






# A Classification Method of Softwood Species for Building and Interior Decoration Based on Deep Learning

Zhikang Tian <sup>a,c</sup>, Xiaotong Liu <sup>a</sup>, Bei Zhang <sup>b</sup>, Zhedong Ge <sup>a</sup>,  
and Yucheng Zhou <sup>a,\*</sup>

The material properties of softwood species affect the safety of building structures, and wood identification is a key factor in material certification in specific institutions for green building certification. This study investigated an efficient wood species identification algorithm, aiming to provide a reliable method for material selection in construction and decoration industries. Using microscopic cross-sectional images of 36 softwood species applied in construction and decoration as research objects, 11 classic deep learning models were employed for species classification, combined with class activation map analysis to examine the key structural features for species identification. Specifically, the model structure and advantages of Swin Transformer were highlighted, in which hierarchical feature extraction and shifted window attention mechanism enable multi-scale fusion of wood structural features, such as tracheids, within global contexts, thereby improving classification accuracy for wood cross-sectional images. Experimental results showed that the Swin Transformer model achieved the highest classification accuracy of 99.97%, with both precision and recall exceeding 99% and an F1 score of 99%. These findings validate that deep learning networks based on the Transformer framework can achieve reliable image classification performance in wood research.

DOI: 10.15376/biores.21.1.781-798

*Keywords:* Wood identification; Microstructure; Building material; Deep learning

*Contact information:* a: School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, 250101, Shandong, China; b: Wood Value Promotion and Sustainable Development Center, Haidian, 100036, Beijing, China; c: School of Computer and Artificial Intelligence, Shandong Jianzhu University, Jinan, 250101, Shandong, China; \*Corresponding author: zhoyucheng\_2015@163.com

## INTRODUCTION

Softwoods have a wide range of applications in construction and interior decoration due to their structural properties and workability. They are commonly used in load-bearing frameworks such as beams and columns. They are also noted for their beautiful texture in wall panels, flooring and furniture. The properties of different softwood species vary significantly, and their material characteristics directly affect the structural safety of buildings. Some buildings need to use wood certified by the Forest Stewardship Council (FSC) or relevant institutions when obtaining LEED certification for green buildings (U.S. Green Building Council, 2019). Such programs require wood identification as a critical certification criterion, which involves the identification of the type of wood and the origin of the wood. Therefore, accurately identifying tree species not only ensures that the safety of the material complies with engineering specifications and building standards, but also holds significant importance for green building certification.

The differentiation of softwood remains challenging, as documented in studies by Moulin *et al.* (2022) and Gao *et al.* (2023), which highlight that interspecific similarities and intraspecific variability in wood species often induce species confusion. Such misclassification outcomes directly lead to product quality deviating from expected standards. The research on the identification of wood species has continued, which has progressed from traditional expert assessment based on structural characteristics to modern computer vision techniques employing large-scale image recognition. Technological advances have significantly enhanced the accuracy and efficiency of classification methods.

Oktaria *et al.* (2019) developed a wood identification system that classifies wood species based on macroscopic images and taxonomic names. Their method employs a Convolutional Neural Network (CNN) model as the classification algorithm. Comparative experiments are conducted with classical deep learning architectures, including AlexNet, ResNet, and GoogLeNet. The results demonstrated that AlexNet achieved optimal performance with 96.7% classification accuracy. Lens *et al.* (2020) conducted a comprehensive study using South American wood image datasets, achieving 95.6% classification accuracy for 112 wood species through a CNN architecture integrating multiple feature descriptors including texture analysis, Gabor filters, and Histogram of Oriented Gradients (HOG). He *et al.* (2021) proposed an integration framework incorporating three CNNs, which was trained on two macroscopic wood image datasets. After evaluating 9 distinct CNN architectures, the model demonstrated 100% perfect classification on their proprietary dataset and 98.81% accuracy on external datasets. These results show that macroscopic images can potentially replace microscopic analysis for wood species identification, while confirming the efficacy of deep CNN integration for this application. Moulin *et al.* (2022) developed a CNN-based system that achieved over 90% accuracy in classifying 23 Brazilian commercial wood samples from 2,448 images. However, the model showed limited discriminative capability between *Toona ciliata* and *Khaya ivorensis*, primarily due to the high structural similarity, in addition to the wood variability within these species. Gao *et al.* (2023) compared four CNN models for tree species identification using annual ring characteristics. Even though the GoogleNet achieved the highest accuracy of 96.7%, some interspecies confusion persisted in the classification results. The experiment revealed that the recognition accuracy at the family and genus levels was higher than that at the species level, and the identification accuracy of broad-leaved tree species was higher than that of coniferous tree species. Similar research findings were confirmed in the paper by Silva *et al.* (2017). In 2024, Liu *et al.* identified 27 endangered wood species based on the ResNet model, achieving a recognition accuracy of 98%. Also in 2024, Zheng *et al.* proposed the WoodGLNet model based on a pyramid attention mechanism, with an identification accuracy of 99% on 20 types of wood species CT dataset. The analysis shows that the introduction of an attention mechanism in the deep learning model allows for greater focus on the important features of wood and flexible handling of the relationship between wood species and image textures.

The Transformer model, first proposed by Vaswani *et al.* (2017), revolutionized natural language processing tasks through its self-attention mechanism, outperforming traditional deep learning models. Subsequently, researchers extended its application to digital image processing. A pivotal advancement occurred in 2020 when Dosovitskiy *et al.* from Google introduced the Vision Transformer (ViT), which leveraged self-attention to capture global image information and demonstrated exceptional performance in computer vision classification tasks.

In this study, 36 softwood specimens' images suitable for construction and decoration were selected from 112 wood specimens developed by Martin *et al.* (2023) as the research objects. Based on the Transformer architecture model, with its multi-scale feature fusion and attention mechanism, it overcomes the limitations of the CNN method in capturing subtle discriminative features, achieving a high classification accuracy of 99.7%. Contrast experiments were conducted using 11 classic models from recent years, aiming to select the most scientific classification model for determining the type of softwood, providing a reliable method for the accurate application of softwood species in construction and decoration.

## EXPERIMENTAL

### Materials

The classification of wood species is critically important for both architectural construction and furniture manufacturing, as different species possess distinct material properties that necessitate selection based on specific applications. Compared to hardwood, softwood demonstrates superior characteristics as raw material for manufacturing wooden structural building components, including rapid growth rate, ease of processing, relatively low density yet high strength, as well as excellent thermal and acoustic insulation properties. This study therefore focuses on 36 softwood species can be used in construction and interior decoration as experimental materials to investigate species identification methods. Microscopic examination of softwood structures reveals the presence of pores, rays, and annual rings in transverse sections. Although these anatomical features exhibit high similarity across different species, subtle differences exist in pore diameter, cell wall thickness, ray distribution density, and resin canal characteristics - all of which constitute critical parameters for traditional wood species identification (Phillips 2008). Figure 1 presents the micrographs of transverse sections from the 36 investigated wood species.



