Using machine learning to predict the density profiles of surface-densified wood based on cross-sectional images

Benedikt Neyses1 · Alexander Scharf1

Received: 2 August 2021 / Accepted: 21 April 2022 / Published online: 4 June 2022
© The Author(s) 2022

Abstract
Over the past decades, the surface densification of solid wood has received increased attention. However, the inhomogeneous density distribution in the densification direction might be a challenge with regard to process control within a large-scale production process, as the density profile governs many relevant properties of surface-densified wood. Currently, the measurement of density profiles relies on sensitive X-ray equipment and is difficult to integrate into an on-line process. Hence, in this study, three machine learning approaches were applied to predict the density profiles of surface-densified Scots pine specimens, only based on visual image acquisition—a technology that is ubiquitous in the wood industry: partial least squares (PLS) regression, artificial neural networks (ANN), and convolutional neural networks (CNN). The machine learning models were trained on images of the specimen cross-sections as input data, and X-ray density profiles as output data. There were 1850 observations, and the model performance was evaluated on external test sets. The models had mean absolute percentage errors of the predicted values between 9 and 18%; the CNN achieving the smallest error (9.24%). A deeper analysis of the data revealed that the ANN approach performed inconsistently between observations. PLS regression predicted the main density peak to a high accuracy but could not model other features. Only the CNN could reliably model the main density peak, wide growth rings, and the important region between the specimen surface and the main density peak. The ability of the models to generalise to untypical new data was improved by augmentation of the training data.

1 Introduction

Wood densification, i.e., the compression of solid wood in the transversal direction, has been a research subject since the beginning of the twentieth century, and densified wood products have been available for purchase for a similar period (Kutnar et al. 2015). For the most part, the purpose of wood densification has been the improvement of the mechanical properties, such as hardness, abrasion resistance or bending strength. Usually, the mechanical properties increase to a similar degree to the achieved increase in density (Kamke 2006; Kutnar et al. 2008; Fang et al. 2012), and this makes the application of densification particularly attractive to low-density and often low-value wood species.

Conceptually, the densification of solid wood is simple: the wood is plasticised—often by moisture and heat—and then mechanically compressed in a hot press. However, due to several practical problems, densified wood products remain as niche products with low annual production volumes (Jones et al. 2019). One of those problems is the low production process speed, which is caused by the reliance on batch processes and by the focus on bulk densification, where the whole cross-section of a board is densified uniformly (Neyses 2019). The former can be overcome by a continuous densification process, an approach currently under development at Luleå University of Technology in Sweden (Neyses 2019), while the latter has been addressed by research into so-called surface densification, where only the region just beneath the wood surface is compressed.

Surface densification was first studied by Tarkow and Seborg (1968), but only in the past two decades has it experienced increased interest (Pizzi et al. 2005; Gong et al. 2010; Rautkari et al. 2011; Laine 2014; Neyses 2019). For many applications of densified wood, it is not necessary to have an increased density throughout the whole cross-section of the component. For example, wooden flooring only needs a
hard and wear-resistant surface, and in fact, this approach is typical for engineered wooden flooring, where a top layer made from a high-density wood species is glued to a substrate made from a low-density wood species. Similarly, for uses with high bending loads, the largest stresses are in the outermost layers of the tension zone. Previous studies have shown that sufficient improvements of the mechanical properties can be achieved with surface densification (Rautkari et al. 2013; Neyses et al. 2020). As only the region close to the wood surface needs to be affected by the process, the resources required for densification can be reduced, which is a crucial aspect for the large-scale production of densified wood products.

One disadvantage of surface densification is the increased complexity in the relationship between the density distribution and the resulting mechanical properties, with the density profile (DP) in the densification direction being the guiding parameter that governs the properties of the surface-densified wood. For this reason, it is important to have accurate and high-resolution knowledge of the DP during the production of densified wood products. Typically, the DP of wood is measured by X-ray densitometry, which provides density data at a spatial resolution of approx. 0.05 mm. Such equipment is, however, fairly costly and difficult to implement as an on-line system. In contrast, it is much less challenging to implement ordinary computer-based vision systems in the production line, which are ubiquitous in, for example, the sawmill industry.

For this reason, it was hypothesised whether simple cross-sectional images contain enough information to predict the DP of surface-densified wood with the help of modern machine learning approaches. More specifically, darker pixels would correlate to a high density, whereas lighter pixels would correlate to a low density, in accordance with the perceived differences in brightness between latewood and earlywood. During the surface densification process, primarily low-density cells are densified, i.e., earlywood, which should decrease the perceived brightness in those areas of the cross-section, resulting in a similar brightness indicative of densities associated with latewood.

