

Machine Learning-Based Prediction of Thermal Conductivity in Bio-Composite Materials

Hafida Er-rradi^{1*}, *Rachida Idchabani*¹ and *Imane Boucheфра*²

¹ Higher National School of Mines Rabat, Morocco

² Molecular Chemistry, Materials and Catalysis Laboratory, University Sultan Moulay Slimane, Beni Mellal, Morocco

Abstract. Bio-composite materials are increasingly used in the construction sector due to their low thermal conductivity and potential to improve building energy efficiency. Traditionally, thermal conductivity is evaluated through experimental testing or theoretical modeling, which can be time-consuming and costly. This study investigates the use of machine learning techniques to predict the thermal conductivity of bio-composite materials using 93 experimental samples extracted from the literature. Four regression models—Linear Regression, Ridge Regression, Support Vector Regression, and Random Forest—were trained using the same dataset and evaluated through an independent train–test validation strategy. Among the tested models, Random Forest achieved the best predictive performance, with a coefficient of determination $R^2 = 0.885$ and a root mean square error RMSE = 0.042, demonstrating its ability to capture nonlinear relationships in heterogeneous composite materials. These results highlight the potential of machine learning as a complementary tool for rapid estimation of thermal conductivity in sustainable building materials.

1 Introduction

Bio-composite materials represent an innovative and sustainable approach to building construction and thermal insulation. These materials are derived from biological sources such as plant fibers, wood, hemp, straw, and even animal sources, combined with a matrix material that binds the fibers together. The result is materials that not only contribute to thermal insulation but also embody environmental benefits, such as reduced carbon footprint, biodegradability, and the use of renewable resources. The development and characterization of bio-composite materials in building construction has been concern of many researchers in the last decade.

Thermal conductivity of materials is a crucial characteristics to enhance the energy efficiency in building. The thermal insulation properties of bio-composite materials are derived from their inherent structure and composition. The incorporation of bio-based additives derived from plant or animal sources into commonly used matrices has demonstrated a significant reduction in thermal conductivity compared to the binder alone [1]. This low thermal

* Corresponding author: h.er-rradi@enim.ac.ma

conductivity is essential for minimizing heat transfer between the interior and exterior of a building, thereby reducing energy consumption for heating and cooling.

The process of quantifying the equivalent thermal conductivity within porous materials encompasses a multitude of methodologies such as experimental campaign, numerical simulation or existing theoretical models in literature [2]. However, experimental approaches are often marred by high costs, both in terms of the financial investment required for sophisticated equipment and the time-consuming nature of meticulous experimental setup and execution. Numerical methods rely on a profound comprehension of the microstructure and are limited by the computational resources. In addition, the application of theoretical models is constrained by the assumptions and simplifications in their formulations. These models often presume idealized conditions (e.g., uniform particle distribution) that may not always provide accurate predictions across all types of porous materials.

Recent advancements in machine learning have opened new avenues for addressing these challenges. ML-based metamodeling, in particular, offers a promising framework for predicting material properties by learning from data rather than relying solely on physical or chemical laws. Recently, artificial intelligence has become a novel tool to predict the thermal conductivity by developing surrogate models based on machine learning algorithms and data analysis [3] [4] [5].

The present work addresses bio-composite materials for building applications using experimental data extracted from the literature. The contribution of this study lies in the comparative evaluation of several regression algorithms under identical conditions, highlighting the potential of machine learning models for predicting thermal conductivity in heterogeneous bio-composites.

2 Methodology and analysis

2.1 Data collection and analysis:

The dataset used in this study was compiled from experimental results reported in previously published literature on bio-composite materials for building applications. A total of 93 samples were collected from literature, covering a wide range of material configurations. A total of nine different natural materials were considered, including cork [6,7], hemp shiv [8], wood shavings [1], sawdust [9], alfa fiber [10], date palm fiber [11], coffee grounds [12], chicken feathers [13], and corn cob [14], covering a wide range of physical forms such as granular, fibrous, crushed, and non-compacted raw states.

The three most important factors in the field of bio-composite materials are thermal conductivity of the matrix (λ_{matrix}), thermal conductivity of the additive ($\lambda_{\text{additive}}$) and the volume fraction of the additive (v). Although thermal conductivity in bio-composites is strongly influenced by parameters such as porosity, apparent density, fiber structure and orientation, and moisture content, these variables were not consistently reported in the experimental literature used to construct the present dataset. Consequently, the input features were limited to the thermal conductivity of the matrix (λ_{matrix}), the thermal conductivity of the additive ($\lambda_{\text{additive}}$), and the volume fraction (v), which are the most commonly available and physically relevant parameters.

