Original Research Article

JNIRS

Journal of Near Infrared Spectroscopy
2025, Vol. 33(3-4) 77–90

DOI: 10.1177/09670335251369048
journals.sagepub.com/home/jns
Sage

Light gradient-boosting machine model provides biologically realistic predictions of southern pine wood extractives content using near infrared spectroscopy hyperspectral imaging compared to artificial neural network and partial least squares regression models

Sameen Raut[1], Joseph Dahlen[1], Thomas L. Eberhardt[2], Laurence Schimleck[3], Daniel M. Johnson[1], Kier D. Klepzig[4] and Sudhagar Mani[5]

Abstract

Near infrared (NIR) spectroscopy combined with partial least squares (PLS) regression has been widely used to predict the physical, mechanical, and chemical properties of wood. Recently, advanced modeling techniques such as deep learning and machine learning models using spectral data have increasingly been used to predict variability in wood physical and mechanical properties, but studies on chemical properties are scarce. Herein, southern pine wood samples were acquired from multiple height levels from 119 trees to yield 1007 pith-to-bark radial strips. The strips were scanned on an NIR hyperspectral imaging system and the spectral data aligned to ring-level measurements. Acetone soluble extractives content was measured on 50-mm radial length samples with 262 data points assigned to training a PLS regression model, an artificial neural network (ANN) model, and a light gradient-boosting machine learning model (LGBM). The models were then evaluated on the test dataset (140 samples). The fit statistics were good for all models, although the LGBM model (RMSE = 2.1%) performed better than the PLS regression (3.3%) or ANN (2.6%) models. The models were subsequently used to predict extractives content for 31,314 rings. The LGBM model provided biologically realistic values for all rings, whereas the PLS regression model predicted 9648 negative values, and the ANN model extrapolated well beyond the measured values (max predicted = 69.4% vs max measured = 53.2%). This study highlights real benefits of using deep learning and machine learning models compared to PLS regression models for NIR hyperspectral imaging datasets.

Keywords

deep learning, loblolly pine, longleaf pine, machine learning, wood quality

Introduction

Nondestructive evaluation (NDE) of wood is widely implemented in forestry and forest products research and manufacturing processes.[1–4] The use of near infrared (NIR) spectroscopy (part of the electromagnetic spectrum in the range 800 to 2500 nm) as an NDE tool for wood properties characterization, quality control, and end-use performance assessment has been widely studied and found to be effective.[5–8] Applications of NIR spectroscopy in wood science include, but are not limited to, the characterization of wood physical, mechanical, chemical, and morphological properties (fiber length and width), surface characteristics, and species identification.[9–24] Many NIR studies have been conducted using southern pines given their global economic importance.[25–32]

Partial least squares (PLS) regression and principal component analysis (PCA) are the traditional modeling techniques used with NIR spectral data.[7,33] PLS regression is a statistical modeling technique that deconvolves both X

[1]Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA, USA
[2]Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI, USA
[3]Wood Science and Engineering, Oregon State University, Corvallis, OR, USA
[4]The Jones Center at Ichauway, Newton, GA, USA
[5]College of Engineering, University of Georgia, Athens, GA, USA

Corresponding author:
Joseph Dahlen, Warnell School of Forestry and Natural Resources, University of Georgia, 180 E Green St, Athens, GA 30602, USA.
Email: jdahlen@uga.edu

and Y matrices by projecting them onto new axes under the constraint that the decomposition best describes the covariance of the spectral and reference variables.[34,35] In PLS regression, similar to PCA, the first few latent variables account for a significant portion of the variability in X, but unlike PCA the decomposed X matrix is regressed against a Y scalar.[35] However, more advanced modeling techniques such as artificial neural networks (ANNs), light gradient boosting machines (LGBMs) and others have been utilized in recent years with success.[9,10,17,25,26,36–38]

ANNs are biologically inspired mathematical models capable of handling complex regression and classification tasks with a good degree of prediction accuracy.[39] ANNs are structured into input, hidden, and output layers which are made up of neurons interconnected with weighted links. Optimization of the number of hidden layers and the number of neurons in each of the hidden layers is necessary depending on factors including the task and the size of the training dataset. Adding more hidden layers and more neurons in each hidden layer makes a model more complex and prone to overfitting the training dataset, leading to weaker predictive performance on a new dataset not incorporated in model development. Therefore, it is important to adjust the model's architecture, including the number of hidden layers and neurons along with the learning rate to ensure that the model generalizes more accurately to the dataset instead of memorizing it.

LGBM is a newer implementation of gradient-boosting decision trees, and for some datasets it has been shown to reduce training times by a factor of 20 while still achieving the same level of accuracy as the conventional gradient-boosting decision trees.[40] LGBM can better manage large data sets and large number of features to significantly surpass extreme gradient boosting (XGBoost) and stochastic gradient boosting in ranking, regression, classification, and many other machine learning problems in terms of higher computational speed and lower memory usage.[40]

As noted by Schimleck et al.,[41] the practical significance of the predictive performance of deep learning and machine learning models has not been fully assessed for wood and such models have largely been utilized only as a comparison to PLS regression models with respect to their fit statistics. A broader question when using NIR calibrations for predicting properties is whether the models yield biologically realistic values when the properties are predicted from a broad sample set. Here the objective was to acquire a suitable dataset of measured reference data for acetone soluble extractives content[42] and use the data to train a deep learning model (ANN) and a machine learning model (LGBM) and compare the results to a PLS regression model. The calibrated models were then used to predict ring-level extractives content for samples scanned on the near-infrared hyperspectral imaging system. This study utilized longleaf pine (*Pinus palustris*) and loblolly pine (*Pinus taeda*) collected from naturally regenerated and plantation stands in the southeastern United States, two species that owing to their similarities, are grouped together as southern pines. The overall goal was to develop robust models that are not species, age, and stand origin specific, and have the potential to be readily applied to predict extractives content in southern pine wood.

Materials and methods

Sample locations, tree selection and felling

A total of 12 stands were sampled, nine longleaf pine and three loblolly pine stands.[43] Eleven stands were located in southwest Georgia and one stand was in northwest Florida. The age of the stands ranged from 30 to 69 years old with a median age of 36 years. In each stand an inventory was conducted and from each stand with 10 straight and visible defect-free trees being selected from across the diameter distribution. The diameter at breast height of the trees (mean = 28.0 cm, standard deviation = 3.6 cm) and total height (mean = 23.6 m, standard deviation = 2.5 m) were typical of stands that would be clearcut in a commercial forestry operation. The trees were felled using a chainsaw and delimbed. From each tree, approximately 5 cm thick cross-sectional disks were collected at heights of 0.15 m (stump base), 0.61 m, 1.37 m (breast height), 2.6 m, 5.3 m, 9.9 m, 14.8 m, and at heights where the over bark diameter was 10.1 cm, and 7.6 cm. If the over bark diameter of 10.1 cm occurred at a height above 19.8 m, an additional disk was cut at 19.8 m. A total of 1007 cross-sectional disks were collected and transported to the Wood and Fiber Quality Laboratory at the University of Georgia, Athens, GA, USA and frozen (−20°C) until further processing.

Pith to bark sample preparation and sample dimensions

Pith-to-bark radial strips were prepared from each disk according to a protocol developed by Dahlen et al.[44] to provide matched samples targeting wood property measurements as follows: (1) 8.5-mm longitudinal sample suitable to measure ultrasonic velocity (USV) (hereinafter referred to as the USV sample only to be consistent with sampling schematic used in Ref. 44), (2) 2-mm longitudinal sample suitable to measure specific gravity (SG) by X-ray densitometry (hereinafter referred to as the X-ray sample), and (3) 15-mm longitudinal sample available for other measurements/analyses (Figure 1). For the present study, the 15-mm sample was cut into three samples in the tangential direction, with one being 2.5-mm thick in the transverse direction for acetone-soluble extractives studies (hereinafter referred to as the acetone extractives sample) (Figure 1); note that all samples have adhering poplar wood "core holder" strips serving several functions, particularly sample stabilization during cutting.

Hyperspectral image collection

Hyperspectral images were collected using a push broom near infrared (NIR) hyperspectral camera (Specim FX17,

Figure 1. Left image shows the acetone-soluble extractives sample (2.5-mm transverse thickness), ultrasonic velocity sample (8.5-mm longitudinal thickness), and the X-ray sample (2-mm longitudinal thickness); right image shows the same samples, but the transverse view of the acetone extractives sample. Note the aligned growth rings for the samples and the yellow poplar strips on the edges of the samples which were removed prior to determining extractives content.

