RESEARCH | Article in Press

Machine learning based on physically relevant descriptors for estimating optical transmittance and haze in transparent wood

D. Nuvoli · A. Bifulco · A. Casciello · A. Mariani · A. Cabras · G. Venusti · D. Pugliese · M. Terzi · Giulio Malucelli

Received: 14 July 2026 / Accepted: 11 November 2026
© The Author(s) 2026

Abstract

Transparent wood (TW) has emerged as a promising class of sustainable optical materials due to its combination of high optical transmittance, tunable haze, low density, and mechanical resistance within a bio-based, hierarchical structure. Despite the rapid development of TW systems for energy-related and optical applications, optimizing their optical properties is still largely based on a *trial and error* approach. This is because light transport is governed by the complex interplay of wood microstructure, refractive index matching, thickness, and processing-induced heterogeneities.

In this work, we evaluate a machine-learning framework based on physically relevant descriptors for estimating the optical transmittance and haze of TW systems. The curated dataset comprised 95 samples spanning different wood species, polymer-infiltration strategies, thicknesses, refractive index values, and optical responses. For haze estimation, the primary evaluation was based on complete-case data containing only experimentally reported haze values, whereas median-based imputation was examined separately as a sensitivity analysis of missing-data treatment. Three supervised regression algorithms (i.e., multilayer perceptron, random forest, and gradient boosted trees) were evaluated. Optical transmittance was estimated with comparatively lower error than haze, which remained more challenging because of the limited information on scattering-related microstructure.

Therefore, the results should be interpreted as an exploratory assessment of predictive relationships within the available data domain, rather than as a framework for predicting optical properties purely from material composition, processing, and structural descriptors.

Keywords Transparent wood · Optical transmittance · Haze · Light scattering · Machine learning · Supervised algorithms

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

1 Introduction

Being one of the most recent materials invented and developed, transparent wood (TW) is gaining a wide interest from both a scientific and technological point of view, as it conjugates multifunctional features with sustainability [1–3]. Unlike conventional transparent materials such as glass and amorphous polymers, TW retains the anisotropic hierarchical organization of natural wood [4] while exhibiting significant visible-light transmission [5]. This combination of structural and optical functionality makes it a particularly attractive platform for applications in energy-efficient building envelopes, solar energy harvesting, light diffusion, smart windows, and optoelectronic components [1, 6–11].

The fabrication of TW generally relies on a top-down approach in which the native wood template is chemically modified to reduce absorption and scattering of the incident light. To this aim, lignin, which is the main chromophore component of wood, is either selectively removed through chemical processes or bleached (hence chemically deactivating its light-absorbing moieties - chromophores). The resulting cellulose-rich porous scaffold is then infiltrated with a monomer or a mixture of monomers that finally undergo polymerization or curing, according to their chemical functionalities, within the wood channels. It is important to note that the choice of infiltrated monomer(s) is directly related to its perfect match with the refractive index of the modified wood. This way, it is possible to mitigate internal refractive index discontinuities and stabilize the brittle delignified structure, thereby not only providing enough mechanical strength to the wood scaffold but also enabling efficient light transmission through a material that is naturally opaque [12–16]. Although the basic concept of TW fabrication is currently well established, the optimization of optical performance is still a challenging and demanding issue: indeed, the final optical transmittance and haze do not depend only on the chemistry of the delignification process or the choice of infiltrating monomer(s), but also on a complex combination of different factors comprising wood thickness, residual lignin, wettability/adhesion at the interface between wood channels and infiltrated monomer(s), and presence of micro-voids and porosity due to inefficient infiltration processes. All these factors may significantly impact the optical behavior of the resulting TW, even for samples having the same material and chemical composition (i.e., same pristine wood, delignification process, monomer(s) infiltration and polymerization/curing) [5]. TW's optical performance mostly relies on the evaluation of its total optical transmittance and haze. The former measures the fraction of incident light that passes through the material, while the latter quantifies the fraction of transmitted light that is diffusely scattered [17]. These two parameters do not show antagonism but are strictly dependent on the destination of the designed TW materials. In particular, such applications as daylighting systems and photovoltaics significantly benefit from a high haze, while limited haze (and high optical transmittance) is demanded in e.g., glazing applications, where visual clarity is usually needed. Therefore, the design of TW materials requires a rationally conceived strategy that must consider both optical transmittance and haze concurrently [18]. From a physical point of view, several factors govern the light transport in TW materials, namely: refractive index matching, optical path length, and structural heterogeneity. Any difference in refractive index between the cellulose-rich cell walls and the infiltrated polymer mainly accounts for scattering phenomena taking place at the interface. Also, the sample thickness is important, as it controls the cumulative probability of absorption and multiple scattering events along the light propagation path. At the same time, such microstructural defects as pores, interfacial debonding, and local density variations can remarkably affect (increase) the haze.

These factors make it difficult to establish robust structure-optical-property relationships in real TW samples and require extensive experimental testing [9, 19]. Data-driven approaches are, therefore, attracting growing interest in materials science [20–23] and wood science and technology [24–29]. Machine-learning (ML) methods can identify patterns in heterogeneous datasets and support property estimation. ML approaches have already been applied to selected properties of transparent wood. For example, Martinka et al. [30] employed neural networks, a class of ML models, to analyze the fire behavior of epoxy-based transparent wood. However, ML applications specifically addressing the estimation of the principal optical descriptors of TW, namely optical transmittance and haze, remain limited.

The aim of this work is to explore the potential and current limitations of a ML approach based on physically relevant descriptors for estimating optical transmittance and haze in TW. A small, heterogeneous dataset was assembled to evaluate whether these optical responses can be conditionally estimated from refractive index, thickness, and the complementary optical descriptor. Accordingly, the proposed framework should be regarded as a conditional estimation approach rather than a fully ex ante material design model, since estimation of one optical property requires knowledge of the complementary optical response. For haze estimation, the primary analysis was restricted to samples with experimentally reported haze values, whereas median-based imputation was examined separately as a sensitivity analysis of missing-data treatment. Overall, the study investigates whether physically relevant descriptors and the coupled optical response can support optical property estimation within the available data domain, while clearly considering the limitations imposed by dataset size, data completeness, and heterogeneity.

2 Materials and methods

2.1 Preparation of TW samples

2.1.1 Materials

Balsa wood with a nominal initial thickness of 1.0 mm was purchased from matèria PRIMA (Spain). This value refers to the pristine wood before delignification, monomer infiltration, and polymerization. Sodium chlorite (purity = 80%, d = 2.43 g/cm3, MW = 90.44 g/mol), glacial acetic acid (purity ≥ 99.7%, d = 1.049 g/cm3, MW = 60.05 g/mol), isobornyl acrylate (IBA, purity ≥ 85%, d = 0.986 g/cm^3, MW = 208.30 g/mol, containing 200 ppm of monomethyl ether hydroquinone, MEHQ as inhibitor), isobornyl methacrylate (IBMA, purity = 85%, d = 0.983 g/mL, MW = 222.32 g/mol), 2-hydroxyethyl methacrylate (HEMA, purity ≥ 99%, d = 1.07 g/mL, MW = 130.14, containing ≤ 50 ppm MEHQ as inhibitor), tetraethylene glycol diacrylate (TEGDA, technical grade, d = 1.11 g/mL, containing 100–150 ppm hydroquinone, HQ, and 150–200 ppm MEHQ as inhibitors), benzoyl peroxide (BPO, with 25% H2O, MW = 242.23 g/mol), and diphenyl(2,4,6-trimethylbenzoyl)phosphine oxide (TPO, purity ≥ 96.5%, MW = 348.37 g/mol) were purchased from Sigma-Aldrich. Sodium acetate (purity > 99%, MW = 82.04 g/mol) was purchased from Fluka Chemicals. Absolute ethanol (purity = 99.9%, d = 0.79 g/mL, MW = 46.07 g/mol) and acetone (purity = 99%, d = 0.79 g/mL, MW = 58.08 g/mol) were purchased from VWR Chemicals. Before use, benzoyl peroxide was recrystallized from a chloroform/methanol solution (1:2 v/v), and inhibitors were removed from HEMA, TEGDA, and IBA using an alumina column. The remaining reagents were used as received.

