an online system.

Research objective 3: Building a machine-learning assisted key

Machine learning will be used at several stages in the development and application of a second type of key: a machine-learning assisted key for wood identification. Machine learning is expected to be able to partly (or for a selected number of species even fully) automate the identification process. In a first step, the identification problem will be understood as a multi-class classification problem where the input is an endgrain image and the output is the taxon name. This is a traditional image recognition problem and (deep) convolutional neural networks (CNN) achieve state-of-the-art performance here. Species with distinct visual features will probably be identifiable by these networks. Therefore, as a first step an image will be fed to this network, and the prediction will be verified for correctness by an operator (by comparing it with an image from the reference library of that species). However, given the large number of species and the within-species variability, a correct identification will not be feasible at this point. Nevertheless, if several species can be detected with a high degree of certainty by the predicted class probabilities of the network, a dedicated, image-specific visual key can be generated on-the-fly that already filters out irrelevant species and simplifies the subsequent identification process. Another way in which machine learning can partly automate the identification process is by detecting specific (anatomical) structures (e.g., the type of porosity) that are used in the decision process of the illustrated keys build on the Xper² and Xper³ platforms. If a specific structure that is associated with a split in the identification key is detected, the split can be made automatically without the need for human interference. On the other hand, if a structure is detected but cannot be reliably categorized, this structure can be highlighted automatically such that the user is aware of its presence.

Machine learning will also be used for assisting the development of a machine-learning assisted key for wood identification. Due to the similarity between dichotomous keys and decision trees from the field of machine learning (such as CART or C4.5), these algorithms can be used to propose a general structure of the key with a minimal number of splits, minimal/maximal depth and constrained width. The decision trees found using these algorithms will be complemented by a series of additional trees that are inferred using constraint programming (Verhaeghe *et al*., 2019) and allow a less greedy and more flexible approach. If the key is used by novices in the field of wood identification, it is very likely that classification mistakes will be a result of incorrectly identified anatomical structures. The latter approach allows, for instance, to postpone the selection of anatomical traits that are hard to observe by a non-expert to deeper splits in the tree. In a second step, data on the (potential) faulty identification of traits will be included in the tree construction process using decision tree inference procedures that can handle uncertainty in the features (Satya Prakash *et al*., 2012).

A first neural network will be trained to detect if any artefacts are present in the images that the user must be aware of, and should not be identified, by mistake, as a relevant anatomical feature. We will therefore use the presence / absence 'tag' of a certain artefact attached to each image to train. If not sufficiently successful, on-image annotations, on a selection of images covering sufficient variability, of the actual artefacts will be performed. Then, we will follow a classical approach, by only using the image itself and train existing network architectures to classify the images. Given the vast number of species, it can be expected that the overall success rate will not be satisfying.

A second network will be trained on the actual anatomical features, using the formal descriptions. If not sufficiently successful, on-image annotations on a selection of images and focusing on the anatomical features that define a critical decision node in the illustrated key, will be used as well. For each of the selected structures (the artefacts and the macroscopic features seen on end-grain surfaces), a separate object detector will be built using deep convolutional neural network (CNN) techniques that are trained on the annotated images. For the image classification tasks pre-trained networks (VGG-16, ResNet-50, Inception-ResNets) will be used as a basis that, relying on the transfer learning paradigm, will be trained using the collected image data (using common data-augmentation strategies). When annotations are available (for instance of defects/cracks, or infestation with a fungus), a (fast) region-based CNN will be used to detect and localize these artifacts in new images and alert the user for closer inspection. These networks will be implemented using the Python library Keras. This means that there is no need to provide a complete segmentation of the image, which can be problematic in the presence of artifacts. The user will be given the opportunity to add or remove annotations and attempts will be made to extract additional descriptors from the partially annotated image, including the spatial arrangement of the identified structures.

Existing software will be used for the annotations to build a classifier per species that can assist in annotation to speed up the process (for instance using WEKA).

Position of the project within the state of the art

SmartwoodID combines human and machine vision. The machine learning method will enhance wood identification success and speed, even for non-experts, by tackling two major problems of classic identification techniques based on visual hand-lens inspection. A first problem is the fact that wood anatomical features can be highly variable, which hampers the development of dichotomic identification keys used for routine identification. Secondly, many features seen on wood surfaces are artefacts (fissures, traces of mechanical damage, fungi and insect attacks) that are not typical for the material and not always easy to distinguish from diagnostic characteristics by the untrained eye. Automated classification techniques can tackle these two difficulties. First results are promising (Hermanson & Wiedenhoeft, 2011; Rosa da Silva *et al.*, 2017; Ravindran *et al.*, 2018; Verly Lopez *et al.*, 2020) and led to the development of a field deployable machine-vision wood identification system for 150 species, called the Xylotron (Forest Products Lab, Madison, USA, see also Ravindran *et al.*, 2019). The Tervuren Xylarium offers the unique opportunity to provide the robust reference database needed for machine learning by valorizing a vast collection of specimens with a focus on the most important Central African timber species.

Opportunities for national and international collaboration

International organizations like CITES and the European commission are asking for more robust and readily available timber identification techniques. International collaboration is possible with institutions like the Wood Identification and Screening Center (U.S. Forest Service International Programs, Corvallis OR, USA), U.S. Fish and Wildlife Forensics Laboratory (Ashland OR, USA). Regular updates will be provided towards the Global Timber Tracking Network, a platform grouping all actors within timber tracking.