Machine learning is a rapidly growing field with an ever-increasing number of useful applications, especially in image/vision-related tasks, and strong progress has been made over the past decade (Zhang and Zhou 2013; Alzubaidi et al. 2021). However, in the field of wood science, there have only been few studies exploiting the capabilities of modern machine learning approaches (Gu et al. 2010; Iliadis et al. 2013; Hu et al. 2019; Demir et al. 2021). Traditionally, algorithms for problem solving have consisted of specific instructions, written by the programmer—so-called rule-based systems. Consider however, the classification of animals in an image. For humans it is trivial to distinguish a dog from a cat, but to manually program an algorithm that can perform this task reliably on thousands of images is virtually impossible. Enter modern machine learning: the programmer provides the data, the desired goal, and a rough framework, and based on that information the algorithm searches for the optimal program to solve the problem at hand (Karpathy 2017; Alzubaidi et al. 2021).

Thus, the objective of this study was to build models that can predict high-resolution DPs from cross-sectional images of densified wood. The purpose is to enable on-line DP measurements with regular computer vision systems, to be used in the production of densified wood products—something that is not possible today. To do so, three different types of machine learning algorithms were applied, optimised, and evaluated in terms of their prediction performance on an external test set: partial least squares (PLS) regression, artificial feed-forward neural networks (ANNs), and convolutional neural networks (CNNs).

PLS was chosen as it is a well-established prediction method for noisy and multivariate datasets that contain a high level of collinearity between multiple input and output variables (Wold et al. 2001). Conceptually, PLS regression reduces the data into a set of uncorrelated components, with the components maximising the co-variance between the input and output data. Afterwards, least squares regression is performed on the PLS components to predict the outputs from the input data. The other two approaches belong to the category of deep learning, which, being proposed in the 1960s, is not a new paradigm (Schmidhuber 2015), but only with more recent advancements in computing power has it been possible to successfully apply such algorithms to large multivariate datasets (LeCun et al. 1998). Deep learning is a subset of machine learning, in which multi-layered neural networks learn from vast amounts of data, and it is inspired by the information processing patterns found in the human brain. For image-related applications, the most commonly used deep learning approaches are ANNs and CNNs, with the latter currently being the most popular architecture. In theory, both approaches can approximate any function—such as DPs—to arbitrary accuracy, given sufficiently good training data (Heinecke et al. 2020; Zhou 2020). In contrast to ANNs, CNNs have the ability to learn spatial relationships in image data (Jaderberg et al. 2015), which might help to model anatomical features (i.e., the growth rings) and generalise to a larger range of input data.

2 Materials and methods

2.1 Specimen preparation

Defect-free Scots pine (Pinus sylvestris L.) sawn timber with a maximum growth-ring angle of 20° relative to the tangential surface was used. 464 specimens with dimensions
of 50 × 21/18.5 × 50 mm (longitudinal × radial × tangential) were cut from the sapwood regions and conditioned to equilibrium moisture content at 20 °C and 65% RH. For the surface densification treatment, the specimens were randomly distributed into eight groups of 58 specimens, as shown in Table 1. The oven-dry density ranged from 425 kg m⁻³ to 578 kg m⁻³ and the average moisture content was 13%. Both properties were estimated through measurements on additional matched specimens to avoid complete drying of the specimens prior to densification.

The densification of the specimens was done in a laboratory hot press (Fjellman Press AB, Mariestad, Sweden) with one heated platen. Compressing the specimens with two different initial thicknesses to the same target thickness of 17 mm resulted in the two different compression ratios. The specimens were placed with the tangential bark-side surface towards the heated platen. 10 s of low contact pressure raised the temperature of the specimen surface region, thereby plasticising it. The pressure was then increased (3–5 MPa) to densify to the target thickness, which was secured by mechanical stops. All specimens were kept under pressure and heat for 240 s, regardless of the other process parameters. Before releasing the pressure, the plate was cooled until it reached a temperature of 35 °C (240–360 s, depending on the pressing temperature).

Each densified specimen was split lengthwise into two pieces, 50 mm × 17 mm × 14.5 mm in size (Fig. 1). By cutting off the bulged-out sides of the specimens, the cuboid shape, preferred for densitometry, is restored while also securing even specimen dimensions. Afterwards, the cross-sectional surfaces were sanded with grit 80 glasspaper.

### 2.2 Data collection and dataset preparation

In this study, the image scans of the cross-sectional specimen surfaces represented the X-data (input), with greyscale pixel values as X-variables. The X-data was used to train machine learning models that could predict the Y-data (output), i.e., the DPs, which consisted of density measurement points as Y-variables. Figure 1 shows an overview of the data collection and the preparation of the datasets performed before training the machine learning models.

