

Article

Evaluation of a Computer Vision Algorithm for Wood Defect Detection and Quality-Class Assignment of Birch (*Betula* spp.) Veneer Logs Under Industrial Conditions

Jānis Magaznieks *, Mareks Millers and Agris Gabrāns 

VMF LATVIA SIA, Skaistkalnes iela 1, LV-1004 Riga, Latvia; millers@vmf.lv (M.M.); gabrans@vmf.lv (A.G.)

* Correspondence: magaznieks@vmf.lv

Abstract

The quality grading of birch (*Betula* spp.) veneer logs, an economically important raw material in the Baltic region, is still largely manual, causing pricing inconsistencies, and automated grading under industrial conditions remains rarely studied. We evaluated deep-learning computer vision models for recognising wood defects and assigning a quality class on a sorting line. The algorithm and a scaler evaluated 153 logs using the same images. The recognition differed considerably between the defects: the central core-zone defects were recognised well (recall: core discolouration 97%, pith and drying cracks 91%, core rot 88%), while the dead and rotten knots and the less conspicuous defects were recognised poorly (knots 44%, ingrown bark 52%). The quality class assigned by the algorithm coincided with that of the scaler in 57% of the cases, and the algorithm tended to assign a lower class, mainly because of the over-indication of the core rot. For reference, manual grading itself was not fully repeatable: line scalers agreed with a detailed control scaler in 79% of the cases (69 to 88% between scalers). Automated grading is already accurate for core discolouration and pith and drying cracks, while knot detection remains the main limitation for industrial deployment.

Keywords: birch; roundwood; wood defects; quality grading; automated grading

1. Introduction

Birch (*Betula* spp.) is one of the most important commercial tree species in the Baltic region, and in recent years its economic role is increasing, as in Latvia more plywood and veneer producing factories have been opened and the demand for birch veneer logs has been growing. In 2025 Latvia exported veneer and plywood for approximately EUR 423 million, which is one of the most important wood product export categories for the country [1]. Latvia, together with Finland and the other Baltic states, is among the main European producers of birch plywood, so the quality of the birch veneer logs that supply this industry has an economic weight beyond the national market. The main raw material for this production is birch veneer logs, and their quality, to a large extent, determines the value of the final product.

The price of each veneer log is determined on the basis of its volume and its quality class. At the roundwood sorting line the quality class is assigned according to the wood defects that are visible on the log side surface and on the log end faces, such as knots, pith and drying cracks, ingrown bark, dead-bark wounds, core discolouration and rot. In Latvia the quality requirements for birch veneer logs are defined in the standard LVS



Academic Editor: René
Alexander Herrera Díaz

Received: 22 June 2026

Revised: 27 July 2026

Accepted: 28 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

80-7:2024 [2]. The type, size and location of the defect with respect to the peeling cylinder (the cylindrical core that is rotary-peeled into the veneer) determine the quality class of the log, and the quality class in turn determines the price that is paid between the timber buyer and the seller.

Despite this importance, the sorting of round timber in many mills is still carried out manually [3]. This work is physically demanding, the decision has to be made quickly for each log, and the result depends on the experience of the scaler. In the related task of manual lumber grading, the accuracy of the human graders has been reported to be only 48 to 75% and to decrease with the production speed and fatigue [4]. Therefore, the manual evaluation has been found insufficiently consistent for a high throughput production [5,6]. Automated systems have a clear advantage in such conditions, because they keep a constant performance over time and can process the information faster and without interruption [7]. These limitations have promoted the development of computer vision systems as a practical alternative to the manual grading, which have been shown to exceed the manual inspection both in accuracy and consistency [6].

In the wood quality assessment, the use of deep learning has increased considerably in the last years, growing from practically zero before 2016 to nearly a half of the publications in the period from 2023 to 2026, while classical computer vision methods still make up the majority [8]. Convolutional neural networks (CNNs) have become especially common for the inspection of wood defects, because they learn the relevant visual features directly from the image data. Both the single-stage detectors such as You Only Look Once (YOLO) and the two-stage models such as Faster Region-based CNNs are used for industrial defect detection [5,9]. Such systems are already used in industrial inspection and grading of wood, for example, in the real-time surface inspection of wood-based panels [7] and in the validated automated grading of sawn hardwood lumber [4,10]. Image-based systems for the sorting of roundwood have likewise been developed [3]. Where the size and the extent of a defect are needed and not only its location, segmentation networks such as U-Net assign each pixel of the image to a defect or to the background, which is the information that grading according to the standard requires. More recent studies have combined red–green–blue (RGB) camera data with the three-dimensional point cloud of the log surface, and the fusion of the two has been shown to give a better knot detection than either of them separately [11]. For the log scaling and grading tasks, the deep learning models have likewise performed better than the classical algorithms where a direct comparison was possible, although comparison between the studies is difficult [12].

At the same time the application of these models in industrial conditions has several difficulties. The defects are rare in comparison with the total log surface, which creates strongly imbalanced training data sets, in which the defect covers only a small part of each image, and the catalogue of the recognised defects is often revised once the automatic inspection begins to operate [7]. A more essential difficulty for the defect recognition on round timber with bark is the high natural variation of the bark surface, because of which a defect can be confused with the features of the bark or a real defect can stay unnoticed. This makes the image based detection of the knots difficult even for the human observers [11]. These difficulties concern in the first place the knots and the other defects on the bark-covered side surface, whereas the defects of the core, such as the discolouration and the rot, appear on the freshly cut end face and have a different visual character. The difficulty with the bark is especially expressed for birch (*Betula* spp.), the bark of which has a particular colour and texture that complicates the automated recognition of the bark itself [13], and therefore, very likely, also of the defects. Birch has received comparatively little attention in the automated grading literature, which is mostly devoted to the coniferous species, and the few studies on birch concern the sawn or processed timber rather than the logs with bark

(for example [14]). The studies that evaluate defect recognition on the bark-covered round timber under industrial conditions, in particular for birch, are still scarce. To our knowledge, no previous study has evaluated an optical computer-vision system for defect detection and quality-class assignment of bark-covered birch veneer logs under real industrial conditions, which is the gap that the present study addresses.

This study evaluates the defect-recognition and grading component of a larger automated system that also determines the tree species and the remaining bark amount [13]. The aim of this study was to evaluate the performance of the prepared deep-learning-based computer vision models in the recognition of the wood defects and in the assignment of the quality class of the birch veneer logs under real industrial conditions, comparing the result of the algorithm with the evaluation made by a scaler using the same visual data. In particular, the study examines: (i) how reliably the different wood defects that determine the quality class are recognised on the birch veneer logs, and which of them limit the automated grading; and (ii) how accurately the quality class is assigned in comparison with the scaler. For this purpose, the algorithm was evaluated on the birch veneer logs of an operating sorting line, and the results give a baseline of the present possibilities of the automated quality evaluation of the birch veneer logs and show the directions in which further development is needed.