The thermal conductivity of the bio-based additives ($\lambda_{\text{additive}}$) ranges from 0.033 to 0.20 $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$, reflecting the diversity of material composition and structure.

Prior to model training, the dataset was examined for consistency and outliers. All variables were normalized to improve numerical stability and model convergence. To evaluate the predictive performance and generalization capability of the machine learning models, the dataset was randomly divided into 80% for training and 20% for testing.

2.2 The correlateion between variables and the target

Investigating the relationship between the predictors and the thermal conductivity within composite materials, figure 1 reveal distinct relationships. Volume fraction v demonstrates a negative correlation with thermal conductivity, evident results expressed by the diminution of thermal conductivity with increasing the amount of additive [10,11,15] . Notably, the thermal conductivity of the matrix λ_{matrix} exhibits a positive correlation with the thermal conductivity, aligning with theoretical expectations that the matrix's conductive properties significantly affect the composite's overall thermal characteristics. For the thermal conductivity of the additive $\lambda_{additive}$, there's a wide spread in the values of the thermal conductivity for similar values of the additive's thermal conductivity. The clustering of data points observed indicates that the relationship is influenced by other variables, notably the volume fraction.

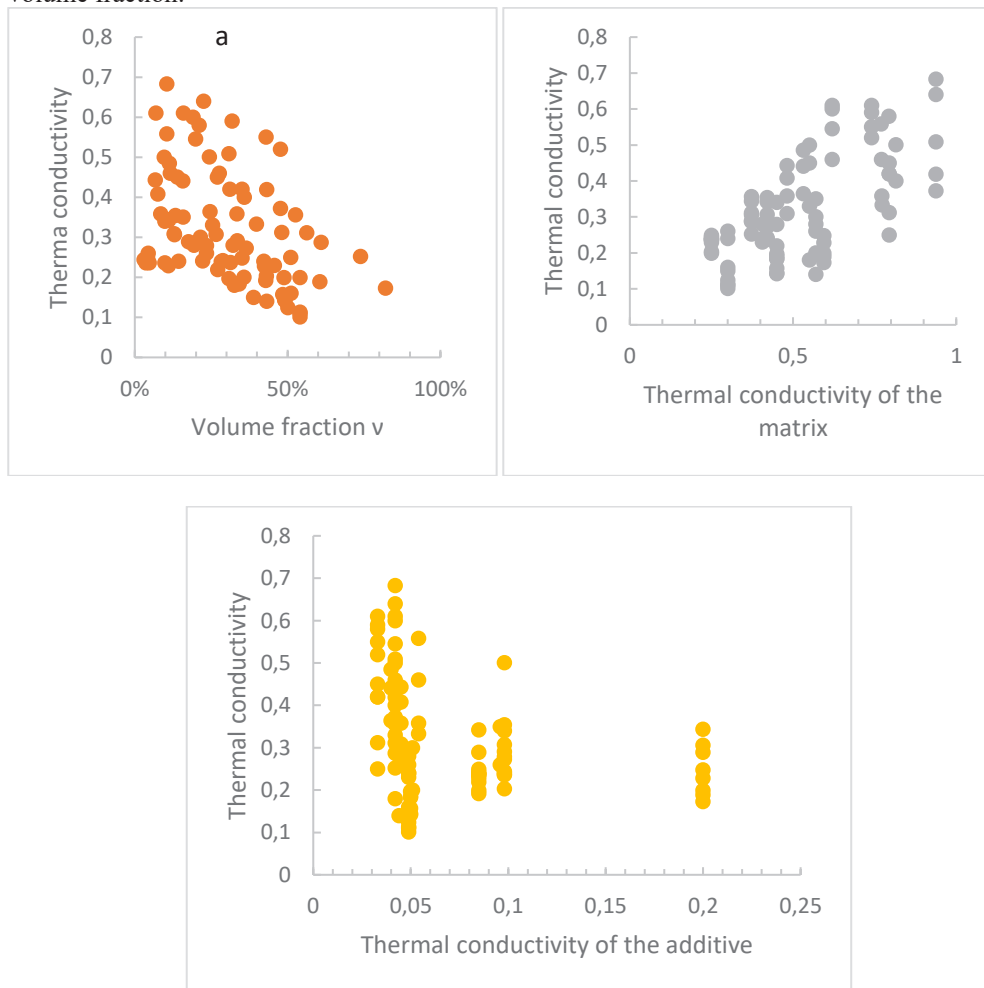


Figure 1: relationship between several factors and thermal conductivity

2.3 Machine learning algorithms:

In this study, we examined four different regression algorithms: Linear Regression, Ridge Regression, Support Vector Regression (SVR), and Random Forest. These models were selected because they represent a broad spectrum of learning strategies, ranging from simple linear relationships to complex nonlinear patterns. Linear Regression and Ridge Regression provide interpretable baseline models, with Ridge adding regularization to improve stability in the presence of multicollinearity. SVR introduces a flexible margin-based approach capable of capturing nonlinear behaviors through kernel functions. Random Forest, as an ensemble of decision trees, offers strong predictive power and robustness by averaging multiple models to reduce variance.

Two statistical coefficients were used to evaluate the effectiveness of the models in estimating the target variable. The coefficient of determination (R^2) measures how well the predicted values explain the variability of the actual data, with higher values indicating better predictive accuracy. The Root Mean Square Error (RMSE), on the other hand, quantifies the average magnitude of prediction errors, giving more weight to large deviations. An R^2 value of 1 indicates perfect correlation between the predicted and actual values, while a value closer to 0 suggests that the model fails to accurately capture the variance of the data. In the context of this study, a higher R^2 value signifies a metamodel's superior capability in predicting thermal conductivity.

$$R^2 = 1 - \frac{(\sum_{i=1}^n (y_i - \hat{y}_i)^2)}{(\sum_{i=1}^n (y_i - \bar{y})^2)} \quad \text{eq.1}$$

Where:

y_i is the actual observed outcome,

