
DEEP LEARNING METHODOLOGY FOR THE IDENTIFICATION OF WOOD SPECIES USING HIGH-RESOLUTION MACROSCOPIC IMAGES

David Herrera-Poyatos ^{*}, Andrés Herrera-Poyatos ^{*}, Rosana Montes ^{*}, Paloma de Palacios [†], Luis G. Esteban [†], Alberto García Iruela [†], Francisco García Fernández [†], Francisco Herrera ^{*}

^{*} Andalusian Institute of Data Science and Computational Intelligence (DaSCI), University of Granada, Spain.
Emails: {divadhp, andreshp, rosana}@ugr.es, herrera@decsai.ugr.es

[†] Escuela Técnica Superior de Ingeniería de Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid.
Emails: {paloma.depalacios, a.garciai, luis.garcia, francisco.garcia}@upm.es

June 18, 2024

ABSTRACT

Significant advancements in the field of wood species identification are needed worldwide to support sustainable timber trade. In this work we contribute to automate the identification of wood species via high-resolution macroscopic images of timber. The main challenge of this problem is that fine-grained patterns in timber are crucial in order to accurately identify wood species, and these patterns are not properly learned by traditional convolutional neural networks (CNNs) trained on low/medium resolution images.

We propose a Timber Deep Learning Identification with Patch-based Inference Voting methodology, abbreviated TDLI-PIV methodology. Our proposal exploits the concept of patching and the availability of high-resolution macroscopic images of timber in order to overcome the inherent challenges that CNNs face in timber identification. The TDLI-PIV methodology is able to capture fine-grained patterns in timber and, moreover, boosts robustness and prediction accuracy via a collaborative voting inference process.

In this work we also introduce a new data set of macroscopic images of timber, called GOIMAI-Phase-I, which has been obtained using optical magnification in order to capture fine-grained details, which contrasts to the other datasets that are publicly available. More concretely, images in GOIMAI-Phase-I are taken with a smartphone with a 24x magnifying lens attached to the camera. Our data set contains 2120 images of timber and covers 37 legally protected wood species.

Our experiments have assessed the performance of the TDLI-PIV methodology, involving the comparison with other methodologies available in the literature, exploration of data augmentation methods and the effect that the dataset size has on the accuracy of TDLI-PIV.

Keywords

wood species identification; deep learning; convolutional neural networks; patch-based inference voting classification.

1 Introduction

The Food and Agriculture Organization of the United Nations (FAO) estimates that 420 million hectares of forest have been lost to deforestation between 1990 and 2020 (FAO, 2020), with agricultural or livestock expansion accounting for nearly 90 percent of global deforestation (EU Regulation 2023/1115). This leads to timber illegal logging activities that accounts for 15-30% of the global trade and is estimated to be worth between US\$50-152 billion (Barber &

Canby, 2018). Furthermore, between 30 and 50% of tropical timber traded globally comes from illegally logged forests (Lawson, 2014).

To counteract this issue, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) has been instrumental in promoting responsible and sustainable international timber trade (CITES, 2018). Additionally, the European Union has established critical regulations, including the EU Forest Law Enforcement, Governance and Trade (FLEGT) Plan¹, the European Union Timber Regulation (EUTR)² and, in 2023, the EU Deforestation Regulation (EUDR). Due to the visual similarity between many species, including those allowed and prohibited in trade, effectively enforcing these regulations demands trained personnel for initial identification of potentially protected timber. This must then be confirmed through microscopic analysis in specialised laboratories. This presents a significant challenge for governments, considering the time constraints and the high costs associated with specialised laboratories and trained personnel. Therefore, it is essential to provide border agents and authorities with a tool capable of efficient and precise early warning in order to detain shipments at origin and destination until they are recognised by qualified personnel. Such tool would lead to a significant reduction in illegal timber trafficking, thus, greatly decreasing deforestation and protecting tree endangered species.

The use of artificial intelligence, specifically machine learning (ML) techniques, has emerged as a promising approach to combat illegal logging and control the timber market. Numerous initiatives have attempted to automate the identification of wood species using ML techniques. These efforts can be categorised based on two distinct data models: identification based on anatomical data of timber (Esteban, Fernández, de Palacios, Romero, & Cano, 2009; Esteban et al., 2017; He, Marco, Soares, Yin, & Wiedenhoef, 2019), and identification based on computer vision and macroscopic images of timber, see (Hwang & Sugiyama, 2021) for a in-detail review. The latter approach has been the focus of a large number of publications in the past three years: (de Geus, Backes, Gontijo, Albuquerque, & Souza, 2021; Fabijańska, Danek, & Barniak, 2021; Figueroa-Mata, Mata-Montero, Valverde-Otárola, Arias-Aguilar, & Zamora-Villalobos, 2022; İsmail Kırbaş & Çifci, 2022; Chun et al., 2022; Kim et al., 2023; Nguyen-Trong, 2023; Urbano, Vargas-Cañas, & Mariño, 2023; Zheng et al., 2024). An overview of these approaches will be presented in Section 2.2.

It is widely recognised that high-resolution images can provide crucial information to identify timber, given the unique, yet intricate, fine-grained visual patterns present in different wood species (Maggiori, Tarabalka, Charpiat, & Alliez, 2017). Moreover, relying on anatomical data for automated identification requires significant more work by the user and is prone to human-error. Consequently, the research focusing on macroscopic images emerges as a more promising avenue. However, existing studies in this domain are subject to several limitations:

- There is a notable scarcity of high-quality extensive datasets that include a diverse range of wood species, compounded by the absence of a large public database to standardise results. We have found 6 datasets of macroscopic images of timber publicly available online, and all of them use different methods for image acquisition. Moreover, 5 out of these data sets contain images for 21 (or less) different species, whereas the data set with the largest number of different species (41) consist of images with low resolution (300x300 pixels). See Table 1 in Section 3 for a description of these datasets.
- Consistent image acquisition of timber poses a substantial challenge in developing machine learning models, and the quantity and quality of these images directly impacts the accuracy of wood species identification.
- Additionally, the full potential of high-resolution imagery has not been adequately leveraged in previous studies, where images are down-scaled to a maximum resolution of 300x300 pixels, thus, resulting in the under-utilisation of fine-grained patterns, which are crucial for accurate prediction.

Given these gaps, there is considerable potential for enhancing these tools to suit real-world applications effectively. Notably, the use of patch-based techniques has shown significant advantages in processing high-resolution images across various computer vision fields, as shown in (Cao, Li, Chen, & He, 2019; Hirra et al., 2021). Building on this foundation, we propose Timber Deep Learning Classification with Patch-based Inference Voting methodology, abbreviated TDLI-PIV methodology, a novel methodology that incorporates a multi-stage process specifically designed for the identification of wood species in timber.

During the training phase, the TDLI-PIV methodology trains Convolutional Neural Networks (CNNs) on subimages of high-resolution images, called *patches*, leveraging the availability of high-resolution macroscopic images of timber and the concept of patching. This technique enables CNNs to capture fine-grained patterns in timber images. In the prediction phase, the model’s robustness and accuracy are boosted through a patch-inference collaborative voting process. The code developed for this work is publicly available on <https://github.com/ari-dasci/S-GoIMAI>.

¹FLEGT Plan <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52003DC0251>

²European Union Timber Regulation https://ec.europa.eu/environment/forests/timber_regulation.htm.

In conjunction with this proposal, as we mentioned in the abstract, we provide a new dataset of macroscopic images of timber that covers 37 legally protected wood species. This dataset is called GOIMAI-Phase-I (Montes et al., Jun 4, 2024) and has been developed as part of the GOIMAI project, which focuses on the identification of protected wood species. GOIMAI-Phase-I consists of 2,120 high-resolution images that have been acquired utilising a smartphone with a 24x magnifying lens attached to the camera. The use of a magnifying lens allow us to pick up small patterns in timber in a portable and economical manner. This database has been made publicly available as an open-source resource to facilitate further research in this critical area, and can be downloaded from <https://doi.org/10.5281/zenodo.10903036>. For a comprehensive description GOIMAI-Phase-I, we refer to Section 3.

Our experimental analysis shows the superior performance of the TDLI-PIV methodology compared to other approaches discussed in the literature, as detailed in Section 5. Furthermore, we have assessed the influence of various parameters on our methodology, including the impact of data availability. This aspect is crucial, given the difficulty of acquiring multiple macroscopic images of timber coming from protected wood species. Therefore, methodologies that can achieve accurate results with minimal training data are preferred. Our proposal excels in this regard.

