Automated Anatomical Analysis of Wood Cross Sections Using Macroscopic Images

Khanh Nguyen-Trong[1], Thanh Nhan Nguyen-Thi[2]
Intelligent Computing for Sustainable Development Laboratory (IC4SD)[1]
Faculty of Information Technology, Posts and Telecommunications Institute of Technology, Hanoi, Vietnam[1,2]

***Abstract*—Wood anatomical features are crucial in forestry science, traditionally relying on manual inspection of wood cross-sections. This conventional method is time-consuming, subjective, and dependent on expert experience. Recent advancements in deep learning offer high accuracy but often operate as black-box models, lacking interpretability and struggling with out-of-distribution challenges under real-world variations. To address these limitations, we propose a two-stage framework combining deep-learning-based image classification and explicit anatomical feature analysis, directly extracting expert-recognized morphological attributes such as pore size, frequency, and spatial arrangement from macroscopic images. By quantifying these anatomical descriptors, our framework yields transparent, OOD-robust features that can be directly fed into downstream species-identification models, thereby enhancing future classification accuracy while preserving interpretability. An end-to-end implementation integrates data acquisition, automated feature extraction, and interactive visualization, making the methodology practically applicable in both laboratory and field settings.**

***Keywords*—Wood species identification; wood anatomical analysis; segmentation; Mask R-CNN; DenseNet**

I. Introduction

Wood anatomical analysis is fundamental to forestry management, biodiversity conservation, and accurate identification of timber species. In routine practice, specialists inspect anatomical features on the transverse surface—such as vessel (pore) arrangement, ray structure, and axial parenchyma—to determine the taxon. Especially in many developing countries, samples for wood-species identification are shipped to a central laboratory, where trained analysts view the cross-sections under optical magnifiers or microscopes and match their observations to standardized references (e.g., the IAWA list [1]). This manual workflow is time-consuming, labour-intensive, and inherently subjective because accuracy depends on the expert's experience.

In recent years, numerous research efforts have explored automated or semi-automated machine learning approaches for wood species analysis, often reporting high accuracy on benchmark datasets [2], [3], [4]. For example, Pratondo *et al.* [4] studied and compared various machine learning techniques for classifying three similar wood types—jeungjing, puspa, and suren—using k-nearest neighbors, support vector machines, decision trees, random forests, and Inception-v3. Their experiments showed that the random forest classifier achieved the highest accuracy at 90.67%.

Bello *et al.* [2] proposed combining Mask R-CNN with a modified residual network, attaining a notable 92% accuracy in wood species identification. Such advancements may be surprising to lay readers, as they adapt state-of-the-art deep learning methods—commonly seen in broader computer vision tasks—to the specialized context of wood identification. Nguyen-Trong [3] further introduced an approach based on convolutional neural networks to analyze wood images at multiple magnification levels, providing a comprehensive discussion of the research framework, methodology, and findings for deeper insight into its effectiveness.

Despite these promising results, many of these solutions operate as black-box image classifiers with limited interpretability regarding which anatomical features underlie the final decision. To address these limitations, our work proposes a strategy to analyze and extract important anatomical features commonly used by experts, such as pore size, pore arrangement, and pore frequency, directly from macroscopic images. Traditional wood anatomy is grounded in systematically categorizing xylem anatomical features, which often facilitates identification at the genus level. However, fully capturing every detail in the xylem structure using only visual judgment and accumulated expertise can be challenging. Consequently, several approaches now focus on quantifying anatomical markers such as cell wall thickness and cell count, in contrast to conventional inspection methods that remain somewhat subjective [5].

Therefore, this approach not only aids in identification by leveraging anatomically meaningful descriptors but also enhances explainability, as the model outputs explicit morphological attributes that correlate with standard wood anatomy references.

The main contributions of this study are as follows:

- We propose a two-stage methodology that first classifies macroscopic images into wood versus non-wood categories, and then segments Vessel-Centric Regions (VCRs), defined as pore lumina along with immediately adjacent perivascular parenchyma, if present.
- We employ DenseNet121-based feature extraction for robust wood image classification and utilize Mask R-CNN for precise segmentation of anatomical features, enabling detailed quantitative analysis.
- We provide an end-to-end implementation integrating mobile-based image classification, remote anatomical segmentation, and visualization, demonstrating practical utility for both laboratory and field applications.

The remainder of this study is organized as follows: Section II reviews recent literature on automated wood identification. Section III details our proposed two-stage framework, including data preprocessing and anatomical feature extraction.

Section IV presents experimental setups and quantitative evaluations. In Section V, we demonstrate an end-to-end application that integrates the proposed method into a real-world workflow. Finally, Section VI concludes the study and discusses future research directions.

II. Related Works

Recent research in wood species analysis has explored various approaches using machine learning and computer vision. Regarding methodologies, two main categories can be distinguished: i) traditional machine learning, including k-NN [6], SVM [7], and ANN [8], which often require handcrafted features, and ii) deep learning methods, such as CNN-based architectures [3] (VGG [9], ResNet [10], SqueezeNet [11], DenseNet [12], etc.), which automatically extract features from image data. Studies have utilized both microscopic and macroscopic images, with a trend towards more accessible macroscopic imaging techniques [3], [13].

Many recent works favor deep learning because of its higher performance when sufficient data is available. Silva et al. [14] demonstrated high success rates (88 to 90%) in classifying 77 Central-African timber species using microscopic images and texture-based features. Advancements in deep learning have shown promise, with He et al. [15] achieving 100% accuracy on a dataset of 8 wood species using an ensemble of convolutional neural networks. Herrera-Poyatos et al. [13] proposed a patch-based classification method for high-resolution macroscopic images, addressing challenges in capturing fine-grained patterns. Nguyen-Trong [3] evaluated various CNN architectures across different magnification levels, finding DenseNet121 to perform best on 20x magnified images.

Despite these advances, most studies primarily focus on classification tasks (i.e., assigning a species label to an image). The anatomical analysis of wood, particularly the segmentation and measurement of pores (e.g., solitary pores, pore multiples, pore clusters), parenchyma arrangements, or other microfeatures, remains relatively under, explored at the macroscopic scale. However, these features are crucial for a detailed understanding of wood anatomy and can bolster both automated classification accuracy and the broader morphological characterization of new wood samples. For example, identifying whether pores are solitary or clustered (and quantifying pore size, frequency, and distribution) directly contributes to more accurate species identification and deeper insight into the wood's anatomical structure.

