Journal Pre-proof

Explainable machine learning for multiscale thermal conductivity modeling in polymer nanocomposites with uncertainty quantification

Bokai Liu, Pengju Liu, Yizheng Wang, Zhenkun Li, Hongqing Lv, Weizhuo Lu, Thomas Olofsson, Timon Rabczuk

PII: S0263-8223(25)00457-X
DOI: https://doi.org/10.1016/j.compstruct.2025.119292
Reference: COST 119292

To appear in: *Composite Structures*

Received date : 16 January 2025
Revised date : 2 April 2025
Accepted date : 15 May 2025

Please cite this article as: B. Liu, P. Liu, Y. Wang et al., Explainable machine learning for multiscale thermal conductivity modeling in polymer nanocomposites with uncertainty quantification. *Composite Structures* (2025), doi: https://doi.org/10.1016/j.compstruct.2025.119292.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier Ltd.

Journal Pre-proof

Explainable Machine Learning for Multiscale Thermal Conductivity Modeling in Polymer Nanocomposites with Uncertainty Quantification

Bokai Liu[a,b], Pengju Liu[b], Yizheng Wang[c,a], Zhenkun Li[d], Hongqing Lv[e], Weizhuo Lu[b], Thomas Olofsson[b], Timon Rabczuk[a,*]

[a] *Institute of Structural Mechanics, Bauhaus-Universität Weimar, Marienstr. 15, D-99423 Weimar, Germany*
[b] *Intelligent Human-Buildings Interactions Lab, Department of Applied Physics and Electronics, Umeå University, 901 87 Umeå, Sweden*
[c] *Department of Engineering Mechanics, Tsinghua University, Beijing, 100084, China*
[d] *Department of Civil Engineering, Aalto University, Rakentajanaukio 4 A, 02150 Espoo, Finland*
[e] *Yantai Research Institute of Harbin Engineering University, Harbin Engineering University, Yantai, 264000, China*

Abstract

Graphene-based polymer nanocomposites show great potential for thermal management, but accurately predicting their thermal conductivity remains challenging due to multiscale structural complexity and parameter uncertainty. We propose an innovative approach integrating interpretable stochastic machine learning with multiscale analysis to predict the macroscopic thermal conductivity of graphene-based polymer nanocomposites. Our bottom-up framework addresses uncertainties in meso- and macro-scale input parameters. Using Representative Volume Elements (RVEs) and Finite Element Modeling (FEM), we compute effective thermal conductivity through homogenization. Predictive modeling is powered by the XGBoost regression tree-based algorithm. To elucidate the influence of input parameters on predictions, we employ SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), providing insights into feature interactions and interpretability. Sensitivity analyses further quantify the impact of design parameters on material properties. This integrated method enhances prediction accuracy, reduces computational costs, and bridges data-driven and physical modeling, offering a scalable solution for designing advanced composite materials for thermal management applications.

Keywords: Thermal properties, Stochastic multi-scale modeling, Interpretable Integrated Learning, Sensitivity Analysis, Polymeric graphene-enhanced composites (PGECs)

1. Introduction

Nowadays Polymeric graphene-enhanced composites (PGECs) have attracted significant attention for their exceptional heat transfer properties like thermal conductivity, achieved by incorporating graphene as inclusion into polymers matrix. Graphene's high intrinsic thermal conductivity dramatically enhances the thermal properties of polymers, which are typically poor heat conductors [1, 2]. This enhancement is crucial for industries like electronic devices, aerospace and mechanical engineering, thermal management, automotive industry,civil engineering, and energy storage devices[3]. In electronic devices, PGECs improve heat dissipation, enhancing reliability and lifespan [4], while in energy storage, such as battery electrodes or supercapacitors, increasing energy storage capacity and device lifespan [5]. The excellent heat transfer capability of graphene, due to its flexibility and large surface area, facilitates enhanced thermal conductivity in PGECs, which can be optimized by adjusting concentration of graphene fillers, the shape, and thee size [6]. the selection of the polymer matrix plays a crucial role, with polyethylene-based PGECs showing high thermal conductivity resulting from the low thermal resistance at the interface between the filler-matrix.

To understand the underlying mechanisms, we employ multi-scale modeling, a computational approach that integrates fine and coarse scales to connect properties and study material behavior. This method bridges the gap from micro-structures to marco behaviors in materials, from the atomic level to the macroscopic level [7]. This approach facilitates the understanding of how microscopic defects can propagate to affect macroscopic performance, the role of grain boundaries in determining thermal and electrical properties, and the interaction between different phases within a composite material. This comprehensive multiscale analysis provides valuable insights for designing materials with tailored properties for specific applications.

Considering the inherent uncertainties in properties, materials behaviors, manufacturing processes, and environmental conditions, stochastic multiscale modeling is essential. These uncertainties can contribute to inconsistency in material performance and behavior, affecting thestructures' safety and reliability in engineering. Stochastic multiscale modeling helps in identifying critical factors contributing to material failure, optimizing material design for enhanced durability,

*Corresponding Authors. E-mail address: timon.rabczuk@uni-weimar.de

and developing more robust engineering practices. By integrating probabilistic data with physical modeling, this method enhances the predictability and resilience of materials used in complex engineering applications, ultimately leading to safer and more efficient designs.

Various studies have explored the use of graphene and its derivatives to enhance the thermal properties of polymeric materials. Luo et al. conducted a molecular dynamics study to investigate the enhancement of thermal energy transport across graphene/graphite and polymer interfaces [8]. They found that the addition of graphene improved the thermal conductivity of the composite material. Zain-ul-abdein et al. presented an experimental and computational investigation of thermal conductivity enhancement in a bakelite-graphite composite [9]. The study demonstrated the effect of graphite addition on the effective thermal conductivity of the composite material. Similarly, Zhang et al. explored the mechanical and thermal properties of hierarchical composites enhanced by pristine graphene and graphene oxide nanoinclusions [10]. In a study by Wang et al., silver nanoparticle-deposited boron nitride nanosheets were used as fillers in polymeric composites to enhance thermal conductivity [11]. The results showed that the composite with silver nanoparticle-deposited boron nitride nanosheets outperformed the one with boron nitride nanosheets alone, due to lower thermal contact resistance among the interfaces. Olowojoba et al. reported on the thermal and mechanical properties of in situ thermally reduced graphene oxide/epoxy composites [12]. The study highlighted the potential of graphene as a filler material in the development of multifunctional polymeric composites. Guo et al. prepared graphene/poly(vinylidene fluoride) composites to improve thermal conductivity, which could lead to the further development of thermal conductive polymeric materials [13]. Fang et al. proposed a method to bond graphene foam with polydimethylsiloxane to fabricate composites with high thermal and mechanical properties [14]. The results showed considerable improvement in thermal conductivity and insulativity, making these composites suitable for heat management in electronic devices. Tu et al. reported on the enhancement of latent heat and thermal conductivity in polyethylene glycol-based composites, demonstrating significantly improved thermal conductivities [15]. Finally, Azizi et al. investigated the performance improvement of EPDM and EPDM/silicone rubber composites using modified fumed silica, titanium dioxide, and graphene additives [16]. The incorporation of these additives led to a significant increase in thermal stability and thermal conductivity of the composites. Overall, these studies highlight the potential of graphene-enhanced composites to improve thermal conductivity in polymeric materials.

The thermal conductivity of composites at the macroscopic level is widely acknowledged to be greatly affected by various factors. For instance,Shokrieh et al. introduced a stochastic multiscale model for CNT composites, considering uncertain parameters like CNT length, orientation, agglomeration, curvature, and dispersion [17]. Similarly, Vu-Bac et al. have contributed [18], and we proposed an uncertainty analysis method for PNC stochastic modeling [19] [20] [21] [22].

While stochastic multiscale modeling is capable of directly predicting the intended values, it is computationally expensive, especially when conducting extensive simulations for a wide range of input parameters or during iterative optimization processes. To mitigate the high computational burden associated with stochastic multiscale models, surrogate methods have been devised for propagating uncertain parameters across scales [23] [24]. Machine learning has emerged as a powerful surrogate models tool in the field of composite materials, offering new perspectives and accelerating the discovery of functional composites [25]. Gu et al. applied machine learning algorithms to predict mechanical properties of composites, such as toughness and strength, showcasing the efficiency and accuracy of this approach [26]. Similarly, Wei et al. took a different approach by using machine learning methods to predict the effective thermal conductivities of composite materials, deviating from traditional methods based on physical understanding of heat transfer mechanisms [27]. Chen et al. summarized recent progress in the applications of machine learning to composite materials modeling and design, highlighting the potential of different ML algorithms to accelerate composite research [28]. Dabetwar et al. focused on damage classification of composites using machine learning techniques, aiming to improve condition monitoring and health management strategies of composite materials [29]. Khan et al. reviewed the use of machine learning for damage assessment of smart composite structures, emphasizing the importance of discriminative features and various machine learning algorithms for detecting, quantifying, and localizing damage [30]. Marani et al. utilized machine learning to predict the compressive strength of PCM-integrated cementitious composites, achieving superior prediction accuracy and providing insights into materials science aspects [31]. Kharb et al. applied machine learning to analyze and predict the erosion behavior of silicon carbide reinforced polymer composites, using SVM and MLR approaches to validate optimized response characteristics [32]. Furtado et al. introduced a methodology to generate design allowables of composite laminates using machine learning, focusing on Legacy Quad Laminates and double-double laminates for composite aerostructures [33]. In our previous research, we also used a hierarchical multi-scale model with physical laws (Physics-Informed), so called Physics-Informed Neural Networks (PINNs), to obtain thermal properties and optimize material design by linking microscale phenomena with macroscale properties [34].

While these studies primarily evaluate the performance of algorithms and models, they significantly lack in explaining the internal mechanisms and interpretability. The models are unable to scientifically explain their predicted results. A common critique of these data-driven techniques in properties prediction and materials design is their characterization as black-box models, lacking transparency and failing to offer understanding into the foundational physical processes that control the issue under examination. Researchers are striving to develop more transparent machine learning (ML) models to better understand their relationships and sensitivities among different scales with input and output parameters. One effective method is employing explainable techniques in AI w.r.t XAI, which can strive towards enhance the interpretabil-

Journal Pre-proof

ity and transparency of data-driven training models [35]. A notable approach in this realm is SHAP (SHapley Additive exPlanations), which provides local explanations for predictions by quantifying the contribution of each feature compared to the dataset average. Another valuable method in Explainable AI (XAI) is LIME (Local Interpretable Model-agnostic Explanations), which generates interpretable explanations for individual predictions by approximating the local decision boundary of any complex model with a simpler, interpretable model [36]. Those methods not only enhance the interpretability of machine learning models but also facilitates deeper insights into how individual features influence model predictions across diverse applications and datasets. This enables us to handle larger datasets and explore parameter spaces that may be computationally prohibitive for direct stochastic modeling as well as better interpretation of ML models and is becoming increasingly popular in practice [37, 38].

This study introduces a novel approach by incorporating explainable AI (XAI) techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations). These methods enhance the transparency of the machine learning models, allowing for deeper insights into how different input features influence the predicted thermal conductivity. This shift in focus from pure prediction accuracy towards interpretable and transparent models is a key novelty of this work.

Furthermore, by combining stochastic multiscale modeling with machine learning and explainable AI, this research presents a comprehensive approach that not only predicts thermal conductivity with high accuracy but also provides actionable insights into the factors driving these predictions. This step towards transparent and interpretable models represents a significant advancement over previous methodologies, offering better control and understanding of material design processes for PGECs.

Based on the above highlights, this study aims to pioneer a machine learning model with uncertainty analysis in a multi-scale approach that provides precise predictions of the thermal conductivity of nanocomposites containing Polymer and graphene. It seeks to establish a clear and interpretable quantitative relationship by leveraging advanced techniques such as the SHAP and LIME methods and sensitivity analysis. By integrating these approaches, the research aims to enhance understanding of how different factors influence thermal conductivity in PGECs, thereby advancing the optimization and design of these innovative materials for various industrial applications. Building upon our previously published paper [39], this study at first enhances the interpretability of the data model using SHAP and then ally LIME to explore the interpretability and finally introduces sensitivity analysis to quantify the significance of the physical model. These additions bolster the credibility of the overall model predictions. The article outlines the general methodology (Section 2), introduces materials (Section 3), discusses data-driven machine learning models and inner interpretability (Section 4), presents simulation analysis and numerical results (Section 5), and concludes (Section 6).