2.1.2 Delignification procedure

Balsa wood was cut into 4 × 4 cm^2 pieces. The samples were immersed in a water buffer solution at pH 4.6, prepared by dissolving 0.4 g of sodium acetate and 0.3 mL of acetic acid per 100 mL of solution. A volume of 12 mL of buffer solution was used for each gram of wood and sodium chlorite was added at a concentration of 1% (w/v). The delignification process was carried out at 80 °C for 18 h [31, 32]. Following the treatment, the samples were cleaned in an ultrasonic bath using a sequential washing cycle: three consecutive steps with distilled water, three with ethanol, and a final three with acetone. Each individual step was carried out for 30 min.

2.1.3 TW preparation

TW containing copolymers of IBA and IBMA. 0.3 wt% TPO (UV initiator, λ_{max} = 380 nm) was added to the solution containing different amounts of IBA and IBMA monomers (χIBA = 0, 0.25, 0.50, 0.75, 1.00). The

acetone-stored delignified wood was then immersed in this mixture to infiltrate for 24 h. Photopolymerization of the infiltrated samples was performed by exposing the wood pieces to UV radiation ($\lambda = 395$ nm) for up to 15 min. The process utilized an OmniCure® LX500 UV LED Spot Curing System Controller (Excelitas Technologies) equipped with Intelli-Lamp® technology.

TW containing copolymers of HEMA and TEGDA. 0.3 mol% BPO (thermal initiator) was added to the solution containing different amounts of HEMA and TEGDA monomers (χHEMA = 0, 0.25, 0.50, 0.75, 1.00). The acetone-stored delignified wood was then immersed in this mixture to infiltrate for 3 h under vacuum and the 18 h at ambient pressure. Thermal polymerization of the infiltered samples was performed in oven at 80 °C for 4 h.

After delignification, monomer infiltration, and polymerization, the final TW specimens were thicker than the initial 1.0 mm balsa sheets because of processing-induced swelling and the formation of thin polymerized resin layers on both external surfaces. Therefore, all thickness values reported in Table 1 and Table S1 refer to the final specimens and were measured after polymerization.

Moreover, the apparent density of some representative TW formulations (i.e., IBA-IBMA-0.50 and HEMA-TEGDA 0.50 samples) was calculated as the ratio between the dry sample mass, measured using an analytical balance, and the external sample volume. The latter was calculated from the measured length, width, and average thickness of small TW square specimens (approximately 1 cm × 1 cm × 2 mm). The results are collected in Table S1, together with the mass increase, referred to the mass of the untreated balsa wood specimen before delignification and monomer infiltration. As shown in Table S2, polymer infiltration and polymerization substantially increased the apparent density of both TW systems compared with the native balsa. The mass increase confirms the substantial uptake of the monomer formulations during processing. Although the final thickness increased to approximately 2 mm because of swelling and the formation of polymer-rich surface layers, the final apparent densities remained close to 1 g cm^{-3}, indicating an effective filling of the wood structure. The slightly higher apparent density of the HEMA-TEGDA-based samples is consistent with the higher density of its copolymer matrix and/or a more compact polymer-filled structure.

2.2 Optical characterization of TW samples

Optical transmittance and haze were measured using an ultraviolet-visible-near-infrared (UV-Vis-NIR) spectrophotometer (Agilent Cary 5000) equipped with an integrating sphere, enabling the collection of both directly transmitted and forward-scattered light. Transmittance was measured in the 300–900 nm spectral range, while haze was evaluated in the visible region (380–780 nm) according to ASTM D1003 [33].

Table 1 Optical transmittance and haze measured at 550 nm through the specimen thickness, perpendicular to the longitudinal wood-channel direction, for TW samples with different compositions. Optical transmittance values are reported as mean ± standard deviation of three repeated spectral measurements performed on the same sample under identical experimental conditions. For HEMA-TEGDA copolymers, the HEMA molar fraction is indicated. For IBA-IBMA copolymers, the IBA molar fraction is indicated

Wood template	Type of infiltrating monomer and molar fraction in copolymers	Average thickness (mm)	Optical transmittance (%)	Haze (%)
Balsa	HEMA	1.85	81.787 ± 0.004	69.86
Balsa	TEGDA	2.13	74.288 ± 0.009	80.44
Balsa	HEMA-TEGDA-0.25	1.78	81.549 ± 0.003	78.98
Balsa	HEMA-TEGDA-0.50	2.07	80.830 ± 0.005	77.45
Balsa	HEMA-TEGDA-0.75	2.12	81.862 ± 0.001	74.59
Balsa	IBA	2.08	71.920 ± 0.002	86.44
Balsa	IBMA	2.01	69.021 ± 0.002	71.99
Balsa	IBA-IBMA-0.25	1.96	78.817 ± 0.002	59.22
Balsa	IBA-IBMA-0.50	1.98	80.038 ± 0.007	66.43
Balsa	IBA-IBMA-0.75	2.02	78.866 ± 0.005	60.13

A 100% transmission reference (WHITE, W) and a 0% baseline (DARK, D) were recorded prior to measurements. Samples were mounted at the entrance port of the integrating sphere, ensuring complete coverage and perpendicular alignment with the incident beam. Instrumental parameters, including spectral bandwidth, data interval, and scan rate, were kept constant for all measurements to ensure consistency and reproducibility.

The total transmitted signal (S), including both ballistic and scattered components, was acquired, and the spectral transmittance (T) was calculated as:

$$T = \frac{S - D}{W - D} \tag{1}$$

Haze was determined using the same integrating sphere configuration following ASTM D1003 through four measurements: (i) T_1, reference signal without sample using a white standard; (ii) T_2, total transmitted light through the sample; (iii) T_3, instrumental scattering measured without sample using a light trap; and (iv) T_4, diffuse transmitted light measured with the sample and light trap.

The haze value was calculated as:

$$\text{Haze} = \frac{T_d}{T_t} = \left(\frac{T_4}{T_2} - \frac{T_3}{T_1}\right) \times 100\% \tag{2}$$

where T_d is the diffuse transmittance and T_t is the total transmittance. All measurements were performed with the incident beam propagating through the specimen thickness, i.e., perpendicular to the longitudinal direction of the wood channels. For optical transmittance, three repeated spectral measurements were acquired for each sample at the same measurement position and under identical experimental conditions, and the reported transmittance values correspond to their mean. Haze was determined from a single measurement sequence comprising the four spectra (T_1–T_4) required by the ASTM D1003 procedure.

n denotes the refractive index of the TW specimens. It was determined through the sample thickness, normal to the specimen surface and perpendicular to the wood channel direction, corresponding to the direction used for optical characterization. A measurement parallel to the channels was not pursued, as specimens perpendicularly cut to the channel direction had poor mechanical integrity.