In literature, anatomical analysis has been widely applied across medical imaging to enhance classification accuracy and provide deeper insights into structural characteristics. This approach is crucial for various applications, including surgical planning, disease progression tracking, and treatment planning [16]. Recent advancements in image segmentation techniques have been employed to define anatomical structures in CT images, particularly for COVID-19 diagnosis [17]. Image analysis techniques are essential for improving the quality of medical images and extracting features that lead to better classification results and more accurate diagnoses [18]. Deep learning methods, such as CNNs, have shown remarkable performance in anatomy-specific classification of medical images. By training on large datasets of CT scans and employing data augmentation techniques, CNNs can achieve high accuracy in classifying anatomical structures, with error rates as low as 5.9% [19].

Techniques like CNNs have demonstrated remarkable success in automatically delineating complex anatomical features from medical images, thereby improving both the efficiency and accuracy of medical assessments [20].

Despite the proven efficacy of anatomical analysis in fields like medical imaging, its application within the domain of wood identification remains relatively underexplored, particularly at the macroscopic level. Traditional wood identification methods often rely on manual examination of features such as pore arrangement, parenchyma patterns, and other microstructures. However, these manual approaches are time-consuming and subject to human error. While some studies have employed computer vision techniques for wood species classification, they predominantly focus on assigning species labels without delving into the segmentation and quantitative analysis of anatomical features [21].

Integrating anatomical analysis into wood identification holds significant potential. By employing advanced image segmentation techniques to automatically identify and measure key features, such as distinguishing between solitary pores and pore clusters, and quantifying their size, frequency, and distribution, researchers can achieve more accurate species identification and gain deeper insights into wood anatomy. This approach mirrors the successes seen in medical imaging and suggests a promising avenue for enhancing wood identification processes through the application of sophisticated computer vision methodologies.

Building upon this research context, our work aims to develop a two-stage approach for automated anatomical analysis of wood cross-sections using macroscopic images. First, a classifier (trained with a DenseNet121 feature extractor and an GMM) determines whether an input image indeed represents a magnified wood cross-section. Next, images confirmed to be wood cross-sections proceed to a detection and segmentation model to identify and segment pores, followed by computing various anatomical descriptors (i.e., average pore size, pore distribution, solitary versus multiple pore frequency). The overall objective is to integrate these methods into a system deployable on resource-constrained devices, such as smartphones, thereby facilitating affordable, efficient, and accurate wood anatomical analysis in both laboratory and field settings.

III. Materials and Methods

In this section, we present the materials and methods underlying our automated wood anatomical analysis. First, we describe our approach for classifying macroscopic images into wood versus non-wood categories. Next, we outline the segmentation of Vessel-Centric Regions (VCRs) using Mask R-CNN. Finally, we detail the extraction of anatomical descriptors, including VCR size, frequency, and spatial arrangement, essential for characterizing different wood species.

A. Wood Macroscopic Image Classification

Binary classification is arguably the most widely documented branch of supervised machine learning, especially in

domains where both positive and negative classes are naturally bounded by the same medium (e.g., "ham versus spam" in email). However, complications emerge when applying this framework to image classification tasks.

Consider the challenge of building a model to determine whether an image shows a magnified wood cross-section. We may have a substantial collection of wood cross-section images for the "wood" class, but "not wood" examples can be drawn from virtually any image category—making it nearly infinite in scope. This is unlike email filtering, where even "spam" emails are still emails, sharing a common format and feature space with "ham".

There are multiple ways to tackle this. One approach is one-class classification, in which a model is trained primarily (or exclusively) on positive-class data—in this case, images of wood cross-sections. When such a model encounters a novel input, it determines whether the input sufficiently resembles the positive class ("wood") or should be flagged as an outlier ("not wood"). Techniques such as autoencoders, One-Class SVM, or Isolation Forest can be used in this context.

However, if we do have access to certain negative examples—i.e., images known to contain no wood cross-sections—then more traditional binary classification methods can be applied. A common strategy is to perform feature extraction via a pretrained CNN (e.g., DenseNet121), reduce the dimensionality (e.g., with PCA), and then train a conventional machine learning model (e.g., SVM, Gaussian Mixture with Isotonic Regression) on these features. This latter approach benefits from explicitly learning to discriminate between "wood" and "not wood", but requires a representative set of negative examples.

In our work, we adopt transfer learning for feature extraction, followed by testing multiple classification schemes. This allows us to balance performance needs with the availability of negative data. By leveraging proven CNN architectures trained on vast image datasets (e.g., ImageNet), we can efficiently generate high-level embeddings that are then used by simpler classifiers. This technique proves especially practical in real-world scenarios, where the negative class (all images that are "not wood") is both diverse and difficult to comprehensively represent.

1) Feature extraction with DenseNet: To handle this classification task, we adopt a transfer learning approach that relies on DenseNet121, a well-known convolutional neural network pretrained on large-scale image datasets (e.g., ImageNet). Specifically, we remove the original classification layer (i.e., the fully connected top) and leverage the network's final feature maps (global average pooled) as a 512-dimensional feature vector for each input image.

All images are first resized to 224×224 and standardized (e.g., pixel values divided by 255) to match the input layer requirements of DenseNet121. Next, the pretrained model processes each image to produce the condensed feature representation. This step provides a robust and discriminative embedding of the original imagery without the need to train a full CNN from scratch.

2) Dimensionality reduction and scaling: Following feature extraction, we scale and apply Principal Component Analysis to reduce dimensionality. Dimensionality reduction serves two main purposes:

- It helps mitigate noise or redundant information in the extracted feature space.
- It substantially decreases training time for subsequent classification models, as high-dimensional feature vectors are compressed into a more manageable number of principal components.

3) Classification approaches: Once the feature vectors are extracted and reduced via PCA, we evaluate several classification methods, including Gaussian Mixture Models (GMM) with Isotonic Regression, One-Class SVM, and Isolation Forest.

GMM with Isotonic Regression can leverage a small validation set containing both positive (wood) and negative (non-wood) examples. By modeling the distributions of each class, the GMM can estimate the likelihood that a new image belongs to the learned "wood" distribution, and Isotonic Regression refines the decision threshold for improved calibration.

In scenarios with little or no labeled negative data, one-class classifiers such as One-Class SVM and Isolation Forest are valuable. Trained primarily (or exclusively) on images known to be wood, these models learn a boundary around the "wood" manifold in the feature space. At inference time, any sample that falls outside this boundary is considered "not wood".

