Interpretable machine learning for predicting high humidity adsorption in tropical wood fibres: a case study from French Guiana

HAJJAR Hamzah [1], MICHAUD Franck[1], BAUDRIT Cédric[2], FERNANDEZ Christophe[2], MOREAU Jérôme[1]

[1] Laboratoire Innovation Matériau Bois Habitat Apprentissage (LIMBHA), Ecole Supérieure du Bois, 7 Rue Christian Pauc, F-44306 Nantes, France
[2] INRAE, USC 1368, Institut de Mécanique et d'Ingénierie (I2M), Université de Bordeaux, 351 Cours de la Libération, F-33405 Talence, France
hamzah.hajjar@esb-campus.fr

Keywords: Wood fibre insulation; machine learning; Moisture adsorption; Bio-based materials; French Guiana

Context and objective

The growing demand for sustainable construction materials in French Guiana has brought renewed attention to bio-based insulation panels manufactured from underutilized wood resources. However, their application in tropical environments remains challenging due to their sensitivity to moisture uptake, which directly affects dimensional stability, mechanical performance, and biological durability (fungal decay, termite attack). The interaction between wood fibres and ambient humidity is a critical factor in predicting long-term panel behaviour under tropical conditions where relative humidity can fluctuate between 50% and 98% (Kartal et al 2013).

Understanding and predicting the moisture adsorption capacity of wood fibres is therefore essential, as it influences panel performance during use. Specifically, adsorption at high relative humidity (e.g., 95% RH) is a key indicator for potential degradation risks in tropical construction, representing worst-case scenarios related to prolonged exposure to humid conditions.

Previous studies have shown that fibre properties such as granulometry, density, lignin content, hemicellulose content, ash content, and extractives composition play significant roles in moisture behaviour (Najafi et al 2011). However, there is a lack of existing mathematical or empirical models that can reliably link fibre properties to moisture adsorption, which highlights the need for a data-driven approach.

This study aims to (1) better understand the relationships between the chemical and physical properties of wood fibres and their water adsorption behaviour under high relative humidity (95%), to (2) predict the resulting moisture content, and to (3) identify the most influential features in the model's predictions.

This work contributes as a decision support system using fibre properties to estimate high-risk moisture adsorption behaviour (95% RH), supporting early-stage material selection for durable bio-based insulation in humid climates.

Material and methods

Data Description

This study uses a dataset of wood fibres derived from residual biomass resources in French Guiana. The fibres were sourced from three main classes of materials: industrial by-products from sawmills, fast-growing species from agricultural clearings, and high-potential plantation species. The dataset includes both chemical features (such as lignin, hemicellulose, soluble compounds, ash, inorganics, and extractives contents) and physical features (granulometry fractions and bulk density at 12% moisture content).

The output variable is the moisture adsorption at 95% relative humidity (Adsorption 95%A), which was measured using Dynamic Vapor Sorption (DVS). This specific humidity level was selected because it represents critical conditions for bio-based insulation materials in tropical environments, where prolonged high humidity can affect material durability.

Tab. 1: Input features and target variable for adsorption classification at 95% RH

Feature	**Symbol**	**Description**
Lignin	L	Lignin content (% of dry mass)
Hemicellulose	H	Hemicellulose content (% of dry mass)
Soluble compounds	S	Soluble compounds (% of dry mass)
Ash Content	A	Total ash content (%)
Inorganics	I	Inorganic matter excluding ash (%)
Hydrophilic Extracts	E_{phil}	Polar extractives (% of dry mass)
Hydrophobic Extracts	E_{phob}	Non-polar extractives (% of dry mass)
Fine Fraction	F_f	0–100 µm particle size (% mass fraction)
Medium Fraction	F_m	200–500 µm particle size (% mass fraction)
Coarse Fraction	F_c	>1000 µm particle size (% mass fraction)
Density	D	Bulk density at 12% moisture content (kg/m³)
Adsorption at 95% RH	Y	Moisture content at 95% RH (% - target)

Machine Learning Model and Workflow

Given the limited size and variability of the dataset, the problem was formulated as a binary classification task, aiming to predict whether a fibre exhibits high or low adsorption at 95% relative humidity (RH). The median adsorption value was used as the threshold to ensure balanced classes.

