Cent. Eur. For. J. 72 (2026) 171–184 DOI: 10.2478/forj-2026-0011

CENTRAL EUROPEAN Forestry Journal

ORIGINAL PAPER

Automatic and non-destructive inner wood quality assessment of wood logs using 3D computed tomography and machine learning

Gokul Kottilapurath Surendran[1†], Tomáš Bucha[2†], Tomáš Gergeľ[2], Ondrej Vacek[2, 3], Martin Mokros[4*]

[1]*Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, CZ- 165 000 Prague, Czech Republic*

[2]*National Forest Centre, Centre of Excellence of Forest-based Industry – LignoSilva, T. G. Masaryka 2175/22, SK- 960 01 Zvolen, Slovak Republic*

[3]*Technical University in Zvolen, Faculty of Wood Sciences and Technology, T. G. Masaryka 24, SK- 960 01 Zvolen, Slovak Republic*

[4]*University College London, Department of Geography, North-West Wing, Gower Street, UK-WC1E 6BT London, United Kingdom*

Gokul Kottilapurath Surendran, who believed in, "Do your duty; that's your beauty," and lived by those words. He was a passionate researcher in Artificial Intelligence, with interests in machine learning and computer vision. His Ph.D. research at the Czech University of Life Sciences Prague focused on developing machine learning approaches for tree species and wood classification. His love for forests and photography greatly influenced his research. He was always inspired by nature to develop new techniques for research, as reflected in his M.Phil. work on a "Natural Inspired Learning Path Recommendation System for Students in E-learning Platforms". During his Master's studies, he also explored brain tumor image analysis. He is remembered by his family and friends for his kind, humorous, and dedicated nature.

Abstract

Determining the quality of wood logs without cutting poses a significant challenge for boosting the productivity of wood suppliers and processors. Our study used 3D Computed Tomography scanning to inspect European Oak wood logs for inequalities. We scanned 26 oak logs with a resolution of 10 × 1 × 1 mm, producing an average of 390 to 430 images per log. These images were visually analysed and categorised into classes: Normal wood, Dead knot, Live knot, Rot, and Split. In the first phase, we investigated classification methods using machine learning for the two categories: non-defective and defective wood. In the second and third phases, we expanded our analysis to include all key defect classes: split, rot, dead and live knots. We utilised Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), k-Nearest Neighbor (k-NN), AdaBoost, and Gradient Boosting (GB). In the initial classification (non-defective and defective wood), we achieved 90% and 93% testing accuracy with DT and RF. Then, we aimed to improve accuracy by replacing the feature extractors and using a bagging voting classifier. The third phase used two feature selection approaches, LASSO feature selection and Mutual Information, and we obtained 88%, 95%, and 90% testing accuracies.

Key words: wood log classification; machine learning; 3D CT scanner; non-destructive log classification

Editor: Vojtěch Ondrejka

1. Introduction

Wood and wood products, including furniture, musical instruments, and barrels, are integral to our daily existence. In this competitive industry, ensuring quality is crucial. Quality checks have been conducted differently depending on the intended use of the wood logs. In earlier times, quality was assessed manually through

*Corresponding author. Martin Mokros, e-mail: m.mokros@ucl.ac.uk

© 2026 Authors. This is an open access article under the CC BY 4.0 license.

visual inspection by operators, focusing on external characteristics (Wollin & Vaughan 1947). This quality assurance process adhered to specific manual inspection and guideline protocols (Wengert & Meyer 1994; Malan 2003; Ridley-Ellis et al. 2016). This method was typically employed at logging sites, where technology could not be utilised. Here, visual characteristics such as knots, holes, and splits are critical in determining the wood logs' quality. Manual inspections are time-consuming and necessitate expertise and knowledge of standards, ensuring the remover of the assessor's subjective perspective. Nonetheless, the true quality of logs becomes apparent only after being processed in sawmills, leading to a reduced portion of higher-quality assortments and diminishing the economic value of the timber delivered.

The use of automated, technology-driven evaluations of wood log quality can ensure the efficient and optimal utilisation of wood resources. By incorporating technology into log quality assessments, researchers have employed computer vision techniques to automate the visual inspection process. In computer vision algorithms, photographs (input data) undergo segmentation, color, and texture analysis to assess the quality of logs to a certain extent. As early as the nineties, Sobey (1991), used neural networks and texture patterns to automate the grading of radiata pine timber using images taken with a CCD camera. Statistical measurements were calculated and fed into neural networks to classify areas as either clear or containing features. Feature types were then identified using thresholding techniques to segment knots, pith, bark, bluestain, holes, and wane. Results indicated that texture measurements enhanced classification accuracy compared to tonal measures alone. Alapuranen & Westman (1992) introduced a new hierarchical vector-connected component (HVCC) multi-stage region growing segmentation method to identify defective plywood boards with greater success rates. Their experiments demonstrated that HVCC was most effective at separating clear wood from defects, while texture-based methods were inadequate for detecting internal defects. Marszalec & Pietikaeinen, (1993) adopted a non-segmentation approach, calculating tonal properties and texture measures for sub-image regions, and then applying statistical pattern recognition. They discovered that greyscale and color were the most significant features, potentially replacing computationally intensive texture measures. Lee et al. (2003) proposed identifying wane in rough hardwood lumber using profile and greyscale images obtained from laser and video cameras. The authors addressed noise and debris in laser-based imaging by combining column-wise image statistics, selective smoothing, and surface shape analysis to identify wane boundaries. Their method, using least-square fitting and threshold-based techniques, increased wane detection accuracy and optimised edge and trim cuts to maximise the board's value and enhance quality control in lumber production. In light of the latest research in this field, a study conducted by Wimmer Georgand and Schraml in 2021 (Wimmer & Schraml 2021) introduced a dual-phase convolutional neural network (CNN) model that integrates Mask R-CNN for segmenting log ends and a triplet loss-based SqueezeNet for recognition. This method, validated on multiple datasets, achieves high recognition accuracy (EER $\leq$ 3.4%) while eliminating the need for manual alignment, demonstrating a fully automated system with strong generalisation potential in real-world applications. Another study (Pyörälä et al. 2019) explores terrestrial laser scanning (TLS) to assess log geometry and wood quality, indicating that measurements of diameter, taper, and volume obtained from TLS closely match sawmill data, with taper negatively correlated with density and knottiness. Although the research identifies TLS as a reliable tool for pre-harvest evaluation, difficulties in estimating sweep and moderate correlations between geometry and quality suggest there is room for advancement. A new method (Yu et al. 2023) integrates Mask R-CNN instance segmentation with a stereo camera to provide an effective solution for identify log outlines and measure dimensions according to Chinese standards. With a detection success rate of 98.2% and diameter errors reduced to just 5.7 mm, this approach is recognised for its reliability in challenging environments, outperforming ellipse-fitting and other conventional methods. These findings highlight advancements in automatic wood measurement, addressing issues such as complex backgrounds, the inefficiency of manual labor, and errors found in traditional techniques.

Utilising acoustic methods (Dickson et al. 2005; Wang et al. 2007; Amishev & Murphy 2008; Mohammadali Saadatnia Aliakbar Enayati & Moradian 2016; Olaoye & Ojo 2022), it is possible to assess the quality of standing trees and logs to some extent. While this approach offers a more detailed visualisation of the internal structures of wood logs (Gejdoš et al. 2023), it is time-intensive and not ideal for lengthy wood logs.

