PHYSICS-INFORMED MACHINE LEARNING FOR PREDICTING OPTICAL TRANSMITTANCE AND HAZE IN TRANSPARENT WOOD

D. Nuvoli[1], A. Bifulco[2], D. Pugliese[3], A. Casciello[2], M. Terzi[4], A. Mariani[1], G. Malucelli[6]

[1] Department of Chemical, Physical, Mathematical, and Natural Sciences, University of Sassari, Via Vienna 2, Sassari 07100, Italy
[2] Department of Chemical, Materials and Production Engineering, University of Naples Federico II, 80125 Naples, Italy
[3] Division of Applied Metrology and Engineering, National Institute of Metrological Research (INRiM), Torino 10135, Italy
[4] Division of Quantum Metrology and Nanotechnologies, National Institute of Metrological Research (INRiM), Torino 10135, Italy
[5] Dept. of Applied Science and Tecnology, Politecnico di Torino, Viale T. Michel 5, 15121, Alessandria, Italy
e-mail: giulio.malucelli@polito.it

Introduction

Transparent wood (TW) has emerged as one of the most promising sustainable materials for advanced optical and energy-related applications. By combining the hierarchical structure of wood with transparent polymers, TW exhibits a unique balance of optical transmittance, light diffusion, low density, mechanical robustness, and thermal insulation. These characteristics make it attractive for smart windows, solar energy harvesting systems, energy-efficient buildings, and optoelectronic devices [1,2].
Despite significant advances in TW fabrication, optimization of optical properties is still largely based on empirical approaches. Optical transmittance and haze depend on a complex combination of variables including sample thickness, refractive index matching, residual lignin content, infiltration quality, and microstructural heterogeneity. As a result, establishing predictive structure–property relationships remains challenging [4].
Machine learning (ML) offers a powerful strategy for identifying hidden correlations within complex datasets and accelerating materials development. However, ML approaches have rarely been applied to transparent wood systems [3]. The present work proposes a physics-informed ML framework aimed at predicting optical transmittance and haze from physically meaningful descriptors, thereby supporting the rational design of transparent wood materials.

Results and discussion

A dataset containing 95 transparent wood samples was assembled from the scientific literature. The dataset covered different wood species, polymer infiltration systems, thicknesses, refractive indices, and optical performances. Particular attention was devoted to the selection of descriptors directly related to light transport phenomena, including thickness, refractive index, transmittance, and haze.
Because literature datasets are often incomplete and heterogeneous, robust preprocessing procedures were implemented. After performing an exploratory statistical analysis of the dataset, missing values were reconstructed through median-based imputation strategies, as the median was found to be more robust to outliers and better suited to the limited number of available observations than mean-based alternatives. Controlled Gaussian data augmentation was subsequently employed to increase statistical robustness without compromising physical consistency. Three supervised learning algorithms were investigated, namely (see Figure 1):

- Multilayer Perceptron (MLP);
- Random Forest (RF);
- Gradient Boosted Trees (GBT).

The predictive performance revealed a clear distinction between the two target properties. Optical transmittance was predicted with relatively high accuracy and stability across different preprocessing strategies. This result is consistent with the underlying physics, since transmittance is mainly controlled by macroscopic variables such as optical path length and refractive index matching.
Conversely, haze prediction proved significantly more challenging. Haze originates from complex multiscale scattering mechanisms and is strongly affected by structural features that were not explicitly included in the dataset, such as porosity, lumen morphology, local density fluctuations, and interface quality. Consequently, larger prediction errors were observed for haze regardless of the selected model.

Among all investigated algorithms, Random Forest consistently exhibited the most robust and reliable performance. Its ability to handle nonlinear relationships, noisy datasets, and limited data availability makes it particularly suitable for transparent wood research, where experimental datasets are often small and heterogeneous. Gradient Boosted Trees achieved competitive results, whereas the Multilayer Perceptron showed greater sensitivity to dataset size and preprocessing conditions. Overall, Random Forest provided the most reliable and consistent performance across the investigated scenarios, whereas Gradient Boosted Trees achieved the highest prediction accuracy in augmented configurations. On the other side, the Multilayer Perceptron with k-NN (k-Nearest Neighbors algorithm) appears significantly affected by the limited amount of available training data, resulting in less stable predictive performance.

The study also highlights the importance of data quality in materials informatics. While data augmentation improved transmittance prediction, its benefits for haze were limited, indicating that future improvements will require richer descriptors capable of directly representing microstructural characteristics.

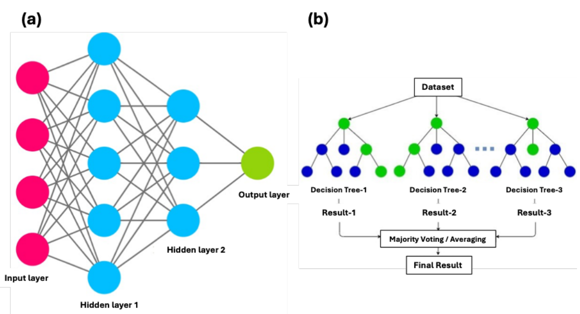

Figure 1. Some of the machine learning models employed for the prediction of Haze and Optical Transmittance.

Conclusions

A physics-informed machine learning framework was successfully developed for predicting optical transmittance and haze in transparent wood systems. The results demonstrate that physically meaningful descriptors can provide accurate predictions of optical transmittance, while haze prediction remains limited by the absence of explicit microstructural information.

Among the investigated models, Random Forest provided the best overall performance, confirming the suitability of ensemble tree-based methods for small and heterogeneous materials datasets. The proposed approach demonstrates the potential of machine learning to accelerate transparent wood development and establishes a foundation for future data-driven design strategies integrating optical characterization, microstructural analysis, and materials informatics.

References

[1] C. Chen et al., *Nat. Rev. Mater.*, vol. 5 642–666 (2020)

[2] G. Malucelli et al., *Nat. Rev. Mater.* (2026) 10.1038/s41578-026-00901-x

[3] A. Mariani and G. Malucelli, *Next Mater.*, vol. 5 5, 100255 (2024)

[4] D. Pugliese and G. Malucelli, Polymers, vol. 17, 3276 (2025)

[5] Li et al., *Chem. Rev.* vol.126, 6776-6803 (2026)

Acknowledgments

The authors gratefully acknowledge financial support from the AI-TRANSPWOOD project, funded under the European Union Horizon Europe programme (HORIZON-CL4-2023-RESILIENCE-01-23, Grant Agreement No. 101138191). This project is co-funded 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 the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.