Specim Spectral Imaging Ltd, Oulu, Finland) fitted with a focusing lens (Specim OLET 17.5 f/2.1) (Figure 2). The instrument collects 224 wavelengths from 931 to 1718 nm at approximately 3.5 nm intervals. The working distance of the camera from the sample was fixed at 230 mm. Two sets of tungsten halogen lamps, each containing 3 bulbs, illuminate the sample at an angle of 45° positioned 240 mm from the sample. Sample motion and image acquisition was controlled using application software (Specim Lumo) supplied with the system. For intensity calibration, a dark current (with the camera's shutter closed) and a white reference were collected before acquiring each image. The scanning speed was set to 11.2 mm/s and the exposure time was set to 5.25 ms such that the white reference maximum value was approximately 85% of the detector range (0 to 4095, bit depth = 12) to maximize the linear dynamic range of the instrument. Samples were scanned in a windowless, dark laboratory space.

The USV samples were conditioned in a temperature and humidity-controlled room (22°C and 52% RH) to reach a moisture content (MC) of around 10% (dry-basis). The conditioned samples were placed on an aluminum sample tray that was anodized black, and the strips from each tree were scanned at one time, with the disks arranged from left to right and top to bottom based on their collected position within the tree, starting with the stump (0.15 m) height sample (Figure 2). Each raw image is 640 pixels wide and 2100 pixels long which translates to approximately 175 mm in width and 547 mm in length captured (each pixel in the image represents approximately 0.2731 mm by 0.2371 mm).

Hyperspectral image processing

The following hyperspectral image processing steps were carried out in Python version 3.9 (Python Software Foundation, https://www.python.org/) on the Spyder interface[45] using the libraries NumPy,[46] OpenCV,[47] pandas,[48] SciPy,[49] and Spectral Python (SPy).[50] Each image was calibrated with the measured dark and white reference signals:

$$R = \frac{R_0 - D}{W - D} \quad (1)$$

where R is the relative diffuse reflectance, R_0 is the measured diffuse reflectance, D is the dark current reference signal, and W is the white reference signal.

The samples were isolated from the background and from each other through several steps. Data collected in the range 931–942 nm was removed due to poor image quality. The standard deviation of each pixel relative diffuse reflectance spectra was found to yield one value per pixel,

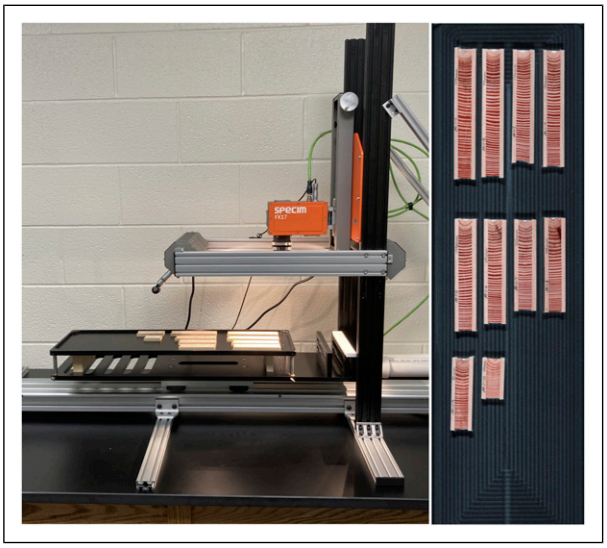

Figure 2. Left: Specim FX17 near-infrared hyperspectral imaging setup (samples scanned in a dark room) showing ten samples from one tree being imaged; right: resulting false color image (R = 1131 nm, G = 1414 nm, and B = 1438 nm).

and the image was blurred using a 5 × 5 median filter. Each pixel with the standard deviation value less than 0.03 was considered background and set to 0. A mask was then created whereby pixels with zero value were the background and the other pixels were converted to 1 via thresholding. The contours for the image were found using hierarchical contour retrieval models.[51] Any contour with area less than 5000 was removed and then the remaining contours were ranked from left to right and top to bottom to yield the samples from the first to last disk.

K-means clustering with 2 clusters and a maximum of 60 iterations was conducted with the first cluster (value = 0) indicating the background and the second cluster (value = 1) indicating the sample. A bounding rectangle using the convex hull was found for the sample and then the sample was rotated such that the image was rectangular and aligned with bark towards the left and pith on the right. Each radial strip image was cropped by 5 pixels from both the bark and pith sides to reduce the edge effects. Each radial strip image was then cropped to 29 pixels (~7.92 mm) from the middle part of the image to remove the poplar core holder and eliminate any wood material which could have glue present (Figure 3). The region of interest was saved for each sample as a tiff file. Pixel values from each file were averaged in the *Y*-axis (tangential) to give one spectrum for every pixel along the length of the sample (*X*-axis, radial) and saved to a CSV file.

Assignment of spectral data to annual rings after alignment with X-ray densitometry data

The X-ray radial strips were immersed in acetone for 24 h at room temperature to remove extractives. Post extraction, the samples were conditioned in a temperature and humidity-controlled room (22°C and 52% RH) to reach a MC around 10%. The conditioned samples were scanned on a QTRS-01X Tree Ring Scanner (Quintek Measurement Systems, Knoxville, TN, USA) with the X-ray beam directed through the transverse face at a radial resolution of 0.04 mm. Ring counts were performed with a SG threshold of 0.480 g cm^{-3} to separate earlywood from latewood on the tree ring scanner after which they were checked manually for accuracy. The 1007 USV radial strips had 31,314 rings.

To assign the spectral data into annual rings, the NIR and X-ray signals were aligned using the normalized cross-correlation method.[52,53] The hyperspectral data was converted to a pseudo-SG value using the signal from the 966 and 1541 nm wavelengths:

$$X = \log_{10}\frac{1}{\frac{R_{1541}}{R_{966}}} \quad (2)$$

where X is the pseudo-SG value, and R_{966} and R_{1541} are the diffuse reflectance values for these wavelengths. Because each image pixel represents a lower resolution (approximately 0.2731 mm) than each X-ray data point

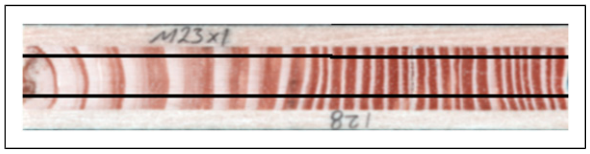

Figure 3. Pseudo color image showing how a hyperspectral image of each radial strip was processed to isolate the region of interest for spectral data acquisition. Area within the solid black lines indicates the region of interest.

(0.04 mm), the X-ray data was averaged to the lower resolution. Both the SG and the image data were then normalized by:

$$Z - score = \frac{X - \mu}{\sigma} \quad (3)$$

where Z-score is the normalized pseudo-SG value (ranging from 0 to 1), X is the pixel or SG value, μ is the overall mean for the sample, and σ is standard deviation. The normalized values from the images and the X-ray densitometer were then aligned whereby the image was shifted from left to right and the highest cross correlation value found. A plot for each sample was then constructed with the normalized values overlaid onto each other to confirm the image data was accurately matched with the X-ray densitometry data (Figure 4). The aligned data was then averaged within the annual ring boundaries to yield ring level spectral data, hence, one spectrum of 220 relative diffuse reflectance values from 945 nm to 1718 nm at 3.5 nm intervals for each growth ring.