Consequently, research has shifted towards more advanced methods, including X-ray applications and various computed tomography techniques (Breinig et al. 2015; Jori Uusitalo Olli Ylhäisi & Makkonen 2018). The operational use of CT scanners necessitates efficient and quick classification of defects in wood logs. For instance, Krähenbühl et al. (2012) employed machine learning techniques such as image segmentation and thresholding to detect pith, then used intensity variations between successive CT slices along with a specific intensity histogram to identify knots in wood logs. Khazem et al. (2023) introduced a deep CNN comprising three end-to-end trainable pipelines. Initially, the pipeline is dedicated to center detection using a CNN architecture with a regression head. The second and third pipelines deal with contour estimation and knot detection, akin to binary segmentation based on an Encoder-Decoder architecture. This study showed encouraging results

across different tree species. There is an especially urgent need for algorithms to reliably detect defects in broadleaved trees, which have not been adequately developed so far.

In recent years, significant contributions have been made to the field of automatic defect detection using CT images. In 2012, Boukadida et al. (2012), introduced an innovative algorithm for detecting pith, utilising 17 different species to demonstrate the algorithm's versatility with various tree species and wood traits. Longuetaud et al. (2012), study concentrated on identifying knots in softwood samples, with their model successfully detecting knots even in low-contrast and noisy CT images. Additionally, the 2020 research by Gazo et al. (2020) featured rapid pith detection in hardwood logs. Furthermore, the latest study by Vacek et al. in 2024 presented a combined method for the automatic classification of wood species and pith detection from CT log images, delivering encouraging outcomes across multiple wood species.

Moreover, the adoption of cutting-edge technologies highlights a case study carried out in a forestry company in Slovakia (Gejdoš & Gergeľ 2022). This research investigated the techniques and standards employed for categorising wood, including factors like species, quality, size, and intended usage. During the experiments, it was found that 36.3% of the logs were classified the same by both their method and the enterprise's quality assessment method, whereas 63.7% of the logs were classified differently. The majority of the logs were downgraded to lower quality categories.

The literature reveals a gap in research concerning the automated detection of multiple defects, particularly from CT scan images. Additionally, there are only specialised software tools available for detecting these defects collectively, but they lack reproducibility in their methods. To address these challenges, we implemented a groundbreaking approach: using three-dimensional Computed Tomography (3D CT) scanning technology, which delivers highly detailed cross-sectional images enabling precise defect detection at a macroscopic scale. Our goal is to enhance understanding of wood quality assessment through the development of machine learning algorithms such as support vector machines, random forests, decision trees, k-nearest neighbor, adaboost, and gradient boosting. The novelty of this work lies in three aspects. First, it operates directly on real industrial 3D CT data acquired from a MiCROTEC CT Log scanner for European Oak, a hardwood species for which automated internal-defect detection remains substantially less developed than for softwoods. Second, it implements a fully software-independent and reproducible classification pipeline, rather than relying on the closed algorithms embedded in the scanner manufacturer's software. Third, it combines CNN-based feature extraction with LASSO and Mutual Information feature selection to discriminate several defect classes (dead knot, live knot, rot, and split) simultaneously, extending earlier CT-based studies that have largely targeted a single internal feature such as pith or knots. The study concentrated on accomplishing two primary objectives:

- The study examines the ability of machine learning algorithms to distinguish between wood that is defective and non-defective.
- The subsequent study aims to categorise all major defect types, including splits, rot, and both dead and live knots.

This paper is structured in the following way: the second section offers a detailed explanation of the operational principles of 3D CT scanners and presents an overview of the specific 3D CT scanner used in this study. This is followed by the stages of data preprocessing and a broad overview of the data format. The third section presents the outcomes obtained using different machine learning algorithms and models. The final sections address the discussion and conclusion of the proposed research.

2. Materials and methods

This chapter is organised into distinct sections to offer more details on the techniques for data collection, dataset preparation, and algorithm selection. The initial subsection (2.1) explains the operational principles of a 3D CT scanner. The following section details the operating principles of the MicroTec CT scanner employed in our research. Subsection 2.3 describes the data collection procedure, mainly how the dataset is created for computer vision applications. The concluding section provides an overview of the label assignment on the collected datasets and the machine learning models employed for classification tasks, along with the system configurations considered.

2.1. Principles of 3D CT scanning

X-ray computed tomography is commonly used in the medical field. Recently, it has been extensively used for the non-destructive evaluation of industrial components and materials (3D Computed Tomography – See through Your Parts Non-Destructively 2024; Home – Comet Yxlon 2024; Industrial CT & X-Ray Solutions from ZEISS 2024; Industries 2024; MiCROTEC – CT Log 2024). This technology operates on the principle of X-ray absorption. As X-rays pass through a wood log from different angles, they interact with tissues, with the level of absorption depending on tissue density. Denser tissues, such as moister wood (like knots or sapwood), absorb more X-rays compared to less dense tissues, such as heartwood in freshly cut coniferous wood. Using these

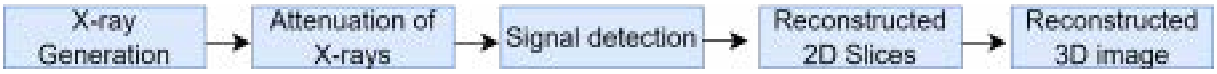


Fig. 1. Semantic workflow of a 3D CT scanner: This illustrates the different stages involved in creating 3D images using CT scanner data.

absorption details, CT scanners can create detailed images of the internal configurations of the wood log.

Refer to the semantic workflow diagram (Fig. 1). The apparatus employs revolving X-ray machines to create intricate images of an object's internal structure. During this procedure, as the X-ray machine moves around the object, it emits a series of X-ray beams from various angles. On the opposing side of the scanner, detectors capture the radiation passing through the object, resulting in a detailed image of the scanned section of the object. Thereafter, a specialised computer stores this data and processes the images to form a three-dimensional representation of the object.

The upcoming subsection will offer a comprehensive summary of the 3D CT scanner evaluated for this study.

2.2. MiCROTEC's 3D CT log computed tomography scanner

The CT scanner functions on a fixed framework that holds a rotating gantry using a sophisticated large cone beam reconstruction algorithm. In this setup, X-rays illuminate across the length passing through the scanner, with differences in radiation absorption revealing density at various points in the wood. Each segment produces an image showing the density distribution via the tomographic inversion algorithm. The operation uses a 360-degree X-ray computed tomography process, where the equipment rotates around the logs (Fig. 2).

As logs move toward the apparatus, X-ray beams are emitted from every angle (360 degrees), reaching a speed of 180 m/min (590 ft/min) when operating at full capacity, which allows for swift, complete digital 3D reconstruction. This extensive scanning method makes it possible to identify different flaws in the logs, like resin pockets, grain slopes, sound knots, splits, and dead knots. This ability greatly improves our quality control and optimisation efforts in handling large logs.

In our study, we used the 3D CT Scanner (MiCROTEC – CT Log 2024), which is fully integrated into the partially operational scanning line at the National Forest Centre in Zvolen, established in 2023 (illustrated in Fig. 3).

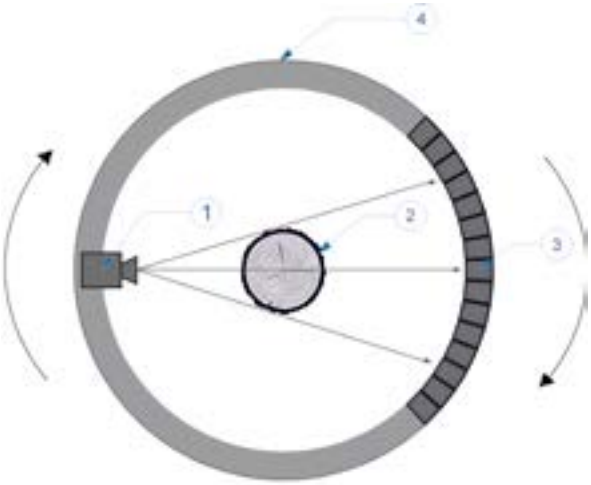


Fig. 2. 3D CT Scanner operational setup: First, the log is positioned in the scanner (2), after which the rotating component (4) circles the logs; during this time, the X-ray source (1) emits beams, and the detector (3) captures the data.