One-Class SVM constructs a decision function tailored to capture the dense region of positive samples in the high-dimensional space. Isolation Forest isolates anomalies by randomly selecting a feature and splitting values between them, effectively "isolating" outliers—i.e., images that do not resemble wood. These one-class methods excel when assembling a truly representative negative dataset is impractical. However, whenever domain knowledge or partial negative examples are available, combining them with positive samples for a more traditional binary classification (e.g., SVM, Random Forest, or GMM with a mixed training set) can boost performance and interpretability.

B. Wood-Pore (Vessel-Centric Region) Segmentation

In diffuse- and ring-porous hardwoods the most conspicuous anatomical element is the vessel (the pore lumen). Depending on species, the lumen may be surrounded by a narrow or wide sheath of perivascular parenchyma; in other cases the lumen is embedded directly in fibre tissue with little or no parenchyma contact. Because both situations occur within the same slide, a consistent segmentation target is essential. Throughout this study, instead of segmenting pore lumina alone—which is challenging due to the close adjacency of lumina and perivascular tissues in many species—we predict vessel-centric regions (VCRs), defined as a pore lumen together with any immediately adjacent perivascular parenchyma, if present.

If two lumina are in contact or share a common PP envelope the region is labelled as a double-VCR; otherwise it is considered a single-VCR. This strategy—which we abbreviate P+PP—allows one annotation protocol to cover solitary pores,

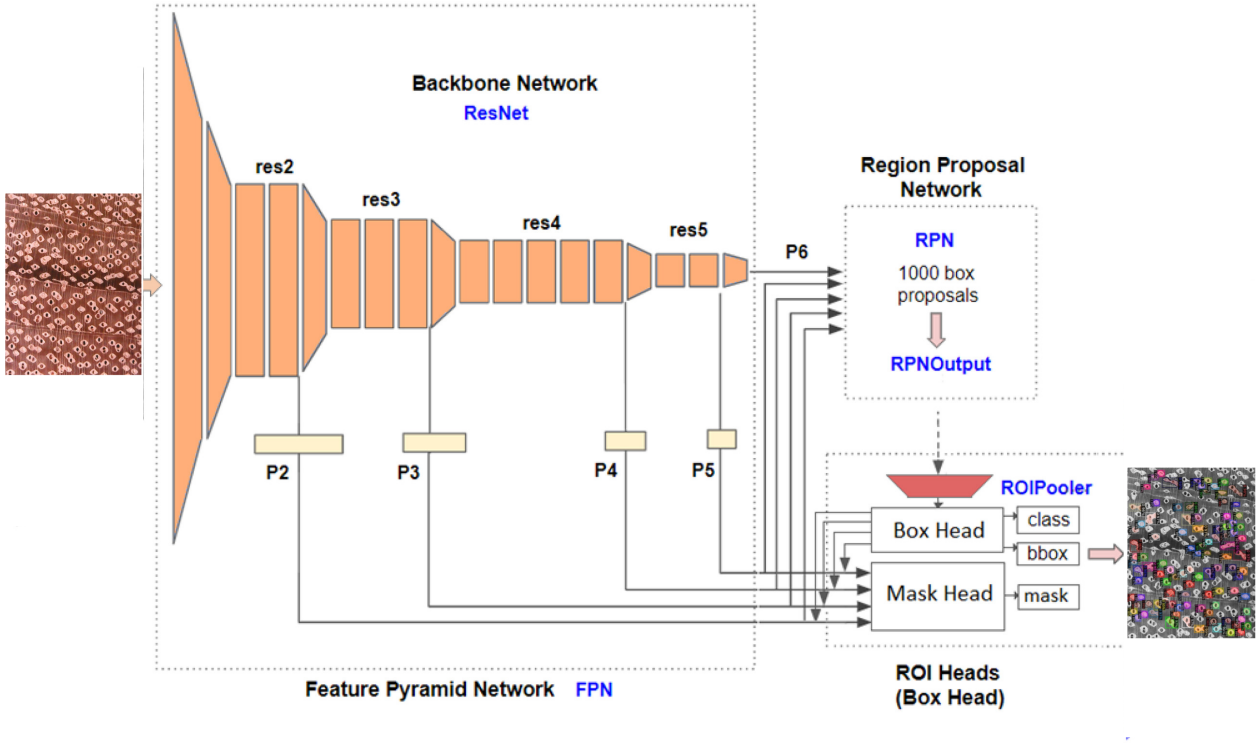


Fig. 1. Mask R-CNN architecture.

radial multiples, and pores embedded in aliform or confluent parenchyma without having to redraw masks when the surrounding tissue changes.

To segment these VCRs in 50× macroscopic images we employ Mask R-CNN [22], which couples a region-proposal network with a fully convolutional mask head and thus yields, in a single forward pass, bounding boxes, class labels (single vs. double) and pixel-accurate masks for every VCR detected.

Fig. 1 presents the architecture of the Mask R-CNN using in this study. The model employs a ResNet-50 backbone, which extract hierarchical features from an input RGB image at multiple scales. These features are processed by a Feature Pyramid Network, which generates multi-scale feature maps. This multi-scale representation captures anatomical details of varying sizes, such as small and large pores, by combining high-resolution features from early stages (e.g., P2 with H/4 × W/4) with semantically rich features from deeper stages (e.g., P6 with H/64 × W/64).

The Region Proposal Network leverages these FPN feature maps to generate candidate regions (box proposals) that may contain pores. Using an anchor-based approach, the RPN applies convolutional operations to propose regions of interest (RoIs). These proposals are then aligned with the feature maps using the RoI Pooler, which employs RoI Align—a precise pooling mechanism that preserves spatial accuracy by avoiding quantization errors.

The aligned features are fed into the RoI Heads, which consist of two components:

- Box head: Performs object classification (e.g., identifying a region as a pore) and bounding box regression to refine the coordinates of each detected object.
- Mask head: Generates binary instance masks for each pore, delineating their exact boundaries through convolutional layers.

This dual-head approach enables the model to detect, classify, and segment individual pores in a single pass. The final output is a segmented image where each pore is identified with a bounding box and a corresponding mask, facilitating detailed analysis.

Table I provides a detailed summary of the Mask R-CNN architecture employed in this study.

