



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/239434/>

Version: Accepted Version

Article:

Al-Hadidi, H., Abuzayed, I.H., Zhang, C. et al. (2026) Micromechanical modelling and machine learning approaches for predicting effective properties of composite materials. Composite Structures, 375. 119767. ISSN: 0263-8223

<https://doi.org/10.1016/j.compstruct.2025.119767>

© 2025 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in Composite Structures is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Micromechanical Modelling and Machine Learning Approaches for Predicting Effective Properties of Composite Materials

Hisham Al-Hadidi^{a,b,*}, Ibrahim H. Abuzayed^a, Chao Zhang^c, Jose L. Curiel-Sosa^a

^a*School of Mechanical, Aerospace and Civil Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 4DT, United Kingdom*

^b*Marine Engineering Department, Military Technological College, Al Matar Road, Al Seeb, Muscat, 111, Oman*

^c*School of Mechanical Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu, 212013, China*

Abstract

Accurately predicting the effective properties of composite materials is critical for their design and optimisation in advanced applications. This study integrates micromechanical modelling and machine learning to develop novel efficient predictive frameworks for composite properties. Using Representative Volume Elements (RVEs) with periodic boundary conditions, a dataset of effective moduli was generated through finite element simulations. Machine learning models, including Random Forest, Gradient Boosting, and Support Vector Regressor (SVR), were tested, with Random Forest performing best in capturing complex material behaviours. A surrogate model coupled with active learning was developed for the first time in this method to enhance computational efficiency, iteratively refining predictions by targeting regions of high uncertainty. Validation against theoretical models and experimental data highlights the superior accuracy of the proposed framework, particularly in predicting transverse and shear properties. This research established a robust methodology combining micromechanics and machine learning, paving the way for advanced composite material design with reduced computational cost and enhanced predictive precision.

Keywords: Micromechanics, Representative Volume Element (RVE), Periodic Boundary Conditions (PBC), Machine Learning, Artificial Intelligence, Surrogate Modelling, Active Learning

*Corresponding author: hsal-hadidi1@sheffield.ac.uk; hisham.alhadidi@mtc.edu.om

1. Introduction

Composite materials, renowned for their superior mechanical properties such as high strength-to-weight ratios and enhanced durability, have become indispensable across various industries including aerospace, automotive engineering, construction, and sports equipment. However, accurately predicting their structural properties remains a significant computational challenge. This complexity arises from the intricate relationships between process, structure, property, and performance inherent in composites, necessitating sophisticated analytical and numerical tools to model them effectively [1].

To address these challenges, numerous analytical, semi-analytical, and finite element (FE) methods have been developed over the years to predict the effective properties of composite materials. Early foundational works by Voigt [2] and Reuss [3] introduced average methods for polycrystals under assumptions of uniform stress or strain, establishing upper and lower bounds for elastic moduli. Hill [4] later refined these methods, providing more enhanced accuracy, while Hashin and Shtrikman [5, 6] introduced variational principles to derive even tighter bounds, considering representative volume elements (RVEs) extracted from periodic heterogeneous composites.

Despite these advancements, many models rely on simplifying assumptions such as homogeneous, parallel, and continuously reinforced fibers; no slippage between fiber and matrix; and uniform stress and strain conditions [7, 8]. These assumptions often do not hold in real-world applications, leading to discrepancies between predicted and observed behaviours. Recognising this, Suquet [9] highlighted that applying uniform strain or stress conditions at RVE boundaries may not capture the true behaviour of composites, advocating for the implementation of periodic boundary conditions (PBCs) to enhance modelling accuracy.

In response to the need for more precise predictive models, recent efforts have focused on benchmarking damage modelling practices for unidirectional (UD) fiber-reinforced polymer (FRP) composite laminates [10]. These efforts underscore the importance of considering multiscale structural features and establishing robust relationships between the constituent microstructure and the overall composite structure. Multiscale modelling methods for FRP composites typically involve three distinct length scales: macroscopic, mesoscopic, and microscopic, each characterised by unique geometric features [11]. By integrating analyses across these scales, detailed

insights into the material’s properties and performance can be obtained, facilitating a comprehensive understanding of the composite’s behaviour under various conditions [12]. This multiscale approach is essential for accurately predicting mechanical behaviour and failure mechanisms, ultimately improving the design and optimisation of composite materials for specific applications.