This paper is organised as follows. In Section 2 we describe the wood species classification problem, explaining the current ML approaches for this problem. In Section 3 we introduce the GOIMAI-Phase-I dataset. In Section 4 we introduce a description of the TDLI-PIV methodology. Results of the experiments conducted are then analysed in Section 5. Finally, in Section 6, we gather the main takeaways of this research and provide a perspective on future improvements.

2 Wood Species Identification and related work

In this section, we provide a description of the timber identification problem and the state of the art using machine learning. In Subsection 2.1 we introduce CITES, the legislation that specifies a list of legally protected woods. In Subsection 2.2 we present the current ML based approaches that have been proposed for the timber identification problem and discuss their limitations.

2.1 CITES Regulations for Trade in Endangered Wood Species

The mentioned CITES, the Convention on International Trade in Endangered Species of Wild Fauna and Flora, is a global treaty designed to regulate and monitor the international trade of endangered plant and animal species. One of the key objectives of CITES is to safeguard forests from being over-exploited and depleted, particularly those that are popular for their use in furniture, instruments, and other consumer goods (CITES, 2018).

CITES lists species of softwoods and hardwoods for which trade is closely monitored and controlled. The majority of these species are slow-growing, making them particularly vulnerable to over-harvesting. Some notable woods on the CITES list include *Brazilian rosewood*, *ebony*, and *mahogany*, which are prized for their unique aesthetic qualities but have been subject to habitat destruction and over-exploitation in recent years.

To trade in CITES-listed woods, companies must comply with strict guidance and regulations, which includes obtaining permits, ensuring that the wood has been legally harvested, and providing documentation of species and origin. Recent statistics have revealed that the implementation of CITES regulations has helped to reduce the illegal timber trade, preserve natural habitats, and even bolster local economies. However, timber traffickers are aware of how difficult it is to identify timber micro or macroscopically, which is why there are many entry and exit points in the European Union where protected timber are able to pass through controls without difficulty.

2.2 Machine learning based approaches for timber identification

As we have mentioned in the introduction, there are two main lines of research for ML-based timber identification: classification based on easily measurable anatomical data, and classification based macroscopic images (i.e. high-resolution images of timber that present any details that can not be seen with the naked eye).

Using anatomical data, Esteban et al. differentiated the timbers from the species *Juniperus cedrus* and *J. phoenicea* var. *canariensis* using artificial neural networks (ANN), reaching an accuracy level of 92% (Esteban et al., 2009). Moreover, they also managed to distinguish between *Pinus sylvestris* and *Pinus nigra* woods using ANN, with an accuracy level of 81% (Esteban et al., 2017). Similarly, Tuo et al. were able to differentiate *Swietenia macrophylla* from *Swietenia mahagoni* using a support vector machine (SVM) that processed anatomical data, achieving a 90% accuracy level (He et al., 2019). These results are very limited due to the fact that the data sets used only contain two different wood species, which contrasts to the aforementioned CITES list of species of softwoods and hardwoods for which trade is closely monitored.

Regarding the use of macroscopic images, ensuring consistent macroscopic image acquisition is a significant challenge that can affect the accuracy of the identification. A first step to overcome this challenge may be using a high-resolution digital camera at a distance of 15–20 cm, allowing for digital zoom. This is the approach followed by (Barmpoutis, Dimitropoulos, Barboutis, Grammalidis, & Lefakis, 2018), who have built an open-source dataset with 4272 images for 12 specimens (<https://doi.org/10.5281/zenodo.2545611>). The main issue with this approach is the lack of resolution and detail of the macroscopic images obtained. The model proposed in (Barmpoutis et al., 2018) crops and scales images to 400×400 pixels, extracts features through multidimensional texture analysis, and performs classification based on these features via a multi-class SVM model, achieving 91.47% accuracy of identification of the 12 most common Greek wood species. For another proposal based on SVM, see (Souza et al., 2020).

A more promising approach for macroscopic image acquisition is combining widely available digital cameras with optical magnifications. An example of such technology is XyloScope device, created by Hermanson et al. (Hermanson, Dostal, Destree, Wiedenhoef, et al., 2019). XyloScope has been used to collect the 15 most common commercial species in Ghana in order to build the Xylotron (Ravindran et al., 2019; Ravindran, Thompson, Soares, & Wiedenhoef, 2020) model, based on deep residual networks. Xylotron achieves an accuracy level of 97% using images obtained from the XyloScope in their laboratory. However, the accuracy dropped to 72% in real-world circumstances.

Other researchers have attempted to employ macroscopic images to timber to identify a larger number of wood species. Tang et al. reported an 87.3% success rate in identifying 100 common commercial species in Malaysia using their mobile app, MyWood-ID, which combines a 20x magnifying lens and a custom CNN SqueezeNet model as the CNN to classify macroscopic images of wood (Tang, Tay, Siam, & Lim, 2018). Their data set consists of 101,546 images from 1,919 specimens for 100 timber species, making (Tang et al., 2018) the largest study carried out to date.

In a later work, Verly Lopes et al. were able to differentiate timber of ten North American hardwood species with a 92.6% accuracy using Inception-ResNet v4 as CNN on macroscopic images obtained by coupling a 14x lens to a smartphone (Verly Lopes, Burgreen, & Entsminger, 2020). The data set in this case contains 1869 unique images from the 10 hardwood species, and the accuracy results are validated with a 5-fold cross validation.

The promising results presented in (Tang et al., 2018), (Ravindran et al., 2019) and (Verly Lopes et al., 2020) arguably placed CNN in the landscape of timber identification. Previous methods were mostly based on SVM classifiers or extraction of textures. In the overview (de Geus et al., 2021) and the methodology comparison (Hwang & Sugiyama, 2021), the authors confirm the superior performance of CNNs for timber identification. Since then, the findings of (Tang et al., 2018), (Ravindran et al., 2019) and (Verly Lopes et al., 2020) have been replicated in several studies, using different types of image acquisition protocols and neural network architectures: (Figueroa-Mata et al., 2022; İsmail Kırbaş & Çifci, 2022; Chun et al., 2022; Kim et al., 2023; Nguyen-Trong, 2023; Urbano et al., 2023; Zheng et al., 2024). However, at the core, all these proposals are intrinsically the same methodology, applying data augmentation to the original data set, down-scaling the images to the input size of the CNN, and fine-tuning the CNN architecture, as it is standard in CNN applications (Gómez-Ríos et al., 2019; Harkat, Nascimento, Bernardino, & Ahmed, 2023; Kiran Pandiri, Murugan, & Goel, 2024). Arguably, the only approach that has departed from this methodology is the one proposed by (Fabijańska et al., 2021), where, instead of taking a photograph, the authors used a scanner to acquire macroscopic images from cores of wood, leading to a resolution of 1984×92 pixels. Due to this unconventional resolution, the authors had to split these images into blocks/patches before processing them with CNNs and, then, aggregate the results of the CNN evaluation. Their dataset is available online (see Table 1) and consists of 312 images, containing 14 different European wood species. Accuracy is excellent, reaching 98.7%, although one could argue that further experiments are needed to validate this method given the small size of the dataset. The main downside of this methodology is that it is invasive and too time-consuming in practice due to the need to extract cores of wood from the timber stored at border control.

In spite of the hype on CNN-based models for timber identification, they are not accurate enough to see real-world application. Indeed, in order to achieve real world applicability we need to minimise false positives as much as possible as otherwise a large amount of timber will be kept on hold until it can be reliably tested in a specialised laboratory or checked by an expert.

Why does this happen (high number of false positives)? Current approaches based on CNN seem to be unable to reach higher accuracy in classification due to the fact that essential fine-grained patterns are lost or pixelated when images are downscaled to the required input size of the CNN architectures. Therefore, resulting models miss out on crucial details of the timber, which, as shown in (de Palacios et al., 2020), are needed to distinguish among wood species, especially for protected species under conventions like CITES. Image magnification via lenses reveals intricate features not visible with traditional lenses, associated to growth rings, vessel and parenchyma arrangement and presence of tyloses and deposits in vessels. These patterns can be easily missed or misinterpreted by conventional image analysis techniques.

Summarising, timber classification presents a unique set of challenges, notably the need to accurately identify fine-grained patterns that can occur across the timber surface, and requires of specific machine learning developments that boost the accuracy of CNNs.