2. Methodology of research

We present a novel machine learning-based multi-scale stochastic model represents a significant advancement in computational modeling and analysis. By integrating stochastic multi-scale modeling and machine learning methodologies, we offer a comprehensive framework capable of addressing complex phenomena spanning various length scales. The hierarchical structure of our approach, as illustrated in Fig. 1, facilitates seamless connections between meso and macro scales. Beginning with a bottom-up paradigm, we ensure the robust propagation of information while effectively managing uncertainties inherent in multi-scale systems. This hierarchical framework orchestrates the integration of outputs from finer scales as inputs for subsequent coarser scales, enabling a holistic understanding of the system under study. Moreover, the utilization of machine learning algorithms adds a layer of sophistication to our model, empowering it to analyze and interpret outputs from the stochastic multi-scale framework with unprecedented depth and accuracy. By leveraging the capabilities of machine learning, we unlock new avenues for insight generation and knowledge discovery within complex multi-scale systems.

This comprehensive methodology encompasses three pivotal stages:

1) **Bottom-up modeling**: Beginning with detailed modeling at the fine scale, ensuring comprehensive understanding and information transfer.

2) **Stochastic modeling**: Incorporating stochastic elements to capture inherent uncertainties and variations across scales.

3) **Data-driven methods**: Leveraging machine learning techniques to analyze and extract meaningful insights from the complex multi-scale data generated.

The overall methodological framework is illustrated in Figure 2. The framework integrates stochastic multiscale modeling with explainable machine learning techniques for accurate and interpretable prediction of the effective thermal conductivity of polymer nanocomposites.

Figure 1: Multi-scale modeling scheme

Figure 2: Integrated framework combining stochastic multiscale modeling and explainable machine learning.

3. Stochastic multi-scale modeling

3.1. Multi-scale approach

A hierarchical multiscale method is employed in a bottom-up approach, where the materials information is transferred progressively from finer scales to coarser scales. This method ensures that detailed data from microscale analyses inform and refine the models at meso and macro scales. Figure 3 illustrates the flowchart of this process. In the following sections, we will describe the models used at each length scale, detailing how they integrate and contribute to the overall multiscale framework. This approach allows for a comprehensive understanding of the material properties, ensuring that microscale phenomena accurately influence the larger scale behaviors and predictions.

Figure 3: Multi-scale modeling flowchart

We apply continuum models in meso-scale with Representative Volume Elements (RVEs). This multi-scale modeling approach can effectively represent properties by incorporating a limited number of inclusions/fillers inside cubic matrix. A commonly used cubic RVE, depicted in Figure 4, incorporates graphene fillers, which are approximated as disks. To automate the generation of these RVE models, we employ Abaqus platform integrated with a custom Python script. This script, which operates according to a three-dimensional non-collision rules implemented by C++ programming[6], ensures accurate placement and distribution of the fillers within the RVE. The positioning of the graphene fillers is determined

according to the probability density functions (PDFs) of the input parameters, ensuring a realistic and statistically representative model. This integration of advanced software and custom algorithms allows for precise modeling and analysis of the composite material, ultimately enhancing the accuracy and reliability of the multi-scale simulation results.

To ensure physical consistency and eliminate artificial boundary effects during the finite element simulations of the RVE, periodic boundary conditions (PBCs) are applied. These boundary conditions are essential for capturing the behavior of an infinite periodic composite and for accurate homogenization of the material's effective thermal properties.

In the mechanical domain, corresponding nodes on opposite faces of the cubic RVE are constrained such that their displacements remain compatible, thereby ensuring continuity across boundaries. For the thermal simulations, a uniform heat flux is applied to one face of the cube, while the opposite face receives an equal and opposite flux, resulting in a consistent temperature gradient across the domain.

The implementation of PBCs is automated through a custom Python script integrated within the Abaqus environment. The script uses multi-point constraint (MPC) techniques to apply the necessary coupling between node pairs on opposite faces of the RVE. This automated framework supports high-throughput RVE generation and simulation with varying filler distributions, aspect ratios, and volume fractions.

At the meso-scale, we study graphene sheet agglomerations and dispersions, common with high aspect ratios and volume fractions. To quantify agglomeration, we use a two-parameter method generating spheres as gathered zones (Fig 5), labeled as 'inclusions', dividing the space into $V_{Graphene}^{inclusion}$ and $V_{Graphene}^{matrix}$ [40]:

$$V_{Graphene} = V_{Graphene}^{inclusion} + V_{Graphene}^{matrix} \tag{1}$$

where the graphene fillers located in the matrix and represented by the terms $V_{Graphene}^{inclusion}$ and $V_{Graphene}^{matrix}$, respectively.

Besides, the degree to which the graphene fillers cluster together within the composite material and the uniformity of filler distribution throughout the matrix should also be considered. So we define them in two indices - agglomeration index ξ and the dispersion index:

$$\xi = \frac{V_{inclusion}}{V}, \qquad \zeta = \frac{V_{Graphene}^{inclusion}}{V_{Graphene}} \tag{2}$$

The agglomeration shown in an index ξ indicates the filler's volume fraction with respect to cubic RVE's volume size, serving as an indicator of how much of the composite's volume is occupied by graphene fillers. The dispersion presented in an index ζ measures the graphene sheets' volume size within fillers with respect to the plates' volume, reflecting degree of dispersion of the graphene within the inclusions. Uniform disk distribution, where there is no agglomeration, is observed if $\xi = \zeta$. This equality signifies that the graphene fillers are evenly dispersed throughout the RVE. Conversely, if $\xi > \zeta$, it suggests an uneven distribution of disks within the RVE, indicating that the graphene fillers are clustering together, leading to agglomeration. This agglomeration can negatively affect the composite's properties, such as reducing thermal conductivity and mechanical strength [41, 42].

The initial step in determining an appropriate RVE size involves employing data boosting called sample enlargement method (SEM). The SEM technique entails progressively boosting the cubic elements' size until the effective thermal conductivity under homogenization converges to a predefined value. Averaging the thermal conductivity data across numerous samples can be regarded as the convergence criterion, shown:

$$\langle R \rangle = \frac{1}{M} \sum_{K=1}^{M} R^{(K)} \tag{3}$$

In this context, this criterion ensures that the averaged homogenized thermal conductivity has stabilized, indicating that the RVE size is sufficiently large to capture the material's representative behavior. Regarding the equation, $R^{(k)}$ donates the current data e within the k-th RVE, with M denoting the entire amount of RVEs. After averaging the ensemble, it is essential to meet a convergence criterion to establish an appropriate size as representative:

$$\left| \frac{\langle R^{(K+1)} \rangle - \langle R^{(K)} \rangle}{\langle R^{(K)} \rangle} \right| < Tol = 1\% \tag{4}$$

Here in this equation, $R^{(k+1)}$ contributes to the $k+1-th$ representative volume element. The convergence is typically assessed by monitoring the changes in thermal conductivity values as the RVE size increases and ensuring that these changes fall below a predefined threshold. This rigorous process guarantees the accuracy and reliability of the RVE model, making it a critical step in multiscale modeling and simulation.

Then we start to discuss the heat transfer problem which is shown below and governed by

$$C_f \frac{\partial \theta}{\partial t} + \nabla \cdot \boldsymbol{q} - Q = 0 \tag{5}$$

Journal Pre-proof

In this equation, θ denotes the absolute temperature reflecting the material's thermal state. C_f is the heat capacity quantifying the intensity and distribution of internal heat sources. C_f is the heat capacity determining the material's ability to store thermal energy. $\boldsymbol{q}$ is the heat flux vector describing the rate and direction of heat flow within the material.

Due to the quasi-steady cases, by simplifying the analysis, the time-dependent term $C_f \frac{\partial \theta}{\partial t}$ is typically disregarded. By substituting Fourier's law inside, we can derive a more manageable form for steady-state heat transfer scenarios:

$$div(\kappa \nabla \theta) + Q = 0 \qquad in \quad \Omega \tag{6}$$

where κ represents the thermal conductivity, θ is the temperature field, and Q denotes the internal heat source. The natural boundary conditions are defined as:

$$q_n = -\boldsymbol{q} \cdot \boldsymbol{n} = \overline{q} \qquad on \quad \Gamma_q \tag{7}$$

where $\boldsymbol{n}$ is the normal vector to the boundary Γ_q, and $\overline{q}$ is the specified heat flux at this boundary.

To derive the weak form of this equation, we multiply the governing equation by a test function $\delta\theta$ and integrate over the domain Ω. This process transforms the partial differential equation into an integral form that is more suitable for numerical methods such as the finite element method (FEM). The weak form is expressed as follows:

$$\int_{\Omega} \kappa \nabla \theta \cdot \nabla \delta\theta d\Omega = -\int_{\Gamma_q} \delta\theta \overline{q} d\Gamma + \int_{\Omega} \delta\theta Q d\Omega \quad \forall \delta\theta \in \nu_0 \tag{8}$$

Here, θ is the trial function representing the temperature field, and $\delta\theta$ is the test function belonging to the function space ν_0. This formulation ensures that the temperature distribution satisfies the heat transfer equation in an average sense across the domain.

Figure 6 illustrates the application of two distinct preset heat fluxes from each single one direction across the representative cube, generating a corresponding temperature gradient. Then, we can obtain the homogenized effective thermal conductivity of composite by employing Fourier's law. Fourier's law is given by:

$$\mathbf{q} = -\boldsymbol{\kappa} \cdot \nabla \theta \tag{9}$$

$$\text{with} \quad \boldsymbol{\kappa} = \begin{Bmatrix} \kappa_{xx} & 0 & 0 \\ 0 & \kappa_{yy} & 0 \\ 0 & 0 & \kappa_{zz} \end{Bmatrix} \tag{10}$$

In a composite material, the macroscopic thermal conductivity is determined by ensuring $\kappa_{xx} = \kappa_{yy} = \kappa_{zz}$, reflecting isotropic behavior [6]. This is achieved by imposing boundary conditions at different cube's edges and solving heat equation's weak form. The composite conductivity κ obtained from this process serves as the meso-scale output, providing essential information for further multiscale modeling and analysis.

Figure 4: The RVE cube and Meshing

Figure 5: The agglomeration and dispersion in the Cubic RVE

At the macroscopic scale, we adopt a larger, homogenized structure to account for uncertainties. This process discretizes the entire material domain and randomly distributes cubes with varying thermal properties within the macroscale domain, each representing discretized RVEs [43, 44]. These cubes are derived from the mesoscale simulation, throughout

the macroscale terrain, as shown in 7. While the Finite Element Method (FEM) could theoretically be applied at the macro-scale, we opt for the rule of mixture due to its computational efficiency. The Voigt model is employed to interconnect multiple Representative Volume Elements (RVEs). This approach is characterized by:

$$k_{\text{eff}} = \bar{k} = \frac{\sum_i k_i V_i}{\sum_i V_i} \tag{11}$$

In this scenario, k_{eff} is the macroscopic effective thermal conductivity; k_i represents the thermal properties of the i-th RVE cube. The weighting factor V_i, denoted as the volume fraction, influences the contribution of each cube.

Figure 6: Applying heat flux in both sides

Figure 7: The material region in macro-scale modeling

The Voigt model is employed here to calculate the macroscopic effective thermal conductivity because it offers a computationally efficient method that assumes uniform temperature gradients across different phases, making it suitable for composite materials with stochastic multiscale properties. This method estimates an upper bound for the effective thermal conductivity, making it suitable for composite materials like ours, where each RVE has varying thermal properties. While more detailed methods like the Finite Element Method (FEM) could provide higher accuracy, the Voigt model is preferred for its simplicity and efficiency in large-scale, uncertainty-driven simulations. The computational efficiency is crucial due to the stochastic nature of our multiscale modeling approach. By leveraging the volume fractions of each RVE, the Voigt model provides a balance between computational cost and accuracy, making it ideal for the stochastic nature of the multiscale modeling used in this study.

3.2. Stochastic modeling

At the material properties scale, variations in thermal conductivity, tensile strength, and elasticity can significantly affect the performance of the final product. At the microstructural level, the arrangement and distribution of grains, phases, and defects within the material can lead to unpredictable behavior under stress or thermal conditions [45, 46, 47]. Furthermore, manufacturing processes introduce another layer of uncertainty, as variations in processing parameters such as temperature, pressure, and cooling rates can result in inconsistencies in material properties and microstructures. So those uncertainties across various scales, including material properties, microstructures, and manufacturing processes are examined in this study shown in Table 1. Meanwhile, Boundary conditions are handled in a deterministic manner to maintain consistency and predictability. In contrast, stochastic analysis is employed for uncertain input parameters by assigning probability density functions (PDFs) to each parameter, which outline their means and variances.