A significant challenge in multiscale finite element modelling is balancing computational load with accuracy. To mitigate demands, researchers have often simplified RVEs into unit cells by assuming repeating fibre patterns, such as square or hexagonal packing [13, 14, 15, 16]. While these approaches provide acceptable estimates of effective elastic properties in homogenisation [17], they may underestimate local damage onset and ultimate failure predictions in microscale modelling due to neglecting microstructural variability and material property distributions. Analytical models have followed a similar path: closed-form solutions for idealised uniaxial fibre arrays, such as those derived by Guinovart-Díaz et al. [18], offer fast and accurate predictions under assumptions of perfect alignment and hexagonal packing, but their applicability becomes limited when fibre distributions are random or variability is significant. More recent efforts, such as those by Ramírez-Torres et al. [19], have introduced inelastic distortions to better capture variability in fibre-reinforced composites. Building on these advances, the present framework retains unidirectional continuous fibres for tractability yet moves beyond analytical models by embedding randomly distributed fibre arrangements and periodic boundary conditions to better capture microstructural variability.

Given these challenges, a growing need exists to optimise computational methods for predicting composite material properties. Machine learning (ML) models present a promising solution to improve prediction efficiency without compromising accuracy. Recent studies have demonstrated the breadth of ML applications in composites, including optimisation of filler selection and mechanical performance [20], prediction of impact behaviour in layered and hybrid composites [21, 22], and comprehensive reviews highlighting the versatility of ML for property prediction, optimisation, and uncertainty quantification [23]. Deep learning architectures such as convolutional neural networks and U-Nets have been applied to predict local stress and strain fields [24, 25], while artificial neural networks have been used to accelerate the evaluation of electromechanical properties in piezoelectric composites [26]. Other approaches include ML-driven identification of composite damage mechanisms [27], recurrent neural networks for nonlinear elasto-plastic behaviour in multiscale simulations [28], and methods to predict transverse strength of UD-CFRP based on

fibre distribution [29]. Collectively, these works underscore the potential of ML in composite modelling but also highlight that many methods require large datasets or intensive training. To address these limitations, active learning strategies have been developed to reduce dataset requirements while preserving predictive accuracy.

Active learning, successfully implemented in various fields [30, 31, 32], enhances surrogate modelling by introducing an iterative refinement process that strategically improves model performance. It selectively identifies regions of high uncertainty in the model’s predictions and focuses computational resources on these areas, improving accuracy and efficiency. This targeted strategy reduces the need for extensive simulations by adding new data points where the model’s predictions are less certain.

Building upon the need for accurate and efficient modelling of composite materials, this paper sets forth two primary objectives. The first is to predict the effective moduli of composite materials by exploring a variety of machine learning (ML) models, to identify the most precise and reliable predictor. The second objective focuses on leveraging the top-performing machine learning model to develop a surrogate modelling framework enhanced with active learning techniques, aiming to further reduce computational costs. To the best of the author’s knowledge, such a method has not been previously applied for this purpose. By accomplishing these goals, we aim to establish a robust and efficient framework that provides dependable predictions of composite material properties.

2. Generation of Representative Volume Element (RVE)

Numerical modelling methods are widely utilised to predict the thermo-mechanical properties fiber-reinforced composites. A critical prerequisite for these methods is the accurate generation of Representative Volume Elements (RVEs) that capture the composite’s microstructural features. RVEs provide a robust framework for linking microscale characteristics to macroscale behaviour, enabling reliable prediction of effective material properties which is an essential aspect of multiscale modelling. Consequently, RVE generation techniques have become a key area of focus in the scientific community [33, 34, 35].

2.1. Algorithms for RVE Generation

To accurately simulate the microstructure of unidirectional carbon fiber-reinforced polymers (UD CFRPs) for composite analysis, researchers have developed various algorithms to generate artificial Representative Volume Elements (RVEs) with random

fiber packing [33, 36, 37]. These methods can be broadly categorised into two primary approaches: image-based techniques and algorithmic generation methods.