Density profiles in the radial direction were obtained with a Dense-Lab-X densitometer (Electronic Wood Systems GmbH, Hameln, Germany) with a spatial resolution of the X-ray beam of 0.050 mm, and a step length of 0.066 mm, resulting in 257 measurement points per specimen. The mass and dimensions of each specimen were recorded prior to the density profile measurements to determine the bulk density as a reference.

1856 greyscale photo scans of the cross-sectional specimen surfaces (2 scans per specimen) were taken with a Ricoh MP C5504ex photo scanner (Ricoh Company Ltd, Tokyo, Japan) at a spatial resolution of 600 dpi. After correcting for rotational errors in the positioning of the specimens, this resulted in a resolution of 401 × 300 pixels (radial × tangential direction) for each cross-sectional image. The scans were saved in the TIFF format, transformed into matrices of grey values with the same resolution, and then normalised to the bulk density of the respective specimen. Three specimens were discarded because of faulty photo scans, resulting in a dataset with 1850 observations. To obtain a Y-matrix with a constant length of rows, the DPs were cut by 20 density measurement points from the undensified end, leaving 237 measurement points as Y-variables.

Prior to applying the machine learning algorithms, the X and Y data points were compiled into three different datasets, with pixel values as X-variables (input) and density measurement points as Y-variables (output) (Table 2). Two types of data compression were carried out to reduce the computational demands during model training. For the dataset Normal S, a 300 × compression of the photo scan data was carried out by mean averaging all values in the tangential direction, resulting in matrices of 401 × 1 pixels. For the dataset Normal L, a 4 × compression of the photo scan data was carried out by mean averaging blocks of 4

| Group ID | Mean density before densification, at 13% MC (kg m⁻³) | CR (%) | T (°C) | Preheating time (s) | Press closing speed (mm min⁻¹) | Pressing time (s) |
|----------|------------------------------------------------------|--------|--------|---------------------|-------------------------------|-------------------|
| D1       | 574 ± 48                                             | 8.1    | 150    | 10                  | 2.7                           | 240               |
| D2       | 574 ± 51                                             | 8.1    | 200    | 10                  | 2.7                           | 240               |
| D3       | 571 ± 40                                             | 8.1    | 150    | 10                  | 1.0                           | 240               |
| D4       | 573 ± 49                                             | 8.1    | 200    | 10                  | 1.0                           | 240               |
| D5       | 574 ± 48                                             | 19.0   | 150    | 10                  | 2.7                           | 240               |
| D6       | 558 ± 44                                             | 19.0   | 200    | 10                  | 2.7                           | 240               |
| D7       | 574 ± 50                                             | 19.0   | 150    | 10                  | 1.0                           | 240               |
| D8       | 573 ± 52                                             | 19.0   | 200    | 10                  | 1.0                           | 240               |

The process parameters are based on a full factorial design with three factors and two levels each. Number of specimens per group = 58. CR compression ratio, T temperature, and MC moisture content
pixels in the tangential direction, resulting in matrices of 401 × 75 pixels. To increase the variation in the data and prevent bias towards certain positions of the main density peak in the training data, the cross-sectional images and the density profiles were mirrored around the tangential axis, resulting in 1850 additional observations for the Combined L dataset. Afterwards, the 401 × 75 pixel-value matrices of the datasets Normal L and Combined L were unwrapped to
create datasets with exactly one row of X- and Y-variables per specimen, which is the required dataset format for the PLS and ANN algorithms. Before splitting the datasets into training sets, validation sets and test sets, the order of the observations was randomised. On a general level, all three machine learning methods applied in this study were trained on the training sets, which contain both the X- and Y-data. Afterwards, the models were optimised based on the validation sets. After saturating the prediction performance through optimisation, the models were evaluated on the external test sets, where the model had to predict the density profiles (Y-data) of unknown specimens, only fed with the photo scans (X-data). The predicted density profiles were then compared with the true density profiles determined by X-ray densitometry.

2.3 Partial least squares (PLS) regression

PLS regression was applied to all three datasets, using the SIMCA 15 software (Sartorius AG, Göttingen, Germany). As PLS regression does not have the same iterative character as the deep learning approaches, the observations from the training and validation sets were compiled into one dataset. No scaling of the datasets was carried out. The PLS algorithm ran seven rounds of cross-validation to determine the latent variables, also known as PLS components. For each Y-variable, a list of coefficients was computed, assigning a weight to each X-variable. In other words, the coefficients represent how strongly the greyscale values in each pixel of an image correlated to the measured density at a specific measurement point. The predicted value for a density measurement point was then calculated by the sum of the weighted greyscale values and the constant Y_avg. The use of the constant Y_avg makes PLS regression highly dependent on the variation in the training set. A greater variation in the training set thus shifts more importance to the coefficients at the expense of the constant Y_avg.

