Advances in Timber Identification Using Deep Learning: A Review of Convolutional Neural Network Models

Kuryati Kipli, and Muhammad Zakwan Suhaimee

Machine Vision (MV) software has emerged as a powerful tool for timber species identification, offering significant advantages including species-level accuracy, cost-effectiveness, and the elimination of human errors. The development of MV software relies on three major supporting technologies: computer vision, machine learning (ML), and deep learning (DL). This paper provides an in-depth exploration of the role of DL, with a particular focus on convolutional neural networks (CNNs), in enhancing MV software for timber identification. The potential of CNN architectures is examined in detail, including a review of commonly used CNN models and their effectiveness in identifying different timber species. This paper also discusses current practices in developing CNN models for integration into MV software for timber identification tasks, highlighting the standards and procedures researchers should follow to ensure optimal performance and reliability. Additionally, the challenges associated with implementing CNN models for timber identification are addressed. They include limited model usability due to the diversity of timber species and geographical variation. Variability in methodologies, imaging devices, and data processing approaches across studies further complicates the comparison and integration of results. This paper emphasises the need for standardised practices and further empirical research to address these inconsistencies and improve CNN-based MV systems for timber identification.

DOI: 10.15376/biores.21.3.Kipli

Keywords: Convolutional neural networks; Deep learning; Machine vision software; Timber identification

Contact information: Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Sarawak, Malaysia;
** Corresponding author: kkuryati@unimas.my*

INTRODUCTION

Timber identification is a crucial field supporting wood science and forest sustainability, as it enables the determination of the origin and species of timber, and plays a significant role in protecting endangered species from illegal trade. For example, Ramin, which is listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), has continued to be illegally traded in European and United States markets (Kleinschmit *et al.* 2016). Therefore, accurate timber identification is essential for the protection of endangered species.

According to the proceedings of the IAWA-IUFRO International Symposium 2019, timber identification plays a vital role in addressing illegal logging (Koch *et al.* 2019). Illegal logging is often associated with false claims of timber species, volume, and quality to reduce or evade the payment of legally prescribed taxes and fees (NEPCon 2018). Timber identification ensures the accurate declaration of a wood’s genus or species and its origin, thereby facilitating the enforcement of international regulations and conventions such as CITES (Koch *et al.* 2019). As a result, a variety of timber identification methods are currently utilised in the field.

There are currently seven timber identification methods, and these can be further categorised into five groups: wood anatomy, genetics, stable isotopes, mass spectrometry, and near-infrared (NIR) spectroscopy (Schmitz *et al.* 2020). Wood anatomy includes macroscopic anatomical analysis, microscopic anatomical analysis, and machine vision (MV) software (Schmitz *et al.* 2020). These three methods are similar in that they utilise the three anatomical planes of timber, which are the transverse, tangential, and longitudinal planes (Schmitz *et al.* 2020). However, several key differences exist among these methods.

Macroscopic and microscopic anatomical analyses are manual methods in which the timber identification process is performed by humans (Schmitz *et al.* 2020). Meanwhile, MV software represents a digitised method in which the identification process is performed by software. Next, the magnification level utilised in macroscopic and microscopic anatomical analyses differs substantially. Macroscopic analysis involves observations made with the unaided eye or with the aid of a magnifying lens (Schmitz *et al.* 2020). In contrast, microscopic analysis requires a high level of magnification and often involves laboratory procedures (Siam *et al.* 2023).

Macroscopic anatomical analysis, microscopic anatomical analysis, and MV software share both similarities and differences, and each method has its own strengths and limitations. In the case of macroscopic anatomical analysis, it is generally the fastest evaluation method as it does not require complex laboratory procedures. A macroscopic lens is the most commonly used tool in macroscopic analysis (Nordahlia and Lim 2016; Timber Portal, Sarawak Timber Industry Development Corporation 2019). Additionally, this method is practical for initial screening and preliminary assessment of wood characteristics (Dormontt *et al.* 2015). However, macroscopic anatomical analysis requires trained wood anatomists and faces difficulties in achieving species-level identification (Schmitz *et al.* 2020). Nevertheless, identification of timbers at the genus level is generally achievable (Dormontt *et al.* 2015).

In the case of microscopic anatomical analysis, its strengths include the availability of readily accessible wood identification frameworks supported by extensive databases such as InsideWood, Commercial timbers, Softwoods, Forest Species Database (FSD), and Microscopic Wood Anatomy of Central European species (Schmitz *et al.* 2020). These databases serve as reference databases for microscopic wood assessment. However, this method is constrained by the requirement for laboratory settings for wood sample preparation, as described by Siam *et al.* (2023). Additionally, microscopic anatomical analysis faces difficulties in achieving species-level identification, although identification at the genus level is generally reliable (Dormontt *et al.* 2015; Schmitz *et al.* 2020; Siam *et al.* 2023).

The MV software-based approach employs software to identify timber species. Specifically, the technologies most commonly utilised are computer vision (CV) and machine learning (ML) (Schmitz *et al.* 2020). The MV models can be configured to perform timber classification either using on-site hardware or *via* cloud-based interfaces (Schmitz *et al.* 2020). This method shows a viable potential as an alternative timber identification tool, as it has demonstrated the potential in achieving species-level identification (Schmitz *et al.* 2020). Additionally, MV-based identification does not require trained wood anatomists, as the process is automated through software, thereby reducing human errors. Nevertheless, this method has several limitations, including the absence of extensive field testing and verification, limited access to relevant scientific expertise for developing and evaluating MV tools, and insufficient reference materials and/or databases (Dormontt *et al.* 2015; Schmitz *et al.* 2020).

In addition to the three aforementioned methods, genetic approaches are also applicable to timber identification tasks. These methods involve analysing the genetic code of tree species (Dormontt *et al.* 2015). Common genetic approaches used in timber identification are deoxyribonucleic acid (DNA) barcoding, population genetics and phylogeography, and DNA fingerprinting (Dormontt *et al.* 2015). Importantly, genetic approaches have demonstrated the ability to identify timber at the species level. Nevertheless, these approaches have considerable drawbacks, including high costs, time-consuming procedures, resource-intensive requirements, and the need for high-quality DNA samples (Dormontt *et al.* 2015; Schmitz *et al.* 2020). In addition, stable isotope analysis is employed to determine the geographical origin of timber rather than to identify timber species. This technique uses unique isotopic signatures that vary across regions, thereby allowing for precise geographical identification (Dormontt *et al.* 2015; Watkinson *et al.* 2022).

In addition, mass spectrometry is capable of identifying timber at the species level (Dormontt *et al.* 2015). Direct analysis in real-time (DART) time-of-flight mass spectrometry (TOFMS) is widely used in timber identification, as it is equipped with well-developed databases and adjoined software (Schmitz *et al.* 2020). This technique analyses phytochemicals present in the heartwood of a tree in real-time (Cody *et al.* 2005). Through the application of statistical analysis, chemical profiles representing the unique phytochemical composition of heartwood are generated, allowing researchers to identify timber at various taxonomic levels (Dormontt *et al.* 2015). However, this approach requires substantial resources and specialised equipment (Zhang *et al.* 2019). Additionally, the cost of DART-TOFMS instrumentation is high (Dormontt *et al.* 2015).

Finally, NIR spectroscopy is a commonly used method for timber identification (Dormontt *et al.* 2015; Schmitz *et al.* 2020). NIR analyses the phytochemicals of a tree, resulting in the generation of unique wood absorption spectra when the wood is exposed to NIR light (Dormontt *et al.* 2015). This enables researchers to elucidate the properties of the wood and perform timber identification (Dormontt *et al.* 2015; Schmitz *et al.* 2020). Moreover, handheld NIR spectroscopy has been shown to achieve species-level identification during field testing (Snel *et al.* 2018; Raobelina *et al.* 2023). However, the price of NIR spectroscopy has been reported to be costly, ranging from USD 1,000 to USD 75,000 (Schmitz *et al.* 2020).

In summary, macroscopic anatomical analysis, microscopic anatomical analysis, and MV software utilise wood anatomical features in timber identification tasks. However, the first two methods are noted for their challenges in identifying timber down to the species level, while the latter method is reported to have potential for achieving species-level identification. Likewise, other methods, such as genetics, mass spectrometry, and NIR, are also noted for their capability to identify timber down to the species level. However, these methods have significant drawbacks, primarily their high cost. Additionally, genetics and mass spectrometry require substantial resources during the sample preparation process. Therefore, among all methods, MV software stands out as a reliable alternative tool for timber identification.

Although it comes with certain limitations, the utilisation of software in timber identification is a viable approach, as it eliminates human error. In addition, this method also possesses species-level identification capability. Furthermore, unlike mass spectrometry and NIR, MV software does not require specialised devices. Instead, the MV model has the option to perform timber classification *via* a cloud-based interface (Figueroa-Mata *et al.* 2022). Finally, unlike genetics and mass spectrometry, MV software does not require substantial resources for its development. Most of the effort in developing MV software is focused on model development, which primarily requires datasets and technical knowledge (Hwang and Sugiyama 2021).

Recognising the importance of accurate timber identification and the potential advancements offered by MV technology, this paper presents a comprehensive review of MV software for timber identification. This review highlights key aspects of the application of MV in timber species identification. Firstly, it examines various methods of timber identification, critically evaluating and summarising the advantages and limitations of each method. Second, it explores MV software and its supporting technologies. Third, this paper reviews the potential of convolutional neural networks (CNNs) as models for MV software and summarises reported findings on the performance of CNN models in timber identification tasks. Additionally, it examines and synthesises current practices adopted by researchers in developing CNN models for timber identification. Finally, this paper offers a critical discussion of the underlying challenges associated with the development of CNN models for timber identification tasks.