Before starting stochastic modeling, a large number of data samples need to be obtained through reasonable sampling methods. Here we apply the Latin Hypercube Sampling (LHS) method to obtain dataset which is a statistical technique used to efficiently model the stochastic behavior of systems with multiple uncertain input parameters [48]. It aims to enhance the accuracy and efficiency of sampling compared to traditional methods such as Monte Carlo Sampling [49]. The LHS method works by generating a preset design matrix of size $N \times m$, including m input parameters as well as N denotes intervals that divide the cumulative probability curve. Each input parameter's range is divided into N equally probable intervals, ensuring that each interval is sampled exactly once. This stratification guarantees that the entire range of each input parameter is explored thoroughly. The standard deviation, mean, and variance of the output parameters are obtained by conducting those probability density functions (PDFs) and then providing a statistical description of the system's behavior under uncertainty. Subsequently, the design matrix containing a large number of data sets generated by LHS is imported into the materials model to obtain the actual target outcome values, following the method described by [50]. To further alleviate computational expenses associated with uncertainty analysis, surrogate models are employed. These surrogate models, often constructed using machine learning techniques or simplified physical approximations, serve

Journal Pre-proof

as efficient proxies for the more complex and computationally intensive physical models. By providing rapid evaluations, surrogate models allow for extensive exploration of the input parameter space without the need for repeated, costly simulations of the full physical model.

Table 1: Model uncertainties

Scale	Inputs	mean	standard deviation	Type of distribution	Sources
Meso	Thermal conductivity of fillers	3978.85	580.79		Suchismita, et al. [51]
	Thermal conductivity of Matrix	0.252806	0.098237		A. Moisala et al. [52]
	Interface resistance	20.2356	5.9294		M.Freitag et al.[53]
	Aspect ratio	59.1911	123.2185		Khoa Bui et al. [54]
	Agglomeration index	0.55105	0.26127		Nam Vu-Bac et al. [18]
	Dispersion index	0.55010	0.26154		Nam Vu-Bac et al. [18]
Macro	Volume fraction	0.054164	0.025811		M.Shokrieh,et al.[17]

4. Data-driven method

4.1. Materials and Dataset Preparation

The database for machine learning should contain at least training set and test set. In this case, the raw materials from dataset, obtained from FE simulations in physical models at the meso-scale in RVE and MATLAB macroscopic numerical analysis. All those results are divided into an training set with 80% portion and a 20% test set. Normalization of this data is essential to standardize it, reducing computational costs and enhancing robustness. The performance and accuracy of the models are assessed using metrics like the root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE), all of which are derived from model residuals, shown below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{ri} - Y_{pi})^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_{ri} - Y_{pi})^2}{\sum_{i=1}^{N}(Y_{ri} - Y_{mean})^2} \tag{13}$$

$$MAE = \frac{\sum_{i=1}^{N}|Y_{ri} - Y_{pi}|}{n} \tag{14}$$

Journal Pre-proof

Lastly, interpretability need to be considered and discussed according to previous steps. Shapley Additive Explanations as an interpretable approach is applied to make transparent and explain the model predictions by providing a unified measure of the contribution of each input parameter to the output predictions, allowing for a clear understanding of how different factors influence the model's behavior, focusing on feature importance and Shapley values. By assigning Shapley values to each feature, SHAP quantifies their individual impact, highlighting which parameters are most critical in driving the model's decisions.

4.2. Cross Validation and Hyper-parameters tuning

After data preparation, It is also necessary to consider parameters optimization during model training. Cross-validation (CV) is pivotal in machine learning approach, offering a reliable method for hyperparameter estimation and model construction. By segregating data into training and validation sets, CV guards against overfitting, ensuring the model's generalizability. Among CV techniques, K-Fold cross-validation reigns supreme. Here, the dataset is divided into k subsets or 'folds', with the model trained iteratively on k-1 folds and validated on the remaining fold. This iterative process yields a comprehensive evaluation of model performance. By aggregating metrics across folds, K-Fold CV provides a robust estimate of model efficacy on unseen data, especially beneficial for smaller datasets. Moreover, K-Fold CV helps alleviate bias and variance by testing the model across multiple subsets, enhancing its reliability in real-world applications.

Hyper-parameter tuning is a critical aspect of machine learning model development, where optimal values for parameters that cannot be learned during training, known as hyper-parameters, are selected. This process significantly impacts model accuracy and performance. However, it can be challenging and time-consuming due to the iterative trial-and-error nature of identifying the best hyper-parameter values. Despite these challenges, hyper-parameter tuning is essential for maximizing model efficacy by carefully balancing model complexity and performance. In this study, we utilize Particle Swarm Optimization (PSO) for hyper-parameter tuning, a method that has shown success in previous research. The PSO algorithm continuously minimizes the Sum of Squared Errors (SSE) during the tuning process, employing a 10-fold cross-validation approach to ensure robustness [20][21]. The equation is present:

$$SSE = \frac{1}{N}\sum_{i=1}^{N}(Y_{ri} - Y_{pi})^2 \tag{15}$$

where N is the number of output parameters; Y_{ri} is the required output parameters and Y_{pi} is the predicted $i-th$ values. In this study, a swarm size of 450 is selected based on our previous study, as well as the inertia weight ω set to 1 and the cognitive and social coefficients c_1 and c_2 set to 2.0 each, as referenced in [21]. These parameters are chosen to optimize the balance between exploration and exploitation in the optimization process, ensuring effective convergence to the global optimum while maintaining diversity within the swarm.

4.3. Regression-Tree-based approach: XGBoost

The divide-and-conquer strategy is applied in tree-based model to iteratively split data based on if-then conditions, with key elements including tree depth, complexity, and segmentation points. However, one single regression tree often fails to capture the full data complexity, leading to potential overfitting or underfitting issues. To address this, XGBoost, an advanced ensemble learning method, combines multiple trees to enhance robustness and accuracy, incorporating regularization techniques and efficient processing to manage complex data interactions and high-dimensional feature spaces.

XGBoost, short for eXtreme Gradient Boosting, stands as a cornerstone in modern machine learning, renowned for its exceptional performance in predictive tasks across diverse domains. Leveraging a gradient boosting framework coupled with decision trees, XGBoost offers unparalleled accuracy and efficiency, making it a popular choice among data scientists and researchers alike. In this section, we delve into the methodology behind XGBoost, exploring its robustness, versatility, and ability to handle complex datasets. We detail its underlying principles, parameter optimization strategies, and model evaluation techniques, showcasing how XGBoost empowers practitioners to tackle real-world challenges effectively. Through this comprehensive introduction, we aim to provide insights into the power and potential of XGBoost in driving advancements in predictive modeling and data-driven decision-making.

In recent years, XGBoost has revolutionized the landscape of machine learning with its unparalleled performance and versatility. Standing at the forefront of predictive modeling techniques, XGBoost combines the strengths of gradient boosting algorithms with decision trees, unleashing a formidable tool for data analysis and predictive tasks [55]. With its ability to handle diverse datasets, ranging from structured to unstructured data, XGBoost has become a go-to choice for data scientists and researchers across industries.

In this section, we embark on a journey to uncover the intricate mechanisms that drive XGBoost's success. We delve into its underlying principles, which enable it to seamlessly blend the predictive power of multiple weak learners into a robust ensemble model. Through a detailed exploration of XGBoost's architecture and optimization strategies, we shed light on how it efficiently learns complex patterns from data while mitigating overfitting. Alg. 1 shows a pseudo code of the XGBoost. Moreover, XGBoost's interpretability and feature importance analysis capabilities provide invaluable insights into model behavior, empowering practitioners to understand the driving factors behind predictions. We explore techniques

Journal Pre-proof

for interpreting XGBoost models, such as SHAP (SHapley Additive exPlanations) values and partial dependence plots, which enhance transparency and trust in the model's decisions. Furthermore, we discuss XGBoost's scalability, illustrating its ability to handle large-scale datasets with ease, making it suitable for real-world applications where speed and efficiency are paramount.

Through this comprehensive exploration, we aim to showcase not only the technical prowess of XGBoost but also its transformative impact on data-driven decision-making and predictive analytics. Join us as we unravel the power of eXtreme Gradient Boosting and unlock new possibilities in the realm of machine learning.

Algorithm 1 XGBoost

```
Require: The number of iterate K;
         The Depth of trees D;
Ensure: Prediction Y(x)
function XGBoost(data, labels, num_rounds, max_depth, learning_rate):
Initialize model:
EnsembleModel = [ ]
BaseModel = BaseDecisionTree()
for round = 1 to num_rounds do
  Compute gradient and hessian for each data point:
  for round = 1 to num_rounds do
    Compute gradient and hessian using labels[i] and predicted label from ensemble model
  Endfor
  Fit base decision tree to gradient and hessian:
  BaseModel.fit(data, gradient, hessian, max_depth)
  Compute shrinkage:
  shrinkage = learning_rate / (1 + round)
  Update ensemble model:
  EnsembleModel.append(BaseModel)
  Update predicted labels using the newly added decision tree
Endfor
return EnsembleModel
```

4.4. Explainable Artificial Intelligence

Explainable artificial intelligence (XAI), or other words Interpretable machine learning (IML), delves into understanding and clarifying the decision-making processes of machine learning models, particularly crucial when these decisions have far-reaching implications [56][57]. Through various techniques such as feature importance analysis, model visualization, and rule-based models, IML methods enhance transparency and interpretability. By instilling trust in the models, these methods facilitate the utilization of machine learning in critical domains, ensuring informed decision-making[58][59].

4.4.1. Shapley Additive Explanations

Shapley Additive Explanations, known as SHAP, is a sophisticated method for explaining machine learning methods by assigning importance values to each feature based on its contribution to the model's predictions. Derived from cooperative game theory, SHAP values provide a fair and consistent way to attribute the output of the model to its input features, ensuring that the sum of the SHAP values matches the model's prediction [60]. Positive SHAP values indicate that the feature contributes positively to the prediction, pushing the prediction higher than the baseline. Conversely, negative SHAP values indicates that the feature contributes negatively, pulling the prediction lower than the baseline. This approach not only quantifies feature importance but also offers insights into how each feature influences the model's decisions, making it a powerful tool for understanding and explaining complex machine-learning models. This framework of SHAP value and components is depicted in Fig. 8.

The SHAP method is a versatile tool for explaining the outputs of any data-driven technique such as machine learning, regardless of the learning strategy or training algorithm used. It employs a sampling-based strategy to compute SHAP values, which determine the individual contribution of every single feature to the entire outcomes by generating numerous combinations of features in input parameters and computing predictions individually. This approach provides both global and local interpretability, making it especially valuable in fields requiring high transparency. [61]. These contributions are quantified as Shapley values, ultimately defining the explanation model $g(x0)$:

$$g\left(\boldsymbol{x}'\right) = \varphi_0 + \sum_{i=1}^{M} \varphi_i \boldsymbol{x}'_i \tag{16}$$

Journal Pre-proof

Figure 8: The framework of Shap Additive Explanations

Figure 9: The framework of Local Interpretable Model-agnostic Explanations

In this scenario, M denotes the features amount in the domain, while φ_0 remains constant if entire input values are zero. Each feature i includes an Attribution value indicated by φ_i. x' refers to the simplified input parameters' vector derived from initial database's x variables. Employing game theory principles, the SHAP method assigns the contribution of every single feature to the outcomes depicted by the Shapley value. x_0 represents a specific input instance for which the SHAP values are calculated. $g(x_0)$ defines the explanation model that quantifies the contribution of each feature to the prediction made for this particular instance x_0. SHAP values provide insights on both a global and local level, with x_0 highlighting the model's interpretability for individual samples.

Those details regarding interpretation need to fulfill specific criteria, which include:

• Local accuracy: Ensuring local accuracy entails that the explanation model's output aligns with the initial data model's values for the given input parameters under examination:

$$g\left(\boldsymbol{x}'\right) = \varphi_0 + \sum_{i=1}^{M} \varphi_i \boldsymbol{x}'_i = \varphi(\boldsymbol{x}) \tag{17}$$

Here, $\varphi(\boldsymbol{x})$ denotes preset initial ML data model, such as decision tree-based data-driven model in this part.