2.3 Scanning electron microscopy

The morphological and microstructural features of the prepared TW samples were assessed using a ZEISS EVO 50 XVP SEM (Oberkochen, Germany) at 10 kV of source potential; before SEM analysis, the samples were coated with a conductive layer of gold.

2.4 Dataset construction and scientific rationale

The identification of the main properties related to TW is essential for evaluating its performance in real-life use cases. However, as mentioned in the Introduction, experimental measurement of these properties can be costly and time-consuming. In the following sections, the application of different ML regression models is reported. Such ML models were employed to estimate the values of optical transmittance and haze for a specific TW sample. To achieve this, an estimation framework was developed to address key challenges in ML, including limited data availability and missing values in the initial dataset.

2.4.1 Dataset description and missing data imputation strategy

The overall dataset comprised 95 transparent wood (TW) samples, including 85 samples collected from the scientific literature and 10 samples prepared and experimentally characterized in the present study. Of these, 94 samples constituted the development dataset used for model training and validation (Table S1), whereas

one experimental sample (IBA-IBMA-0.75) was retained separately for the final case-specific estimation exercise (Table S3). The samples prepared and tested in the present study are labelled "EXP" in the corresponding tables. The dataset spans a broad range of wood templates, polymer-infiltration systems, thicknesses, and optical responses. Each sample was described using numerical variables selected for their relevance to light transport. Specifically, refractive index, thickness, and one optical descriptor were used as predictors, whereas the complementary optical descriptor was used as the target variable. Because optical transmittance and haze are physically coupled descriptors of light transport, each was used as an input to estimate the other. The resulting framework is therefore conditional: optical transmittance is estimated using haze as one of the available predictors, whereas haze is estimated using optical transmittance. This configuration was adopted to investigate the reciprocal information carried by the two coupled optical responses within the available dataset. Considering the complete collection of 95 observations, 33 samples (34.7%) belonged to the balsa wood template, whereas 62 samples (65.3%) belonged to other wood templates. Within the 94-sample development dataset used to train and validate the models, 32 observations (34.0%) were balsa and 62 (66.0%) were non-balsa. For haze estimation, samples with experimentally reported haze values constituted the primary analysis dataset. Samples with missing haze values were excluded from this complete-case configuration, so that model evaluation was based exclusively on experimentally measured target values. Median-based imputation was investigated separately to assess the sensitivity of model behaviour to different treatments of missing haze data and was not used to define the primary estimates of haze-estimation performance.

Table S1 contains missing haze values for samples for which no experimental haze measurement was reported. Given the small dataset size and the consequent risk of overfitting, missing-data treatment was examined cautiously. Median-based imputation strategies were selected because they are comparatively robust to outliers in small datasets [34]. For the sensitivity analyses, missing haze values were replaced using the following strategies:

- **Global median**: haze median computed over the whole dataset.
- **Group median**: haze median computed within a specific group of similar samples (i.e., samples with the same infiltrating monomer/polymer). This strategy aims to provide a more representative estimate for each sample.
- **Hierarchical median**: missing haze values were replaced using a multi-level strategy in which the median was first computed within the most specific group (i.e., samples sharing the same infiltrating monomer/polymer). If insufficient data were available (i.e., fewer than three samples per group), a progressively broader grouping (i.e., wood template) was considered. If this grouping also contained insufficient data, the global median was used as the final fallback.

The three median-based strategies generated three imputed dataset versions (Tables S4-S6), which were analyzed exclusively as secondary sensitivity analyses of missing-data treatment. Imputed haze values were not interpreted as experimental observations or as evidence of primary haze-estimation performance.

2.4.2 Data augmentation and variability modeling

In addition to the imputation analysis, Gaussian perturbation was applied exclusively as a controlled regularization and sensitivity-analysis procedure within the training subset. This procedure does not generate new independent physical observations and was not used as a substitute for experimental data. Small perturbations were applied to the input features (up to 10% of their standard deviation) and to the target variables (up to 3% of their standard deviation) to assess model sensitivity to controlled numerical variation while preserving the original feature structure. Because these perturbation amplitudes were not derived from a systematic uncertainty budget across the literature sources, they were not interpreted as representations of experimental measurement uncertainty. An "Is augmented?" column was included in Tables S7-S9 to discriminate original observations from synthetically perturbed training instances.

Data augmentation was applied only after the training/validation split and exclusively to the training subset. A unique identifier (Original_Sample_ID) was assigned to each original observation to track its synthetic derivatives.

The random 70/30 partition was defined at the level of the original observations, and synthetic instances were subsequently included only when their corresponding Original_Sample_ID belonged to the training subset. Consequently, the validation subset contained only original, non-augmented observations, and no synthetic derivative of a validation observation was included during model training. The target sample was also retained in its original, non-augmented form. This procedure was adopted to prevent information leakage between training and validation data.

The augmentation was applied as follows:

- Input parameters: $x_{aug} = x + N(0,\ 0.1 \bullet \sigma_x)$.
- Target parameters: $y_{aug} = y + N(0,\ 0.03 \bullet \sigma_y)$.

where:

- x_{aug} are the augmented samples related to input parameters.
- y_{aug} are the augmented samples related to target parameters.
- N is the set representing the randomic Gaussian noise.
- σ_x and σ_y are the standard deviations of the features computed on the dataset.

Data augmentation was applied only to samples with experimentally available feature values; conversely, no augmentation was performed when haze had been imputed. This restriction avoids propagating uncertainty associated with statistical estimates. Each imputed dataset was treated independently, and a fixed random seed was used to ensure reproducibility. While the original dataset contained 94 instances, the augmented one includes 224 instances derived from the original observations. During model evaluation, however, synthetic instances were included only when the corresponding Original_Sample_ID belonged to the training subset. Therefore, data augmentation increased the density of the training data without increasing the number of independent experimental observations.

2.5 ML models: multilayer perceptron, random forest, and gradient boosted trees

Three regression models (i.e., multilayer perceptron (MLP), random forest (RF), and gradient boosted trees (GBT)) were evaluated for estimating optical transmittance and haze in a transparent balsa-wood sample infiltrated with an IBA-IBMA mixture at a 0.75–0.25 molar ratio (Table S3), which was selected as the target sample (Fig. 1). These algorithms were selected because they represent complementary modelling paradigms for nonlinear structured data. Their application enabled comparison between neural-network- and tree-based regression approaches, while the influence of the different preprocessing configurations, including imputation and data augmentation, was examined separately through sensitivity analyses.

MLP is a feedforward neural network that approximates nonlinear relationships through hidden layers and nonlinear activation functions. During training, internal weights are adjusted by backpropagation to minimize prediction error [35]. It was included because optical transmittance and haze may depend on nonlinear interactions among the available descriptors; however, neural networks are sensitive to limited sample size, feature scaling, and overfitting. Due to this sensitivity, a k-nearest-neighbours (k-NN) algorithm was used as a preliminary

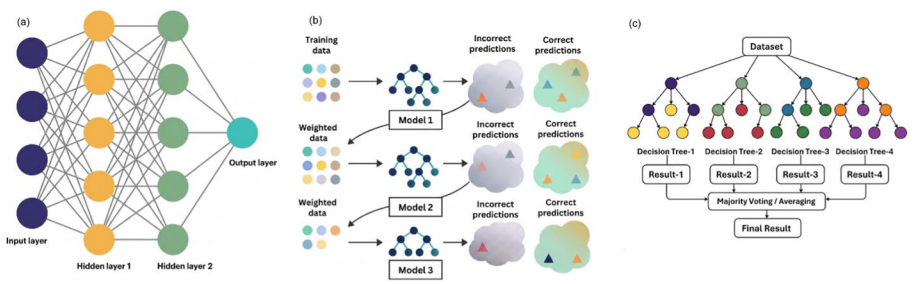

Fig. 1 ML models employed for the estimation of optical transmittance and haze

local-selection step to identify observations most similar to the target sample before MLP-based estimation. The purpose of this step was to restrict the local learning context to chemically and physically related TW systems rather than to introduce k-NN as an additional independent regression model [36]. Accordingly, MLP with kNN is treated throughout the validation and estimation analyses as a single modelling configuration.

