FIBER QUALITY PREDICTION USING NIR SPECTRAL DATA: TREE-BASED ENSEMBLE LEARNING VS DEEP NEURAL NETWORKS

*Vahid Nasir**

Postdoctoral Fellow
University of British Columbia
and
Courtesy Faculty
Department of Wood Science and Engineering
Oregon State University
E-mail: vahid.nasir@oregonstate.edu

Syed Danish Ali

Research Associate
USDA Forest Service
Forest Products Laboratory
Madison
and
Department of Biological Systems Engineering
University of Wisconsin-Madison
E-mail: syed.ali2@usda.gov

Ahmad Mohammadpanah

Assistant Professor
Department of Mechanical Engineering
Loyola Marymount University
E-mail: Ahmad.Mohammadpanah@lmu.edu

Sameen Raut

PhD Candidate
E-mail: sameen.raut@uga.edu

Mohamad Nabavi

Graduated Student
E-mail: m.nabavi84@gmail.com

Joseph Dahlen

Associate Professor
Warnell School of Forestry and Natural Resources
University of Georgia
E-mail: jdahlen@uga.edu

Laurence Schimleck

Professor
Department of Wood Science and Engineering
Oregon State University
E-mail: laurence.schimleck@oregonstate.edu

(Received March 2023)

* Corresponding author

Wood and Fiber Science, 55(1), 2023, pp. 100-115
https://doi.org/10.22382/wfs-2023-10

Abstract. The growing applications of near infrared (NIR) spectroscopy in wood quality control and monitoring necessitates focusing on data-driven methods to develop predictive models. Despite the advancements in analyzing NIR spectral data, literature on wood science and engineering has mainly utilized the classic model development methods, such as principal component analysis (PCA) regression or partial least squares (PLS) regression, with relatively limited studies conducted on evaluating machine learning (ML) models, and specifically, artificial neural networks (ANNs). This could potentially limit the performance of predictive models, specifically for some wood properties, such as tracheid width that are both time-consuming to measure and challenging to predict using spectral data. This study aims to enhance the prediction accuracy for tracheid width using deep neural networks and tree-based ensemble learning algorithms on a dataset consisting of 2018 samples and 692 features (NIR spectra wavelengths). Accordingly, NIR spectra were fed into multilayer perceptron (MLP), 1 dimensional-convolutional neural networks (1D-CNNs), random forest, TreeNet gradient-boosting, extreme gradient-boosting (XGBoost), and light gradient-boosting machine (LGBM). It was of interest to study the performance of the models with and without applying PCA to assess how effective they would perform when analyzing NIR spectra without employing dimensionality reduction on data. It was shown that gradient-boosting machines outperformed the ANNs regardless of the number of features (data dimension). All the models performed better without PCA. It is concluded that tree-based gradient-boosting machines could be effectively used for wood characterization utilizing a medium-sized NIR spectral dataset.

Keywords: Convolutional neural network (CNN), deep learning, ensemble learning, gradient-boosting (TreeNet), near infrared (NIR) spectroscopy, random forest, wood.

INTRODUCTION

Smart quality control (QC) and characterization of wood materials requires developing intelligent monitoring systems through combining nondestructive evaluation (NDE) methods with data-driven techniques. NDE tools play a crucial role in fast and reliable data acquisition when it comes to wood QC and properties monitoring. Common NDE methods applied to wood materials include color measurement, X-ray computed tomography (X-ray CT), thermography, wave propagation, and near infrared (NIR) spectroscopy, among which, wave propagation methods and NIR spectroscopy have been widely studied in the literature (Nasir et al 2022). NIR spectroscopy can be used to predict wood properties directly related to sample chemistry (Schimleck et al 2019) through analyzing absorption or reflectance in the NIR range. It has also been widely applied to predict wood density, MC, mechanical properties, such as modulus of elasticity and modulus of rupture (Tsuchikawa 2007; Tsuchikawa and Schwanninger 2013). Apart from predicting wood properties, NIR spectroscopy has been used for wood identification, wood classification, and QC of thermally modified wood (Tsuchikawa and Kobori 2015; Willems et al 2015).

Despite the growing opportunities that NIR spectroscopy offers for wood characterization and QC, there are still limitations associated with this technique. Apart from the need for a robust calibration model, spectral data acquisition in industrial settings imposes practical challenges because of noise related to temperature and humidity variations (Hein et al 2017). Specifically, there is a significant gap between the performance of NIR spectroscopy in laboratory vs industrial conditions, where the high performance achieved under laboratory-controlled conditions using multivariate data analysis may not necessarily account for the high variation in data acquired in real time in a manufacturing environment (Hein et al 2017).

NIR spectral data are typically processed using a pretreatment (bandwidth selection, smoothing, normalization, etc.), feature extraction and/or selection, and final decision-making, which mainly involves using classification or regression models. Most studies on the applications of NIR spectroscopy in wood materials have focused on dimensionality reduction for feature extraction through applying principal component analysis (PCA) regression, or partial least squares (PLS) regression (Sandak et al 2016). There is a gap in the literature about advanced data-driven

methods applied to NIR spectral data analysis. Feature engineering could be a crucial step toward designing a decision-making model with improved accuracy over classical techniques. The main feature selection techniques (filter, wrapper, and embedded methods) are reviewed in the literature (Chandrashekar and Sahin 2014), including an overview on the variable selection in multivariate analysis of NIR spectra (Yun et al 2019). PCA does use employ dimensionality reduction; however, feature extraction via PCA yields a new set of variables with no clear physical meaning (Mao 2005). An alternative approach could be feature selection in the paradigm of supervised learning. Examples include utilizing heuristic search methods, such as genetic algorithm for variable selection in predicting the wood properties, such as pulp yield using NIR spectral data (Ho et al 2022; Zhen et al 2022).

The performance of a data-driven predictive model depends on both the feature engineering (extraction or selection) technique and the choice of regression or classification model for final prediction. Moreover, it also significantly impacted by the size and complexity of the dataset (Nasir and Sassani 2021). Machine learning (ML), and specifically, artificial neural networks (ANNs), have been utilized to deal with datasets of higher complexity to better unveil hidden patterns within the data. Superior performance of ANNs over PLS regression is reported in the literature with a sample size in the range of 172-480 (Watanabe et al 2014; Costa et al 2019; Ayanleye et al 2021). Variable (feature) extraction was typically performed using PCA and the reduced datasets were then fed into ANNs (Costa et al 2019; Nasir et al 2019; Ayanleye et al 2021). While the reported accuracy could be impacted by the size and complexity of the data, it highlights the importance of employing ML for enhancing the prediction accuracy of NIR-based models. Yet, literature in wood science and engineering still lacks comparative studies on the predictive performance of models developed using different ML and feature engineering techniques.

While traditional ML models rely heavily on the knowledge of user(s) for feature extraction and selection, deep learning (DL) models exhibit built-in feature engineering characteristics (Miotto et al 2018; Wang et al 2018), enabling them to model complex nonlinear relationships in big data (Nasir and Sassani 2021) through extracting complex high-level abstractions as data representations (Najafabadi et al 2015). Convolutional neural networks (CNNs) are among the most commonly used types of DL models and have been combined with NIR hyperspectral imaging for wood species identification (Kanayama et al 2019). One-dimensional (1D) CNN has also been applied to softwood species classification (Yang et al 2020) and QC of Chinese zither panels (Huang et al 2019).