• Missingness: The corresponding attribution value will be zero when a feature is absent in a sample, ensuring that only present features contribute to the prediction.

$$x'_i = 0 \Rightarrow \varphi_i = 0 \tag{18}$$

• Consistency: In the Shapley value, consistency is crucial for reliable model interpretation. When comparing two models namely ϕ and ϕ', their feature attribution data should vary consistently, preserving the relative importance of features and ensuring trustworthy insights. The equation is shown below:

$$\phi_{x'}\left(z'\right) - \phi_{x'}\left(z'\backslash i\right) \geq \phi_x\left(z'\right) - \phi_x\left(z'\backslash i\right) \Rightarrow \varphi_i(\phi, x) \geq \varphi_i\left(\phi', x\right) \tag{19}$$

Here in the equation, the notation z' depicts a subset of the input x', while $z'\backslash i$ indicates the subset excluding the i-th feature, setting $z'_i = 0$. This approach is crucial for calculating Shapley values, as it helps determine the impact of each feature's presence or absence on model predictions, ensuring accurate and interpretable feature attribution.

To secure a singular solution for $g(\Delta)$, it's essential to constrain the three properties mentioned earlier. Theoretically, only one value of ϕ_i aligns with these criteria, leading to a singular result for a equation.

$$\varphi_i(\phi, x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!} \left[\phi_x\left(z'\right) - \phi_x\left(z'\backslash i\right)\right] \tag{20}$$

Here, z' is a subset of x', denoted by $z' \subseteq x'$. The notation $|z'|$is essential for calculating Shapley values to determine feature importance accurately, indicating the count of non-zero entries in subset z'.

Directly solving Eq. 20 presents a computational challenge due to the multitude of potential feature subsets. To overcome this, various approximation techniques, such as TreeSHAP, have emerged for efficiently computing the Shapley value. These computed values play a pivotal role in explaining the model's output, enabling the creation of individual interpretation plots for diverse samples. In these plots, positive Shapley values are depicted in red, while negative ones are shown in blue, signifying the feature's impact on the model output. Furthermore, the Shapley Value sheds light on the significance of different features in final model predictions, with feature dependence plots offering insights into feature-target variable relationships. These plots illustrate the marginal effect of one single feature on the prediction when keeping the rest of the features constant, providing clear insights into the feature's impact. They assist in interpretation by displaying feature values on the x-axis and the corresponding predicted values of the target variable on the y-axis. Additional visual aids like shaded areas or error bars often accompany these plots to indicate prediction uncertainty.

Journal Pre-proof

4.4.2. Local Interpretable Model-agnostic Explanations

Local Interpretable Model-agnostic Explanations, which stands for LIME , is a popular methodology used to explain the predictions of machine learning models. Developed by Ribeiro, Singh, and Guestrin in 2016, LIME aims to provide insights into the behavior of complex models by approximating them with simpler, interpretable models locally around the prediction of interest [62].

LIME can be applied to any machine learning model, regardless of its complexity or type. This includes neural networks, ensemble methods, and even linear models. Instead of trying to explain the entire model, LIME focuses on explaining individual predictions [59]. It creates local approximations around the specific prediction to provide insights into why the model made that particular decision. LIME can use simple, interpretable models like linear models or decision trees to approximate the behavior of the complex model in the local vicinity of the prediction [63]. These models are easier for humans to understand and interpret. The entire framework of LIME in this work is presented in Figure 9. We mainly use LIME to consider global importance,feature interactions, and local interpretation.

Normally, LIME starts by perturbing the input data around the instance being explained. For a given prediction, it generates several new data points by slightly altering the original data. Then the complex model is used to predict the outcomes for these perturbed data points. This step helps in understanding how the model's predictions change with slight variations in the input. The perturbed data points are weighted based on their similarity to the original instance. Points closer to the original instance receive higher weights, emphasizing the local behavior of the model. A simple, interpretable model (like a linear model) is then fitted to the perturbed data points, using the weights assigned in the previous step. This model serves as a local approximation of the complex model [64]. The coefficients or structure of the interpretable model provide insights into which features are most important for the prediction. These explanations are presented in a human-understandable format [65].

Overall feature importance is determined by averaging the contributions of each feature across multiple samples. This approach helps identify which features are most influential in the model's predictions, providing a foundation for model interpretation and optimization. Let $f(x)$ be the model's prediction function, x_i be the i-th feature, and S be the set of samples. The overall feature importance $I(x_i)$ is defined as:

$$I(x_i) = \frac{1}{|S|} \sum_{x \in S} \text{contribution}(x_i, x)$$

where $\text{contribution}(x_i, x)$ represents the contribution of feature x_i in sample x. In LIME, feature importance is indicated by the weights in local explanations.

LIME explains the local behavior of complex models by generating new samples around the instance to be explained and training a simple, interpretable model (e.g., a linear model) on these samples to approximate the complex model's behavior in that local region [66].

Given a complex model f and an instance x to be explained, LIME generates neighborhood samples Z and trains a simple model g with the loss function:

$$L(f, g, \pi_x) + \Omega(g)$$

where L measures the difference between f and g on the neighborhood samples, π_x is the weighting of the neighborhood samples, and Ω is a regularization term for the complexity of g.

Specifically, a linear model g can be represented as:

$$g(z') = w_0 + \sum_{i=1}^{d} w_i z'_i$$

where z' are the neighborhood samples and w_i are the weights indicating the importance of feature i.

Feature interaction analysis examines the combined effect of two or more features on the model's predictions. By identifying interactions between feature pairs, we can gain deeper insights into how the model leverages multiple features for its predictions. Interaction strength $I(x_i, x_j)$ is defined as the difference between the joint contribution of features x_i and x_j and the sum of their individual contributions:

$$I(x_i, x_j) = \text{contribution}(x_i, x_j) - (\text{contribution}(x_i) + \text{contribution}(x_j))$$

where $\text{contribution}(x_i, x_j)$ is the joint contribution of features x_i and x_j. In LIME, interaction strength is calculated by analyzing the weights of feature pairs in local explanations.

SHAP is used for detailed analysis of feature importance and interactions at a global scale and for individual predictions. LIME is employed to verify specific predictions and provide a simpler explanation in cases where domain experts may prefer straightforward, localized insights over global consistency. Using both methods enhances the robustness of our interpretability framework. By comparing the results from SHAP and LIME, we can validate that the explanations provided are consistent and reliable. LIME serves as a complementary tool to ensure that localized behaviors observed in SHAP explanations hold under a different interpretability method.

Journal Pre-proof

4.5. Sensitivity Analysis

Quantitative analysis (QA) employs mathematical and statistical modeling to understand the mechanisms behind behaviors, aiming to represent reality in numerical terms. In this article, we use local and variance-based sensitivity analysis (SA) to rank the importance of model inputs in contributing to output variability. Variance-based SA systematically varies input parameters to measure their impact on the model's output, decomposing the variance into components that reflect the contribution of each parameter or their combinations. This method is particularly useful when inputs are highly correlated or interact in complex ways.

Using variance-based SA for our quantitative analysis provides detailed insights into how each input parameter affects the model's output, identifying the most influential factors. This information is crucial for optimizing models and systems, as well as informing policy decisions by highlighting the most impactful factors [67]. By understanding both the main effects and interactions of parameters, we can achieve more accurate models, better decision-making, and more efficient designs, ensuring that we focus on the most critical aspects of the studied system.

4.5.1. First-order sensitivity indices

In this work, Key Performance Indicators (KPIs) in sensitivity analysis include the numerical values of the first-order sensitivity indices S_i and the total effect sensitivity indices S_{Ti}. According to sensitivity analysis theory, these indices relate to a physical model's response function, expressed as $\boldsymbol{Y} = f(X_1, X_2, \ldots, X_k)$. As outlined by [68], the first-order sensitivity indices are derived from this framework:

$$S_i = \frac{V_{x_i}[E_{X_{\sim_i}}(\boldsymbol{Y}|X_i)]}{V(\boldsymbol{Y})} \tag{21}$$

In this context, $V_{x_i}[E_{X_{\sim_i}}(\boldsymbol{Y}|X_i)]$ represents the main effect of $X_i$on the output variable in the mathematical model. The denominator, $V(\boldsymbol{Y})$, is the unconditional variance of $\boldsymbol{Y}$. The term $V_{x_i}[E_{X_{\sim_i}}(\boldsymbol{Y}|X_i)]$ indicates the variance of the expected value of $V(\boldsymbol{Y})$ when conditioned on X_i meaning X_i is fixed at a specific value $X_{j_i}, j = 1, \ldots\ldots, N;$, N is the number of samples.

4.5.2. Total effect sensitivity indices

Theoretical modeling reveals the insufficiency of depending solely on the first-order sensitivity index, as it provides only a partial assessment of model variance bias. Consequently, there arises a crucial need to broaden our analysis beyond the first order, in order to account for higher-order coupling effects and effectively mitigate this source of error. As a response to this imperative, an extension to the first-order sensitivity index is proposed, culminating in the formulation of the total effect index, as elaborated in [68]:

$$S_{T_i} = \frac{E_{x_{\sim_i}}[V_{X_i}(\boldsymbol{Y}|\boldsymbol{X_{\sim i}})]}{V(\boldsymbol{Y})} = 1 - \frac{V_{x_{\sim_i}}[E_{X_i}(\boldsymbol{Y}|\boldsymbol{X_{\sim i}})]}{V(\boldsymbol{Y})} \tag{22}$$

At the molecular position, $V_{x_{\sim_i}}[E_{X_i}(\boldsymbol{Y}|\boldsymbol{X_{\sim i}})]$ represents the variance derived from the expected value of all parameters $\boldsymbol{Y}$ except X_i denoted subsequently as $\boldsymbol{X_{\sim i}}$. This variance characterizes the first-order effect of $\boldsymbol{Y}$ except X_i on the model output, excluding the contribution of X_i.

The total effect index S_{T_i} encapsulates the collective impact of input parameters X_i on the output. This index signifies the summation of the first-order term and all higher-order terms, portraying the comprehensive influence of the input variables on the output.

$$S_{T_i} = S_i + S_{i,\sim_i} = 1 - S_{\sim_i} \tag{23}$$

where $S_{\sim_i}$ is a sum, including all parameters' sensitivity indices except of i.

Achieving accurate values for S_i and S_{T_i}, as described in equations [21] and [22], necessitates a large number of samples, which can be computationally intensive, especially when data is limited. To address this, we utilize a surrogate model, denoted as $\boldsymbol{Y}$, to approximate and depict the responses of the true physical model.

5. Numerical analysis and discussion

5.1. Results in Multi-scale modeling

Figure 10 illustrates the study in convergence, i.e., how the predictive outcomes varies with the RVE size, showing a trend that stabilizes at a specific value as it approaches a certain point. This convergence indicates RVE size at which

further increases do not significantly affect the predicted thermal conductivity, suggesting that an optimal RVE size has been reached.

To provide a more comprehensive understanding, Figures 12 and 13 present the temperature gradient distribution in 'Cube' when the heat flow flows through the X-axis. Figures 14 and 15 indicate the situation of Y-axis while Figures 16 and 17 illustrate the Z-axis. Figures 12 14 16 focus on the temperature distribution within the individual graphene, while Figures 13 15 17 show the temperature distribution across the composite material. These visualizations highlight how heat is transferred through different components of the material, further validating the model's accuracy and effectiveness in predicting thermal properties.

Figure 10: Convergence Study: Thermal conductivity versus RVE size

Figure 11: The predictive performance of XGB in test set

Table 2: The comparison in experiments and prediction

Literature	Materials Properties		Thermal conductivity (W/mK)		
	κ_m	V_f	Experiments	Simulation	Percent Error
Zhengqing Yang [69]	graphene (rGO)/epoxy ($0.178W/mK$)	5%	1.239	1.3721	9.7%
Yun Seon Lee [70]	rGO/epoxy ($0.23W/mK$)	5%	1.76	1.9273	8.6 %
		10%	2.56	2.8005	8.5%
Jinrui Gong [71]	graphene/PI $0.25W/mK$)	12%	3.73	3.5478	10.77%

Additionally, we compare the values found in the literature with those presented in Table 2. This table contains outputs computation through stochastic multi scale simulation. By examining these comparisons, we can evaluate the accuracy and reliability of our model. The alignment or discrepancies between our model's predictions and the literature values provide insights into the model's performance and highlight areas for potential refinement. This comprehensive comparison underscores the robustness of our approach and its applicability in accurately predicting the properties under study. The stochastic multi-scale modeling process, when executed on a MacBook Pro (15-inch, 2017) equipped with a 2.8 GHz quad-core Intel Core i7 CPU, is about to require approximately 8 hours. This extended runtime is attributed to the system's limited processing power, with only four cores available, and limited GPU acceleration. The system specifications include a Radeon Pro 555 GPU with 2 GB VRAM, supplemented by an Intel HD Graphics 630 with 1.5 GB shared memory, 16 GB of 2133 MHz LPDDR3 RAM.