**Fig. 1.** The transverse section of 36 softwood species

**Table 1.** Properties and Uses of 36 Kinds of Softwood

No.	Species	Timber Property	Buildings Applications
1	<i>Agathis beccarii</i>	low to moderate density, not exceptionally strong in strength	joinery, paneling, lightweight frame
2	<i>Araucaria angustifolia</i>	moderate density, heartwood nondurable	stair treads, plywood, interior trim, joinery
3	<i>Cephalotaxus drupacea</i>	tough and resilient, corrosion-resistant	ornamental in architectural and garden design; hedge
4	<i>Cephalotaxus harringtonia</i>	hard, moderate strength and durability	ornamental in architectural and garden design; decorative strips
5	<i>Torreya nucifera</i>	high density, hard, resistant to corrosion	ancient architectural maintenance
6	<i>Calocedrus decurrens</i>	moderate hardness, soft, lightweight, resistant to decay and insects	exterior construction, fencing, interior and exterior siding
7	<i>Chamaecyparis formosensis</i>	moderate to high range density, highly durability of resisting to decay and insects	temples; door, window frame
8	<i>Chamaecyparis pisifera</i>	moderately dense, decay and moisture resistance	palace, temples; joinery
9	<i>Cupressus arizonica</i>	medium density and strength, resilient, resistant to decay	fence posts, sashes, doors, blinds
10	<i>Cupressus lindleyi</i>	medium density, resistant to moisture	frame, general purpose in constructions
11	<i>Fitzroya cupressoides</i>	strong, moderately durable and resistant to decay	interior trim, flooring, decking, furniture, joinery
12	<i>Larix laricina</i>	medium density, resistance to decay	frame, general purpose in construction
13	<i>Larix leptolepis</i>	moderate density and strength, solid, resistance to decay and pests	lumber, material for structural panel (e.g. Brett Stapel)
14	<i>Larix sp.</i>	high strength, tough and durable	frame, general purpose in construction
15	<i>Tetraclinis articulata</i>	medium-density, good hardness, durable	ancient buildings (e.g. beams, rafters and pillars)
16	<i>Widdringtonia cupressoides</i>	medium-low density, resistant to decay	building material theoretically, but now has been protected
17	<i>Abies religiosa</i>	soft, poor durability against decay and insects	decorative items, sheathing for interiors, furniture
18	<i>Abies vejarii</i>	low hardness, poor resistance to decay and insect attacks	building material theoretically, but now has been protected
19	<i>Cedrus atlantica</i>	very durable, resistance to insects	joists, scaffolding, decorative veneers, furniture
20	<i>Cedrus libani</i>	lightweight, resistant to decay and insects	joinery, ancient religious buildings (e.g. beams, columns)
21	<i>Cedrus sp.</i>	moderate density, lightweight, resistance to decay and insects	constructing raised beds, reinforcement material in gypsum boards
22	<i>Keteleeria fortunei</i>	resistant to water and moisture, moderate tensile and compressive strength	general purpose in theory but now it has been protected
23	<i>Picea abies</i>	medium density, high strength, low to medium durability	LVL, general purpose in construction
24	<i>Pinus arizonica</i>	medium texture, good strength and durability	particleboard, rafter, joists, handicrafts
25	<i>Pinus caribaea</i>	high strength, good load-bearing capacity, durable	building frame structure, floors, doors, plywood, furniture
26	<i>Pinus elliottii</i>	moderate density, good strength and durability	building frame structure, particleboard
27	<i>Pinus greggii</i>	low to moderate density, low durability	construction and fence posts



No.	Species	Timber Property	Buildings Applications
28	<i>Pinus maximinoi</i>	moderately heavy, relatively high density, moderately durable	frame structure, joinery, plywood, general housing components
29	<i>Pinus taeda</i>	density varies with tree age, moderate decay resistance, excellent hardness	frame, general purpose in construction
30	<i>Pseudotsuga macrolepis</i>	hard, good workability, moderate resistance to decay	general purpose theoretically, but now listed as endangered
31	<i>Tsuga canadensis</i>	moderately dense, poor resistant to decay	CLT, plywood, framing, etc.
32	<i>Tsuga sp.</i>	moderate density, moderate compressive strength, low-durable for decay resistance	lightweight framing, plywood
33	<i>Podocarpus lambertii</i>	moderate strength, relatively hard, heartwood not very resistant to decay	fence posts, joinery
34	<i>Taxus baccata</i>	high density, strong and hard, resistance to decay and insect	architectural and garden design, ancient buildings restoration
35	<i>Sequoia sempervirens</i>	moderately hard, low to medium density, resistant to decay and insect	trim, veneer, exterior furniture, lumber
36	<i>Taxodium distichum</i>	resistant to rot, moderately heavy, moderately strong and hard.	heavy construction, siding, interior tri

The 36 wood species can be classified into 22 genera in 7 families, with their material properties and representative applications in construction and interior decoration summarized in Table 1.

Araucariaceae wood exhibits moderate hardness, typically featuring smooth surfaces, excellent adhesive properties, and easy rotary cutting, making the plantation timber a suitable raw material for interior decorative products such as plywood.

Cephalotaxaceae species has remarkable density, toughness, and elasticity, making it an ideal choice for load-bearing components in construction. But due to resource conservation, it is confined to the realm of architectural design, serving as decorative materials and elements in landscape design.

Cupressaceae timbers are valued for their stable, aesthetically pleasing coloration and durable hardness and moisture resistance, commonly employed in traditional timber-framed buildings, serving as load-bearing components (e.g. beams and columns) and interior elements including staircases, windows and panels.

Pinaceae timbers serve as the preferred material in the construction industry. They have been widely used in various aspects of construction and indoor decoration due to their high renewability, low cost, and stable supply chain, and they are the core raw material for modern structural wooden materials, such as Laminated Veneer Lumber (LVL) and Cross-Laminated Timber (CLT). Furthermore, these timbers feature a high carbon sequestration rate, making them more suitable for application in green buildings.

Podocarpaceae species can be used in interior decorations such as high-end furniture panels, premium flooring, due to their distinctive grain patterns and stable coloration. Also it meets the requirements for small-scale buildings construction, but their high cost restricts application.

Taxaceae timbers exhibit high decay resistance and exceptional toughness, making them theoretically suitable for load-bearing components, such as beams. However, they are not mainstream construction materials even though their timber is valued for strength and aesthetic appeal in woodworking. This is because most species in this family are endangered, and lawful harvesting requires FSC certification. Nowadays, they are

predominantly applied in architectural landscape design and ancient architecture restoration.