METHODOLOGY

To ensure a comprehensive and transparent synthesis of the current state of wood species identification and ML applications, a systematic literature search was conducted. This section details the search strategy, eligibility criteria, and the multi-stage selection process employed.

Literature Search Strategy

The search was designed to capture a broad spectrum of research published between 2005 and 2025, a period characterised by the rapid evolution of digital image processing and deep learning in forestry. The search focused on peer-reviewed journals, conference proceedings, preprints, and grey literature to mitigate publication bias. The primary search was executed across the following databases and institutional repositories:

Table 1. Databases and Platforms

Source Category	Targeted Databases and Platforms
Scientific Databases	Scopus, Web of Science (WoS), ScienceDirect, IEEE Xplore, ACM Digital Library
Specialized Publishers	Springer Nature (*Wood Science and Technology*), Elsevier (*Smart Agricultural Technology*), Frontiers Media, MDPI (*Forests*), Brill (*IAWA Journal*), De Gruyter (*Holzforschung*)
Preprints and Repositories	arXiv.org, ResearchGate
Institutional Archives	Forest Research Institute Malaysia (FRIM), Sarawak Timber Industry Development Corporation (STIDC), Preferred by Nature (NEPCon)

The search string utilised a combination of Boolean operators (AND/OR) to link domain-specific terminology with computational methods: Search String: (“wood identification” OR “timber identification” OR “wood species classification”) AND ("machine learning" OR "deep learning" OR "computer vision" OR "neural network") OR ("macroscopic identification" OR "microscopic wood anatomy" OR "DART-MS" OR "NIRS" OR "stable isotope analysis").

Inclusion and exclusion criteria

Rigid eligibility criteria were established to ensure that the reviewed studies were technically relevant and provided verifiable data. To be included in this review, studies were required to propose or evaluate methods, algorithms, or frameworks designed specifically for wood or timber species identification. Furthermore, the selected literature had to explicitly utilise computer vision, machine learning, or deep learning architectures. It was also imperative that the incorporated studies provided empirical results accompanied by clear performance metrics, such as accuracy or F1-scores, to allow for an objective assessment of the proposed models. Finally, the scope of inclusion was restricted to peer-reviewed articles, conference papers, and technical reports published by recognised forestry institutions to maintain the academic rigor of the review.

Conversely, several criteria were established to exclude literature that did not align with the specific technical focus of this study. Research on general forest management, ecology, or silviculture was excluded if it lacked a clear diagnostic identification component. Additionally, papers were omitted if they lacked sufficient methodological detail or failed to provide adequate descriptions of the datasets utilised. To prevent duplication of data and ensure the highest quality of sources, redundant records, such as preprints for which a subsequent peer-reviewed version was available, were removed from the selection pool. Finally, publications written in languages other than English or those for which the full-text manuscript remained inaccessible were excluded from the final synthesis.

Study Selection Procedure

The selection of literature for this study followed a sequential and iterative screening process, wherein each retrieved source was evaluated individually against the specific research objectives. The procedure commenced with database querying, which involved executing the predefined keyword combinations across the specified list of databases. Following the initial retrieval, a sequential evaluation was conducted for each search result. This phase consisted of a rigorous three-step manual verification process prior to the inclusion of any study in the review. First, a preliminary appraisal of the titles and abstracts was performed, followed by deduplication to eliminate overlapping records sourced from multiple platforms. Subsequently, a comprehensive content relevance check was conducted to rigorously verify that the full text of the articles satisfied all established inclusion criteria. Finally, upon successfully passing these screening phases, the eligible studies were formally included, downloaded, and systematically archived in a local directory for subsequent reference and synthesis.

SUPPORTING TECHNOLOGIES OF MV SOFTWARE

Computer Vision

Computer vision (CV) is a field of study that aims to enable computer systems to see, identify, and understand the visual world like human perception. Over time, researchers have developed a wide range of algorithms to address various visual perception tasks, such as object detection and object recognition (Feng *et al.* 2019). As a result, CV is widely utilised in various domains: (i) optical character recognition (OCR), (ii) machine inspection, (iii) retail, (iv) warehouse logistics, (v) medical imaging, (vi) self-driving vehicles, (vii) photogrammetry, (viii) match move, (ix) motion capture (mocap), (x) surveillance, and (xi) fingerprint recognition and biometrics (Szeliski 2022). Additionally, CV has been applied in the field of timber identification, leading to the approach now commonly referred to as CV-based wood identification.

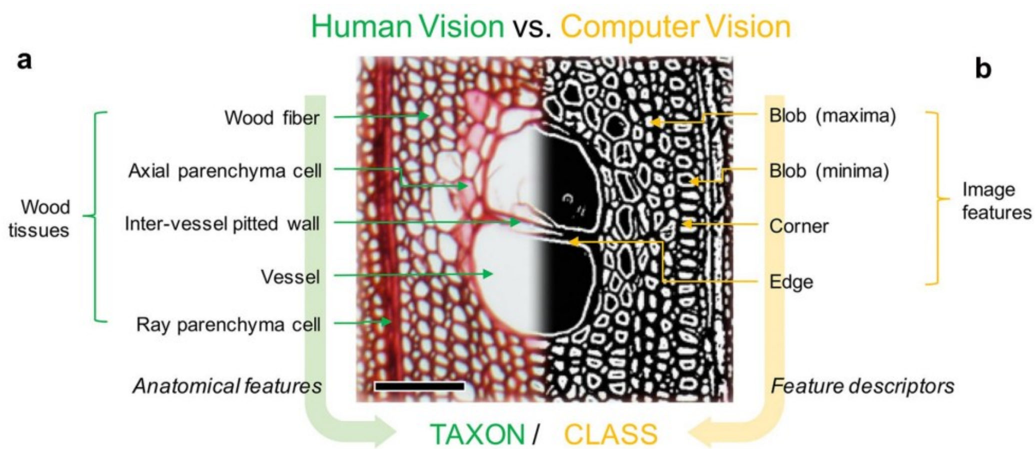

Fig. 1. Comparison: Human *vs.* computer vision features (Hwang and Sugiyama 2021)

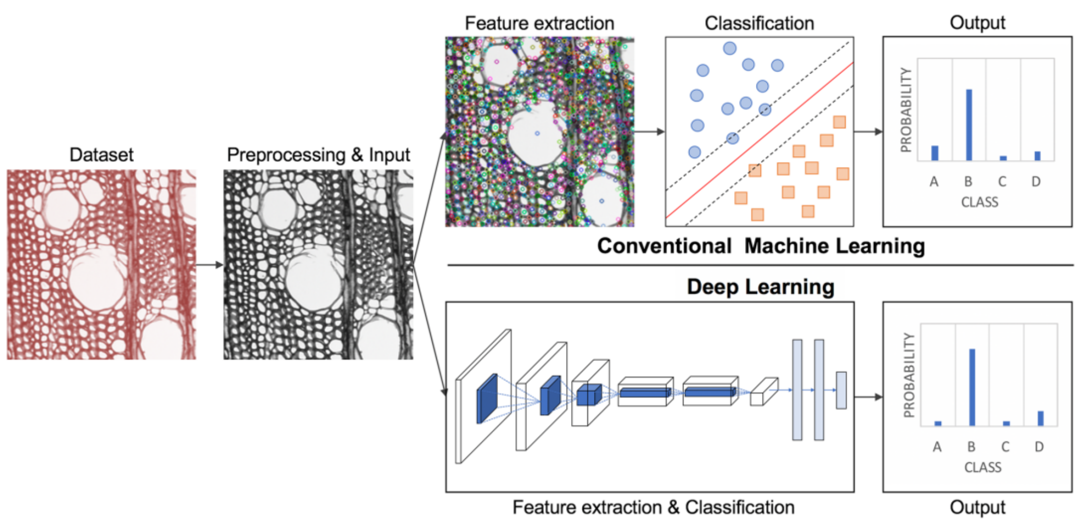

Fig. 2. General workflow of conventional ML and DL (Hwang and Sugiyama 2021)

According to Silva *et al.* (2022), CV-based wood identification is a system that integrates CV and artificial intelligence (AI). In the context of wood identification, CV detects and extracts relevant wood features, such as points, blobs, and edges (Fig. 1), while AI applies ML algorithms to learn and classify wood species based on the extracted features (Hwang and Sugiyama 2021). According to Hwang and Sugiyama (2021), CV-based wood identification can be divided into conventional ML and deep learning (DL) methods. In conventional ML, feature extraction and classification are performed independently. Meanwhile, DL performs feature extraction and classification jointly within a single process, as depicted in Fig. 2 (Hwang and Sugiyama 2021). It is important to note that DL is a subset of ML that emerged as a result of advancements in ML techniques (Shinde and Shah 2018). The color differences shown in Fig. 1 and Fig. 2 are intended for sample purposes only and do not imply that CV systems are limited to grayscale or binary images. In practical computer vision applications, images may be processed in various color spaces (*e.g*., RGB, grayscale, HSV, or binary) depending on the analysis objectives. For wood anatomy studies, image transformations are commonly applied to enhance the visibility of anatomical structures and improve feature extraction.