TABLE I. SUMMARY OF MASK R-CNN ARCHITECTURE

Component	Description
Backbone	ResNet-50 with stages res1 to res4, extracting hierarchical features at multiple scales
Feature Aggregation	Feature Pyramid Network (FPN) generating multi-scale feature maps P2 to P6
Region Proposal Network	Generates candidate regions (box proposals) from FPN feature maps using anchor-based method
RoI Align	Precise pooling with RoI Align, aligning region proposals to feature maps
Classification Head	Box Head with fully connected layers for object classification and bounding box regression
Mask Head	Convolutional layers generating binary instance masks for each detected pore

C. Wood Anatomical Feature Analysis

After segmenting VCRs from macroscopic wood cross-section images, we compute a series of anatomical descriptors to characterize and differentiate wood species. These descriptors—such as size, frequency, and spatial arrangement—are

well-established in wood anatomy and provide key diagnostic information. Because our VCR masks may include either the pore lumen alone or the lumen together with immediately adjacent perivascular parenchyma (when visually inseparable). All subsequent analysis is performed over these vessel-centric regions rather than the pore lumen alone.

1) VCR classification and counting: Each segmented VCR is categorized based on proximity and grouping patterns to reflect anatomical arrangements: solitary, multiple, or clustered. This classification follows adjacency checks, where two or more VCRs are considered grouped if they share a boundary or are positioned in close radial or tangential proximity.

a) Solitary VCR: A vessel-centric region that is spatially isolated.

b) Multiple/Clustered VCRs: Two or more VCRs appearing together, either in radial multiples or compact clusters.

Counting each category provides a basic anatomical fingerprint. For example, the predominance of solitary vessels versus radial multiples can suggest different taxonomic characteristics.

2) VCR Size measurement: As images are acquired at a fixed magnification (50×), we apply a known pixel-to-micrometer conversion factor to translate geometric measurements into physical units.

From each segmented VCR mask, we compute:

- Area: Total pixels within the VCR mask, converted into square micrometers.
- Equivalent Circular Diameter (ECD):

$$d = \sqrt{\frac{4A}{\pi}} \tag{1}$$

  where, A is the area of the VCR mask.

- Optional Shape Descriptors: Including elliptical fits (major or minor axis lengths) when more detailed morphology is needed.

These measurements provide a robust morphological overview of the segmented regions. By using VCRs—especially in diffuse-porous species where parenchyma halos often co-occur with vessels—we ensure more consistent and anatomically meaningful descriptors, even when vessel and parenchyma are difficult to distinguish in 50× imagery.

Fig. 2. Example images used in the classification task. The first row shows wood cross-section images at various magnifications. The second row contains non-wood images collected from general sources such as magnified surfaces of flowers, fruits.

IV. Experiments

A. Datasets

We employed two separate datasets for wood versus non-wood classification and segmentation tasks. For the wood/non-wood classification task, as illustrated in Fig. 2, we created two distinct subsets:

1) Wood dataset: This dataset aggregates macroscopic wood images collected from five distinct sources detailed in [3], including: i) VN_26, our self-collected dataset, ii) WRD_21, a Southeast Asian wood dataset published by Sun [23], iii) BFS_46, the Brazilian flora species dataset [24], iv) BD_11, a Brazilian dataset from 2021 [25], and v) PCA_11,

featuring wood species from the Pacific and Colombian Amazon regions [26]. In total, this dataset contains 20,000 images captured at magnifications ranging from 10x to 50x.

2) Non-Wood dataset: Approximately 9,000 images were gathered from online sources, depicting various subjects such as vehicles, landscapes, humans, and magnified images of non-wood objects.

The combined dataset was split into three subsets: 50% images for training, 20% for validation, and 30% for testing. Notably, we allocated a larger portion of the data to the test set to allow for more stable and representative performance evaluation, particularly given the diversity of both wood and non-wood images. This strategy was chosen to minimize variance in evaluation results and to better reflect real-world deployment conditions, where unseen data is common.

For the segmentation task, we utilized our self-collected subset (VN_26) exclusively at 50x magnification. Due to the extensive annotation required for segmentation, where the Labelme tool was employed to draw polygons around each pore, the labeling process was notably labor-intensive.

As presented, we clarify that every polygon corresponds to a VCR:

- Pore-only VCR: the lumen and its wall when no perivascular parenchyma collar is visible at the chosen magnification.
- Pore + PP (P+PP) VCR: the pore lumen together with the contiguous PP sheath if such a sheath clearly encircles the vessel(s).[1]

Hence, our masks deliberately "over-segment" to include PP whenever it is morphologically inseparable from the vessel on 50 × imagery, thereby preserving the original mask geometry without the need to re-annotate.

In total, 1,012 images were annotated, comprising approximately 34,500 single-VCR annotations and about 11,000 double-VCR annotations. Typically, each macroscopic wood image contained around 40 to 60 VCRs, highlighting the complexity and meticulous nature of the annotation work.

Given the substantial annotation effort involved, we adopted an iterative labeling strategy, progressively annotating smaller subsets of images while continuously evaluating the segmentation model's performance. Annotation was halted when model accuracy stabilized, ensuring efficient use of annotation resources without sacrificing model effectiveness. Additionally, the emphasis on single and double VCR facilitated focused improvements in segmentation accuracy, ultimately enhancing the robustness and practical usability of the resulting segmentation model.

B. Wood Macroscopic Image Classification

Because our objective is to detect if an image represents a magnified wood cross-section, we split data into training/validation/test sets at the feature level. In many real-world scenarios, it can be beneficial to have a small, curated set of "likely negatives" (e.g., images of plant bark or random textures) to help tune decision thresholds.

For each candidate model, hyperparameters (e.g., the number of PCA components, kernel parameters for SVM, and the number of estimators in Isolation Forest) are tuned based on performance metrics. In this study, we evaluate classification performance by standard metrics, accuracy, precision, recall, and F1-score, when negative examples are available. For one-class approaches, we track metrics such as the false-positive rate on non-wood samples to gauge how well the model excludes irrelevant data.

By combining DenseNet121, based feature extraction with robust classification techniques, be the binary or one-class, we ensure a flexible pipeline capable of handling the open-ended nature of "wood versus non-wood" classification. High-quality embeddings, coupled with dimensionality reduction and tailored classification algorithms, allow us to accurately filter and flag images for subsequent segmentation and anatomical feature analysis.

Fig. 3, Fig. 4, and Fig. 5 show the confusion matrices obtained from these models. In each matrix, the rows (actual labels) and columns (predicted labels) correspond to two classes: "0" (non-wood image) and "1" (wood cross-section image).

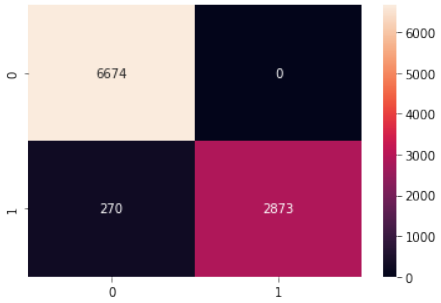


Fig. 3. Confusion matrix - OneSVM.