This study aims to perform a comparative study between the performance of some of the ML and DL models for prediction of the properties of wood-based materials. Since most published studies on data-driven methods applied to NIR spectral data deal with small sample sizes, it also aims to investigate applying ML and DL to a medium-sized dataset. The target property for prediction is tracheid width, which in addition to other tracheid morphological characteristics (diameter, length, and wall thickness) largely determine pulp fiber quality and paper performance. Evaluation typically occurs on macerated samples that are examined with microscope or optical imaging system; however, more rapid techniques are required for routine incorporation into tree improvement programs. NIR spectroscopy offers an alternative approach and fiber/tracheid properties have been estimated in both hardwoods (Inagaki et al 2012; Pereira et al 2015) and softwoods (Schimleck and Evans 2004; Via et al 2004; Nabavi et al 2018; Dahlen et al 2021). Fiber/tracheid length models have generally demonstrated strong performance. Similarly, models for wall thickness measured on solid samples have also performed well; however, models for tracheid diameter, whether it be measured by SilviScan on solid samples (Evans 1994) or on macerated samples, have been noticeably weaker. Therefore, the study investigates options for enhancing the prediction accuracy of tracheid

width. Two neural network (NN)-based models, including the multilayer perceptron (MLP) and 1D-CNN, are utilized. Their performances will be compared with those obtained from tree-based ensemble methods; random forest, TreeNet gradient-boosting, extreme gradient-boosting (XGBoost), and light gradient-boosting machine (LGBM). The study evaluates the impact of applying PCA and compares that with the case of developing models using the full range of NIR data (no dimensionality reduction). The objective is to improve prediction performance by designing models capable of handling medium-sized NIR spectral datasets while analyzing the full range of NIR features with embedded feature selection characteristics.

EXPERIMENTS AND DATA ACQUISITION

ML study was performed on samples utilized by Nabavi et al (2018), which represents one of the-largest spectral datasets on wood materials and was comprised of 1842 loblolly pine (*Pinus taeda* L.) radial strips from 225 trees and 99 stands. The samples were taken from the Wood Quality Consortium (WQC) baseline study (Jordan et al 2008; Antony et al 2010), which totaled 134 stands representing six physiographic regions (Gulf Coastal Plain, Hilly Coastal, North and South Atlantic Coastal Plains, Piedmont, and Upper Coastal Plain) across the Southeastern United States. At each stand, three trees were felled and 25-mm thick disks were cut at 1.5 m intervals from the base to a top diameter of 25 mm. A breast height (1.37 m) disk was also collected and four book-matched pith-to-bark radial strips (12 mm longitudinally × 12 mm tangentially) were cut from these disks. Jordan et al (2008) and Nabavi et al (2018) provide additional detail regarding the collection and preparation of these samples. The stands were largely sampled in the early 2000s and many of the breast height disks had been used in various WQC research projects that followed their collection, hence some samples were unavailable for study (Nabavi et al 2018). Several of these disks also had blue stain and were excluded from the study (Nabavi et al 2018). Following the sample selection procedure, the selected sections were cut from the radial strips into individual 10-mm sections using a razor blade. The sections were further cut into 2-mm sections using a razor blade. The samples were then macerated with 20 mL of 50% hydrogen peroxide, 30 mL of water, and 50 mL glacial acetic acid at 60°C for 48 h (Franklin 1945). After 48 h, the tracheids and chemicals were cooled to room temperature, then the tracheids were separated from the pulping chemicals using a Buchner funnel, rinsed with approximately 1.9 L of water, and then the acidic spent chemicals and the rinse water were neutralized with sodium carbonate. The rinsed tracheids were then diluted with approximately 1 L of water prior to analysis. The macerated samples were analyzed using a TechPap MorFi Compact Fiber & Shive Analyzer with a 4 µm resolution camera (Techpap SAS, France) to assess tracheid properties. The equipment measured the number of tracheids, frequency of tracheids, tracheid length, and width using image analysis. For each sample, approximately 2500 tracheids were analyzed. The width-weighted tracheid width was used for the calculations instead of the mean value to minimize the effect of fines (Carvalho et al 1997). Width-weighted tracheid width is calculated by:

$$W_l = \frac{\sum_{n_i w_i^2}}{\sum_{n_i w_i}} \quad (1)$$

where W_l is width-weighted tracheid length, n_i is number of tracheids in the ith class, and w_i is the mean width of the ith class (Carvalho et al 1997). Prior to the collection of NIR spectroscopic data, each radial sample was marked from pith to bark into 10-mm sections on the transverse face. Subsequently, NIR diffuse reflectance spectra were collected from the radial–longitudinal surface of each pith-bark strip for the same 10-mm sections using a FOSS NIRSystems Model 5000 scanning spectrophotometer (FOSS NIRSystems, Inc., Laurel, MD) fitted with a diffuse reflectance static module. A custom-made autosampler consisting of a Parker servo motor and controller, and a linear stage provided precise sample motion and a Teflon mask with a window 5-mm high × 10-mm wide to limit the area scanned (Jones et al

2007). NIR spectra were collected at 2 nm resolution over the wavelength range 1100-2500 nm. The instrument reference was a ceramic standard provided by the manufacturer. A total of 11,428 NIR spectra were collected from the 1842 samples.

Data Analysis

The preliminary analysis indicated the effectiveness of applying second derivative to the NIR spectra with left and right gaps of 4 nm. Also, the data were smoothed using the Savitzky–Golay approach (Savitzky and Golay 1964). All models in this study were fit on the data after pretreatment. A total of 70% of the data ($N = 1404$) were used for model training while the remaining 30% ($N = 614$) were used to test the model. The test dataset was not used in the model development until the final model was selected and thus, maintained its independence.

Models here were fit using two approaches. The first approach extracted features by first applying PCA to reduce data dimensionality. The reduced dataset was then fed into the predictive ML models. The second approach fit the ML models using the full spectral data, and the two methods compared. Six different types of deep neural networks (DNNs) and tree-based ensemble algorithms were used in this stud. Python and Minitab statistical software were used in this study for the data analysis.

Deep neural networks. MLP known as a universal approximator (Hornik et al 1989) is one of the most widely used NNs for regression and classification. The model consists of input, hidden, and output layers (Fig 1). The number of hidden layers and neurons in each layer should be defined by the user or tuned using an optimization technique for hyperparameter tuning. Having several hidden layers could result in a deep model with higher complexity. MLP NN with a feedforward architecture is among the most basic types of deep models with a series of feedforward fully connected layers. The error backpropagation was used by applying the "Adam" optimization algorithm (with default parameter settings), which is a stochastic gradient descent optimization method

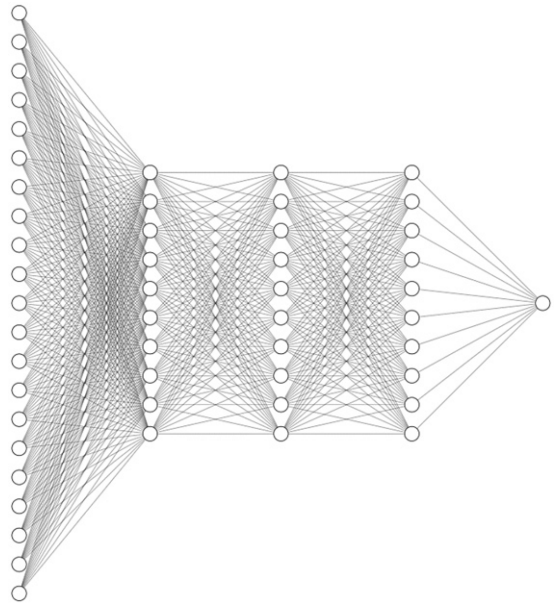

Figure 1. A schematic representation of the multilayer perceptron model used in this study with three hidden layers each having 10 neurons. The feed forward model with error backpropagation for hyperparameters tuning is among the most widely used ANNs. The shown sample network has 21 input parameters representing the case, in which PCA is applied to the dataset.