Fig. 3. The MiCROTEC 3D CT scanner employed in our research. Logs are transported by conveyor belts, scanned by the CT scanner, and subsequently transferred to the second conveyor belt.

We established particular scanning parameters: the X-ray tube operated at a voltage of 200 kV and a current of 14.4 mA, with the scan running at 195 rpm. The logs advanced towards the scanner at a rate of 5 meters per minute, and the scans featured a spatial resolution of 10 mm in the longitudinal dimension and 1 × 1 mm in the transverse dimension.

Once the scanning is successfully completed, we must preprocess the data using Microtec's software (MiCROTEC – CT Log 2024) for preprocessing and visualisation. This tool performs an in-depth analysis and then detects characteristics and attributes of the scanned logs, such as knots, splits, ring shakes, among others.

Although the error detection algorithm for conifers is adjusted to a degree that permits operational use, the precision in identifying errors for deciduous trees is lower. In our research, the data was exported in TIFF image format (Fig. 4.) to facilitate quicker examination after scanning. Furthermore, it is important to highlight that each tree log requires a manual export process. As each log is scanned separately and all sections are processed within the CT scanner, the individual scanned images are compiled into unified TIFF files. For our experiment, we need to extract each image from a single TIFF file. To fulfill this need, a MATLAB script is utilised to systematically divide the complete TIFF image into separate JPG images.

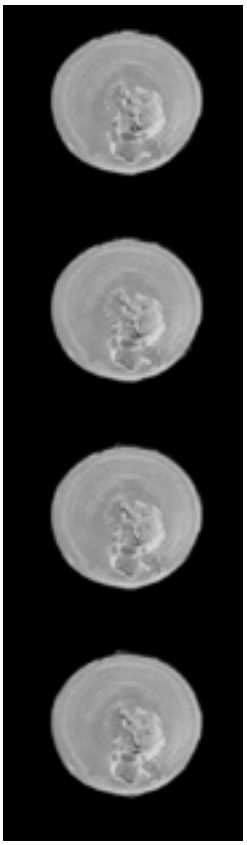


Fig. 4. Log scan in TIFF format: This illustration depicts just a small segment (4 sections) of a single log post-saving.

2.3. Label assignment and algorithm selection

This section will discuss the labeling phase and the algorithms used for this research. The produced data is crucial for our machine learning classification experiments because we intend to carry out classification tasks independently, without utilising separate software. During this experimentation phase, we have successfully generated two distinct datasets consisting of Oak species (26 individual logs). The first dataset includes two categories: "Normal" with 1,690 instances and "Defective" with 8,034 instances, providing a total of 9,722 images. The second dataset consists of 713 images divided into five categories: "Dead Knot" (119), "Live Knot" (177), "Normal log" (81), "Rot" (184), and "Split" (152). Fig. 5 below offers a detailed overview for a better understanding of the dataset.

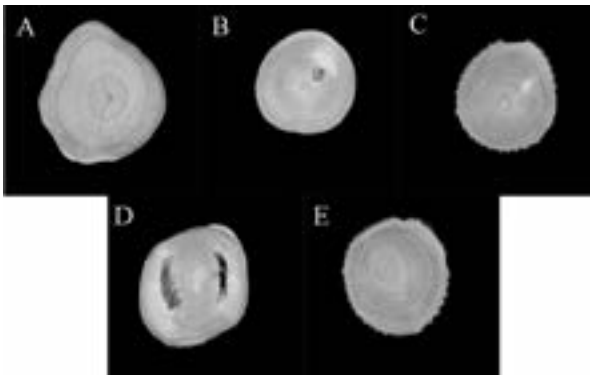


Fig. 5. Categories of the dataset: The initial log (A) displays an initial defect, a crack on an Oak wood log, while the second log (B) indicates the Death Knot. The third image (C) illustrates a Live Knot, and the fourth one (D) depicts Rot. Image E shows a log with no defects, labeled as the Normal log.

The proposed study is structured in three phases (Fig. 6). In the first phase, the study was divided into two experiments. In the initial experiment, we focused on two classes; all the defects mentioned earlier were combined into one category called defective, with the second class labeled as normal. This research utilises two machine learning algorithms – Random Forest and Decision Tree – necessitating the selection of suitable feature extraction methods. We employed the Gray Level Co-occurrence Matrix (GLCM) (Haralick 1973) as the feature extractor, with predefined parameters for both classifiers. We investigated five specific textural features: contrast, dissimilarity, homogeneity, correlation, and energy. To extract these features, five GLCM windows ([11] [0], [30] [0], [40] [0], [50] [0], & [60] [0]) with zero radians were used. After feature extraction, all features were applied to the machine learning algorithms for the classification process. In this experiment, 80% of the dataset was used for training, while the remainder was set aside for testing. The 80/20 partition was

performed at the image level, with individual CT cross-sectional images assigned at random to the training and testing sets. Because many closely spaced cross-sections are extracted from each log, cross-sections originating from the same log may therefore occur in both partitions; the effect of this image-level splitting on the reported accuracies is addressed in the Limitations and Future Work section.

In the next section of this phase, we are dealing with a second dataset. Instead of two classes, we now have multiple classes, which are listed below:

i. Dead knot
ii. Live knot
iii. Normal log
iv. Rot
v. Split

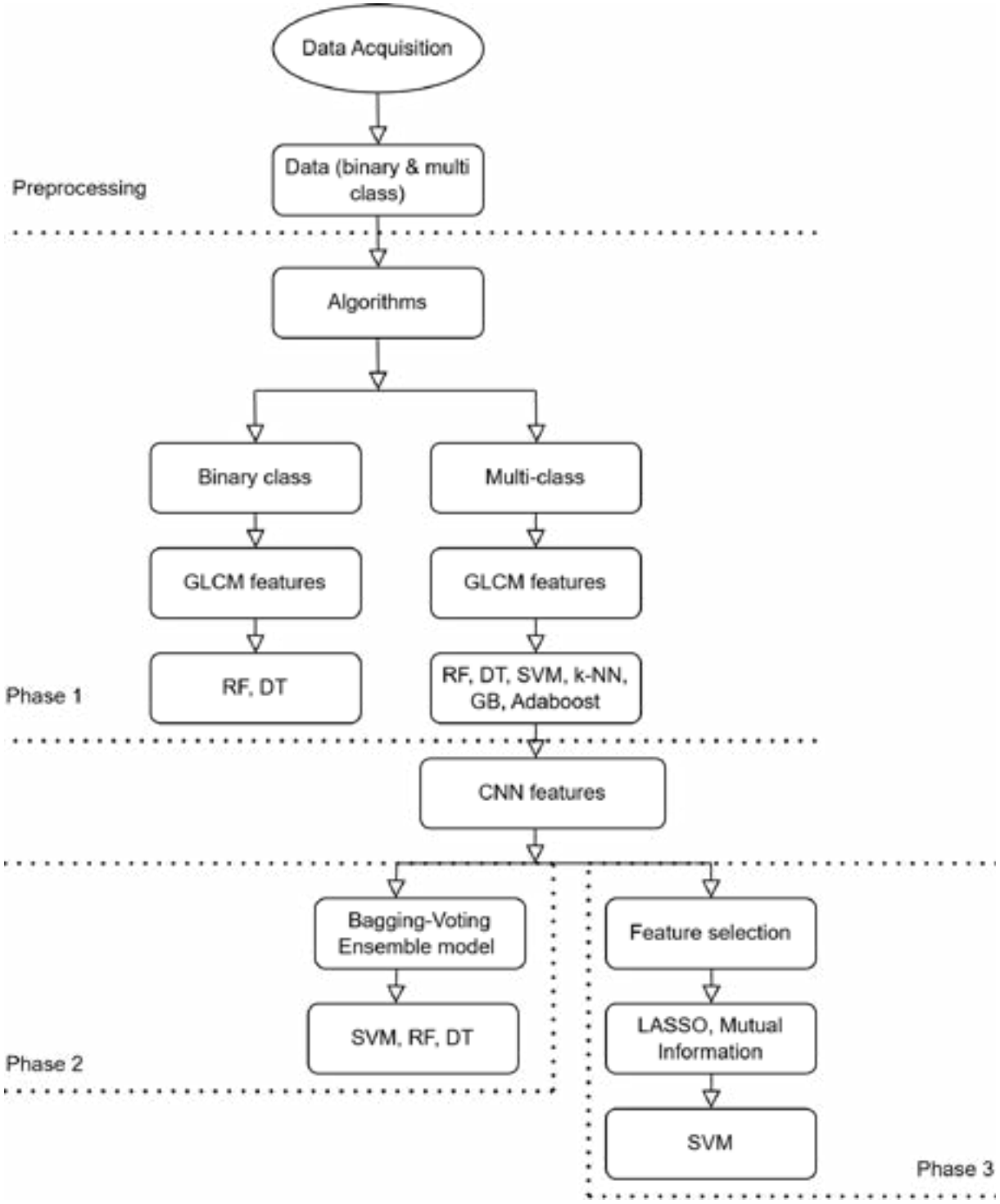


Fig. 6. Proposed study design.

Each class signifies a distinct category in the classification procedure, thereby expanding the breadth and complexity of our analysis. In this regard, we have opted to examine images that have just one defect. The primary reason for this is that having multiple defects in a single image shifts the task from a classification problem to the more complex challenge of detection. Additionally, during the preliminary phase of our research, all defective categories were combined into one class. Hence, we are viewing this detection task as a significant focus for our future work.

In this sub-phase, GLCM is also used for feature extraction. As in the initial experiment, 80% of the dataset is allocated for training, while the remaining 20% is reserved for testing. Additionally, this phase incorporates six distinct machine learning algorithms into the classification workflow.

i. Random Forest (RF)
ii. Support Vector Machine (SVM)
iii. k-Nearest Neighbors (k-NN)
iv. Decision Tree (DT)
v. AdaBoost
vi. Gradient Boosting (GB)

We anticipate that relying solely on classical machine learning algorithms for this classification reduces the likelihood of achieving high accuracy. The primary limitation lies in the complexity of input data. Since all image categories are in greyscale, distinguishing between various defects and normal classes poses a significant challenge.

In the previous experiments, we used GLCM as a feature extractor but hypothesise that it may not always deliver high accuracy. To address this issue and improve classification performance, we focused on adopting alternative methods. Initially, we replaced the GLCM feature extractor with Convolutional Neural Networks (CNN). For this, we employed a five-layer convolutional neural network with a 3 × 3 kernel and 2 × 2 max-pooling in each intermediate layer for feature extraction. The resulting feature maps from the convolutional layers are displayed in Fig. 7.

Following the feature extraction process, we implemented two additional phases for classification. The first model (phase 2) combined bagging (Skurichinars & Duin 1998) and voting ensemble models (Kotsiantis & Pintelas 2004; Leon et al. 2017) for the classification task. In this model, SVM, DT, and RF algorithms create the ensemble. We first apply bagging techniques to these algorithms and then use the voting ensemble model after the bagging process.

In the third phase of study, we applied two feature selection algorithms to the CNN feature sets. First, we used the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996) feature selection algorithm with the SVM algorithms for classification. Finally, we applied the Mutual Information (Nguyen et al. 2016; Bhuyan et al. 2023) algorithm in combination with an SVM classifier for feature selection. For model evaluation, we used four accuracy metrics: Precision, Recall, F1-score (Powers 2011) and Accuracy, which we calculated using the 'score' method in the scikit-learn library. The equation for calculating accuracy using the score method is provided below.

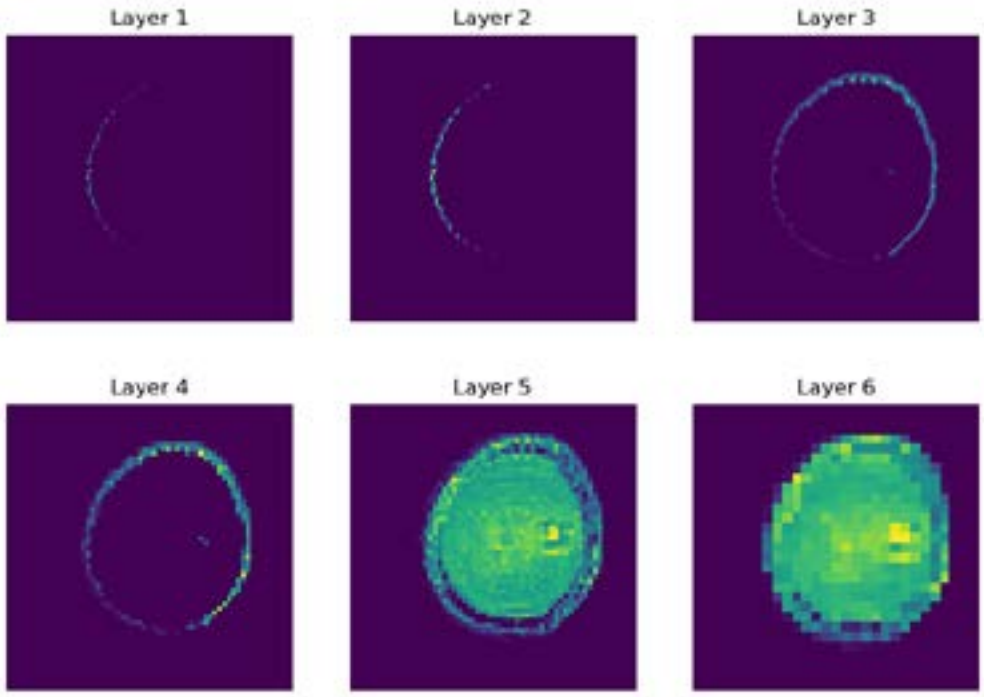


Fig. 7. CNN feature maps: The Figure presents feature maps from five convolutional layers (1–5) and the flattened layer (layer 6) on a random input image. Each layer determines edges and patterns, showcasing the progression of feature extraction.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad [1]$$

In this experiment, we used two different computers. For the first phase, we employed a desktop computer with an Intel 11th generation i7 processor and 64 GB of RAM. In the second and third research phase, we had less extensive data compared to the first phase, so we used a standard laptop, specifically the Dell XPS 13, equipped with an Intel 12th generation i7 processor and 16 GB of RAM.

The following section presents the results achieved on these datasets using various machine learning algorithms.

3. Results

3.1. Results of the first experiment in the initial phase

In the first experiment phase of the research, we used a dataset with two classes. We initially grouped all defects and categorised them into two classes: the 'Defective' class, which indicates the presence of defects, and the 'Normal' class, which signifies the absence of defects. Using DT and RF classifiers for the classification task, we achieved testing accuracies of 90% and 93% respectively. The RF algorithm delivered better accuracy, while the DT also performed effectively on these datasets. Table 1 provides the detailed accuracy metrics.

Table 1. Results of the first experiment (Phase 1): Classification of defect presence or absence on wood logs.