1) OneSVM: Although OneSVM achieves an impressive outcome of zero false positives (i.e., no non-wood images were misclassified as wood), it still has 270 false negatives. In other words, it fails to correctly identify 270 wood images as wood. This indicates a tendency toward under-classification of wood samples, which can be problematic in scenarios where missing a true wood specimen is undesirable.

2) Isolation forest: Isolation Forest exhibits a similar pattern to OneSVM, with zero false positives and a slightly higher number of false negatives (279). This small increase in false negatives yields a slight reduction in overall accuracy, but the difference between Model OneSVM and Isolation Forest performance is minimal in most practical contexts.

3) GMM: GMM stands out in that it achieves a significantly lower number of false negatives (only 2), thus correctly identifying nearly all true wood images. However, it does have a small number of false positives (1). From an overall performance standpoint, GMM demonstrates a much stronger capability to detect wood images without sacrificing non-wood accuracy by any large margin.

[1]In diffuse-porous and ring-porous taxa where parenchyma halos are diagnostic, this choice is anatomically meaningful because the pore + PP structure functions as a single macro-texture cue.

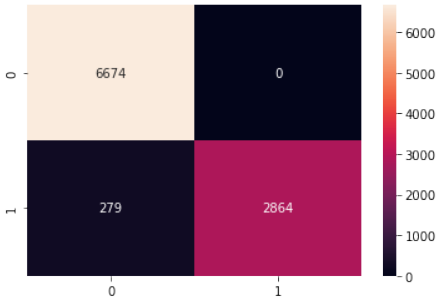


Fig. 4. Confusion matrix - IsolationForest.

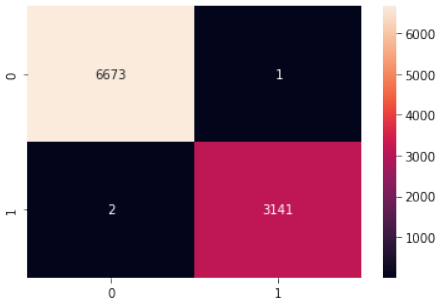


Fig. 5. Confusion matrix - GMM.

TABLE II. COMPARISON OF CLASSIFICATION METRICS FOR THE THREE MODELS ON THE TESTSET

Model	Accuracy	Precision	Recall	F_1-score
Isolation Forest	97%	98%	95.5%	96.5%
OneSVM	97%	98%	95.5%	97%
GMM	99%	99%	99%	99%

Table II compares the performance of three models. Isolation Forest and OneSVM have similar accuracy but struggle with higher false negatives, which drags down their recall. GMM, on the other hand, has almost no false negatives, boosting its recall and giving it a better F_1-score. In wood identification, missing a true wood sample (a false negative) is a bigger deal than accidentally labeling a non-wood image as wood (a false positive). So, keeping false negatives low is critical. With only 2 false negatives out of 3141 true positives, GMM stands out as the best option.

The results prove that combining transfer learning with solid classification can nail the task of separating wood from non-wood images. Isolation Forest and OneSVM play it safe (no false positives, but they miss more wood samples), while GMM strikes a great balance between recall and precision. In a two-step system, where classification feeds into pore segmentation, GMM makes sure most true wood images move forward. Looking ahead, tweaking feature extraction or anomaly detection could cut down errors even more and boost generalization.

C. Wood VCR Segmentation

Accurately segmenting individual wood pores from macroscopic images is inherently challenging due to the close adjacency of pore lumina and surrounding perivascular parenchyma. As discussed previously, we therefore adopted the concept of Vessel-Centric Regions, defined as a pore lumen together with any immediately adjacent PP tissue if present. Annotating each VCR required meticulously drawing polygons around each region using the Labelme tool, making this a particularly labor-intensive process. Each macroscopic wood image typically contained around 40 to 60 VCRs, leading to an estimated total of approximately 32,500 annotated VCRs across our 1,012-image dataset. These annotations comprised about 23,500 single-VCR and 9,000 double-VCR instances.

Given the substantial annotation effort involved, we adopted an iterative annotation strategy, incrementally annotating subsets of images and evaluating model performance continuously. Annotation was halted when the model's performance stabilized, thus efficiently balancing annotation cost and model accuracy.

Additionally, we augmented the dataset by cropping each annotated image into two smaller segments, effectively enhancing training-data diversity without additional annotation work. The numbers of annotated images and corresponding experimental results are summarized in Table III.

Two experimental scenarios were conducted to evaluate the performance of our segmentation model under different conditions, detailed as follows.

1) Scenario 1 Effect of training set size: In Scenario 1, we assessed accuracy improvements resulting from incremental increases in annotated training data. We employed the iterative annotation strategy outlined above, progressively adding labeled subsets until segmentation performance stabilized.

Six dataset sizes were evaluated, summarized in Table III, with each set split into training, validation, and testing subsets at a 70:15:15 ratio. Models were trained for 10,000 epochs, with a batch size of 512 and a learning rate of 10^{-4}.

Table III presents the accuracy achieved at various stages of this iterative process. Fig. 6 shows training progress, including accuracy and loss trends for classification and segmentation on both training and validation sets. Key observations include:

TABLE III. SCENARIO 1: SEGMENTATION ACCURACY WITH INCREASING TRAINING SET SIZES

No.	Train Images	Val Images	Test Images	Accuracy (%)
1	78	16	16	69.55
2	388	84	84	93.378
3	500	108	108	93.923
4	528	114	114	91.193
5	682	146	146	84.629
6	708	152	152	91.393

- Incremental Annotation versus Accuracy: Initially, the accuracy was low (69.55%) with only 78 annotated training images. Accuracy increased significantly to 93.378% and 93.923% as annotations reached 388 and 500 images, respectively, indicating the effectiveness of additional labeled data in early stages.