RF combines estimates from multiple decision trees built on random subsets of observations and features [37]. It is well suited to heterogeneous and noisy datasets because it can capture nonlinear relationships while remaining relatively robust to limited sample size.

GBT uses a sequence of decision trees in which each learner corrects the residual error of the preceding one [38]. This approach can provide accurate estimates for structured tabular data, although careful hyperparameter tuning is required to limit overfitting.

2.5.1 Estimation workflow: from data processing to property estimation

The ML workflow was implemented using KNIME Analytics Platform, an open-source environment for visual, node-based data analysis [39]. The workflow comprised four main phases: preprocessing, hyperparameter optimization, evaluation, and estimation (Figure S1). Missing-value imputation and the generation of synthetic observations were performed outside KNIME, whereas the assignment of synthetic observations to the training subset was controlled within the modelling workflow using the Original_Sample_ID associated with each observation.

Within the workflow, non-numeric or non-informative features were removed, and Z-score normalization was applied to the scale-sensitive MLP configuration. Hyperparameters were optimized using a KNIME parameter-optimization loop, exploring the number and depth of trees for RF and GBT, and the learning rate, momentum, and number of training iterations for MLP.

The dataset was divided using a random 70/30 training/validation split at the individual-sample level. The partition was defined using the original observations before the inclusion of synthetic instances. Following partitioning, synthetic instances were associated with their corresponding original observations through Original_Sample_ID and were included exclusively when the original observation belonged to the training subset. Therefore, the validation subset contained only original, non-augmented observations. Samples from the same reference, polymer family, wood template, or experimental series were not constrained to remain in the same subset. Accordingly, the reported validation metrics represent random-split interpolation performance within the available dataset and do not provide a strict assessment of extrapolation to independent TW systems.

Model performance was quantified using the mean absolute error (MAE) on the validation subset. MAE was preferred to mean absolute percentage error because percentage-based errors can become unstable when the target assumes small values. After optimization, the models were retrained using the corresponding training configurations and applied to the IBA-IBMA-0.75 target sample (Table S3). Because related IBA/IBMA systems are represented in the dataset, this application should be interpreted primarily as a case-specific within-family estimation exercise rather than as an assessment of extrapolation to an independent material family. All preprocessing operations derived from the training data were applied consistently to the target sample. The median-imputed and augmented configurations were evaluated separately as sensitivity analyses to examine the influence of preprocessing choices on model behavior.

2.5.2 Complete-case cross-validation and sensitivity analyses

In addition to the original train-validation analysis, model robustness was assessed through leave-one-out cross-validation (LOOCV). In each iteration, one eligible observation was excluded from model training and used exclusively for validation, while the remaining observations were used for model fitting. The procedure was repeated until each eligible observation had served once as the validation sample, and the resulting out-of-fold predictions were pooled to calculate the overall mean absolute error (MAE). This analysis reduces the dependence of performance assessment on a single random train-validation partition.

For haze estimation, the primary model evaluation was performed using complete-case LOOCV, including only samples with experimentally reported haze values and applying neither haze imputation nor synthetic data augmentation. Consequently, all haze values used as targets in this primary evaluation corresponded to experimental measurements. The resulting complete-case LOOCV performance therefore provides the primary estimate of haze-estimation performance in this study.

LOOCV was also performed for MLP with kNN, RF, and GBT across the median-imputed and augmented configurations, which were evaluated exclusively as sensitivity analyses to investigate the influence of missing-data treatment and controlled numerical perturbation on model behavior. For augmented configurations, the same identifier-based separation was applied within each LOOCV iteration. When an original observation was held out for validation, all synthetic instances carrying the same Original_Sample_ID were excluded from the corresponding training fold, whereas synthetic observations associated with the remaining original samples were retained for model fitting. Thus, neither the held-out observation nor any synthetic derivative of that observation was available to the model during the corresponding LOOCV iteration. The LOOCV results for haze and optical transmittance are reported in Tables S14 and S15 of the Supporting Information, respectively.

Finally, the three modelling configurations trained on the complete-case dataset without haze imputation or synthetic data augmentation were applied to the IBA-IBMA-0.75 target sample to provide a case-specific assessment. Because related IBA/IBMA formulations are represented in the training data, this analysis should be interpreted as a within-family estimation exercise rather than as an assessment of extrapolation to an independent material family. The resulting estimates and absolute errors are reported in Tables S16 and S17 of the Supporting Information.

To investigate the potential influence of the unequal representation of wood templates, an additional subgroup analysis was performed using out-of-fold predictions obtained from the unmodified dataset, without applying missing-value imputation or synthetic data augmentation. The analysis was conducted separately from the primary complete-case LOOCV benchmark. The analysis started from the unmodified 94-observation development dataset. No missing-value imputation or synthetic augmentation was applied. Subgroup error evaluation was restricted to the 66 observations with experimentally available haze values, comprising 28 balsa and 38 non-balsa samples. For each of the three investigated modelling configurations (MLP with kNN, RF, and GBT), the mean absolute error was calculated separately for these two groups, considering both optical transmittance and haze estimation. The wood-template classification was used exclusively for performance analysis and was not included among the predictive features. The separate IBA-IBMA-0.75 target sample was excluded from this analysis. The corresponding subgroup results are reported in Table S18 of the Supporting Information.

3 Results and discussion

3.1 Experimental optical properties and morphology of TW systems

The wavelength-dependent optical transmittance and haze are shown in Figs. 2 and 3. All samples exhibit a rapid increase in transmittance above ~350–400 nm, followed by a plateau in the visible region, indicating reduced absorption and a scattering-dominated transport regime.

In the HEMA/TEGDA systems (Fig. 2a), visible-range transmittance is consistently high, with hybrid formulations showing performance comparable to or slightly higher than that of the pure components, particularly referring to TEGDA. A similar trend is observed for the IBA/IBMA systems (Fig. 3a), where mixed compositions display higher transmittance than the corresponding homopolymers, consistent with improved optical homogeneity. The haze spectra (Figs. 2b and 3b) decrease with increasing wavelength and remain relatively high, confirming a substantial forward-scattering contribution. In both systems, pure polymers generally show higher haze, whereas mixed formulations generally show lower scattering. A weak but reproducible inflection around 600 nm is observed in several spectra. Because no intrinsic absorption is expected in this region, this feature is unlikely