Scientific collaboration and exchanges of experiences are possible with research teams working in the same domain and this as well in Europe, the USA, Brazil and Japan.

Reference list

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.

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.

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

Barefoot, A.C., Hankins, F.W., 1982. Identification of modern and tertiary woods. Oxford University Press.

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.

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

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.

Butterfield, B., 2012. Three-dimensional structure of wood: an ultrastructural approach. Springer Science & Business Media.

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.

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.

Chudnoff, M., 1984. Tropical timbers of the world. US Department of Agriculture, Forest Service.

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.

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.

Deklerck, V., 2019. National treasure: valorisation of the Federal Xylarium in Belgium for timber identification and wood technology.

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.

Dierickx, S., Genbrugge, S., Beeckman, H., Hubau, W., Kibleur, P., Van den Bulcke, J., 2024. Non-destructive wood identification using X-ray μCT 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.

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.

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.

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.

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.

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.

Goodell, B., Nielsen, G., 2023. Wood Biodeterioration, in: Springer Handbook of Wood Science and Technology. Springer, pp. 139–177.

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, 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.

Hermanson, J.C., Wiedenhoeft, A.C. (2011). A brief review of machine vision in the context of automated wood identification systems. IAWA Journal, 32(2):230–250.

Hirschberger, P. (2008). Illegal wood for the European market. In An Analysis of the EU Import and Export of Illegal Wood and Related Products; Griesshammer, N., Schulmeister, A., Eds.; WWF: Frankfurt am Main, Germany, p. 45.

Hoare, A., 2015. Tackling illegal logging and the related trade. What Prog. where next 79.

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.

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.

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.

Ilic, J., 1993. Computer aided wood identification using CSIROID. IAWA J. 14, 333–340.

ITTO, 2021. Biennial review and assessment of the world timber situation 2019–2020.

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.

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.

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.

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.

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.

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., 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.

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.

LaPasha, C.A., Wheeler, E.A., 1987. A microcomputer based system for computer-aided wood identification. IAWA J. 8, 347–354.

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.

Lowe, A.J., Dormontt, E.E., Bowie, M.J., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoeft, 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.

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.

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.

May, C. (2017). Transnational Crime and the Developing World.

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.

Ministère de l'environnement et de développement durable. (2017). Guide opérationel. Série Généralité – n°2. Liste des essences forestières de la RDC. DIAF.

Mishler, B.D., 2000. Deep phylogenetic relationships among "plants" and their implications for classification. Taxon 49, 661–683.

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).

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.

NS, I.B., 1989. IAWA list of microscopic features for hardwood identification.

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

Partnership., C.B.F., 2005. The forests of the Congo Basin: a preliminary assessment.

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.

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., 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):1–10.

Ravindran, P., Ebanyenle E., Ebeheaky, A.A., Abban, K.B., Lambog, O., Soares, R., Costa, A., Wiedenhoeft, A.C. (2019). Image Based Identification of Ghanaian Timber Using the XyloTron ; Opportunities, Risks and Challenges. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

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.

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., Wiedenhoeft, A.C., 2015. Review of macroscopic features for hardwood and softwood identification and a proposal for a new character list. IAWA J. 36, 208–241.

Schmidt, O., 2006. Wood and tree fungi. Springer.

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.

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

Sosef, M.S.M., Dauby, G., Blach-Overgaard, A., van der Burgt, X., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield, J., Droissart, V., Duarte, M.C., Engledow, H., Fadeur, G., Figueira, R., Gereau, R.E., Hardy, O.J., Harris, D.J., de Heij, J., Janssens, S., Klomberg, Y., Ley, A.C., Mackinder, B.A., Meerts, P., van de Poel, J.L., Sonké, B., Stévart, T., Stoffelen, P., Svenning, J.-C., Sepulchre, P., Zaiss, R., Wieringa, J.J., Couvreur, T.L.P. (2017). Exploring the floristic diversity of tropical Africa. BMC Biology, 15:15.

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.

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.

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.

Tacconi, L., 2012. Illegal Logging:" Law Enforcement, Livelihoods and the Timber Trade". Routledge.

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.

Tardif, J.C., Conciatori, F., 2015. Microscopic examination of wood: Sample preparation and techniques for light microscopy. Plant Microtech. Protoc. 373–415.

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.

United Nations Environment Programma, INTERPOL. (2013). Green Carbon, Black Trade: illegal logging, tax fraud and laundering in the world's tropical forests, 72p.

University of Wisconsin–Madison, 2025. USDA Forest Service Xylarium [WWW Document]. URL https://herbarium.wisc.edu/collections/wood/ (accessed 3.26.25).

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.

Vander Mijnsbrugge, K., Beeckman, H., 1992. Knowledge modelling for a wood identification system. Silva Gandav. 57.

Verly Lopes, D.J., Burgreen, G.W., Enstminger, E.D. (2020). North American Hardwoods Identification Using Machine-Learning. Forests, 11(3), 298.