Sample selection for measuring extractives content

The aligned USV radial strip spectral data was averaged for every 50-mm increment in length from pith to bark to give one spectrum per sub-sample. The 50-mm length was selected since it provided enough dry ground material from the acetone extractives sample for the measurement of other chemical properties not mentioned in this study. Some sub-samples were shorter than 50 mm and in these cases the spectrum reflected the actual length. The 1007 USV radial strips produced 2525 spectra, each representative of the sub-sample. The relative diffuse reflectance (R) spectra were transformed to pseudo absorbance (A):

$$A = log_{10}\left(\frac{1}{R}\right) \quad (4)$$

Second derivative spectra pretreatment with left and right gaps of 4 points using the Savitzky-Golay approach[54] was applied to the pseudo absorbance data which resulted in 212 X-variables (Figure 5). This single pretreatment was used because the ANN and LGBM models take a considerable amount of time for hyperparameter tuning, and this pretreatment has consistently produced good results across a wide range of samples and wood properties. The 50-mm USV sub-samples were split into the training and

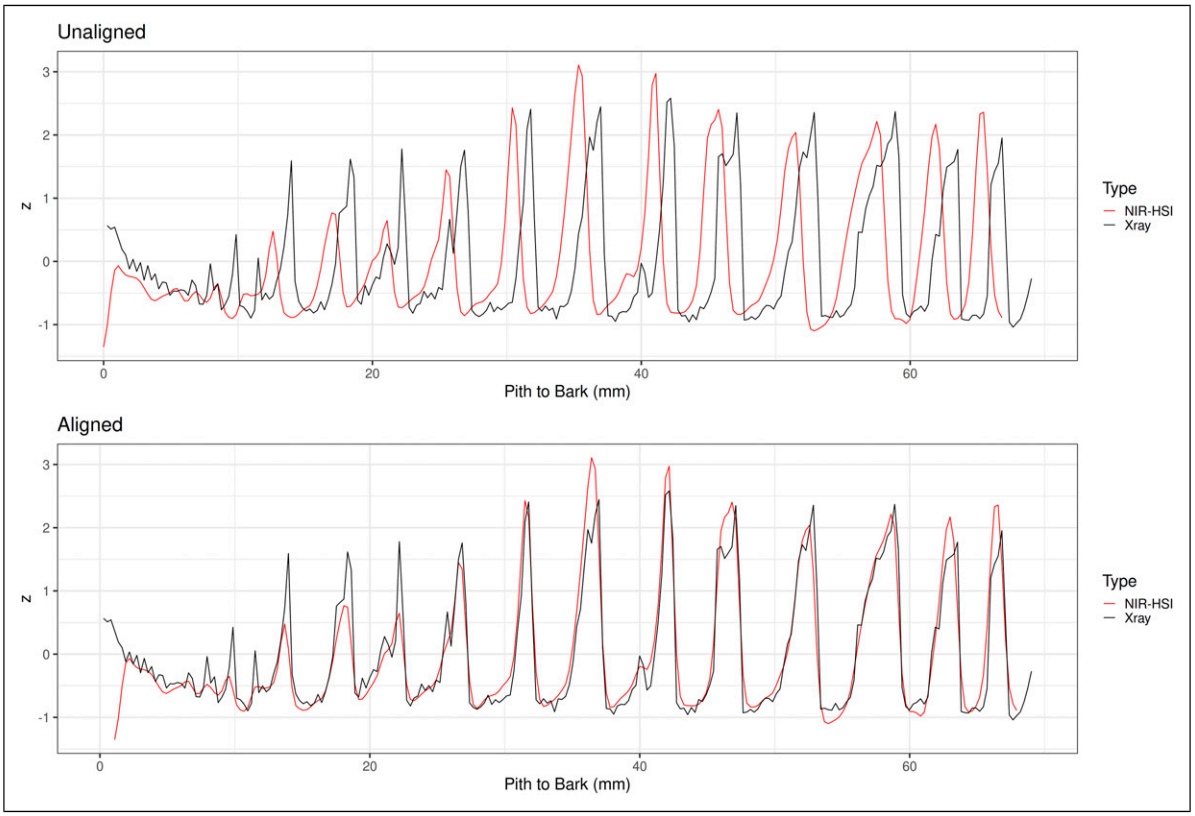

Figure 4. Unaligned ultrasonic velocity radial strip spectral data to X-ray densitometry data (top) versus aligned data after cross correlation (bottom).

test sets using the Duplex sample selection technique (modified Kennard-Stone)[55,56] using 3 principal components on the second derivative spectra. The most unique samples are selected based on Mahalanobis distance[57] (whereby the first most unique spectrum is assigned to the training set, the second to the test set, the third to the training set, and so on) to ensure that the data are representative of the population and the training and test sets are independent. Samples from an individual tree were assigned to either the training or testing dataset but not both sets. The actual 50-mm long acetone extractives samples matching the corresponding first 262 and 140 unique USV sub-sample spectra formed the reference training and the test set (a 65/35 split), respectively. The 31,314 ring-level second derivative pseudo absorbance spectral data was used as a prediction set.

Measuring reference extractives content (%)

The selected acetone extractives radial strips were split using a razor blade at 50 mm radial intervals to generate the sub-samples for reference measurements. The yellow

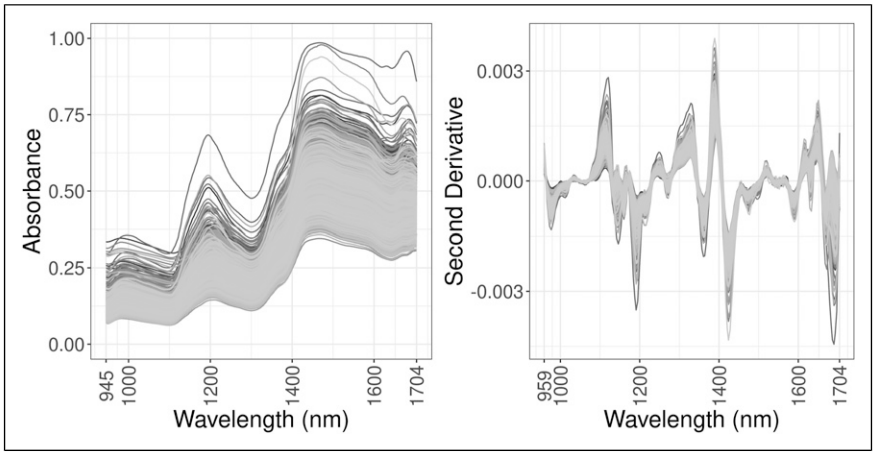

Figure 5. Left image shows pseudo absorbance spectra; right image shows spectra after applying the second derivative treatment using the Savitzky-Golay approach.

poplar wood glued to each sub-sample was then removed from each sub-sample. The 402 acetone extractives sub-samples were oven dried at 103°C until they reached constant weight (0% MC), placed in a desiccator, cooled, and their mass recorded. Note that oven drying and the steps leading up to oven drying will result in a loss of the volatile extractives. The samples were then extracted in a Soxhlet apparatus filled with acetone for 1 week, with the longer than typical extraction process being the result of these samples not having been ground (Figure 6). They were then oven dried again, cooled in a desiccator, and their mass again recorded. The difference before and after extraction gave the total amount of extractives (in grams) that was removed from a sample. The extractives content (%) in each 50-mm long sub-sample was calculated as follows:

$$\text{Extractivescontent}(\%) = \frac{\text{Weight}_{\text{OvenDry}} - \text{Weight}_{\text{OvenDryExtracted}}}{\text{Weight}_{\text{OvenDry}}} \times 100 \quad (5)$$

where $\text{Weight}_{\text{OvenDry}}$ is the oven dry weight before extraction, and $\text{Weight}_{\text{OvenDryExtracted}}$ is the oven dry weight after extraction is shown in Figure 6.