For such classification problem, several algorithms were tested, as illustrated in Fig. 1, including Logistic Regression, K-Nearest Neighbours (KNN), Random Forest, Support Vector Machine (Linear), Naive Bayes, Decision Tree, MLP Classifier, XGBoost, and CatBoost. Among these models, Logistic Regression was selected as it demonstrated the most consistent and reliable performance in this specific case.

This method is particularly suitable for small datasets, as it offers simplicity, robustness, and interpretability. Logistic regression outputs probabilities rather than binary labels, allowing the model to quantify the likelihood of high-water uptake based on the input features. Additionally, the learned coefficients directly represent the magnitude and direction of influence of each variable on the adsorption behaviour, providing transparent insights that can guide material design and optimization (Hosmer et al 2013).

Prior to model training, the dataset was standardized using z-score normalization to ensure all features contributed equally, regardless of their original scales. This prevents features with larger numerical ranges—such as density or granulometry—from dominating the model's learning process.

For model evaluation, a Leave-One-Out Cross-Validation (LOO-CV) procedure was used. This method is well-suited for small datasets, as it maximizes data usage by training the model on all samples except one in each iteration, ensuring an unbiased estimate of generalization error (Arlot and Celisse 2010).

The model's performance was assessed using standard classification metrics, including accuracy, along with a confusion matrix (Fig. 2) to visualize the prediction outcomes. This metric provides a comprehensive view of the model's reliability and practical relevance for supporting decision-making (Caruana and Niculescu-Mizil 2006)