2.4 Artificial feed-forward neural network (ANN)

An ANN (Fig. 2) consists of input and output “layers”, with hidden layers of neurons in between. Neurons are nodes in a network, where each neuron in a layer is connected to all neurons in the preceding and succeeding layers, and weights are assigned to all connections between neurons. In simple terms, a neuron receives the weighted sums of all incoming connections as an input and computes the output by running the input value through an activation function (Jain et al. 1996). After the training data has been fed through the network once, the error between the calculated output and the true output—in this case DPs—is determined. The error information is fed through the network backwards to adjust the weights between the neurons with an approach called backpropagation, and different algorithms can be used to optimise this procedure (Rojas 1996). If appropriate model parameters are chosen, the error between the predicted and true outputs will decrease with every iteration—the model is learning. The iterations are also called epochs. ANNs are straightforward to implement, but they cannot extract local features from an image (LeCun et al. 1998), which means that the ability of the model to generalise well on new data is strongly influenced by the amount of variation within the training data set.

The ANN method was applied to all three datasets. The models were built and trained using the TensorFlow 2.5.0 open-source machine learning platform (Google Brain, Mountain View, California, USA), supported by the Keras open-source machine learning library (François Chollet, Google LLC, Mountain View, California, USA), all running on the Python programming language, and within the Google Colab cloud computing platform (Google LLC, Mountain View, California, USA). Different ANN architectures were iteratively tested, with the following ranges of hyperparameters being tried: number of hidden layers: 2, 3, 4; neurons per hidden layer: 128, 256, 512, 1024, 2048;
batch sizes: 10, 20, 40, 80 observations; optimisers: SGD, AdaGrad, Adam; learning rates: 0.001, 0.005, 0.01. Finally, the architectures shown in Table 3 provided the best performance in terms of mean absolute error (MAE). The final models were trained with a batch size of 20 observations, the mean squared error as the loss function, the ‘Adam’ optimiser (learning rate = 0.001), and the MAE as the performance metric. The activation function for all layers was ‘ReLU’, except for the output layer where a linear activation function was used.

### 2.5 Convolutional neural network (CNN)

ANNs have densely connected layers, which means that each neuron sees all the data from the previous layer, i.e., they always consider the entire image. CNNs on the other hand, consist of several sparsely connected convolutional layers that look at specific parts of an image (Fig. 3), which allows them to extract local features (Lawrence et al. 1997; LeCun et al. 1998). Once a feature is extracted, it can be detected even if it appears in another part of an image. A generic CNN consists of the following steps: convolution, pooling/subsampling, and classification/regression. In the convolutional layers, many filters with different arrangements of pixel values are moved over an input image, and in every position, the pixel values of the image and the filter are multiplied. Each filter thus generates a response map which contains the extracted features. In the pooling step, the dimensionality of each response map is reduced, while still retaining the most important information. These blocks of convolutional and pooling layers can be combined in many ways to create different CNN architectures. For image classification or regression, usually one or more densely connected layers are added, and in the present study the number of neurons in the output layer is equal to number of density measurement points of the DPs.

The CNN method was applied to the datasets Normal L and Combined L. Before training the models, the X-variable array for each observation had to be reshaped into the original 401 × 75 pixel-value matrix, as the convolutional layers of a CNN are conceived to run over a 2-dimensional matrix. For this reason, the CNN method was not applied to the dataset Normal S.

The CNN models were built and trained on the same platform as the ANN models. Several existing architectures were tested: LeNet-5 (LeCun et al. 1998), AlexNet (Krizhevsky et al. 2012), VGG-16 (Simonyan and Zisserman 2014), and ResNet 50 (He et al. 2016). The following ranges of hyperparameters were tried: batch sizes: 10, 20, 40, 80 observations; optimisers: SGD, Adam; learning rates: 0.001, 0.005, 0.01. Finally, the ResNet 50 model architecture provided the best performance. It was originally conceived for image classification, and not for regression, as required for this study; hence, a linear activation function was used on the output layer. In addition, the two fully connected layers fc1 and fc2 before the output layer were changed to 512 neurons, respectively. The final models (Table 4) were trained with a batch size of 20 observations, the mean squared error as the performance metric.

| Model name         | Input neurons | Neurons hidden layer 1 | Neurons hidden layer 2 | Neurons hidden layer 3 | Output neurons | Trainable parameters |
|--------------------|---------------|-------------------------|------------------------|------------------------|----------------|----------------------|
| ANN Normal S       | 401           | 256                     | 256                    | 256                    | 237            | 295,405              |
| ANN Normal L       | 30,075        | 1024                    | 1024                   | 1024                   | 237            | 33,139,949           |
| ANN Combined L     | 30,075        | 1024                    | 1024                   | 1024                   | 237            | 33,139,949           |

Trainable parameters refer to the number of elements the algorithm can adjust to improve its performance.