Algorithm	Training accuracy (%)	Testing accuracy (%)	Precision score	Recall score	F1-Score
Decision Tree	100	90	0.902	0.903	0.903
Random Forest	100	93	0.932	0.934	0.931

The confusion matrix analysis highlights a significant number of misclassifications, primarily due to the difficulty of differentiating between normal and defective logs, as their appearances in the CT scanner are only slightly distinct. In the DT results (Fig. 8 (A)), 257 images from the defective class were correctly classified, while 100 were misclassified as normal. Similarly, 1501 images from the normal class were correctly identified, with 87 misclassified. In contrast, the RF classifier showed improved performance, misclassifying only 92 images from the defective class and 36 from the normal class (Figure 8 (B)). These results demonstrate that the RF classifier achieved more accurate classifications overall.

Results of the second experiment in the initial phase

In the second experiment, we focused on classifying the five categories described in subsection 2.4. During this phase, we directly applied machine learning algorithms (as outlined in subsection 2.4) to the dataset. Among all the algorithms evaluated, only the RF algorithm achieved higher accuracy, delivering 81% accuracy in testing. Analysing the class-wise performance revealed that the classifier demonstrated effective classification overall but encountered notable misclassifications across three main classes (Fig. 9).

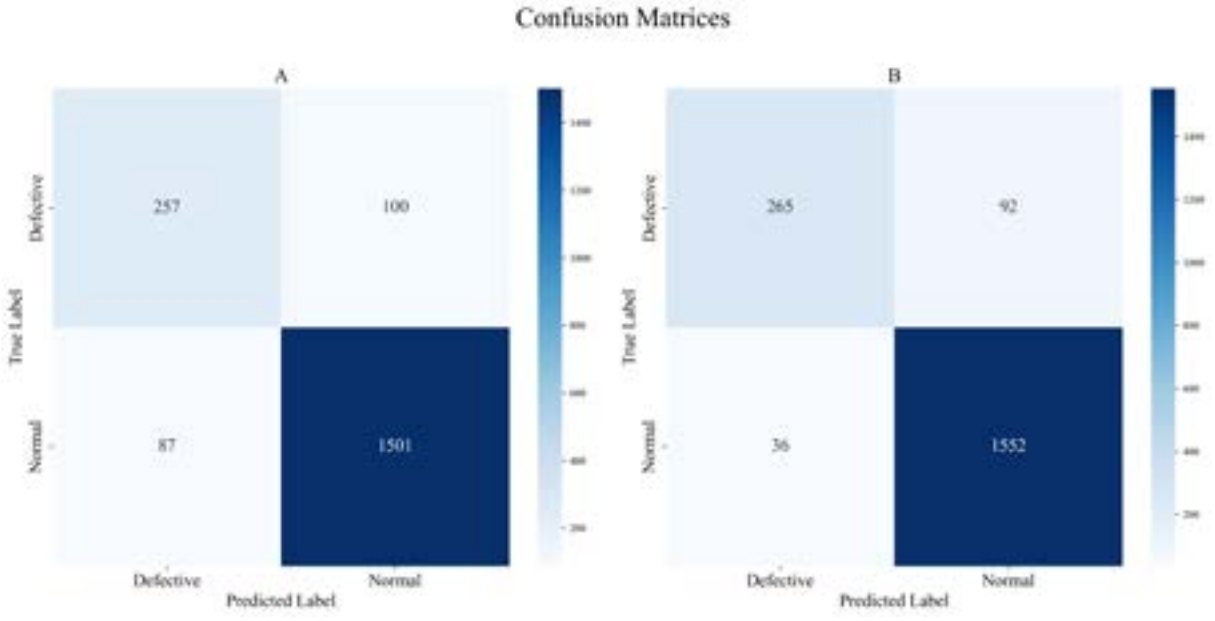


Fig. 8. Confusion matrices for first phase classification (first experiment): (A) Confusion matrix results for the decision tree classifier; (B) Confusion matrix results for the random forest classifier.

Six images from the "Dead knot" class were misclassified as "Live knot"; and two images were classified as "Rot." For the "Live knot" class, there were significant misclassifications, with most errors involving images being classified as "Dead knot." Additionally, two images from this class were misclassified as "Rot," and one as "Split." For the third class, "Normal," seven images were incorrectly classified as "Live knot". For the "Rot" class, the predictions were entirely accurate, with no misclassifications. In the final class, "Split," some images were incorrectly classified as "Live knot," and one image was classified as "Normal." Despite these misclassifications, the RF classifier showed strong overall performance.

Fig. 10 compares the accuracies of various classifiers, revealing that two classifiers did not meet commonly

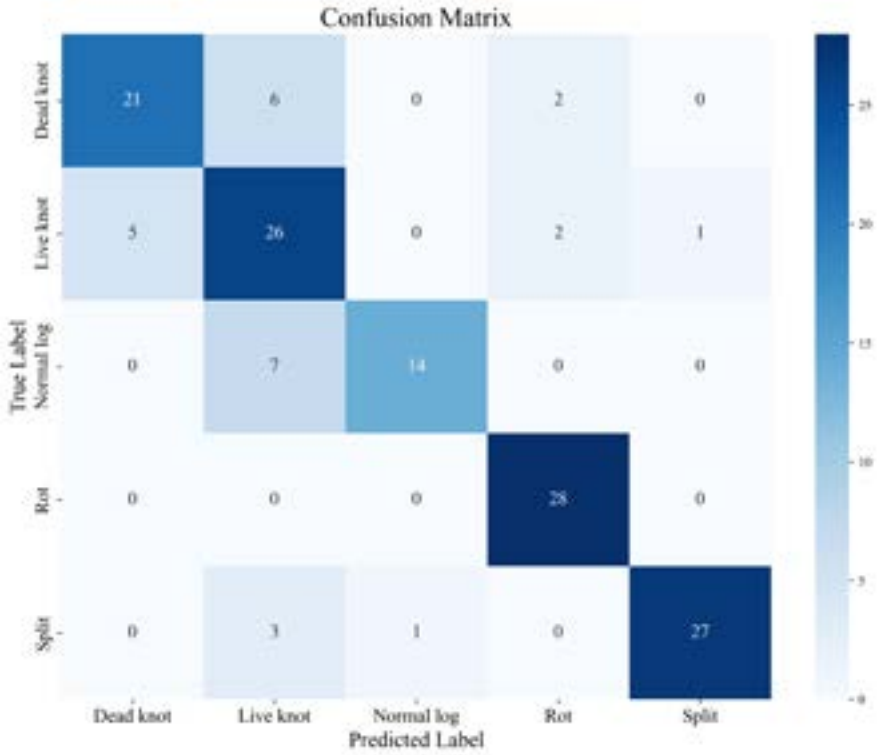


Fig. 9. Confusion matrix of random forest classifier (Second phase classification): This analysis focuses on classifying five distinct classes, including the normal class, rather than distinguishing between normal and defective.