2. Materials and Methods

2.1. Materials and the Study Site

The study was carried out on birch (*Betula* spp.) veneer logs at an operating roundwood sorting line of a commercial veneer mill in Latvia. The visual data of the round timber were obtained with line-scan cameras for the curved side surface of the log and area-scan cameras for the two log end faces, while the logs were moving on the sorting line, without stopping the conveyor. The cameras were placed in a closed room with the LED lighting, so that the lighting conditions did not depend on the time of the day. The visual data were collected during all the seasons of the year, in order to cover the different conditions of the birch veneer logs. The birch veneer logs measured on the sorting line ranged in top diameter from about 13 to 48 cm, in butt diameter from about 15 to 57 cm and in length from about 2.5 to 3.3 m, based on the control-measured logs. The 153 evaluation logs were a defect-containing subset of this material and were not separately measured for size. The logs were freshly cut, green roundwood.

The quality class of the birch veneer logs was assigned according to the Latvian standard LVS 80-7:2024 [2], which divides the birch veneer logs into the first, the second and the third quality class and the non-conforming material (i.e., rejects). The recognition of the tree species and the evaluation of the remaining bark amount are the preconditions for the automated quality assessment. The evaluation of the remaining bark amount was prepared and described in the previous study [13] and is therefore not repeated here.

2.2. The Observers and the Design of the Comparison

Four evaluations of the birch veneer logs were used in this study:

- The control scaler: A scaler who evaluates each log physically and in detail and sees the whole (100%) surface of the log. The control scaler works in the log yard, away from the sorting line, and is not limited by the sorting line speed, so about two to three minutes can be spent on each log, which is an approximate operational figure and not a measured time. The evaluation of the control scaler is the most detailed human assessment.
- The line scaler: The scaler who operates the sorting line in the production environment and registers the wood defect that determines the quality class. The whole side surface

is exposed as the log is turned along the line, so the line scaler can in principle see all of it, but the decision has to be made within a few seconds under the production speed;

- The remote scaler: An experienced control scaler (one of the authors) who evaluated the same visual data (the image sets) as the computer vision algorithm and therefore had the same information as the algorithm (about 75% of the side surface and the log end faces);
- The computer vision algorithm: The prepared mathematical models that evaluated the same visual data as the remote scaler.

In this study the reference for evaluating the algorithm is the remote scaler, because the remote scaler assessed exactly the same images as the algorithm, so that any difference reflects the algorithm and not a difference in the visible information. The control scaler provides the most detailed human assessment and, together with the line scaler, is used only to characterise the consistency of the manual grading. Two comparisons were made. The agreement between the control scaler and the line scaler was used to characterise the consistency of the manual quality class assignment under the production conditions. The evaluation of the wood defect recognition and of the quality class assignment by the algorithm was made against the remote scaler, who evaluated the same images as the algorithm. In this way the comparison is not influenced by the different amounts of the visible surface. Before the evaluation the remote scaler had calibrated his image-based reading on other birch veneer logs, which he had inspected physically at the site and had also seen on the images. His reading of the image sets is therefore grounded in the direct physical experience of such logs, while the 153 evaluation logs themselves were assessed only on the images, as by the algorithm. No single set of logs passed through all four evaluations together, because the algorithm was not yet ready to be run when the control assortments were collected. The comparison therefore rests on two paired evaluations, each anchored on a control–scaler standard: the control scaler for the manual-consistency comparison, and the remote scaler, an experienced control scaler and one of the authors, for the algorithm. In total the manual assessment involved one control scaler, seven line scalers and one remote scaler, and the sample sizes of the two comparisons are given in Section 2.5.

2.3. The Wood Defects and Their Evaluation

The wood defects that determine the quality class of the birch veneer logs and their evaluation are defined in the standard LVS 80-7:2024 [2]. The defects are evaluated either on the log end face (the freshly sawn cross-section) or on the curved side surface of the log. The evaluation of each defect with respect to the peeling cylinder—the cylindrical part of the log that is rotary-peeled into the veneer—is essential (Figure 1), because a defect that does not reach the peeling cylinder has a smaller influence on the quality class. The evaluated wood defects, the surface on which they are graded and the characteristics that the standard uses for the grading are summarised in Table 1. These characteristics are the basis of the grading in the standard, and in the present study the recognition of the wood defects was evaluated as the presence or the absence of each defect (Section 2.6).

These seven wood defects were chosen for the evaluation because they are the ones that most often determine the quality class of the birch veneer logs. In the operational records of the control scaler (Section 2.5) they make up about 87% of all the registered wood defects and the majority of the cases in which the quality class is lowered. The other defects of the standard were not included for different reasons. Some of them are an extrinsic damage and not a property of the wood (the mechanical and the harvester damage of the side surface and the end face). Some are a purely dimensional or stem-form characteristic that is measured and not recognised as a defect (for example the curvature, the butt swell

and the dimensions of the log). The sound knots are recognised but do not lower the quality class, and the remaining intrinsic defects occur only rarely, each in less than about 3% of the logs (for example mineral inclusions and frost or storage defects). The dimensional reduction of the log was outside the scope of the present study.

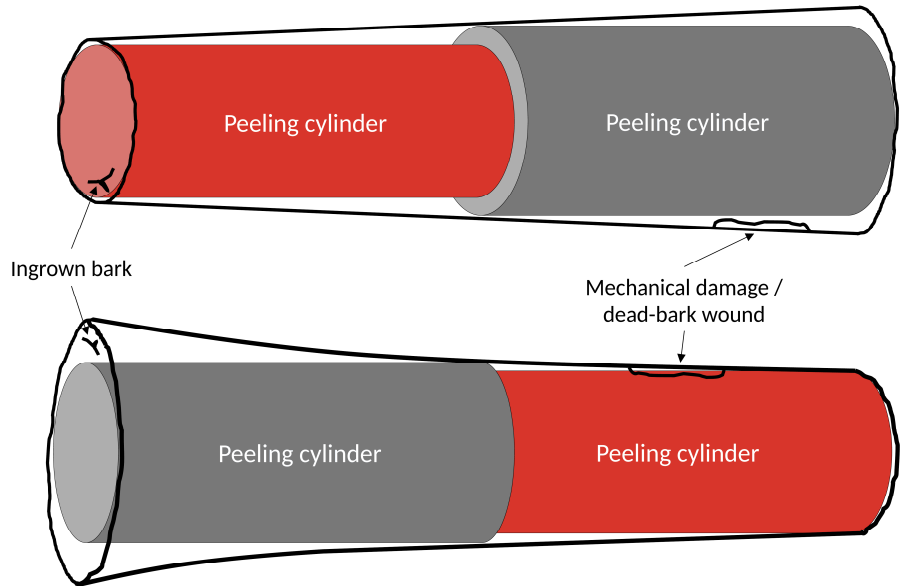


Figure 1. Schematic of a birch veneer log and its peeling cylinder. Each veneer log is processed as two consecutive peeling sections (shortlogs), and the cylindrical core of each section is rotary-peeled into the veneer. This peeling cylinder is shown in grey where it is intact and in red where a defect reduces its usable diameter. A defect affects the quality class mainly when it reaches the peeling cylinder: an ingrown bark on the end face, or a dead-bark wound or mechanical damage on the side surface, reduces the usable part and lowers the quality class, whereas a defect that stays outside it has a smaller influence.

Table 1. The evaluated wood defects of the birch veneer logs, the surface on which they are graded according to LVS 80-7:2024, and the characteristics for the grading.