---

**Fig. 3** Generic structure of a convolutional neural network

© Springer
loss function, the ‘SGD’ optimiser (learning rate = 0.001), and the mean absolute error as the performance metric.

### 2.6 Analysis of the model performance

After training the models, the predicted and true density values of the test sets were exported into Microsoft Excel (Microsoft Corporation, Redmond, Washington, USA) for analysis. The performance of the models was assessed by calculating the coefficient of determination ($R^2$) and the mean absolute percentage error (MAPE) between true and predicted density values, and by examining the predicted and true density profiles of specific specimens, especially those with pronounced growth ring peaks and other peculiarities.

### 3 Results and discussion

In this study, three different machine learning approaches were used to create models that predicted DPs from cross-sectional photo-scans of surface-densified wood specimens. Table 5 shows the mean $R^2$ and MAPE within the test sets between the true and predicted values of all models. The ANN and CNN models were trained for the number of epochs that gave the lowest MAPE on their respective test sets. For the PLS approach, the number of iterations was not a tuneable hyperparameter.

As there are no established methods for the estimation of density profiles (DP) of surface-densified wood that are not relying on X-ray technology (which is used as the ground truth in this study), it is arbitrary to set boundaries of what constitutes a good or a bad performance—it very much depends on the use case. That said, it seems reasonable to consider models that have an MAPE of around 10% or better to be performing well.

The $R^2$ values indicate that all models trained on the Normal L dataset and tested on mirrored data performed badly, to the point where they can be considered unusable. This means that none of the models is capable of generalising to new data that is very different from all the training and validation examples, such as DPs with the main density peak on the opposite site of the specimen. For this reason, the Combined L dataset was created, which dramatically increased the variation within the training data. Still, the performance of the models trained on the Combined L data-set was slightly worse than the best models trained on the Normal datasets.

However, merely looking at the overall $R^2$ or MAPE does not provide sufficient information to properly assess the models, and for this reason, Fig. 4 shows the MAPE of the best performing models of each machine learning approach as a function of the density measurement points, with the first measurement point corresponding to the specimen surface. The prediction performance of all models was bad for

### Table 4 Overview of the applied CNN models

| Model name   | Model architecture | Input shape | Neurons fc 1 layer | Neurons fc 2 layer | Output neurons | Trainable parameters |
|--------------|--------------------|-------------|-------------------|-------------------|----------------|---------------------|
| CNN Normal L | ResNet 50          | 401, 75, 1  | 512               | 512               | 237            | 38,593,133          |
| CNN Combined L | ResNet 50      | 401, 75, 1  | 512               | 512               | 237            | 38,593,133          |

Trainable parameters refer to the number of elements the model can adjust to improve its performance.

### Table 5 Overview of the trained machine learning models after evaluation with the test sets

| Model Name     | Trained epochs/iterations | Observations test set | $R^2$  | MAPE (%) |
|----------------|---------------------------|------------------------|--------|----------|
| PLS Normal S   | 150                       | 185                    | 0.63   | 11.99    |
| PLS Normal L   | 302                       | 185                    | 0.73   | 9.49     |
| PLS Normal L   | 302                       | 555 from mirrored data | 0.03   | 21.86    |
| PLS Combined L | 164                       | 555                    | 0.61   | 11.67    |
| ANN Normal S   | 1600                      | 185                    | 0.62   | 11.83    |
| ANN Normal L   | 4000                      | 185                    | 0.32   | 18.05    |
| ANN Normal L   | 4000                      | 555 from mirrored data | 0.15   | 31.68    |
| ANN Combined L | 3000                      | 555                    | 0.59   | 11.88    |
| CNN Normal L   | 400                       | 185                    | 0.74   | 9.24     |
| CNN Normal L   | 400                       | 555 from mirrored data | 0.04   | 22.40    |
| CNN Combined L 100| 100                      | 555                    | 0.63   | 11.32    |
| CNN Combined L 400| 400                      | 555                    | 0.67   | 10.28    |

$R^2$ R squared coefficient of determination, MAPE mean absolute percentage error.
the first 6–8 measurement points, which was caused by two phenomena. Firstly, there was always some misalignment of the specimens in the densitometer, and as a result for the first few measurements points the X-ray beam was attenuated partially by air and wood. Secondly, the image scans were not perfectly aligned, either, with slight rotational variations between specimens. Both phenomena created noise in the first few X- and Y-variables, which was impossible to model. Around density measurement points 25–45, the MAPE is the lowest for all models, which is the location of the main density peak for most of the specimens, due to the surface densification procedure. Accordingly, the datasets contained a high amount of systematic data in this region, which had a positive effect on the model performance.

To further analyse the performance of the machine learning models, several representative observations from the test sets were chosen to highlight certain characteristics of DPs of surface-densified wood. The true and predicted DPs of these observations are plotted in Figs. 5–9.