Vlam, M.; Arjen de Groot, G.; Boom, A.; Copini, P.; Laros, I.; Veldhuijzen, K.; Zakamdi, D.; Zuidema, P.A. (2018). Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature. Biological conservation, 220, 262–271.

von Arx, G., Crivellaro, A., Prendin, A.L., Čufar, K., Carrer, M., 2016. Quantitative Wood Anatomy—Practical Guidelines. Front. Plant Sci. 7.

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.

Wiedenhoeft, A.C., 2020. The XyloPhone: toward democratizing access to high-quality macroscopic imaging for wood and other substrates. Iawa J. 41, 699–719.

Wilber, M.J., Scheirer, W.J., Leitner, P., Heflin, B., Zott, J., Reinke, D., Delaney, D.K., Boult, 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.

Wodzicki, T.J., 2001. Natural factors affecting wood structure. Wood Sci. Technol. 35, 5–26.

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.

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.

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.

3. METHODOLOGY

The methodological approach of SmartWoodID consists of a study of the wood anatomical structure of tropical timber species, including the construction of a feature database, the development of illustrated keys for classification and applying machine learning techniques.

Wood anatomy

Some of the biological and chemical characteristics of lignified material are typical for the botanical clade and they can be used accordingly to identify the taxon. The biological characteristics allowing taxon identification are summarized in a list of wood anatomical features. The successful application of identification techniques relies on extensive databases of trustworthy reference material. Such scientific reference material is typically present in a xylarium curated within a governmental institute of natural history research.

A wood anatomical assessment is a visual method most often aided by a range of levels of magnification using different techniques (hand lenses, optical and electronic microscopes, X-ray computed tomography). The structure of cells and tissues are being visualized and coded often with the help of a list of standardized and formal features established by the International Association of Wood Anatomists (IAWA), an expert organization existing since the 1930s. There are lists of macroscopic features (Ruffinato & Crivellaro, 2019), hardwood features (IAWA Committee, 1989), softwood features(IAWA Committee, 2004 & 1989), bark features (Angyalossy *et al.*, 2016), and vessel features (Hemling *et al.*, 2018). Optical microscopic assessments can be done by incident light equipment, but most often transmitted light is used at different magnifications. Therefore, microtomic sections in the principal directions (transverse, tangential and radial) need to be made. Most xylaria curate permanent slide collections. Some of the observations are ideally made on macerations where the individual cells can be seen dissociated from their tissue context. Wood anatomical species descriptions based on the standardized IAWA terminology are stored in wide-ranging databases like the online database InsideWood (Wheeler, 2011), a database of hardwood and softwood species, both living and fossil, that contains next to the feature codes also microscopic images. The InsideWood database has evolved to a very complete information source of a wide range of species and is as such an indispensable means in identification work. A very robust identification key for commercial timbers is the Delta-Intkey system (Richter and Dallwitz, 2000 onwards). The classical wood anatomical assessment is based on a good expert knowledge of the wood anatomical features which often hampers the use of the methodology by untrained staff. For this reason, the concept of visual keys has been introduced (Kirchoff *et al.*, 2011), where extensive databases of reference images are constructed that are systematically be compared with images from unknown samples. These manual identification approaches are expected to be gradually complemented or even replaced by computer vision to speed up the process and exclude a certain degree of subjectivism (Hermanson & Wiedenhoeft, 2011).

A database will be constructed with standardized descriptions of the visible features making use of the macroscopic feature list published by Ruffinato & Crivellaro (2019). The approach will result in a classification of the material in species groups. A positive identification of the botanical species will be rather exceptional because a tropical tree flora typically consists of several species with similar wood anatomy belonging to the same botanical genus. For species identification, other techniques need to be applied, such as observations at higher magnification on longitudinal thin sections and chemical analysis of the metabolites.

Machine learning

Over the last decades, machine learning has gained a lot of interest as a tool to extract relevant information. Machine learning is a subfield of computer science that uses computational and statistical models to automatically search for patterns present in the data without being explicitly programmed. This process is often called "learning", and its success is determined by the training on reference data. Within the field of machine learning, we will also explore the potential of deep learning networks. Deep learning is a subfield of machine learning known for its autonomous pattern recognition. Machine learning and specifically deep learning can have a high predictive power, if enough data are provided to train the network properly (Liu *et al*., 2017; Kiranyaz *et al*., 2019). Deep learning methods are gaining interest because of their excellent performance in many computer vision and speech processing applications.

The wood anatomical approach will result in information-rich images and formal descriptions of features, representing a large amount of data. Both the images and formal descriptions will be used to train machine learning algorithms (including deep learning networks).