Wood Defect	Graded On	Characteristics for Grading
Dead and rotten knots	Side surface	Type (healthy / dead / rotten); knot diameter, measured in the narrowest direction without the bark
Pith and drying cracks	End face	Width and length; whether the crack splits the side surface (cracks wider than 2 mm)
Felling and bucking cracks	End face	Visible on the butt end face, continuing in the longitudinal direction
Ingrown bark	End face	Largest thickness; circumferential angle; diameter; effect on the peeling cylinder
Dead-bark wound (In Latvian: saussāns)	Side surface/end face ¹	Effect on the peeling cylinder
Core discolouration	End face	Occurrence; diameter
Core (forest) rot	End face	Occurrence; diameter

¹ The dead-bark wound lies on the side surface but is graded by its effect on the peeling cylinder, assessed where the wound reaches the end face.

In addition to these per-defect characteristics, the quality class of a log depends on how large each defect is in relation to the peeling cylinder. Table 2 gives the indicative size allowances of the first, the second and the third quality class for the knot and core defects that are graded chiefly by their size, summarized from LVS 80-7:2024 [2].

Table 2. Indicative quality-class size allowances for the knot and core defects of the birch veneer logs that are graded chiefly by their size (summarized from LVS 80-7:2024 [2]).

Wood Defect	Class I	Class II	Class III
Dead and rotten knots	Not permitted	≤ 40 mm	≤ 80 mm
Core (forest) rot	Not permitted	Not permitted	≤ 50 mm
Core discolouration	≤ 70 mm	Permitted	Permitted

The values are a simplified summary of the main size limits and do not reproduce the full standard. Logs with defects exceeding the third-class limits are graded as non-conforming (reject).

2.4. Image Acquisition and the Computer Vision Algorithm

The visual data were obtained with two camera systems (Figure 2). The curved side surface of the log was recorded with three colour line-scan cameras (about 8000 pixels per line, downsampled to about 2040 pixels for the analysis), placed to the left, to the right and above the log. Together they cover about 75% of the side surface, while the underside, which rests in the conveyor trough, is not imaged. The line-by-line recording during the movement of the log on the conveyor was triggered by an incremental encoder. The two log end faces were recorded with two colour area-scan cameras with a global-shutter sensor (2464×2056 pixels). All the cameras were fitted with 50 mm lenses. The side-surface images are built line by line during the movement of the log into one long, narrow colour strip, while the end-face images are single frames. In the analysed data set each such strip was about $2040 \times 12,500$ pixels and the end-face images were 2464×2056 pixels. The visual data were processed on a workstation with an Intel Core i9-13900K processor, 32 GB of memory and an NVIDIA GeForce RTX 3060 (12 GB) graphics processor.



Figure 2. Examples of the visual data of a birch veneer log. (a) A part of the side-surface image obtained with a line-scan camera. (b) An end-face image obtained with an area-scan camera.

The visual data were processed automatically by the prepared computer vision models, which were developed by the development team. All the models are CNNs: the segmentation models use the U-Net architecture [15] and the species classification model uses the EfficientNet architecture [16], and the models were trained and run with the PyTorch framework. Separate models are used for the area-scan (end-face) images and for the line-scan (side-surface) images. The recognition of the wood defects in the present study is based on two of these models—a segmentation model for the end-face defects and a segmentation model for the side-surface knots—while the recognition of the tree species and the evaluation of the remaining bark amount are performed by the other models of the system and were not evaluated in the present study. The evaluation of the remaining bark amount has been described in a previous study [13]. The quality class is not predicted directly by the models: the wood defects and their characteristics that are recognised by

the models are passed to a separate rule-based module, which assigns the quality class according to the requirements of the standard LVS 80-7:2024.

For the preparation of the models the annotated images were divided, for each model, into the training, the validation and the test set in the proportion 80/15/5. The end-face defect model and the knot model were trained on 1033 and 417 annotated images, respectively (the line-scan images are processed in tiles, so the effective number of the training images of the knot model is larger). The annotation was carried out by one person throughout the whole project, in a dedicated in-house web application developed by the development team, which also serves as the interface for viewing the detections of the algorithm. Each wood defect was annotated as a polygon (a segmentation mask) according to the definitions of the standard LVS 80-7:2024, and the annotations were afterwards reviewed and corrected by the same person after discussions with the development team. The 153 birch veneer logs that were used for the evaluation (Section 2.5) were not part of the training data and were therefore independent from it.

The images were stored in PNG format. The models were trained at a reduced resolution, for example 320×320 pixels for the end-face defect model and 1024×1024 pixel tiles for the knot model, whereas the production images were 2464×2056 pixels for the area-scan cameras and about 2040 pixels wide for the line-scan cameras. Each segmentation model applies a confidence threshold and a minimum-size filter that removes the detected objects with a very small area. The development team evaluated the models internally with the Intersection-over-Union (IoU) and the Dice coefficient for the segmentation and with the F1 value for the classification. These internal metrics are separate from the log-level comparison with the remote scaler that is reported in this study.

2.5. The Data Sets Used for the Evaluation

- The full population. The visual data were collected for 1,673,568 birch veneer logs on the operating sorting line.
- The operational validation. A total of 2789 control assortments were inspected by the control scaler at the processing site. Of these, 2674 could be compared with the corresponding evaluation of the line scaler (for 115 assortments the sorting line data was not available). At the sorting line each log is graded not as a whole but as two consecutive peeling sections (i.e., shortlogs, *iskluči*), each of which receives its own quality class and price. The agreement of the quality class between the line scaler and the control scaler was therefore evaluated per shortlog on 5341 shortlogs, and it was 79% overall, varying from 69 to 88% between the seven individual line scalers, each of whom graded between about 200 and 1400 shortlogs.
- The wood defect and the quality class evaluation. A total of 153 birch veneer logs were evaluated both by the algorithm and by the remote scaler using the same images. This data set was intentionally selected so that it contains wood defects, in order to allow a meaningful defect-level analysis rather than a population-level estimation, and it includes more reject material (18% of the evaluated logs) than the production flow, where about 5% of the shortlogs are graded as reject. Because the 153 logs were intentionally enriched with defects, the precision and the overall class agreement depend on how often the defects occur, so these values describe this defect-rich subset and would be different on the production flow. The recall does not depend on how often the defects occur, so it can be transferred more directly to the production flow. These 153 logs were not among the images used to train the models.