(Kingma and Ba 2014) to adjust the weight of neurons in the training process. The architecture of MLP (number of hidden layers and neurons) is shown to significantly impact the performance of the model. Three hidden layers were used in the MLP each having 10 neurons. The first hidden layer used "Swish" activation function while the remaining two used rectified linear unit ("ReLU"). Model training was done with a batch size of 32 and 25% of the training data were separately used as validation data to evaluate loss and model metrics at the end of each epoch. L2 regularization with a value of 1 e-3 was employed to prevent overfitting. Early stopping method was used to monitor the validation loss, which stopped model training if the validation loss did not improve over 40 epochs. Thus, the model training was stopped after 425 epochs on the dataset without applying PCA and after 521 epochs on the dataset reduced using PCA.

The performance of MLP NN was compared with that obtained from CNN. CNN is an effective ML algorithm that requires significantly fewer parameters than traditional DNNs. This is due to the convolution operation, sparse connectivity, and weight sharing in the network, resulting in much fewer preprocessing operations compared with other neural networks (Liu 2018). CNNs are a subset of DL models, which is a class of ML that uses multilayered ANNs to deliver state-of-the-art accuracy in tasks, such as object detection, speech recognition, language translation, image classifications, and so on (Alzubaidi et al 2021). 1D-CNNs are similar to 2-D CNNs, but are used mainly on 1D signals. In 1D-CNN, the convolutional kernel/filter moves in just one direction to calculate the output, and the output is a 1D signal. Modeling using 1D-CNN performs convolution operation on data and extracts significant features from the raw input data. Convolution involves sliding the kernel over the input signal, which is also known as a shift-compute procedure. The architecture of a typical CNN model is shown in Fig 2. The architecture of the 1D-CNN model consists of 72 filters, each with a size of 9, and rectified linear unit (ReLU) as an activation function for the convolution layer. The convolution layer's output is then flattened, followed by two dense layers having ReLU as an activation function with 64 and 32 neurons, respectively. To prevent overfitting, the first dense layer employs L2 regularization with a value of 1e-5. The model was constructed using the Adam optimizer with a learning rate of 0.0005 and a batch size of 27. During the training of the model, an early stopping technique is used to prevent overfitting by terminating the training if there is no improvement in validation loss for 30 epochs.

Tree-based ensemble learning. Ensemble learning is an approach to ML, which aims to enhance predictive performance through combining the prediction of multiple models that could be considered as weak learners. While it can work for a variety of algorithms, decision trees are commonly used in this method. The intuitive nature of decision trees make them easy for interpretation. Decision trees were used for predicting the mechanical properties and check formation in weathered timber (Nasir et al 2021a, c; van Blokland et al 2021a, b) using the classification and regression trees (CART) algorithm (Steinberg and Colla 2009). Decision trees have some advantages, such as using a white-box model that is easy to understand, require little preprocessing of data, and can handle both numerical and categorical data as well as missing data. However, they are prone to overfitting by creating very complex trees, which suffer from the generalization issue. This highlights the importance of the pruning phase

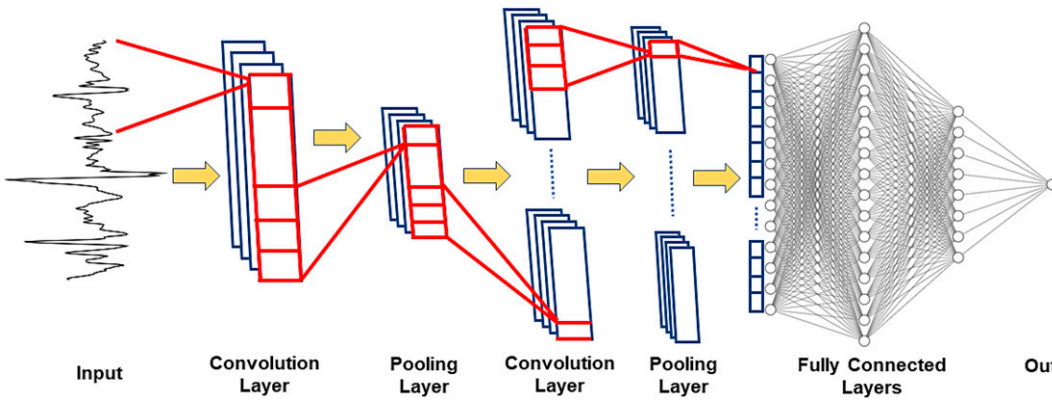

Figure 2. A sample representation of a typical 1D-CNN showing the convolution and pooling layers. The down-sampled data are fed into a fully connected layer(s) for predicting the output. The number of convolution and pooling layers along with its hyperparameters should be defined by the user or optimized.

and imposing conditions, such as defining the maximum number of terminal nodes or tree depth. Also, decision trees may become unstable due to variation in the data. That said, decision trees can benefit from ensemble learning algorithms. Two of the most common ensemble algorithms, bagging and boosting, were used in this study. Both methods had embedded feature selection characteristics that enables ranking of the relative importance of the features (wavelengths, in this study) through considering the contribution of the nodes having that feature when finding an optimal decision (Nasir et al 2021b).

Employing the bagging approach, random forest uses a subset of data to train multiple decision trees and aggregates the decisions of the trees known as weak learners through a voting scheme for the final prediction (Liaw and Wiener 2002). Random forest has been applied to the estimation of mechanical properties of wood fiber insulation boards (Schubert et al 2020). It has also used in wood machining for monitoring tool wear classification (Nasir et al 2021d) and identifying frozen and green lumber during circular sawing of logs (Nasir et al 2020). The random forest in this study contained 3000 trees and 80% of the training dataset was randomly chosen to identify bootstrap sample size. Additionally, a minimum number of three cases was specified to split an internal node. Finally, the number of predictors for node splitting was set to be the root mean root square of the total number of predictors. Figure 3 shows a schematic representation of a random forest model.