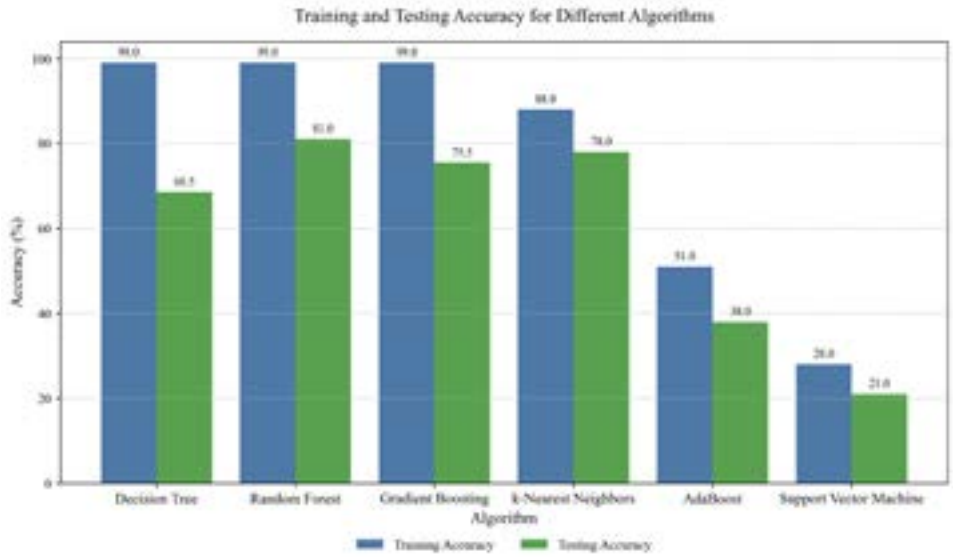


Fig. 10. Accuracy of classical machine learning algorithms: Six machine learning algorithms are applied to the second dataset. The plot illustrates the corresponding training and testing accuracies for each algorithm.

accepted accuracy standards. The first, the SVM classifier, showed notably lower performance, achieving only 28% training accuracy and 21% testing accuracy, indicating an inability to learn from the dataset. The second, AdaBoost, performed similarly poorly with 51% training accuracy and 38% testing accuracy.

Four other classifiers achieved standard accuracy levels. Notably, the DT and GB classifiers matched the RF classifier's training accuracy. However, the DT classifier reached only 68.5% testing accuracy, while the GB classifier achieved 75.5% testing accuracy. The k-NN classifier produced different results, with 88% of training accuracy and 78% testing accuracy, performing better on the multiclass classification problem. Upon reviewing the results, the RF classifier clearly outperforms the others, achieving 99% training accuracy and 81% a testing accuracy.

3.2. Results of phase two and three experiments

In the preceding phase of the research, we observed that SVM and AdaBoost classifiers failed to achieve standard accuracy (Fig. 10), primarily due to the complexity of the input data. However, the ensemble models designed to enhance the performance of classical machine learning algorithms yielded promising results. The initial bagging voting classifier model achieved 99% training accuracy and 88% testing accuracy (Fig. 11), along with 88% accuracy for precision, recall, and f-1 scores.

The feature selection models applied to the CNN feature sets (phase 3) also exceeded expectations. The LASSO model achieved the highest testing accuracy of 95% and a training accuracy of 99%. The mutual information model also delivered strong results, with 96% training accuracy and 90% testing accuracy. Compared to the previous results (GLCM), the SVM algorithm showed improvements, surpassing both RF and DT classifiers.

This phase of research demonstrates that selecting appropriate feature extractors and feature selection methods can significantly enhance the accuracy of classical machine learning algorithms. Additionally, ensemble models contribute to improved classification accuracy, although their effectiveness depends on the specific algorithms employed.

4. Discussion

Assessing the quality of wood logs without cutting or splitting them presents a major challenge in enhancing the yield for wood suppliers and processors. This is crucial for identifying costly, high-quality wood logs needed for particular uses, such as crafting musical instruments, barrels, and similar products. Ensuring the quality of wood logs also significantly contributes to cutting down carbon emissions and minimizing wood waste.

The advanced technology based on CT scanning offers superior outcomes for non-destructive testing of wood logs. Nonetheless, the implemented CT technology exclusively supported the machine learning algorithms

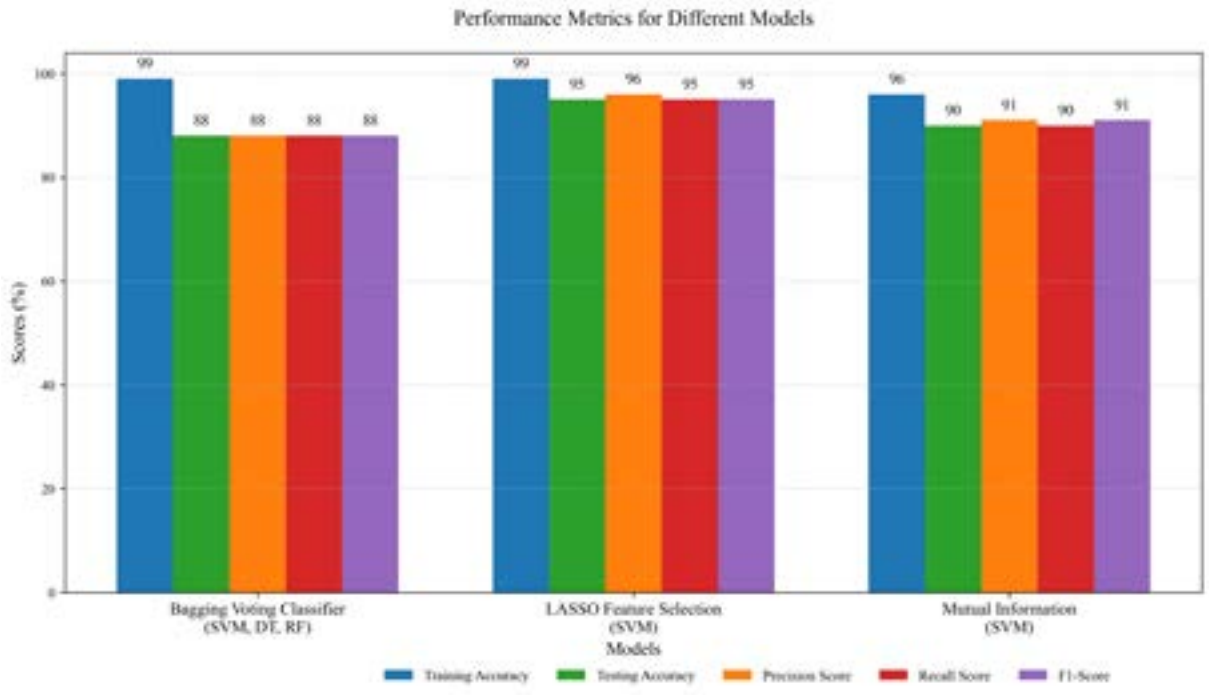


Fig. 11. Performance metrics for various models: The Figure presents the accuracy metrics for the bagging ensemble and feature selection models.

incorporated within the software (MiCROTEC – CT Log 2024), representing another significant issue of this advanced technology. Additionally, independent research beyond software development for classification/detection has centered solely on detecting single defects rather than a comprehensive approach. Studies such as Boukadida et al. 2012, and Gazo et al. 2020, present robust algorithms for identifying internal wood features like pith. Furthermore, Breinig et al. 2015, and Longuetaud et al. 2012, emphasise the effectiveness of CT scanning in detecting knots and grain patterns. However, researchers did not concentrate on identifying normal wood and various wood defects simultaneously.

Our research focused on using a distinct machine learning phase instead of relying on preset software, allowing us to experiment with various algorithms. To meet this goal, we first converted the scanned dataset into multiple jpg images from the original TIFF file. Subsequently, we organised our classification into two phases. In the initial phase, we utilised only two algorithms, DT and RF, and had just two categories: normal and defective. The benefit of this method is the ease and speed of constructing a training database, where each scan is marked with a 0 (denoting an image without a defect) or a 1 (indicating an image with a defect). Another benefit is the ability to visually evaluate defects using free image processing software. This method is applicable for wood grading when determining the presence or absence of a defect is adequate. It can also be used to swiftly discard scans without defects, allowing continued focus on those with defects.