2.6. Evaluation of the Results and Statistical Methods

For each wood defect, the presence or absence indicated by the algorithm was compared with the evaluation of the remote scaler, and a confusion matrix was built from the true positives (TPs), false positives (FPs) and false negatives (FNs). The recognition of each wood defect was characterised by the recall, the precision and the F1 value, defined in Equations (1)–(3): the recall is the proportion of the defects evaluated by the remote scaler that were recognised by the algorithm, the precision is the proportion of the defects indicated by the algorithm that were confirmed by the remote scaler, and the F1 value is the harmonic mean of the precision and the recall. The comparison was made at the level of the presence or absence of each wood defect and of the resulting quality class. The accuracy of the individual measured values defined in the standard (for example the knot diameter or the crack width) was not evaluated. The reference evaluation was a visual grading of the images and recorded only the presence of each defect and the assigned quality class, not the individual measured values. There was therefore no measurement reference against which the values produced by the algorithm could be compared.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \quad (2)$$

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}), \quad (3)$$

The assignment of the quality class was characterised by the proportion of the logs for which the algorithm and the remote scaler assigned the same class, by the four-class confusion matrix, and by the direction of the disagreement (a lower or a higher class). In addition to the proportion of the matching logs, the agreement was summarized with Cohen's kappa [17]. To show the uncertainty that follows from the limited number of the logs, a 95% Wilson score confidence interval (CI) [18] was calculated for each recall and precision and for the overall class agreement. Whether the disagreements between the algorithm and the remote scaler were directed systematically towards a lower or a higher class, rather than being symmetric, was tested on the paired quality-class table with the Bowker test of symmetry [19], and the difference between the two class distributions was tested with the Stuart–Maxwell test of marginal homogeneity [20,21]. Because the 153 logs were an intentionally defect-enriched set and not a random sample of the production flow (Section 2.5), these indicators characterise the agreement on this particular set of the logs rather than estimating the parameters of the whole population, and for the rare wood defects the number of the logs is small, so their precision and F1 value should be interpreted with caution. The evaluation of the wood defects on the images by the remote scaler was carried out in the same dedicated web application, and the data handling, the metric calculations and the figures were prepared in Microsoft Excel and in Python 3.11. During the preparation of this study, the authors used Claude (Claude Opus 4.8, Anthropic, San Francisco, CA, USA) to assist with writing and checking the code used for the data handling, the calculations and the figures. The authors designed the analysis, verified all the results, and take full responsibility for them.

3. Results

3.1. Recognition of the Wood Defects

For the evaluation of the wood defect recognition, 153 birch veneer logs were used, for which the visual data were assessed both by the prepared computer vision algorithm and by a scaler who evaluated the same image sets (i.e., the remote scaler). The evaluation of the remote scaler was used as the reference. Among these logs the most frequently found

wood defects were core discolouration (in 63% of the logs), pith and drying cracks (58%) and dead and rotten knots (34%), while ingrown bark (14%), dead-bark wounds (7%), core rot (5%) and felling and bucking cracks (5%) were found less often.

For each wood defect, the recognition was characterised by the recall (the proportion of the defects evaluated by the remote scaler that were also recognised by the algorithm), the precision (the proportion of the defects indicated by the algorithm that were confirmed by the remote scaler) and the F1 value (Table 3, Figure 3).

Table 3. Recognition of the wood defects of the birch veneer logs by the computer vision algorithm, compared with the remote scaler (153 logs), with the 95% Wilson confidence intervals of the recall and the precision. Ordered by the recall.

Wood Defect	Scaler-Positive Logs	Recall, % (95% CI)	Precision, % (95% CI)	F1
Core discolouration	96	97 (91–99)	91 (84–95)	0.94
Pith and drying cracks	88	91 (83–95)	96 (90–99)	0.94
Core (forest) rot	8	88 (53–98)	33 (17–55)	0.48
Dead-bark wound	11	64 (35–85)	78 (45–94)	0.70
Ingrown bark	21	52 (32–72)	65 (41–83)	0.58
Dead and rotten knots	52	44 (32–58)	77 (59–88)	0.56
Felling and bucking cracks	7	43 (16–75)	23 (8–50)	0.30

Scaler-positive logs: the number of logs out of the 153 log total in which the remote scaler recorded a defect; the basis for the recall.

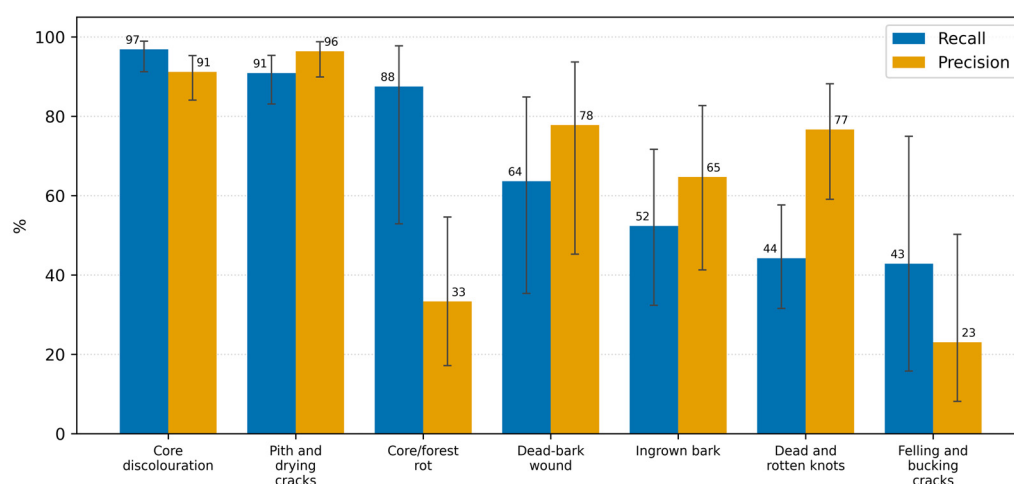


Figure 3. Recall and precision of the computer vision algorithm for each wood defect, compared with the remote scaler (153 logs). The error bars show the 95% Wilson confidence intervals.

The recognition differed considerably between the wood defects. Core discolouration, pith and drying cracks and core rot were recognised well, with the recall from 88 to 97%, while dead-bark wounds, ingrown bark, dead and rotten knots and felling and bucking cracks were recognised considerably worse, with the recall from 43 to 64%. The precision was high for core discolouration and pith and drying cracks (91 and 96%), but low for core rot and for felling and bucking cracks (33 and 23%, see Section 3.3).

3.2. Recognition of Knots

As knots are one of the most important wood defects for the quality class of birch veneer logs, the recognition of dead and rotten knots was evaluated separately (Figure 4). On the side surfaces where the algorithm indicated a knot, the remote scaler confirmed a knot in 77% of the cases (precision). However, the algorithm recognised only 44% of the dead and rotten knots that were evaluated by the remote scaler (recall), so more than a half

of the knots were not recognised. At the same time, on the side surfaces that were free from the knots, the algorithm indicated a knot only in about 7% of the cases. Thus, the main problem in the knot recognition is not the false indication of knots but the missed knots.