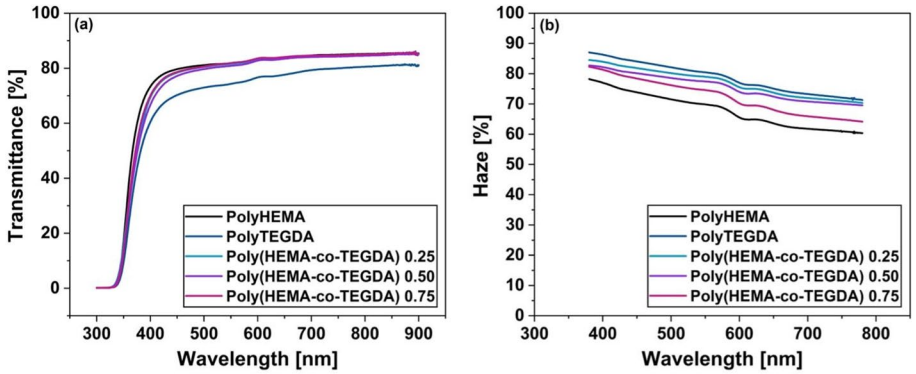

Fig. 2 Spectral optical properties of transparent wood based on HEMA/TEGDA systems: (**a**) optical transmittance and (**b**) haze as a function of wavelength. For copolymers, the HEMA molar fraction is indicated in the legend

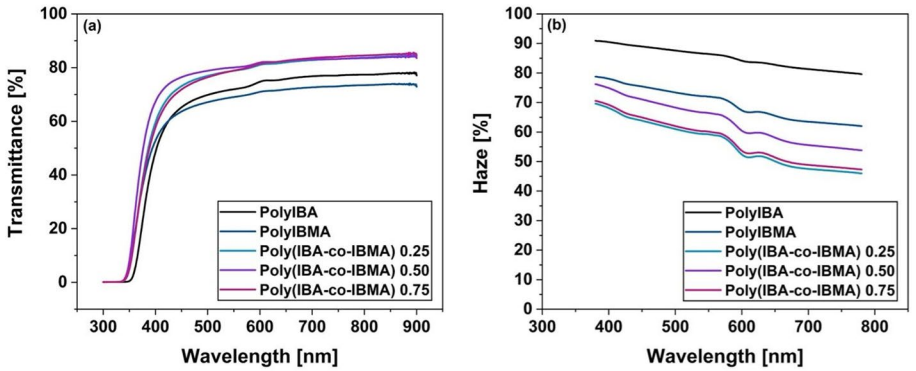

Fig. 3 Spectral optical properties of transparent wood based on IBA/IBMA systems: (**a**) optical transmittance and (**b**) haze as a function of wavelength. For copolymers, the IBA molar fraction is indicated in the legend

to originate from residual lignin, whose absorption is mainly confined below 450 nm [40]. It is more cautiously attributed to wavelength-dependent scattering associated with refractive index mismatch and multiscale structural heterogeneity [41]. To enable direct comparison among compositions, optical transmittance and haze values were extracted at 550 nm and are collected in Table 1.

All samples exhibit high optical transmittance (69–82%), confirming a substantial reduction in internal attenuation. HEMA-based and hybrid HEMA-TEGDA systems show the highest values, whereas IBA and IBMA systems exhibit slightly lower transmittance. Haze ranges from 59 to 86%, reflecting the intrinsic scattering character of transparent wood. Pure IBA shows the highest haze, whereas IBA-IBMA mixtures show lower haze, demonstrating that polymer composition substantially affects the measured light-diffusion behavior.

Representative SEM micrographs of the cross-section of TW prepared with the IBA/IBMA 0.50 formulation are shown in Fig. 4. The SEM micrographs show extensive filling of the wood channels by the polymer and close interfacial contact between the infiltrated polymer phase and the wood cell walls, with no evident large interfacial gaps in the observed regions.

(a).

3.2 Estimation of optical transmittance and haze by ML models

For haze estimation, the primary performance assessment was based on complete-case LOOCV, including only samples with experimentally reported haze values and excluding both haze imputation and synthetic data augmentation. Under this configuration, the MAEs were 27.518, 23.597, and 22.826% points for MLP with kNN, RF, and GBT, respectively (Table S14). These results provide the primary estimates of haze-estimation performance because all target values correspond to experimental measurements. They also indicate that haze remains challenging to estimate using the descriptors available in the present dataset.

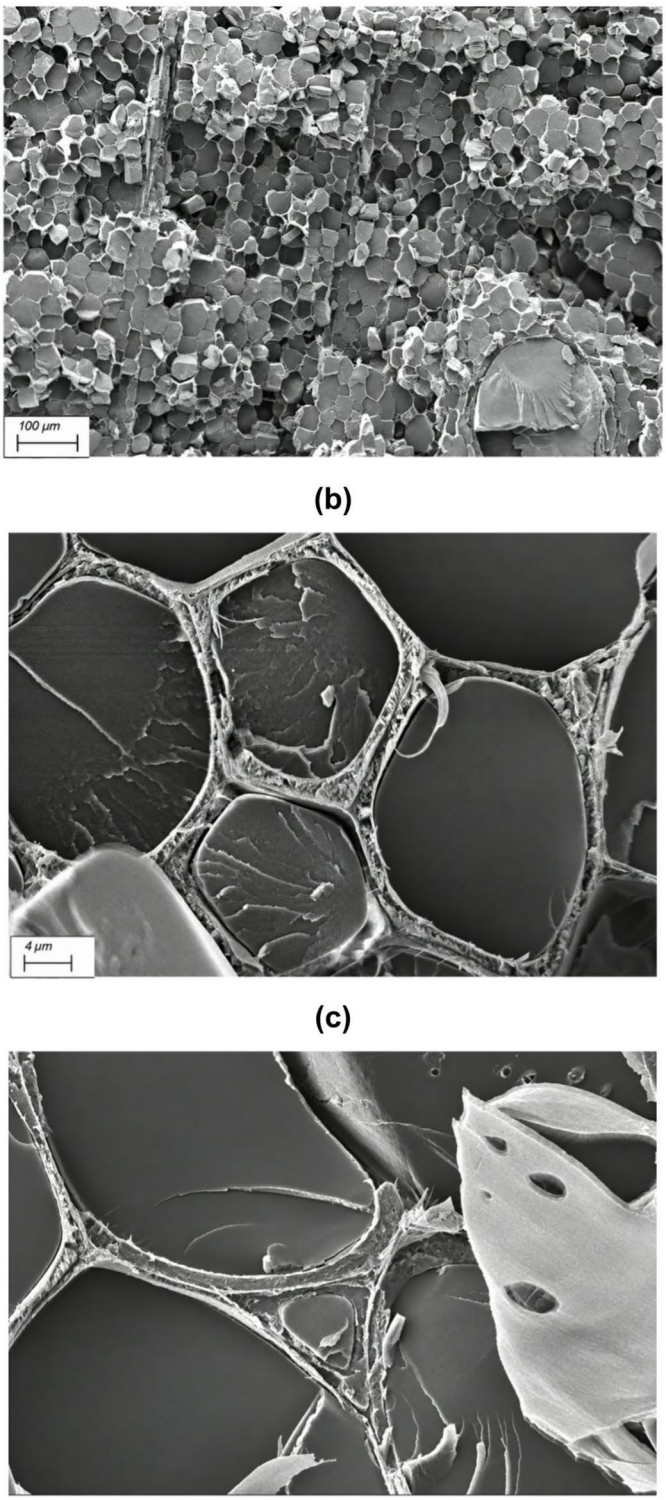

Fig. 4 SEM micrographs of the cross-section of TW based on the IBA-IBMA 0.50 formulation at different magnifications: (**a**) overview of the wood cellular structure; (**b**, **c**) polymer-filled channels and the polymer/cell-wall interface. The micrographs indicate extensive wood channel filling and continuous interfacial contact

Median-imputed and augmented configurations were subsequently examined as sensitivity analyses to assess the influence of missing-data treatment and controlled numerical perturbation on the model behavior. In the original 70/30 random-split analysis, results for optical transmittance estimation are reported in Table S10, whereas those for haze estimation are shown in Table S12. Within these sensitivity-analysis configurations, RF showed comparatively consistent haze-estimation MAE values, reaching a minimum of 12.135. Although this value is

lower than the errors obtained in the primary complete-case LOOCV analysis, it should not be interpreted as evidence of superior primary haze-estimation performance because these configurations include statistically imputed haze values. Rather, the difference underlines the sensitivity of model performance to missing-data treatment. For optical transmittance, GBT achieved the lowest MAE in an augmented configuration (10.635). These results should also be interpreted cautiously because augmentation increases the number of training instances rather than the number of independent experimental observations, and the random validation split does not assess extrapolation to independent material families.