Machine Learning

According to Silva *et al.* (2022), conventional ML is a software that recognises and learns patterns from pre-processed input images to define the descriptive structure of unknown training datasets. In the context of wood identification, conventional ML comprises three stages, namely image acquisition, image datasets, and image processing. Image acquisition involves building the required dataset to train the algorithm, which consumes a considerable amount of time and effort (Hwang and Sugiyama 2021). However, readily available databases provide large amounts of wood samples.

Some of these databases include CAIRO and FRIM, which contain stereogram images of commercial hardwood species in Malaysia; LignoIndo, which contains stereogram images of commercial hardwood in Indonesia; and Federal University of Parana (UFPR), which contains macroscopic and micrograph images of Brazilian woods (Hwang and Sugiyama 2021). However, despite the availability of such databases, the lack of free access to global wood image datasets hinders the advancement of CV-based wood identification systems (Silva *et al.* 2022).

The image processing stage of conventional ML consists of two independent steps, which are feature extraction and classification. Feature extraction requires a preliminary step, known as pre-processing, to reduce computational complexity (Hwang and Sugiyama 2021). Various techniques are used in pre-processing, including grayscale conversion, image cropping, image scaling, filtering, and sharpening (Hwang and Sugiyama 2021; Silva *et al.* 2022). The pre-processed dataset is then divided into three subsets, which are training, validation, and test sets, before undergoing the feature extraction process (Silva *et al.* 2022).

Feature extraction is a process by which important features of wood samples are detected and extracted. In wood identification, the most commonly selected features are texture features and local features (Hwang and Sugiyama 2021). Texture features refer to visual patterns that contain information about the spatial arrangement of pixel intensities in an image. Meanwhile, local features refer to structural elements such as points, corners, and edges (Hwang and Sugiyama 2021). Accordingly, a wide range of feature extraction algorithms has been employed, including the gray-level co-occurrence matrix (GLCM), basic gray-level aura matrix (BLGAM), local binary pattern (LBP), and scale-invariant feature transform (SIFT) (Silva *et al.* 2022) (Hwang and Sugiyama 2021).

Finally, the dataset undergoes the classification process, where the extracted features are learned by classification models to establish classification rules that are applied to unknown datasets (Hwang and Sugiyama 2021; Silva *et al.* 2022). In addition, the three most preferred classifiers in wood identification studies are k-nearest neighbor (KNN), support vector machine (SVM), and artificial neural networks (ANNs) (Hwang and Sugiyama 2021).

Deep Learning

Deep learning is a subset of ML that has been developed during the advancement of ML (Shinde and Shah 2018). It is one of the most prominent and promising branches of ML, as it has the capacity to learn from and process large volumes of data (Silva *et al.* 2022). According to Silva *et al.* (2022), five DL architectures have been identified: ANNs, deep neural networks (DNNs), recurrent neural networks (RNNs), deep reinforcement learning (DRL), and CNNs. Following this, ANN and CNN are mostly implemented in the context of wood identification. Deep learning enables automated timber identification through a structured, data-driven pipeline consisting of four key stages: image acquisition, preprocessing, model training, and evaluation. High-quality datasets were first collected at either macroscopic (*e.g.*, smartphone-based cross-sections) or microscopic levels, where factors such as lighting, resolution, and surface preparation critically influenced model performance. The images were then standardised through preprocessing steps such as cropping, resizing, and normalization. They were then augmented using transformations to improve generalisation and reduce overfitting. During model training, CNNs automatically learned hierarchical features from labelled data, progressing from basic edges to complex anatomical characteristics such as vessel structures, ray patterns, and growth rings. Finally, the model's reliability was assessed by partitioning the dataset into training, validation, and testing subsets, ensuring unbiased evaluation using unseen data and providing a realistic measure of its practical applicability in timber identification.

According to Hwang and Sugiyama (2021), CNNs are developed by applying filtering techniques to ANNs, which makes CNNs more effective for image processing (Hwang and Sugiyama 2021). Numerous studies on wood identification have utilised CNN architectures, leading to varying levels of performance across CNN models. Notably, classification accuracy in wood identification using CNN architectures was high for most reported models (Hwang and Sugiyama 2021). However, CNNs require large numbers of labeled wood images, which are often difficult to obtain (Hwang and Sugiyama 2021).

The CNNs achieved a breakthrough during the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. The CNNs outperformed traditional methods in image classification (Russakovsky *et al.* 2015). Additionally, CNNs have been reported as a common approach in the field of timber identification (Silva *et al.* 2022). Consequently, exploring and harnessing CNNs for timber identification tasks can significantly enhance MV software. The development of MV software for timber identification requires a core component known as the model, which serves as the fundamental element of the system (Schmitz *et al.* 2020). Integrating CNNs as models within MV software can further improve the accuracy and efficiency of the timber identification process.

POTENTIAL OF CNN ARCHITECTURES IN TIMBER IDENTIFICATION

Various CNN architectures have been explored by researchers, resulting in a wide range of performance outcomes. Nevertheless, the majority of studies consistently demonstrate the effectiveness of CNN-based approaches for accurate timber species identification. Table 2 summarises key information from previous studies in a clear and organised manner. It includes details about the datasets used, image processing methods, and data preparation and splitting procedures. In addition, Table 2 shows the CNN models and training settings used in each study, along with the evaluation metrics for performance comparison. Finally, the key findings highlight the main contributions of each study.

Table 2. CNNs in Timber Identification

Authors	Dataset		Image Processing			Dataset splitting	CNN Architectures						Evaluation metrics	Key findings
	Total images	Image acquisition device	Image pre-processing	Image size	Image type		CNN type	Models	Optimizer	Learning rate	Batch size	Epochs		
(Wu *et al.* 2021)	52716, 3000	Microtec Goldeneye 300 Multi-Sensor Quality Scanner. Smart-phone (OnePlus3) with 20x	Flip, Rotation, Gray-scale, Up-sampling, Patch division, Resizing, Cropping	224x224, 300x300	Composite	Stratified K-Fold Cross Validation (k=5), training, validation and testing (7:1:2), training and testing (5:1)	Base-line	ResNet50, DenseNet 121, MobileNet V2, VGG16, VGG19, Xception, Inception V3	Adam and SGD	0.045 to 0.025, 0.001	64, 64	30 to 10, to 5	Recall, Accuracy, Precision, Macro F1, Confusion, F1 score	Single model Acc = 95% ResNet50 + LDA and KNN Acc = 98% (ensemble model)
(Olschofsky & Köhl 2020)	300	Konika-Minolta Bizhub Scanner c224 (600 dpi)	None	263x263	Normal, high resolution	Bootstrapping (5 folds), training and testing (4:1)	Base-line	Inception V3	-	-	-	-	Accuracy, Confusion matrix, Probability threshold-ing	Acc without probability thresholding = 98% Acc with probability thresholding = 93%
(Lens *et al.* 2020)	2240	UFPR Database (Olympus CX40 microscope (100x magnification)	None	1024x768	Microscopic	Stratified 10-fold Cross Validation, training and validation (9:1)	Base-line	AlexNet, VGG16, GoogleNet, ResNet101	-	0.001	-	-	Accuracy	ResNet101 achieved highest Acc: Acc at species level = 96% Acc at genus level = 96%

(de Geus *et al.* 2020)	n. a	Stereo-microscope Olympus SZX7 with 20x MF.	Rotation, Mirroring, Bright-ness jittering, Colour noise	299x299	Stereoscopic	Training, validation and testing (8:1:1)	Base-line	ResNet, DenseNet, SqueezeNet, Inception V3	SGD	-	-	Early stopping	Accuracy	DenseNet achieved highest Acc = 98.75% Other architectures Acc > 97%.
(Sun *et al.* 2021)	-	Magnifying glass	-	-	Macroscopic	-	Custom	ResNet50 (+ LDA and KNN)	-	-	-	-	Diffp, DiffF1, Accuracy, Precision, Recall, F1 score	Acc = 99.6% accuracy F1 score = 0.994 Diffp Acc = 0.006 DiffF1 score = 0.0009.
(Figueroa-Mata *et al.* 2022)	147742	5 Megapixel Celestron 20X MF electronic stereo-scope	Shifting, Rotation, Flipping, Zooming, Patch division	224x224	Stereoscopic	Training, validation and testing (7:2:1)	Custom	VGG16 pre-trained with ImageNet. Softmax layer removed. Add GAP layer, 2 dropout layers and 2 dense layers.	SGD	0.01	32	-	MRR, Accuracy, Voting Rule	Top-1 Acc = 76.5% Top-3 Acc = 80.3%
(Moulin *et al.* 2022)	-	Microscope 50x MF.	Shifting, Flipping, Resizing, Rotation, Patch division	646x486, 2448	Microscopic	Stratified K-Fold Cross Validation (k=5), training and validation (4:1)	Custom	Custom CNN inspired by AlexNet	Adam	0.01	-	Early stopping	Precision, Accuracy, Recall, F1 score, Confusion matrix	Acc > 90% F1 scoreup to 0.99 for some instances.
(Tang *et al.* 2018)	2100000	Smart-phone with 20x macro lens	Rotation	256x256	Macroscopic	Training and testing (6:1)	Custom	SqueezeNet (modified) +1	SGD	0.002	64	30	Recall, Accuracy, Fl-score, Precision	Top-2 and Top-1 Acc: Species-level (77.52%, 87.29%), Family-level (84.04%, 90.77%).