Taxodiaceae wood features straight grain and excellent workability, with densities ranging from low to medium and moderate strength. The heartwood is highly resistance to decay, which makes it suitable for wetland construction. With low economic costs, it serves as a favorable raw material for non-load-bearing components and decorative elements.

## METHODOLOGY

The experimental procedure comprises three parts. a) Dataset creation: establishing the dataset of the selected 36 softwood species cross-sectional microscopic images. b) Experimental model: 11 classification models were selected for training respectively, and the Swin-Transformer model is illustrated as a sample. c) Prediction evaluation: based on the prediction accuracy of each model, model performance is evaluated against ground truth labels and classification metrics.

### Dataset Construction

The experimental image dataset was sourced from the Wood Anatomy Laboratory at the Federal University of Paraná, Curitiba, Brazil. The dataset includes 36 species of softwood, with the image collection process consisting of three steps. First, the wood was immersed in boiling water for an extended period to soften it. It was subsequently sliced into thin sections of 25 microns in thickness. The wood sections were subjected to a triple staining procedure employing astra blue, chrysoidine, and acridine red to differentially highlight cellulose and lignified structures, then dehydrated through a graded ethanol series to remove moisture for microscopic observation. Finally, high-resolution digital images of the wood structure were captured under an Olympus Cx40 microscope and saved in PNG format. For each species, 20 cross-sectional images of the wood were collected, with image dimensions of 1024px × 768px. To ensure an adequate number of wood cross-sectional images for deep learning model training, data augmentation was performed on the images in the dataset. Each image was cropped into four independent sub-images without overlapping pixels, with each sub-image measuring 384px × 384px. These sub-images were then proportionally resized to 224px × 224px. Further augmentation was applied by rotating the images 90°, 180°, and 270°, as well as flipping them vertically and horizontally. After data augmentation, the number of images for each species increased from the original 20 to 480, resulting in a total of 17,280 images across the 36 wood species, as shown in Fig. 2. Then, the images were randomly split into training and testing sets in an 8:2 ratio, resulting in 13,824 images for the training set and 3,456 images for the testing set. Because the stained wood cross-sectional images were 8-bit, three-channel true-color PNG images, to eliminate the influence of color tone on species classification, the images were converted to grayscale to obtain single-channel 8-bit grayscale images. The pixel values of the grayscale images were replicated to generate three-channel true-color images, where the RGB pixel values were equal.

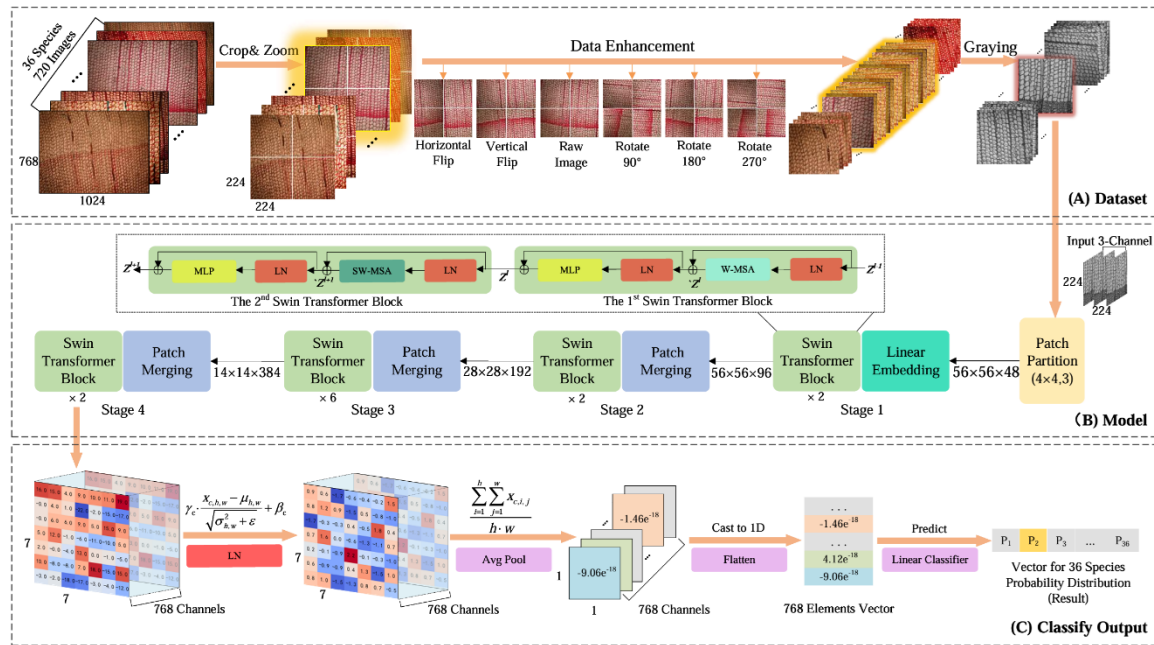


Fig. 2. The classification process for 36 softwood species

## Experimental Procedures

The employed model was the Swin-Transformer, which consists of Patch Partition (PP), Linear Embedding (LE), Swin-Transformer Block (STB), Patch Merging (PM), and a Classifier. Patch Partition divides the input image into non-overlapping patches, which serve as the input of model. Linear Embedding linearly maps each image patch into a high-dimensional space before passing it to the STB. Patch Merging is applied at different stages of the Swin-Transformer to reduce the resolution of the feature maps while increasing the feature dimension. This module involves concatenation, linear projection, and pooling operations. The Classifier, as the final module of the model, introduces a classification head to map the module's output to the final class labels. Among these components, the STB is the core of the model. It includes Layer Normalization (LN), which normalizes the input, and Window-based Multi-head Self-Attention (W-MSA), which applies multi-head self-attention within divided windows. A Multilayer Perceptron (MLP) is attached after each attention layer. Shifted Windowing is employed to create overlapping windows by shifting the window positions, thereby enhancing cross-window interactions while reducing computation. Dropout is used for regularization to prevent overfitting.