While the training process of weak learners in the bagging method is independent and in parallel, it is sequentially in the boosting method as the performance of a prior model is modified in the subsequent one (Fig 4). This approach has been used for wood species identification (Sun et al 2021), tool wear classification during wood machining (Nasir et al 2021d), predicting the properties of wood composites (Carty et al 2015), and developing an online color classification model for solid wood flooring (Zhuang et al 2021). In the TreeNet gradient-boosting machine, a random portion of the training data are utilized to calibrate a CART model with the maximum number of terminal

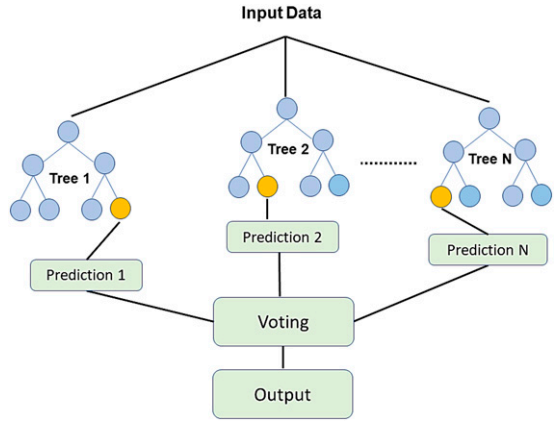

Figure 3. Schematic representation of a random forest model showing multiple decision trees trained separately using a random subset of data. Then a voting scheme is used for the final prediction.

nodes or the depth of the tree predefined. The maximum number of terminal nodes per tree was set to 12 and the minimum number of cases allowed for a terminal node was set to six. The subsample fraction in this study was set to 0.25. The CART model should then update based on the loss function. However, the update shrinks by a learning rate, which was set to 0.1 following a trial and error process during preliminary analysis. Subsequently, additional CARTs are added to improve the error for a specified number of iterations that defines the number of trees in the model (Modeler 2019) (set to 5000 in this study). Finally, the number of predictors for node splitting was set to be the root mean root square of the total number of predictors.

XGBoost is a scalable and distributed gradient-boosting algorithm for various ML problems (Chen and Guestrin 2016). It uses decision trees as base models and trains them in a stagewise manner, creating a strong composite model by combining multiple weak models. XGBoost is designed to improve the model generalization and handle sparse data efficiently. In this optimal model, the number of estimators was set to 1455, the maximum depth of the trees in the model was set to 5, the learning rate was defined as 0.027, the subsample rate was equal to 0.33, and

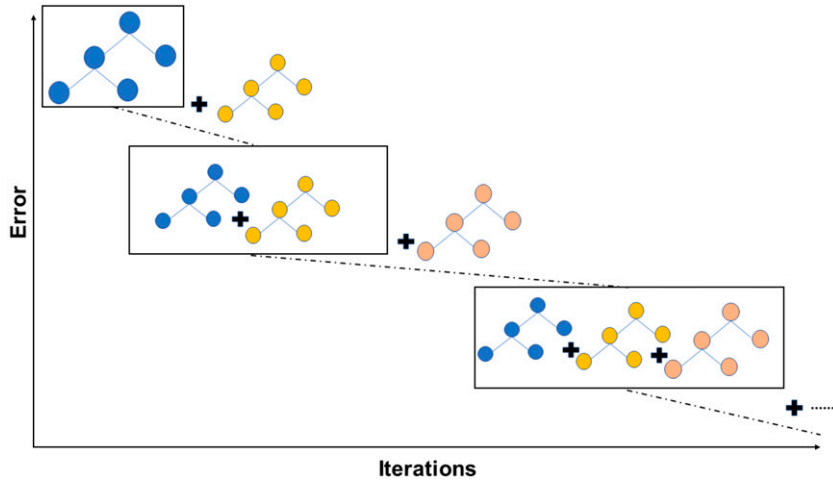

Figure 4. In TreeNet model, the gradient-boosting algorithm performs the training process sequentially, in which subsequent models correct the performance of prior models. The process starts with training a classification and regression tree (CART) model. It is then updated and CART models are sequentially added for a specified number of iterations lowering the error.

colsample_bytree and gamma were also set to 0.25 and 0.50, respectively. The colsample_bytree is set to 0.25 determines 25% of the columns are randomly chosen for each tree, and a gamma value of 0.5 means that a node will only be split if it results in a reduction of the loss function greater than or equal to 0.5, which potentially prevent overfitting and enhance the model's generalization ability.

LGBM is also an efficient, scalable, and optimized tree-based learning algorithm using a histogram-based decision tree learning algorithm that reduces communication overhead and minimizes memory usage, resulting in faster training times and improved performance on large datasets (Ke et al 2017). Its strong predictive power and efficiency have made it popular in various applied ML tasks. In the optimal model, the bagging_fraction and bagging_freq, which respectively determine the subsample ratio and frequency of bagging during training, were set to 0.42 and 0.53, respectively. It indicates that 42% of the data are sampled for each tree, and bagging is applied every 53 iterations in the training process. Also, the L1 and L2 regularization terms on the weights (by adding a penalty term to the loss function to reduce the magnitude of the weights and prevent overfitting) were 0.00028 and 0.043, respectively. The learning rate was set to 0.011, and no restriction was put on the maximum depth of the tree allowing the model to learn complex interactions between features until all leaves are pure or until the num_leaves parameter is reached; for instance, in this case, the number of leaves is set to 95. Finally, the number of estimators was equal to 1600. In this study, the optimal hyperparameters for XGBoost and LGBM were selected using the Python API Optuna (Akiba et al 2019). The Tree-Structured Parzen Estimator (TPE) sampler in Optuna was used, which is a Bayesian optimization-based sampler that uses a probabilistic model to guide the search process of hyperparameters. The TPE uses a tree-structured representation of the search space to model the probability distribution of the target function and generates new samples in regions likely to contain optimal solutions.

RESULTS AND DISCUSSION

Figure 5 shows the distribution for tracheid width for the 2018 samples measured. The histogram shows that the tracheid width of the samples had an average of 40.67 μm with a standard deviation of 2.75 μm. The normality test using the Ryan–Joiner method yielded a p-value < 0.01

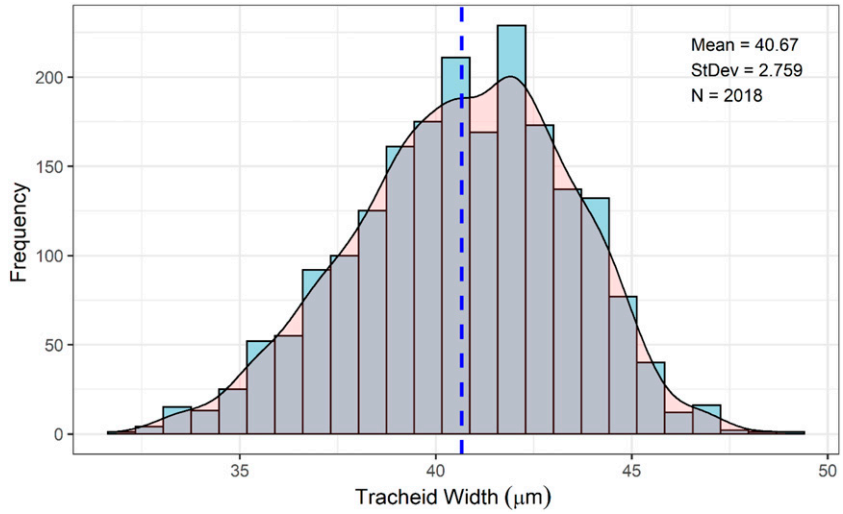

Figure 5. Histogram showing the distribution of tracheid width.

rejecting the null hypothesis of normal distribution. It can be seen that the measured tracheid width varies in the range of 31-49 μm. Table 1 summarizes tracheid width descriptive statistics for the training and test data. The average and standard deviation of the tracheid width in both the training and test data are very similar to that of the total dataset indicating appropriate data splitting.

Prior to any analysis, the input dataset was standardized by rescaling the data to show an average of 0 with standard deviation of 1. Subsequently, PCA was performed on the input NIR data for dimensionality reduction. Figure 6 indicates the contribution of PCs to the data and reveals that using 21 PCs can explain 95% of variation in the data. This reduced the size of the input dataset to a matrix of 2018×21. Table 2 shows the summary of the performance (mean square error [MSE] and coefficient of determination [R^2]) obtained from the deep and ensemble models.