Modeling spectral data to extractives content (%)

Partial least squares (PLS) regression. PLS regression models were developed on the training dataset (N = 262), a maximum of 15 factors were considered with four interleaved cross-validation segments, where the final number of factors was determined based on the root mean square error (RMSE) and the coefficient of determination of cross-validation (r_{cv}^2). The training models were then tested for performance on the test set (N = 140) by calculating the RMSE and coefficient of determination of prediction (r_p^2). PLS regression for this study was implemented in R statistical software[58] using the R Studio interface[59] along with tidyverse collection of packages,[60] pls,[61] prospectr,[62] signal,[63] and gridExtra.[64]

Artificial neural networks (ANNs). The ANN model consisted of one input layer, four hidden layers, and one output layer. The number of neurons in each layer was 32, 32, 16, 8, 8, and 1, respectively. The total number of trainable parameters for this model was 8617. To determine this architecture, the number of hidden layers was systematically varied from 1 to 10, and neuron counts of 4, 8, 16, 32, 64, 128, 256, or 512 were tested for the input and each of the hidden layers. All the layers except the output layer used rectified linear unit (ReLU) as an activation function whereas the output layer utilized the default 'linear' activation function. L1 regularization with a value of 1e-9 was used in all four hidden layers to prevent overfitting. The first two hidden layers utilized 'RandomNormal' initializer whereas the last two layers used the default initializer ('GlorotNormal'). For error backpropagation, 'Adam' optimizer was used with default settings (learning rate = 0.001, beta_1 = 0.9, beta_2 = 0.999, and epsilon = 1e-7). Mean squared logarithmic error (MSLE) was the loss function used. The second derivative spectra which are the input (X) for the ANN model were scaled using the StandardScaler() function which transforms the dataset to have a mean of 0 and a standard deviation of 1 for all three datasets (training = 262, test = 140, and prediction =

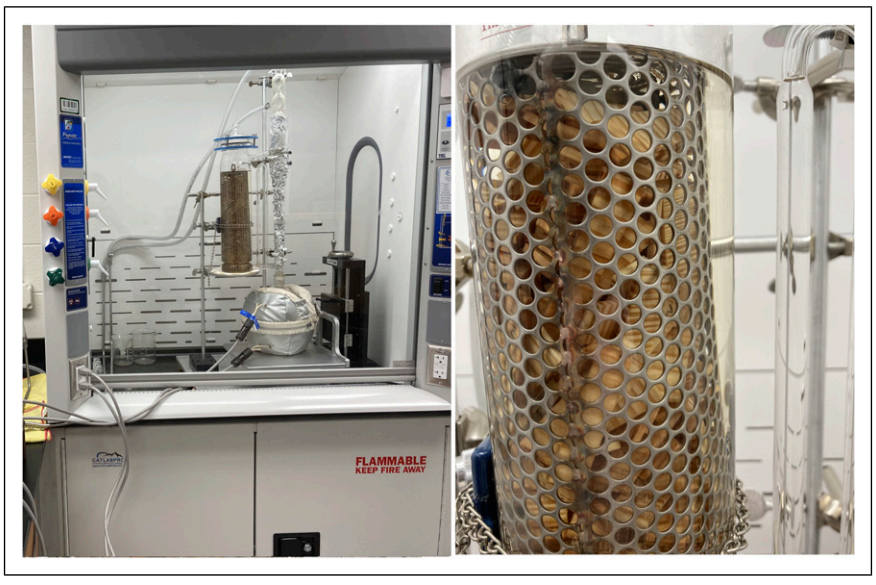

Figure 6. Left: Soxhlet apparatus used to exhaustively remove acetone soluble extractives from 50-mm long sub-samples; right: close up view of the Soxhlet apparatus showing submerged samples.

31,314). Training of the ANN model was done on the training dataset with a batch size of 32 and 1/4th of the training data was allocated as cross-validation data to assess loss and model performance metrics after every epoch (one complete iteration over the entire dataset). Early stopping was employed to prevent overfitting by stopping the model training if there was no improvement in validation MSLE over 40 epochs. Therefore, model training was stopped after 86 epochs. The performance of this model was tested on the test set (N = 140) by calculating the coefficient of determination (R^2) and root mean squared error (RMSE). ANN and LGBM models for this study were implemented in Python (Python Software Foundation, https://www.python.org/) with libraries TensorFlow,[65] Keras,[66] LGBM,[40] Scikit-learn,[67] and SHAP.[68,69] The ANN and LGBM models were fit to extractives defined on an extractive free basis and then converted to dry-basis.

Light gradient boosting machines (LGBMs). The LGBM model hyperparameters were found using the Optuna Bayesian optimization framework.[70] Optuna does an automated search for optimal hyperparameters using the Tree-structured Parzen Estimator which is its default optimization algorithm. Tree-structured Parzen Estimator models a probability distribution for the hyperparameters and selects new sets of hyperparameters that have a higher probability of improving performance giving them an edge over traditional random and grid search methods.[71,72] To implement Optuna, a regression study was created, and the direction was set to minimize the cross-validation RMSE (training and cross-validation split of 80/20) with the number of trials set to 320,000. The second derivative spectra which are the input (X) for the LGBM model were scaled using the Normalizer() function which transforms each row (treated as a vector) in all three datasets (training = 262, test = 140, and prediction = 31,314) to have a magnitude of 1. After exhausting the number of trials, the study returned the hyperparameters from the best trial having the lowest cross-validation RMSE. The hyperparameters returned by Optuna are 'num_leaves = 133', 'lambda_l1 = 0.261', 'lambda_l2 = 0.000419', 'feature_fraction = 0.369', 'bagging_fraction = 0.999', 'bagging_freq = 3', 'min_child_samples = 3', 'learning_rate = 0.0825', 'max_depth = 15'. The total number of estimators 'n_estimators' was set to 1600. The performance of the LGBM model fitted with these tuned hyperparameters was tested on the test set (N = 140) by calculating R^2 and RMSE.

The top 10 most important features (wavelengths) for the LGBM model were found using SHAP (SHapely Additive exPlanations)[68,69] which assigns an importance value to each feature for a particular prediction. SHAP takes a game-theory approach to explain the output of a machine learning model and provides significant theoretical improvement over the Shapley Value estimation method with regards to feature consistency and model stability across different machine learning models.[73,74] The 10 most important wavelengths were assigned to their respective structural components (if any) and functional groups as per Schwanninger et al.[75] Measured extractives content (%) in the training and test datasets (N = 402) were then regressed using a simple linear model against second derivative absorbance values for each of these 10 important wavelengths and the model fit statistics (R^2 and RMSE) reported for the 10 individual regression models. A final multiple linear model was also fitted by regressing the second derivative values from all the 10 important wavelengths against the measured extractives content (%).

Annual ring extractives content prediction. The PLS regression, ANN, and LGBM models were then used to predict extractives content (%) to 31,314 ring-level second derivative spectral data, thus providing extractives content estimates for each growth ring in all of the 1007 pith-to-bark USV radial strips that were scanned on the near-infrared hyperspectral imaging system. Hence the models were evaluated over a broad range of samples which allows us to evaluate the biological realism of the predictions for southern pine.

Extractives prediction on image for visualization. The LGBM model was used to predict the extractives content on the NIR-HSI images to visualize the spatial variation within the samples. Each pixel in the NIR-HSI image was converted to the second derivative using the procedures previously described, converted from 3D to 2D, and then the LGBM model was used to predict the extractives content. The data was converted back to the input dimensions, and then saved as a png file.

Results and discussion

The summary statistics for the extractives content (%) in the training and test set split by Duplex method is shown in Table 1, along with the overall data. The mean values were slightly different between the training (10.7%) and test (7.1%) sets however their range was very similar. Half of the measurements ranged from 0% to 10% extractives with the majority within this group being less than 5%.

Modeling results for PLS regression, ANN, and LGBM

The fit statistics for the PLS regression, ANN, and LGBM models are given in Table 2. The optimal number of factors for the PLS regression model was four, which was able to explain 92% of the variability in the training set (r_{cv}^2 = 0.92) and 88% of the variability in the test set (r_p^2 = 0.88). The training and test errors (RMSE) associated with the PLS regression extractives content model (%) were 3.0% and 3.3%, respectively (Figure 7). Where the PLS regression model is really deficient for predicting extractives content (%), is the negative predictions for measured extractives content (%) values that are very close to zero. The PLS regression model had 36 negative predictions for the training dataset with the minimum value being −5.8%. Similarly for the test dataset, the PLS regression model had

Table 1. Summary statistics of the measured extractives content (%) for the training, test, and the total dataset (training and test).

Set	N	Mean (%)	SD (%)	Min (%)	Max (%)
Training	262	10.7	11.0	0.0	53.2
Test	140	7.1	9.4	0.2	53.2
Total	402	9.4	10.6	0.0	53.2

30 negative predictions out of 140, with the most negative value being −8.5%. The problem with negative values was not observed for the ANN and LGBM models which are more advanced and have displayed good performances in modeling complex, non-linear relationships. The performance of the ANN model improved on the PLS regression model with training (R^2) and testing (r^2) values of 0.94 and 0.92, respectively, and the RMSE values were also lower. The LGBM model was the best model among the three with training and testing coefficients of determination of 1.00 and 0.95, and RMSE of 0.0% and 2.1%, respectively. The LGBM overfits to the training data but is still able to perform very well on the independent test data. This shows that the LGBM model was able to improve on the PLS regression and ANN models in explaining variability in extractives content (%). ANN and LGBM models, in this study, recognized that negative extractives content (%) values are not possible and therefore their predictions of NIR spectral data cannot be negative.