3 GOIMAI-Phase-I: a wood species dataset through magnifying lenses in laboratory

A huge challenge in this area of research is the lack of publicly available data sets. Since researchers have to gather data for their experiments by themselves, new studies are very costly and time consuming, which has significantly impacted the speed of ML advancements in this field, as well as collaboration among research teams. Given the nuances of timber detection, we believe that it is mandatory to have access to publicly available high quality data in order to accelerate the arrival of ML based tools to market.

In this paper we introduce the named GOIMAI-Phase-I dataset (Montes et al., Jun 4, 2024), which has been developed as part of the GOIMAI project (<https://goimai.es>), a collaborative effort among the Andalusian Institute of Research in Data Science and Computational Intelligence, the School of Forestry Engineering and Natural Resources from the Technical University of Madrid, and Spanish Wood Trade and Industry Association (AEIM). This project co-funded by the Spanish Ministry of Agriculture and the Fisheries and Food and the European Agricultural Fund for Rural Development (FEADER). The aim of the project is to provide a tool for professionals in the sector, customs officers and law enforcement agencies, especially Spain’s Nature Protection Service (SEPRONA), that will aid EUTR and CITES compliance, and allow them to raise an early warning when they suspect a shipment contains illegally traded timber.

The GOIMAI-Phase-I dataset is publicly available as an open-source resource to facilitate further research in this critical area (<https://doi.org/10.5281/zenodo.10903036>). GOIMAI-Phase-I contains 2120 macroscopic images of timber from 37 different CITES wood species. These images has been captured with 3000x4000 pixel resolution in RGB, using a smartphone camera and optical 24x magnification; see Figure 1.

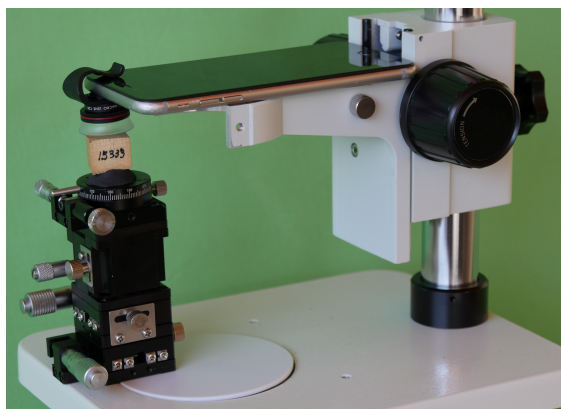


Figure 1: Set-up to capture macroscopic images of timber.

Figure 2 showcases a few examples of the images in GOIMAI-Phase-I. They show the distinct characteristics of each species of wood, including colour, pattern, and presence of vessels (holes). For instance, the presence of holes in the wood, which varies in quantity and size across different species, can be easily observed in Figure 2d. In contrast, Figure 2a shows a species with no holes. Additionally, each wood species displays a unique colour and pattern of lines (rays). For instance, Figure 2b exhibits darker lines with wider spacing, while the other images feature lighter lines with a smaller gap between them.

In GOIMAI-Phase-I there is a high variation between the number of images in each class; between 10 and 130 images. The distribution among classes of GOIMAI-Phase-I simulates a real-world scenario, where some wood species are significantly more rare than others. From a theoretical point of view, this imbalance should make the classification problem much harder. However, as we will see, this is not a problem for TDLI-PIV, our classification methodology based on patches and voting. On average, the class size in our data set has 57 images. Figure 3 shows the list of wood species covered in GOIMAI-Phase-I and the distribution of data for all these 37 classes.

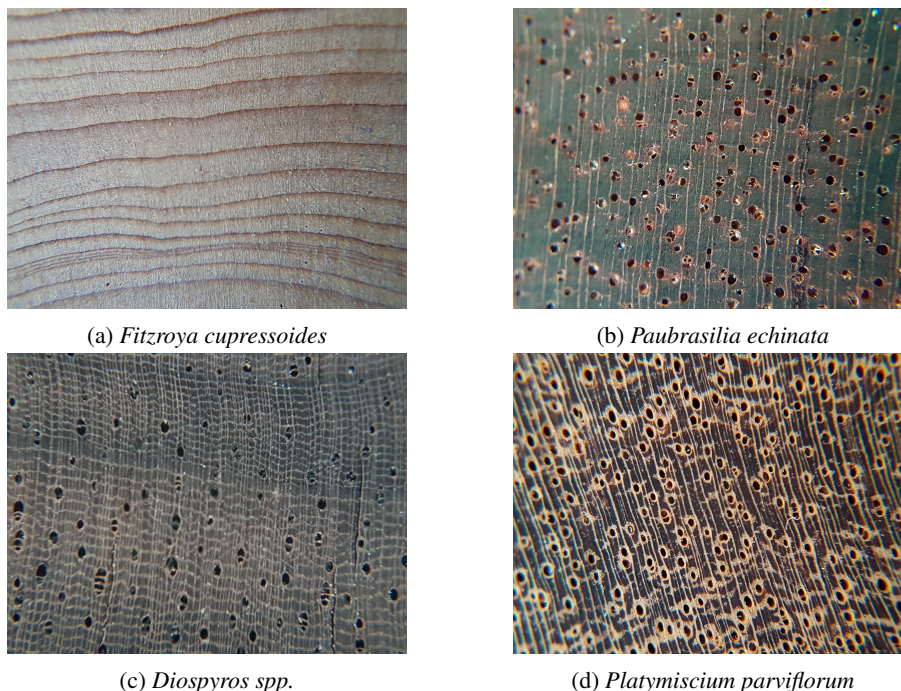


Figure 2: Examples of high-resolution images of timber from different wood species in the GOIMAI-Phase-I dataset.

To conclude this section, in Table 1 we compare GOIMAI-Phase-I to all the publicly available datasets that we have found in the literature. Each one of these datasets follows a different methodology for image acquisition, which we discuss below. Starting with the dataset Wood Auth³, images have been captured using a camera placed 20cm away from timber. Even though this is convenient, the resulting images have low resolution and, thus, do not accurately preserve the fine-grained details that are essential to accurately classify timber. A similar methodology is followed in the dataset Tropical Forest Species⁴, where images have been zoomed in with digital magnification. Nonetheless, this does not resolve the low resolution issue (640x480 pixels). A better approach is placing the camera 1cm away from the piece of timber, as it is done in the dataset Forest Species Database⁵. As a consequence, the resulting resolution is much higher (3264x2448 pixels). However, in practice, the images are slightly blurry and lack clarity of details. To avoid this issue, we have used a smartphone camera with x24 optical magnification in GOIMAI-Phase-I, leading to high-quality images as shown in Figure 2. A similar approach was followed in the dataset WoodRecognition⁶. Unfortunately, these images have been down-scaled to 300x300 pixels before making the dataset public. The dataset CAW⁷ uses a low-cost portable microscope to acquire images, which is easier to standardise among users. However, resolution of the obtained images is also lacking (640x480 pixels). Finally, a drastically different approach is that of the Tree Species Recognition dataset, where the authors cut cores of wood and scanned them in order to obtain high-resolution images where fine-grained details are visible, see (Fabijańska et al., 2021). Even though this approach is interesting from a research point of view, it is intrusive and too time consuming in practice. Overall, the methodology followed in GOIMAI-Phase-I presents the best balance between quality of images, portability and cost.

³The Wood Auth dataset can be downloaded from <https://zenodo.org/records/1216746>. See (Barmoutis et al., 2018) for the associated publication.

⁴The Forest Species Database can be downloaded from <https://web.inf.ufpr.br/vri/databases/forest-species-database-macroscopic/>. See (Filho, Oliveira, Nisgoski, & Britto, 2014) for the associated publication.

⁵The Wood Recognition dataset can be downloaded from <https://github.com/sunyongke/woodRecognition> and has been published by Y. Sun. For a paper using this dataset, see (Nguyen-Trong, 2023).

⁶The Tropical Forest Species dataset can be downloaded from <http://dx.doi.org/10.17632/yzzcbyvgnh.3>. See (Cano Saenz, Ordoñez Urbano, Gaitan Mesa, & Vargas-Cañas, 2022) for the associated publication.

⁷Tree Species Recognition dataset can be downloaded from <https://github.com/afabijanska/TreeSpeciesRecognition>. See (Fabijańska et al., 2021) for the associated publication.