5.2. Results in machine learning

The initial outcome of the optimization process identifies the best settings for XGBoost. Key parameters include the complexity and the minimum samples per split, which are set to 0, 5, and 10. The tuning also determines the number of estimators to be 100, the learning rate to be 0.03, the maximum number of regression trees to be 10,000, the interaction depth to be 7, and the maximum tree depth to be 4. A detailed summary of these selected hyperparameters can be found in Table 3

Figure 12: Temperature distribution ($^{\circ}C$) of inner plates (graphene) within RVEs (X-axis)

Figure 13: Temperature distribution ($^{\circ}C$) of composites (graphene mixed with epoxy) within RVEs (X-axis)

Figure 14: Temperature distribution ($^{\circ}C$) of inner plates (graphene) within RVEs (Y-axis)

Figure 15: Temperature distribution ($^{\circ}C$) of composites (graphene mixed with epoxy) within RVEs (Y-axis)

Table 4 highlights the performance metrics of XGBoost, demonstrating its reliability and effectiveness in predicting the required properties. To further illustrate the model's predictive accuracy, scatter plots of its predictions are presented. Figure 11 shows the scatter plot of XGBoost predictions on the test set, where most points align closely with the Y=T line, indicating high accuracy. This alignment underscores XGBoost's excellence as a regression tree-based model, effectively capturing the relationship between the input features and the target variable. Additionally, the points are densely concentrated within the data interval of [0.5-2.5], suggesting particularly precise predictions within this range. This concentration indicates that the model performs exceptionally well within this specific domain, making it a robust choice for applications requiring high precision in this interval. The scatter plot visually confirms the model's ability to generalize well to unseen data, reinforcing its potential utility in various predictive tasks. The generation of the training database for machine learning (ML), which involves data preprocessing, feature extraction, and dataset preparation, is expected to take approximately 39 minutes (2363s) on the same system (2.8 GHz quad-core Intel Core i7 CPU, Radeon Pro 555 2 GB, Intel HD Graphics 630 1536 MB, 16 GB of 2133 MHz LPDDR3 RAM). The quad-core CPU's limited parallel processing capabilities similarly constrain the task's runtime. The system specifications remain identical, with the reliance on Python-based libraries such as Pandas and Scikit-learn for data handling and preparation.

5.3. Model Interpretations in SHAP method

Three components in interpretability regarding SHAP will be discussed in Individual Interpretations, Global Interpretations, and Feature Dependency to unveil the inside mechanism of workings in opacity.

5.3.1. Global Interpretations

The absolute mean of every SHAP value in single feature is calculated to assess its importance as the start of global interpretability, yielding a standardized bar chart as depicted in Figure 18. This chart reveals the broad influences of different features on predictions. It is worth noting that the results indicate the matrix's thermal conductivity is a crucial

Journal Pre-proof

Figure 16: Temperature distribution ($°C$) of inner plates (graphene) within RVEs (Z-axis)

Figure 17: Temperature distribution ($°C$) of composites (graphene mixed with epoxy) within RVEs (Z-axis)

Table 3: Hyper-parameters tuning

ML method	hyper-parameters	Definition	Interval	Optimum Value
Gradient Boosting Machine	n	Number of estimators	[0,150]	100
	N	The maximum number of regression tree	[100,10000]	10000
	N_d	Maximum depth of tree	[1,10]	4
	λ	The learning rate	[0.001,0.99]	0.03
	D_{ia}	Interaction depth	[1,10]	7

parameter influencing the thermal conductivity of the composite on a macro scale. In contrast, the dispersion index is relatively less significant. The volume fraction also plays a role in composite thermal conductivity but to a lesser extent. Conversely, agglomeration and dispersion indices exhibit relatively minor effects on the overall outcome. The SHAP summary plot, depicted in Figure 19, presents the distribution of SHAP values for each feature, illustrating their respective contributions to the model's predictions. Each feature listed on the y-axis influences the model output, with SHAP values on the x-axis representing the impact on the model's predictions. Regarding the color gradient, from blue (low) to red (high) in the SHAP plot, illustrating each feature's impact on predictions. For instance, in the thermal conductivity of matrix and volume fraction, high values (red) generally push the model output higher, whereas low values (blue) pull it lower, reinforcing the influence of high feature values on increasing thermal conductivity.

Analyzing Figure 19, we note the thermal conductivity of matrix show a significant spread of SHAP values, indicating it has a strong influence on the model's predictions. High values (marked in red) tend to increase model output, suggesting that a higher thermal conductivity of matrix significantly increases predicted composite thermal conductivity. Volume Fraction also displays a broad distribution, meaning it is relatively influential. High values contribute positively, though the impact is somewhat less pronounced than the thermal conductivity of matrix. The distribution of Aspect Ratio here is moderate. Higher aspect ratios seem to positively influence the model output, though the effect is more centered around zero, indicating a balanced influence with both positive and negative contributions. Kapitza Resistance has a balanced impact, with low values (blue) generally having a negative impact on predictions and high values contributing positively. High thermal conductivity of graphene correlates with a positive impact on model output, as indicated by the red values on the positive side. Agglomeration Index and Dispersion Index have a limited impact on the predictions, as shown by the narrow distribution of SHAP values around zero.

5.3.2. Local interpretation

Besides, SHAP values also enable us to gain insights into individual interpretations for each sample. Figures 20 represents there distinctive samples that can be further elucidated. The length of each bar reflects the magnitude of these changes as well as the colors show the different directions. The red bars in the figure indicate features contributing to to a boost in outcomes in the base value, while blue bars lead corresponding features in contrast.

In sample 51, we observe the corresponding thermal conductivity of matrix and inside fillers graphene positively impact critical feature predictions, whereas the agglomeration index, volume fraction, and aspect ratio negatively affect the final prediction. Turning to sample 75, we note that dispersion index and aspect ratio positively impact final model

Journal Pre-proof

Table 4: ML models performance in Prediction

ML model	Sets	R^2	RMSE	MAE	MAPE (%)
Gradient Boosting Machine (XGBoost)	Training	0.9960895358598272	0.0495465858010725	0.029409395870822447	2.6910865129030777
	Test	0.8470324779142511	0.28600361472317265	0.11576731435512813	0.11576731435512813

output, while the agglomeration index, thermal conductivity of matrix, and volume fraction exhibit somewhat negative impacts on the model outputs. As for sample 80, the agglomeration index and the volume fraction have positive influence on the prediction while polymer and graphene's thermal conductivity negatively affect the model outcomes.

5.3.3. Feature interactions

We examine outcomes of the SHAP in every feature relative to whole dataset, gaining deeper insights into how individual feature values impact ML models training. Figure 22 (a)-(g) illustrates the SHAP dependence plots for different features, showing their impact on predicting the thermal conductivity of the composite material. In Figure 22 (a), there is a clear inverse relationship between the SHAP value and the thermal conductivity of matrix value, suggesting that as the thermal conductivity of matrix value increases, its impact on the model's output diminishes. High values of the thermal conductivity of matrix (toward the right) generally decrease the SHAP value, indicating a reduced influence on predicted thermal conductivity at higher thermal conductivity of matrix levels. In Figure 22 (b), The plot shows a scattered pattern with a relatively lower range of SHAP values at high volume fractions, suggesting that increasing volume fraction stabilizes its influence on the predictions. The color gradient (with high values in red and low in blue) indicates potential interactions with the "Thermal Matrix" feature, suggesting a compounded effect when both thermal conductivity of matrix and volume fraction values are high. In Figure 22 (c), The aspect ratio plot shows that higher aspect ratios tend to have a balanced impact on model predictions, with SHAP values clustered around zero. The color coding reveals that high volume fraction values (red) amplify the positive impact of higher aspect ratios on the model's output. In Figure 22 (d), This plot indicates a generally balanced contribution of Kapitza to the SHAP values, with both positive and negative impacts around a central range. The thermal conductivity of matrix values (color-coded) show an interaction effect where higher Kapitza values and high thermal conductivity of matrix values together increase the model's predictive output. In Figure 22 (e), Thermal graphene exhibits a dispersed SHAP value pattern, indicating that high thermal conductivity of graphene has a slightly positive impact on the prediction, though not as dominant as other features. The thermal conductivity of matrix interaction is visible, with higher thermal conductivity of matrix values slightly increasing the impact of thermal graphene on model predictions. In Figure 22 (f), Higher agglomeration index values seem to positively influence the model's output, as indicated by the upward trend of SHAP values with increasing agglomeration index. High volume fraction (red dots) amplifies the agglomeration index's positive influence, showing that the interaction between these factors is relevant in affecting the model's predictions. In Figure 22 (g), The dispersion index has an overall weak effect on SHAP values, although a slight upward trend in SHAP values is observed with higher dispersion index values. The color gradient shows an interaction with the thermal conductivity of matrix, where high thermal conductivity of matrix values (red) moderately amplify the positive SHAP values associated with higher dispersion index levels.

Additionally, Figure 21 shows the SHAP interaction values among different parameters, highlighting how pairs of features interact to influence the model's prediction of composite thermal conductivity. Each row corresponds to a primary feature, while each column represents the interaction with another feature. SHAP interaction values on the x-axis indicate the strength and direction of the combined effect of each feature pair on the model's output, with positive values increasing and negative values decreasing the prediction. The thermal conductivity of matrix feature shows strong interactions with itself and volume fraction, indicated by the wide spread of SHAP values in these cells. High values of the thermal conductivity of matrix tend to increase the model's predictions, especially when interacting with other high values (in red). Volume fraction also exhibits significant interactions, particularly with itself and with the thermal conductivity of matrix, implying that these two features together are crucial in shaping the model's predictions. The spread of values suggests that higher volume fractions positively influence the prediction, especially in combination with high thermal conductivity of matrix values. Aspect ratio has moderate interactions, mainly with itself and to a lesser extent with other features. The distribution of SHAP values around zero suggests that aspect ratio has a balanced, less extreme effect on predictions compared to thermal conductivity of matrix and volume fraction. Kapitza shows a relatively balanced interaction profile, indicating that it moderately affects predictions. Higher values of Kapitza, especially in combination with high values of other features, can enhance model output, as seen in the interaction with thermal conductivity of matrix and volume fraction. The thermal conductivity of graphene has a minimal but generally positive interaction effect. Its interaction with the thermal conductivity of matrix and volume fraction slightly increases the predictive output, but overall, it has a narrower impact range. Both the agglomeration and dispersion indices show limited influence on the model output, as seen by their narrow distributions around zero. This suggests they have minor interactions with other features and a negligible effect on the overall predictions.

Figure 18: Global interpretations by SHAP values (SHAP feature importance)

Figure 19: Global interpretations by SHAP values (SHAP summary plot)

Figure 20: Individual interpretations for different Samples

Figure 21: SHAP interaction value among different parameters

5.4. Model Interpretations in LIME

In this section, we present the results from LIME, focusing on global interpretations, feature interactions, and local interpretations of samples. These insights will help us understand the overall behavior of the model, identify significant feature interactions, and explain individual predictions.

5.4.1. Global Interpretations

The Fig 23 illustrates that Thermal Matrix and Volume Fraction are the most important features in the model's predictions. Thermal conductivity of matrix has the highest average importance, approximately 0.35, while Volume Fraction follows with an average importance of around 0.25. This indicates that these features play a crucial role in the model's decision-making process.

Aspect Ratio and Kapitza also demonstrate significant average importance, with values around 0.10 and 0.05, respectively, highlighting their substantial roles in the model. In contrast, Agglomeration Index and Dispersion Index have lower average importance, about 0.05 and 0.02, respectively, suggesting a relatively smaller impact on the model. Thermal Graphene shows the lowest average importance at approximately 0.01, indicating minimal influence on the model's predictions.

The chart depicts the distribution of feature importance across all selected samples. Higher importance signifies greater consistency and impact on the model's predictions. This visualization aids in understanding the features the model relies on, providing a basis for feature selection and potential model improvements.

Compare the results from Fig 18 and Fig 23, both LIME and SHAP identify Thermal Matrix and Volume Fraction as the most influential features in determining thermal conductivity. However, LIME ranks Volume Fraction as the most significant feature, followed closely by Thermal Matrix, while SHAP places thermal conductivity of matrix as the dominant feature with a larger margin of importance. This consistency in highlighting the two key features across both methods reinforces their significance in the model. Additionally, features like Agglomeration Index and Dispersion Index remain less important in both interpretations, suggesting that these features have a relatively minor impact on the model's predictions.