Rationale w.r.t. literature and state-of-the-art

The large economic value of wood stimulated the development of identification techniques using machine-learning-based image analysis. Semi-automated annotation attempts often rely on general purpose software tools (pixel classifiers such as Fiji with Weka-plugin (Arganda-Carreras *et al*., 2017), or Ilastik (Sommer *et al*., 2011)) that are not tailored to the analysis of wood sections and do not take advantage of the power of modern deep learning approaches. More recently, several studies used deep learning approaches for wood identification (Ravindran *et al*., 2018; Verly Lopes *et al*. , 2020)). However, the common approach in these works does not rely on training data that has been annotated and therefore only indirectly allows to identify the diagnostic characteristics. It has been shown, however, that a reliable identification of some botanical characteristics can be achieved using deep learning approaches (Meeus *et al*., 2019). The proposed approach builds upon the latter, where advanced deep learning models are used to learn from annotated data to provide more reliable and explainable identifications. A successful application of deep learning depends on an image database of reference material that captures enough of the biological variability required to successfully carry out classification. Martins *et al*. (2013) built a database containing 2,240 microscopic images, covering 112 species (20 images per class with each class corresponding to a species). Paula Filho *et a*l. (2014) did the same but for macroscopic images. Their dataset consisted of 2,942 samples from 41 Brazilian species. Hafemann *et al*. (2014) used the dataset collected by Paula Filho *et al*. (2014) and reached an accuracy of 95.77 %. Figueroa-Mata et al. (2018) transformed the images and obtained a dataset of 57,024 images which resulted eventually in 98.03 % accuracy. Yadav *et al*. (2015) selected 1,500 images covering 75 hardwood species from the dataset of Martins *et al*. (2013), they considered several models and obtained performance of 97.67 % accuracy for grayscale and 98.40 % for RGB images. Ravindran *et al*. (2018) tried to classify ten Meliaceae species from six genera using a pretrained CNN. At species level, accuracies of 88.7 % were obtained and 97.5 % at genus level. *Cedrela* species could not be discriminated, and *Khaya* species were classified in wrong genera. Machine learning is promising as a screening tool to wood classification, but still needs an expert opinion for final conclusion. The expert should either conclude whether species cannot be discriminated based on their wood anatomy or an identification has not been possible because the database did not cover the complete biological variability which can be substantial as is the case for the Khaya genus for example.

The SmartwoodID methodological approach (Figure 1) is innovative due to the use of long transverse surfaces. These cover a higher variability compared to standardized sectioning blocks. In addition, the constructed database, with a focus on a species-rich region with major import to other countries, will include information on artefacts in the transverse surfaces which can then be filtered out. The combination with deep learning will allow us to present a machine-learning assistant, a new and robust timber classification technique based on quality reference data coming from the Tervuren Xylarium.

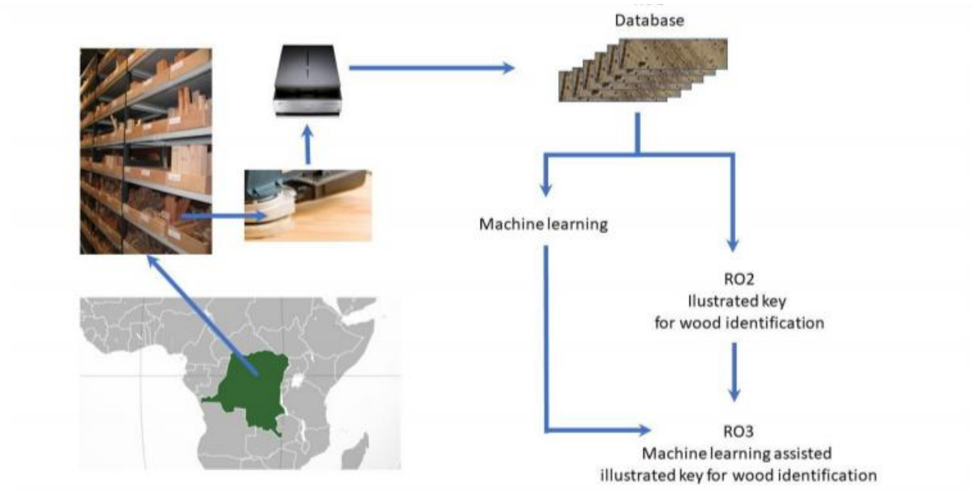

Figure 1: SmartwoodID flow chart. ~5,000 specimens of ~1,000 timber species from the Democratic Republic of the Congo, curated at the Tervuren xylarium, will be polished and scanned with a flatbed scanner. A database (Research objective 1 - RO1) will be constructed with images and text-based descriptions of the macroscopic features and information on artefacts visible on end-grain surfaces. Illustrated keys will be constructed with commonly used software for taxon identification (RO2). The image and artefacts chapters of the database will be used for machine learning to verify whether the identification process can be automated, possibly for a selected number of species. Machine learning will also be used to assist the development of advanced identification keys (RO3).

Using a multi-step approach, the user of the illustrated key will be further supported by the machine learning assistant to separate artefacts and wood anatomical diagnostic features and will help the actual identification. Contrary to existing approaches, both macroscopic features as well as on-image annotations will be used to train machine learning algorithms at key positions in the illustrated key.

Workpackages

WP1: preparing, digitizing and annotating the material (research objective 1): C, P3

Task 1.1 *Preparation and scanning of the material: C*

We adapted a technique developed by Cerre (2016) to prepare high quality end-grain surfaces of large quantities of wood samples. A rotating sanding machine mounted on a fixed stage will be used with gradually finer grain abrasive paper. This will result in perfectly flat surfaces that will allow to be scanned with a flatbed scanner at 1200 dpi. A typical image will cover the wood anatomical end-grain structure of a surface of 7 cm long and 1 cm large, but some of the xylarium specimens have smaller or larger surfaces. These digital images cover more within-tree variability compared to sections usually used in identification which are typically 1 cm² large or less. The scans also contain quantitative information (RGB values) on the natural color of the wood.