In the initial stage of our study on distinguishing between defective and normal logs, we attained 90% accuracy using the DT classifier and 93% with the RF classifiers. From the existing literature, we discovered a Ph.D. dissertation by Ursella (2021) that employed neural networks for the classification of dead and sound knots. Their research demonstrated an overall accuracy of 86.5%. Specifically, for sound knot classification, they accurately identified 88.5% but mistakenly categorised 16.4% as dead knots. Regarding dead knots, 11.4% were incorrectly identified as sound knots, whereas 83.5% were accurately identified. When comparing these findings with our study, we observe that for the recall of the positive class (Defective/Dead), our DT model attained an accuracy of 72%, while our RF model achieved 74.2% accuracy. Conversely, their CNN model reached 83.5% accuracy. In terms of the recall of the negative class (Sound/Normal), our DT and RF models recorded accuracies of 94.5% and 97.7%, respectively, compared to 88.5% accuracy by the CNN model. In summary, our models exhibit superior accuracy, especially concerning overall accuracy.

In the second phase of the research, six machine learning algorithms were tested. When analysing the results, we find that the algorithms provide no accuracy greater than 85%. We are expecting the main reason that the classical machine learning algorithms cannot obtain more relevant feature sets without using appropriate feature extractors. However, in the case of deep neural networks, algorithms can identify suitable set of features without input/guiding. In addition, two algorithms, SVM and AdaBoost, were found to have no average accuracy. In SVM, the classifier cannot define the most suitable hyperplane. Since the dataset is in greyscale format, we did not use segmentation or thresholding operations, which could be the main reason for this low accuracy. In the case of AdaBoost, the classifier did not have enough features to perform proper training processes. It is important to note that we used GLCM for the feature extraction process. This may be the primary reason for the low accuracy of these classifiers. To verify the hypothesis, we need to implement deep neural networks and various feature extractors for this classification task.

Consequently, we used CNN to extract features and developed three new models in the subsequent stages (Bagging Voting Classifier; LASSO, Mutual Information feature selection) for the classification task. We achieved the highest test accuracy of 95% with the LASSO-SVM model. The study demonstrates that machine learning algorithms can achieve standard accuracy in classifying different defect classes, even with the default parameter settings on 3D CT scanner data. This suggests that 3D CT scanner datasets are more effective for quality assessment tasks than other datasets. Additionally, we obtained higher accuracy by using CNN as a feature extractor rather than GLCM. Furthermore, there is a significant potential for surpassing 95% accuracy with various deep learning algorithms, warranting further investigation.

The performance differences across models can be attributed mainly to the feature representation rather than to the classifiers alone. With hand-crafted GLCM descriptors, the greyscale CT cross-sections offered only a limited set of textural cues, and the SVM in particular could not define an effective separating hyperplane in this low-dimensional, weakly discriminative feature space, which explains its poor multiclass performance. Replacing GLCM with CNN-derived features supplied a richer, hierarchical representation of edges, ring structure and local density patterns, while the subsequent LASSO and Mutual Information feature selection retained the most informative and least redundant of these features. This combination allowed the same SVM classifier to surpass the RF and DT baselines, indicating that the improvement arose from better feature extraction and selection rather than from the choice of classifier. Defect type also influenced accuracy: rot and split, which produce large, spatially contiguous density changes, were classified almost perfectly, whereas dead and live knots – visually similar in greyscale cross-sections were the main source of confusion, consistent with the misclassifications seen in the confusion matrix (Fig. 9). Relative to the dead/sound knot classification of Ursella (2021), which reported 86.5% overall accuracy, our CNN-with-feature-

selection pipeline reached up to 95% testing accuracy across a broader set of defect classes, underlining the contribution of combining real industrial CT data with a reproducible, software-independent feature-selection workflow.

Several limitations of this study should be acknowledged. First, the dataset was derived from a relatively small sample of 26 European Oak logs from a single site and scanner, which constrains the diversity of defect appearances and limits the generalisability of the trained models to other species, growth conditions and acquisition settings. Second, the multiclass dataset is modest in size and somewhat imbalanced across the five categories, and it does not cover the full range of internal features encountered in practice (for example resin pockets, ring shakes or fungal staining), so the models are validated only on the defect types represented here. Third, as noted in the Materials and methods, the 80/20 partition was performed at the image level; because many near-duplicate cross-sections are extracted from each log, this carries a risk of optimistic bias if correlated slices from the same log appear in both the training and testing sets, and the reported accuracies should be read with this caveat in mind. Finally, the results rest on classification accuracy against visually assigned labels and have not yet been verified against independent ground truth such as subsequent sawing or sawmill grading outcomes, so the practical value of the predictions for operational sorting remains to be confirmed.

Additionally, the primary limitation of this study was that we did not concentrate on identifying multiple defect classes from a single image; rather, we concentrated on individual defect classes. As a future goal, we aim to implement an automatic model that recognises all defects from a single log image and displays the defect boundaries along with the corresponding defective class labels.

Future work should address these limitations along several lines. To obtain more robust and reproducible performance estimates, the evaluation should adopt log-level (grouped) cross-validation, in which all cross-sections from a given log are kept within a single fold, thereby removing any leakage of correlated slices between training and testing. The classification pipeline should also be benchmarked against deeper architectures, such as ResNet and other modern convolutional networks, and extended from two-dimensional cross-sections towards volumetric (3D CNN) models that exploit the continuity of defects along the log axis. Moving from classification to detection, instance-segmentation approaches such as Mask R-CNN would allow several defects to be localised and labelled within a single image, together with their boundaries, which is the natural next step beyond the single-defect-per-image setting used here. Finally, assembling larger and more balanced multi-species datasets, and validating the predictions against subsequent sawing or sawmill grading outcomes, would strengthen both the accuracy and the practical relevance of the method for integration into operational log-sorting lines.

5. Conclusion

This study thoroughly investigated log quality classification by utilising a variety of well-established machine learning algorithms. The main aim of this research is to emphasise the potential uses of 3D CT scanning technology within the wood industry. For this analysis, two different datasets were used: one with a simpler class structure (referred to as the first dataset) and another with a multiclass framework (known as the second dataset).

In this study, we leveraged 3D CT scanning technology for classifying the quality of wood logs. This groundbreaking approach speeds up the classification process and aids in saving time and reducing waste, leading to a more efficient use of wood resources. Typically, machine learning tasks like this require high-powered computing resources and the know-how of specialised professionals. Nonetheless, we conducted this research using standard systems. Additionally, we improved classical machine learning algorithms by implementing superior feature extractors and integrating bagging voting classification with feature selection techniques. This opens a new research path in utilising advanced technology with minimal computational resources. It should be emphasised that, operationally, the proposed method performs internal defect mapping and defect-class prediction on CT cross-sections rather than a complete quality assessment: full quality grading, for instance for barrel or musical-instrument production, additionally depends on the location, size and continuity of defects relative to the intended product. The classifications produced here are therefore best understood as an automated input to such grading rather than a stand-alone quality verdict. In practical terms, integrating this software-independent and reproducible pipeline into industrial CT-based sorting lines could support faster, more objective and more consistent grading of hardwood logs, reduce subjective downgrading and processing waste, and make low-cost, standard-hardware defect screening more accessible to forestry enterprises and sawmills.

Acknowledgements

This study was funded by the European Commission under the LignoSilva project [Grant Agreement #101059552] through the Horizon Europe – Teaming for Excellence program and Slovak Research and Development Agency under contract No. APVV-20-0118 Ligno3Dscan.

References

Alapuranen, P., Westman, T., 1992: Automatic visual inspection of wood surfaces. Proceedings. 11th IAPR International Conference on Pattern Recognition, Netherlands, The Hague, p. 371–374.

Amishev, D., Murphy, G. E., 2008: Implementing Resonance-Based Acoustic Technology on Mechanical Harvesters/Processors for Real-Time Wood Stiffness Assessment: Opportunities and Considerations. International Journal of Forest Engineering, 19:48–56.