Figure 22: Feature dependence plots

Figure 23: Overall Feature of Importance in LIME method

Figure 24: Interaction strength distribution among overall factors

However, the importance of the Kapitza and Aspect Ratio features varies slightly between the two methods. In the LIME results, Aspect Ratio appears to play a more significant role, ranking higher than Kapitza, whereas SHAP shows Kapitza as slightly more influential than Aspect Ratio. Furthermore, the impact of Thermal Graphene is minimal in both methods, but LIME ranks it slightly higher than SHAP. These differences likely stem from the distinct ways that LIME and SHAP quantify feature importance - LIME provides local explanations for individual predictions, whereas SHAP calculates global feature importance based on the overall contribution across all data points.

5.4.2. Feature interactions

By analyzing the importance of feature interactions, we can gain a deeper understanding of the model's logic in the prediction process, enhancing its interpretability and usability. A common pattern in these interactions is the combination of specific ranges of one feature with specific ranges of another, such as the frequent pairing of different ranges of Volume Fraction and Thermal Matrix. This indicates that the joint distribution of these features significantly impacts the model's predictions.

Figure 25: Overall Feature interactions in Lime method

The Figure 25 displays the combined importance of feature pairs, with high combined importance indicating a key role in the model's decision process. As shown in the figure, the interaction between $0.05 <$ Volume Fraction ≤ 0.08 and Thermal Matrix ≤ 0.18 is the most critical, with a combined importance close to 1.2. This suggests that this feature interaction has the greatest influence on the model's predictions. Other important interactions include $0.25 <$ Thermal Matrix ≤ 0.34 and Volume Fraction ≤ 0.03 and $0.05 <$ Volume Fraction ≤ 0.08 and $0.18 <$ Thermal Matrix ≤ 0.25, both with a combined importance of around 0.8.

Additional significant feature pairs are $0.25 <$ Thermal Matrix ≤ 0.34 and Volume Fraction > 0.08, $57.82 <$ Aspect Ratio ≤ 79.86 and Thermal Matrix > 0.34, and $37.16 <$ Aspect Ratio ≤ 57.82 and Thermal Matrix > 0.34, with their combined importance ranging from 0.5 to 0.7.

Aspect Ratio, Thermal Matrix, and Volume Fraction frequently appear in these important interaction pairs, highlighting their substantial impact on model predictions. This interaction pattern suggests an underlying regularity in the data, indicating interdependence between features.

Interaction intensity profiles provide insights into the interactions between features and their importance to model predictions. The Fig 24 reveals that the frequency distribution of interaction strength presents a right-skewed long-tail distribution, with most interaction strengths concentrated in the lower value range. Interaction strengths between 0.0 and 0.1 have the highest frequency, occurring more than ten times. This indicates that most feature pairs exhibit small interaction effects.

As the intensity of interaction increases, the frequency gradually decreases, showing that there are fewer feature pairs with strong interactions. Although the frequency of feature pairs with interaction strengths between 0.6 and 1.2 is low, these pairs have a greater influence on model predictions. These high-interaction-strength features may significantly impact model predictions under specific circumstances and warrant further study and attention.

From the distribution trend perspective, the Kernel Density Estimation (KDE) curve further illustrates the distribution of interaction strength. The curve peaks near 0.1 and then gradually decreases, indicating that while most feature

pairs have low interaction strength, there are still a few pairs with high interaction strength.

Through the results from Figure 26, we can better understand the performance of feature interactions in different samples. We selected five samples to illustrate the relevant results.

Figure 26(a): The local impact of $0.05 <$ Volume Fraction ≤ 0.08 and Thermal Matrix ≤ 0.18. It is evident that this feature has the largest local impact on sample 8, which is significantly higher than in other samples. The local effects on other samples are relatively low and stable, indicating that this feature generally has a small impact but plays a significant role in specific samples.

Figure 26(b): The local effect of $0.25 <$ Thermal Matrix ≤ 0.34 and Volume Fraction ≤ 0.03. In sample 9, this pair of features has the largest local impact, suggesting a significant influence on model predictions for this sample. The impact on other samples is lower and relatively stable, indicating a limited impact in most cases.

Figure 26(c): The local effect of $0.05 <$ Volume Fraction ≤ 0.08 and $0.18 <$ Thermal Matrix ≤ 0.25. This feature pair has a large local impact on samples 0 and 8, indicating a significant contribution to model predictions in these samples. The impact on other samples is lower and more evenly distributed, showing a relatively small impact in most samples.

Figure 26(d): The local impact of $0.25 <$ Thermal Matrix ≤ 0.34 and Volume Fraction > 0.08 on samples 3, 4, and 8. This pair of features has a larger local impact on these samples, indicating a greater contribution to model predictions. The impact on other samples is lower and unevenly distributed, highlighting significant variability in the impact among different samples.

Figure 26(e): The local effect of $57.82 <$ Aspect Ratio ≤ 79.86 and Thermal Matrix > 0.34. The local effects are largest on samples 1 and 3, indicating a significant impact on model predictions in these samples. The impact on other samples is low and stable, suggesting a relatively small impact in most cases.

Figure 26: Selected 5 top most important feature pairs and local impact on different samples

The feature interaction heat map reveals the interrelationships between features. By analyzing these relationships, the model can be better understood and optimized, improving both its predictive power and interpretability. Specific pairs of highly interactive features highlighted in the heat map deserve further study and attention. The features displayed in the heatmap were not pre-screened, as the purpose of this analysis was to provide a comprehensive view of all potential feature interactions in the model.

Journal Pre-proof

Figure 27: Heatmap of feature interactions

As can be seen in Figure 27, which displays three levels of interaction intensity:

- **High-Strength Interaction**: The interaction strength between *Thermal Matrix (thermal conductivity of matrix)* and *Volume Fraction* is close to 1 (0.99), indicating a very strong interactive relationship in the model. The interaction strength between *Thermal Matrix (thermal conductivity of matrix)* and *Aspect Ratio* is also relatively high (0.53), signifying a significant mutual influence between these two features.
- **Medium-Strength Interaction**: The interaction strength between *Volume Fraction* and *Aspect Ratio* is 0.41, suggesting a notable influence between them. Additionally, the interaction strength between *Agglomeration Index* and *Aspect Ratio* is 0.36, indicating a moderately strong interaction between these features.
- **Low-Strength Interaction**: For most feature pairs, the interaction strength is low, close to 0, indicating a small or negligible interaction effect between these pairs.

The interaction strength between *Thermal Matrix* and *Volume Fraction* almost reaches the maximum value, indicating a very close relationship in model predictions. This strong interaction may help optimize the model's performance. The interaction strength between *Thermal Matrix* and *Aspect Ratio* is 0.53, suggesting that these two features can jointly affect the model's prediction results in some cases. The interaction strength between *Volume Fraction* and *Aspect Ratio* is 0.41, implying that, to a certain extent, the combination of these two features influences the model's predictions. Similarly, the interaction strength between *Agglomeration Index* and *Aspect Ratio* is 0.36, indicating that their interaction has a notable impact on model predictions under certain circumstances. The interaction strength between other feature pairs is lower, indicating that these pairs have less or no significant influence on each other.

5.4.3. Local interpretation of samples

The analysis results in the Table 5 show that when the model predicts different samples, the contributions and interactions of various features differ significantly. Predicted values range from 0.792079 to 2.919846, indicating substantial

Journal Pre-proof

variation in the model's predictions across different samples. Different features contribute differently in each sample, with some features having positive contributions in certain samples and negative contributions in others. *Thermal Matrix*, *Volume Fraction*, and *Aspect Ratio* are the most influential features, significantly impacting multiple samples. In several samples, some features, such as *Thermal Matrix* and *Aspect Ratio*, contribute negatively to the predicted value.

Table 5: Local interpretation (Feature importance) of 9 selected samples

Sample Index	Predicted Value	Thermal Graphene	Thermal Matrix	Kapitza	Aspect Ratio	Volume Fraction	Agglomeration	Dispersion
103.0	1.361851	0.056089	-0.582026	-0.027901	0.028747	0.672186	0.000000	0.000000
104.0	1.177865	0.000000	-0.337557	0.056307	0.183590	-0.412122	0.000000	0.041460
92.0	0.917865	0.000000	-0.342848	-0.025765	-0.170581	-0.289699	0.020721	0.000000
89.0	1.579804	0.000000	-0.369224	0.000000	-0.150841	0.686172	-0.072517	0.031017
97.0	1.147424	0.000000	-0.122192	-0.009913	-0.112237	0.054622	-0.054379	0.000000
30.0	2.919846	0.000000	0.999995	-0.053861	-0.094192	0.682349	-0.074107	0.000000
123.0	1.048928	0.000000	-0.126275	0.000000	-0.138444	-0.411122	-0.075657	0.093561
91.0	0.792079	0.000000	-0.570817	-0.074600	0.041033	-0.403543	-0.054885	0.000000
117.0	0.905157	-0.016489	-0.393365	-0.042748	-0.153768	0.066687	0.000000	0.000000

As illustrated in the Figure 28, the feature importance distribution varies across samples. For instance, in Sample 1, *Volume Fraction* and *Thermal Matrix* are the main features, while in Sample 2, *Volume Fraction* and *Aspect Ratio* dominate.

The red bars in the chart represent the negative contribution of features to the predicted value, whereas the green bars represent positive contributions. For example, in Sample 1, *Volume Fraction* contributes negatively, while *Thermal Matrix* contributes positively. Additionally, the chart highlights the impact of feature interactions on prediction results. In Sample 3, the combination of *Volume Fraction* and *Thermal Matrix* has a large negative contribution to the predicted value.

To understand the model's decision-making basis for individual samples, we deeply analyze which features have a significant impact in specific situations in different samples:

- **Sample 1**: *Volume Fraction* ≤ 0.03 and *Thermal Matrix* > 0.34 have a significant positive impact on the predicted value, while *Thermal Graphene* ≤ 3963.19 has a negative impact.
- **Sample 2**: *Volume Fraction* > 0.08 and *Thermal Matrix* ≤ 0.34 have a significant negative impact, while *Aspect Ratio* ≤ 37.16 has a positive impact.
- **Sample 3**: *Thermal Matrix* ≤ 0.34 and *Volume Fraction* ≤ 0.08 have a significant negative impact, while *Agglomeration Index* ≤ 0.55 has a positive impact.
- **Sample 4**: *Volume Fraction* ≤ 0.03 and *Thermal Matrix* ≤ 0.34 have a significant negative impact, while *Aspect Ratio* > 79.86 has a positive impact.
- **Sample 5**: *Thermal Matrix* > 0.25 and *Aspect Ratio* ≤ 79.86 have a significant negative impact, while *Volume Fraction* ≤ 0.05 has a positive impact.
- **Sample 6**: *Thermal Matrix* ≤ 0.18 and *Volume Fraction* ≤ 0.03 have a significant positive impact, while *Agglomeration Index* ≤ 0.55 has a negative impact.
- **Sample 7**: *Volume Fraction* > 0.08 and *Aspect Ratio* ≤ 79.86 have a significant negative impact, while *Dispersion Index* ≤ 0.55 has a positive impact.
- **Sample 8**: *Thermal Matrix* > 0.34 and *Volume Fraction* > 0.08 have a significant negative impact, while *Aspect Ratio* ≤ 57.82 has a positive impact.
- **Sample 9**: *Thermal Matrix* ≤ 0.34 and *Aspect Ratio* > 79.86 have a significant negative impact, while *Volume Fraction* ≤ 0.05 has a positive impact.

From the Table 5 and Figure 28, we can also comprehensively analyze that the characteristics of *Thermal Matrix* and *Volume Fraction* have significant effects in multiple samples, with their influence being either positive or negative. This indicates their varying effects in different contexts. Additionally, the combination of different features significantly impacts the predicted value. For instance, the interaction between *Thermal Matrix* and *Volume Fraction* has a notable effect in multiple samples, highlighting the crucial role of feature interactions in the model.

Moreover, the importance and direction of different features vary across samples, indicating the model's complex behavior in different scenarios. For example, *Aspect Ratio* has a positive impact in some samples (such as Sample 2) and a negative impact in others (such as Sample 5).

Figure 28: Local interpretation of 9 selected samples

Ultimately, the balance of contributions from positive and negative features in each sample determines the predicted value. This balance reflects the model's comprehensive consideration of each feature.