\hat{y}_i is the predicted outcome from the model,

\bar{y} is the mean of the observed data, and n is the number of observations.

RMSE is the square root of the mean squared error. It measures the standard deviation of the residuals, providing insights into the concentration of the data around the line of best fit.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{eq.2}$$

Where:

n is the number of observations,

y_i is the actual observed value,

\hat{y}_i is the predicted value by the model.

In this study, a train–test split strategy was adopted to evaluate the predictive capability of the machine learning models. The dataset was divided into independent training and testing subsets. Although k-fold cross-validation and hyperparameter optimization can further improve robustness, the main objective of this work was to perform a comparative evaluation of commonly used regression algorithms under the same conditions. All models were trained using identical datasets and default hyperparameters to ensure a fair comparison.

3 Results and discussion:

Figure 2 illustrate the results of the performance metrics of used metamodels. In this study, we assessed the performance of four models, including Linear Regression (LR), Ridge Regression (RR), Support Vector Regression (SVR), and Random Forest (RF). Notably, Random Forest (RF) emerged as the standout performer, exhibiting the highest coefficient of determination (R^2) value of 0.885, indicating its exceptional ability to explain approximately

88,5 % of the variance in the data. Furthermore, RF demonstrated the lowest root mean squared error (RMSE), underscoring its superior predictive accuracy and robustness in capturing nonlinear relationships and interactions within the dataset. Similar trends have been reported by Li et al. [5], where ensemble-based algorithms outperform linear and kernel-based models in predicting thermal conductivity of heterogeneous materials.



Figure 2: Performance metrics for trained ML algorithms

Furthermore, figure 3 visualize a comparative analysis of different predictive models in estimating the thermal conductivity of bio-composite materials. Linear models exhibit a degree of predictive accuracy, as denoted by their respective proximity to the line of perfect prediction within the scatter plots. However, they demonstrate limitations in encapsulating the non-linear intricacies in the dataset. In contrast, the Support Vector Regression (SVR) presents a marked improvement in the central range of values, with notable deviations at the extremities, suggesting challenges in modeling the full spectrum of the dataset's variance. Random Forest model display a superior alignment with the ideal predictive line, indicating a robust capacity for modeling the complex, non-linear interactions present in the thermal properties of bio-composite materials. The superior performance of the Random Forest model can be attributed to its ability to capture nonlinear relationships and complex interactions between input variables. Unlike linear and kernel-based models, Random Forest relies on an ensemble of decision trees that can model hierarchical decision rules and variable interactions without assuming a predefined functional form. This characteristic is particularly well suited for bio-composite materials, whose thermal behavior results from heterogeneous structures and non-uniform phase distribution