The specific implementation is illustrated in Fig. 2(B). The Swin-Transformer receives a  $224\text{px} \times 224\text{px}$ , 3-channel wood cross-section image as input. The component PP utilizes 3-channel  $4\text{px} \times 4\text{px}$  non-overlapping pixel blocks to divide the image into  $56 \times 56$  tokens. These tokens are generated by transforming the  $4\text{px} \times 4\text{px}$  pixel blocks into 48-dimensional vectors, with each vector corresponding to a  $1\text{px} \times 1\text{px}$  patch's RGB channel pixel values. In the first stage, the downsampling module employs LE to perform linear transformations on the image patches after partitioning, mapping them into a high-dimensional feature space, producing feature vectors of a certain dimension. This allows the image patches to be input into subsequent Transformer modules in a more abstract and representative form. In stages 2 to 4, PM components are used to downsample the feature maps by merging adjacent patches, doubling the channel count to reduce the feature map size. The output feature map's width and height are reduced to half of their original values,

thereby decreasing the computational complexity of subsequent layers while increasing the receptive field to capture global information. The specific process is as follows: the initial  $56 \times 56 \times 48$  token sequence is passed through Stage 1, which includes LE and two STBs, outputting a  $56 \times 56 \times 96$  feature map. In Stage 2, which includes PM and two STBs, the output is a  $28 \times 28 \times 192$  feature map. In Stage 3, after passing through PM and six STBs, the output becomes a  $14 \times 14 \times 384$  feature map. Finally, Stage 4, with PM and two STBs, outputs a  $7 \times 7$  feature map with 768 channels. The STB serves as the core of the model and undergoes two operations to enhance computational efficiency and feature extraction capacity. In the first STB, the input feature map is embedded into the LE *via* linear projection. After embedding, the high-dimensional image blocks are normalized by LN and input into the W-MSA module, which computes self-attention over the window regions of the input feature map. The output is then combined with the input through residual skip connections. After LN normalization and processing by a MLP, the resulting feature map is added to the original input, producing the output feature map. The second STB is similar to the first, with the only change being the replacement of the W-MSA module with the Sliding Window Multi-Head Self-Attention (SW-MSA) module, which calculates self-attention within the sliding window to enhance inter-window information interaction. Other structural components and processing orders remain unchanged. The alternating W-MSA and SW-MSA modules in this design facilitate both local and global information interaction within the model.

The feature map output from the model contains the key feature matrix of tree species, serving as input to the classification module ‘Classify Output’ for wood species prediction, as shown in Fig. 2(C). The LN processes the feature map according to Eq. 1. Global average pooling (AvgPool) operates on the 768-channel  $7 \times 7$  feature map as defined in Eq. 2. The 768-channel  $1 \times 1$  feature map is flattened into a 1-dimensional vector of length 768, which is fed into the linear classifier. The outputs  $p_1, p_2, p_3 \dots p_{36}$  represent the class probabilities for each wood species. Equations 1 and 2 are as follows:

$$LN(x_{c,h,w}) = \gamma_c \cdot \frac{x_{c,h,w} - \mu_{h,w}}{\sqrt{\sigma_{h,w}^2 + \varepsilon}} + \beta_c \quad (1)$$

$$AvgPool_c = \frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} x_{c,i,j}}{h \cdot w} \quad (2)$$

## Experimental Environment and Evaluation Index

A high-performance computing system was employed for wood image classification tasks, configured with the following specifications: Central Processing Unit (CPU): 13th Gen Intel® Core™ i9-13900HX (2.20 GHz); Memory: 32 GB RAM; Graphics Processing Unit (GPU): NVIDIA GeForce RTX 4080 with 16 GB GDDR6X VRAM, featuring 9,728 Compute Unified Device Architecture (CUDA) cores. The system operated on Microsoft Windows 11, utilizing the PyCharm integrated development environment. The classification framework was implemented in Python 3.9, leveraging PyTorch 2.0.1 and CUDA Toolkit v11.8 for deep learning model training and inference.

To evaluate the performance of the wood species recognition models, the following metrics were used: Accuracy, Precision, Recall, and F1 score. Accuracy reflects the model’s overall ability to correctly identify wood species and is an important evaluation metric for the recognition capability of the model, especially for identifying unknown



species. Precision refers to the proportion of samples predicted as positive that are actually positive, which measures the accuracy of the model when predicting a particular tree species category. This metric is particularly important for the correct identification of softwood species. Recall indicates the proportion of actual positive samples that the model correctly identifies, and it assesses the model's ability to recognize all actual positive samples, serving as a key metric for evaluating model sensitivity. The F1 score is the harmonic mean of Precision and Recall, offering a more comprehensive performance evaluation, especially in the context of the complexity and imbalance of wood species classification. Equations 3 through 6 are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

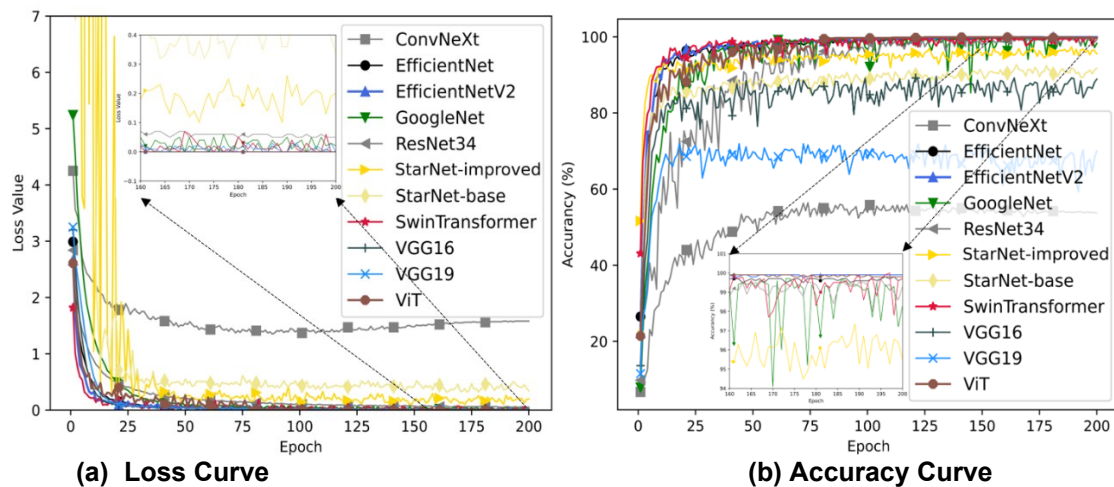
$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F_1score = \frac{2Precision \times Recall}{Precision + Recall} \quad (6)$$

The TP denotes the number of samples where the model predicts a positive class and the actual class is also positive, indicating the number of correctly identified wood species. The FP refers to the number of samples where the model predicts a positive class but the actual class is negative, indicating that samples of other wood species are misclassified as the target wood species. The TN represents the number of samples where the model predicts a negative class and the actual class is also negative, meaning that samples of non-target wood species are correctly identified as non-target. The FN signifies the number of samples where the model predicts a negative class but the actual class is positive, *i.e.*, samples of the target wood species are misclassified as other wood species.

## RESULTS AND DISCUSSION

The dataset consisting of 36 cross-sectional images of softwood species was used to train 11 deep learning models: VGG16, VGG19, GoogleNet, ResNet34, ConvNeXt, StarNet-base, StarNet-improved, EfficientNet, EfficientNetV2, ViT, and Swin-Transformer. The training set was employed to optimize the model weights, while the test set was used to evaluate the species classification ability of the models. All models adopt the Adam optimizer, with the learning rate set to 0.0001 and the batch size set to 32. The experiment was conducted over 200 iterations, yielding training curves, accuracy, and loss values. From Fig. 3(a), it can be observed that, except for StarNet-base and StarNet-improved, which exhibited substantial fluctuations before the 25<sup>th</sup> iteration, all other models demonstrated only minor fluctuations during the early stages of training.