⁸CAW (Commercial Amazon wood specie dataset) can be downloaded from https://www.facom.ufu.br/~backes/wood_dataset.php. See (de Geus et al., 2021) for the associated publication.

Dataset	Species	Num. images	Image Resolution	Acquisition
Wood Auth ³	12	8544	400x400	Camera (20cm from timber)
Forest Species Database ⁴	41	2942	3264x2448	Camera (1cm from timber)
Tropical Forest Species ⁵	11	10792	640x480	Camera with digital magnification
Wood Recognition ⁶	21	1126	300x300	Cellphone with magnifying glass
CAW ⁷	11	440	640x480	Portable microscope
Tree Species Recognition ⁸	14	316	1984x92	Scanner (on wood cores)
GOIMAI-Phase-I (ours)	37	2120	4000x3000	Camera with optical magnification

Table 1: List of public datasets of macroscopic wood images.

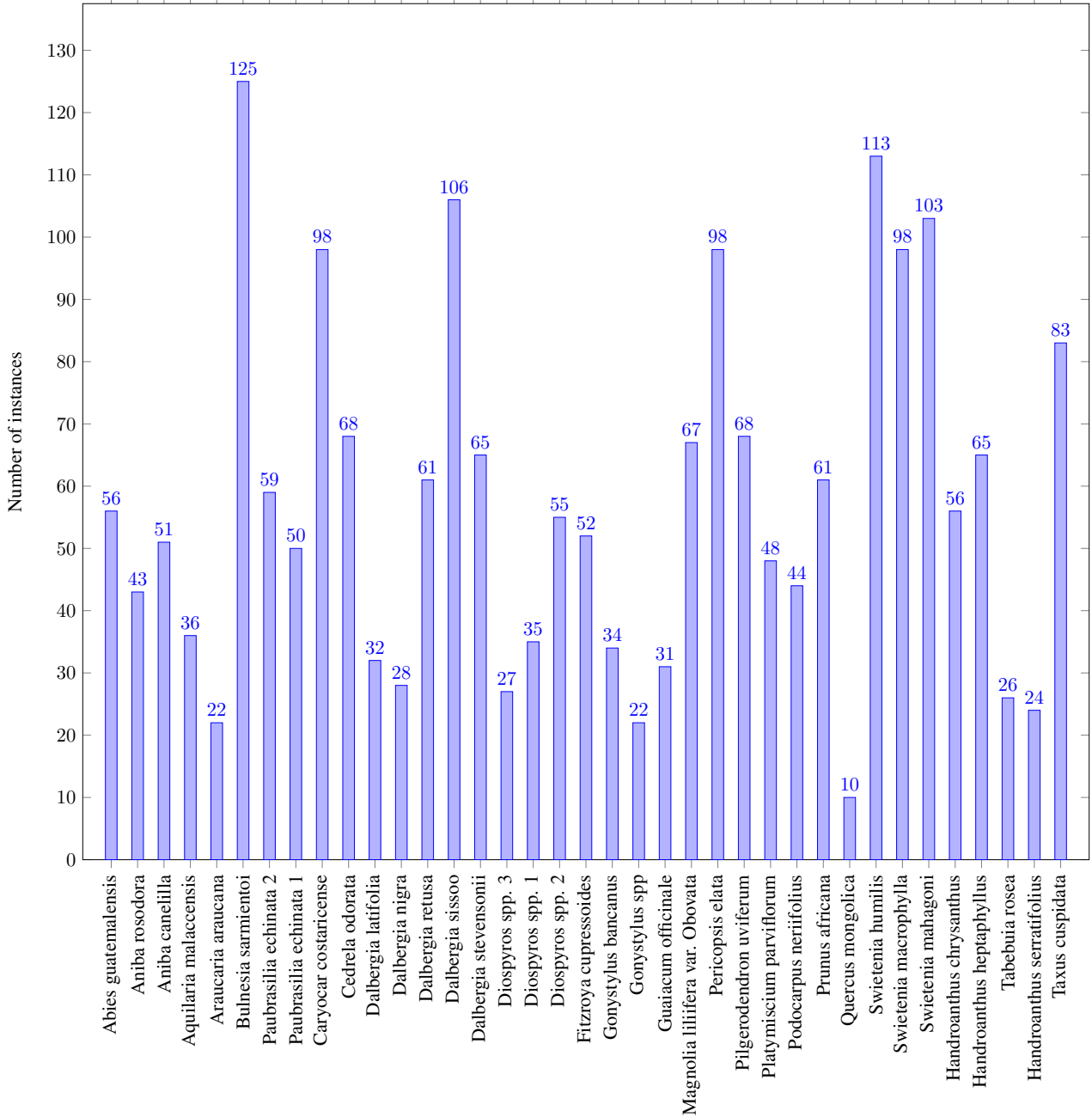


Figure 3: Number of images in each class of the database.

4 TDLI-PIV methodology: Timber Deep Learning Identification with Patch-based Inference Voting

It is well known that traditional CNNs may struggle to detect fine-grained patterns when the original images are massively down-scaled to the network input size. It is necessary to come up with new approaches that are able to exploit high resolution in order to extract fine-grained patterns and, thus, obtain an edge in timber classification. Our proposal for this problem, namely the TDLI-PIV methodology, enhances CNN on high-resolution images with fine-grained patterns. The TDLI-PIV methodology consists in training a CNN on subimages of the high-resolution image, called *patches*, following six stages divided into two phases, training and prediction. Figure 4 visually illustrates these phases and stages.

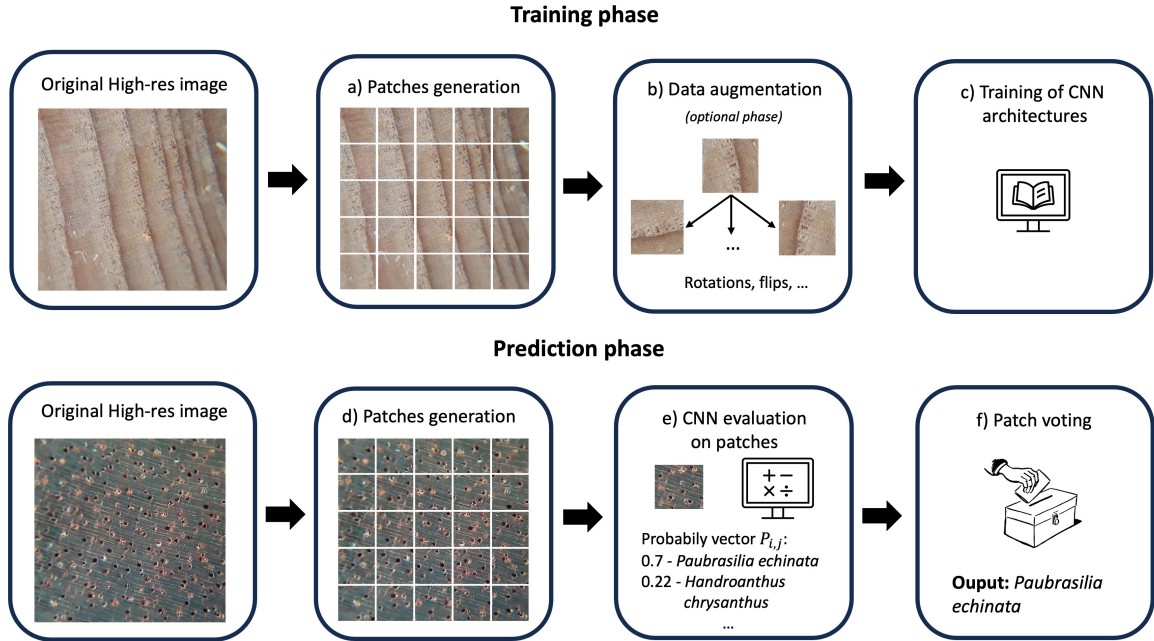
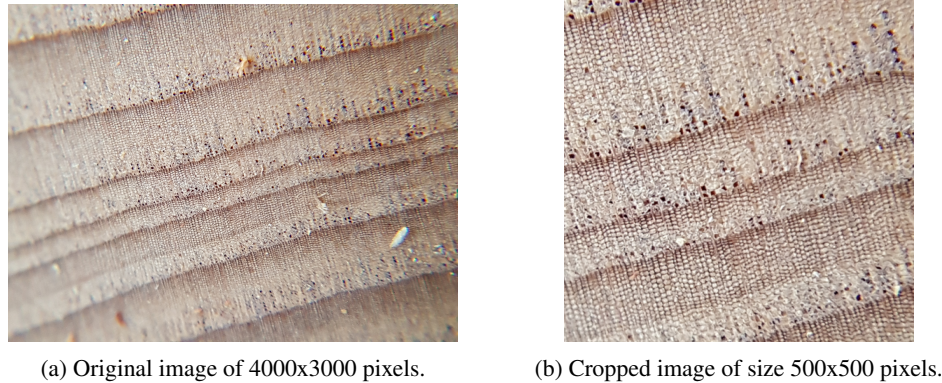


Figure 4: The TDLI-PIV methodology.