Gierlinger et al.[14] reported r_{cv}^2 of 0.86 and root mean standard error of cross-validation (RMSECV) of 0.32% from their PLS regression model developed using 63 wood powder samples from *Larix* sp. The range of acetone-soluble extractives in their study was very narrow, 0.8% to 3.6%, compared to this study. So and Eberhardt[32] reported prediction r^2 of 0.78 and SEP of 2.51% for PLS regression models developed to predict acetone-soluble extractives in longleaf pine wood meal. Most NIR studies on wood that have used advanced modeling techniques have focused on model development for wood physical and mechanical properties, or for wood identification, with limited studies done on wood chemical properties. To the authors' knowledge, the only study in the literature that applies deep learning to wood chemistry data is Zhang et al.[76] who measured extractives, lignin, and holocellulose content in wood samples collected from different species and developed adversarial transfer learning (ATL) models using NIR spectral data collected from both solid wood blocks and ground material. Solid wood is anisotropic which results in higher variability in the collected spectra due to spatial inhomogeneity. Ground material, at the expense of a time-consuming additional sample preparation step, is more homogenous, and the spectral data collected is thus more representative of the actual sample. They concluded that more advanced ATL models were able to adapt and make reasonable chemical composition predictions despite spectral differences resulting from different species and particle size (solid block or ground material). The more advanced ANN and LGBM models employed in this study likely accounted for the spatial inhomogeneity inherent in collecting spectral data from solid wood, ensuring that all predictions remained above zero.

The PLS regression, ANN, and LGBM models were then used to predict ring-level extractives content (%) for the 31,314 rings of the 1007 pith-to-bark USV radial strips (Table 3 and Figure 8). PLS regression predicted values had the lowest mean (4.1%, SD of 9.4%) whereas the ANN predicted values had the highest mean (6.1%, SD of 10.0%), and the LGBM predicted values were in the middle (5.4%, SD of 9.3%). PLS regression predicted a total of 9648 negative values, nearly 30% of the predictions, with the lowest value predicted being −28.9%. The ANN model here did not predict any negative values, but it predicted a maximum value of 69.4%, which is not a realistic value for extractives considering the maximum value in the measured dataset was 53.2% (Table 1). The ANN model produced 171 predictions out of 31,314 for which the extractives content values were greater than the maximum value measured, with 27 greater than 60%. The LGBM model predicted values from 0.1% to 52.6%, which is slightly constrained compared to the measured range on the 50 mm sub-samples (0% to 53.2%).

Table 2. Model fit statistics for the partial least squares (PLS) regression, artificial neural network (ANN), and light gradient boosting machine (LGBM) models fitted to the training set and evaluated on the test set for extractives content (%). The summary statistics of the model predicted values along with the number of negative predictions from each of these models is also shown.

Set	Variable	PLS regression	ANN	LGBM
Training (N = 262)	R^2	0.92	0.94	1.00
	RMSE (%)	3.0	2.6	0.0
	Mean (%)	10.7	10.8	10.7
	SD (%)	11.1	11.4	11.0
	Min (%)	−5.8	0.4	0.0
	Max (%)	48.6	52.4	53.1
	No. of negative predictions	36	0	0
Test (N = 140)	r^2	0.88	0.92	0.95
	RMSE (%)	3.3	2.6	2.1
	Mean (%)	6.0	6.5	6.8
	SD (%)	9.8	9.0	9.1
	Min (%)	−8.5	0.4	0.4
	Max (%)	49.2	53.5	51.5
	No. of negative predictions	30	0	0

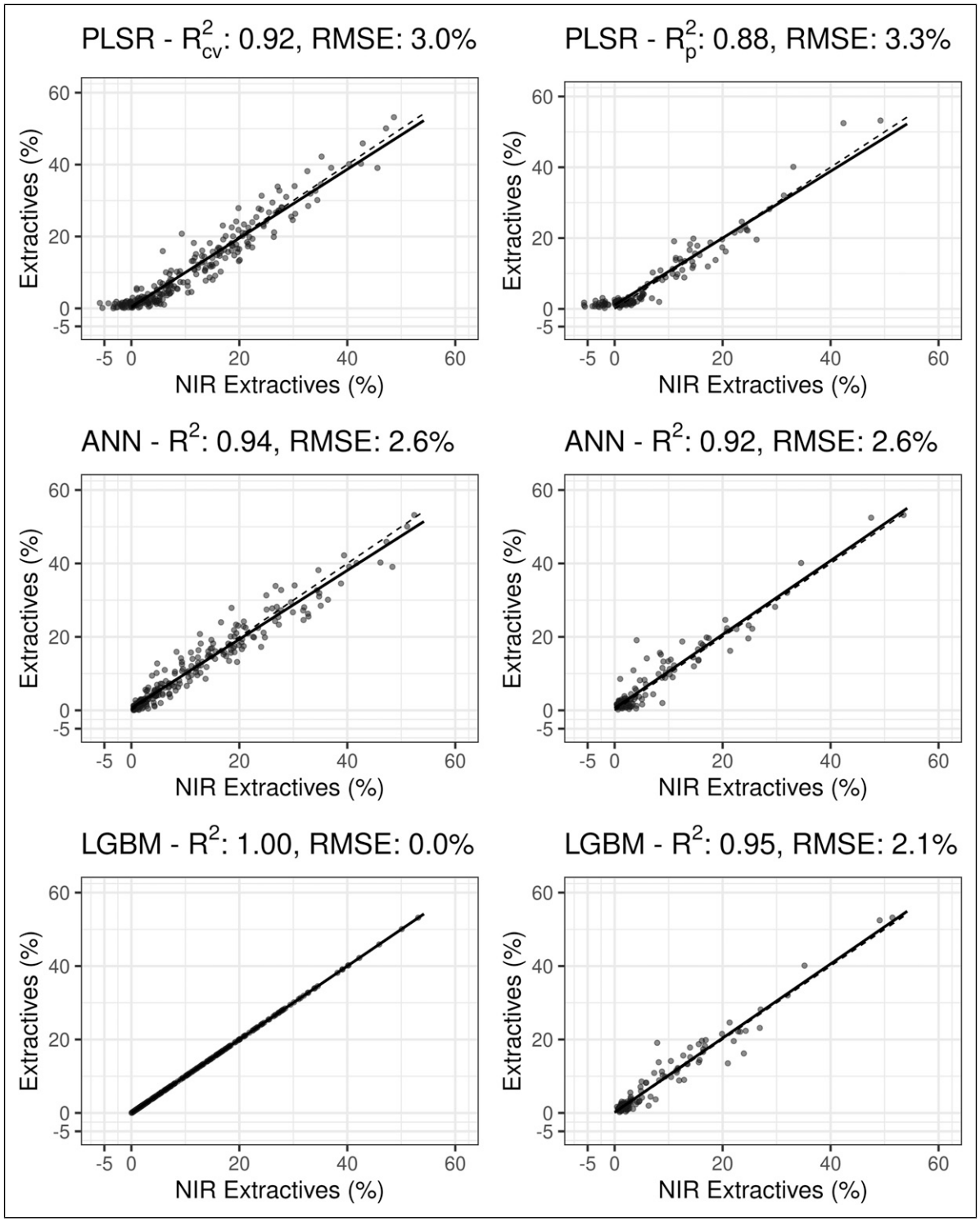

Figure 7. Measured versus near infrared (NIR) predicted extractives content (%) for training (left) and test datasets (right) from the partial least squares regression (top), artificial neural network (middle), and the light gradient boosting machine (bottom) models. The black dashed line denotes the line of equivalence.

SHAP (SHapely Additive exPlanations) which assigns an importance value to each feature for a particular prediction was used to identify the top 10 wavelengths which had the most impact on extractives content (%) predictions from the LGBM model (Table 4). Measured extractives content (%) in the training and test datasets (N = 402) were regressed against second derivative absorbance values for each of these 10 important wavelengths and linear model fit statistics (R^2 and RMSE) reported in Table 4. It is necessary to note that the most important wavelength returned by SHAP for LGBM will not necessarily be the most important for linear regression as well, which is what was found here. The wavelength 1378 nm was the most influential for the LGBM model whereas it was the third

Table 3. Summary of the ring-level extractives content (%) predictions made by the partial least squares regression, artificial neural network, and light gradient boosting machine models developed in this study.