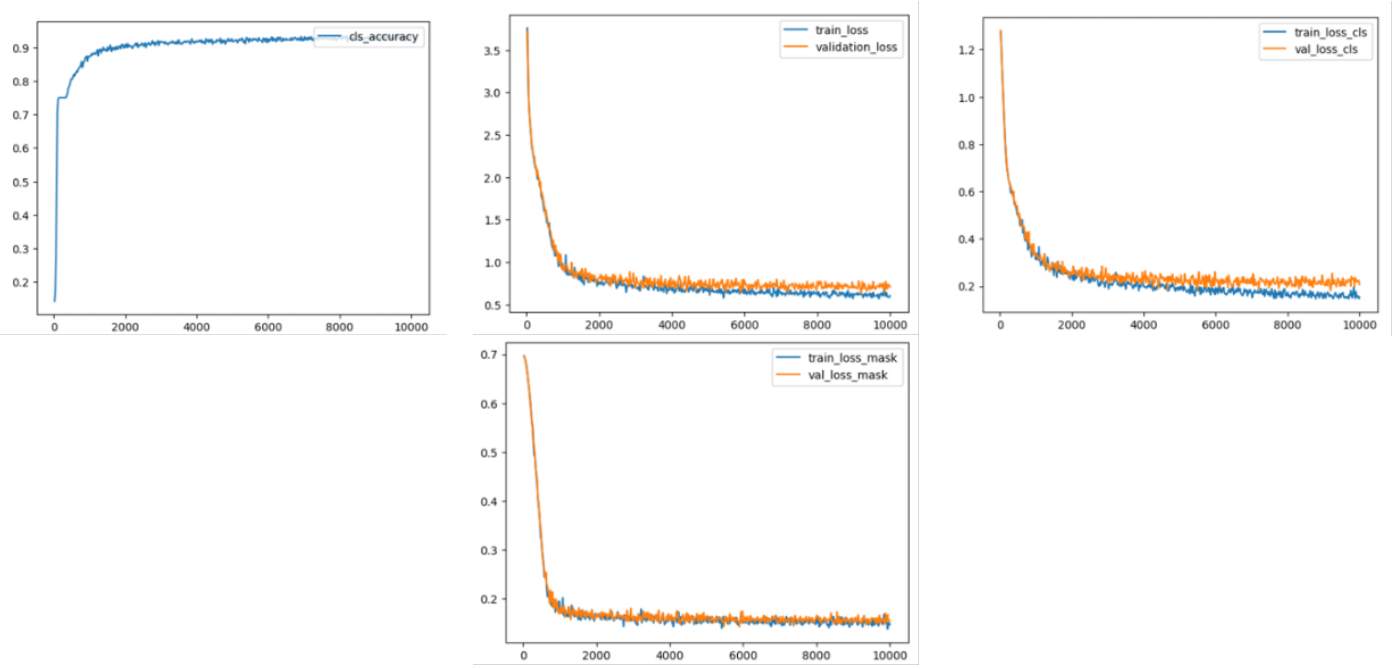


Fig. 6. Experiment 1 – Training loss and accuracy progresses with 500 training images.

- Incremental Annotation and Accuracy: Initially, accuracy was relatively low (69.55%) with just 78 annotated images. However, significant improvements were observed when annotations increased to 388 (93.378%) and 500 images (93.923%), demonstrating clear early-stage benefits of additional labeled data.
- Performance Saturation and Overfitting: Accuracy peaked at 500 annotated images. Beyond this point, accuracy fluctuated, declining to 84.629% at 682 images before improving slightly again at 708 images (91.393%). This indicates potential saturation and overfitting when adding excessively diverse annotations beyond an optimal point.
- Data Augmentation Effectiveness: The augmentation strategy of cropping annotated images successfully enhanced data diversity, mitigating accuracy declines observed at larger dataset sizes.

These results emphasize the practical effectiveness of iterative annotation and augmentation strategies for maximizing segmentation accuracy while minimizing annotation overhead. Future work may investigate adaptive annotation and augmentation strategies to further enhance model performance.

2) Scenario 2 Effect of number of epochs: In the second scenario, we investigated the impact of training duration on segmentation accuracy by varying the number of epochs (5,000; 10,000; and 15,000). This experiment employed 500 training images, 108 validation images, and 108 test images. Other training parameters remained constant (image size: 480×640, batch size: 512, learning rate: 10^{-4}). Table IV summarizes the accuracy results.

The results highlight several key insights:

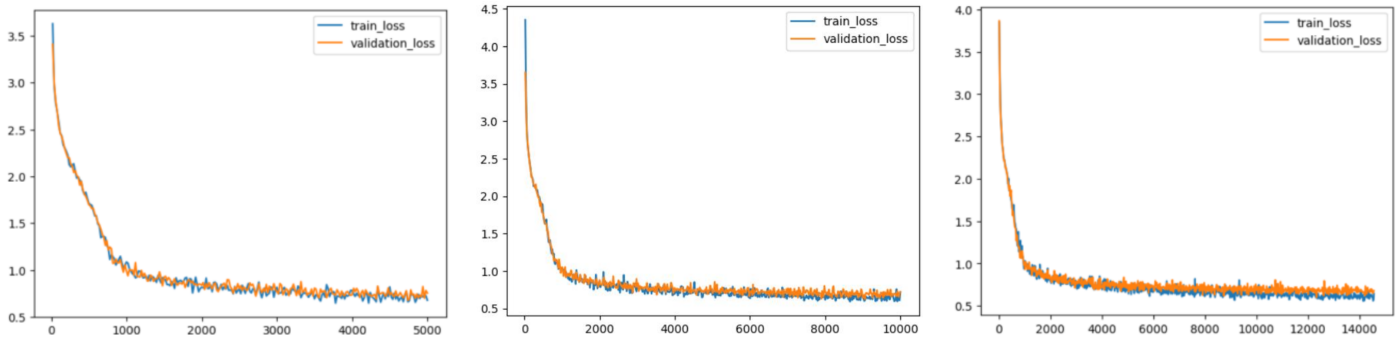
- Optimal performance at 10,000 epochs: Peak accuracy (93.923%) occurred at 10,000 epochs, slightly outperforming 5,000 epochs (93.677%). Further training to 15,000 epochs led to decreased accuracy (91.877%), indicating potential overfitting.
- Loss curve behavior: Analyzing loss curves (Fig. 7) revealed a consistent downward trend in loss, with diminishing accuracy returns beyond 10,000 epochs. Continued training beyond this point tended toward overfitting noise rather than learning meaningful anatomical patterns.
- Practical training recommendations: Balancing computational efficiency and accuracy, training around 10,000 epochs offers optimal segmentation performance without unnecessary resource expenditure.

TABLE IV. SCENARIO 2: TEST ACCURACY BY NUMBER OF TRAINING EPOCHS

Epochs	Accuracy (%)
5,000	93.677
10,000	93.923
15,000	91.877

Combining the findings of Scenarios 1 and 2, we observe the following:

- Optimal annotation size: A moderate dataset size (e.g., 500 images) yields optimal performance. Larger datasets may introduce complexity, potentially reducing accuracy without careful management.
- Balanced epoch selection: Training duration of approximately 10,000 epochs best balances accuracy gains against risks of overfitting.