To jointly assess validation performance and case-specific estimation performance on the target sample (i.e., balsa wood infiltrated with IBA-IBMA-0.75; Table S3), a quadrant-based analysis was performed. The results are shown in Figs. 5 and 6, where the validation MAE is represented on the y-axis and the Absolute Error (AE) obtained for the target sample on the x-axis. The lower-left quadrant identifies configurations combining low validation error with low target-sample estimation error. The quadrant analysis showed that models exhibiting low validation errors did not necessarily achieve the lowest estimation errors for the target sample, highlighting the importance of considering both validation-set performance and case-specific estimation performance when evaluating model behavior.

Across the investigated preprocessing sensitivity analyses, RF showed comparatively stable behavior, whereas GBT achieved the lowest errors in selected augmented configurations, particularly for the estimation of optical transmittance. However, the relative performance of the models depended on the preprocessing strategy, target property, and validation configuration. Therefore, differences among RF, GBT, and MLP should be interpreted as dataset- and workflow-dependent rather than as evidence of general superiority of a particular algorithm. The comparatively poorer performance of MLP is likely to rely on the limited amount of available training data and the higher data requirements of neural-network models.

LOOCV was also exploited to assess model behavior across the median-imputed and augmented configurations (Tables S14 and S15). For augmented configurations, identifier-based separation was maintained in each LOOCV iteration: all synthetic instances associated with the held-out Original_Sample_ID were excluded from the corresponding training fold. This procedure prevented direct information leakage between the held-out observation and its synthetic derivatives. Nevertheless, augmentation does not increase the number of independent

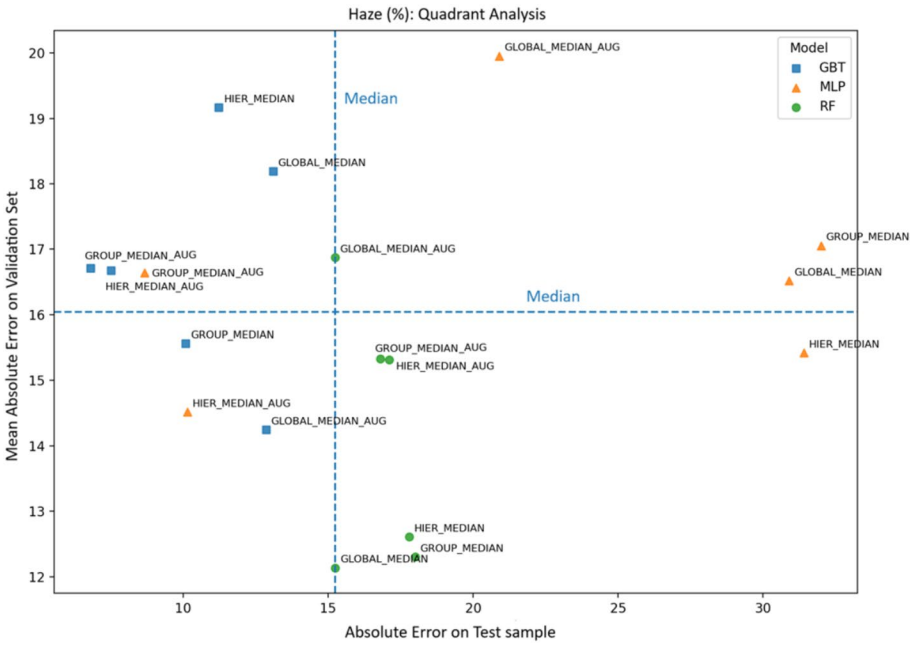

Fig. 5 Quadrant-based comparison of ML model performance for haze estimation. The horizontal and vertical dashed lines represent the median Absolute Error (AE) on the target sample and the median Mean Absolute Error (MAE) on the validation set, respectively. Models located in the lower-left quadrant exhibit both low validation error and low target-sample estimation error, representing the most balanced configurations in terms of validation-set performance and case-specific estimation accuracy

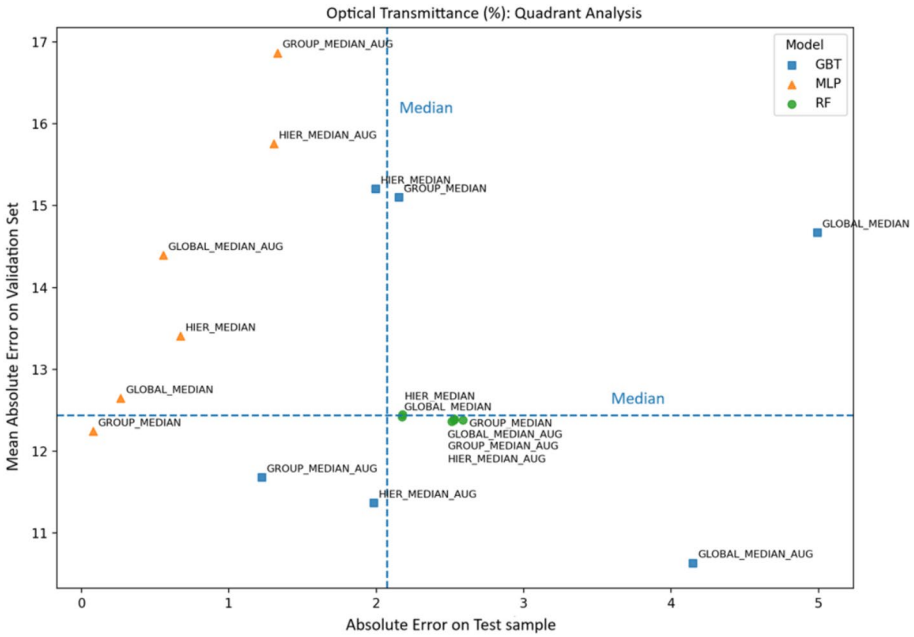

Fig. 6 Quadrant-based comparison of ML model performance for optical transmittance estimation. The horizontal and vertical dashed lines represent the median Absolute Error (AE) on the target sample and the median Mean Absolute Error (MAE) on the validation set, respectively. Models located in the lower-left quadrant exhibit both low validation error and low target-sample estimation error, representing the most balanced configurations in terms of validation-set performance and case-specific estimation performance

experimental observations: therefore, the resulting performance should be interpreted as a sensitivity assessment of increased training-data density rather than as evidence derived from additional experimental samples.

For optical transmittance estimation, LOOCV on the complete-case dataset yielded MAEs of 15.610, 11.578, and 13.100% points for MLP with kNN, RF, and GBT, respectively (Table S15). Across the sensitivity-analysis configurations, lower errors were observed in selected augmented datasets, with GBT reaching a minimum LOOCV MAE of 9.654 for the global-median augmented dataset. Because augmented instances were restricted to the training fold in each LOOCV iteration, these results indicate that controlled perturbation can influence model performance within the represented data domain. However, the observed reduction in error should not be interpreted as evidence of improved extrapolation to unseen material families because augmentation does not introduce new independent experimental systems.