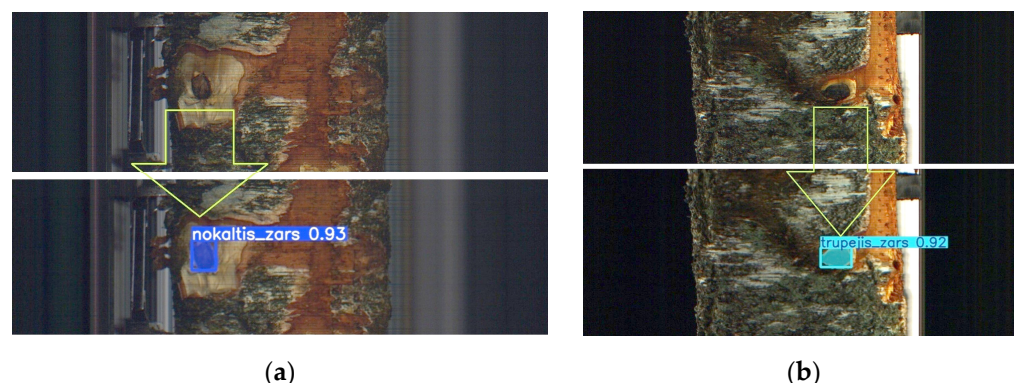


Figure 4. Recognition of the knots on the side surface of the birch veneer logs by the computer vision algorithm: (a) a dead knot and (b) a rotten knot. In each panel the same part of the log side surface is shown without the detection (**top**) and with the detection (**bottom**), where the arrow points to the recognised knot and the box gives the predicted class and confidence value (the class labels are in Latvian: nokaltis zars—dead knot; trupējis zars—rotten knot). For better visibility in print, the brightness and contrast of the images were increased uniformly across the whole image. No features were added, removed or selectively altered. The detection was performed by the algorithm on the original images.

3.3. Over-Indication of Core Rot

For some wood defects the algorithm indicated a defect more often than the remote scaler. The clearest case was core rot, for which the precision was only 33%, indicating a high number of false positives: although the recall was high (88%), two thirds of the core rot indications of the algorithm were not confirmed by the remote scaler (14 not confirmed against seven confirmed). A similar but less certain over-indication was seen for felling and bucking cracks (precision 23%). With only seven such logs, this value is uncertain.

3.4. Assignment of the Quality Class

Overall, the algorithm assigned the same quality class as the remote scaler for a little more than half of the logs, and where the two differed the algorithm more often assigned a lower class than a higher one. The quality class assigned by the algorithm was compared with the quality class of the remote scaler for the same 153 logs (Table 4, Figure 5). According to the remote scaler the logs were distributed between the quality classes as follows: I—23%, II—51%, III—8% and the non-conforming material (reject)—18%. According to the algorithm the distribution was I—26%, II—33%, III—20% and reject—21%.

Table 4. Quality class assigned by the computer vision algorithm against the quality class of the remote scaler (153 logs). The rows show the class assigned by the remote scaler and the columns the class assigned by the algorithm, and the diagonal (in bold) shows the coinciding evaluations.

Remote Scaler Class	Algorithm-Assigned Class				Total
	I	II	III	Reject	
I	26	6	3	0	35
II	13	35	20	10	78
III	0	7	5	1	13
Reject	0	3	3	21	27
Total	39	51	31	32	153

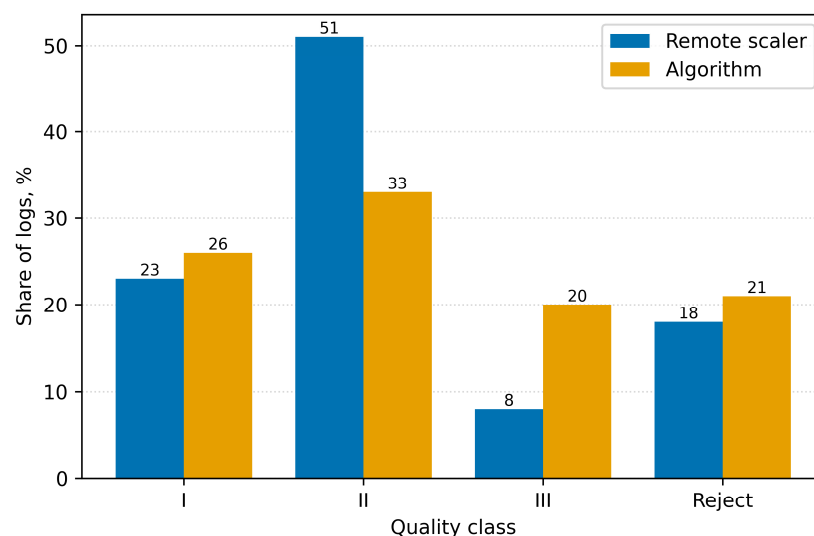


Figure 5. Quality-class distribution of the 153 birch veneer logs according to the remote scaler and according to the algorithm.

The quality class of the algorithm coincided with the quality class of the remote scaler in 57% of the cases, with a 95% Wilson confidence interval from 49 to 64% and a Cohen's kappa of 0.40 (95% CI 0.29 to 0.51), which is a fair agreement [17]. The algorithm assigned a lower quality class than the remote scaler in 26% of the cases and a higher quality class in 17% of the cases, that is, the algorithm tended to lower the quality class more often than to raise it. The class distribution assigned by the algorithm differed systematically from the distribution of the remote scaler, and the disagreements were not symmetric (Bowker test of symmetry, $\chi^2 = 16.61$, $df = 5$, and $p = 0.005$, and Stuart–Maxwell test of marginal homogeneity, $\chi^2 = 14.11$, $df = 3$, and $p = 0.003$). Most of the shift was in the second-class logs that were reassigned to the third class. The largest changes were in the second quality class: of the 78 logs that the remote scaler evaluated as the second class, only 35 (45%) were also assigned to the second class by the algorithm, while 13 were raised to the first class, 20 were lowered to the third class and 10 were assessed as non-conforming material.

The lowering tendency had also the opposite side. Of the 27 logs that the remote scaler evaluated as non-conforming material, 21 (78%) were also rejected by the algorithm, but six (22%) were assigned to the second or the third quality class, that is, they would be passed further as a usable material. At the same time, of the 32 logs that the algorithm rejected, 11 (34%) were in fact a usable material according to the remote scaler (10 of the second class and one of the third class).

3.5. The Manual Quality Class Assignment as a Reference

To give a reference for the manual grading itself, the quality class agreement between the line scaler (human), who works under the production conditions, and the control scaler, who evaluates each log in detail, was 79% overall and varied from 69 to 88% between the individual line scalers (evaluated per shortlog). This value is not directly comparable with the agreement of the algorithm reported above, because it was obtained on a different and larger set of logs and was evaluated per shortlog and not per whole log. In addition, the 79% comes from a production-like flow while the 57% comes from the defect-enriched set of 153 logs, so at the production prevalence of defects, the precision of the algorithm would most likely be lower rather than only different. It shows that even the manual quality class assignment is not fully repeatable. This human-to-human agreement is therefore the natural benchmark for the algorithm: because even a careful manual grading of the same logs is not fully repeatable, a perfect coincidence cannot be expected from the algorithm

either, and its 57% agreement with the remote scaler has to be read against this level of the manual consistency rather than against an ideal of 100%.