**Fig. 3.** Training loss and accuracy curves for each classification model

**Table 2.** The Performance Indicator for 11 Models

No	Model Name	Input Size	Best Accuracy	Training Time (s)	Prediction Time (ms)	#Params (M)	#FLOPs (G)
1	SwinTransformer; (Liu <i>et al.</i> 2021)	224 × 224	99.97%	$1.826 \times 10^4$	154.69	27.55	4.38
2	EfficientNetV2; (Tan <i>et al.</i> 2021)	224 × 224	99.94%	$2.919 \times 10^4$	143.13	20.22	2.88
3	ViT; (Dosovitskiy <i>et al.</i> 2021)	224 × 224	99.88%	$3.568 \times 10^4$	206.77	86.42	16.88
4	EfficientNet; (Tan <i>et al.</i> 2019)	224 × 224	99.79%	$1.130 \times 10^4$	123.44	4.05	0.41
5	GoogLeNet; (Szegedy <i>et al.</i> 2015)	224 × 224	99.73%	$1.828 \times 10^4$	121.88	10.41	1.6
6	ResNet; (He <i>et al.</i> 2015)	224 × 224	99.71%	$1.542 \times 10^4$	118.44	21.3	3.68
7	StarNet-improved	224 × 224	97.35%	$1.024 \times 10^4$	112.92	3.45	0.43
8	StarNet-base; (Ma <i>et al.</i> 2024)	224 × 224	92.67%	$0.940 \times 10^4$	110.83	2.68	0.43
9	VGG16; (Simonyan <i>et al.</i> 2015)	224 × 224	89.55%	$2.387 \times 10^4$	151.25	134.41	15.52
10	VGG19; (Simonyan <i>et al.</i> 2015)	224 × 224	71.99%	$2.749 \times 10^4$	167.29	139.72	19.68
11	ConvNeXt; (Liu <i>et al.</i> 2022)	224 × 224	56.36%	$2.567 \times 10^4$	130.73	27.85	4.49

After the 50<sup>th</sup> iteration, the loss values of all models stabilized and approached zero, indicating fast convergence and strong performance. Figure 3(b) presents the classification accuracy of the 11 models. VGG19 and ConvNeXt achieved accuracy below 80% even after 200 training epochs, highlighting their poor classification performance for wood species. The reason for this could be attributed to the high number of layers in both models,

which may have hindered the adequate training of key features, resulting in gradient vanishing during training and deteriorating model performance. In contrast, the other models exhibited good classification accuracy, with Swin-Transformer, EfficientNetV2, and ViT maintaining high accuracy levels after approximately 150 training epochs.

To further evaluate the classification capability of 11 models for wood species, five evaluation metrics were provided: maximum accuracy, training time, inference time, model parameters, and floating-point operations per second. The experimental results are shown in Table 2. The classical convolutional neural network classification methods, such as VGG16 and VGG19 recognition accuracy of 89.55% and 71.99% respectively. This indicates that the deeper VGG19 model is more prone to the vanishing gradient phenomenon during the iterative update of convolutional kernel weight parameters in the back-propagation process, so that the accuracy rate is reduced. The Resnet model with the introduction of residual blocks effectively solves the problem of vanishing gradients with an accuracy rate of its up to 99.71%. However, increasing model depth is not as effective as multi-scale feature fusion. Taking GoogleNet as an example, this model does not employ residual blocks but uses the Inception module as the basic unit. It outputs feature maps through multi-scale convolution and pooling operations in multiple parallel branches. This design enables the model to integrate feature information from multiple receptive fields and perform species identification, which results in a wood species classification accuracy of 99.73%. The high-dimensional feature interaction in the Inverted Bottleneck (IB) module of EfficientNet and EfficientNetV2, as well as the hierarchical feature extraction of the Swin Transformer model, both adopt multi-scale feature fusion, with their accuracies all exceeding 99.79%.

In addition to multi-scale feature fusion, the spatial attention mechanism is also a key module for the model to identify wood species. To mitigate the interference of stain color in microscopic cross-sectional wood images, this study employed RGB images with consistent grayscale values for corresponding pixels across three channels during dataset creation. Such images can render the channel attention mechanism in models ineffective, thereby limiting identification accuracy. EffectiveNet and ViT utilized this mechanism, with accuracy rates of 99.79% and 99.88% respectively. In contrast, the convolutional computation in EfficientNetV2 reduced reliance on the channel attention mechanism, boosting accuracy to 99.94%. StarNet-base is a lightweight convolutional neural network with the fewest parameters and the shortest training time. The StarNet-improved, an enhancement of the StarNet baseline model, improves the original architecture by focusing on its third stage. In this stage, each even-numbered Demo Block incorporates a depthwise convolution following the Star Operation. Specifically, a spatial attention mechanism is inserted immediately before the depthwise convolution to enhance feature extraction efficiency. Compared to StarNet-base, the training time of StarNet-improved increased nearly 9%, but its accuracy improved 4.68%, demonstrating that the spatial attention mechanism plays a crucial role in enhancing the model's ability to distinguish wood species. The self-attention mechanism established by ViT directs the model to consider the dependencies between image regions both spatially and across channels. Through training, the model's discrimination mechanism learned the feature distribution of wood species *via* spatial attention, resulting in a high classification accuracy of 99.88%. Swin Transformer achieved the highest wood species classification accuracy, reaching 99.97%. In contrast to ViT's self-attention mechanism, Swin Transformer uses a SW-MSA, which enables communication between windows by exchanging information across shifts, allowing the multi-head attention to monitor changes in features such as tracheid distribution, ray

density, and cell wall thickness. This mechanism plays a critical role in tracking continuous wood anatomical features, and SW-MSA also enhances the edge detection of growth ring boundaries in the microscopic images of wood species. Moreover, the model only utilized spatial attention, without the interference of channel attention, effectively addressing the issue of wood species classification based on microscopic images of wood cross-sections.

ConvNeXt achieved an accuracy of only 56%, the lowest among the 11 models. The authors tried various optimizers for ConvNeXt but still failed to achieve satisfactory results. The primary cause of this phenomenon lies in the  $4 \times 4$  downsampling operation of the “Patchify” layer, which induces the loss of low-level texture features. This mechanism impedes the model’s capacity to acquire critical characteristics, including wood species-specific texture patterns and tracheid morphology. This indicates that ConvNeXt is ill-suited for the classification of wood cross-sectional images.