This section is organised as follows. In Section 4.1 we describe the training phase of the TDLI-PIV methodology, and in Section 4.2 we describe the prediction phase. Finally, in Section 4.3 we discuss the benefits of our proposal.

4.1 Training phase

Stage a: patches generation. Patching techniques offer a promising solution for enhancing the performance of wood species classification models. Figure 5 shows an example of an original high-resolution image of timber alongside a cropped image of size 500x500 pixels, which was obtained by cropping the original image at its centre. As we can see in this example, fine-grained patterns arise when we zoom in macroscopic images. These patterns are not taken into account in previous methodologies as they are lost when down-scaling macroscopic images to the input size of CNNs, thus, missing out on crucial characteristics of different wood species. For example, the popular CNN architecture EfficientNetV2 has a maximum input size of 480x480 pixels. In order to use all the information available in macroscopic images of timber, the TDLI-PIV first constructs a new dataset of patches by splitting the original high-resolution macroscopic images into patches using a grid, see Figure 4, stage a), for an example. Thus, some hyper-parameters arise when putting this training methodology into practice. First of all, we have to decide the size of the grid used to extract patches. In our study we use a 6x8 grid (so our 3000x4000 images give rise to 48 patches with resolution 500x500). The choice of size of the grid depends on the specific requirements of the task at hand. A smaller grid size, such as 3x4, may be preferred when dealing with larger, more prominent features, whereas selecting a larger grid size, such as 12x16, could be advantageous for tasks demanding a higher level of granularity and detail. In the case of timber identification, we have found that, for our dataset, the 6x8 grid presents the best balance between preserving fine-grained details and allowing larger features to appear in each patch.



(a) Original image of 4000x3000 pixels.

(b) Cropped image of size 500x500 pixels.

Figure 5: Comparison between the original image (a) and the cropped image (b).

Stage b: data augmentation. As a second step in the training phase, Figure 4 includes the possibility of applying data augmentation to patches in order to further increase the size and diversity of the training set. We highlight that the patch generation stage significantly enlarges the size of the data set and, thus, data augmentation is less critical in the TDLI-PIV methodology as opposed to previous approaches for timber identification. The diversity introduced by data augmentation may still be helpful, taking into account that there is a trade off with training time. Indeed, in Section 5 we analyse the effect of including data augmentation (or not) in the TDLI-PIV methodology, concluding that a simple data augmentation protocol has a positive impact in TDLI-PIV performance. This protocol consists in rotations and flips:

- Rotations have been applied to each patch, introducing variations in their orientations at 90, 180, 270 degrees. This step multiplies the size of the training set times 4. Figure 6 visually shown the impact of applying a 90 degrees rotation to an image. The comparison between the two images serves as a visual representation of the impact of rotation in line-shaped patterns, thus, highlighting the importance of incorporating rotation as a transformation technique in the data augmentation process. We note that, when generating the training set, rotating patches is equivalent to first rotating the high-resolution images 90, 180, 270 degrees, and then applying a grid to extract patches of the rotated images. The latter implementation is more efficient and, thus, is the one that we use in our code.
- The next step is applying a flip to all the previous generated patches. A flip is a symmetry with respect to the horizontal or vertical axis. To avoid further increasing the size of the data set (which after patches generation and rotations has been increased times 192), we apply a horizontal flip with probability 0.5 and, afterwards, a vertical flip with probability 0.5 (these two choices are independent). The original patch in the data set is then substituted by the resulting image after this random computation. That is, for each image, we maintain the original image with probability 0.25, and we substitute it by a flipped one with probability 0.75 (with one or two symmetries applied), so the size of the data set stays the same. Figure 6 visually shows the impact of flipping process on a patch.



(a) Cropped image of size 500x500 pixels from Figure 5.

(b) Image with rotation of 90 degrees and size 500x500 pixels.

(c) Image with horizontal flip and size 500x500 pixels.

Figure 6: Comparison between the cropped image from Figure 5, its 90 degrees rotation and a horizontal flip.

Overall, data augmentation is an effective method for increasing the size of the original dataset, improving the accuracy of models trained and providing robustness due to several reasons: diversity, overfitting reduction, improved generalisation, adaptability to CNN architectures, among others.

Stage c: training of CNNs architectures. To tackle the timber classification problem, we selected various CNN architectures, which will be compared in Subsection 5.3. The CNN architectures chosen for this task are Inception-ResNet v2 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017), and EfficientNet v2 (Tan & Le, 2021) B0-B3 (also known as versions XS, S, M, L). Note that each architecture has a different input size (largest one among these has input size 480x480 pixels) and, thus, patches must be re-scaled accordingly. Since our patches are 500x500 pixels, this involves a small down-scaling, which significantly contrasts to the large down-scaling performed in previous approaches where the input of the CNN is the whole image of timber. Here we briefly introduce the architectures InceptionResNet v2 and EfficientNet v2.

- InceptionResNet v2 (IRv2), proposed by Szegedy et al. in their paper "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning" (Szegedy et al., 2017), is a CNN architecture that builds on the Inception family of architectures and incorporates residual connections. These residual connections replace the filter concatenation stage of the Inception architecture. This architecture has been used in various tasks such as image classification and machine translation.
- EfficientNet V2 was proposed by Mingxing Tan and Quoc V. Le in their paper "EfficientNetV2: Smaller Models and Faster Training" (Tan & Le, 2021). These networks have faster training speed and better parameter efficiency than previous EfficientNet models. To develop this family of models, they use a combination of training-aware neural architecture search and scaling, to jointly optimise training speed and parameter efficiency. The models were searched from the search space enriched with new operations such as Fused-MBConv. EfficientNetV2 models train much faster than state-of-the-art models and are up to 6.8x smaller.

At this mid-stage, it is clear that our proposal presents the advantage of significantly increasing the size and diversity of the data sets where CNNs are trained, all this while performing a significantly less drastic down-scaling. Our method leverages the concept of patching and the availability of high-resolution images in order to detect the fine-grained patterns that characterise wood species.

4.2 Prediction phase

The second phase of our methodology deals with the evaluation of the trained CNNs on new high-resolution macroscopic images of timber. This second phase of TDLI-PIV methodology is visually depicted in Figure 4 as the prediction phase. Let us consider a new high-resolution image to evaluate. There are three stages in the prediction process:

Stage d: patches generation. We perform a grid division on the high-resolution image, getting the associated patches. The grid dimensions must match those used in the learning phase so that the size of the patches matches those that the model has been trained on. Recall that we perform a 6x8 grid division, obtaining 48 patches with 500x500 pixels. Taking into account the position of each patch in the original image, these patches can be seen as a matrix of patches, so each patch is determined by a pair (i, j) with $i \in \{0, 1, \dots, 6\} \times \{0, 1, \dots, 8\}$.

Stage e: CNN evaluation on patches. The CNN model is evaluated on each one of the patches, obtaining a vector of probabilities $P_{i,j}$ for each patch (i, j) . This vector indicates which wood species (in our particular case) the patch is more likely to belong to. Thus, we have to obtain a tensor of probabilities that our prediction will be based on.

Stage f: Patch inference voting. In order finish the inference process for prediction, it is necessary an aggregation the information from all these vectors of probabilities. For this aggregation, we perform a majority voting scheme. The class with the highest number of votes among all the patches is chosen as the output of our model. The patch (i, j) votes the class with the highest probability in the vector $P_{i,j}$.

4.3 Methodology analysis and benefits

The proposed methodology has been designed to exploit the advantages of having access high resolution macroscopic images and the fact that some wood species are only accurately discernible at a fine-grained level. The benefits of this methodology are summarised and described below:

- By extracting patches to construct the training set, we can create a larger, more diverse and comprehensive dataset for training the classification models.