4. Discussion

4.1. *The Recognition Is Better for the Core-Zone Defects than for the Knots and the Less Conspicuous Defects*

The main result of this study is that the recognition of the wood defects by the computer vision algorithm was not uniform, and it followed the visual character of the defect. The wood defects of the central, core part of the log—the core discolouration, the core rot and the pith and drying cracks—were recognised well, because they appear as a large area of a changed colour or as the cracks that start from the pith, that is, as a clear and large signature in the central part of the log end face. The dead and rotten knots, the less conspicuous wood defects (the ingrown bark and the felling and bucking cracks) and the dead-bark wounds were recognised considerably worse. For the knots, which are evaluated on the bark-covered side surface, the main reason is the difficulty the algorithm has with the birch bark (Section 4.2). For the ingrown bark and the felling and bucking cracks, the most probable reason is that these wood defects are smaller and have a less distinct visual signature than the large colour defects of the core. It should be noted, however, that the present data do not contain the size of the defects, so this explanation could not be tested directly. The dead-bark wound is a separate case. Although it had the highest recall of this group (64%), it lies on the side surface and was evaluated only where it reaches the log end face (Section 4.5), so its recognition is limited by where it can be observed rather than by a less distinct appearance. The uneven recognition is in agreement with the general observations that the natural variation of the surface and the bark makes the image-based defect detection difficult [11], and that the log grading combines the different tasks of the log end grading and the log side grading [12]. The end-face model finds the core-zone defects, which were recognised well, while the side-surface model finds the knots, which are still the weakest part, so the two models add different things to the final quality class. The reference here was given by the same scaler who labelled the training data, which is discussed as a limitation in Section 4.5.

4.2. *The Knots and the Effect of the Birch Bark*

The dead and rotten knots are one of the most important wood defects for the quality class, and their recognition showed the most characteristic pattern. When the algorithm indicated a knot, the indication was correct in 77% of the cases (precision), which is close to the precision of the surface knot detection reported for sawn wood by Lopes et al. [22] (precision 0.77). However, the algorithm recognised only 44% of the dead and rotten knots evaluated by the remote scaler, so more than a half of the knots were not recognised, while the false indication of the knots on the knot-free surfaces was low (about 7%). Therefore, the main difficulty is not that the bark is mistaken for a knot, but that the real knots are not noticed. This is in agreement with the literature on bark-covered logs: Reich et al. [11] report that on bark-covered logs the detection accuracy of a single image modality remains low (the average precision of the RGB stream alone was much lower than that of the fusion with the three-dimensional data), and that the bark makes the image-based detection of the knots difficult even for human observers. Nguyen et al. [23] similarly found that bark causes both the false indications and the missed defects, and that even after a special modelling of the bark, a part of the defects remained unnoticed. It should be noted that in our study the remote scaler evaluated the same images as the algorithm and still found the knots that the algorithm did not recognise. The missed knots therefore reflect a limitation

of the algorithm itself in recognising the knots on the bark-covered surface, and not only their limited visibility in the images.

4.3. The Assignment of the Quality Class and Its Practical Consequences

As described in Section 3.4, the algorithm agreed with the remote scaler on the quality class for most of the logs and, where they differed, assigned a lower class more often than a higher one. The most important single cause of this lowering was the over-indication of core rot, for which two thirds of the indications were not confirmed by the remote scaler (precision 33%), because a non-existing defect moves the log into a lower quality class. The cause of this over-indication was not established in the present study and would require a separate analysis of the model and its training data. However, several causes are possible. Core rot appears on the end face as a dark, discoloured core zone that is visually similar to the core discolouration, which was the most frequent defect in our data, so the model may confuse the two. Core rot was also rare in the evaluation set (8 logs), and rare classes are known to be learned less stably by the convolutional models [24]. Several directions could therefore reduce the over-indication without changing the grading rules. Adding “negative” (distractor) classes, so that the model learns to reject the rot-like features that are not rot, such as a heavy discolouration and the mineral inclusions, has been shown to reduce false detections in the other tasks [7,23]. Enlarging and balancing the rot training data would address its scarcity. A texture descriptor such as the local binary pattern improved the optical detection of rot in felled stems [25]. Tuning the confidence and the minimum-size threshold of the rot class may also reduce false detections. However, this does not explain all of the lowering: in about a third of the lowered logs the algorithm had not indicated any defect that the remote scaler did not also see, so the lower class most likely follows not from an invented defect but from a jointly detected defect being graded as more severe, for example measured as larger or as reaching the peeling cylinder. As the individual measured values were not evaluated in this study (Section 2.6), this remains a probable explanation rather than a measured result. The missed knots act in the opposite direction, because a log with an unnoticed knot can be evaluated as better than it is, and this explains a part of the cases where the algorithm assigned a higher quality class. As a result, the second quality class, which is the most frequent class, was the most affected: only 45% of the logs evaluated by the remote scaler as second class were also assigned to the second class by the algorithm. Because the quality class itself is assigned by a fixed rule module that implements the standard (Section 2.4), these errors originate mainly in the detection and the measurement of the wood defects rather than in the grading rules.

Both directions of the error have a practical meaning, and they are not equal from the point of view of the value. On the one hand, of the logs that the remote scaler rejected, 22% were assigned by the algorithm to a usable quality class, which means that a part of the unsuitable logs would be passed further. On the other hand, of the logs that the algorithm rejected, 34% were in fact a usable material, which means that a part of the suitable veneer logs would be unnecessarily rejected and their value would be lost. For the practical use of the algorithm the reduction of the over-indication of the core rot is therefore one of the most important tasks. It also follows that the algorithm is best applied as a decision-support aid to the line scaler rather than as an autonomous replacement, because its current tendency to under-grade would, if applied directly, unnecessarily reduce the value of the usable logs.

4.4. Comparison with the Other Studies and the Scale of the Evaluation

Automated image-based grading of logs has so far been studied mostly for the coniferous species and on the basis of the end-face (cross-sectional) images or the external geometry, and the reported accuracy of recent CNN models is about 80% for quality sorting

from the end-face images [3], whereas image-based classification of logs by the market value reached a considerably lower accuracy of about 40 to 55% [26]. These results are, however, not directly comparable with our study, because they concern the other species, a smaller number of the quality classes and the end-face images, while in the present study a four-class quality assignment of the birch veneer logs was made on the basis of the defects on both the end face and the bark-covered side surface under industrial conditions. The lower coincidence of the quality class (57%) is therefore expected, and it lies within the range of the values reported for the image-based log quality and value classification. Optical detection of rot on the end face is in itself a difficult task: a CNN recognised root rot on cross-section (stem end) images of felled Scots pine stems with about 63% accuracy, which increased to about 70% with an additional local binary pattern texture operator [25] (although for a different species and for root rot). The over-indication of core rot observed in the present study, with a high recall (88%) but a low precision (33%), is consistent with this difficulty of the optical assessment of the rot, even though its specific cause in our data was not established (Section 4.3). The limited repeatability of the manual grading itself has likewise been reported for sawn timber, where manual grading reached only a moderate quality yield against an optimal grade and was not fully repeatable between runs [27], which is consistent with the incomplete repeatability of the manual grading observed in the present study (Section 3.5).

To our knowledge, no previous study has evaluated the optical computer vision defect detection and quality grading of bark-covered birch veneer logs under such industrial production conditions, which is the main contribution of this study. This study was carried out under real industrial conditions rather than in a laboratory: the algorithm operates on the full production flow of an operating sorting line (the visual data were recorded for more than 1.6 million logs during the study), although the detailed comparison with the remote scaler was made on a subset of 153 logs, so the accuracy values characterise this subset and not the whole production flow. Previous studies were carried out mostly on sawn wood or in the laboratory, and although some of them used larger labelled data sets (for example 9454 boards [4,10] and 5549 auction log images [26]), none evaluated the bark-covered birch veneer logs on an operating production line.

4.5. Limitations

Several limitations have to be taken into account.