(a) Loss progress at 5,000 epochs (b) Loss progress at 10,000 epochs (c) Loss progress at 14,000 epochs

Fig. 7. Training loss curves of segmentation models at different epoch milestones (5k, 10k, and 15k).

- Model robustness and generalization: Despite varying training conditions, accuracy consistently ranged between 84% and 94%, demonstrating robustness. Future improvements may involve advanced augmentation, regularization techniques, or specialized segmentation architectures.

These results confirm that strategic adjustments of annotation quantity and training epochs significantly affect segmentation performance, offering practical guidance for efficient annotation and training protocols.

V. Demonstration of End-to-End Wood Anatomical Analysis

To illustrate the practical value and usability of our models, we have implemented an end-to-end system for wood anatomical analysis. This system combines a lightweight Android mobile application with a remote server for advanced segmentation tasks, thereby demonstrating how classification and segmentation can be integrated into a real-world workflow.

A. System Overview

1) Mobile application - Frontend: The user-facing component is an Android application that performs on-device image classification to determine whether an input image depicts a magnified wood cross-section. Users can capture new images via the device camera or select existing images from the phone's gallery. For smartphones lacking sufficient native magnification, an external macro lens attachment can be used to acquire clear macroscopic wood images.

Upon pressing the "Analyze" button, the app uses a lightweight classifier trained as described in Section IV-B to check locally if the input represents a magnified wood cross-section. If the classification result is positive, the app forwards the image to the remote server for detailed segmentation and vessel-centric anatomical analysis.

2) Server segmentation - Backend: On the server side, our Mask R-CNN model (described in Section IV-C) receives the uploaded image and segments VCRs —each defined as a pore lumen plus its adjacent perivascular parenchyma, when PP is present. Specifically, the model segments and categorizes each identified region into:

- Single VCRs (solitary pore lumen $\pm$ PP),
- Double VCRs (two adjacent pore lumina sharing a common PP sheath or directly contacting each other).

From these segmented regions, the server computes a set of meaningful anatomical descriptors, including:

- Total number of single and double VCRs,
- Average VCR size in micrometers (derived from known magnification levels),
- Minimum and maximum sizes of VCRs,
- Inter-VCR distance metrics (minimum, maximum, and average distances),
- Summary statistics, such as spatial distribution and frequency patterns of VCRs.

B. Illustrative Results

To validate the end-to-end pipeline under realistic conditions, we tested the application using various wood species commonly encountered in Vietnam. Images were captured at magnifications of 10$\times$ and 50$\times$ under variable lighting conditions. The system consistently demonstrated:

- Reliable on-device classification, efficiently filtering out non-wood images,
- Robust server-based segmentation, clearly delineating VCRs even in complex anatomical arrangements,
- Rapid retrieval of detailed anatomical metrics, allowing users immediate insight into wood anatomy and supporting species identification.

Fig. 8 provides sample screenshots showing segmented VCR results and anatomical summaries for species such as Gõ Pachy, Cãm Hòng, and Ebiara at different magnifications. Users can view segmentation overlays directly in the app interface and access detailed textual descriptions of segmented regions.

The demonstrated architecture highlights a practical division of tasks:

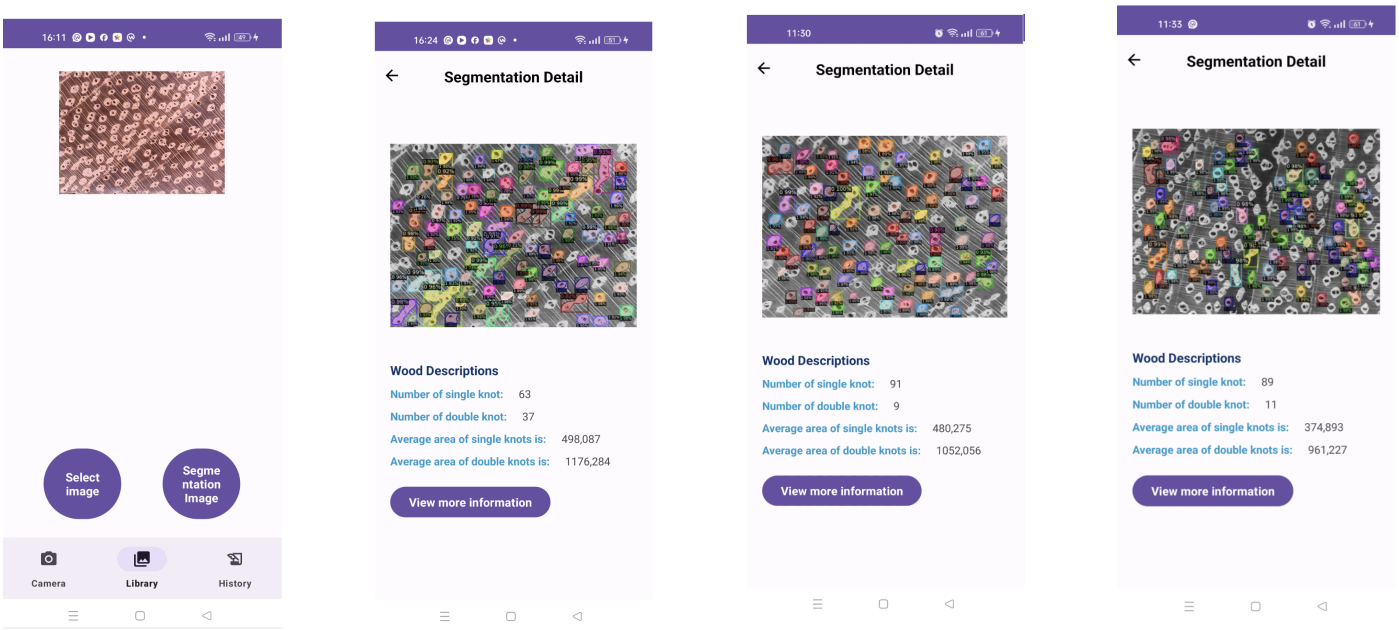


Fig. 8. Android app for end-to-end wood anatomical analysis.

- A compact classifier deployed on Android devices quickly verifies the suitability of images for further analysis,
- A computationally intensive segmentation model hosted on a remote server performs detailed VCR segmentation and returns comprehensive anatomical insights.

Overall, our end-to-end system provides a convenient, low-cost solution enabling wood analysts, researchers, and industry professionals to rapidly quantify critical anatomical features from macroscopic images. Future enhancements may include offline segmentation capabilities for fieldwork environments with limited connectivity or more advanced user interfaces for annotating and verifying newly encountered wood species. Additionally, this pipeline can be extended towards fully automated wood species identification by leveraging large language models to reason over extracted anatomical descriptors and incorporate established taxonomic knowledge.