The lowest R^2 (test data) was obtained using random forest model followed by the MLP NN. 1D-CNN showed a slightly higher R^2 than MLP NN. Compared with the DNN models used in this study, the gradient-boosting machines (TreeNet, XGBoost, and LGBM) showed superior performance resulting in higher R^2 and lower MSE. The lowest error and highest R^2 was obtained from the LGBM model followed by TreeNet. Figure 7 shows the R^2 variation with respect to the number of trees in the TreeNet model. Although the model was expanded to comprise 5000 trees, the optimal performance (considering overfitting) was chosen when having 2159 trees in the model. Compared with DNNs, the gradient-boosting machine algorithms show overfitting on the training data, yet, yielded the best performance on the test data.

The relatively higher performance of LGBM model compared with other boosting methods, such as XGBoost, could be linked to its tree growth strategy, which is leaf-wise (it is level-wise for XGBoost) resulting in growing the tree in a selective manner, leading to smaller and faster models as compared with XGBoost. It is

Table 1. Descriptive statistics of the tracheid width for the training and test data.

Dataset	*N*	% of *N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Training	1404	69.57	40.65	2.71	32.59	38.85	40.80	42.60	48.92
Test	614	30.43	40.71	2.86	31.86	38.81	40.82	42.77	47.58

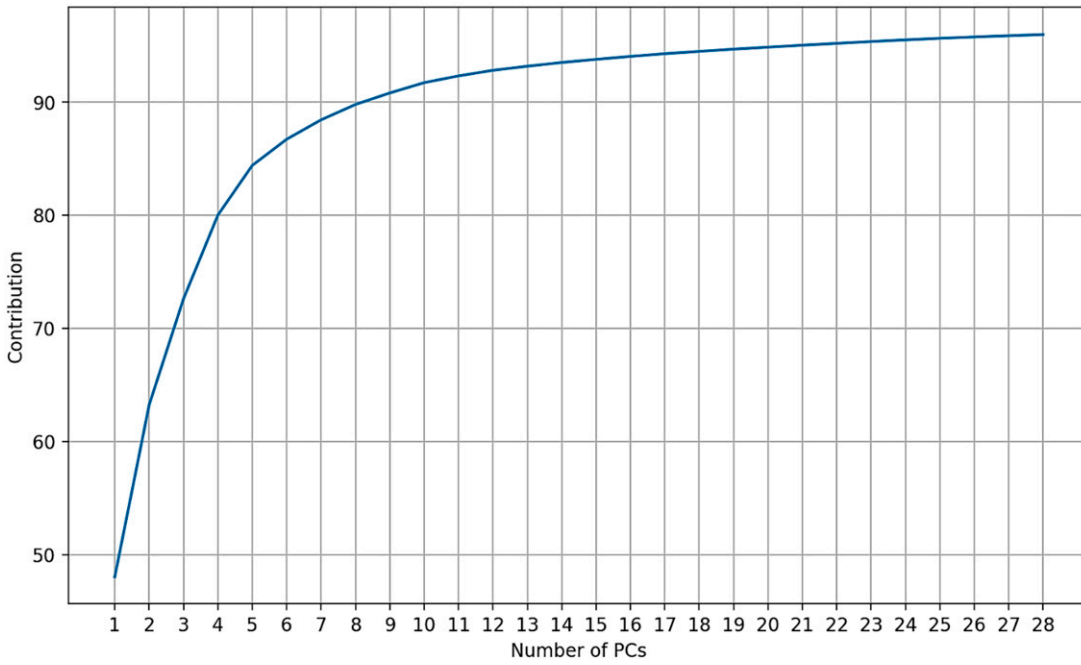

Figure 6. Contribution of PCs to the data.

observed that LGBM performs slightly better than XGBoost, which could be attributed to either a better hyperparameter tuning or a better model fitting. As the hyperparameters of both LGBM and XGBoost are different, it could be that the optimal hyperparameters for LGBM are more suitable for the given dataset. Additionally, the LGBM model is a better fit for the data and was able to capture more complex relationships within it.

Table 2. Performance summary of the deep and ensemble models for prediction (PCA is applied).

	Performance (R^2)			
	MSE		R^2	
Models	Train	Test	Train	Test
MLP NN	2.88	3.17	0.61	0.61
1D-CNN	2.53	3.45	0.63	0.62
Random Forest	3.05[a]	3.39	0.58[a]	0.58
TreeNet	0.54	2.94	0.93	0.64
XGBoost	0.11	2.99	0.99	0.63
LGBM	0.27	2.86	0.96	0.65

[a]Results indicating the out-of-bag performance.

The spectral data were also fed into the ML models without applying PCA. Table 3 summarizes MSE and R^2 obtained from the different models when full-range spectral data were directly used for prediction. The lowest R^2 and highest MSE (test data) corresponded to MLP NN. Random forest and 1D-CNN revealed similar performance while both slightly outperformed the MLP NN. Similar to Table 2, it can be seen that gradient-boosting algorithms outperformed the NNs and random forest. The highest R^2 of 0.69 was achieved using the LGBM followed by TreeNet and XGBoost, which showed similar performance ($R^2 = 0.68$). Figure 8 shows the variation in R^2 vs the number of trees in the TreeNet model. Optimal performance considering R^2 of 0.68 and overfitting was obtained using a model with 1959 trees. It can be seen that feeding the full range of NIR data into the models without applying PCA enhances the performance of the models.

All of the ML models fit here had improved fit statistics compared with the R^2 of 0.61 (test data) achieved by Nabavi et al (2018) for tracheid width

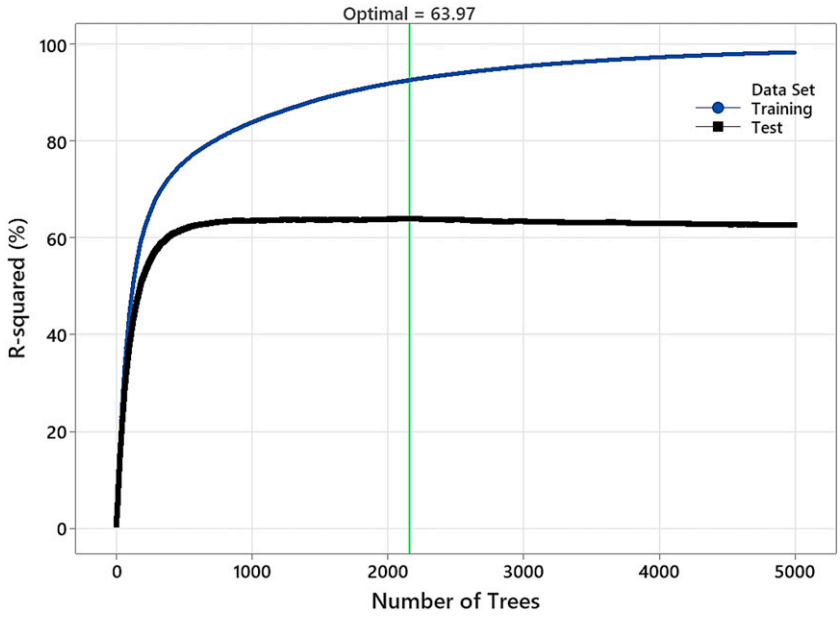

Figure 7. Variation in the R^2 vs the number of trees in the TreeNet model (PCA was applied to the input data).

prediction using PLS regression. The slightly higher performance of CNN compared with MLP could be associated with the high-level data abstraction capability of CNN. While MLP could be considered a DNN model having three hidden layers in its structure, it is one of the most basic types of NNs. Representation of raw data are learned in a DL model, such as CNN, through multiple levels of abstraction (LeCun et al 2015; Miotto et al 2018). Thus, one can conclude that high-level data abstraction and the built-in feature engineering capability of CNN during its learning process results in superior performance compared with a more traditional model, such as MLP NN.