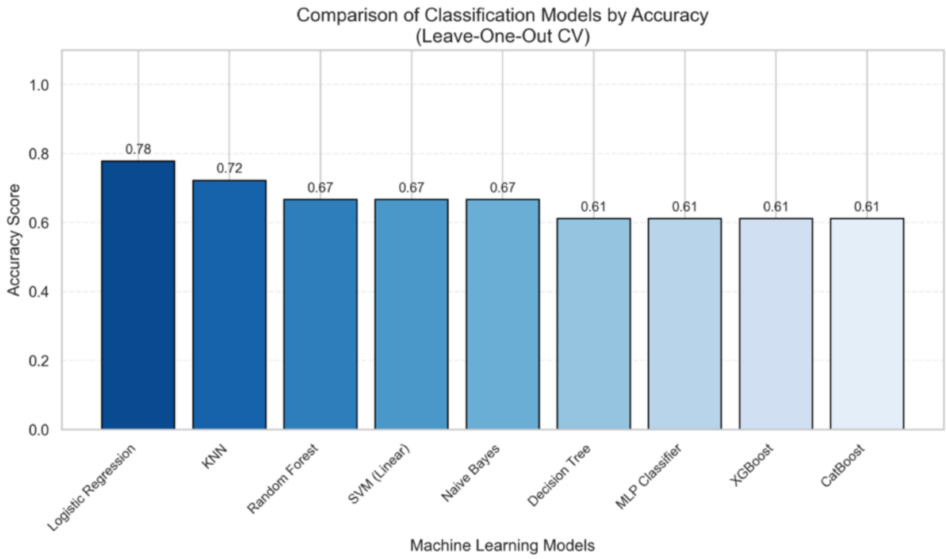

Fig. 1: Comparison of Machine learning Model's performance

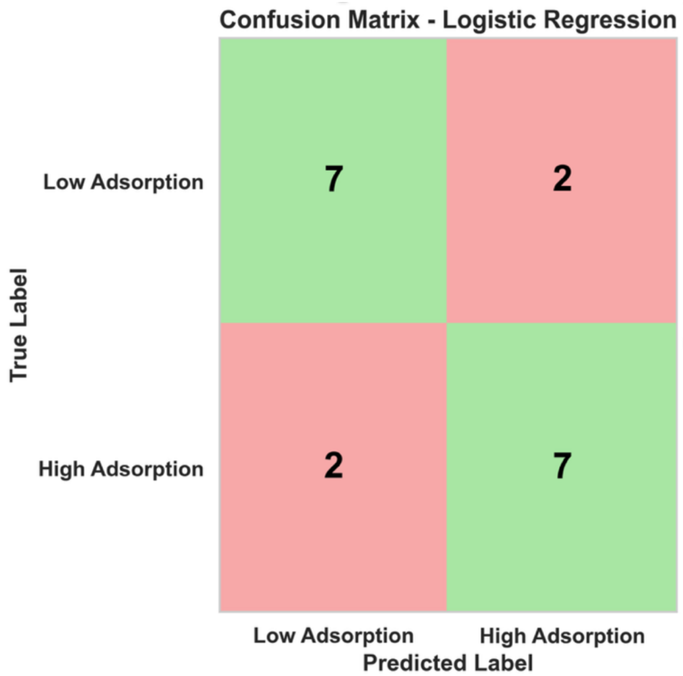

Fig. 2: Confusion matrix of the logistic regression model

Results and Discussion

The logistic regression model successfully classified fibre water adsorption behaviour at 95% relative humidity (RH), achieving an accuracy of 78%, with balanced precision and recall of 78% for both classes (high and low adsorption). This performance indicates that the model can reliably distinguish between fibres with higher and lower moisture uptake under critical tropical conditions.

Fig. 2 illustrates the model's predictive performance, showing its ability to distinguish between high and low adsorption classes. The classifier correctly predicted 14 out of 18 cases, corresponding to a balanced accuracy of 78%, with 2 misclassifications in each class.

Logistic regression coefficients were directly obtained from the trained model using scikit-learn (version 1.6.1) with the bilinear solver and L2 regularization. Each coefficient indicates how a feature contributes to the log-odds of high-water adsorption. We used the values of these coefficients (Fig. 3) to rank the features by their relative importance high moisture content, following standard practice in linear model interpretation (Hosmer et al 2013).

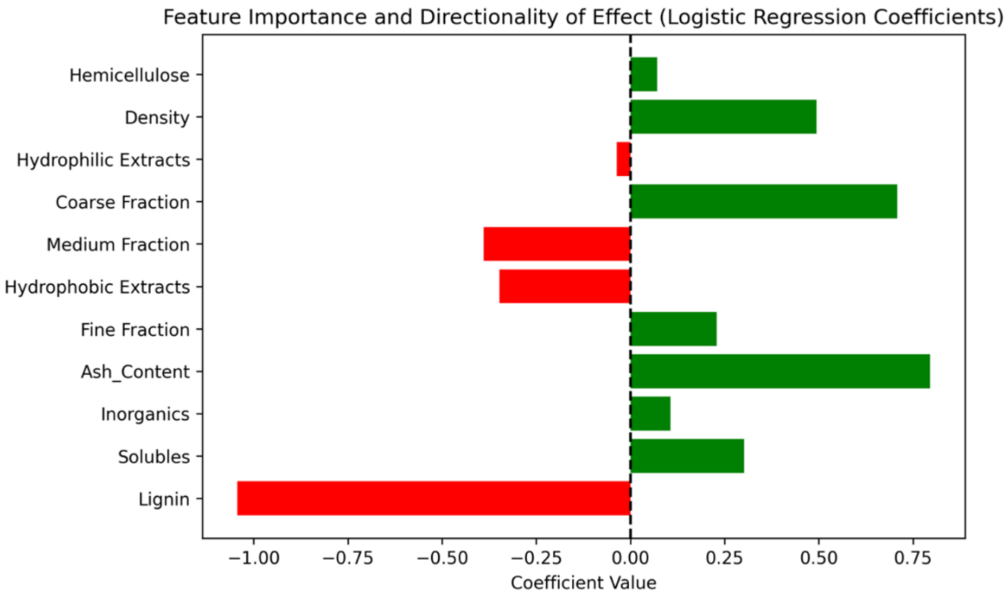

Fig. 3: Feature importance for predicting high moisture content using logistic regression

Key findings aligned with the literature

The model's feature importance results confirm several well-established relationships in wood science:

- Lignin content has the strongest negative effect on (high moisture content) water adsorption at 95% RH, consistent with its hydrophobic role in limiting fibre swelling and moisture uptake (Zhou et al 2016).
- Hydrophobic extractives reduce water adsorption, acting as natural moisture barriers by filling capillaries and blocking water access (Mantanis et al 1994), in our model, the low coefficient associated with extractive content suggests that their impact on the probability of high-water adsorption is limited.

- Soluble compounds and hemicellulose have a positive influence, consistent with their hygroscopic behaviour and role in increasing moisture absorption (Engelund et al 2013, Mantanis et al 1994).
- Ash content shows a positive effect on adsorption in our model, which may be explained by its role in increasing material porosity and surface area, enhancing adsorption efficiency (Li et al 2022).