First, the evaluation of the algorithm was made against a single remote scaler who assessed the same visual data as the algorithm. This makes the comparison fair, because both the algorithm and the remote scaler had the same information, but the reference is the visual judgement of one person and not a physical measurement of the logs. A second remote scaler was not used, so the agreement of the reference with another equally experienced evaluator could not be estimated, and the coincidence values should therefore be understood as the agreement between the algorithm and a careful expert evaluation of the same images.

Second, the training images were labelled by a single experienced control scaler, and this was the same person who later graded the logs as the remote scaler, that is, as the reference. The models were therefore trained on his labels and then compared with his own grading, which could make the agreement look higher than it really is. This effect is small. The quality class is not decided by the models but by fixed rules from the defects they find, and the 153 evaluation logs were not among the training images, so the models had not seen them before. The agreement was also only 57%, and the algorithm mostly gave a lower class, so the models did not simply repeat his grading. Better future work would use a reference scaler who did not label the training data.

Third, the comparison concerns the detection of the wood defects and the resulting quality class, and not the accuracy of the individual measured characteristics defined in the standard (for example the knot diameter or the crack dimensions). Their combined effect is, however, reflected in the agreement of the quality class.

Fourth, the 153 logs used for the defect and quality class evaluation were selected so that they contain the wood defects, and they include more logs of the reject class (18% of the evaluated logs) than the production flow, where about 5% of the shortlogs are rejected, so the values describe a defect-rich subset and should not be transferred directly to the whole population. In addition, for the rare defects (the core rot, the felling and bucking cracks, and the dead-bark wounds) the number of the logs is small, so their precision and F1 value have a wide uncertainty.

Fifth, the visual data of the 153 logs could not be linked to the measurements of the control scaler, who sees the whole surface of the log, so the algorithm could not be compared with the full physical evaluation. In production, the quality class is also assigned to each shortlog (the peeling section) of the log separately. At the time the evaluation data set was prepared, however, the algorithm assigned a single quality class to the whole log, so both the algorithm and the remote scaler assessed each log as a whole. The grading of each shortlog separately by the system was introduced later, and an evaluation per shortlog remains a task for the further development. The dead-bark wound was evaluated on the end-face images, where its effect on the peeling cylinder is observable, so the wounds that remained on the side surface without reaching a log end were outside the present evaluation.

Finally, this study was carried out on a single sorting line and with one image acquisition system, so the results may differ on the other sorting lines and with the other cameras and the lighting. The models are specific to birch and would require a separate training and validation before they are used for other tree species.

A comparison with the three-dimensional point-cloud data was not carried out in this study, so it remains a direction for future work rather than a present result.

4.6. Practical Implications and Further Development

For practical use, it can be concluded that the algorithm already gives reliable support for the core discolouration and the pith and drying cracks, which can be assessed automatically. The main limitation is the recognition of the dead and rotten knots, which are hidden by the bark of the birch on the side surface. The less conspicuous wood defects (the ingrown bark, the dead-bark wounds and the felling and bucking cracks) are also recognised less reliably.

Further development should therefore be directed, first, to improving the recognition of the knots by enlarging and improving the training data. Combining colour images with three-dimensional (point cloud) data of the surface could be explored as a further direction, since this has been shown to improve knot detection on bark-covered logs [11] and has also been used for the three-dimensional recognition of wood defects [28]. Adding the surface-profile sensors to the sorting line would make such a comparison possible and is a feasible next step for the system. Imaging a larger part of the side surface may give a further gain. Second, the over-indication of the core rot, which is the main cause of the unnecessary lowering of the quality class, should be reduced. Third, the accuracy of the measured characteristics of the detected defects (their size and their position with respect to the peeling cylinder) should be improved, since a part of the quality class disagreements most likely came not from the missed or the falsely indicated defects but from a more severe measurement of the jointly detected defects (Section 4.3). Evaluating this measurement accuracy directly would require an independent reference of the defect

dimensions, measured both on the images and physically on the logs, which was outside the scope of the present study, and it remains a task for further work. It should also be kept in mind that even the manual quality class assignment is not fully repeatable—in the present project the agreement between the line scaler and the control scaler varied from 69 to 88% between the individual scalers—so the aim of the automated system is not a perfect coincidence but a consistent and reliable support of the line scaler.

5. Conclusions

1. The recognition of the wood defects of the birch veneer logs by the computer vision algorithm differed considerably between the defects. The defects of the central, core part of the log—the core discolouration, the pith and drying cracks and the core rot—had a high recall (from 88 to 97%), while the dead and rotten knots and the less conspicuous defects (the ingrown bark, the dead-bark wounds and the felling and bucking cracks) had a considerably lower recall (from 43 to 64%).
2. For the dead and rotten knots, the main problem is not the false indication of the knots (the precision was 77% and the false indication on the knot-free surfaces was about 7%), but the missed knots: 56% of the knots evaluated by the remote scaler were not recognised, which is explained by the birch bark, on which the knots are difficult for the algorithm to distinguish.
3. The quality class determined by the algorithm coincided with the evaluation of the remote scaler in 57% of the cases. The algorithm assigned a lower quality class more often than a higher one. The most important single cause was the over-indication of the core rot (precision 33%), while a part of the lowering probably followed from a more severe measurement of the real defects, although this was not quantified. As a result, a part of the usable veneer logs would be unnecessarily rejected and a part of the non-conforming logs would be passed further.
4. The algorithm was evaluated under real industrial conditions, on an operating sorting line, on a set of 153 birch veneer logs, and the present results provide a baseline for the further development of the system.
5. For the further development it is recommended to improve the recognition of the knots, in the first place, by enlarging and improving the training data. Combining the colour images with the three-dimensional data of the surface could be explored as a further direction. Imaging a larger part of the log side surface may give a further gain. It is also recommended to reduce the over-indication of the core rot, which is the main cause of the unnecessary lowering of the quality class.
6. Overall, the automated grading of the birch veneer logs under industrial conditions is feasible, but at present it is limited by the detection of the defects on the bark-covered surface, above all the dead and rotten knots.

Author Contributions: Conceptualization, J.M.; methodology, J.M. and M.M.; validation, M.M. and A.G.; formal analysis, M.M. and A.G.; investigation, M.M.; resources, J.M.; data curation, M.M.; writing—original draft preparation, M.M. and A.G.; writing—review and editing, J.M., M.M. and A.G.; visualization, M.M. and A.G.; supervision, J.M.; project administration, J.M.; and funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Forest Sector Competence Centre of Latvia project “Determining the quality of round timber using machine learning algorithms, measured individually with an automatic measuring device” (Recovery and Resilience Facility, 5.1.1.2.i.0/1/22/A/CFLA/007). The APC was funded by VMF LATVIA SIA.

Data Availability Statement: The data presented in this study are not publicly available owing to commercial confidentiality, as they were collected under production conditions at a commercial

veneer mill. They may be available from the corresponding author on reasonable request. The computer vision models and their source code are proprietary to VMF LATVIA SIA and are not publicly available.