(Bello *et al.* 2023)	4272	WOOD-AUTH dataset (standard digital imaging)	None	512x512	Macroscopic	Training and testing (7:3)	Hybrid	Mask RCNN-ResNet +1	-	0.001	200	5	AP, Recall, Precision, IoU, MAP	mAP of 0.92 outperforms existing work methods on Based Wood Identification
(Holmström *et al.* 2023)	4930	Samsung Galaxy A50 (25MP main camera)	Image segmentation	512x512	Normal	Training, validation and testing (296:98:99)	Hybrid	DenseNet20, EfficientNetB4, InceptionResNetV2, Inception V3, MobileNet V2, Xception, ResNet152V2	Adam	0.0001	32	120	Rank-1 Accuracy	Rank-1 Acc random test set rotation= 84% Acc with fixed rotation = 91%

Dataset

Wood samples preparation

Images of wood samples were obtained either from readily available databases or from raw wood specimens, which underwent various processing steps before image acquisition. For wood samples obtained from online databases, sample preparation was not required, as the wood sample images are already well-prepared and ready for ML development. This approach is exemplified in Bello *et al.* (2023) and Lens *et al.* (2020), who obtained wood sample images from the WOOD-AUTH and UFPR databases, respectively.

Meanwhile, other studies performed their own wood sample collection. The collected wood samples were in raw form and thus required certain processing steps to prepare them for imaging. The processing steps are generally divided into two categories, depending on the type of images the researchers intended to acquire: laboratory processes and non-laboratory processes. Generally, laboratory processes involve a series of procedures. For instance, Brazilian wood samples in Moulin *et al.* (2022) were softened by boiling in deionised water. The softened wood samples were then sectioned using a microtome to a thickness of 20 µm. Finally, the specimens were dehydrated *via* a series of alcohol concentrations and eventually colored using safranine dye (Moulin *et al.* 2022). In contrast, non-laboratory processes are generally simpler and include any combination of cleaning, sectioning, cutting, drying, sanding, and polishing (Tang *et al.* 2018; de Geus *et al.* 2020; Olschofsky and Köhl, 2020; Sun *et al.* 2021; Wu *et al.* 2021).

Laboratory-based preparation methods may improve feature visibility and support more accurate species-level classification. However, such procedures increase sample preparation complexity and cost, which may limit the scalability of datasets required for ML and DL applications. In contrast, non-laboratory approaches enable more efficient large-scale image acquisition and are better suited for practical field deployment. Therefore, the selection of an image acquisition protocol should consider the trade-off between classification performance and dataset scalability.

Image acquisition

After processing, the wood samples were used for imaging. The selection of imaging devices depended on the type of image the researchers intended to acquire for training CNN models. Commonly used image types in timber identification are macroscopic image, X-ray computed tomography (CT) image, stereogram image, and micrograph image, as depicted in Fig. 3 (Hwang and Sugiyama 2021).

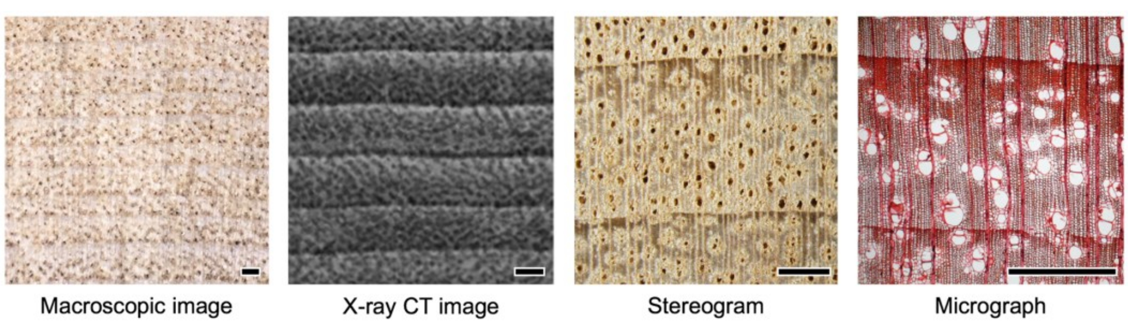

Fig. 3. Most used image types in timber identification (Hwang and Sugiyama 2021)

Figueroa-Mata *et al.* (2022) and de Geus *et al.* (2020) utilised stereogram images, which required specialised image acquisition devices. Meanwhile, Moulin *et al.* (2022) utilised micrograph images captured using a microscope at 50× magnification. In addition, Sun *et al.* (2021) and Tang *et al.* (2018) employed macroscopic images, which can be obtained using a smartphone camera equipped with a macroscopic lens.

In contrast, Tang *et al.* (2018) and Sun *et al.* (2021) captured macroscopic images using smartphones equipped with 20× macroscopic lens. However, different types of smartphones were utilised in both studies.

Image Processing

Image pre-processing and augmentation

Once the wood sample images were captured, the next step was image pre-processing. The raw images typically require pre-processing to remove unnecessary parts in the image. For instance, Sun *et al.* (2021) cropped each image to a size of 300 × 300 pixels, focusing on the central area of the image, as it was clearer and less fuzzy than other areas. Following image pre-processing, the pre-processed images were augmented to increase the size of the dataset, reduce overfitting, and address class imbalance (Figueroa-Mata *et al.* 2022). Researchers have applied various image augmentation processes to timber images, including rotation, flipping, mirroring, shifting, resizing, brightness adjustment, color noise jittering, and patch division (Tang *et al.* 2018; Moulin *et al.* 2022).

Image size

Input image resolution plays a critical role in CNN performance. As shown in Table 2, the pixel size or resolution of the input images used in the existing studies varies. The studies by Bello *et al.* (2023) and Holmström *et al.* (2023) used 512 x 512 pixels, Figueroa-Mata *et al.* (2022) and Wu *et al.* (2021) used 224 x 224 pixels, Moulin *et al.* (2022) used 646 x 486 pixels, Sun *et al.* (2021) used 300 x 300 pixels, Olschofsky and Köhl (2020) used 263 x 263 pixels, Lens *et al.* (2020) used 1024 x 768 pixels, de Geus *et al.* (2020) used 299 x 299 pixels, and Tang *et al.* (2018) used 256 x 256 pixels. The resolution of input images in six studies ranges from 224 to 300 pixels, while the remaining four studies use resolutions of 512 pixels and above. While there seems to be no standard resolution for input images in developing deep learning models, it should be noted that the resolution of the input image plays a pivotal role in determining model performance (Richter *et al.* 2021; Tan and Le 2019).

Higher resolution generally improves accuracy by preserving more discriminative features (Tan and Le 2019; Richter *et al.* 2021), while downscaling significantly degrades performance across architectures such as VGG16, ResNet, and EfficientNet. However, this relationship is not strictly linear. Richter *et al.* (2021) showed that even when image detail remains constant (*e.g.*, upscaling CIFAR-10 from 32×32), performance can still improve, suggesting that factors beyond feature detail contribute to accuracy. Conversely, excessive upscaling (*e.g.*, 224×224 to 1024×1024) degrades performance, suggesting that overly large input sizes may introduce inefficiencies or noise. Therefore, an optimal resolution range is crucial, balancing feature representation and model efficiency.

CNN Model

CNN is a type of deep learning model. The basic building blocks of CNN consist of three types of layers: convolution, pooling, and fully connected layers. Convolution and pooling layers are responsible for feature extraction, while the fully connected layer maps the extracted features to the final output (Reza *et al.* 2026; Zhao *et al.* 2024). The section discusses CNN architectures that are utilised in existing timber identification tasks, as summarised in Table 2.

Once the dataset is prepared, model development occurs. Model development is the process of building the CNN model and may involve several approaches. These approaches include building a model from scratch, directly applying readily available CNN models using high-level application programming interfaces (APIs), which are considered baseline models in this paper, customising the layers of baseline models to create custom models, or integrating two different baseline models to form hybrid models.

The CNN architectures utilised in the timber identification tasks listed in Table 2 can be grouped into three categories: baseline, custom, and hybrid. Baseline refers to the original architecture without any modifications. Custom involves modified architectures, where certain layers are altered by the researchers. In contrast, hybrid combines two distinct CNN architectures into a single model.

Model training/testing

Model training is typically a recurring process that involves multiple tests with different configurations to determine the optimal configuration for the developed model.

Dataset splitting is the process of splitting a dataset into subsets, and it is crucial in deep learning model development to avoid over-fitting (Babaei *et al.* 2025). Table 2 indicates the use of different ratios for different subsets. It is observed that datasets are split into three variations: (i) training, testing, and validation, (ii) training and testing, and (iii) training and validation.

The training set is used to build the model *via* multiple model parameters, while the validation set is used to fine-tune the model parameters (Xu and Goodacre 2018; Ying 2019). On the other hand, the testing set is used to evaluate the model's performance after the training and validation phases, as it consists of "unseen" data that was not used during model development (Xu and Goodacre 2018; Ying 2019).

Most studies allocate the largest portion of data for training, commonly using 70/30 or 80/20 splits for small datasets, while larger datasets often adopt higher training ratios (*e.g.*, ≥90%) to enhance model learning. Validation and testing splits are typically similar or balanced (*e.g.*, 70/15/15 or 80/10/10), reflecting consistent practices across recent works (Muraina 2022).