Table 3. Performance summary of the deep and ensemble models for prediction (PCA is not applied and the entire spectral data were directly fed into the models).

	Performance			
	MSE		R^2	
Models	Train	Test	Train	Test
MLP NN	2.73	3.03	0.63	0.63
1D-CNN	1.73	2.95	0.77	0.64
Random Forest	2.81[a]	2.92	0.61[a]	0.64
TreeNet	0.36	2.64	0.95	0.68
XGBoost	0.16	2.57	0.99	0.68
LGBM	0.09	2.56	0.99	0.69

[a]Results indicating the out-of-bag performance.

The higher performance of gradient-boosting algorithms over NNs is aligned with the benchmark study of Grinsztajn et al (2022), in which tree-based models, such as XGBoost, outperformed the NNs on medium-sized data (approximately 10 K samples). The competitive performance of tree-based models even when dealing with an irregular pattern in the target function and uninformative features makes them powerful tools for regression and classification tasks on tabular datasets, while MLP-like architectures are noticeably affected by uninformative features (Grinsztajn et al 2022). This could explain the lower performance of MLP in this study. The superior performance of gradient-boosting model over random forest was also reported when classifying tool wear during wood machining (Nasir et al 2021d). However, further studies on datasets with different sample size and complexity is needed to benchmark this observation.

The embedded feature selection nature of tree-based ensemble learning models resulted in

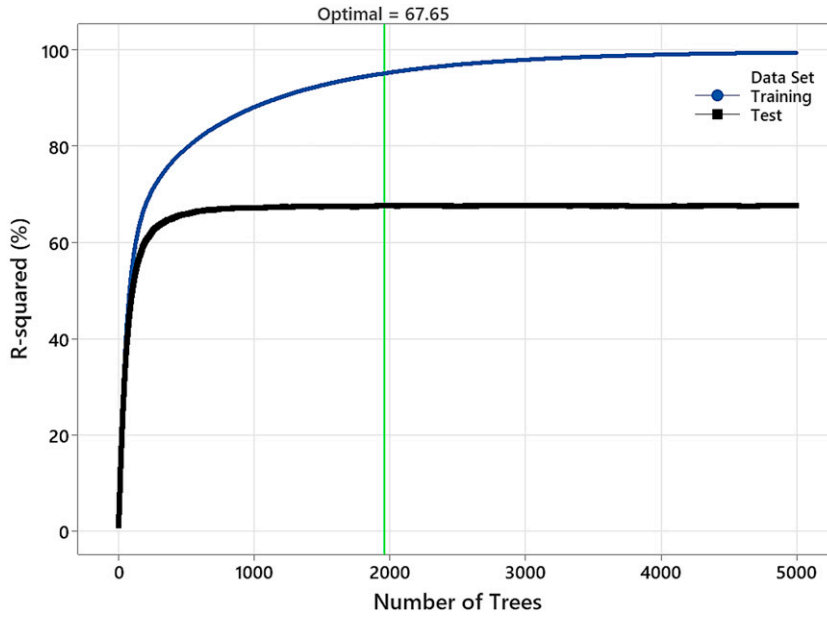

Figure 8. Variation in the R^2 vs the number of trees in the TreeNet model (PCA was not applied to the input data).

identifying the relative importance of features (wavelengths). For example, the results showed that all 692 wavelengths were important in the TreeNet model, while the random forest model had 516 important predictors. The top 10 wavelengths for the TreeNet and random forest models are summarized in Figs 9 and 10. In both, the relative importance of features are shown, where the most important was assigned a score of 100, with the remaining scaled accordingly.

Band assignments reported by Schwanninger et al (2011) for wood components were used to determine, where possible, the origins of the features shown in Figs 9 and 10. Four wavelengths identified for the TreeNet model (including the top three) were in the range 1540-1546 nm and this region is associated with first overtone O-H stretch vibrations in cellulose. Wavelengths at 1542 and 1544 nm were also identified as the top two important features in LGBM and XGBoost

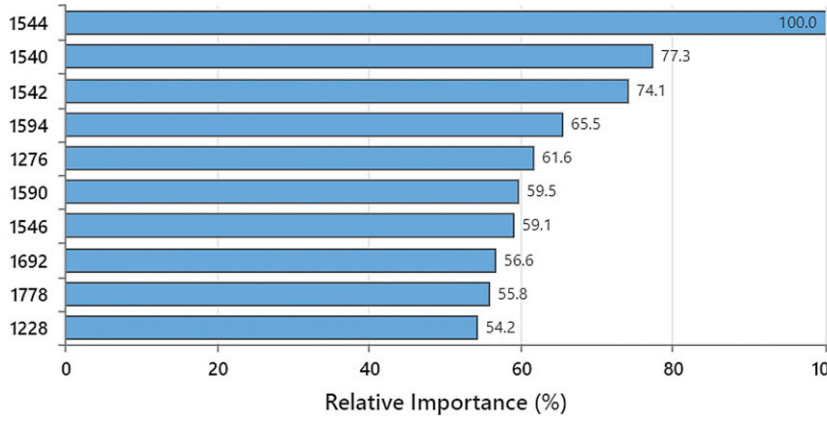

Figure 9. Top 10 important wavelengths contributing to the obtained predictive TreeNet model.

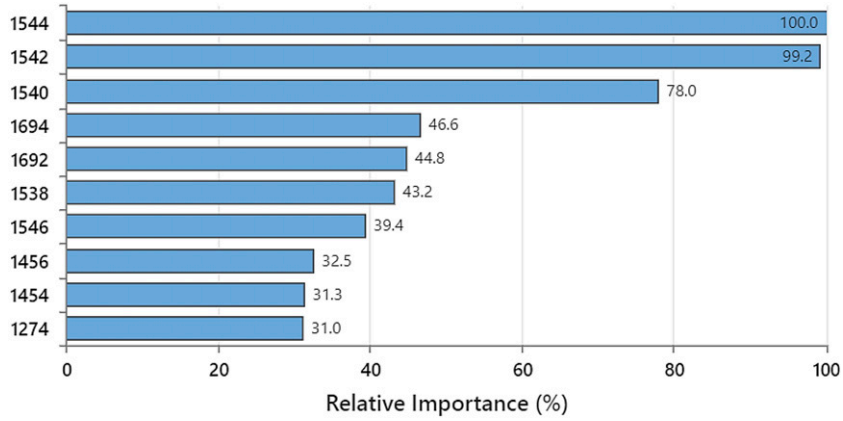

Figure 10. Top 10 important wavelengths contributing to the obtained predictive random forest model.

models. Wavelengths at 1590 and 1594 nm were also among the most important and bond vibrations in this region (1579-1597 nm) share the same origin. Wavelengths in both regions have been assigned to specific bond types in cellulose (Schwanninger et al 2011), which are summarized in Table 4.