Deliverables: 5000 wood samples with polished surfaces (1.1.1.), 5000 digital images obtained through scanning (1.1.2.).

Task 1.2. *Description of the visual features: annotation of the images: C, P3*

The descriptions of the samples will be based on the list of macroscopic features (Ruffinatto *et al*. 2015) and the RGB values of the images. 31 of the standardized features are visible on a typical high-resolution scan. Artefacts due to biological or mechanical impacts that do not have a diagnostic value will also be coded.

Deliverable: A specimen-based database with the collected observations (1.2.1.).

WP2: Development of illustrated keys (research objective 2): C, P3

Task 2.1. *Implementation of text-based and illustrated classification keys: C, P2, P3*

A database with predominantly microscopic images of 60 commercially important species and associated visual identification keys are in use by the RMCA, especially in the context of training of Central African scientists. The platform used is Xper² because this allows easily to create offline systems. SmartwoodID will add more species to the existing system. A parallel system will be developed where text-based characters are maximally being replaced by images of these characters. This will enable a direct comparison of images of unknown wood with reference images collected in Task 1. Furthermore, we aim at building an online Xper³ system and the development of interactive teaching software for Central African wood species, which will aid experts as well as students of wood anatomy.

Deliverable: Illustrated classification keys (2.1.1.) and interactive teaching software (2.1.2.).

WP3: Applying machine learning techniques: P4, P5

Task 3.1. *Using machine learning on image material: P5*

This task consists of applying machine learning tools on the available images, without the textbased descriptions of the wood anatomical features.

Deliverable: list of species, artefacts and features automatically recognizable (3.1.1.).

Task 3.2. *Implementation of a machine learning assistant identification system (research objective 3): P4*
At this stage the machine learning will be applied on the complete database consisting of as well images as text-based descriptions of the human observations of the end-grain surfaces.

Deliverable: user-friendly interface for wood identification (3.2.1.).

WP4: Coordination, project management and reporting: C
The project partners together with the hired staff will meet on a regular base (once a month) to discuss and plan the practical work and scientific progress and make maximally use of the multiple valorization opportunities. Once a year the project team will report to and meet with the followup committee.

Deliverables: coordinated project team (4.1.1.), administration and budget followed-up (4.2.1.), project reports (4.3.1.).

WP5: Data management: P3, C
The collected images and the associated text-based descriptions will be incorporated into the online database of the Tervuren xylarium (http://xylarium.africamuseum.be).

A selection of the images (species that are not yet covered by InsideWood) will be transferred to the North Carolina State University to be uploaded into the InsideWood database (https://insidewood.lib.ncsu.edu/search?0).

The interactive identification keys will be linked to the database (https://congobasincarbon.africamuseum.be/): Xper² as downloadable package, Xper³ as online system.

Deliverables: database of project results (5.1.1.), uploaded data in the Tervuren xylarium database (5.2.1., selection of images in InsideWood (5.2.2.).

WP6: Valorization, dissemination, exploitation of results: C, P1, P2, P3, P4, P5
Project results will be valorized in as well scientific contexts as on forums where mainly policy items are being discussed. We expect to be able to produce an image atlas covering the timber species of the DRC (possible editor European Journal of Taxonomy or a printed book edited by the Royal Museum for Central Africa). We plan to write a research paper on the application of deep learning techniques on large end-grain surfaces of tropical timbers. We plan a presentation of project results at scientific fora convened by the International Association of Wood Anatomists.

Other fora include:

- Presentation annual Forest Legality Week at the World Resources Institute (Washington D.C.; USA)

- Presentation of the classification keys for master students in DRC, Gembloux and Gent

- Presentation of the classification keys at the ATIBT (https://www.atibt.org/en/

- See valorization plans (4.3)

Deliverables: picture atlas (6.1.1.), paper on keys (6.1.2.), group training in DRC (6.2.1.), training of Belgian students (6.2.2.)

Reference list

Angyalossy, V., Pace, M.R., Evert, R.F., Marcati, C.R., Oskolski, A.A., Terrazas, T., Kotina, E., Lens, F., Mazzoni-Viveiros, S.C., Angeles, G., Machado, S.R., Crivellaro, A., Rao, K.S., Junikka, L., Nikolaeva, N., Baas, P. (2016). IAWA List of Microscopic Bark Features. IAWA Journal, 37(4):517-615.

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S. (2017). Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification. Bioinformatics, Aug1:33(15):2424-2426.

Cerre, J.-C. (2016). Incident light microphotography at high depth of focus. IAWA Journal, 37(3):506-510.

Figueroa-Mata, G., Mata-Montero, E., Valverde-Otárola, J.C., Arias-Aguilar, D. (2018). Automated Image-based Identification of Forest Species: Challenges and Opportunities for 21st Century Xylotheques. IEEE International Work Conference on Bioinspired Intelligence (IWOBI), San Carlos, p. 1-8.

Hafemann, L.G., Oliveira, L.S., Cavalin, P. (2014). Forest Species Recognition Using Deep Convolutional Neural Networks. 22nd International Conference on Pattern Recognition, Stockholm, p. 1103-1107.