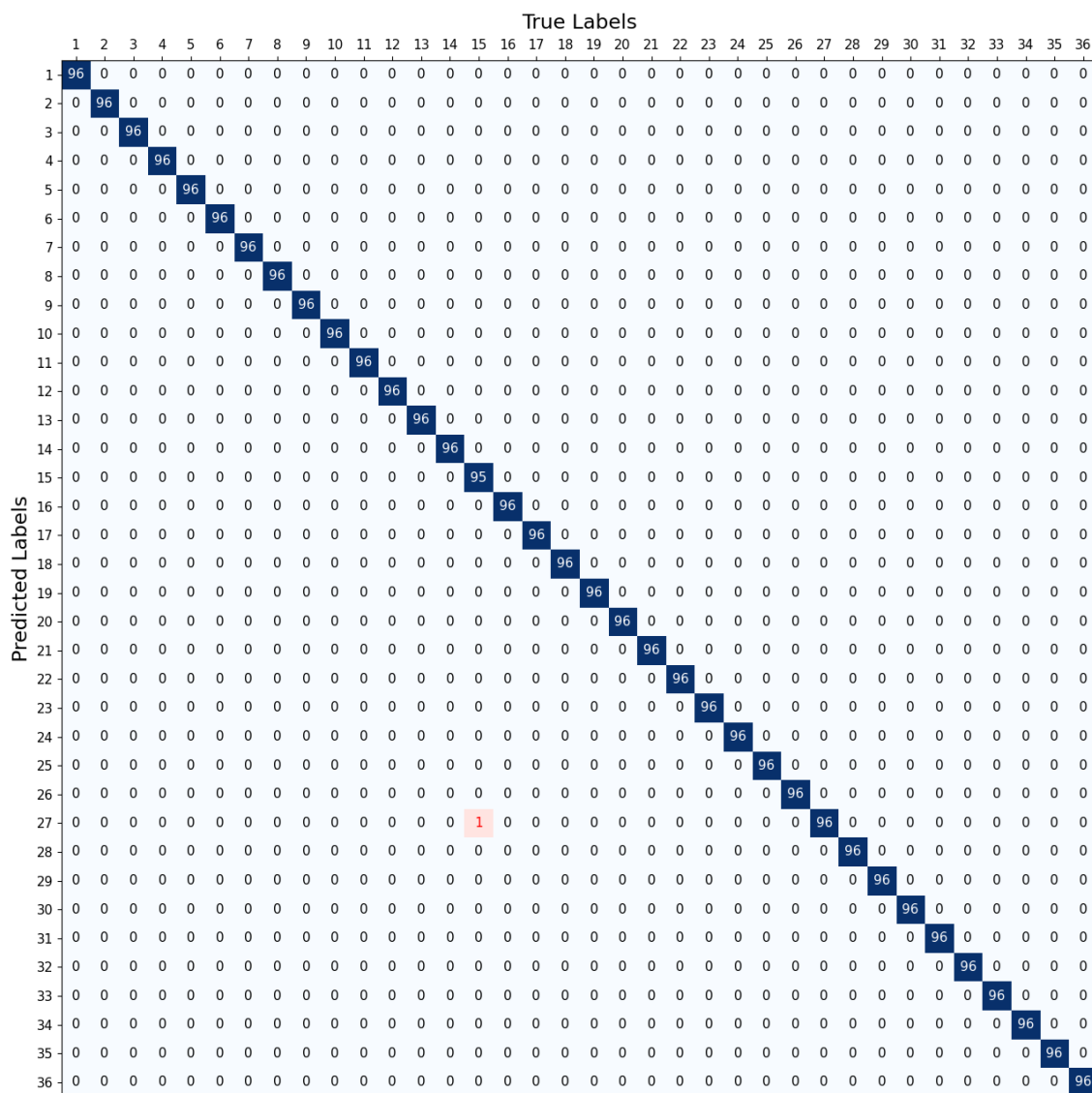


Fig. 4. Confusion matrix of the 36 softwood species test set using Swin-Transformer



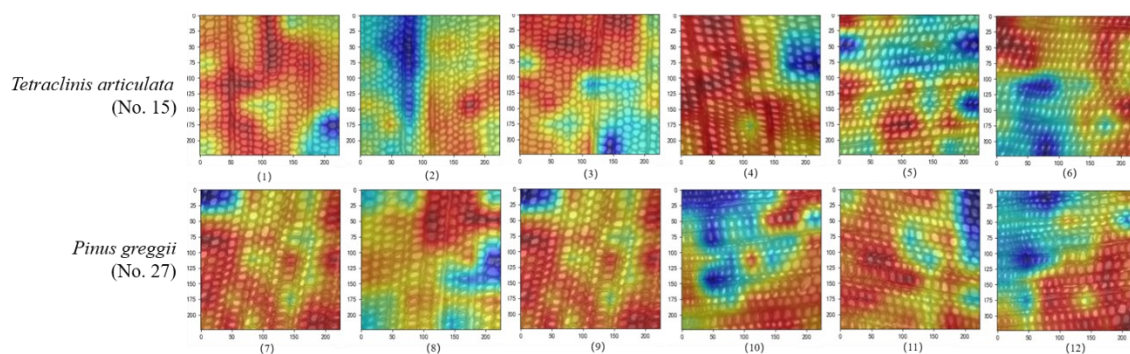
Wood species identification was investigated using 36 wood specimens, with each specimen in the test set represented by 96 images, totaling 3,456 images. Figure 4 presents the classification performance of Swin-Transformer on the test set using a confusion matrix. The numerical labels on the X-axis of the confusion matrix correspond to the true species categories, while the Y-axis represents the predicted species categories. The numbers along the diagonal of the confusion matrix, from the top left to the bottom right, represent the number of images of a particular wood species that were correctly classified, where a value of 96 signifies that all test images of that species were accurately identified. The off-diagonal entries, marked in red, denote misclassified images. The prediction results for species No. 15 (*Tetraclinis articulata*) show that 95 images were correctly classified, while one image was misclassified as species No. 27 (*Pinus greggii*). To evaluate the performance of the Swin-Transformer model, this study employed Accuracy, Precision, Recall, and F1-score as metrics. The model achieved the highest accuracy, with detailed results presented in Table 3.