	Prediction set (N = 31,314)					
Model	Mean (%)	SD (%)	Min (%)	Max (%)	No. of negative predictions (%)	No. of predictions more than maximum measured
PLS regression	4.1	9.4	−28.9	65.0	9648 (30%)	30
ANN	6.1	10.0	0.4	69.4	0 (0%)	171
LGBM	5.4	9.3	0.1	52.6	0 (0%)	0

most influential (out of the 10 selected) for the linear model with an R^2 of 0.77 and RMSE of 5.1%. The wavelength 1262 nm had the best fit statistics (R^2 = 0.86, RMSE = 4.0%) which is not assigned to any functional groups or structural components (similar to wavelengths 1265 nm and 1269 nm which likely represent the same functional groups). There were two wavelengths, 1385 nm and 1382 nm which are associated with isolated OH groups and C-H combinations but are not associated with any structural component, which is logical because extractives are a non-structural wood component. In *Larix* sp., Gierlinger et al.[14] reported 1654 nm associated with acetone extractives and taxifolin to be suitable for extractives model development which is similar to 1651 nm (Table 4) A model based on 1651 nm was able to predict 73% of the variation in extractives content with RMSE of 5.5%. When all the top 10 wavelengths returned by SHAP were regressed against measured extractives content (%), the resultant linear model reached an R^2 of 0.93 and an RMSE of 2.7%.

Figure 9 shows a pseudo color image (top), and the predicted extractives content using the LGBM model (bottom) for one sample having a high amount of extractives. For this particular sample, the concentration of extractives ranges from 35-45% in the first 120 pixels near the pith, but for older rings the extractive deposition becomes more variable and localized in specific areas. Towards the bark the extractives content is near 0%.

The excessively large number of negative predictions based on the PLS regression model was unexpected. Giroud et al.[77] developed NIR PLS regression models for SilviScan measured wood properties (density, microfibril angle, modulus of elasticity) and then predicted radial variation on 30,159 tree cores containing 114,729 growth rings from 6 different species. The resulting radial profiles were consistent with the literature and scaling the radial predictions to disk level estimates proved reliable when compared to measured data. Dahlen et al.[78] report within-tree tracheid length and width models developed from PLS regression NIR calibration models and they found logical predictions relative to measured values. Here using PLS regression, it was found that extrapolation well beyond the minimum and maximum measured values occurred. The differences found here for extractives relative to Giroud et al.[77] and Dahlen et al.[78] likely relate to how the properties are formed and the fact that extractives content can be zero, whereas other wood properties always have a non-zero value. Wood density (extractive free), microfibril angle, modulus of elasticity, tracheid length, and tracheid width properties are formed when the cells are elongating and when the cell wall is being lignified. In contrast, the extractives content for sapwood is relatively low, and deposition of extractives occur during the conversion from sapwood to heartwood, or as an active defense to an injury; both

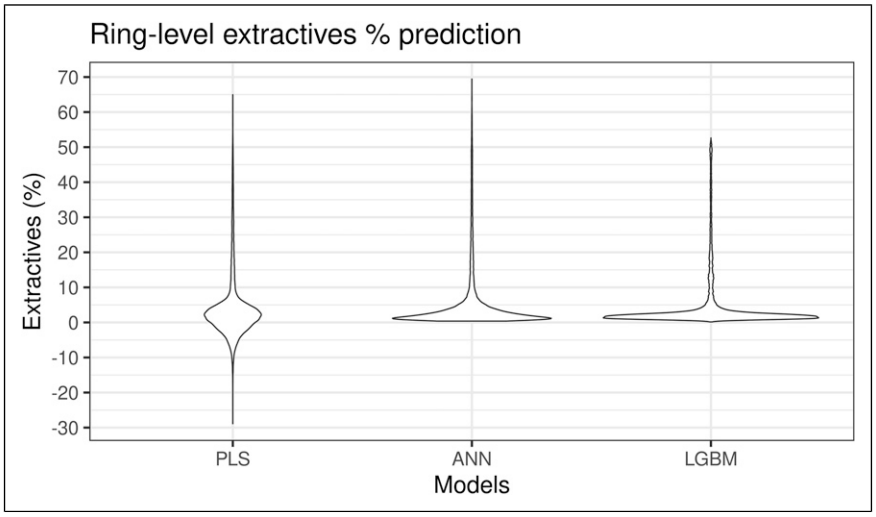

Figure 8. A violin plot showing the distribution of model predicted extractives content (%) for 31,314 ring-level spectral data for the partial least squares regression, artificial neural network, and light gradient boosting machine models.

Table 4. Wavelengths of importance found through SHAP (SHapely Additive exPlanations) regressed against all measured extractives content (%) values (N = 402).

	Extractives content (%)			
	(N = 402)			
Wavelength (nm)	Structural component	Functional group	R^2	RMSE (%)
1378	Hemicellulose/all	$-CH_3$	0.77	5.1
1385	-	-OH, C-H	0.50	7.5
1265	NA	NA	0.84	4.2
1382	-	-OH, C-H	0.65	6.2
1339	Hemicellulose/all	$-CH_3$	0.70	5.8
1269	NA	NA	0.76	5.2
1651	(Extractives)	C-H	0.73	5.5
1184	Lignin/cellulose	$-CH_3$	0.74	5.4
1262	NA	NA	0.86	4.0
1188	Lignin/cellulose	$-CH_3$	0.69	5.9
All 10 wavelengths			0.93	2.7

Note. Band assignments from Schwanninger et al.[75] NA – Not Assigned.

processes are extremely variable relative to the more ordered cell elongation and lignification processes.

Unlike the PLS regression models, the ANN and LGBM models did not result in prediction of negative values. The ANN model predictions extrapolated beyond the measured range, whereas the LGBM constrained the maximum values. Tree-based models such as LGBM do not extrapolate, and assuming that the training dataset covers the entire range of values this interpolation only behavior can be beneficial. We do anticipate that for some rings the extractives content would be subtly higher than what was measured on the training dataset and test dataset, and thus some extrapolation in the predictions is probably prudent, but how much more? Extractives content is reported on a dry-basis and not an extractive-free basis, and thus it is important to consider that the 69.4% maximum value predicted by the ANN model represents a substantial increase in the amount of extractives predicted versus 53.2%. Consider that for one g of wood, a 53.2% extractives content represents 1.14 g of extractives compared to 2.27 g of extractives for 69.4%. The maximum value predicted by the PLS regression model was 65.0%, which corresponds to 1.86 g of extractives, again a substantial increase compared to 1.14 g. It is concluded that the most realistically predicted extractives content across the entire range of possible values was from the LGBM model. Alternatively, both the PLS regression and ANN models could be used with filters to reduce the amount of extrapolation. If negative predicted values are set to zero, and values greater than the measured maximum are set to that value (53.2%), the PLS regression average and standard deviation are 5% and 8.6% (originally 4.1% and 9.4%), whereas the ANN is 6.1% and 9.9% (originally 6.1% and 10.0%). The PLS regression filtered predictions resulted in the most changes compared to the unfiltered predictions due to the large number of negative predictions, whereas the ANN model mean and SD are effectively unchanged due to the large number of predictions (31,314 annual rings).

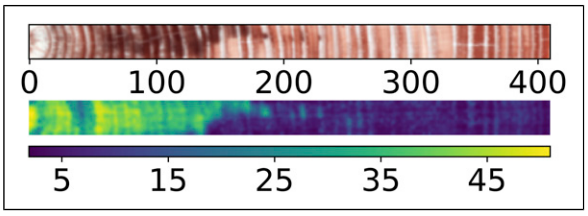

Figure 9. Pseudo-color image (top) sample with predicted extractives content using the light gradient boosting machine model (bottom).