Helming, S., Olbrich, A., Heinz, I., Koch, G. (2018). Atlas of Vessel Elements: Identification of Asian Timbers. IAWA Journal, 39(3), 249-352.

Hermanson, J.C., Wiedenhoeft, A.C. (2011). A brief review of machine vision in the context of automated wood identification systems. IAWA Journal, 32(2):230–250.

IAWA Committee (Richter, H.G., Grosser, D., Heinz I. & Gasson, P., Eds). (2004). IAWA list of Microscopic Features for Softwood Identification. IAWA Journal, 25: 1-70.

IAWA Committee (Wheeler, E.A., Baas, P. & Gasson, P., Eds). (1989). IAWA list of Microscopic Features for Hardwood Identification. IAWA Journal, 10: 219-332.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J. 1D Convolutional Neural Networks and Applications: A Survey. Electrical Engineering and Systems Science, 9th May, p. 20.

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.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234:11-26.

Martins, J., Oliveira, L.S., Nisgoski, S., Sabourin, R. (2013). A database for automatic classification of forest species. Machine Vision and Applications, 24(3):567-578.

Meeus, L., Huang, S., Devolder, B., Dubois, H., Martens, M., Pižurica, A. (2019). Deep Learning for Paint Loss Detection with a multiscale, translation invariant network. 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, p. 158-162.

Paula Filho, .L.P., Oliveira, L.S., Nisgoski, S., Britto Jr., A.S. (2014). Forest species recognition using macroscopic images. Machine Vision and Applications 25, 1019–1031.

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):1–10.

Richter, H.G., Dallwitz, M J. (2000 onwards). Commercial timbers: descriptions, illustrations, identification, and information retrieval. In English, French, German, Portugese, and Spanish, version: 9th April 2019. delta-intkey.com'.

Ruffinato, F., Crivellaro, A. (2019). Atlas of Macroscopic Wood Identification, with a special focus on timbers used in Europe and CITES-listed species. ISBN: 978-3-030-23565-9, Springer Nature Switzerland.

Sommer, C., Straehle, C., Köthe, U., Hamprecht, F.A. (2011). Ilastik: Interactive learning and segmentation toolkit. IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, p. 230-233.

Verly Lopes, D.J., Burgreen, G.W., Enstminger, E.D. (2020). North American Hardwoods Identification Using Machine-Learning. Forests, 11(3), 298.

Wheeler, E.A. (2011). InsideWood – A Web resource for hardwood anatomy, 32(2):199–211.

Yadav, A.R., Anand, R.S., Dewal, M.L., Gupta, S. (2015). Multiresolution local binary pattern variants based texture feature extraction techniques for efficient classification of microscopic images of hardwood species. Applied Soft Computing, 32:101-112.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

WP1: preparing, digitizing, and annotating the material (research objective 1):

Task 1.1 & 1.2

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. 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 was published in Oxford Academic in 2023.

Species selection:

- A selection was made and 954 timber species from the DRC were selected to be part of the core collection.

General progress (1.1.1 / 1.1.2 / 1.2.1):

- Core collection:
  - Sanding 100% finished (1.1.1)
  - Scanning 100% finished (1.1.2)
  - Annotations 100% finished (1.2.1)
  - Density measurements 100% finished (1.2.1)
  - Pictures 100% finished (1.2.1)
- Extended collection:
  - Sanding 100% finished (1.1.1)
  - Scanning 100% finished (1.1.2)
  - Annotations 100% finished (1.2.1)
  - Density measurements 100% finished (1.2.1)
  - Pictures 100% finished (1.2.1)

WP2: Development of illustrated keys (research objective 2):

Task 2.1. Implementation of text-based and illustrated classification keys:

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.

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. The identification key has been built on 954 Congolese timbers using Xper2. (2.1.1 / 2.1.2)

WP3: Applying machine learning techniques:

Task 3.1. Using machine learning on image material:

To investigate whether visual information not captured by standard descriptors could improve identification, CNN models trained on raw cross-sectional images were explored. 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. (3.1.1)

It was investigated 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 (3.1.1).

Critical factors in building effective training databases were determined. 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. (3.1.1)

Furthermore, we 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. (3.1.1)

Task 3.2. Implementation of a machine learning assistant identification system (research objective 3):
Recognizing that real-world identifications often integrate anatomical descriptors and visual impressions, it was 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 code was made for enabling automated identification with a baseline graphical user interface. (3.2.1).

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. (3.2.1).

WP4: Coordination, project management and reporting:
The project coordinator met with the hired staff on a regular basis (once a week) to discuss and plan the practical work and scientific progress and make maximally use of the multiple valorisation opportunities.

The weekly meetings with the staff and the meetings between partners on a regular basis provide a coordinated project team (4.1.1.). A report was written after every meeting with partners in order to ensure good follow-up on the topics discussed (4.3.1.). Administration and budget followed-up was performed by the project coordinator and partner 4 (4.2.1.).