**Table 3.** Wood Classification Indicators for the Best Accurate Model

Species (No.)	Accuracy	Precision	Recall	F1 Score
<i>Agathis beccarii</i> (1); <i>Araucaria angustifolia</i> (2); <i>Cephalotaxus drupacea</i> (3); <i>Cephalotaxus harringtonia</i> (4); <i>Torreya nucifera</i> (5); <i>Calocedrus decurrens</i> (6); <i>Chamaecyparis formosensis</i> (7); <i>Chamaecyparis pisifera</i> (8); <i>Cupressus arizonica</i> (9); <i>Cupressus lindleyi</i> (10); <i>Fitzroya cupressoides</i> (11); <i>Larix laricina</i> (12); <i>Larix leptolepis</i> (13); <i>Larix sp.</i> (14); <i>Widdringtonia cupressoides</i> (16); <i>Abies religiosa</i> (17); <i>Abies vejarii</i> (18); <i>Cedrus atlantica</i> (19); <i>Cedrus libani</i> (20); <i>Cedrus sp.</i> (21); <i>Keteleeria fortune</i> (22); <i>Picea abies</i> (23); <i>Pinus arizonica</i> (24); <i>Pinus caribaea</i> (25); <i>Pinus elliottii</i> (26); <i>Pinus maximinoi</i> (28); <i>Pinus taeda</i> (29); <i>Pseudotsuga macrolepis</i> (30); <i>Tsuga canadensis</i> (31); <i>Tsuga sp.</i> (32); <i>Podocarpus lambertii</i> (33); <i>Taxus baccata</i> (34); <i>Sequoia sempervirens</i> (35); <i>Taxodium distichum</i> (36);	1.00	1.00	1.00	1.00
<i>Tetraclinis articulata</i> (15)	1.00	1.00	0.99	0.99
<i>Pinus greggii</i> (27)	0.99	0.99	1.00	0.99

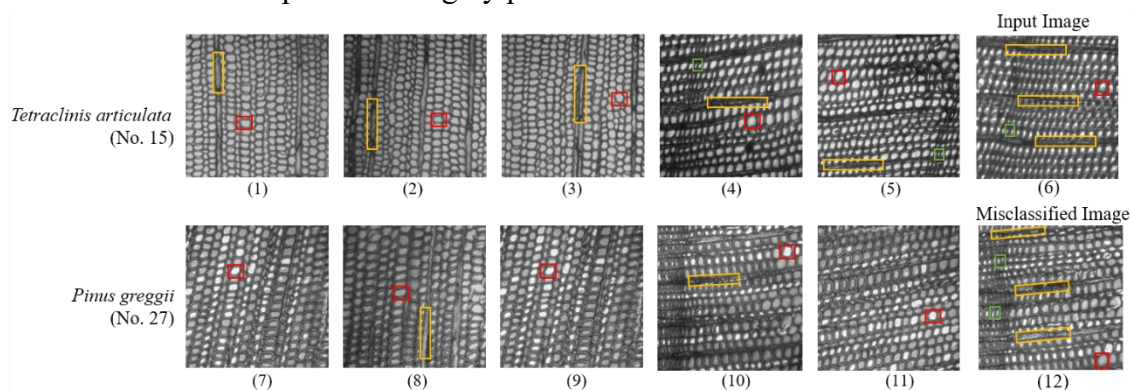
The Swin-Transformer model achieved an accuracy of 99.97% in classifying the cross-sectional microscopic images of 36 types of softwood, with both precision and recall exceeding 99%, and an F1 score of 99%. With the exception of species No. 15, all other 35 species achieved an accuracy of 1. Out of 3,456 images, 3,455 were correctly identified, with only one image being misclassified. Upon analyzing the misclassification of species No. 15 (*Tetraclinis articulata*), it was found that the wood texture closely resembled the image features of species No. 27 (*Pinus greggii*) in the training set. The W-MAS and SW-MAS mechanisms in Swin-Transformer are capable of capturing low-dimensional structures, such as the wood texture, tracheids, and rays, in the cross-sectional images, while also integrating contextual semantic information at higher dimensions to extract features for classification.

In view of the misclassification phenomenon when species No. 15 were classified by Swin Transformer method, the authors randomly selected the cross-section images of species No. 15 and species No. 27 and used Gradient-weighted Class Activation Mapping (Grad CAM) to generate Class Activation Maps. The distribution of model-activated regions in the image is shown in Fig. 5, where the red color is the region activated by the model (*i.e.*, the high correlation region), which represents the key focus structure of interest to the model; and the blue color is the non-interested structure (*i.e.*, the low correlation region), to which the model pays less attention. It can be noted that the high-correlation regions are mainly concentrated in wood rays and tracheids, especially focusing on the structure of that of the earlywood, indicating that they are an important basis for judging the tree species category.



**Fig. 5.** Class activation map of *Tetraclinis articulata* (No. 15) and *Pinus greggii* (No. 27)

To further investigate the cause of the misclassification, a comparison was made between the cross-sectional texture structures of species No. 15 and the misclassified species No. 27. As shown in Fig. 6, the upper row displays images of *Tetraclinis articulata* from the test set, while the lower row shows images of *Pinus greggii* from the training set. The textures of these images are strikingly similar, indicating that the cross-sectional image features of these two species are highly prone to confusion.



**Fig. 6.** Images of *Tetraclinis articulata* (No. 15) and *Pinus greggii* (No. 27)

A detailed analysis revealed that the ray cell walls of *Tetraclinis articulata* are thin, with earlywood tracheid diameters ranging from 18 to 30  $\mu\text{m}$ . The cross-sectional shape of the tracheid is polygonal, rectangular, and square, while the cross-sectional shape of latewood tracheid is oval and square. In contrast, the ray cell walls of *Pinus greggii* are

also thin, with earlywood tracheid diameters similar to those of *Tetraclinis articulata*. The cross-sectional shape of the tracheid is predominantly oval or circular, with some being polygonal, while the latewood tracheid cross-sections are elongated oval or rectangular. Upon examining these images, three common features were identified. The left half of the image contains one annual ring, exhibiting curved diagonal lines. Three rays traverse the image horizontally, with uniform spacing. The tracheid diameters around the annual ring are smaller, and the size and shape in the other regions are similar. These features were also observed in the misclassified images, and the ray arrangement of the two species was identical. Although the Swin-Transformer model did not achieve perfect accuracy in classifying the 36 softwood species, its classification accuracy was already exceptionally high, demonstrating outstanding performance.

## CONCLUSIONS

1. This paper describes the application background of softwood in the construction and decoration fields, and presents cross-sectional images of 36 softwood species. It proposes the use of deep learning methods to classify softwood, providing an effective approach for the identification of softwood species in the construction and decoration industries.
2. The experiment applied 11 kinds of image classification models proposed in recent years for comparative study. Data enhancement techniques were used, including rotating, flipping, and non-overlapping cropping to enlarge the dataset, to explore the classification performance of these models on the 36 kinds of architectural softwood microscopic cross-sectional image dataset. Experimental results showed that 6 models achieved over 99.0% classification accuracy, verifying the approach of deep learning neural network can be applied to the identification of softwood species. Among them, Swin Transformer achieved the highest accuracy of 99.7%, with only 1 misclassification in 3,456 test images. The misidentification was attributed to the morphological similarity of ray cells and tracheids between *Tetraclinis articulata* and *Pinus greggii*. Class activation maps generated by Grad-CAM revealed that the Swin Transformer model primarily focused on tracheid and ray structures, confirming their critical role in wood identification.
3. The Swin Transformer model incorporates hierarchical feature extraction and a shifted window attention mechanism. The former fuses feature at different scales, while the latter organizes feature information in regions of interest within the global context of input data, enabling model training to achieve high identification accuracy. By analyzing the classification mechanisms and experimental results of each model, it was found that introducing multi-scale feature fusion and spatial attention mechanisms enhances the ability to capture critical structural information in wood microscopic images, effectively improving model identification accuracy. This will provide research directions for further improving deep learning-based wood classification methods.