- The down-scaling performed to match the patches size to the CNN input size is much less drastic than in traditional approaches. For example, in our case study, if this input size is 300x300 pixels, the down-scaling to train CNNs on the full image (3000x4000 pixels) has factors (0.1, 0.075), whereas our 500x500 pixels patches are down-scaled with a factor (0.6, 0.6). Therefore, crucial fine-grained patterns are still present in the input images and CNNs can learn to detect them for more accurate classification, whereas traditional full-image analysis methods may overlook these patterns due to down-scaling.
- Data augmentation is an effective method for increasing the size of the original dataset, improving the accuracy of models trained and providing robustness due to several reasons: diversity, overfitting reduction, improved generalisation, adaptability to CNN architectures, among others.
- The majority voting mechanism adds an additional layer of reliability and robustness to the classification process. Since timber patterns can vary greatly in distribution –some being uniformly spread while others are clustered– majority voting ensures that the final classification reflects the most prevalent patterns across the entire piece of timber.

5 Experimental framework, results and analysis

In this section, we present the experimental results obtained in the GOIMAI-Phase-I dataset. We conducted a series of studies to evaluate the performance of the TDLI-PIV methodology, including a comparison with other methodologies of the literature. To do that, first we introduce our experimental setup in Section 5.1. In Section 5.2 we analyse the performance of the TDLI-PIV methodology, including experiments on the input size of the CNN architecture, the contribution of the patch inference voting stage to the overall performance and the contribution of the data augmentation stage. In Section 5.3 we compare our methodology with other literature proposals. Finally, in Section 5.4 we address the impact of the dataset size for TDLI-PIV methodology performance, concluding that TDLI-PIV is suitable for datasets where samples of each class are scarce thanks to the extra information extracted via patch generation.

5.1 Experimental setup

For experiments we use the GOIMAI-Phase-I dataset, introduced in Section 3. To ensure statistical significance of our results, we perform a 5-fold cross-validation in every experiment. We point out that the high-resolution images are split into the five folds before preprocessing, so in every experiment the folds are the same.

The CNN models used in this study have been trained for 50 epochs, being the model from the last epoch the one selected for analysis. This approach is motivated by a desire to capture the model’s state at the conclusion of the training process.

We would like to highlight an important aspect of the experimental study. In this study, we consider two versions of the GOIMAI Phase-I dataset: the original dataset containing 2120 images, and a reduced dataset comprising only 25% of the original, 530 images. The goal for using this reduced dataset is to investigate how the methodology and the models perform with varying dataset sizes, where some classes may have very few examples. This analysis is motivated by a potential extension of the data set to include more wood species and the difficulty of acquiring examples of certain wood species that are illegal to trade.

5.2 TDLI-CPIV methodology analysis for CNNs models and subimages selection

In this subsection we analyse the TDLI-CPIV methodology performance considering a 6x8 grid, resulting in subimages of 500x500 pixels. The 6x8 grid offers a balance between granularity and inclusion of larger patterns; each patch covers a significant area, enabling the model to capture broader context while performing minimal-downscaling to the input size of the CNN architecture, thus, preserving fine-grained patterns.

The choice between this grid configuration and others depends on the specific requirements of the task at hand. The 6x8 grid might be preferred when dealing with larger, more prominent features, whereas selecting a smaller patch size could be advantageous for tasks demanding a higher level of granularity and detail. We must note that experiments with 12x16 grids were made, resulting in patches of 250x250 pixels. Results with this approach were not promising, showcasing that there is a loss of information with such small patches for this classification problem. It is preferable to down-scale the 500x500 patches to the input size of the CNN architecture, and, thus, this is the approach followed in the rest of this experimentation.

In our study, we used different models of EfficientNet V2, recognised as state-of-the-art CNNs. These models allow us to assess performance variations with different input sizes as well as patch sizes. Table 2 shows the input size that we have used for each model.

Model	Input size (pixels)
EfficientNet V2 B0	224x224
EfficientNet V2 B1	240x240
EfficientNet V2 B2	260x260
EfficientNet V2 B3	300x300

Table 2: Input size for the EfficientNet V2 models.

First, we have trained these models following the learning phase showcased in Figure 4, except for the data augmentation step in order to determine the performance of the various EfficientNet V2 models. The architecture EN V2 B3 shows the best results as one would expect since less details are lost when down-scaling. The fact that models with larger input size perform better supports our hypothesis that down-scaling has a negative impact in timber identification.

Models	Accuracy					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
EN V2 B0-w/oDA	0.998	0.973	0.963	0.978	0.975	0.977
EN V2 B1-w/oDA	0.988	0.973	0.992	0.96	0.968	0.976
EN V2 B2-w/oDA	0.973	0.98	0.993	0.996	0.975	0.983
EN V2 B3-w/oDA	0.981	0.988	0.995	0.979	0.986	0.986

Table 3: Cross-validation results for grid 6x8, patch-based voting. Models labeled w/oDA indicate no data augmentation was used during their training.

In order to analyse the influence of patch voting inference in the accuracy of the model, we have evaluated the CNN models obtained after training on a single central crop of 500x500 pixels. Table 4 shows the results. We see a significant drop of accuracy with respect to the results presented in Table 3, concluding that patch voting inference is a key step in the TDLI-CPIV methodology. This is not a surprise, as we can see patch voting as an ensemble, which would be expected to improve robustness of our model.

Models	Accuracy					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
EN V2 B0-w/oDA	0.983	0.967	0.946	0.965	0.962	0.965
EN V2 B1-w/oDA	0.973	0.965	0.963	0.948	0.944	0.959
EN V2 B2-w/oDA	0.958	0.973	0.984	0.975	0.96	0.97
EN V2 B3-w/oDA	0.962	0.983	0.988	0.955	0.974	0.972

Table 4: Cross-validation results for grid 6x8, evaluating the central crop. Models labeled w/oDA indicate no data augmentation was used during their training.

Finally, to conclude this section we present the results of the full TDLI-CPIV methodology in Table 5. The difference here with respect to Table 3 is that we include the data augmentation stage. The architecture used is EfficientNet V2 B3. The results showcase that data augmentation can indeed further improve performance and, thus, we recommend its inclusion in our methodology. Remarkably, only 13 images out of 2120 have been misclassified by our model, which supports the viability of our methodology for practical application in border control.

Models	Accuracy					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
TDLI-CPIV-EN V2 B3	0.993	0.998	1.0	0.993	0.988	0.994

Table 5: Cross-validation results of the TDLI-CPIV methodology, using EfficientNet V2 (EN V2 B3).

5.3 State-of-the-art models for comparison

In this section we compare TDLI-CPIV to other state-of-the-art models for timber identification on macroscopic images (using the dataset GOIMAI Phase-I), and analyse the obtained results. Recall that in Subsection 2.2 we surveyed other approaches published in the literature, concluding that the application of CNN to timber identification is the most

successful line of research. When it comes to the application of CNNs to datasets where images have been acquired using optical magnification, we highlight two publications that develop successful CNN models: (Tang et al., 2018) and (Verly Lopes et al., 2020). In a nutshell, (Tang et al., 2018) and (Verly Lopes et al., 2020) apply state-of-the-art CNN architectures together with data augmentation. As pointed out in Subsection 2.2, papers published after the work of (Tang et al., 2018) focus on a different method for image acquisition or change the CNN architecture/data augmentation protocol. In order to perform a fair comparison with our proposal, we have trained both data augmentation methodologies of (Tang et al., 2018) and (Verly Lopes et al., 2020) using EfficientNet V2, the architecture that we have used in TDLI-CPIV.

5.3.1 Analysis of the model proposed by Verly Lopes et al., 2020

In the context of (Verly Lopes et al., 2020), a meticulous set of image augmentation techniques was applied to enhance the robustness and generalisation capabilities of a deep learning model. Notably, images were first cropped to 3000x3000 pixels. Subsequent augmentations, including random zoom, random flip, random rotation, and resizing to standardise input dimensions, were used to produce 20 images from each original to increase the training dataset. In the following we shortly describe them:

- **Random Zoom:** This augmentation randomly adjusts the height of training images within a range of 80% to 120% of their original height while maintaining the original width. The purpose is to introduce variability in scale, aiding the model in learning to recognise objects at different zoom levels.
- **Random Flip:** With a 50% probability, this augmentation introduces a horizontal flip to training images. By incorporating left and right orientation variations, the model becomes more adept at handling diverse scenarios, contributing to improved performance on real-world data.
- **Random Rotation:** Training images undergo a random rotation, with angles ranging from -45 to 45 degrees. This augmentation aims to expose the model to various viewpoints, enhancing its ability to generalise across different orientations and angles.
- **Resize:** To standardise input dimensions before training a CNN, training images are resized to a fixed dimension of 299 pixels in width and 299 pixels in height.