Figure 5 shows a typical observation where the corresponding specimen has highly angled growth rings (10–20°), resulting in a density profile that is homogeneous in density, except for the main peak in the densified region. In general, all models performed best on such DPs. For the potential use case of on-line quality and process control, such a smooth DP is ideal, because the only distinct features are the position and shape of the property-defining main density peak. This finding demonstrates a typical behaviour of the PLS regression and ANN algorithms, which are strongly influenced by common systematic variations among the observations in the training data: as mentioned before, most density peaks caused by the surface densification are located at roughly the same distance from the specimen surfaces, while the region between the surface and the main density peak (measurement points 1–25) is less systematic across the training data. As a result, the predicted DPs are much smoother than the true one obtained by X-ray densitometry, and this would most
likely lead to an overestimation of relevant mechanical properties, such as the hardness or bending strength.

Figure 6 presents an observation with a rather shallow and wide density peak in the densified area, and wide, distinct growth rings, which are almost perfectly parallel to the tangential surface in the corresponding specimen. It can be argued that any features beneath the main density peak are irrelevant for the properties of surface-densified wood products, but it demonstrates an important conceptual difference between the PLS and ANN approaches on the one side, and the CNN approach on the other side. The growth rings of the observation shown in the figure were wide enough to be treated as systematic variation by all tested approaches when trained on the Normal L dataset. When the models were trained on the Combined L dataset—which also contains the mirrored observations—the PLS and ANN models were no longer able to predict these features, whilst the CNN model was. The region where the growth rings are located in this particular observation overlaps with the region where the main density peaks are located in the mirrored data. Since the PLS and ANN algorithms rely on systematic variation among the training observations at particular regions in the density profiles, these regions are hence dominated by the main density peak, and not by the growth rings as in the Normal L dataset. In contrast to that, the CNNs can extract localised features from an image, regardless of their positions within an image. A growth ring is still detected as a growth ring, even if half of the training observations have the main density peak located in roughly the same location. More importantly, with regard to potential use cases, the same phenomenon can be observed for the first 20 density measurement points beneath the specimen surface, where only the CNN model could improve its performance after being trained on the Combined L dataset. The other approaches performed worse after adding mirrored observation to the training set. The corresponding regions on the MAPE curves (Fig. 4) support this.

An even more extreme instance of the phenomenon described in the previous paragraph is illustrated in Fig. 7, where two distinct growth rings overlap in the corresponding DP, which generated a very wide and strong peak on the opposite side of the main density peak. Again, the PLS model trained on the Normal L dataset could pick up this density peak to some degree, but when the model was trained on the Combined L dataset, the peak completely disappeared. In contrast to that, the CNN could still model...
the peak, and with more training epochs the prediction performance improved. The same applied to the first 25 density measurement points. The ANN prediction is not shown, as it performed worse than the other two approaches.

Figure 8 shows an observation with very distinct but narrow growth rings, which even the CNN models could not predict. Most likely, the training data was too weak for the algorithm to extract such fine-grained features. A slight offset between the pixel positions in the image scans and density measurement points would break down any reliable relationship. There is some indication that this scenario occurred in this study. For many of the density profiles in the test set that exhibited narrow growth rings, the CNN model predicted distinct density fluctuations, but with the amplitude offset along the x-axis. The question arises whether this failure to predict narrow growth rings is a feature or a bug for most potential use cases. Such narrow features are likely to be irrelevant to any wood properties of interest in surface densification—in fact, for the purpose of on-line quality and process control, they could be considered a distraction from the main density peak.

Figure 8b indicates why the MAPE between true values and predictions increases in the undensified regions of the specimens (Fig. 4). Here, the training data is dominated by the growth rings, which when too narrow could not be modelled by any of the tested algorithms, thereby dragging down the prediction accuracy on the whole test set.

To evaluate the limits of the tested machine-learning approaches in terms of their capabilities to generalise to new observations that are very different from those contained in the training data, all three approaches were trained and validated on the Normal L dataset, but then tested only on mirrored data. Unfortunately, neither the PLS, ANN, nor CNN approach could handle such different data; all algorithms...
always predicted the main density peak location on the opposite side of the specimen—in accordance with the training data (Fig. 9), and consequently, the predicted DPs were useless. For the PLS and ANN models this outcome was expected, as their architectures prevent them from extracting localised features from the data. Any feature they are trained on will always be connected to the position of that feature in the image. While the CNN architecture is able to extract localised features independent of their positions in an image, for the problem at hand, the model needed to know where to place the extracted features in order to build a DP prediction. This is the main reason why the last two hidden layers in the ResNet 50 model have the same architecture as ANNs (He et al. 2016). For this reason, at least some number of observations in the training data needs to be similar to any new observations the model may be confronted with. Previous studies support this finding and report strong improvements of the prediction performance after increasing the variation in the training data through different ways of data augmentation, such as mirroring or rotating images (Papageorgiou et al. 1998; Lienhart and Maydt 2002; Yang et al. 2019).