WP5: Data management
Digitizing the database for public display (1.2.1 / 5.1.1):

All scans in the SmartWoodID collection and metadata (country of origin, taxonomy, etc.) are available on a private link, where they can be used in an IIIF environment (Database URL: https://hdl.handle.net/20.500.12624/SmartWoodID_first_edition). (5.2.1)

A selection of the images (species that are not yet covered by InsideWood) were be transferred to the North Carolina State University to be uploaded into the InsideWood database (https://insidewood.lib.ncsu.edu/search?0). This consists of 1908 specimens and encompasses

781 unique wood species (from which 342 wood species were not yet published on InsideWood (anno 2022)). (5.2.2)

WP6: valorisation, dissemination, exploitation of results:
The picture atlas is being developed in collaboration with Gembloux Agro-Bio Tech (Université de Liège), Passage des Déportés 2, 5030 Gembloux, Belgium. This will be published in the course of 2026 (6.1.1).

A chapter was dedicated in the PhD dissertation (defended and successfully obtained on 2/09/25) to the implementation of identification keys. This will be submitted to a peer-reviewed journal in autumn 2025. (6.1.2)

A woodshop is being constructed in partner research centres in Luki & Yangambi. Local samples will be processed in the same manner as the core and extended collections, to further complete the database of the core collection. A group training in DRC has taken place for two weeks from 21/02/2022 till 04/03/2022. The session focused on task 1.1 to convey the method of sample preparation, sanding a scanning to the researchers at Yangambi research centre in DRC. The group training in DRC (from 21/02/2022 till 04/03/2022) was considered successful as the local researchers and staff were able to produce polished endgrain surfaces of samples of high quality. (6.2.1)

Annual lectures are given to UGent students for the Wood Anatomy course to introduce them to the concept of wood identification keys using macroscopic cross-sectional feature (2022/2023/2024) (6.2.2).

5. DISSEMINATION AND VALORISATION

PARTICIPATION/ORGANISATION OF SEMINARS (NATIONAL/INTERNATIONAL)

A poster was presented at the online international conference Gapsym14: Africa and its ecologies: people and nature in the age of climate change organized in November 2021 by Ruben De Blaere. The title of the paper was:" Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation –Why digitizing wood collections and how."

A presentation was given at the international conference Closing meeting of the Ambitus project in June 2022 in Paris, France by Ruben De Blaere. The title of the presentation was: "Wood identification techniques for combatting illegal logging: Importance, issues, techniques, applications on wood anatomical assessment and the importance of wood collections"

A presentation was given at the 6th Annual Meeting on Plant Ecology and Evolution & COBECORE meeting

29th September 2022 in September 2022 in the Meise Botanic Garden, Belgium by Ruben De Blaere. The title of the presentation was: "Wood identification techniques for combatting illegal logging: Importance, issues, techniques, applications on wood anatomical assessment and the importance of wood collections."

A joint workshop with Kévin Lievens was given the following day at interested researchers in the field of Plant Sciences. The participants were taught how to use the classification key to distinguish African timber genera using macroscopic features on the cross section. In addition, a presentation and demonstration were given on a prototype CNN, capable of classifying African timbers on the genus level. (6.2.2)

SmartWoodID was presented with live demonstration in the Africamuseum on 17/09/2023 in light of September science month by employee Ruben De Blaere.

SmartWoodID was presented with live demonstration in the Africamuseum on 1/10/2023 in light of OpenBedrijvendag by employee Ruben De Blaere along with private tours in the Tervuren wood collection.

Employee Ruben De Blaere attended the online Joint Research Discussion on Developing Integrated Timber Data for Xylaria Networking to stay up-to-date on international collaborations concerning new developments for forensic timber research, digitisation of xylaria, data sharing, and applying deep learning on such data. The meeting was held on the 7/12/2024.

A short oral by poster will be presented at the 26th IUFRO world congress in Stockholm in June 2024 by Ruben De Blaere. The title of the poster is Valorizing xylaria using computer vision-based wood identification: a case study on the INERA-Yangambi xylarium.

Collaborations

Perspectives for collaborations are considered whenever opportunities arise and are explored if the common goals are compatible with the workload and do not compromise the objectives of this project.

1. Collaborations were attempted with a company specializing in building apps capable of recognizing wood species, Agritix (Compatibility with deliverable 3.2.1). A small number of scans containing Asian timbers has been shared and researched by them, with a success in recognition of the species. Further collaboration was halted when the workload and pay-off were out of balance.
2. Collaborations with University of Antananarivo Madagascar were initiated to explore the potential of distinguishing Malagasy rosewood, palisander and ebony timber on species level with look-a-like timbers (3.1.1). Further cooperation is pending, until concrete plans of action and data sharing are determined.
3. A collaboration was initiated with the São Carlos Institute of Physics at the University of São Paulo, Brazil to develop and study the potential of convolutional neural networks in recognizing wood anatomical features (3.1.1). The study will aim to develop one (or multiple) model(s) that incorporates multi-task learning to simultaneously extract features and classify the timbers based on the annotations and scans of the SmartWoodID core and extended collection (2.1.2). This collaboration shall commence from April 2024 and has as a distinct target to be published in co-authorship (6.1.2).

Volunteers

The Royal Museum for Central Africa (RMCA) engaged a broad audience in the valorisation of the wood biology collections through physical Citizen Science (CS) investigation actions. Considering this, the SmartWoodID collection was digitized using fast and accessible tools to simultaneously:

- Test the speed and ease of handling for non-experts (6.2.1/6.2.2)
- Create a reference image collection for publication in the Tervuren wood collection database (5.2.1), InsideWood database (5.2.2), image material for classification keys (2.1.1) and a picture atlas on Congolese timbers (6.1.1).
- Foresee future studies and publications on higher resolution images then the current scans of the SmartWoodID collection, to develop and embed A.I. into custom built software that can automatically quantify wood anatomical features beyond the boundaries in the standardized features (3.1.1).