The complete-case models were also applied to the IBA-IBMA-0.75 target sample to provide a case-specific within-family assessment. Optical transmittance estimation yielded absolute errors of 0.739, 3.940, and 0.190% points for MLP with kNN, RF, and GBT, respectively (Table S16), whereas haze estimation yielded absolute errors of 16.333, 16.761, and 18.576% points, respectively (Table S17). These results are consistent with the broader observation that haze is more difficult to estimate using the currently available descriptors. Because related IBA/IBMA formulations are represented in the dataset, this analysis should be interpreted as a case-specific within-family estimation exercise rather than as evidence of extrapolation to an unseen material chemistry.

Taken together, these analyses establish complete-case LOOCV as the primary assessment of haze-estimation performance because it fully relies on experimentally measured haze values. The median-imputed and augmented configurations provide complementary sensitivity analyses of missing-data treatment and controlled numerical perturbation, respectively, and are not interpreted as independent evidence of improved haze-estimation performance. The lower errors observed in some imputed configurations relative to the complete-case analysis highlight the sensitivity of model performance to target availability and preprocessing, while the higher complete-case errors underline the limitations imposed by the small number of experimentally complete observations and by the limited representation of scattering-related descriptors.

The potential influence of wood-template imbalance was further investigated by separately evaluating prediction errors for balsa and non-balsa observations (Table S18). The analysis included 28 balsa and 38 non-balsa observations. For haze estimation, the MAEs obtained for balsa samples were 30.255, 14.085, and 26.525% points for MLP with kNN, RF, and GBT, respectively, compared with 25.498, 28.545, and 32.614% points for

non-balsa samples. For optical transmittance, the corresponding MAEs were 10.582, 9.256, and 9.733% points for balsa samples and 19.314, 13.288, and 15.579% points for non-balsa samples, respectively. The subgroup analysis revealed differences in predictive performance between balsa and non-balsa observations, with their magnitude and direction depending on the target property and modelling configuration. In particular, all three models exhibited lower optical-transmittance prediction errors for balsa samples. For haze estimation, RF and GBT also showed lower errors for balsa, whereas MLP with kNN exhibited the opposite trend. These findings suggest that wood-template representation and the heterogeneity of the available experimental data may influence model performance, although the observed differences cannot be attributed exclusively to dataset imbalance. The subgroup analysis complements the previously reported LOOCV results by providing additional information on the distribution of prediction errors across wood templates. Nevertheless, the findings should be interpreted cautiously because the non-balsa category includes different wood templates, several of which are represented by only a small number of observations. Furthermore, sample-level LOOCV does not establish generalization to entirely unseen wood templates or independent material families.

Overall, the results support the feasibility of using supervised ML to investigate conditional relationships between the available descriptors and the optical responses of TW, while also highlighting substantial methodological limitations. Optical transmittance was generally estimated with lower error than haze, whereas haze remained more sensitive to data completeness and preprocessing choices. The present framework should therefore be regarded as an exploratory, conditional estimation approach rather than as a fully validated design tool. In particular, estimation of each optical property relies on knowledge of the complementary optical response, and the current analyses primarily quantify sample-level interpolation within the heterogeneous material domain represented by the available dataset. They do not establish generalization to entirely unseen polymer families or TW chemistries. Such extrapolation would require larger experimentally complete datasets, physically richer descriptors, prospective experimental testing, and group-aware validation based on sufficiently populated and consistently defined material families.

4 Conclusions

This study investigates the use of physically relevant descriptor within supervised ML for estimating the optical properties of TW. Using a curated literature-based dataset and systematic preprocessing, predictive models were evaluated for optical transmittance and haze. A key finding is the difference in predictability between the two properties. Optical transmittance generally exhibited lower prediction errors, whereas haze was more difficult to estimate, consistent with its greater sensitivity to microstructural heterogeneity and scattering mechanisms not represented in the available feature set. Among the evaluated algorithms, RF showed comparatively stable performance across several preprocessing configurations, whereas GBT achieved the lowest errors in selected configurations, highlighting the suitability of ensemble tree-based methods for the small, noisy, and heterogeneous datasets typical of transparent wood research. The results also emphasize the importance of preprocessing. In particular, median-based imputation enabled assessment of missing haze values, although the resulting predictions should be interpreted as sensitivity analyses rather than definitive evidence because imputed values are not experimental measurements. Selected augmented configurations yielded lower optical transmittance prediction errors, particularly for GBT, whereas the effect of augmentation on haze prediction was less consistent. Because augmentation increases training data density without introducing additional independent experimental observations, these results should be interpreted as evidence of preprocessing sensitivity within the represented data domain rather than as proof of improved extrapolative performance. The additional leave-one-out cross-validation analysis provides a complementary assessment that reduces dependence on a single train-validation partition. In particular, the complete-case haze baseline, restricted to experimentally measured haze values and excluding synthetic augmentation, provides a conservative assessment independent of both target imputation and synthetic observations. The higher haze errors obtained under this configuration further highlight the limitations

associated with the small number of experimentally complete samples. The corresponding optical transmittance analysis confirmed that predictive performance remains model dependent. Overall, these comparisons show that preprocessing effects should be interpreted on a model- and property-specific basis rather than being assumed to systematically improve predictive performance. At the same time, LOOCV remains a sample-level validation strategy; therefore, the present results do not establish extrapolative generalization to completely unseen material families.

Overall, these findings indicate that machine learning can complement experimental and theoretical approaches by identifying predictive relationships within heterogeneous TW datasets and supporting the prioritization of future experimental investigations. Future work should expand the dataset and incorporate descriptors related to microstructure, porosity, residual lignin or chromophores, density, refractive-index mismatch, void fraction, scattering-related imaging features, and processing conditions. Integrating imaging-derived descriptors, standardized optical measurements, and physics-informed learning approaches could improve both predictive performance and interpretability.

The present framework remains conditional since the estimation of each optical property requires the complementary optical response as an input; therefore, it cannot currently be exploited for fully ex ante materials optimization.

In future work, once models capable of estimating optical properties only from compositional, processing, and microstructural descriptors have been developed and rigorously validated, Bayesian optimization could be explored to propose formulations that maximize optical transmittance while maintaining application-specific haze levels. A key limitation of the present study is that model performance may not be maintained under leave-one-family-out validation. If entire polymer families, wood templates, or experimental series are excluded from training, prediction accuracy may decrease substantially, indicating that the framework is currently more suitable for interpolation within represented material families than extrapolation to new TW systems. Therefore, future studies should adopt stricter group-aware validation strategies, such as leave-one-reference-out or leave-one-polymer-family-out testing, to assess generalization beyond the material families represented in the dataset.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43939-026-01095-3.

Author contributions A. B., A.C. (Angelo Casciello), D.N., A.M., D.P. and G. M.: Resources, investigation, data curation, writing—original draft, reviewing. A.C . (Alessia Cabras) and G. V. prepared the transparent wood samples. D. P. and M. T. performed the optical characterization of transparent wood samples. All authors reviewed the manuscript.

Funding The AI-TRANSPWOOD project, HORIZON-CL4-2023-RESILIENCE-01-23 (Grant Agreement 101138191), is gratefully acknowledged. This project is co-founded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

Data availability The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Declarations

Ethics approval This research does not involve human or animal subjects, their data and/or biological materials.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Non-financial interests The authors declare no non-financial interests.