## ACKNOWLEDGMENTS

The authors are grateful for the support of the three projects, as listed below. The Natural Science Foundation of Shandong Province, China (ZR2024MC185), Funded Project under the Open Fund of Key Research Bases (XTYC202401), and the Taishan Scholar Advantage Characteristic Discipline Talent Team Project of Shandong Province of China (2015162). In particular, we would like to thank Dr. Jefferson Martins, with the Laboratory of Wood Anatomy and the Laboratory of Vision Robotics and Imaging at the Federal University of Parana in Curitiba, Brazil, for providing wood specimens images used to carry out research.

## Conflicts of Interest

The authors declares that there is no competing interest.

## REFERENCES CITED

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X. H., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., *et al.* (2021). “An image is worth 16x16 words: Transformers for image recognition at scale,” in: *International Conference on Learning Representations (ICLR 2021)*, Vienna, Austria, pp. 1-22. <https://doi.org/10.48550/arXiv.2010.11929>
- Gao, X., Yang, L. X., and Chen, Z. J. (2023). “Convolutional neural network tree species identification based on tree-ring radial section image features,” *Chinese Journal of Applied Ecology* 34(01), 47-57. <https://doi.org/10.13287/j.1001-9332.202301.001>
- He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2016) “Deep residual learning for image recognition,” in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2016.90>
- He, T., Mu, S. B., Zhou, H. K., and Hu, J. G. (2021). “Wood species identification based on an ensemble of deep convolution neural networks,” *Wood Research* 66(01), 01-14. <https://doi.org/10.37763/66.1.0114>
- Lens, F., Liang, C., Guo, Y. H., Tang, X. Q., Jahanbanifard, M., 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(04), 1-21. <https://doi.org/10.1163/22941932-bja10029>
- Liu, S. J., Zheng, C., Wang, J. J., Lu, Y., Yao, J., Zou, Z. Y., Yin, Y. F., and He, T. (2024). “How to discriminate wood of cites-listed tree species from their look-alikes: using an attention mechanism with the ResNet model on an enhanced macroscopic image dataset,” *Frontiers in Plant Science* 15, 1567-1583. <https://doi.org/10.3389/fpls.2024.1368885>
- Liu, Z., Lin, Y. T., Cao, Y., Hu, H., Wei, Y. X., Zhang, Z., Lin, S., and Guo, B. N. (2021). “Swin Transformer: Hierarchical vision transformer using shifted windows,” in: *2021 IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Liu, Z., Mao, H. Z., Wu, C. Y., Feichtenhofer, C., Darrell, T., and Xie, S. N. (2022). “A ConvNet for the 2020s,” in: *2022 IEEE/CVF Conference on Computer Vision and*



- Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 11966-11976.  
<https://doi.org/10.1109/CVPR52688.2022.01167>
- Ma, X., Dai, X. Y., Bai, Y., Wang, Y. Z., and Fu, Y. (2024). "Rewrite the stars," in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 5694-5703.  
<https://doi.org/10.1109/CVPR52733.2024.00544>
- Martins, J., Oliveira, L. S., Nisgoski, S., and Sabourin, R. (2013). "A database for automatic classification of forest species," *Machine Vision and Applications* 24(3), 567-578. <https://doi.org/10.1007/s00138-012-0417-5>
- Moulin, J. C., Lopes, D. J. V., Mulin, L. B., Bobadilha, G. S., and Oliveira R. F. (2022). "Microscopic identification of Brazilian commercial wood species via machine-learning," *CERNE* 28(01), article 2978.  
<https://doi.org/10.1590/01047760202228012978>
- Oktaria, A. S., Prakasa, E., Suhartono, E., Sugiarto, B., Prajitno, D. R., and Wardoyo, R. (2019). "Wood species identification using convolutional neural network (CNN) architectures on macroscopic images," *Information Technology and Computer Science* 04(03), 274-283. <https://doi.org/10.25126/JITECS.201943155>
- Phillips, E. W. J. (2008). "The identification of coniferous woods by their microscopic structure," *Journal of the Linnean Society of London* 52, 259-320.  
<https://doi.org/10.1111/j.1095-8339.1941.tb01390.x>
- Silva, N. R., Ridder, M. D., Baetens, J. M., Bulcke, J. V., Rousseau, M., Bruno, O. M., Beeckman, H., Acker, J. V., and Baets, B. D. (2017). "Automated classification of wood transverse cross-section micro-imagery from 77 commercial Central-African timber species," *Annals of Forest Science* 74(30), 1-14.  
<https://doi.org/10.1007/s13595-017-0619-0>
- Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in: *3<sup>rd</sup> International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, pp. 1-14.  
<https://doi.org/10.48550/arXiv.1409.1556>
- Szegedy, C., Liu, W., Jia, Y. Q., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). "Going deeper with convolutions," in: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Tan, M. X., and Le, Q. V. (2019). "EfficientNet: Rethinking model scaling for convolutional neural networks," in: *Proceedings of the 36<sup>th</sup> International Conference on Machine Learning (ICML)*, Long Beach, CA, USA, pp. 6105-6114.  
<https://doi.org/10.48550/arXiv.1905.11946>
- Tan, M. X., and Le, Q. V. (2021). "EfficientNetV2: Smaller models and faster training," in: *Proceedings of the 38<sup>th</sup> International Conference on Machine Learning (ICML)*, Vienna, Austria, pp. 10096-10106. <https://doi.org/10.48550/arXiv.2104.00298>
- U.S. Green Building Council. (2019). *LEED v4 Reference Guide for Building Design and Construction*, U.S. Green Building Council, Washington, DC, USA
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). "Attention is all you need," *Advances in Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, pp. 5998-6008.  
<https://doi.org/10.48550/arXiv.1706.03762>

Zheng, Z. Z., Ge, Z. D., Tian, Z. K., Yang, X. X., and Zhou, Y. C. (2024). “WoodGLNet: a multi-scale network integrating global and local information for real-time classification of wood images,” *Journal of Real-Time Image Processing* 21(04), 1-15. <https://doi.org/10.1007/s11554-024-01521-w>

Article submitted: April 29, 2025; Peer review completed: May 30, 2025; Revised version received: June 30, 2025; Accepted: November 18, 2025; Published: December 9, 2025.

DOI: 10.15376/biores.21.1.781-798