The *Citizen Rescuers for Collections* (CRESCO) initiative was started up for this reason at the RMCA and employed 5 volunteers, that each performed half-a-day of work per week. The volunteers have signed a contract for 2024 to continue digitisation and aid in the annotation of wood anatomical features to develop deep learning segmentation models, that can automatically detect wood tissue. Those models will be used to study if quantification, beyond the current scope of the IAWA features, allows for a finer identification then possible with the main wood description, as created in this project. This approach will be submitted for publication in the beginning of 2025 with employee Ruben De Blaere as first author, and volunteers and the initiative in acknowledgments (3.1.1/6.).

Internships

Employee Ruben De Blaere followed a five-day internship at the Royal Botanic Gardens, Kew, as a scientist studying the process of DART TOFMS to identify wood from 13/02/2023 till 17/02/2023. Specifically, focus was put on identifying the botanical taxa of Pterocarpus and Afzelia, two protected African timbers. The goal was to research state-of-the-art wood identification techniques to help build a forensic center for wood identification to help combat illegal logging. During the internship, my responsibility was to assist in the wood identification process using DART TOFMS and XYLOTRON techniques. Research was performed on wood specimens of the SmartWoodID collection. The second technique, DART TOFMS, uses mass spectrometry to create chemical fingerprints and can be applied to distinguish timbers. Analysis techniques were taught to identify species by comparing to the reference database of Ed Espinoza and building our own models using random forests and KDA. In conclusion, the internship at the Royal Botanic Gardens, Kew, was a valuable experience, providing technical skills and personal growth experiences. The research conducted on state-of-the-art wood identification techniques will be useful in building a forensic center for wood identification to help combat illegal logging.

Funding was attained through the 4th and final call of the Synthesys+ grant to follow an internship at the Jodrell Laboratory within the Royal Botanic Gardens, Kew from 5/06/2023 till 23/06/2023. It provided the opportunity to contribute to the development of a timber identification key for the Democratic Republic of the Congo (DRC) (2.1.1). Under the guidance of experts Dr. Peter Gasson and Dr. Victor Declerck, work focused on the Economic Botany Collection and World Forest ID Collection. Responsibilities concerned:

- Integration of data from InsideWood and SmartWoodID to create a comprehensive timber identification key, emphasizing standardized IAWA features and macroscopic end-grain features.
- Digitization of specimens in the Economic botany collection, absent from InsideWood and SmartWoodID at the Royal Botanic Gardens, Kew, using methods such as sanding specimens, capturing images with a handheld microscope, and describing macroscopic wood anatomy on the cross-section.

Preliminary conclusions indicated a potential distinguishability of 426 Congolese timber genera using InsideWood with macroscopic end-grain features. Discussions with caretakers and promoters were deemed necessary for validation. This internship at the Jodrell Laboratory has provided a diverse range of experiences, from hands-on wood anatomy work to database integration and digitization. The mission, focused on building a timber identification key for the DRC, has unveiled both challenges and promising preliminary results. The journey has emphasized the complexity of botanical research and the need for collaboration and continuous refinement.

A visit was arranged on 31/01/2024 to the Naturalis Biodiversity centre as a means of networking and sharing intermediary results. The visit featured presentations from staff at the research department of functional traits, a guided tour throughout the collections and laboratories and a visit to the Leiden Natural History Museum.

Courses

Employee Ruben De Blaere followed a Doctoral School Course on Personal Effectiveness to improve efficacy and efficiency further still on 27th and 28th of February 2024.

Employee Ruben De Blaere follows a Doctoral School Course on Advanced Academic English: Conference Skills to improve presenting and effective slide design during the first half of 2024.

Phd dissertation

Employee Ruben De Blaere has successfully defended his dissertation to obtain the titles of Doctor of Bio-science engineering: Natural resources, and PhD on 2/09/2025 at Ghent University. This degree corresponds to level 8 of the Flemish Qualifications Structure, as laid down in the Decree of 30/04/2009 concerning the Qualifications Structure, and to level 8 of the European Qualifications Framework for Lifelong Learning. The dissertation was written on the findings and contents of this project.

6. PUBLICATIONS

Publications

- 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
- R. De Blaere, "Identification of Congolese wood species by use of anatomical images and artificial intelligence", ISBN: 9789463579025, PhD dissertation, defended on 2/09/2025
- 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

- **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.africaplatform.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.
- **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/

7. 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. Describing wood anatomical features was aided by job students Michael Monnoye, Miro Cnops, Senne Suykerbuyk, and volunteers Tibo Deckers and Michèle Florquin. Further acknowledgements are given to Cécile De Troyer, Frank Simoens, Ferre Caron, Olayemi Razaq Saliu, Véronique Maes, Philippe Quintin, Patrick De Snijder, and José Kempenaers for aiding in digitisation as volunteers. Furthermore, Volunteering was enabled by the *Citizen science* initiatives managed by Luiza Mitrache.