The mirroring of the image scans and the corresponding density profiles as a way of data augmentation was shown to be a powerful method of increasing the variation within the training data without the need for more specimens and time-consuming physical experiments. For all three tested approaches, the MAPE slightly increased when trained on the Combined L dataset (Table 5, Fig. 4), but on the upside, they performed well on a much wider range of new test data. The slight increase in overall MAPE was mostly due to less accurate predictions in the regions where the main density peaks were located, but this increase was higher for the PLS and ANN than for the CNN. Again, the reason for this is the reliance of the former on systematic variation among the training data in particular regions of the density profiles. Mixing original and mirrored data reduced such systematic variation.

While the consumption of computing resources to train the models was not recorded in this study, it is clear that the PLS approach had the lowest computational demands for same-sized datasets. On a consumer laptop, even the largest model (PLS Combined L) took less than 2 min to train, only using the central processing units. The ANN and CNN models on the other hand, were trained using the Google Colab platform with dedicated graphics processing units. Especially, the ResNet 50 CNN model had a complex architecture with almost 40 million trainable parameters distributed over 50 layers, which led to a training time of up to 5 h. While such computing requirements were not related to significant costs, it is worthwhile to point out that only relatively small decreases in MAPE could be achieved with a several orders of magnitude higher consumption of computing resources.

One of the main limitations of this study was the specimen size, which was smaller than typical in the wood industry. This can be both an advantage and a disadvantage in terms of applicability of this study when considering real use cases. On the one hand, one image scan would need to be representative of a larger wood volume than it is the case in this study. It was neither investigated how much the DPs vary across the length of a surface-densified piece of wood. On the other hand, the larger cross-sectional area of full-sized wood boards may provide more comprehensive image data for training.

At the end of the day, one must consider how such models will be applied, as this will determine whether the superior capabilities of the CNN approach in some areas outweigh its higher complexity and computational demands. In many cases, it might be sufficient—or even preferred—to obtain accurate predictions of high-level features of a DP, i.e., the main density peak, while ignoring low-level features, such as the growth rings. In such a case, the PLS approach may be considered as sufficiently capable. However, the ResNet 50 model could simply be trained for a smaller number of epochs, while still maintaining its performance advantage in modelling the region between the specimen surface and the main density peak, and in being more robust to variation among new observations.

**Fig. 9** True vs. predicted density profiles and cross-sectional image scan for an observation of the test set with mirrored data, after training the models only on the Normal L dataset, which did not contain mirrored data.
Further improvements of the algorithm’s ability to generalise to new data may be achieved by more sophisticated data augmentation. For example, the cross-sectional images and corresponding density profiles could be separated into several columns, which would then be shifted around randomly, thereby creating an arbitrary number of new observations. In addition, newer architectures, such as Transformers (Dosovitskiy et al. 2020) or pre-trained models could be tested.

4 Conclusion

The objective of this study was to train models that can accurately predict the density profiles (DPs) of surface-densified wood, solely based on images of the cross-sectional surfaces, which would enable on-line quality and process control during the production of densified wood products. Three machine learning approaches were trained and then evaluated on external test sets: partial least squares (PLS) regression, artificial feed-forward neural networks (ANNs), and convolutional neural networks (CNNs).

All three approaches provided what could be construed as decent predictions but behaved in markedly different ways when confronted with distinct inhomogeneities in the DPs, such as growth ring peaks. The best overall performance was shown by the CNN (MAPE = 9.24%) and PLS (MAPE = 9.49%) models, while the ANN only reached a MAPE of 11.83%, and did not perform consistently well.

The PLS approach was easy to implement, did not require particularly high-level hardware for training, and provided highly accurate predictions of the main density peak caused by the surface densification (MAPE < 10%). The undensified regions of the specimens were, however, largely treated as a homogenous material (i.e., no density variation due to early- and latewood) and increasing the ability to generalise to atypical new data came at the expense of the prediction accuracy, especially in the region of the main density peak. The ANN approach had a similar behaviour to the PLS regression, but performed less accurately and less reliably overall. The CNN approach showed the best overall performance, as there was almost no trade-off between the ability to generalise to new data and the prediction accuracy. In addition, it was the only approach that could model wide growth rings in the DPs, and perhaps more importantly, the region between the specimen surface and the main density peak. These characteristics were further improved by training the CNN model on the augmented Combined L dataset.