Image-based methods reconstruct microstructures from cross-sectional images, effectively capturing the real configurations of composites. For example, the microstructure of UD carbon fiber-reinforced composites has been statistically studied based on scanning electron microscopy (SEM) images of cross-sections, leading to the development of algorithms that generate numerical microstructures with equivalent random distributions of fibers [38]. Such methods are highly effective in representing actual microstructural features but are computationally intensive due to the multiple stages required for image acquisition, processing, and analysis especially when numerous configurations are needed for statistical reliability [37, 39, 40]. Additionally, a systematic approach has been employed to determine the RVE of glass fiber/polypropylene (GF/PP) composites using two methods: geometrical two-point probability function analysis and Hill–Mandel kinematic homogenisation [41].

In contrast, algorithmic generation methods create microstructures numerically, often focusing on ensuring statistical similarity to real configurations. These methods utilise statistical functions, like the nearest neighbour distance distribution, to validate the generated structures, aiming to create a Statistically Equivalent RVE (SERVE) [42]. A widely used algorithm in this category is Random Sequential Adsorption (RSA), which places fibers or particles randomly within the RVE while maintaining spatial constraints. This results in a statistically representative distribution of fibers akin to actual composites [33, 43, 44, 45, 46].

RSA involves placing random points within an RVE while ensuring a minimum separation distance between them. For instance, Yang et al. [47] validated the statistical equivalency of simulated microstructures for ceramic matrix composites using RSA. However, RSA can encounter a jamming limit at around 54.7% fibre volume fraction¹ for random sequential adsorption of discs in 2D, this limit is approximately 54.7% [48]., restricting how dense fibres can be packed. Modified RSA algorithms have been developed to address this limitation. Segurado and Llorca [43] used a modified RSA algorithm to generate three-dimensional RVE models of randomly distributed spherical particles. They generated three-dimensional cubic unit cells containing 30 non-overlapping identical spheres, suitable for particle volume frac-

¹The jamming limit refers to the maximum achievable packing density using RSA, beyond which no additional fibres can be added without violating the non-overlap condition.

tions of up to 50%. This modified RSA algorithm allows for higher packing densities while avoiding particle overlap.

Additionally, a shaking technique is sometimes introduced in RSA, where each fibre is slightly displaced to allow higher packing densities². A comprehensive review [49] examined several shaking algorithms aimed at achieving denser configurations in RSA-generated microstructures. Other techniques build upon RSA by adding disturbances to initially regular fibre arrangements, creating more realistic, non-uniform distributions. For example, several studies [50, 51, 52] introduced disturbances into hexagonal or square fibre patterns to simulate randomness.

A three-step Random Microstructure Generation (RMG) algorithm was proposed [53], combining RSA for initial placement, random redistribution of fibers, and additional fiber placement in matrix-rich areas, achieving fiber volume fractions up to 65%. This approach was validated through extensive statistical analysis [54], demonstrating its similarity to real fiber distributions. Moreover, the three-step RMG method has been shown to meet the SERVE criteria proposed in earlier work [42].

A notable approach is the Nearest Neighbour Algorithm (NNA), developed by Vaughan and McCarthy [37]. This algorithm combines experimental and numerical methods, using statistical data to generate distributions that mimic both short- and long-range fiber interactions. Although computationally efficient, NNA sometimes results in a higher frequency of small inter-fiber distances and fewer larger distances than typically observed in actual composite structures. To address this limitation, a modified NNA method was proposed by Wang et al. [55], enhancing the nearest neighbor distribution to create statistically equivalent random RVEs that better reflect real microstructures.

Yang et al. [56] introduced the Random Sequential Expansion (RSE) algorithm, a modification of the Random Sequential Addition (RSA) method that incorporates adjustable distances between fibers to achieve higher volume fractions. The algorithm can be adapted and implemented straightforwardly, RSE gained popularity in the field [57, 58]. Despite these advantages, the algorithm faces significant lim-

²The shaking technique involves small random displacements of fibres after their initial placement using RSA to reduce gaps and increase packing density. A comprehensive review of such algorithms is provided in [49]

itations: it cannot exceed a maximum fiber volume fraction of 65%, and it tends to cause fiber clustering in central regions while leaving matrix-rich areas near the boundaries. These issues arise due to constraints related to the nearest-neighbor distances within the algorithm[56].