Bhuyan, H. K., Saikiran, M., Tripathy, M., Ravi, V., 2023: Wide-ranging approach-based feature selection for classification. Multimedia Tools and Applications, 82:23277–23304.

Boukadida, H., Longuetaud, F., Colin, F., Freyburger, C., Constant, T., Leban, J. M. et al., 2012: PithExtract: A robust algorithm for pith detection in computer tomography images of wood – Application to 125 logs from 17 tree species. Computers and Electronics in Agriculture, 85, 90–98.

Breinig, L., Broman, O., Brüchert, F., Becker, G., 2015: Optimization potential for perception-oriented appearance classification by simulated sawing of computed tomography-scanned logs of Norway spruce. Wood Material Science and Engineering, 10:319–334.

Dickson, R., Joe, B., Johnstone, D., Austin, S., Ribton-Turner, F., 2005: Pre-processing prediction of wood quality in peeler logs grown in northern New South Wales. Australian Forestry, 68:186–191.

Gazo, R., Vanek, J., Abdul_Massih, M., Benes, B., 2020: A fast pith detection for computed tomography scanned hardwood logs. Computers and Electronics in Agriculture, 170:105107.

Gejdoš, M., Gergeľ, T., 2022: Case study of qualitative sorting of raw wood assortments in the conditions of a forestry enterprise in Slovakia. Central European Forestry Journal, 68:232–237.

Gejdoš, M., Michajlová, K., Gretsch, D., 2023: The use of the acoustic tomograph and digital image analysis in the qualitative assessment of harvested timber – case study. Central European Forestry Journal, 69:106–111.

Haralick, R. M., 1973: Textural Features for Image Classification. Studies in Media and Communication SMC-3:610-621. Available at https://www.researchgate.net/publication/302341151.

Jori Uusitalo Olli Ylhäisi, H. R., Makkonen, M., 2018: Predicting probability of A-quality lumber of Scots pine (*Pinus sylvestris* L.) prior to or concurrently with logging operation. Scandinavian Journal of Forest Research, 33:475–483.

Khazem, S., Richard, A., Fix, J., Pradalier, C., 2023: Deep learning for the detection of semantic features in tree X-ray CT scans. Artificial Intelligence in Agriculture, 7:13–26.

Kotsiantis, S. B., Pintelas, P. E., 2004: Bagged Voting Ensembles. In: Bussler, Ch., Fensel, D. (eds.): Artificial Intelligence: Methodology, Systems, and Applications. Berlin Heidelberg, Springer, p. 168–177.

Krähenbühl, A., Kerautret, B., Debled-Rennesson, I., Longuetaud, F., Mothe, F., 2012: Knot Detection in X-Ray CT Images of Wood. ISVC – 8th International Symposium on Visual Computing, 2012, Rethymnon, Greece, p. 209–218.

Lee, S. M., Abbott, A. L., Schmoldt, D. L., 2003: Surface shape analysis of rough lumber for wane detection. Computers and Electronics in Agriculture, 41:121–137.

Leon, F., Floria, S.-A., Bădică, C., 2017: Evaluating the effect of voting methods on ensemble-based classification. 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, p. 1–6.

Longuetaud, F., Mothe, F., Kerautret, B., Krähenbühl, A., Hory, L., Leban, J. M. et al., 2012: Automatic knot detection and measurements from X-ray CT images of wood: A review and validation of an improved algorithm on softwood samples. Computers and Electronics in Agriculture, 85:77–89.

Malan, F. S., 2003: The wood quality of the South African timber resource for high-value solid wood products and its role in sustainable forestry. Southern African Forestry Journal, 198:53–62.

Marszalec, E. A., Pietikaeinen, M., 1993: Color analysis of defects for automated visual inspection of pine wood. In: Wu, F. Y., Dawson, B. M. (eds.): Machine Vision Applications in Industrial Inspection, Proceedings Volume 1907, p. 80–94.

Mohammadali Saadatnia Aliakbar Enayati, K. P., Moradian, M., 2016: Investigation on transversal variation of poplar tension wood quality using ultrasound wave parameters. Wood Material Science & Engineering, 11:201–208.

Nguyen, H. B., Xue, B., Andreae, P., 2016: Mutual information for feature selection: estimation or counting? Evolutionary Intelligence, 9:95–110.

Olaoye, K. O., Ojo, M. O., 2022: Non-destructive acoustic assessment of wood quality in trees and logs and the effects of silvicultural treatments: a review. International Wood Products Journal, 13:156–171.

Powers, D. M., 2011: Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technology, 2, 37–63.

Pyörälä, J., Kankare, V., Liang, X., Saarinen, N., Rikala, J., Kivinen, V.-P. et al., 2019: Assessing log geometry and wood quality in standing timber using terrestrial laser-scanning point clouds. Forestry: An International Journal of Forest Research, 92:177–

187.
Ridley-Ellis, D., Stapel, P., Baño, V., 2016: Strength grading of sawn timber in Europe: an explanation for engineers and researchers. European Journal of Wood and Wood Products, 74:291–306.
Skurichinars, M., Duin, R. P. W., 1998: Bagging for linear classifiers Linear discriminant Generalization error Small sample size Regularization Bagging Instability Bias and variance. Pattern Recognition, 31:909–930.
Sobey, P. J., 1991: Automated optical grading of timber. In: DeShazer, J. A., Meyer, G. E. (eds.): Optics in Agriculture, 1379:168–179.
Tibshirani, R., 1996: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B: Methodological, 58:267–288.
Ursella, E., 2021: In-line industrial computed tomography applications and developments. Ph.D. Thesis, Universita Ca´Foscari Venezia, Venice, Italy, 101p.
Vacek, O., Gergeľ, T., Bucha, T., Gracovský, R., Gejdoš, M., 2024: Automatic Wood Species Classification and Pith Detection in Log CT Images. Forests, 15:2207.
Wang, X., Ross, R., Carter, P., 2007: Acoustic evaluation of wood quality in standing trees. Part I. Acoustic wave behavior. Wood and fiber science, 39:28–38.
Wengert, E. M., Meyer, D. A., 1994: Guidelines for Grading Hardwood Logs. Forestry Facts No. 74, University of Wisconsin-Madison Extension, Department of Forestry, Madison, 6 p.
Wimmer, G., Schraml, R., Hofbauer, H., Petutschnigg, A., Uhl, A., 2021: Two-Stage CNN-Based Wood Log Recognition. In: Gervasi, O., Murgante, B., Misra, S., Garau, Ch., Blečić, I., Taniar, D. et al. (eds.): Computational Science and Its Applications – ICCSA 2021. International Conference on Computational Science and Its Applications, Cagliari, Italy, 13 September 2021. Springer International Publishing, p. 115–125.
Yu, C., Sun, Y., Cao, Y., He, J., Fu, Y., Zhou, X., 2023: A Novel Wood Log Measurement Combined Mask R-CNN and Stereo Vision Camera. Forests, 14:285.

Other sources

3D Computed Tomography – see through your parts non-destructively, 2024. Available at https://www.rx-solutions.com/en (February 15, 2024).
Home – Comet Yxlon, 2024. Available at https://yxlon.comet.tech/ (February 15, 2024).
Industrial CT and X-Ray Solutions from ZEISS, 2024. Available at https://www.zeiss.com/metrology/products/systems/computed-tomography.html (February 15, 2024).
Industries, 2024. Available at https://www.lumafield.com/industries (February 15, 2024).
MiCROTEC – CT Log, 2024. Available at https://www.microtec.eu/en/products/ct-log (February 15, 2024).
Wollin, A. C., Vaughan, C. L., 1947: Guidelines for the scaling and grading of hardwood logs. Available at www.appalachianhardwood.org.