Other factors, such as the optimisation of the algorithm and the integration of dropout layers, also play essential roles in model training. Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (ADAM) are the universally preferred optimisers, with learning rates empirically set between 0.01 and 0.0001. Early stopping is increasingly utilised instead of fixed epoch counts. For instance, Holmström *et al.* (2023) utilised hyperparameter tuning to determine the optimal learning rate for the model. They also reported that introducing a dropout layer decreased model performance. In contrast, Moulin *et al.* (2022) adopted a learning rate reported in previous studies.

Figueroa-Mata *et al.* (2022) evaluated multiple optimisation algorithms, including SGD, ADAM, Adaptive Delta (AdaDelta), and Root Mean Square Propagation (RMSProp), and identified SGD as the most effective optimiser.

Overall, no *de facto* standard for model training can be identified based on current studies.

Model evaluation metrics

Trained models have been evaluated using various standard evaluation metrics. Common metrics employed to evaluate DL models include accuracy, precision, recall, and F1 score. For instance, Tang *et al.* (2018) and Sun *et al.* (2021) implemented these metrics when evaluating their models for timber identification tasks. Precision represents the proportion of true positive predictions among all positive predictions made by the model. Similarly, recall, also known as sensitivity, measures the proportion of true positive predictions among all actual positive instances in the dataset. The F1 score is calculated based on the model's precision and recall of the model, representing their harmonic mean.

In addition to these metrics, the confusion matrix has been widely employed in studies such as Moulin *et al.* (2022) and Wu *et al.* (2021). Confusion matrix is an N × N matrix, where N represents the number of classes. It is useful for evaluating model performance by visualising misclassifications and identifying which classes are most frequently confused.

In addition, Bello *et al.* (2023) employed Intersection over Union (IoU) and mAP, which are metrics commonly used to evaluate the performance of object detection models (Wang 2022; Rezatofighi *et al.* 2019). Bello *et al.* (2023) utilised a hybrid CNN model, integrating the Mask-RCNN architecture, which is an object detection model. Therefore, these two metrics were needed to assess the performance of the developed hybrid model.

In addition, Figueroa-Mata *et al.* (2022) employed top-k accuracy, top-k accuracy with the Voting Rule (VR), and mean reciprocal rank (MRR). The formulas for these metrices are shown in Eqs. 1, 2, and 3, respectively. These metrics are useful when CNN models are trained using image patches as well as complete images. Equation 1 defines the top-k accuracy of the CNN model, considering only patch images, where T represents the total number of patches and I denotes an individual patch. The term $hit(k, I)$ is a Boolean function that indicates whether one of the top-K candidate species in the ranking generated by the model corresponds to the correct identification of patch I.

Equation 2 represents the accuracy of the model when evaluated using complete images of the wood species, rather than subdivided patches. In Eq. 2, C represents the total number of images, and I' represents an individual image. Figueroa-Mata *et al.* (2022) noted that Eq. 2 provides a more realistic evaluation, as it considers the complete wood image instead of patches. For MRR, it is utilised to assess how effectively the model ranks the correct identification for each species in the dataset as shown in Eq. 3. In this context, $|Q|$represents the total number of images, and $rank_i$ denotes the ranking position assigned to the i-th image (Figueroa-Mata *et al.* 2022).

$$Accuracy_k = \frac{1}{|T|}\sum_{I \in T} hit(k, I) \tag{1}$$

$$VRAccuracy_k = \frac{1}{|C|}\sum_{I' \in C} VRhit(k, I') \tag{2}$$

$$MRR = \frac{1}{|Q|}\sum_{i \in Q} \frac{1}{rank_i} \tag{3}$$

Additionally, custom evaluation metrics have also been employed to evaluate the performance of CNN models in timber identification tasks. For example, Sun *et al.* (2021) utilised a custom metric, $Diff_p$, defined in Eq. 4, where P denotes the subset of accuracy, precision, recall, and F1 score. According to Sun *et al.* (2021), this custom metric allows the evaluation of the CNN model's generalisation performance. A small $Diff_p$ value indicates that the model performs well on both the training and testing sets. In contrast, large $Diff_p$ value suggests poor generalisation, indicating that the model may be overfitting to the training set (Sun *et al.* 2021).

$$Diff_P = P_{training} - P_{testing} \tag{4}$$

CNN architectures

Despite the wide range of CNN architectures employed in timber identification tasks, not all architectures performed well in distinguishing timber species. Researchers often seek the optimal model by comparing the performance of different models.

Table 3. Top-performing CNN Architectures in Timber Identification

CNN Architecture	Evaluation Metrics	References
Mask RCNN-ResNet	Precision Recall AP IoU mAP	(Bello *et al.* 2023)
InceptionResNetV2	Rank-1 Accuracy	(Holmström *et al.* 2023)
VGG16 was pre-trained on ImageNet. Softmax layer was removed. Add GAP layer, two dropout layers, and two dense layers.	Accuracy Voting Rule MRR	(Figueroa-Mata *et al.* 2022)
Custom CNN inspired by AlexNet	Accuracy Precision F1 score Recall Confusion matrix	(Moulin *et al.* 2022)
ResNet50	Macro F1 Accuracy Precision Recall Confusion	(Wu *et al.* 2021)
ResNet50 with LDA and KNN	Accuracy Precision Recall F1 score $\mathrm{Diff_p}$	(Sun *et al.* 2021)
InceptionV3	Accuracy Confusion matrix Probability thresholding	(Olschofsky and Köhl 2020)
ResNet101	Accuracy	(Lens *et al.* 2020)
DenseNet	Accuracy	(de Geus *et al.* 2020)
SqueezeNet	Accuracy Recall Precision F1-score	(Tang *et al.* 2018)

For instance, Lens *et al.* (2020) trained multiple CNN models, such as AlexNet, GoogleNet, VGG16, and ResNet101, and found that ResNet101 performed the best.

Similarly, de Geus *et al.* (2020) compared several CNN models, namely InceptionV3, DenseNet, ResNet, and SqueezeNet, which were pretrained on the ImageNet dataset. The study found that DenseNet achieved the highest accuracy among them. Only certain architectures demonstrated outstanding performance, and these architectures are discussed in detail in the subsequent section of this paper.

The CNN architectures listed in Table 3 represent the architectures that performed well in timber identification tasks. For studies utilising a single CNN architecture, that architecture was considered and is listed in Table 3. For studies evaluating multiple CNN architectures, only the best-performing architectures were included.

The first architecture on the list is Mask RCNN-ResNet, a hybrid architecture that combines Mask RCNN and ResNet (Bello *et al.* 2023). The overall architecture configuration is depicted in Fig. 4, in which the ResNet component in this architecture is ResNet101, modified in its filter size. In the original ResNet, the first convolution layer has a filter size of 7 × 7 (He *et al.* 2015). For the modified ResNet used in this study, this filter size was reduced to 3 × 3, as shown in Fig. 5 (Bello *et al.* 2023).

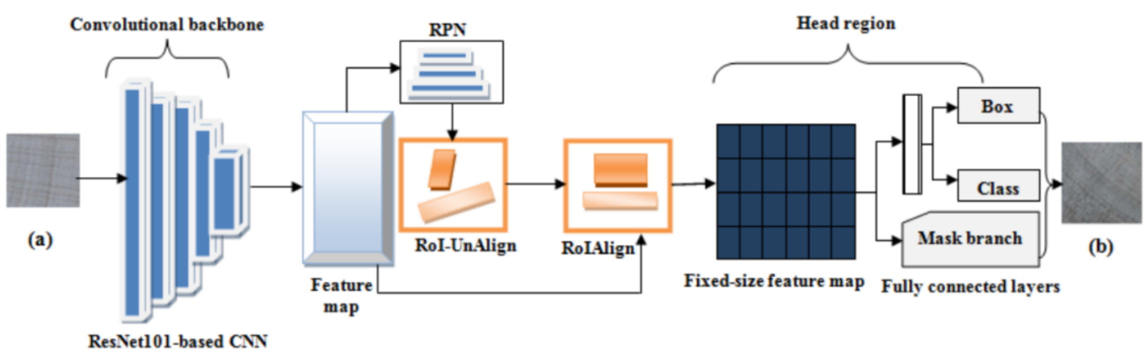

Fig. 4. Mask RCNN-ResNet architecture (Bello *et al.* 2023)

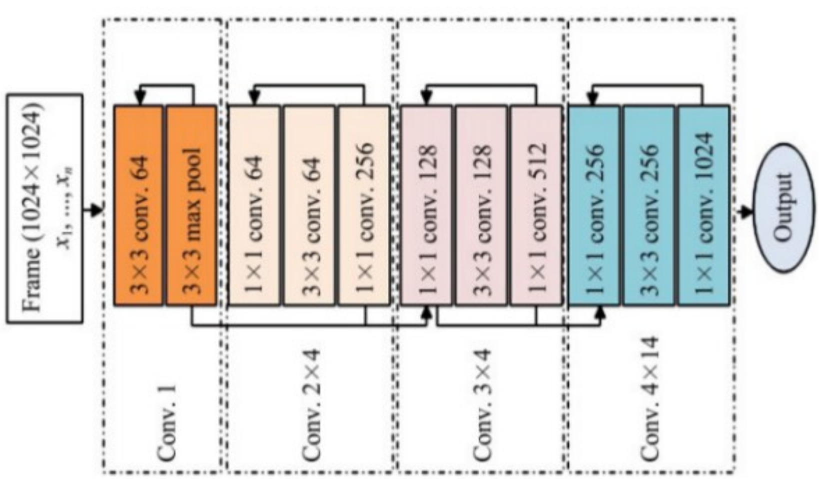

Fig. 5. Modified ResNet architecture (Bello *et al.* 2023)

Mask RCNN-ResNet was reported to achieve a mean average precision (mAP) of 0.92 in classifying the Greek wood species from the WOOD-AUTH database (Bello *et al.* 2023). The integration of the ResNet101 architecture as a feature extractor improved the mAP by 0.02 compared to the Mask R-CNN model alone, which was tested on the same dataset (Bello *et al.* 2023). Additionally, the modified ResNet101 utilized in this study had a smaller filter size to enable the extraction of smaller and more complex features. The inclusion of region proposals in the hybrid model allows the utilisation of multi-scale semantic features (Bello *et al.* 2023), contributing to model scale invariance. Finally, the hybrid model also utilises mask generation for the classification of individual wood (Bello *et al.* 2023), allowing for pixel-based classification.