PLS regression models fit quickly and perform very well in general for most wood properties. These models are optimized by selecting the number of factors (latent variables) and their performance tested using cross-validation, and prediction fit statistics on independent test data. ANN models have the capability to outperform PLS regression models but there are many hyperparameters that need to be tuned to achieve a good model. While searching for an optimal architecture, a larger ANN model with many hidden layers and more neurons in each layer (resulting in a greater number of trainable parameters) attained R^2 values close to 1.0 yet showed poorer performance on the testing dataset. Here, a relatively smaller ANN model was able to generalize better to both datasets. The LGBM hyperparameter tuning run with 320,000 trials took close to 75 h to complete. Consequently, this meant that performing a similar search for each different spectral pre-treatment would either require substantially more time or involve a fewer number of trials which would reduce the chances of finding the optimal hyperparameter set. ANN and LGBM models can perform better than PLS regression models, but unless properly tuned, they typically perform similar or worse.

Conclusions

This study tested the prediction performances of three distinct types of models: partial least squares regression,

artificial neural network, and light gradient boosting machine, for their ability to accurately predict southern pine wood extractives content using NIR spectral data. The models were then used to predict extractives content for 31,314 ring-level spectral data thus providing a realistic extractives content estimate for each growth ring in the 1007 pith-to-bark radial strips scanned on the NIR hyperspectral imaging system, providing a practical application to these models. Many studies in the literature report on improving model fit statistics over the PLS regression models using machine learning and deep learning approaches but few describe using the models for any practical applications. In terms of model fit statistics, PLS regression, ANN, and LGBM were all acceptable, with the LGBM model outperforming the other two models with an r_p^2 of 0.95 and an RMSE of 2.1%. The predictions from these different models however were very different. When the PLS regression model was applied to the ring-level spectral data, the model predicted a total of 9648 negative values. In contrast, the ANN and LGBM models did not yield any negative predictions in any of the datasets. For the ring-level spectral data, ANN however, predicted very high extractives content predictions for some growth rings, with the maximum value predicted being 69.4%, which again is not a realistic value, thus rendering ANN unsuitable for making ring-level predictions for extractives content in southern pines unless the maximum values were capped to the measured dataset. The LGBM model performed better across the entire range of extractives content for southern pines, and therefore is recommended for predicting extractives content using NIR spectral data in southern pines. SHAP (Shapely Additive explanations) identified 10 most important wavelengths with the absorbance at 1262 nm having a very strong linear relationship with extractives content, it alone predicted 86% of the variability with an RMSE of 4.0%. This study reports the first successful application of ANN and LGBM models for predicting extractives content in southern pines using NIR spectral data with strong implications for extractives content predictions from PLS regression models for future studies. The findings presented here also demonstrate that an LGBM model enables more precise and scalable extractives content predictions, with potential for significant implications of wood quality assessment in forest product industries, and future advancements in precision forestry and wood characterization.

Acknowledgments

We would like to thank the reviewers and the editors for their helpful comments and suggestions for improving the manuscript.

ORCID iDs

Sameen Raut https://orcid.org/0000-0002-7088-5946
Joseph Dahlen https://orcid.org/0009-0002-7231-4845
Thomas L. Eberhardt https://orcid.org/0000-0002-8818-4822

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was supported by the USDA Forest Service, Forest Products Laboratory (No. 20-JV-11111134-031), the Jones Center at Ichauway, and the Warnell School of Forestry and Natural Resources at the University of Georgia, and the NIFA McIntire-Stennis project 1023340.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Raut S and Dahlen J. Image analysis to assess wood variability in longleaf pine cross-sectional disks. *Wood Fiber Sci* 2023; 55: 157–175.
2. Ross RJ and Pellerin RF. *Nondestructive Testing for Assessing Wood Members in Structures: A Review*; General Technical Report FPL-GTR-70; Department of Agriculture, Forest Service, Forest Products Laboratory: Madison, WI, USA, 1994; 40p.
3. Ross RJ. Nondestructive testing and evaluation of wood. In: Ross RJ (ed). *Nondestructive evaluation of wood*. 2nd ed. USDA Forest Service, Forest Products Laboratory, 2015, p. 169. Gen. Tech. Rep. FPL-GTR-238.
4. Schimleck L, Dahlen J, Apiolaza LA, et al. Non-destructive evaluation techniques and what they tell us about wood property variation. *Forests* 2019; 10: 728.
5. Tsuchikawa S and Kobori H. A review of recent application of near infrared spectroscopy to wood science and technology. *J Wood Sci* 2015; 61: 213–220.
6. Tsuchikawa S. A review of recent near infrared research for wood and paper. *Appl Spectrosc Rev* 2007; 42: 43–71.
7. Sandak J, Sandak A and Meder R. Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips. *J Near Infrared Spectrosc* 2016; 24: 485–505.
8. Schimleck L, Ma T, Inagaki T, et al. Review of near infrared hyperspectral imaging applications related to wood and wood products. *Appl Spectrosc Rev* 2022; 58: 585–609.
9. Pan X, Yu Z and Yang Z. A multi-scale convolutional neural network combined with a portable near-infrared spectrometer for the rapid, non-destructive identification of wood species. *Forests* 2024; 15: 556.
10. Ayanleye S, Nasir V, Avramidis S, et al. Effect of wood surface roughness on prediction of structural timber properties by infrared spectroscopy using ANFIS, ANN and PLS regression. *Eur J Wood Prod* 2021; 79: 101–115.
11. Haartveit EY and Flæte PO. Rapid prediction of basic wood properties by near infrared spectroscopy. *N Z J Sci* 2006; 36: 393–407.
12. Hoffmeyer P and Pedersen J. Evaluation of density and strength of Norway spruce wood by near-infrared reflectance spectroscopy. *Holz Roh Werkst* 1995; 53: 165–170.
13. Willems W, Lykidis C, Altgen M, et al. Quality control methods for thermally modified wood: COST action FP0904