Recent developments include the discrete element method, which generates RVEs with fibers of varying radii, enhancing the realism of simulated microstructures [59]. Researchers have also compared simulated and real fiber arrangements found in unidirectional composites using statistical descriptors, aiming to validate numerical models against actual microstructures [54]. Reconstructing RVEs from real micrographs plays a crucial role in this validation process, allowing for direct comparison between simulated and actual fiber distributions.

To ensure the generated microstructures are statistically equivalent to real composites, Liu and Ghosh [60] provided a framework for assessing statistical and geometric equivalency of microstructures, including probability functions, size convergence, and fiber radius distribution. This framework offers a standardised method for RVE validation, bridging the gap between simulated models and real-world materials.

While image-based techniques offer high fidelity in capturing real microstructures, they are computationally intensive and less practical for generating numerous configurations. Algorithmic generation methods provide a more efficient alternative, with various algorithms developed to overcome the limitations of RSA and achieve higher fiber volume fractions. Modified RSA algorithms and the incorporation of statistical descriptors have enhanced the ability to generate RVEs that accurately represent the microstructural characteristics of UD CFRPs. Validation through statistical equivalence ensures that the generated RVEs are reliable for composite analysis.

Numerical modelling methods are widely utilised to predict the thermo-mechanical properties of particle-reinforced composites. A critical prerequisite for these methods is the accurate generation of Representative Volume Elements (RVEs) that capture the composite’s microstructure. Consequently, RVE generation methods have become a key area of focus in the scientific community [33, 34, 35].

To address the limitations of the Random Sequential Adsorption (RSA) algorithm in achieving high fiber volume fractions, we adopt the Molecular Dynamics (MD)-based algorithm introduced by Ghossein and Lévesque [61]. In this method, inclusions are initially introduced into the matrix with zero radius and are set in

motion while their radii gradually increase. During this expansion and movement, inclusions may collide with one another or interact with the boundaries of the RVE. When collisions occur, the velocities of the inclusions are adjusted accordingly, and periodic boundary conditions are applied to handle interactions with the RVE edges. This process continues until the target inclusion volume fraction is achieved. The MD-based algorithm efficiently generates RVEs with high volume fractions, reaching values close to the theoretical dense packing limit, all while maintaining low computational costs.

In the next section, a brief explanation and comparison of the RSA and MD-based algorithms, highlighting their respective advantages and applicability in generating realistic microstructures for composite materials.

2.1.1. Random Sequential Adsorption (RSA)

The Random Sequential Adsorption (RSA) algorithm is one of the simplest and most commonly used methods for generating RVEs [62, 63]. In this method, particles are randomly positioned and sequentially added into the RVE, ensuring no overlaps occur with previously placed particles. The condition for each newly added particle i and existing particle j is that the distance between their centers d_{ij} must satisfy:

$$d_{ij} \geq r_i + r_j \quad (1)$$

where r_i and r_j are the radii of particles i and j , respectively. The process continues until the system reaches a maximum particle volume fraction (PVF), also known as the jamming limit. As the system approaches this limit, the coverage at the jamming state is given by:

$$\theta(t) = \theta(\infty) - \frac{c}{t^{1/2}} \quad (2)$$

where t represents the number of attempts to place new particles, and $\theta(t)$ represents the fraction of the surface particles covered at time t . The coverage saturates asymptotically as $t^{-1/2}$ [64], indicating that the system slows down as it approaches the jamming limit. In two dimensions, for disks, the maximum achievable coverage is approximately $\theta(\infty) = 0.547 \pm 0.003$ [64].