The next CNN architecture on the list is InceptionResNetV2, a hybrid model combining the Inception and ResNet architectures. The overall architecture configuration is depicted in Fig. 6.

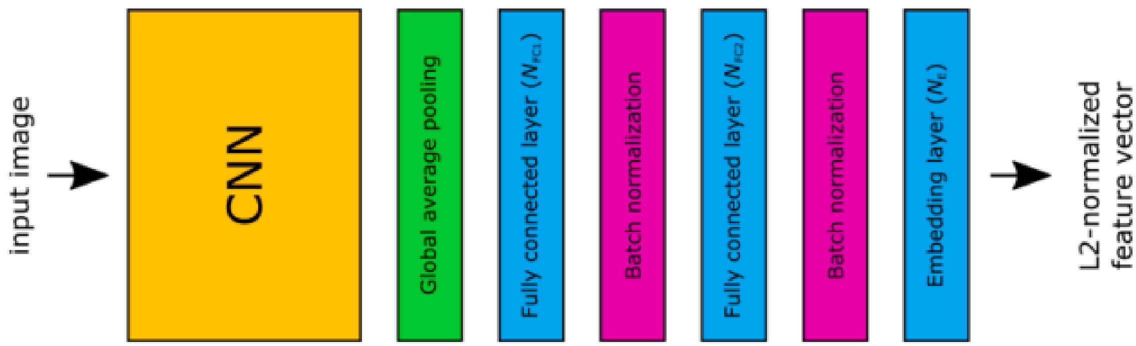

Fig. 6. InceptionResNetV2 (yellow block) with custom layers (Holmström *et al.* 2023)

The InceptionResNetV2 architecture serves as the feature extractor, while custom layers are embedded in the model for classification (Holmström *et al.* 2023). InceptionResNetV2 has been reported to outperform several other CNN architectures in identifying tree logs. Although this study emphasised tree log identification rather than timber species, the identification was particularly based on the evaluation of wood texture. Thus, the findings of this study are deemed relevant for discussion in this paper.

The third CNN architecture on the list is VGG16 with custom layers, including Global Average Pooling (GAP) layer, two dropout layers, and two dense layers (Figueroa-Mata *et al.* 2022). In addition, the softmax layer of the VGG16 model was removed (Figueroa-Mata *et al.* 2022). The resulting model was reported to achieve top-1 and top-3 accuracies of 76.5% and 80.3%, respectively, during the testing phase for identifying Costa Rican native tree species (Figueroa-Mata *et al.* 2022).

The fourth CNN architecture is a custom CNN inspired by the AlexNet architecture. The overall architecture configuration is shown in Fig. 7. The model was reported to achieve an accuracy greater than 90%, with an overall adjusted accuracy of 92.4% (Moulin *et al.* 2022). In addition, the F1 score reached up to 0.99 for some instances (Moulin *et al.* 2022). This custom CNN model was designed to classify Brazilian timber species.

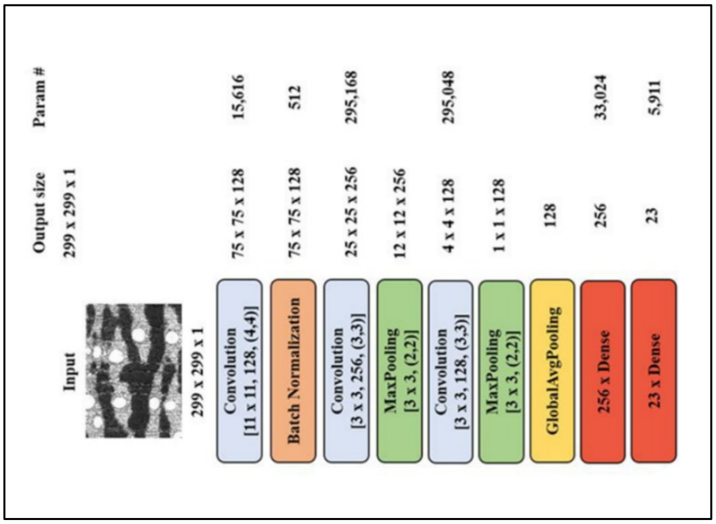

Fig. 7. Custom CNN inspired by AlexNet (Moulin *et al.* 2022)

The fifth architecture on the list is ResNet50, which serves as a baseline model. ResNet50 was introduced by He *et al.* (2015) to address the issue of increasing training error associated with the increment of layers in the network. In addition to ResNet50, numerous other ResNet variations are available, as depicted in Fig. 8.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 8. Variations of ResNet architectures (He *et al.* 2015)

ResNet50 architecture was reported to outperform two other CNN architectures, namely DenseNet121 and MobileNetV2, in identifying 11 common hardwood lumber species (Wu *et al.* 2021). In particular, ResNet50 achieved accuracies of 95.19% and 98.15% for patch and board-level identification, respectively. Nevertheless, DenseNet121 and MobileNetV2 also demonstrated competitive performance. DenseNet121 achieved accuracies of 93.84% and 97.55%, while MobileNetV2 achieved accuracies of 93.52% and 97.12% for patch and board-level identification, respectively (Wu *et al.* 2021).

The sixth architecture on the list is a custom CNN architecture that combines ResNet50 with linear discriminant analysis (LDA) for feature refinement and K-nearest neighbor (KNN) for classification. This combination was reported to achieve 99.6%

accuracy and 0.994 F1 score in identifying 25 rare wood species. Additionally, the study employed custom evaluation metrics, namely $\mathrm{Diff}_{\mathrm{accuracy}}$ and $\mathrm{Diff}_{\mathrm{F1\ score}}$, to evaluate model performance. Small values of these metrics indicate similar model performance on the training and testing sets. In contrast, greater values of these metrics indicate reduced performance on the testing set compared to the training set (Sun *et al.* 2021). The ResNet50 paired with LDA and KNN achieved $\mathrm{Diff}_{\mathrm{accuracy}}$ of 0.006 and $\mathrm{Diff}_{\mathrm{F1\ score}}$ of 0.0009 (Sun *et al.* 2021). Overall, this custom model combination demonstrated superior generalisation performance and accuracy (Sun *et al.* 2021).

The seventh architecture on the list is InceptionV3, which serves as a baseline model. InceptionV3 implements a sparse structure of CNN module, as depicted in Fig. 9 (Szegedy *et al.* 2015). This module enhances computational complexity by factorising convolutions with large filter sizes (Szegedy *et al.* 2015). InceptionV3 was reported to achieve an overall accuracy of 98% in identifying *Cedrella odorata*, a timber species listed under CITES. Additionally, when probability thresholding was applied, the overall accuracy of InceptionV3 decreased to 93%.

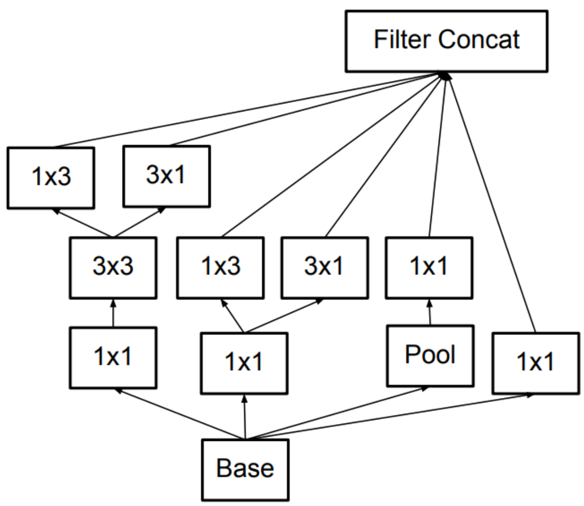

Fig. 9. InceptionV3 module (Szegedy *et al.* 2015)

The next CNN architecture on the list is ResNet101, which serves as a baseline model. ResNet101 is a variation of the ResNet architecture that differs primarily in the number of layers. It consists of 101 layers and, compared to ResNet50, generally has a larger number of parameters and a greater model size. The complete architecture configuration of ResNet101 was illustrated in Fig. 8. ResNet101 was reported to achieve remarkable accuracy, with species-level and genus-level accuracies of 96% and 99%, respectively, using wood species data from the UFPR database (Lens *et al.* 2020).

The ninth architecture on the list is DenseNet, which serves as a baseline model. In DenseNet, each layer receives input not only from its preceding layer but also directly from all previous layers, as depicted in Fig. 10. Additionally, the features are not combined before being passed to a layer; instead, they are concatenated (Huang *et al.* 2016). As a result, DenseNet architecture offers several advantages, including alleviating the vanishing gradient problem, strengthening feature propagation, reducing the number of parameters, and enabling feature reuse (Huang *et al.* 2016). DenseNet was reported to achieve the highest accuracy of 98.75% in identifying 281 Brazilian timber wood species, thereby surpassing the other three CNN architectures evaluated in the same study (de Geus *et al.* 2020).