Of the remaining four wavelengths identified by the TreeNet model, two (1692 and 1778 nm) occurred in regions associated with first overtone C-H stretch vibrations in lignin (1685 and 1698 nm) and cellulose (eg 1780, 1788, and 1790 nm). Similarly, 1228 nm is in a region (1212-1225 nm) assigned to second overtone C-H stretch vibrations in cellulose. The only wavelength without a recognized band assignment arising from a specific wood component was 1276 nm.

Table 4. The specific bond types in cellulose for different wavelengths.

Wavelength (nm)	Specific cellulose
1540	Adsorbed water (strong H-bond with microcrystalline cellulose)
1545	Intramolecular H-bond in cellulose
1548	Crystalline regions in C1. O(3)-H(3)$\cdots\cdots$O(5) intrachain H-bond
1588	Strong O(2)-H(2)$\cdots\cdots$O(6) of cellulose
1591	H-bonds of the cellulose 1_3 phase
1592	Crystalline regions in cellulose
1597	Strongly H-bonded O-H group in cellulose 1_α

The Random Forest model shared many frequently used features with models developed using TreeNet model and wavelengths (1544, 1542, and 1540 nm) associated with first overtone O-H stretch vibrations in cellulose were again the most important. The 1546 nm was also identified (ranked seventh for both models), but its relative importance was lower (39.4% vs 59.1% for TreeNet). Further, an additional wavelength (1538 nm) from this region was identified among the top 10 features. First overtone C-H stretch vibrations in lignin (1685 and 1698 nm) were again important, with two wavelengths (1690 and 1692 nm) identified (vs 1692 nm for TreeNet). Two additional wavelengths (1454 and 1456 nm) were important and occur close to 1447 and 1448 nm, wavelength which arise from first overtone O-H stretch vibrations in lignin. The only wavelength lacking a specific assignment was 1274 nm. The presence of recognized bond vibrations for the majority of important features arising from specific wood components indicate the importance of cellulose and lignin in the development of our tracheid width models. In general, the wavelengths identified for the model types were very similar; however, the random forest model appears to have greater emphasis on lignin owing to the presence of 1454 and 1456 nm and the corresponding absence of 1590 and 1594 nm (the cellulose-related wavelengths important for the TreeNet model).

Overall, methods such as tree-based ensemble models based on full-range NIR data could be

applied for property assessment of a variety of wood products. Not only may this enhance predictive performance of the developed model, but it provides an opportunity for NIR feature ranking and studying the relative importance of wavelengths in contributing to the performance of the predictive model. For example, this study highlighted the importance of wavelengths related to cellulose and lignin in developing a model for predicting tracheid width. Future studies should focus on the role of dataset size when employing different ML and DL models. Further expanding dataset size through employing synthetic data augmentation methods, such as generative adversarial network (GAN), could be an important topic of future studies.

CONCLUSIONS

Overall, the tree-based gradient-boosting machines outperformed the neural network models. LGBM yielded the highest R^2 and lowest error followed by XGBoost and TreeNet, whereas using MLP NN did not result in high performance. All models performed better without PCA when full-range NIR spectra were directly fed into them. The embedded feature selection characteristic of the tree-based ensemble models could be used to study the relative importance of features indicating the importance of cellulose and lignin in the development of tracheid width models. The developed models improved the reported results for tracheid width prediction using a PLS model, demonstrating the potential of DL, and specifically tree-based ensemble models, when applied to a medium-sized NIR dataset. The methodology could be applied to any wood product for which NIR spectral data can be collected and provides better predictive performance and improved insight into the relative importance of individual NIR wavelengths. Since the performance could be impacted by the size of data and its complexity, the proposed models should be tested on other NIR datasets, specifically those of small sample size, which is more typical when collecting data under laboratory conditions. Finally, the models could also be tested on big data; however, this requires large-scale data acquisition to better simulate industrial conditions and account for the high-variability of real-life situations. Future research could study employing synthetic data augmentation methods for increasing the size of datasets and how it impacts the performance of predictive models based on the NIR data spectral.

ACKNOWLEDGMENTS

This research was possible through support from the Wood Quality Consortium (WQC) at the University of Georgia, the National Science Foundation (NSF), Center for Advanced Forestry Systems (CAFS), and the NIFA McIntire-Stennis project 1023340. The authors wish to thank the WQC, CAFS, and NIFA for funding this project.

This work is partially supported by a cooperative agreement, 22-CO-11111137-017 between the US Department of Agriculture Forest Service Forest Products Laboratory and the University of Georgia through the US Department of Energy Award Number: DE-EE0008911 along with the University of Wisconsin.

REFERENCES

Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: A next-generation hyperparameter optimization framework. Pages 2623-2631 *in* Proceedings, 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 4-8, 2019, Anchorage, AK. Association for Computing Machinery, New York, NY.

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Farhan L (2021) Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data 8(1):1-74.

Ayanleye S, Nasir V, Avramidis S, Cool J (2021) Effect of wood surface roughness on prediction of structural timber properties by infrared spectroscopy using ANFIS, ANN and PLS regression. Eur J Wood Wood Prod 79:101-115.

Antony F, Schimleck LR, Daniels RF, Clark A III, Hall DB (2010) Modeling the longitudinal variation in wood specific gravity of planted loblolly pine (*Pinus taeda*) in the United States. Can J For Res 40(12):2439-2451.

Carty DM, Young TM, Zaretzki RL, Guess FM, Petutschnigg A (2015) Predicting and correlating the strength properties of wood composite process parameters by use of boosted regression tree models. For Prod J 65(7-8):365-371.

Carvalho MG, Ferreira PJ, Martins AA, Figueiredo MM (1997) A comparative study of two automated techniques for measuring fiber length. Tappi J 80(2):137-142.

Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16-28.

Chen T, Guestrin C (2016). XGBoost: A scalable tree boosting system. Pages 785-794 *in* Proceedings, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY.

Costa LR, Tonoli GHD, Milagres, FR, Hein PRG (2019) Artificial neural network and partial least square regressions for rapid estimation of cellulose pulp dryness based on near infrared spectroscopic data. Carbohydr Polym 224:115186.

Dahlen J, Nabavi M, Auty D, Schimleck L, Eberhardt TL (2021) Models for predicting the within-tree and regional variation of tracheid length and width for plantation loblolly pine. Forestry 94:127-140.

Evans R (1994) Rapid measurement of the transverse dimensions of tracheids in radial wood sections from *Pinus radiata.* Holzforschung 48(2):168-172.

Franklin G (1945) Preparation of thin sections of synthetic resins and wood-resin composites, and a new macerating method for wood. Nature 155(3924):51-55.

Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on tabular data? arXiv preprint arXiv:2207.08815.

Hein PR, Pakkanen H, Santos AAD (2017) Challenges in the use of near infrared spectroscopy for improving wood quality: A review. For Syst 26(3):eR03.

Ho TX, Schimleck LR, Dahlen J, Sinha A (2022) Utilization of genetic algorithms to optimize loblolly pine wood property models based on NIR spectra and SilviScan data. Wood Sci Technol 56(5):1419-1437.

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2(5):359-366.

Huang Y, Meng S, Zhao P, Li C (2019) Wood quality of Chinese zither panel based on convolutional neural network and near-infrared spectroscopy. Appl Opt 58(18): 5122-5127.