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

References

1. Hu X, Yu R, Wang F, et al. Fabrication, functionalities and applications of transparent wood: a review. Adv Funct Mater. 2023;33:2303278.
2. Jele TB, Andrew J, John M, et al. Engineered transparent wood composites: a review. Cellulose. 2023;30:5447–71.
3. Mariani A, Malucelli G. Transparent wood-based materials: A new step toward sustainability and circularity. Next Mater. 2024;5:100255.
4. Pereira Oliveira Moreira RL, Simao JA, Gouveia RF, et al. Exploring the Hierarchical Structure and Alignment of Wood Cellulose Fibers for Bioinspired Anisotropic Polymeric Composites. ACS Appl bio Mater. 2020;3:2193–200.
5. Pugliese D, Malucelli G. Toward the Optimization of the Optical Behavior of Transparent Wood: Current State of the Art and Perspectives. Polym (Basel). 2025;17:3276.
6. Malucelli G, Mariani A, Nuvoli D. (2026) The promise of transparent wood as a multifunctional energy material. Nat Rev Mater 1–2.
7. Yang H, Wang H, Cai T, et al. Light and wood: A review of optically transparent wood for architectural applications. Ind Crops Prod. 2023;204:117287.
8. Zhang J, Koubaa A, Tao Y, et al. The emerging development of transparent wood: Materials, characteristics, and applications. Curr Rep. 2022;8:333–45.
9. Hai L, Van, Srikanth N, Le TDT, et al. Transparent wood fabrication and applications: A review. Molecules. 2025;30:1506.
10. Pandit KH, Goswami AD, Holkar CR, et al. A review on recent developments in transparent wood: sustainable alternative to glass. Biomass Convers Biorefinery. 2025;15:6331–43.
11. Zhang N, Zhou X, Yin S, et al. Characterization of smart photochromic transparent wood with enhanced optical and mechanical properties. Polym Compos. 2026;47:458–69.
12. Terzopoulou P, Vouvoudi EC, Achilias DS. Delignification as a key strategy for advanced Wood-Based materials: Chemistry, delignification Parameters, and emerging applications. Forests. 2025;16:993.
13. Long S, Zheng D, Zhang M, et al. Preparation of photoluminescent transparent balsa wood with densifying structure for light-scattering windows based on cross-linking of lignocellulose and poly (vinyl alcohol). Int J Biol Macromol. 2025;311:143639.
14. Zhang Y, Zhou X, Zhang J, et al. Preparation process and performance study of transparent veneer treated with three metal salt-based deep eutectic solvent (MDES). Biomass Bioenergy. 2026;207:108774.
15. Lin L, Quan Q, Wang Q, et al. Multi-functional transparent wood: A novel approach to balance rigidity and sensitivity in smart materials. Chem Eng J. 2025;518:164630.
16. Guo M, Pan Y, Xu S, et al. Sustainable, transparent and fire-retardant wood composites. Ind Crops Prod. 2025;233:121332.
17. Zhou L, Liu S, Zhong T. A comprehensive review of optical diffusers: progress and prospects. Nanoscale. 2023;15:1484–92.
18. Yaddanapudi HS, Hickerson N, Saini S, et al. Fabrication and characterization of transparent wood for next generation smart building applications. Vacuum. 2017;146:649–54.
19. Chen H, Montanari C, Yan M, et al. Refractive index of delignified wood for transparent biocomposites. RSC Adv. 2020;10:40719–24.
20. Li Y, Wang S, Wang J, et al. Artificial Intelligence for Materials Science: Transforming Research Paradigms: Focus Review. Chem Rev; 2026.
21. Jiang X, Fu H, Bai Y, et al. Interpretable machine learning applications: a promising prospect of AI for materials. Adv Funct Mater. 2025;35:2507734.
22. Bai X, Zhang X. Artificial intelligence-powered materials science. Nano-micro Lett. 2025;17:135.
23. Zhong X, Gallagher B, Liu S, et al. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204.

24. Hwang S-W. Machine learning-driven research in wood science: from prediction to understanding through the framework of Wood Informatics. J Wood Sci. 2026;72:7.
25. Fathi H, Nasir V, Kazemirad S. Prediction of the mechanical properties of wood using guided wave propagation and machine learning. Constr Build Mater. 2020;262:120848.
26. Feng Y, Mekhilef S, Hui D, et al. Machine learning-assisted wood materials: Applications and future prospects. Extrem Mech Lett. 2024;71:102209.
27. Nasir V, Nourian S, Avramidis S, et al. Classification of thermally treated wood using machine learning techniques. Wood Sci Technol. 2019;53:275–88.
28. Nasir V, Fathi H, Fallah A, et al. Prediction of mechanical properties of artificially weathered wood by color change and machine learning. Mater (Basel). 2021;14:6314.
29. Chen S, Awano T, Yoshinaga A, et al. Flexural behavior of wood in the transverse direction investigated using novel computer vision and machine learning approach. Holzforschung. 2022;76:875–85.
30. Martinka J, Mitterpach J, Štefko T, et al. Fire hazard of epoxy-based transparent wood. J Therm Anal Calorim. 2023;148:9893–907.
31. Montanari C, Ogawa Y, Olsén P, et al. High performance, fully bio-based, and optically transparent wood biocomposites. Adv Sci. 2021;8:2100559.
32. Li Y, Fu Q, Rojas R, et al. Lignin-retaining transparent wood. Chemsuschem. 2017;10:3445–51.
33. American Standard for Testing Materials. (2021) ASTM D1003-21- Standard Test Method for Haze and Luminous Transmittance of Transparent Plastics. i:1–5.
34. Little RJA, Rubin DB. Statistical analysis with missing data. Wiley; 2019.
35. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–6.
36. Atkeson CG, Moore AW, Schaal S. Locally weighted learning. Artif Intell Rev. 1997;11:11–73.
37. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.
38. Friedman JH. (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 1189–232.
39. Berthold MR, Cebron N, Dill F, et al. KNIME-the Konstanz information miner: version 2.0 and beyond. AcM SIGKDD Explor Newsl. 2009;11:26–31.
40. Li Y, Fu Q, Yang X, Berglund L. Transparent wood for functional and structural applications. Philos Trans Math Phys Eng Sci. 2017;376:20170182.
41. Vasileva E, Chen H, Li Y, et al. Light scattering by structurally anisotropic media: a benchmark with transparent wood. Adv Opt Mater. 2018;6:1800999.

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

Authors and Affiliations

D. Nuvoli[1] · A. Bifulco[2] · A. Casciello[2] · A. Mariani[1] · A. Cabras[1] · G. Venusti[1] · D. Pugliese[3] · M. Terzi[4] · Giulio Malucelli[5]

✉ Giulio Malucelli
giulio.malucelli@polito.it

[1] Department of Chemical, Physical, Mathematical, and Natural Sciences, University of Sassari and local INSTM Unit, Via Vienna 2, Sassari 07100, Italy

[2] Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Piazzale Tecchio 80, Napoli 80125, Italy

[3] Division of Applied Metrology and Engineering, National Institute of Metrological Research (INRiM), Strada delle Cacce 91, Torino 10135, Italy

[4] Division of Quantum Metrology and Nanotechnologies, National Institute of Metrological Research (INRiM), Strada delle Cacce 91, Torino 10135, Italy

[5] Department of Applied Science and Technology, Politecnico di Torino, Viale Teresa Michel 5, Alessandria 15121, Italy