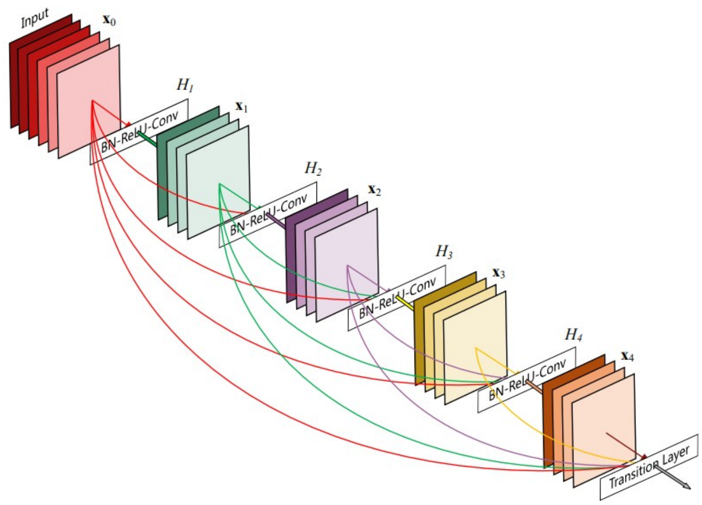

Fig. 10. 5-layer dense blocks with growth rate k = 4 (Wang 2022)

The final architecture on the list is SqueezeNet, which serves as a baseline model. SqueezeNet was first introduced by Iandola *et al.* (2016), with its main hallmark being the ability to achieve high accuracy in image classification tasks while maintaining a significantly reduced model size. Fig. 11 illustrates the SqueezeNet architecture. The architecture begins with a convolution layer (conv1), followed by eight fire modules (fire 2 through 9), and ends with a convolution layer (conv10). SqueezeNet was reported to achieve top-1 and top-2 accuracies of 77.52% and 87.29%, respectively, for species-level identification of 100 Malaysian timbers (Tang *et al.* 2018). In addition, it achieved top-1 and top-2 accuracies of 84.04% and 90.77%, respectively, for genus-level identification of 100 Malaysian timbers (Tang *et al.* 2018).

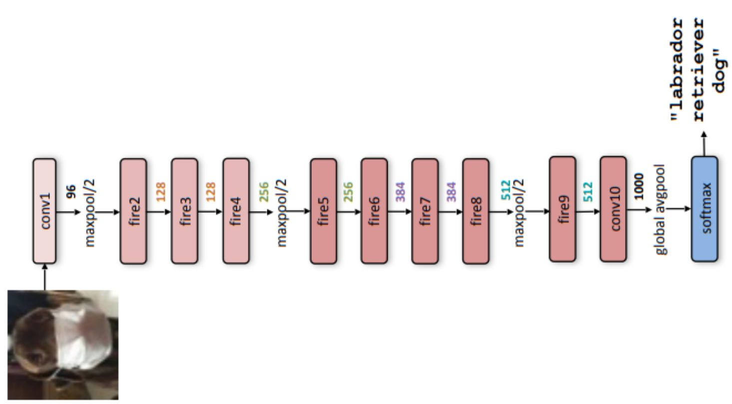

Fig. 11. SqueezeNet architecture (Iandola *et al.* 2016)

CHALLENGES AND FUTURE WORKS IN IMPLEMENTATION OF CNN IN MV SOFTWARE FOR TIMBER IDENTIFICATION

This paper has provided a comprehensive review of current methods in the field of timber identification, the potential of MV software, the utilisation and potential of CNN as models for MV software, and the challenges associated with the development of CNN models for timber identification tasks.

Although the diversity of methods employed in CNN development for timber identification is valuable, it poses challenges for comparing and integrating findings across studies. The immense variations in imaging devices, image types, datasets, pre-processing and data augmentation techniques, CNN models, training procedures, and evaluation metrics make direct comparison of findings from these studies difficult. This diversity highlights the richness of approaches in the field and underscores the need for a unified framework to enable meaningful comparison and integration of findings. Therefore, a comprehensive empirical study would provide valuable validation and establish clearer benchmarks, thereby advancing consistency and reproducibility across studies.

Dataset

There is a lack of standardisation in how wood samples are prepared and captured. Methods range from no preparation using smartphone cameras to highly rigorous microtome sectioning and colouring for microscopic analysis. This variability influences the complexity of the CNN required. Macroscopic, field-acquired images (*e.g.*, from smartphones) often require more robust preprocessing due to inconsistent lighting and environmental noise, whereas microscopic images provide highly structured anatomical data but are impractical for rapid field deployment. Therefore, the field must balance laboratory precision with real-world practicality.

Dataset sizes in the reviewed literature vary widely, ranging from as few as 300 images to as many as 2.1 million. Consequently, data augmentation techniques (*e.g.*, rotation, flipping, and cropping) have become a standard practice rather than an optional step. The heavy reliance on aggressive augmentation suggests that access to large, diverse, and publicly available wood anatomy datasets remains a significant bottleneck. Furthermore, while k-fold cross-validation and standard 70:30 splits are commonly used, researchers rarely evaluate their models on entirely independent, “out-of-distribution” datasets, leaving model vulnerability to domain shift largely untested.

Sun *et al.* (2021) and Tang *et al.* (2018) used macroscopic images, though the CNN models utilised in the two studies were different. Similarly, the sample preparation procedures varied, where Sun *et al.* (2021) used sandpaper and emery paper to flatten and polish the wood samples, whereas Tang *et al.* (2018) only utilised a cutter. Additionally, the data augmentation techniques differed, with Tang *et al.* (2018) applying rotation, brightness adjustment, mirroring, and color noise jittering, while Sun *et al.* (2021) did not implement any of these procedures.

Figueroa-Mata *et al.* (2022) utilised Costa Rican native timber species collected from forests along the Pacific Coast of Costa Rica. Similarly, Tang *et al.* (2018), conducted in Malaysia, utilised 100 commonly traded wood types found in the country. These studies highlight that the usability of CNN models for the timber identification task is constrained by the specific timber species utilised in each study, which, in turn, are determined by the location of CNN model development.

CNN model

Addressing the current challenges in the implementation of CNN in MV software requires a thorough understanding of the gaps identified in previous studies. Several gaps have been highlighted in the development of CNN models for timber identification. First, there is limited usability of CNN models. Given the vast diversity of wood species worldwide, it is practically impossible to develop a one-size-fits-all CNN model that can identify all timbers. Additionally, current studies have indicated that the development of CNN models for timber identification is highly influenced by the location in which the models are developed.

Moreover, current studies exhibit a high degree of flexibility. Different studies employ different CNN models. Bello *et al.* (2023) utilised Mask-RCNN ResNet, a hybrid model for timber identification. Meanwhile, Wu *et al.* (2021) utilised baseline models, including ResNet50, DenseNet121, and MobileNetV2, for timber identification. Similarly, researchers use different imaging devices, resulting in different image types. For example, Moulin *et al.* (2022) utilised a microscope at 50× magnification to capture micrographs or microscopic images.

There appears to be an evolutionary trend in the selected architectures. Earlier studies predominantly utilised baseline or custom models such as SqueezeNet and VGG16. By 2023, the literature demonstrates a distinct shift toward complex hybrid networks (*e.g.*, Mask RCNN-ResNet, InceptionResNetV2). However, ResNet variants remain the dominant foundational backbone across multiple years.

The transition to hybrid models indicates that single-architecture networks may be reaching performance ceilings. Researchers are increasingly prioritising models that combine the strengths of different networks, such as Inception's multi-scale feature extraction and ResNet's ability to mitigate vanishing gradients, to better handle highly complex or poorly illuminated wood grain patterns.

The widespread use of early stopping reflects efforts to prevent overfitting, which remains a high risk given the relatively small baseline datasets often used in timber identification. Hyperparameter selection remains largely empirical, suggesting a need for automated hyperparameter tuning to improve standardisation and establish more reliable performance benchmarks.

Reported accuracies are consistently high, frequently exceeding 95% and occasionally reaching up to 99.6%. While traditional accuracy is the most commonly reported metric, recent studies have increasingly adopted precision, recall, F1-score, and mean average precision (mAP). This trend suggests that an accuracy saturation point may have been reached in closed-dataset environments. As models achieve over 95% accuracy in controlled experimental settings, simple accuracy metrics are no longer sufficient to demonstrate model superiority. Consequently, the focus should shift toward assessing computational efficiency (e.g., inference speed), real-time deployability, and robustness in uncontrolled field environments.

The integration of CNN models into MV software has greatly improved the accuracy and efficiency of timber species identification. As these technologies continue to develop, they are expected to play a crucial role in safeguarding natural resources for the future. This supports biodiversity conservation, prevents illegal logging, and promotes sustainable forestry practices.

Based on the review, the field of deep learning for timber identification has rapidly matured from early proof-of-concept stages into a mature and technically refined research domain. Baseline performance is no longer the primary hurdle; CNNs, particularly ResNet and newer hybrid architectures, have consistently demonstrated the ability to identify wood species with expert-level accuracy under laboratory conditions. However, a critical gap identified in this review is the disconnect between laboratory performance and industrial applicability. The substantial variation in dataset sizes, image acquisition devices, and sample preparation methods highlights a fragmented research landscape that lacks standardised protocols.