This study showed the strengths and weaknesses of different state-of-the-art machine learning algorithms and future studies could focus not only on predicting density profiles, but on properties that are directly relevant to the end-customer, such as different strengths properties. Naturally, the same approach could also be applied to undensified wood products, such as glulam or cross-laminated timber.

Funding Open access funding provided by Lulea University of Technology.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaria J, Fadhel MA, Al-Amidie M, Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8(1):1–74
Demir A, Cakiroglu EO, Aydin I (2021) Determination of CNC processing parameters for the best wood surface quality via artificial neural network. Wood Mater Sci Eng 1–8
Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929
Fang CH, Cloutier A, Blanchet P, Koubaa A (2012) Densification of wood veneers combined with oil-heat treatment. Part II: hygroscopicity and mechanical properties. BioResources 7(1):925–935
Gong M, Lamsan C, Li L (2010) Interactive effect of surface densification and post-heat-treatment on aspen wood. J Mater Process Technol 210(2):293–296
Gu IYH, Andersson H, Vicen R (2010) Wood defect classification based on image analysis and support vector machines. Wood Sci Technol 44(4):693–704
He K, Zhang X, Ren S, Sun J (2016) Deep residual leaning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition 770–778
Heinecke A, Ho J, Hwang WL (2020) Refinement and universal approximation via sparsely connected relu convolution nets. IEEE Signal Process Lett 27:1175–1179
Hu J, Song W, Zhang W, Zhao Y, Yilmaz A (2019) Deep learning for use in lumber classification tasks. Wood Sci Technol 53(2):505–517
Iliadis L, Mansfield SD, Avramidis S, El-Kassaby YA (2013) Predicting Douglas-fir wood density by artificial neural networks (ANN) based on progeny testing information. Holzforschung 67(7):771–777
Jaderberg M, Simonyan K, Zisserman A (2015) Spatial transformer networks. Adv Neural Inf Process Syst 28
Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: a tutorial. Computer 29(3):31–44
Jones D, Sandberg D, Goli G, Todaro L (2019) Wood modification in Europe—a state-of-the-art about processes, products and applications. Florence University Press, Florence
Kamke FA (2006) Densified radiata pine for structural composites. Maderas Ciencia y Tecnología 8(2):83–92
Karpathy A (2017) Software 2.0. Medium. https://karpathy.medium.com/software-2-0-a64152b37c35. Accessed 23 June 2021
Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, Proceedings of the First 12 Conferences 25:1097–1105
Kutnar A, Kamke FA, Sernek M (2008) The mechanical properties of densified VTC wood relevant for structural composites. Holz Roh Werkst 66(6):439–446
Kutnar A, Sandberg D, Haller P (2015) Compressed and moulded wood from processing to products. Holzforschung 69(7):885–897
Laine K (2014) Improving the properties of wood by surface densification. Dissertation, Aalto University
Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: A convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113
LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
Lienhart R, Maydt J (2002) An extended set of haar-like features for rapid object detection. In: Proceedings of the International Conference on Image Processing, IEEE
Neyes B, Karlsson O, Sandberg D (2020) The effect of ionic liquid and superbase pre-treatment on the spring-back, set-recovery and Brinell hardness of surface-densified Scots pine. Holzforschung 74(3):303–312
Neyes B (2019) Surface Densification of Solid Wood: Paving the Way Towards Industrial Implementation. Dissertation, Luleå University of Technology
Papageorgiou CP, Oren M, Poggio T (1998) A general framework for object detection. In: Sixth international conference on computer vision, IEEE
Pizzi A, Leban JM, Zanetti M, Pichelin F, Wieland S, Properzi M (2005) Surface finishes by mechanically induced wood surface fusion. Holz Roh Werkst 63(4):251–255
Rautkari L, Laine K, Laflin N, Hughes M (2011) Surface modification of Scots pine: the effect of process parameters on the through thickness density profile. J Mater Sci 46(14):4780–4786
Rautkari L, Laine K, Kutnar A, Medved S, Hughes M (2013) Hardness and density profile of surface densified and thermally modified Scots pine in relation to degree of densification. J Mater Sci 48(6):2370–2375
Rojas R (1996) The backpropagation algorithm. Neural networks. Springer, Berlin, Heidelberg, pp 149–182
Schmidhuber J (2015) Deep learning in neural networks: An overview. Neural Netw 61:85–117
Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
Tarkow H, Seborg R (1968) Surface densification of wood. For Prod J 18(9):104–107
Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58(2):109–130
Yang X, Yang J, Yan J, Zhang Y, Zhang T, et al (2019) Towards More Robust Detection for Small, Cluttered and Rotated Objects. arXiv preprint arXiv:1811.07126
Zhang ML, Zhou ZH (2013) A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 26(8):1819–1837
Zhou DX (2020) Universality of deep convolutional neural networks. Appl Comput Harmon Anal 48(2):787–794

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.