Authors mention that these augmentations were chosen after careful consideration of the characteristics of the dataset and the desired model performance, validated through rigorous experimentation, resulting in improved model robustness and performance metrics during the training process.

The dataset GOIMAI-Phase-I is used for training two models with this data augmentation protocol, InceptionResnet v2 (noted as VL-IRv2), which is the model chosen in (Verly Lopes et al., 2020), and EfficientNet-V2 B3 (noted as VL-EN V2 B3). A 5-fold cross-validation was conducted to assess the models’ performance. Results are shown in Table 6. Results show a low performance in comparison with TDLI-PIV methodology.

Models	Accuracy					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
VL-IRv2	0.9615	0.9709	0.9612	0.9903	0.9728	0.9728
VL-EN V2 B3	0.9808	0.9417	0.9709	0.9806	0.9709	0.967

Table 6: Evaluation results for InceptionResNet v2 (IR V2) and EfficientNet V2 B3 (EN V2 B3) with Data Augmentation from Verly Lopes et al 2020.

5.3.2 Analysis of the model proposed by Tang et al., 2018

Before delving into the evaluation results, let’s explore the augmentation strategies employed in (Tang et al., 2018). The training dataset underwent extensive augmentation, with each image being augmented using the following techniques:

- **Random Rotation:** Each training image underwent random rotations within the range of -45.0 to 45.0 degrees. This augmentation enhances the model’s ability to handle variations in object orientations.
- **Random Brightness:** Random adjustments to brightness were introduced, with a minimum factor of 0.8 and a maximum factor of 1.2. This diversifies the illumination conditions experienced by the model during training.
- **Random Flip:** Horizontal flips were incorporated with a probability of 1.0. This provides the model exposure to left and right orientation variations, contributing to improved real-world performance.

- **Resize:** Images were resized to a fixed size of 224 pixels in width and 224 pixels in height, as this is the input size of the CNN architecture used (a customised version of SqueezeNet).

Unfortunately the customised version of SqueezeNet use in (Tang et al., 2018) is not available online. Since the authors use an architecture with 224x224 input, we have selected the model EfficientNet V2 B0 for this comparison. We refer to the resulting model trained with the data augmentation protocol of (Tang et al., 2018) as T-EN V2 B0. Table 7 shows the results obtained with this methodology, which does not reach the accuracy levels of any of the models presented in Subsection 5.2.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
T-EN V2 B0	0.943	0.959	0.943	0.959	0.959	0.9526

Table 7: Evaluation results for EfficientNet V2 B0 (EN V2 B0) with Data Augmentation from Tang et al. 2018.

5.4 Impact of Dataset Size on TDLI-PIV methodology performance

In this subsection, we continue to explore the TDLI-PIV methodology, with the aim to study whether the size of the dataset influences the outcome of the proposed methodology. To achieve this, we use a version of the dataset composed of only 25% of the total images from the original dataset. This data set has been sampled maintaining the proportion of instances belonging to each class. Table 8 shows the results of TDLI-PIV on this reduced dataset, as well as those of the methodologies of (Tang et al., 2018), denoted as T-EN v2 B0, and (Verly Lopes et al., 2020), denoted VL-IRv2 and VL-EN v2 B3, following the notation of Section 5.3. These results show that the TDLI-PIV methodology also excels on data sets where classes have few samples. In our opinion, this is due to the fact that the use of patches allows us to extract significantly more information from a single image than traditional approaches. This sort of behaviour is extremely important when extending the model to a larger number of classes, since it is challenging to get many samples of certain wood species that are illegal to trade.

Models	Accuracy					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
TDLI-PIV-EN V2 B3	0.991	1.0	1.0	0.991	1.0	0.996
VL-IRv2	0.9615	0.9709	0.9612	0.9903	0.9728	0.9728
VL-EN v2 B3	0.9808	0.9417	0.9709	0.9806	0.9709	0.967
T-EN v2 B0	0.972	0.991	0.904	0.991	1.0	0.971

Table 8: Cross-validation results for TDLI-PIV and the Tang et al. and Verly Lopes et al. methodologies.

6 Conclusions and future directions

This work has been developed with two primary objectives. First, we aim to demonstrate that fine-grained details in timber are crucial for highly accurate identification with machine learning models. By leveraging high-resolution images in model design, we seek to enhance accuracy in timber identification, which has stagnated in recent years (see Section 2.2 for an overview of the state-of-the-art). Second, we intend to use these findings to advocate for the industry-standard adoption of optical magnification when acquiring images of timber for automated identification.

To achieve these objectives, we have presented two main contributions. First, we introduce the GOIMAI-Phase-I dataset, the first open database of high-resolution macroscopic images of timber acquired with optical magnification and a smartphone camera. This dataset covers 37 legally protected wood species and contains a total of 2,120 images, publicly available for download in (Montes et al., Jun 4, 2024).

Our second contribution is the TDLI-PIV methodology, which enhances convolutional neural networks (CNNs) for classification of high-resolution images when classes are characterised by intricate fine-grained patterns. TDLI-PIV comprises three key elements: the extraction of patches, the implementation of data augmentation techniques, and the application of a patch-based voting system for classification. Our TDLI-PIV methodology exhibits exceptional performance, achieving an accuracy of 99.4% in the GOIMAI-Phase-I dataset, the best reported accuracy of image identification up to date. This contrasts to the accuracy obtained by the methodologies of (Tang et al., 2018) and (Verly Lopes et al., 2020), which reach at most 97.28% in our dataset.

We draw four important conclusions regarding data quality and model development for timber identification:

- Our proposal reaches a level of accuracy that is suitable for real world application, where there is a low tolerance to false positives due to the associated economical and time cost of analysing every piece of timber in a laboratory.
- Our findings underscore the fact that developing techniques that exploit fine-grained patterns is crucial in order to achieve accurate models for timber identification. In order to fuel these techniques, the use of high-resolution images is key moving forward and, thus, one must avoid drastically down-scaling images to the input size of the CNN architecture in timber identification.
- Analysing the results from state of the art models, (Tang et al., 2018) reported an accuracy of 87.3% in identifying 100 common commercial species, while (Verly Lopes et al., 2020) achieved a 92.6% accuracy in classifying ten North American hardwood species. Utilising their data augmentation methods in conjunction with EfficientNetV2 on the GOIMAI-Phase-I dataset, results showed an enhanced performance, reaching approximately 95-97%. This showcases the critical role of high-quality datasets and image acquisition in computer vision, and supports our hypothesis that the use of x24 optical magnification should be the industry standard in timber identification. Indeed, the need of quality and excellence data is well-known and discussed in the literature (Aroyo, Lease, Paritosh, & Schaekermann, 2022).
- Thanks to the use of patches, the TDLI-PIV methodology is able to extract a significant amount of information from each high-resolution image, as evidenced by the consistently excellent results achieved across both the original dataset and the reduced dataset. Even with a smaller dataset, comprising only 25% of the original images, our model achieves an accuracy of 99.6%. This suggests that our approach is capable of generalising well to datasets where classes have very few examples, a crucial factor in real-world application. Indeed, availability of timber of protected species is limited and, thus, few examples are available in current datasets.

Finally, we highlight two important challenges for future research in this field that we should tackle in order to achieve real-world application.

- The next obvious challenge is gathering a public dataset of high-resolution images of timber that comprises most wood species that are used commercially as well as those species listed in CITES. It is necessary to standardise the image acquisition process, and the research community should work together towards such a standardised public dataset.
- Another important challenge for real-world application, is adapting this methodology to images taken with a variety of smartphones, magnifying lenses and natural conditions. That is, with poor sample acquisition (lack of sanding or imperfect cutting) and with poor photograph sample acquisition (bad lighting, bad pulse, ...). Another fact to be considered is that the image acquisition process introduces a significant variability depending on which device has been used to take the photographs, given the variety of cameras in smartphones and post-processing algorithms developed by each brand to improve image quality. Nonetheless, succeeding in such an endeavour would lead to a cheap and practical protocol to automatically identify timber in border controls.