Despite these advancements, a critical gap remains between laboratory success and real-world deployment. The significant variability in dataset sizes, image acquisition methods, and sample preparation reflects a lack of standardisation across studies, limiting the generalisability of reported results. Consequently, future research should prioritise practical applicability over incremental improvements in accuracy. Key directions include improving model robustness against environmental noise and domain shifts, establishing standardised and open-access timber image repositories for benchmarking, and optimising models for lightweight, edge-based deployment. Addressing these challenges is essential to enable reliable, real-time timber identification in operational settings such as forestry monitoring and customs enforcement.

ACKNOWLEDGMENTS

The authors would like to acknowledge Universiti Malaysia Sarawak (UNIMAS) for the support provided towards this publication. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no financial or personal relationships that could be perceived as potential conflicts of interest influencing the work reported in this paper. The authors confirm that there are no patents, stock ownership, employment, advisory board memberships, consultancy roles, or other financial interests related to the work described.

REFERENCES CITED

Bello, R.-W., Oluigbo, C. U., Tiebiri, C. B., Ogidiga, B. D., and Motunrayo, O. (2023). "Wood species identification using mask RCNN-residual network approach," *PRO LIGNO* 19(1), 41-51. https://ssrn.com/abstract=4410849

Cody, R. B., Laramée, J. A., Nilles, J. M., and Durst, H. D. (2005). "Direct analysis in real time (DART) mass spectrometry," *JEOL News* 40(1), 8-12.

de Geus, A. R., Silva, S. F. da, Gontijo, A. B., Silva, F. O., Batista, M. A., and Souza, J. R. (2020). "An analysis of timber sections and deep learning for wood species classification," *Multimedia Tools and Applications* 79(45–46), 34513-34529. https://doi.org/10.1007/s11042-020-09212-x

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

Feng, X., Jiang, Y., Yang, X., Du, M., and Li, X. (2019). "Computer vision algorithms and hardware implementations: A survey," *Integration, the VLSI Journal* 69, 309-320. https://doi.org/10.1016/j.vlsi.2019.07.005

Figueroa-Mata, G., Mata-Montero, E., Valverde-Otárola, J. C., Arias-Aguilar, D., and Zamora-Villalobos, N. (2022). "Using deep learning to identify Costa Rican native tree species from wood cut images," *Frontiers in Plant Science* 13, article 789227. https://doi.org/10.3389/fpls.2022.789227

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Deep residual learning for image recognition," *ArXiv Preprint* 1512, article 03385. http://arxiv.org/abs/1512.03385

Holmström, E., Raatevaara, A., Pohjankukka, J., Korpunen, H., and Uusitalo, J. (2023). "Tree log identification using convolutional neural networks," *Smart Agricultural Technology* 4, article 100201. https://doi.org/10.1016/j.atech.2023.100201

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2016). "Densely connected convolutional networks," *ArXiv Preprint* 1608, article 06993. http://arxiv.org/abs/1608.06993

Hwang, S. W., and Sugiyama, J. (2021). "Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review," *Plant Methods* 17, article 47. https://doi.org/10.1186/s13007-021-00746-1

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *ArXiv Preprint* 1602, article 07360. http://arxiv.org/abs/1602.07360

Kleinschmit, D., Mansourian, S., Wildburger, C., and Purret, A. (2016). *Illegal Logging and Related Timber Trade-Dimensions, Drivers, Impacts and Responses A Global Scientific Rapid Response Assessment Report*, International Union of Forest Research Organizations (IUFRO), Vienna, Austria.

Koch, G., Heinz, I., Haag, V., and Schmitt, U. (2019). "K-1: Wood anatomy - The role of macroscopic and microscopic wood identification to combat illegal logging and trading," in: *IAWA-IUFRO International Symposium: Challenges and Opportunities for Updating Wood Identification*, China, pp. 4–5.

Lens, F., Liang, C., Guo, Y., Tang, X., Jahanbanifard, M., da Silva, F. S. C., Ceccantini, G., and Verbeek, F. J. (2020). "Computer-assisted timber identification based on features extracted from microscopic wood sections," *IAWA Journal* 41(4), 660-680. https://doi.org/10.1163/22941932-bja10029

Moulin, J. C., Lopes, D. J. V., Mulin, L. B., Bobadilha, G. D. S., and Oliveira, R. F. (2022). "Microscopic identification of Brazilian commercial wood species *via* machine-learning," *Cerne* 28, 1-9. https://doi.org/10.1590/01047760202228012978

NEPCon. (2018). *Timber Legality Risk Assessment Malaysia-Peninsular*. www.nepcon.org/sourcinghub

Nordahlia, A. S., and Lim, S. C. (2016). "Introduction to basic wood identification," *Timber Technology Bulletin* 64, 1-12.

Olschofsky, K., and Köhl, M. (2020). "Rapid field identification of cites timber species by deep learning," *Trees, Forests and People* 2, article 100016. https://doi.org/10.1016/j.tfp.2020.100016

Raobelina, A. C., Chaix, G., Razafimahatratra, A. R., Rakotoniaina, S. P., and Ramananantoandro, T. (2023). "Use of a portable near infrared spectrometer for wood identification of four *Dalbergia* species from Madagascar," *Wood and Fiber Science* 55(1), 4-17. https://doi.org/10.22382/wfs-2023-03

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," (http://openaccess.thecvf.com/content_CVPR_2019/papers/Rezatofighi_Generalized_Intersection_Over_Union_A_Metric_and_a_Loss_for_CVPR_2019_paper.pdf), Accessed 01 Jan 2026.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., *et al.* (2015). "ImageNet large scale visual recognition challenge," *ArXiv Preprint* 1409, article 0575. http://arxiv.org/abs/1409.0575

Schmitz, N., Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J. W. B., Antonio Cabezas, J., Chaix, G., Crameri, S., Degen, B., *et al.* (2020). *Overview of Current Practices in Data Analysis for Wood Identification*, Global Timber Tracking Network, European Forest Institute and Thünen Institute, Joensuu, Finland. https://doi.org/10.13140/RG.2.2.21518.79689

Shinde, P. P., and Shah, S. (2018). "A review of machine learning and deep learning applications," in: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Prune, India, pp. 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697857

Siam, N. A., Abdullah, N. A., Anwar Uyup, M. K., Che Amri, C. N. A., Ahmad Juhari, M. A. A., and Talip, N. (2023). "Wood anatomical features of Anacardiaceae from Malaysia," *BioResources* 18(1), 1232-1250. https://doi.org/10.15376/biores.18.1.1232-1250

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

Snel, F. A., Braga, J. W. B., da Silva, D., Wiedenhoeft, A. C., Costa, A., Soares, R., Coradin, V. T. R., and Pastore, T. C. M. (2018). "Potential field-deployable NIRS identification of seven *Dalbergia* species listed by CITES," *Wood Science and Technology* 52(5), 1411-1427. https://doi.org/10.1007/s00226-018-1027-9

Sun, Y., Lin, Q., He, X., Zhao, Y., Dai, F., Qiu, J., and Cao, Y. (2021). "Wood species recognition with small data: A deep learning approach," *International Journal of Computational Intelligence Systems* 14(1), 1451-1460. https://doi.org/10.2991/ijcis.d.210423.001

Szegedy, C., Vanhoucke, V., Ioffe, S., and Shlens, J. (2015). "Rethinking the inception architecture for computer vision," *ArXiv Preprint* 1512, article 00567. https://arxiv.org/abs/1512.00567

Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*, Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-34372-9

Tang, X. J., Tay, Y. H., Siam, N. A., and Lim, S. C. (2018). "MyWood-ID: Automated macroscopic wood identification system using smartphone and macro-lens," in: *ACM International Conference Proceeding Series*, Phuket, Thailand, pp. 37-43. https://doi.org/10.1145/3293475.3293493

Timber Portal, Sarawak Timber Industry Development Corporation (2019). "Timber Portal," Timber Portal Sarawak Timber, (https://timberportal.sarawaktimber.gov.my/), Accessed 01 Jan 2026

Wang, B. (2022). "A parallel implementation of computing mean average precision," *ArXiv Preprint* 2206, article 09504. http://arxiv.org/abs/2206.09504

Watkinson, C. J., Rees, G. O., Gwenael, M. C., Gasson, P., Hofem, S., Michely, L., and Boner, M. (2022). "Stable isotope ratio analysis for the comparison of timber from two forest concessions in Gabon," *Frontiers in Forests and Global Change* 4, article 650257. https://doi.org/10.3389/ffgc.2021.650257

Wu, F., Gazo, R., Haviarova, E., and Benes, B. (2021). "Wood identification based on longitudinal section images by using deep learning. *Wood Science and Technology* 55(2), 553-563. https://doi.org/10.1007/s00226-021-01261-1

Zhang, M., Zhao, G., Guo, J., Wiedenhoeft, A. C., Liu, C. C., and Yin, Y. (2019). "Timber species identification from chemical fingerprints using direct analysis in real time (DART) coupled to Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS): Comparison of wood samples subjected to different treatments," *Holzforschung* 73(11), 975-985. https://doi.org/10.1515/hf-2018-0304

Article submitted: February 7, 2026; Peer review completed: February 28, 2026; Revised version received: April 21, 2026; Accepted: June 3, 2026; Published: July 15, 2026.
DOI: 10.15376/biores.21.3.Kipli