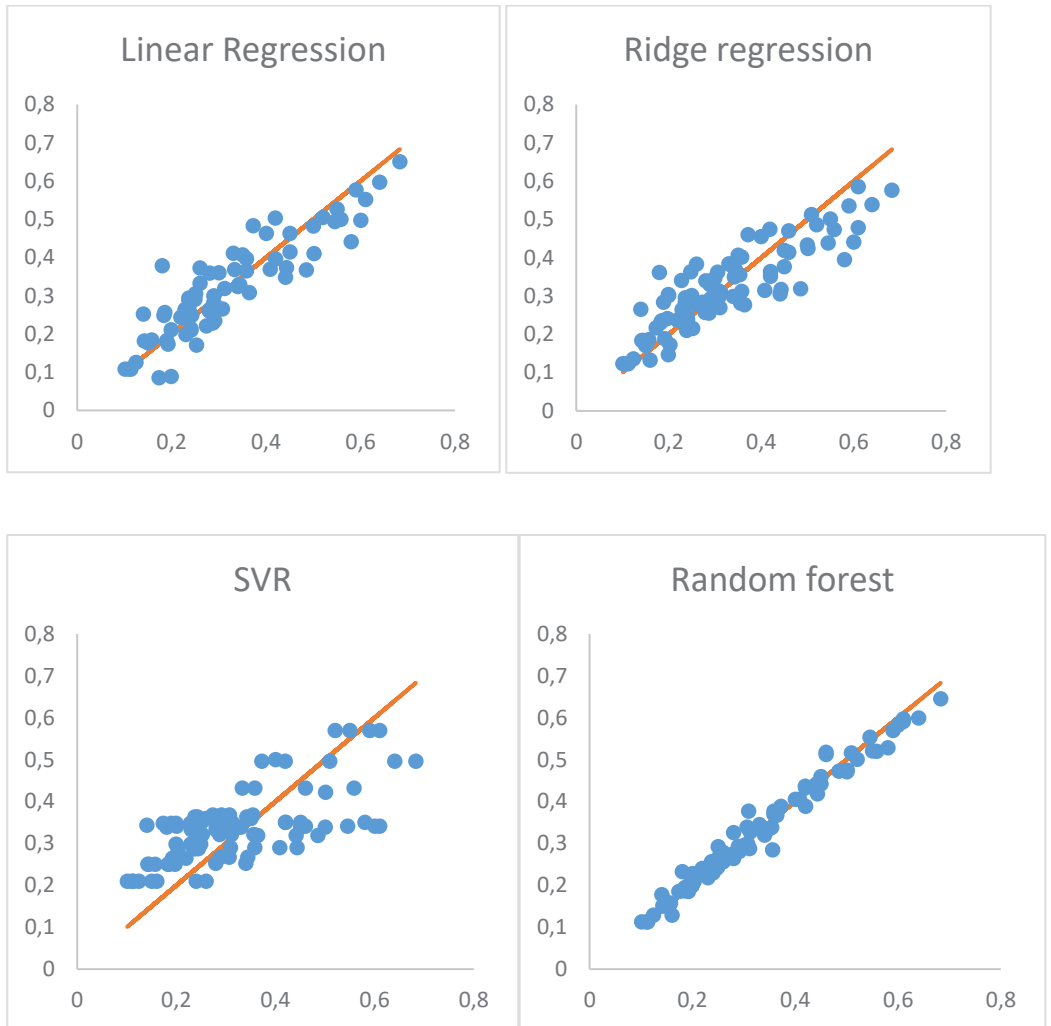


Figure 3: Experimental thermal conductivity versus predicted thermal

Conclusion

This study highlights the potential of machine learning models for predicting the thermal conductivity of bio-composite materials using experimental data from the literature. Among the evaluated algorithms, Random Forest showed the best predictive performance, confirming its ability to capture nonlinear relationships in heterogeneous materials.

However, the proposed approach is limited by the restricted set of input parameters and the validation strategy adopted in this work. From a practical point of view, the developed models

can provide engineers with a rapid estimation tool for preliminary design and selection of eco-friendly building materials.

Future developments will focus on improving model robustness and physical relevance by incorporating additional material descriptors, enhancing model interpretability, and performing a direct comparison with classical analytical models to better assess the added value of machine learning approaches.