Acknowledgments: The authors thank the development team for preparing the computer vision algorithm and the in-house software used in this study. During the preparation of this manuscript, the authors used Claude (Claude Opus 4.8, Anthropic) for the purposes of assisting with drafting and editing the text and with organising and verifying the literature. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: All authors are employed by VMF LATVIA SIA, an independent roundwood measurement company that developed the computer vision system evaluated in this study. This could be perceived as a potential conflict of interest. The authors declare no other conflicts of interest. Under the funding programme, VMF LATVIA SIA is eligible for a higher reimbursement of the project's costs if a set number of publications is produced within a defined period, an incentive that is contingent on publishing and not on the findings reported. The funders had no role in the design of the study, in the collection, analyses or interpretation of data, or in the writing of the manuscript.

References

1. Central Statistical Bureau of Latvia. *Most Important Goods in Latvia's Exports (Table ATD120): Veneer and Plywood (CN 4408, 4412)*; Official Statistics of Latvia: Riga, Latvia, 2026. Available online: <https://data.stat.gov.lv/sq/29545> (accessed on 20 June 2026).
2. *LVS 80-7:2024; Quality Requirements for Round Timber—Part 7: Quality Requirements for Birch Veneer Blocks*. Technical Committee LVS/STK 38. Latvian Standard (LVS): Riga, Latvia, 2024.
3. Achatz, J.; Lukovic, M.; Hilt, S.; Ladrach, T.; Schubert, M. Convolutional neural networks for quality and species sorting of roundwood with image and numerical data. *Expert Syst. Appl.* **2024**, *246*, 123117. [\[CrossRef\]](#)
4. Gazo, R.; Wells, L.; Krs, V.; Benes, B. Validation of automated hardwood lumber grading system. *Comput. Electron. Agric.* **2018**, *155*, 496–500. [\[CrossRef\]](#)
5. Liew, P.Y.; Akbar, M.F.; Ab Wahab, M.N.; Rosdi, B.A.; Fauthan, M.A.; Shrifan, N.H.M.M. The Prospect of Artificial Intelligence-Based Wood Surface Inspection: A Review. *IEEE Access* **2024**, *12*, 84706–84725. [\[CrossRef\]](#)
6. Ali, M.; Hashim, U.R.; Kanchymalay, K.; Wibawa, A.P.; Salahuddin, L.; Rahiddin, R.N.N. A review of recent deep learning applications in wood surface defect identification. *IAES Int. J. Artif. Intell.* **2025**, *14*, 1696–1707. [\[CrossRef\]](#)
7. Stephani, H.; Weibel, T.; Rösch, R.; Moghiseh, A. Challenges and approaches when realizing online surface inspection systems with deep learning algorithms. *Discov. Data* **2023**, *1*, 3. [\[CrossRef\]](#)
8. Goycochea Casas, G.; Ismail, Z.H.; Garcia Leite, H. Computer Vision, Machine Learning, and Deep Learning for Wood and Timber Products: A Scopus-Based Bibliometric and Systematic Mapping Review (1983–2026). *Forests* **2026**, *17*, 112. [\[CrossRef\]](#)
9. Shah, H.; Saarela, E.; Korkeakangas, T.; Pitkääho, T. A comprehensive review of automatic defect detection in wooden surface inspection. *Appl. Comput. Intell.* **2026**, *6*, 1–22. [\[CrossRef\]](#)
10. Wells, L.; Gazo, R.; Del Re, R.; Krs, V.; Benes, B. Defect detection performance of automated hardwood lumber grading system. *Comput. Electron. Agric.* **2018**, *155*, 487–495. [\[CrossRef\]](#)
11. Reich, B.; Kunda, M.; Zolotarev, F.; Eerola, T.; Zemčík, P.; Kauppi, T. Multimodal surface defect detection from wooden logs for sawing optimization. *arXiv* **2025**, arXiv:2503.21367. [\[CrossRef\]](#)
12. Sandvik, Y.J.; Futsæther, C.M.; Liland, K.H.; Tomic, O. A Comparative Literature Review of Machine Learning and Image Processing Techniques Used for Scaling and Grading of Wood Logs. *Forests* **2024**, *15*, 1243. [\[CrossRef\]](#)
13. Magaznieks, J.; Millers, M. Bark Abrasion Evaluation on Birch (*Betula* spp.) Round Timber by Using Machine Learning Algorithm. In *Environment. Technology. Resources, Proceedings of the 16th International Scientific and Practical Conference, Rēzekne, Latvia, 19–20 June 2025*; RTU Press: Riga, Latvia, 2025; Volume 1, pp. 349–353. [\[CrossRef\]](#)
14. Ji, M.; Zhang, W.; Diao, X.; Wang, G.; Miao, H. Intelligent Automation Manufacturing for *Betula* Solid Timber Based on Machine Vision Detection and Optimization Grading System Applied to Building Materials. *Forests* **2023**, *14*, 1510. [\[CrossRef\]](#)
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [\[CrossRef\]](#)
16. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019*; PMLR: Long Beach, CA, USA, 2019; Volume 97, pp. 6105–6114. [\[CrossRef\]](#)
17. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [\[CrossRef\]](#)

18. Wilson, E.B. Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **1927**, *22*, 209–212. [[CrossRef](#)]
19. Bowker, A.H. A Test for Symmetry in Contingency Tables. *J. Am. Stat. Assoc.* **1948**, *43*, 572–574. [[CrossRef](#)] [[PubMed](#)]
20. Stuart, A. A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika* **1955**, *42*, 412–416. [[CrossRef](#)]
21. Maxwell, A.E. Comparing the Classification of Subjects by Two Independent Judges. *Br. J. Psychiatry* **1970**, *116*, 651–655. [[CrossRef](#)] [[PubMed](#)]
22. Lopes, D.J.V.; dos Santos Bobadilha, G.; Grebner, K.M. A fast and robust artificial intelligence technique for wood knot detection. *BioResources* **2020**, *15*, 9351–9361. [[CrossRef](#)]
23. Nguyen, V.-T.; Constant, T.; Colin, F. An innovative and automated method for characterizing wood defects on trunk surfaces using high-density 3D terrestrial LiDAR data. *Ann. For. Sci.* **2021**, *78*, 32. [[CrossRef](#)]
24. Li, D.; Xie, W.; Wang, B.; Zhong, W.; Wang, H. Data Augmentation and Layered Deformable Mask R-CNN-Based Detection of Wood Defects. *IEEE Access* **2021**, *9*, 108162–108174. [[CrossRef](#)]
25. Holmström, E.; Kainulainen, H.; Raatevaara, A.; Pohjankukka, J.; Piri, T.; Honkaniemi, J.; Lehtonen, A. Automatic detection of root rot and resin in felled Scots pine stems using convolutional neural networks. *Int. J. For. Eng.* **2024**, *35*, 153–165. [[CrossRef](#)]
26. Triplat, M.; Lukančič, Ž.; Kavčič, V. Evaluation of Deep Learning Models for Image-Based Classification of Timber Logs by Market Value. *Forests* **2026**, *17*, 518. [[CrossRef](#)]
27. Lycken, A. Comparison between automatic and manual quality grading of sawn softwood. *For. Prod. J.* **2006**, *56*, 13–18.
28. Sioma, A. Assessment of Wood Surface Defects Based on 3D Image Analysis. *Wood Res.* **2015**, *60*, 339–350.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.