2010-2014: thermo-hydro-mechanical wood behaviour and processing. *Holzforschung* 2015; 69: 875–884.
14. Gierlinger N, Schwanninger M, Hinterstoisser B, et al. Rapid determination of heartwood extractives in Larix sp. by means of Fourier transform near infrared spectroscopy. *J Near Infrared Spectrosc* 2002; 10: 203–214.
15. Mora CR and Schimleck LR. On the selection of samples for multivariate regression analysis: application to near-infrared (NIR) calibration models for the prediction of pulp yield in Eucalyptus nitens. *Can J Res* 2008; 38: 2626–2634.
16. Ma T, Inagaki T, Ban M, et al. Rapid identification of wood species by near-infrared spatially resolved spectroscopy (NIR-SRS) based on hyperspectral imaging (HSI). *Holzforschung* 2019; 73: 323–330.
17. Watanabe K, Kobayashi I, Matsushita Y, et al. 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* 2014; 32: 590–596.
18. Zhao R, Huo X and Zhang L. Estimation of modulus of elasticity of Eucalyptus pellita wood by near infrared spectroscopy. *Spectrosc Spectr Anal* 2009; 29: 2392–2395.
19. Schimleck LR, Evans R and Ilic J. Estimation of Eucalyptus delegatensis wood properties by near infrared spectroscopy. *Can J Res* 2001; 31: 1671–1675.
20. Fujimoto T, Yamamoto H and Tsuchikawa S. Estimation of wood stiffness and strength properties of hybrid larch by near-infrared spectroscopy. *Appl Spectrosc* 2007; 61: 882–888.
21. Fujimoto T, Kurata Y, Matsumoto K, et al. Application of near infrared spectroscopy for estimating wood mechanical properties of small clear and full length lumber specimens. *J Near Infrared Spectrosc* 2008; 16: 529–537.
22. Thumm A and Meder R. Stiffness prediction of radiata pine clearwood test pieces using near infrared spectroscopy. *J Near Infrared Spectrosc* 2001; 9: 117–122.
23. Thumm A, Riddell M, Nanayakkara B, et al. Near infrared hyperspectral imaging applied to mapping chemical composition in wood samples. *J Near Infrared Spectrosc* 2010; 18: 507–515.
24. Thumm A, Riddell M, Nanayakkara B, et al. Mapping within-stem variation of chemical composition by near infrared hyperspectral imaging. *J Near Infrared Spectrosc* 2016; 24: 605–616.
25. Nasir V, Ali SD, Mohammadpanah A, et al. Fiber quality prediction using NIR spectral data: tree-based ensemble learning vs deep neural networks. *Wood Fiber Sci* 2023; 55: 100–115.
26. Ali SD, Raut S, Dahlen J, et al. Utilization of synthetic near-infrared spectra via generative adversarial network to improve wood stiffness prediction. *Sensors* 2024; 24: 1992.
27. Dahlen J, Diaz I, Schimleck L, et al. Near-infrared spectroscopy prediction of southern pine No. 2 lumber physical and mechanical properties. *Wood Sci Technol* 2017; 51: 309–322.
28. Via BK, Shupe TF, Groom LH, et al. Multivariate modelling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (Pinus palustris). *J Near Infrared Spectrosc* 2003; 11: 365–378.
29. Via BK, So CL, Shupe TF, et al. Prediction of wood mechanical and chemical properties in the presence and absence of blue stain using two near infrared instruments. *J Near Infrared Spectrosc* 2005; 13: 201–212.
30. Schimleck LR, Jones PD, Clark A, et al. Near infrared spectroscopy for the nondestructive estimation of clear wood properties of Pinus taeda L. from the southern United States. *For Prod J* 2005; 55: 21–28.
31. Nabavi M, Dahlen J, Schimleck L, et al. Regional calibration models for predicting loblolly pine tracheid properties using near-infrared spectroscopy. *Wood Sci Technol* 2018; 52: 445–463.
32. So CL and Eberhardt TL. Chemical and calorific characterisation of longleaf pine using near infrared spectroscopy. *J Near Infrared Spectrosc* 2010; 18: 417–423.
33. Schimleck LR, Matos JLM, Trianoski Ret al. Comparison of methods for estimating mechanical properties of wood by NIR spectroscopy. *J Spectrosc* 2018; 2018: 4823285. DOI: 10.1155/2018/4823285, Epub ahead of print 2018.
34. Geladi P and Kowalski BR. Partial least-squares regression: a tutorial. *Anal Chim Acta* 1986; 185: 1–17.
35. Fernandes A, Lousada J, Morais J, et al. Comparison between neural networks and partial least squares for intra-growth ring wood density measurement with hyperspectral imaging. *Comput Electron Agric* 2013; 94: 71–81.
36. Mora CR and Schimleck LR. Kernel regression methods for the prediction of wood properties of Pinus taeda using near infrared spectroscopy. *Wood Sci Technol* 2010; 44: 561–578.
37. Ho TX, Schimleck LR, Dahlen J, et al. Utilization of genetic algorithms to optimize loblolly pine wood property models based on NIR spectra and SilviScan data. *Wood Sci Technol* 2022; 56: 1419–1437.
38. Ho TX, Schimleck LR and Sinha A. Utilization of genetic algorithms to optimize Eucalyptus globulus pulp yield models based on NIR spectra. *Wood Sci Technol* 2021; 55: 757–776.
39. Haykin S. *Neural networks and learning machines*. Third Edition. Pearson Education Inc, 2009. DOI: 10.1016/B978-0-12-809633-8.20339-7. Epub ahead of print 2009.
40. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Guyon I, Luxburg UV, Bengio S, et al. (eds). Advances in neural information processing systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017. Curran Associates, Inc, pp. 3149–3157.
41. Schimleck LR, Ho TX, Duong DV, et al. Comparison of genetic algorithm optimized and partial least squares regression density models for Acacia auriculiformis. *Wood Fiber Sci* 2024; 56: 51–61.
42. Eberhardt TL and Samuelson LJ. Collection of wood quality data by X-ray densitometry: a case study with three southern pines. *Wood Sci Technol* 2015; 49: 739–753.
43. Sandera M. *Comparing wood quality between naturally regenerated longleaf, planted longleaf, and planted loblolly pine*. MS Thesis. University of Georgia, 2022.
44. Dahlen J, Auty D, Eberhardt TL, et al. Determination of ring-level dynamic modulus of elasticity in loblolly pine from measurements of ultrasonic velocity and specific gravity. *For An Int J For Res* 2023; 96: 588–604.

45. Raybaut P. Spyder-documentation. https://www.spyder-ide.org, accessed 10 May 2023.
46. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature* 2020; 585: 357–362.
47. Bradski G. The OpenCV library. *Dr Dobb's J Softw Tools* 2000; 25: 122–125.
48. McKinney W. Data structures for statistical computing in Python. In: Proceedings of the 9th python in science conference. Austin, Texas, USA, 28 Jun–3 July 2010, pp. 56–61.
49. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020; 17: 261–272.
50. Boggs T. Spectral python (SPy). https://www.spectralpython.net/, accessed 10 May 2023.
51. Serra J. *Image analysis and mathematical morphology*. Academic Press, 1982.
52. Keane RD and Adrian RJ. Theory of cross-correlation analysis of PIV images. *Appl Sci Res* 1992; 49: 191–215.
53. Yoo JC and Han TH. Fast normalized cross-correlation. *Circ Syst Signal Process* 2009; 28: 819–843.
54. Savitzky A and Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 1964; 36: 1627–1639.
55. Kennard RW and Stone LA. Computer aided design of experiments. *Technometrics* 1969; 11: 137–148.
56. Snee RD. Validation of regression models: methods and examples. *Technometrics* 1977; 19: 415–428.
57. Mark HL and Tunnell D. Qualitative near-infrared reflectance analysis using Mahalanobis distances. *Anal Chem* 1985; 57: 1449–1456.
58. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/ (accessed 10 May 2023).
59. RStudio. RStudio: integrated development for R. RStudio, Inc. Boston, MA. https://www.rstudio.com/ (accessed 10 May 2023).
60. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *J Open Source Softw* 2019; 4: 1686.
61. Liland KH, Mevik B-H and Wehrens R. pls: partial least squares and principal component regression. R package version 2.8-1. https://cran.r-project.org/package=pls (accessed 10 May 2023).
62. Stevens A and Ramirez-Lopez L. An introduction to prospectr package. *R package Vignette R package* version 0.2.6 (accessed 10 May 2023). https://cran.r-project.org/web/packages/prospectr/index.html
63. Signal Developers. Signal: signal processing. https://r-forge.r-project.org/projects/signal/, accessed 10 May 2023.
64. Auguie B. gridExtra: miscellaneous functions for 'grid' graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra. accessed 10 May 2023.
65. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI'16), Savannah, GA, USA, 2–4 November 2016. USENIX Association, 2016, pp. 265–283.
66. Chollet F. *Deep learning with Python*. Manning Publications Co, 2018.
67. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–2830.
68. Lundberg SM and Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, et al. (eds) Advances in neural information processing systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017, pp. 4765–4774.
69. Lundberg SM, Erion GG and Lee S-I. Consistent individualized feature attribution for tree ensembles. arXiv preprint. arXiv:1802.03888 2018. DOI: 10.48550/arXiv.1802.03888.
70. Akiba T, Sano S, Yanase T, et al. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining (KDD '19), Anchorage, AK, USA, 4 – 8 August 2019, pp. 2623–2631.
71. Shekhar S, Bansode A and Salim A. A comparative study of hyper-parameter optimization tools. In: 2021 IEEE Asia-Pacific conference on computer science and data engineering (CSDE), Brisbane, Australia, 8–10 December 2021. IEEE, 2021, pp. 1–6.
72. Rimal Y, Sharma N and Alsadoon A. The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, Bayesian, genetic, and optuna algorithms. *Multimed Tool Appl* 2024; 83: 74349–74364.
73. Meng Y, Yang N, Qian Z, et al. What makes an online review more helpful: an interpretation framework using xgboost and shap values. *J Theor Appl Electron Commer Res* 2021; 16: 466–490.
74. Shapley LS. A value for n-person games. In: *Contributions to the Theory of Games II*. Princeton University Press, 1953, pp. 307–317.
75. Schwanninger M, Rodrigues JC and Fackler K. A review of band assignments in near infrared spectra of wood and wood components. *J Near Infrared Spectrosc* 2011; 19: 287–308.
76. Zhang Z, Zhong H, Li Y, et al. Predicting components of pulpwood feedstock for different physical forms and tree species using NIR spectroscopy and transfer learning. *Cellulose* 2024; 31: 551–566.
77. Giroud G, Defo M and Bégin J. Determination of radial profiles of wood properties using a near infrared scanning system. *J Near Infrared Spectrosc* 2021; 29: 24–32.
78. Dahlen J, Nabavi M, Auty D, et al. Models for predicting the within-tree and regional variation of tracheid length and width for plantation loblolly pine. *Forestry* 2021; 94: 127–140.