VI. Conclusion

This study presented a two-stage automated framework for anatomical analysis of macroscopic wood cross-sectional images. Our approach combined image classification using DenseNet121-based features and a Gaussian Mixture Model classifier with segmentation of Vessel-Centric Regions (VCRs) via Mask R-CNN. Defining each VCR as a pore lumen plus any adjacent perivascular parenchyma (PP) allowed effective segmentation despite closely adjacent anatomical structures.

Experimental results confirmed that iterative annotation and data augmentation efficiently balanced annotation effort and segmentation accuracy, achieving optimal results (93.923%) with approximately 500 annotated images trained over 10,000 epochs. An end-to-end demonstration highlighted the practical usability of our system, combining mobile-based classification and remote segmentation. Future directions include offline segmentation for field use and automated species identification leveraging large language models.

References

[1] E. Wheeler, P. Baas, and P. Gasson, "Iawa list of microcopie features for hardwood identification," *IAWA journal / International Association of Wood Anatomists*, vol. 10, p. 219–332, 01 1989.

[2] R.-W. Bello, "Wood species identification using mask rcnn-residual network," *Pro Ligno*, vol. 19, no. 1, 2023.

[3] K. Nguyen-Trong, "Evaluation of wood species identification using cnn-based networks at different magnification levels," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.0140487

[4] A. Pratondo and A. Novianty, "Comparison of wood classification using machine learning," in *2022 IEEE 10th Conference on Systems, Process & Control (ICSPC)*. IEEE, 2022, pp. 308–312.

[5] K. Kobayashi, T. Kegasa, S.-W. Hwang, and J. Sugiyama, "Anatomical features of fagaceae wood statistically extracted by computer vision approaches: Some relationships with evolution," *PLOS ONE*, vol. 14, no. 8, pp. 1–13, 08 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0220762

[6] M. Khalid, E. L. Y. Lee, R. Yusof, and M. Nadaraj, "Design of an intelligent wood species recognition system," *International Journal of Simulation: Systems, Science & Technology*, vol. 9, no. 3, pp. 9–18, 2008.

[7] H.-J. Wang, G.-Q. Zhang, and H.-N. Qi, "Wood recognition using image texture features," *PLoS ONE*, vol. 8, no. 10, p. e76101, 2013.

[8] K. Kobayashi, M. Akada, T. Torigoe, S. Imazu, and J. Sugiyama, "Automated recognition of wood used in traditional japanese sculptures by texture analysis of their low-resolution computed tomography data," *Journal of Wood Science*, vol. 61, no. 6, pp. 630–640, 2015.

[9] I. Kirbas and A. Cifci, "An effective and fast solution for classification of wood species: a deep transfer learning approach," *Ecological Informatics*, vol. 69, p. 101633, 2022.

[10] F. Wu, R. Gazo, E. Haviarova, and B. Benes, "Wood identification based on longitudinal section images by using deep learning," *Wood Science and Technology*, vol. 55, pp. 553–563, 2021.

[11] X. J. Tang, Y. H. Tay, N. A. Siam, and S. C. Lim, "Mywood-id: Automated macroscopic wood identification system using smartphone and macro-lens," in *Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems*, 2018, pp. 37–43.

[12] Y. Sun, Q. Lin, X. He, Y. Zhao, F. Dai, J. Qiu, and Y. Cao, "Wood species recognition with small data: a deep learning approach," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 1451–1460, 2021.

[13] D. Herrera-Poyatos, A. H. Poyatos, R. M. Soldado, P. De Palacios, L. G. Esteban, A. G. Iruela, F. G. Fernández, and F. Herrera, "Deep learning methodology for the identification of wood species using high-resolution macroscopic images," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8.

[14] N. Rosa da Silva, M. De Ridder, J. M. Baetens, J. Van den Bulcke, M. Rousseau, O. Martinez Bruno, H. Beeckman, J. Van Acker, and B. De Baets, "Automated classification of wood transverse cross-section micro-imagery from 77 commercial central-african timber species," *Annals of forest science*, vol. 74, pp. 1–14, 2017.

[15] T. He, S. Mu, H. Zhou, and J. Hu, "Wood species identification based on an ensemble of deep convolution neural networks." *Wood Research*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233792800

[16] M. Hu, J. Zhang, L. Matkovic, T. Liu, and X. Yang, "Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions," *Journal of Applied Clinical Medical Physics*, vol. 24, no. 2, p. e13898, 2023.

[17] S. Zimeras and M. Diomidous, "2d anatomical structure for covid-19 medical images," *Studies in health technology and informatics*, vol. 295, pp. 542–544, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:250175683

[18] M. Berahim, N. A. Samsudin, and S. S. Nathan, "A review: Image analysis techniques to improve labeling accuracy of medical image classification," in *International Conference on Soft Computing and Data Mining*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:2095539

[19] H. R. Roth, C. T. Lee, H.-C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers, "Anatomy-specific classification of medical images using deep convolutional nets," *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 101–104, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:2484308

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, p. 60–88, Dec. 2017. [Online]. Available: http://dx.doi.org/10.1016/j.media.2017.07.005

[21] S.-W. Hwang and J. Sugiyama, "Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review," *Plant Methods*, vol. 17, no. 1, pp. 1–21, 2021.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2018. [Online]. Available: https://arxiv.org/abs/1703.06870

[23] Y. Sun, "Wood Recognition," https://github.com/sunyongke/woodRecognition, 2020, [Online; accessed 02-March-2023].

[24] D. V. Souza, J. X. Santos, H. C. Vieira, T. L. Naide, S. Nisgoski, and L. E. S. Oliveira, "An automatic recognition system of brazilian flora species based on textural features of macroscopic images of wood," *Wood Science and Technology*, vol. 54, no. 4, pp. 1065–1090, 2020.

[25] A. R. de Geus, A. R. Backes, A. B. Gontijo, G. H. Albuquerque, and J. R. Souza, "Amazon wood species classification: a comparison between deep learning and pre-designed features," *Wood Science and Technology*, vol. 55, pp. 857–872, 2021.

[26] D. A. Cano Saenz, C. F. Ordoñez Urbano, H. R. Gaitan Mesa, and R. Vargas-Cañas, "Tropical wood species recognition: A dataset of macroscopic images," *Data*, vol. 7, no. 8, p. 111, 2022.