References

- [1] D. Taoukil, A. El-bouardi, H. Ezbakhe, T. Ajzoul, Thermal proprieties of concrete lightened by wood aggregates, *Res. J. Appl. Sci. Eng. Technol.* 3 (2011) 113–116.
- [2] Maxwell, *A treatise of electricity and magnetism*, 3rd Ed. New-York Dover (n.d.).
- [3] H. Béji, T. Kanit, T. Messenger, Prediction of Effective Elastic and Thermal Properties of Heterogeneous Materials Using Convolutional Neural Networks, *Appl. Mech.* 4 (2023) 287–303. <https://doi.org/10.3390/applmech4010016>.
- [4] H. Wei, S. Zhao, Q. Rong, H. Bao, Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods, *Int. J. Heat Mass Transf.* 127 (2018) 908–916. <https://doi.org/10.1016/j.ijheatmasstransfer.2018.08.082>.
- [5] K.Q. Li, Y. Liu, Q. Kang, Estimating the thermal conductivity of soils using six machine learning algorithms, *Int. Commun. Heat Mass Transf.* 136 (2022). <https://doi.org/10.1016/j.icheatmasstransfer.2022.106139>.
- [6] A. Cherki, Experimental thermal properties characterization of insulating cork–gypsum composite, *Constr. Build. Mater.* 54 (2014) 202–209.
- [7] F.Z. El Wardi, A. Khabbazi, A.B. Cherki, A. Khaldoun, Thermomechanical study of a sandwich material with ecological additives, *Constr. Build. Mater.* 252 (2020) 119093. <https://doi.org/10.1016/j.conbuildmat.2020.119093>.
- [8] A. Hussain, J. Calabria-Holley, M. Lawrence, Y. Jiang, Hygrothermal and mechanical characterisation of novel hemp shiv based thermal insulation composites, *Constr. Build. Mater.* 212 (2019) 561–568. <https://doi.org/10.1016/j.conbuildmat.2019.04.029>.
- [9] M.J. Morales-Conde, C. Rodríguez-Liñán, M.A. Pedreño-Rojas, Physical and mechanical properties of wood-gypsum composites from demolition material in rehabilitation works, *Constr. Build. Mater.* 114 (2016) 6–14. <https://doi.org/10.1016/j.conbuildmat.2016.03.137>.
- [10] M. Touil, A. Lachheb, R. Saadani, M.R. Kabiri, M. Rahmoune, A new experimental strategy assessing the optimal thermo-mechanical properties of plaster composites containing Alfa fibers, *Energy Build.* 262 (2022) 111984. <https://doi.org/10.1016/j.enbuild.2022.111984>.
- [11] M. Boumhaout, L. Boukhattem, H. Hamdi, B. Benhamou, F. Ait Nouh, Thermomechanical characterization of a bio-composite building material: Mortar reinforced with date palm fibers mesh, *Constr. Build. Mater.* 135 (2017) 241–250. <https://doi.org/10.1016/j.conbuildmat.2016.12.217>.
- [12] M. Touil, A. Lachheb, R. Saadani, M. Sanbi, A. Talidi, M. Rahmoune, Experimental investigation on the combined effect of the water mixing ratio and the addition of spent coffee grounds on plaster’s thermo-mechanical properties, *Therm. Sci. Eng. Prog.* 36 (2022). <https://doi.org/10.1016/j.tsep.2022.101488>.
- [13] M. Ouakarrouch, N. Laaroussi, M. Garoum, Thermal characterization of a new bio-composite building material based on plaster and waste chicken feathers, *Renew. Energy Environ. Sustain.* 5 (2020) 2. <https://doi.org/10.1051/rees/2019011>.

- [14] J. Pinto, D. Cruz, A. Paiva, S. Pereira, P. Tavares, L. Fernandes, H. Varum, Characterization of corn cob as a possible raw building material, *Constr. Build. Mater.* 34 (2012) 28–33. <https://doi.org/10.1016/j.conbuildmat.2012.02.014>.
- [15] H. Er-radi, R. Idchabani, M. Oubrek, C. Bojji, L. Boukhattem, A. Doublali, A. Jilbab, Experimental study of composite materials based on ecological wastes, *J. Compos. Mater.* 56 (2022) 3295–3306. <https://doi.org/10.1177/00219983221113029>.