7 Acknowledgements

This work has been partially supported by the Ministry of Science and Technology of Spain under the project PID2020-119478GB-I00. This work has been also partially supported by Programa Nacional de Desarrollo Rural 2014-2020 del Ministerio de Agricultura, Pesca y Alimentación y el Fondo Europeo Agrícola de Desarrollo Rural (FEADER). Thank also to Jodrell Laboratory of the Royal Botanic Gardens, Kew for donation of wood samples.

References

- Aroyo, L., Lease, M., Paritosh, P., & Schaekermann, M. (2022). Data excellence for ai: why should you care? *Interactions*, 29(2), 66–69.
- Barber, C. V., & Canby, K. (2018). *Assessing the timber legality strategy in tackling deforestation*. World Resources Institute, Washington. Available at <https://www.wri.org/research/ending-tropical-deforestation-assessing-timber-legality-strategy-tackling-deforestation>.
- Barmpoutis, P., Dimitropoulos, K., Barboutis, I., Grammalidis, N., & Lefakis, P. (2018). Wood species recognition through multidimensional texture analysis. *Computers and electronics in agriculture*, 144, 241–248.
- Cano Saenz, D. A., Ordoñez Urbano, C. F., Gaitan Mesa, H. R., & Vargas-Cañas, R. (2022). Tropical wood species recognition: A dataset of macroscopic images. *Data*, 7(8).
- Cao, G., Li, J., Chen, X., & He, Z. (2019). Patch-based self-adaptive matting for high-resolution image and video. *The Visual Computer*, 35, 133–147.

- Chun, T. H., Hashim, U. R., Ahmad, S., Salahuddin, L., Choon, N. H., & Kanchymalay, K. (2022). Efficacy of the image augmentation method using cnn transfer learning in identification of timber defect. *International Journal of Advanced Computer Science and Applications*, 13(5).
- CITES. (2018). *Convention on international trade in endangered species of wild fauna and flora*. Available at <https://cites.org/eng/disc/what.php>.
- de Geus, A. R., Backes, A. R., Gontijo, A. B., Albuquerque, G. H. Q., & Souza, J. R. (2021, May 01). Amazon wood species classification: a comparison between deep learning and pre-designed features. *Wood Science and Technology*, 55(3), 857-872.
- de Palacios, P., G. Esteban, L., Gasson, P., García-Fernández, F., de Marco, A., García-Iruela, A., ... González-de Vega, D. (2020). Using lenses attached to a smartphone as a macroscopic early warning tool in the illegal timber trade, in particular for cites-listed species. *Forests*, 11(11), 1147.
- Esteban, L. G., de Palacios, P., Conde, M., Fernández, F. G., García-Iruela, A., & González-Alonso, M. (2017). Application of artificial neural networks as a predictive method to differentiate the wood of *pinus sylvestris* L. and *pinus nigra* arn subsp. *salzmannii* (dunal) franco. *Wood Science and Technology*, 51, 1249–1258.
- Esteban, L. G., Fernández, F. G., de Palacios, P. d. P., Romero, R. M., & Cano, N. N. (2009). Artificial neural networks in wood identification: the case of two juniperus species from the canary islands. *IAWA journal*, 30(1), 87–94.
- Fabijańska, A., Danek, M., & Barniak, J. (2021). Wood species automatic identification from wood core images with a residual convolutional neural network. *Computers and Electronics in Agriculture*, 181, 105941.
- FAO. (2020). *Evaluación de los recursos forestales mundiales 2020 – principales resultados*. Available at <https://doi.org/10.4060/ca8753es>.
- Figueroa-Mata, G., Mata-Montero, E., Valverde-Otárola, J. C., Arias-Aguilar, D., & Zamora-Villalobos, N. (2022). Using deep learning to identify costa rican native tree species from wood cut images. *Frontiers in Plant Science*, 13.
- Filho, P. L. P., Oliveira, L. S., Nisgoski, S., & Britto, A. S. (2014). Forest species recognition using macroscopic images. *Machine Vision and Applications*, 25, 1019–1031.
- Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A., Krawczyk, B., & Herrera, F. (2019). Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications*, 118, 315-328.
- Harkat, H., Nascimento, J. M., Bernardino, A., & Ahmed, H. F. T. (2023). Fire images classification based on a handcraft approach. *Expert Systems with Applications*, 212, 118594.
- He, T., Marco, J., Soares, R., Yin, Y., & Wiedenhoef, A. C. (2019). Machine learning models with quantitative wood anatomy data can discriminate between *swietenia macrophylla* and *swietenia mahagoni*. *Forests*, 11(1), 36.
- Hermanson, J. C., Dostal, D., Destree, J. C., Wiedenhoef, A. C., et al. (2019). *The xyloscope: A field-deployable macroscopic digital imaging device for wood*. United States Department of Agriculture, Forest Service, Forest Products Laboratory.
- Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., ... Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, 9, 24273-24287.
- 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), 47.
- Kim, J.-H., Purusatama, B. D., Saverio, A. M., Prasetia, D., Yang, G.-U., Han, S.-Y., ... Kim, N.-H. (2023). Performance influencing factors of convolutional neural network models for classifying certain softwood species. *Forests*, 14(6).
- Kiran Pandiri, D., Murugan, R., & Goel, T. (2024). Smart soil image classification system using lightweight convolutional neural network. *Expert Systems with Applications*, 238, 122185.
- Lawson, S. (2014). *Consumer goods and deforestation: An analysis of the extent and nature of illegality in forest conversion for agriculture and timber plantations*. Available at <https://www.forest-trends.org/publications/consumer-goods-and-deforestation/>.
- Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). High-resolution image classification with convolutional networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (p. 5157-5160).
- Montes, R., Herrera-Poyatos, D., Herrera-Poyatos, A., Herrera, F., de Palacios, P., De Marco, A., ... G. Esteban, L. (Jun 4, 2024). *GOIMAI DataSet-Phase-I of high-resolution macroscopic images for the identification of wood species*. Available for download at <https://zenodo.org/doi/10.5281/zenodo.11092208>.
- Nguyen-Trong, K. (2023). Evaluation of wood species identification using cnn-based networks at different magnification levels. *International Journal of Advanced Computer Science and Applications*, 14(4).
- Ravindran, P., Ebanyenle, E., Ebeheakey, A. A., Abban, K. B., Lambog, O., Soares, R., ... Wiedenhoef, A. C. (2019). Image based identification of ghanaiian timbers using the xylotron: Opportunities, risks and challenges. In *arxiv preprint arxiv:1912.00296, proceedings of the 33rd conference on neural information processing systems (neurips 2019), vancouver, canada*.

- Ravindran, P., Thompson, B. J., Soares, R. K., & Wiedenhoef, A. C. (2020). The xylotron: flexible, open-source, image-based macroscopic field identification of wood products. *Frontiers in plant science*, *11*, 1015.
- 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 Science and Technology*, *54*(4), 1065–1090.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 31).
- Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096–10106).
- Tang, X. J., Tay, Y. H., Siam, N. A., & Lim, S. C. (2018). Mywood-id: Automated macroscopic wood identification system using smartphone and macro-lens. In *Proceedings of the 2018 international conference on computational intelligence and intelligent systems* (p. 37–43). New York, NY, USA: Association for Computing Machinery.
- Urbano, C. F. O., Vargas-Cañas, R., & Mariño, N. M. D. (2023). *Imaca—automated wood identification system of colombian timber species using convolutional neural networks*. Available at Research Square [<https://doi.org/10.21203/rs.3.rs-3640320/v1>].
- Verly Lopes, D. J., Burgreen, G. W., & Entsminger, E. D. (2020). North american hardwoods identification using machine-learning. *Forests*, *11*(3).
- Zheng, Z., Ge, Z., Yang, X., Liu, X., Qin, L., Wang, X., & Zhou, Y. (2024). A-repvgg: Research on classification algorithms based on deep learning and wood ct images. Available at Research Square [<https://doi.org/10.21203/rs.3.rs-4021077/v1>].
- İsmail Kırbaş, & Çifci, A. (2022). An effective and fast solution for classification of wood species: A deep transfer learning approach. *Ecological Informatics*, *69*, 101633.