5.5. Sensitivity Analysis

We consider quadratic regression and moving least square regression as surrogate models for sensitivity analysis. Then 100,000 Latin hypercube samples within those models are generated. The Si and values are computed using the MLS approximation method and polynomial regression. The reduction in the sensitivity index indicates how closely the surrogate model approximates the 'real' model. Table 6 displays the outcomes from the variance-based methods. Figures 29 and 30 illustrate these trends in first-order indices and total-effect indices, respectively. The main takeaways from the sensitivity analysis can be obtained from those data. From this table, it presents that the total-effect indices and first-order indices show almost no discrepancies, implying no correlations among the input parameters. We can also find here all surrogate models exhibit suitability and predict the same trends, which means the parameter relevance and importance of this physical model are independent of the choice of surrogate models. From the numerical values of Si and STi we can find that, the volume fraction and aspect ratio exert the most significant influence on the overall thermal conductivity of the composite, followed by the thermal conductivity of the matrix and the conductivity of the graphene. Kapitza resistance corresponds to less sensitivity to parameter changes, which means that the interface effect has a smaller impact in the macroscopic composite design process compared with other mesoscopic parameters. Agglomeration index and Dispersion index have the lowest sensitivity and have the smallest impact on the results. The reason may be that the homogenization process of composite materials treats the interior as a homogeneous continuum, thereby reducing the impact of agglomeration and dispersion.

Table 6: First-order and total effects sensitivity indices computed on different surrogate models

Influencing factors	**Quadratic without mixed terms**	**Full Quadratic Regression**	**MLS**
First-order indices $\hat{S}_i$	First-order indices $\hat{S}_i$	First-order indices $\hat{S}_i$	First-order indices $\hat{S}_i$
$\hat{S}_1$: Thermal graphene	$\hat{S}_1$=0.1039	$\hat{S}_1$=0.1039	$\hat{S}_1$=0.1039
$\hat{S}_2$: Thermal matrix	$\hat{S}_2$=0.1217	$\hat{S}_2$=0.1217	$\hat{S}_2$=0.1217
$\hat{S}_3$: Kapitza resistance	$\hat{S}_3$=0.0212	$\hat{S}_3$=0.0212	$\hat{S}_3$=0.0212
$\hat{S}_4$: Aspect ratio	$\hat{S}_4$=0.3301	$\hat{S}_4$=0.3301	$\hat{S}_4$=0.3301
$\hat{S}_5$: Agglomeration index	$\hat{S}_5$=0.0087	$\hat{S}_5$=0.0087	$\hat{S}_5$=0.0087
$\hat{S}_6$: Dispersion index	$\hat{S}_6$=0.0087	$\hat{S}_6$=0.0087	$\hat{S}_6$=0.0087
$\hat{S}_7$: Volume fraction	$\hat{S}_7$=0.4057	$\hat{S}_7$=0.4057	$\hat{S}_7$=0.4057
	$\sum_{i=1}^{5} \hat{S}_i$=0.9999	$\sum_{i=1}^{5} \hat{S}_i$=0.9999	$\sum_{i=1}^{5} \hat{S}_i$=1
Total-effect indices $\hat{S_{Ti}}$	Total-effect indices $\hat{S_{Ti}}$	Total-effect indices $\hat{S_{Ti}}$	Total-effect indices $\hat{S_{Ti}}$
$\hat{S_{T1}}$ Thermal graphene	$\hat{S_{T1}}$=0.1039	$\hat{S_{T1}}$=0.1039	$\hat{S_{T1}}$=0.1039
$\hat{S_{T2}}$ Thermal matrix	$\hat{S_{T2}}$=0.1217	$\hat{S_{T2}}$=0.1217	$\hat{S_{T2}}$=0.1217
$\hat{S_{T3}}$ Kapitza resistance	$\hat{S_{T3}}$=0.0212	$\hat{S_{T3}}$=0.0212	$\hat{S_{T3}}$=0.0212
$\hat{S_{T4}}$ Aspect ratio	$\hat{S_{T4}}$=0.3301	$\hat{S_{T4}}$=0.3301	$\hat{S_{T4}}$=0.3301
$\hat{S_{T5}}$ Agglomeration index	$\hat{S_{T5}}$=0.0087	$\hat{S_{T5}}$=0.0087	$\hat{S_{T5}}$=0.0087
$\hat{S_{T6}}$ Dispersion index	$\hat{S_{T6}}$=0.0087	$\hat{S_{T6}}$=0.0087	$\hat{S_{T6}}$=0.0087
$\hat{S_{T7}}$ Volume fraction	$\hat{S_{T7}}$=0.4057	$\hat{S_{T7}}$=0.4057	$\hat{S_{T7}}$=0.4057
	$\sum_{i=1}^{5} \hat{S_{Ti}}$=1	$\sum_{i=1}^{5} \hat{S_{Ti}}$=1	$\sum_{i=1}^{5} \hat{S_{Ti}}$=1

6. Conclusions

A data-driven technique with multi-scale approach is utilized to accurately predict the Polymeric Graphene-Enhanced Composites' thermal conductivity, covering both meso- and macro-scales through stochastic multi-scale approach, hierarchical modeling, and interpretable machine learning (XAI). Critical parameters include filler and matrix thermal conductivity, aspect ratio, volume fraction, Kapitza resistance, agglomeration index, and dispersion index, with uncertainties quantified using probability density functions. Utilizing XGBoost for modeling and prediction, we achieve reliable results. To enhance interpretability, we apply both SHAP and LIME models to gain insights into model mechanisms and conduct sensitivity analyses to assess the impact of design parameters on material properties.

This approach improves the transparency of both data and physical models, reducing the reliance on extensive analytical modeling and simulations in material structural design. It enhances the credibility of predictions and significantly

Figure 29: First order sensitivity indices

Figure 30: Total effect sensitivity indices

lowers computational costs compared to previous multi-scale stochastic modeling methods. The key conclusions of this study are as follows:

1. The XGBoost algorithm, when coupled with PSO hyperparameter tuning and 10-fold cross-validation, exhibits robust predictive capabilities while maintaining reasonable computational efficiency.

2. Global interpretations under SHAP method indicate that the polymer matrix thermal conductivity presents the most significant influence on final outcomes, followed by the volume fraction as well as aspect ratio, with the dispersion index being the least influential.

3. On a local scale, the ultimate prediction for a particular sample is broken down into a base value and contributions from each feature, facilitating a visible quantification of the effects of individual features.

4. Furthermore, SHAP analysis elucidates how the impact of individual feature on thermal conductivity of composite changes with influence index, providing nuanced and intuitive insights, and aiding in identifying optimal ranges for composite design.

5. The analysis of feature importance in LIME shows that *Thermal Matrix (thermal conductivity of matrix)* and *Volume Fraction* are the most critical features, with average importance values of approximately 0.35 and 0.25, respectively, while other features like *Aspect Ratio*, *Kapitza*, *Agglomeration Index*, *Dispersion Index*, and *Thermal Graphene* have lower importance, indicating varying impacts on the model's predictions.

6. The analysis in LIME reveals that *Thermal Matrix (thermal conductivity of matrix)* and *Volume Fraction*, with average importance values of 0.35 and 0.25 respectively, are the most critical features, and their strong interaction (close to 1) significantly impacts the model's performance, along with notable interactions between *Thermal Matrix (thermal conductivity of matrix)* and *Aspect Ratio* (0.53) and *Volume Fraction* and *Aspect Ratio* (0.41).

7. The varying distribution of feature importance and interaction strengths across samples highlights the model's complexity, with high-strength interactions (e.g., *Thermal Matrix (thermal conductivity of matrix)* and *Volume Fraction*) indicating optimization areas, while most feature pairs exhibit low-strength interactions, suggesting minimal mutual influence.

8. The analysis in LIME shows that *Thermal Matrix (thermal conductivity of matrix)*, *Volume Fraction*, and *Aspect Ratio* are the most influential features, with varying impacts across samples. Feature interactions, particularly between *Thermal Matrix (thermal conductivity of matrix)* and *Volume Fraction*, significantly affect predictions. This variation in feature importance and direction highlights the model's complexity. Ultimately, the balance of positive and negative contributions determines the predicted values.

9. In terms of feature importance, the volume fraction and aspect ratio exert the greatest influence on overall thermal conductivity, followed by the matrix and graphene's thermal conductivity . Conversely, the agglomeration and dispersion indices exhibit lower sensitivity and have a minor impact on the results.

Limitations and Future Work

Despite the promising results achieved by the proposed stochastic multiscale machine learning framework, several limitations remain. The current study relies on idealized microstructures with simplified filler shapes and interfacial conditions, which may not fully reflect the complexity of real composites. Additionally, the model is trained on synthetic data, and further experimental validation is needed to confirm its generalizability. While SHAP and LIME provide valuable interpretability, their reliability may decrease in high-dimensional or correlated feature spaces. Future work will explore

Journal Pre-proof

more robust interpretability techniques, such as counterfactual or causal methods, and investigate alternative learning models like graph neural networks (GNNs) and physics-informed neural networks (PINNs). We also plan to incorporate real-time surrogate modeling and active learning strategies to reduce computational cost and enable more efficient material design exploration.

Acknowledgment

We sincerely appreciate the support of the Royal Swedish Academy of Forestry and Agriculture (KSLA: GFS2023-0131, BYG2023-0007, GFS2024-0155) and the J. Gust. Richert stiftelse, from SWECO, Sweden (2023–00884). This work also receives funding from several sources: the Sweden Kempe Foundation (Kempestiftelserna - Stiftelserna J.C. Kempes och Seth M. Kempes minne), the EU project H2020-AURORAL (Grant agreement ID: 101016854), the Swedish Energy Agency's E2B2 project (Project number: P2021-00248), and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas 2022-01475).

The computations were facilitated by resources from the Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at the High-Performance Computing Center North (HPC2N) under the grant agreement no. 2018-05973 and no. 2022–06725.

Data Availability

All related code used in this study will be made publicly available on GitHub for the benefit of the research community. The repository will be accessible at: https://github.com/jackylbk/ThermalConductivity_PGECs_XAI

References

[1] S. Stankovich, D. A. Dikin, G. H. Dommett, K. M. Kohlhaas, E. J. Zimney, E. A. Stach, R. D. Piner, S. T. Nguyen, R. S. Ruoff, Graphene-based composite materials, nature 442 (2006) 282–286.

[2] X. Huang, X. Qi, F. Boey, H. Zhang, Graphene-based composites, Chemical Society Reviews 41 (2012) 666–686.

[3] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, C. N. Lau, Superior thermal conductivity of single-layer graphene, Nano letters 8 (2008) 902–907.

[4] D. Ghosh, I. Calizo, D. Teweldebrhan, E. P. Pokatilov, D. L. Nika, A. A. Balandin, W. Bao, F. Miao, C. N. Lau, Extremely high thermal conductivity of graphene: Prospects for thermal management applications in nanoelectronic circuits, Applied Physics Letters 92 (2008) 151911.

[5] B. Mortazavi, M. E. Madjet, M. Shahrokhi, S. Ahzi, X. Zhuang, T. Rabczuk, Nanoporous graphene: A 2d semiconductor with anisotropic mechanical, optical and thermal conduction properties, Carbon 147 (2019) 377–384.

[6] B. Mortazavi, M. Baniassadi, J. Bardon, S. Ahzi, Modeling of two-phase random composite materials by finite element, mori–tanaka and strong contrast methods, Composites Part B: Engineering 45 (2013) 1117–1125.

[7] J. Fish, G. J. Wagner, S. Keten, Mesoscopic and multiscale modelling in materials, Nature materials 20 (2021) 774–786.

[8] T. Luo, J. R. Lloyd, Enhancement of thermal energy transport across graphene/graphite and polymer interfaces: a molecular dynamics study, Advanced Functional Materials 22 (2012) 2495–2502.

[9] M. Zain-ul Abdein, S. Azeem, S. M. Shah, Computational investigation of factors affecting thermal conductivity in a particulate filled composite using finite element method, International Journal of Engineering Science 56 (2012) 86–98.

[10] B. Zhang, R. Asmatulu, S. A. Soltani, L. N. Le, S. S. Kumar, Mechanical and thermal properties of hierarchical composites enhanced by pristine graphene and graphene oxide nanoinclusions, Journal of Applied Polymer Science 131 (2014).

[11] F. Wang, X. Zeng, Y. Yao, R. Sun, J. Xu, C.-P. Wong, Silver nanoparticle-deposited boron nitride nanosheets as fillers for polymeric composites with high thermal conductivity, Scientific reports 6 (2016) 19394.

[12] G. B. Olowojoba, S. Eslava, E. S. Gutierrez, A. J. Kinloch, C. Mattevi, V. G. Rocha, A. C. Taylor, In situ thermally reduced graphene oxide/epoxy composites: thermal and mechanical properties, Applied Nanoscience 6 (2016) 1015–1022.

[13] H. Guo, X. Li, B. Li, J. Wang, S. Wang, Thermal conductivity of graphene/poly (vinylidene fluoride) nanocomposite membrane, Materials & Design 114 (2017) 355–363.