Inagaki T, Schwanninger M, Kato R, Kurata Y, Thanapase W, Puthson P, Tsuchikawa S (2012) *Eucalyptus camaldulensis* density and fiber length estimated by near-infrared spectroscopy. Wood Sci Technol 46(1-3):143-155.

Jones PD, Schimleck LR, So CL, Clark A, Daniels RF (2007) High resolution scanning of radial strips cut from increment cores by near infrared spectroscopy. IAWA J 28(4): 473-484.

Jordan L, Clark A, Schimleck LR, Hall DB, Daniels RF (2008) Regional variation in wood specific gravity of planted loblolly pine in the United States. Can J For Res 38(4):698-710.

Kanayama H, Ma T, Tsuchikawa S, Inagaki T (2019) Cognitive spectroscopy for wood species identification: Near infrared hyperspectral imaging combined with convolutional neural networks. Analyst (Lond) 144(21): 6438-6446.

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Liu TY (2017) LightGBM: A highly efficient gradient boosting decision tree. Pages 3146-3154 *in* I Guyon, UV Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, eds. Advances in neural information processing systems. Curran Associates, New York, NY.

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436-444.

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2(3):18-22.

Liu YH (2018) Feature extraction and image recognition with convolutional neural networks. J Phys Conf Ser 1087(6):062032. IOP Publishing.

Mao KZ (2005) Identifying critical variables of principal components for unsupervised feature selection. IEEE Trans Syst Man Cybern B Cybern 35(2):339-344.

Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2018) Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform 19(6):1236-1246.

Nabavi M, Dahlen J, Schimleck L, Eberhardt TL, Montes C (2018) A regional calibration model for predicting loblolly pine tracheid properties using near-infrared spectroscopy. Wood Sci Technol 52(2):445-463.

Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. J Big Data 2(1):1-21.

Nasir V, Ayanleye S, Kazemirad S, Sassani F, Adamopoulos S (2022) Acoustic emission monitoring of wood materials and timber structures: A critical review. Constr Build Mater 350:128877.

Nasir V, Dibaji S, Alaswad K, Cool J (2021a) Tool wear monitoring by ensemble learning and sensor fusion using power, sound, vibration, and AE signals. Manuf Lett 30:32-38.

Nasir V, Fathi H, Fallah A, Kazemirad S, Sassani F, Antov P (2021b) Prediction of mechanical properties of artificially weathered wood by color change and machine learning. Materials (Basel) 14(21):6314.

Nasir V, Fathi H, Kazemirad S (2021c) Combined machine learning–wave propagation approach for monitoring timber mechanical properties under UV aging. Struct Health Monit 20(4):2035-2053.

Nasir V, Kooshkbaghi M, Cool J (2020) Sensor fusion and random forest modeling for identifying frozen and green wood during lumber manufacturing. Manuf Lett 26:53-58.

Nasir V, Kooshkbaghi M, Cool J, Sassani F (2021d) Cutting tool temperature monitoring in circular sawing: Measurement and multi-sensor feature fusion-based prediction. Int J Adv Manuf Technol 112(9):2413-2424.

Nasir V, Nourian S, Zhou Z, Rahimi S, Avramidis S, Cool J (2019) Classification and characterization of thermally modified timber using visible and near-infrared spectroscopy and artificial neural networks: A comparative study on the performance of different NDE methods and ANNs. Wood Sci Technol 53(5):1093-1109.

Nasir V, Sassani F (2021) A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. Int J Adv Manuf Technol 115(9): 2683-2709.

Pereira H, Santos AJA, Anjos O (2015) Fibre morphological characteristics of kraft pulps of *Acacia melanoxylon* estimated by NIR-PLS-R models. Materials (Basel) 9(1):8.

Salford Predictive Modeler by Minitab (2019) Introducing TreeNet gradient boosting machine. Stage College, PA.

Sandak J, Sandak A, Meder R (2016) Assessing trees, wood and derived products with near infrared spectroscopy: Hints and tips. J Near Infrared Spectrosc 24(6): 485-505.

Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36(8):1627-1639.

Schimleck L, Dahlen J, Apiolaza LA, Downes G, Emms G, Evans R, Wang X (2019) Non-destructive evaluation techniques and what they tell us about wood property variation. Forests 10(9):728.

Schimleck LR, Evans R (2004) Estimation of *P. radiata* D. Don tracheid morphological characteristics by near infrared spectroscopy. Holzforschung 58(1):66-73.

Schubert M, Luković M, Christen, H (2020) Prediction of mechanical properties of wood fiber insulation boards as a function of machine and process parameters by random forest. Wood Sci Technol 54(3):703-713.

Schwanninger M, Rodrigues JC, Fackler K (2011) A review of band assignments in near infrared spectra of wood and wood components. J Near Infrared 19(Spec): 287-308.

Steinberg D, Colla P (2009) CART: Classification and regression trees. Page 179 *in* X Wu and V Kumar eds. The top ten algorithms in data mining, Volume 9. CRC Press, Boca Raton, FL.

Sun Y, Lin Q, He X, Zhao Y, Dai F, Qiu J, Cao Y (2021) Wood species recognition with small data: A deep learning approach. Int J Comput Intell Syst 14(1):1451-1460.

Tsuchikawa S (2007) A review of recent near infrared research for wood and paper. Appl Spectrosc Rev 42(1): 43-71.

Tsuchikawa S, Kobori H (2015) A review of recent application of near infrared spectroscopy to wood science and technology. J Wood Sci 61(3):213-220.

Tsuchikawa S, Schwanninger M (2013) A review of recent near-infrared research for wood and paper (Part 2). Appl Spectrosc Rev 48(7):560-587.

van Blokland J, Nasir V, Cool J, Avramidis S, Adamopoulos S (2021a) Machine learning-based prediction of surface checks and bending properties in weathered thermally modified timber. Constr Build Mater 307:124996.

van Blokland J, Nasir V, Cool J, Avramidis S, Adamopoulos S (2021b) Machine learning-based prediction of internal checks in weathered thermally modified timber. Constr Build Mater 281:122193.

Via BK, Stine M, Shupe TF, So CL, Groom L (2004) Genetic improvement of fiber length and coarseness based on paper product performance and material variability—A review. IAWA J 25(4):401-414.

Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: Methods and applications. J Manuf Syst 48:144-156.

Watanabe K, Kobayashi I, Matsushita Y, Saito S, Kuroda N, Noshiro S (2014) Application of near-infrared spectroscopy for evaluation of drying stress on lumber surface: A comparison of artificial neural networks and partial least squares regression. Dry Technol 32(5): 590-596.

Willems W, Lykidis C, Altgen M, Clauder L (2015) Quality control methods for thermally modified wood. Holzforschung 69(7):875-884.

Yang SY, Kwon O, Park Y, Chung H, Kim H, Park SY, Yeo H (2020) Application of neural networks for classifying softwood species using near infrared spectroscopy. J Near Infrared Spectrosc 28(5):298-307.

Yun YH, Li HD, Deng BC, Cao DS (2019) An overview of variable selection methods in multivariate analysis of near-infrared spectra. Trends Analyt Chem 113:102-115.

Zhen Y, Ho TX, Roberts L, Schimleck LR, Sinha A (2022) On the selection of the weighting parameter value in optimizing *Eucalyptus globulus* pulp yield models based on NIR spectra. Wood Sci Technol 56(6):1835-1850.

Zhuang Z, Liu Y, Ding F, Wang Z (2021) Online color classification system of solid wood flooring based on characteristic features. Sensors (Basel) 21(2):336.