Findings requiring careful interpretation

According to the results obtained through this study, coarse particles (>1000 µm) show a positive influence with adsorption, medium particles (200–500 µm) have a negative influence, while fine particles (0–100 µm) exhibit a positive effect. This non-linear pattern does not follow a clear trend and raises questions about the robustness of the observed relationships, suggesting the need for further investigation. Water adsorption of fibres usually follows an increase with lower granulometry (due to higher specific area) which is even more observed with DVS method (due to the very small amount contained in samples). This global observation may be explained by the multispecies characteristic of the sampling. For example, an oak shaving sample (high porosity due to vessels) can absorb more than a lower density wood such as pine flour (Keunecke et al 2008).

Density appears to have a positive influence on water uptake. In wood fibre insulation, higher density boards exhibited increased liquid water absorption, suggesting that denser fibres may facilitate more water uptake due to reduced porosity and increased capillary action (Snow et al 2024). This effect could be enhanced by grinding for small granulometry in particular (and by the counterbalanced usual density influence on adsorption for massive wood samples).

Conclusion and perspective

This study presents a preliminary data-driven approach to predict the water adsorption behaviour of bio-based fibres at 95% relative humidity, using chemical composition and granulometry as input features. The logistic regression model achieved an accuracy of 78%, while also offering interpretable insights into the material parameters driving moisture uptake. Findings confirm the dominant role of lignin and extractives in controlling adsorption, while highlighting the complex interactions between particle size distribution, ash content, and fibre density. Fine morphology and porosity aspects at fibre scale may also induce contradictory effects hidden by limited amounts of data and variability.

Importantly, this work emphasizes that machine learning is not a “magic” solution: the reliability of any model is deeply dependent on data quality, size, and variability. The relatively small and heterogeneous nature of our dataset inevitably introduces uncertainty and may explain some conflicting trends in feature importance, especially where multispecies samples and varying fibre morphologies are involved. Further research is needed to expand the dataset, reduce uncertainty, and validate the model across multiple humidity levels. Future work will also investigate multi-scale approaches to bridge the gap between fibre-level behaviour and panel-scale performance under real tropical conditions.

Acknowledgment

This research was conducted within the framework of the PANTHER project, with financial support from the French National Research Agency (ANR).

The authors would like to express their gratitude to the experts and researchers involved in the PANTHER project for their valuable contributions to data collection, technical discussions, and overall support throughout the study.

References

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection, Statistical Surveys, 4: 40–79.

Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, pp 161–168.

Engelund ET, Thygesen LG, Svensson S, Hill CAS (2013) A critical discussion of the physics of wood–water interactions, Wood Science and Technology, 47(1): 141–161.

Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied Logistic Regression. John Wiley & Sons, 510 p.

Kartal SN, Aysal S, Terzi E, Yılgör N, Yoshimura T, Tsunoda K (2013) Wood and bamboo-PP composites: Fungal and termite resistance, water absorption, and FT-IR analyses, BioResources, 8(1): 1222–1244.

Keunecke D, Sonderegger W, Niemz P (2008) Moisture-dependent sorption behaviour of differently treated wood particles, Wood Science and Technology, 42: 621–637.

Li D, Tao Y, Li S, Wu Y, Wang C, Lv Y, Zhu G, Qiu H, Liu X, Chen C (2022) Porous cage-like microfiber of fly ash magnetic powder (CMS)/polyaniline (PANI) composites with absorption properties, Physica Scripta, 97(8): 085817.

Mantanis GI, Young RA, Rowell RM (1994) Swelling of wood: Part 1. Swelling in water, Wood Science and Technology, 28(2): 119–134.

Najafi A, Khademi-Eslam H (2011) Effects of filler particle type/size on water absorption and swelling of HDPE composites, BioResources, 6(3): 2411–2424.

Snow J, Herzog B, O’Brien L, Li L (2024) Characterization of wood fiber insulation for the development of wood fiber-insulated panels (WIPs) for use in building envelope, BioResources, 19(3): 6142–6159.

Zhou H, Xu R, Ma E (2016) Effects of Removal of Chemical Components on Moisture Adsorption by Wood, BioResources, 11(2): 3110–3122