[14] H. Fang, Y. Zhao, Y. Zhang, Y. Ren, S.-L. Bai, Three-dimensional graphene foam-filled elastomer composites with high thermal and mechanical properties, ACS applied materials & interfaces 9 (2017) 26447–26459.

[15] J. Tu, H. Li, J. Zhang, D. Hu, Z. Cai, X. Yin, L. Dong, L. Huang, C. Xiong, M. Jiang, Latent heat and thermal conductivity enhancements in polyethylene glycol/polyethylene glycol-grafted graphene oxide composites, Advanced Composites and Hybrid Materials 2 (2019) 471–480.

[16] S. Azizi, G. Momen, C. Ouellet-Plamondon, E. David, Performance improvement of epdm and epdm/silicone rubber composites using modified fumed silica, titanium dioxide and graphene additives, Polymer Testing 84 (2020) 106281.

[17] M. M. Shokrieh, R. Rafiee, Stochastic multi-scale modeling of cnt/polymer composites, Computational Materials Science 50 (2010) 437–446.

[18] N. Vu-Bac, T. Lahmer, Y. Zhang, X. Zhuang, T. Rabczuk, Stochastic predictions of interfacial characteristic of polymeric nanocomposites (pncs), Composites Part B: Engineering 59 (2014) 80–95.

[19] B. Liu, N. Vu-Bac, X. Zhuang, T. Rabczuk, Stochastic multiscale modeling of heat conductivity of polymeric clay nanocomposites, Mechanics of Materials 142 (2020) 103280.

[20] B. Liu, N. Vu-Bac, X. Zhuang, X. Fu, T. Rabczuk, Stochastic integrated machine learning based multiscale approach for the prediction of the thermal conductivity in carbon nanotube reinforced polymeric composites, Composites Science and Technology 224 (2022) 109425.

[21] B. Liu, N. Vu-Bac, T. Rabczuk, A stochastic multiscale method for the prediction of the thermal conductivity of polymer nanocomposites through hybrid machine learning algorithms, Composite Structures 273 (2021) 114269.

[22] B. Liu, N. Vu-Bac, X. Zhuang, X. Fu, T. Rabczuk, Stochastic full-range multiscale modeling of thermal conductivity of polymeric carbon nanotubes composites: A machine learning approach, Composite Structures 289 (2022) 115393.

[23] B. Liu, N. Vu-Bac, X. Zhuang, W. Lu, X. Fu, T. Rabczuk, Al-demat: A web-based expert system platform for computationally expensive models in materials design, Advances in Engineering Software 176 (2023) 103398.

[24] B. Liu, W. Lu, Surrogate models in machine learning for computational stochastic multi-scale modelling in composite materials design, International Journal of Hydromechatronics 5 (2022) 336–365.

[25] Y. Xia, C. Zhang, C. Wang, H. Liu, X. Sang, R. Liu, P. Zhao, G. An, H. Fang, M. Shi, et al., Prediction of bending strength of glass fiber reinforced methacrylate-based pipeline uv-cipp rehabilitation materials based on machine learning, Tunnelling and Underground Space Technology 140 (2023) 105319.

[26] G. X. Gu, C.-T. Chen, M. J. Buehler, De novo composite design based on machine learning algorithm, Extreme Mechanics Letters 18 (2018) 19–28.

[27] H. Wei, S. Zhao, Q. Rong, H. Bao, Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods, International Journal of Heat and Mass Transfer 127 (2018) 908–916.

[28] C.-T. Chen, G. X. Gu, Machine learning for composite materials, MRs Communications 9 (2019) 556–566.

[29] S. Dabetwar, S. Ekwaro-Osire, J. P. Dias, Damage classification of composites using machine learning, in: ASME International Mechanical Engineering Congress and Exposition, volume 83501, American Society of Mechanical Engineers, p. V013T13A017.

[30] A. Khan, N. Kim, J. K. Shin, H. S. Kim, B. D. Youn, Damage assessment of smart composite structures via machine learning: a review, JMST Advances 1 (2019) 107–124.

[31] A. Marani, M. L. Nehdi, Machine learning prediction of compressive strength for phase change materials integrated cementitious composites, Construction and Building Materials 265 (2020) 120286.

[32] S. S. Kharb, P. Antil, S. Singh, S. K. Antil, P. Sihag, A. Kumar, Machine learning-based erosion behavior of silicon carbide reinforced polymer composites, Silicon 13 (2021) 1113–1119.

Journal Pre-proof

Journal Pre-proof

[33] C. Furtado, L. Pereira, R. P. Tavares, M. Salgado, F. Otero, G. Catalanotti, A. Arteiro, M. A. Bessa, P. P. Camanho, A methodology to generate design allowables of composite laminates using machine learning, International Journal of Solids and Structures 233 (2021) 111095.

[34] B. Liu, Y. Wang, T. Rabczuk, T. Olofsson, W. Lu, Multi-scale modeling in thermal conductivity of polyurethane incorporated with phase change materials using physics-informed neural networks, Renewable Energy 220 (2024) 119565.

[35] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (xai), IEEE access 6 (2018) 52138–52160.

[36] M. T. Ribeiro, S. Singh, C. Guestrin, Model-agnostic interpretability of machine learning, arXiv preprint arXiv:1606.05386 (2016).

[37] S. Mangalathu, S.-H. Hwang, J.-S. Jeon, Failure mode and effects analysis of rc members based on machine-learning-based shapley additive explanations (shap) approach, Engineering Structures 219 (2020) 110927.

[38] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, A. K. Mohammadian, Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis, Accident Analysis & Prevention 136 (2020) 105405.

[39] B. Liu, W. Lu, T. Olofsson, X. Zhuang, T. Rabczuk, Stochastic interpretable machine learning based multiscale modeling in thermal conductivity of polymeric graphene-enhanced composites, Composite Structures 327 (2024) 117601.

[40] R. Rafiee, A. Eskandariyun, Predicting young's modulus of agglomerated graphene/polymer using multi-scale modeling, Composite Structures 245 (2020) 112324.

[41] X. Li, Y. M. Liu, W. G. Li, C. Y. Li, J. G. Sanjayan, W. H. Duan, Z. Li, Effects of graphene oxide agglomerates on workability, hydration, microstructure and compressive strength of cement paste, Construction and Building Materials 145 (2017) 402–410.

[42] L.-C. Tang, Y.-J. Wan, D. Yan, Y.-B. Pei, L. Zhao, Y.-B. Li, L.-B. Wu, J.-X. Jiang, G.-Q. Lai, The effect of graphene dispersion on the mechanical properties of graphene/epoxy composites, Carbon 60 (2013) 16–27.

[43] M. M. Shokrieh, R. Rafiee, Development of a full range multi-scale model to obtain elastic properties of cnt/polymer composites, Iranian polymer journal 21 (2012) 397–402.

[44] R. Rafiee, V. Firouzbakht, Multi-scale modeling of carbon nanotube reinforced polymers using irregular tessellation technique, Mechanics of Materials 78 (2014) 74–84.

[45] R. Rafiee, R. Shahzadi, Predicting mechanical properties of nanoclay/polymer composites using stochastic approach, Composites Part B: Engineering 152 (2018) 31–42.

[46] R. Rafiee, Influence of carbon nanotube waviness on the stiffness reduction of cnt/polymer composites, Composite Structures 97 (2013) 304–309.

[47] R. Rafiee, M. Sahraei, Characterizing delamination toughness of laminated composites containing carbon nanotubes: experimental study and stochastic multi-scale modeling, Composites Science and Technology 201 (2021) 108487.

[48] R. L. Iman, W. Conover, Small sample sensitivity analysis techniques for computer models. with an application to risk assessment, Communications in statistics-theory and methods 9 (1980) 1749–1842.

[49] H. Keitel, Bewertungsmethoden für die Prognosequalität von Kriechmodellen des Betons, doctoralthesis, Bauhaus-Universität Weimar, 2012.

[50] D. Novák, B. Teplý, Z. Keršner, The role of latin hypercube sampling method in reliability engineering, in: Proc. of ICOSSAR, volume 97, pp. 403–409.

[51] S. Ghosh, W. Bao, D. L. Nika, S. Subrina, E. P. Pokatilov, C. N. Lau, A. A. Balandin, Dimensional crossover of thermal transport in few-layer graphene, Nature materials 9 (2010) 555.

[52] A. Moisala, Q. Li, I. Kinloch, A. Windle, Thermal and electrical conductivity of single-and multi-walled carbon nanotube-epoxy composites, Composites science and technology 66 (2006) 1285–1288.

[53] M. Freitag, M. Steiner, Y. Martin, V. Perebeinos, Z. Chen, J. C. Tsang, P. Avouris, Energy dissipation in graphene field-effect transistors, Nano letters 9 (2009) 1883–1888.

[54] K. Bui, B. P. Grady, D. V. Papavassiliou, Heat transfer in high volume fraction cnt nanocomposites: Effects of inter-nanotube thermal resistance, Chemical Physics Letters 508 (2011) 248–251.

[55] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

[56] C. Molnar, Interpretable machine learning, Lulu. com, 2020.

[57] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Definitions, methods, and applications in interpretable machine learning, Proceedings of the National Academy of Sciences 116 (2019) 22071–22080.

[58] D. Slack, S. Hilgard, E. Jia, S. Singh, H. Lakkaraju, Fooling lime and shap: Adversarial attacks on post hoc explanation methods, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 180–186.

[59] D. Garreau, U. Luxburg, Explaining the explainer: A first theoretical analysis of lime, in: International conference on artificial intelligence and statistics, PMLR, pp. 1287–1296.

[60] S. M. Lundberg, G. G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, arXiv preprint arXiv:1802.03888 (2018).

[61] E. Štrumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, Knowledge and information systems 41 (2014) 647–665.

[62] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135–1144.

[63] M. R. Zafar, N. Khan, Deterministic local interpretable model-agnostic explanations for stable explainability, Machine Learning and Knowledge Extraction 3 (2021) 525–541.

[64] U. Schlegel, D. L. Vo, D. A. Keim, D. Seebacher, Ts-mule: Local interpretable model-agnostic explanations for time series forecast models, in: Joint European conference on machine learning and knowledge discovery in databases, Springer, pp. 5–14.

[65] H. Prabhu, A. Sane, R. Dhadwal, N. R. Parlikkad, J. K. Valadi, Interpretation of drop size predictions from a random forest model using local interpretable model-agnostic explanations (lime) in a rotating disc contactor, Industrial & Engineering Chemistry Research 62 (2023) 19019–19034.

[66] M. Toğaçar, N. Muzoğlu, B. Ergen, B. S. B. Yarman, A. M. Halefoğlu, Detection of covid-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from cnns, Biomedical Signal Processing and Control 71 (2022) 103128.

[67] B. Liu, S. R. Penaka, W. Lu, K. Feng, A. Rebbling, T. Olofsson, Data-driven quantitative analysis of an integrated open digital ecosystems platform for user-centric energy retrofits: A case study in northern sweden, Technology in Society 75 (2023) 102347.

[68] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

[69] Z. Yang, J. Che, Z. Zhang, L. Yu, M. Hu, W. Sun, W. Gao, J. Fan, L. Wang, G. Liu, High-efficiency graphene/epoxy composite coatings with outstanding thermal conductive and anti-corrosion performance, Composites Part A: Applied Science and Manufacturing 181 (2024) 108152.

[70] Y. S. Lee, N. R. Kim, S. K. Park, Y.-i. Ko, Y. Shin, B. Yang, C.-M. Yang, Effects of high-temperature thermal reduction on thermal conductivity of reduced graphene oxide polymer composites, Applied Surface Science 650 (2024) 159140.

[71] J. Gong, Z. Liu, J. Yu, D. Dai, W. Dai, S. Du, C. Li, N. Jiang, Z. Zhan, C.-T. Lin, Graphene woven fabric-reinforced polyimide films with enhanced and anisotropic thermal conductivity, Composites Part A: Applied Science and Manufacturing 87 (2016) 290–296.

Journal Pre-proof

Highlights

- **A multi-scale framework is developed to predict the thermal conductivity of graphene-based polymer nanocomposites, integrating meso- and macro-scale analyses for comprehensive material characterization.**
- **SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) provide actionable insights into the impact of design parameters on material performance through sensitivity analysis.**
- **The Explainable machine learning approach bridges data-driven and physics-based modeling, minimizing reliance on complex analytical simulations in material design.**
- **The proposed methodology significantly enhances predictive accuracy while reducing computational costs compared to traditional stochastic multi-scale modeling techniques.**

Journal Pre-proof

Journal Pre-proof

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Best regards,

Timon Rabczuk

Journal Pre-proof