2.1.2. Molecular Dynamics (MD)-based Method

To overcome the inherent limitations of RSA (e.g., low maximum packing density due to the jamming limit and lack of particle rearrangement), a Molecular Dynamics

(MD)-based algorithm is implemented. In this approach, particles are initially placed in the RVE with a radius $r_0 = 0$, and their size increases dynamically while they are allowed to move and collide. The time t at which two particles i and j collide is determined by solving the following quadratic equation:

$$at^2 + 2bt + c = 0 \quad (3)$$

where the coefficients are defined as:

$$a = \|\Delta \mathbf{v}_n\|^2 - (a_i + a_j)^2 \quad (4)$$

$$b = (\Delta \mathbf{r}_n \cdot \Delta \mathbf{v}_n) - (R_i^n + R_j^n)(a_i + a_j) \quad (5)$$

$$c = \|\Delta \mathbf{r}_n\|^2 - (R_i^n + R_j^n)^2 \quad (6)$$

In these equations, $\Delta \mathbf{r}_n$ is the relative position between the centers of the particles, and $\Delta \mathbf{v}_n$ is the relative velocity. a_i and a_j are the growth rates of the particles, while R_i^n and R_j^n are their radii at time n [61].

When two particles collide, their velocities are updated based on the kinetic energy and growth rates. The new velocities $\mathbf{v}_i^{n+1}$ and $\mathbf{v}_j^{n+1}$ of particles i and j after a collision are calculated using:

$$\mathbf{v}_i^{n+1} = \mathbf{v}_j^n + (a_i + a_j)\hat{\mathbf{u}} + \mathbf{v}_i^\perp \quad (7)$$

$$\mathbf{v}_j^{n+1} = \mathbf{v}_i^n - (a_i + a_j)\hat{\mathbf{u}} + \mathbf{v}_j^\perp \quad (8)$$

where $\hat{\mathbf{u}}$ is the unit vector along the line connecting the centers of the two colliding particles, and $\mathbf{v}_i^\perp$ represents the perpendicular component of the velocity, which remains unchanged during the collision [61].

The MD method continues until the desired PVF is achieved. This approach enables efficient generation of RVEs with high PVFs, often up to and exceeding the theoretical dense packing limit. The PVF achieved through MD can reach values greater than 80%, significantly surpassing the jamming limit of RSA:

$$\phi_{\text{MD}} > \phi_{\text{RSA}} \approx 0.5 \quad (9)$$

By incorporating the MD algorithm after an initial RSA placement, the system can achieve higher volume fractions while maintaining non-overlapping particles.

This leads to greater computational efficiency and more realistic modelling of dense composite microstructures. As shown in Figure 1, the algorithm effectively generates microstructures for different RVE sizes, demonstrating uniform particle distributions at varying scales. Figure 2 illustrates the successful integration of these microstructures into Abaqus, highlighting the mesh distribution across the fibers and matrix within the RVE. A comprehensive analysis of the Molecular Dynamics (MD) algorithm for RVE generation is provided in Ref [61, 65].

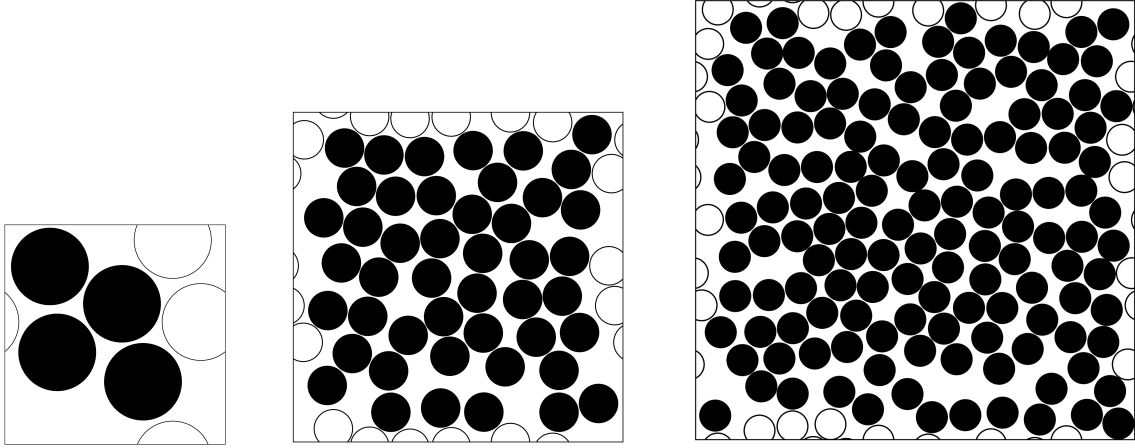


Figure 1: Microstructures generated by the algorithm for 20 μm , 60 μm , and 100 μm .

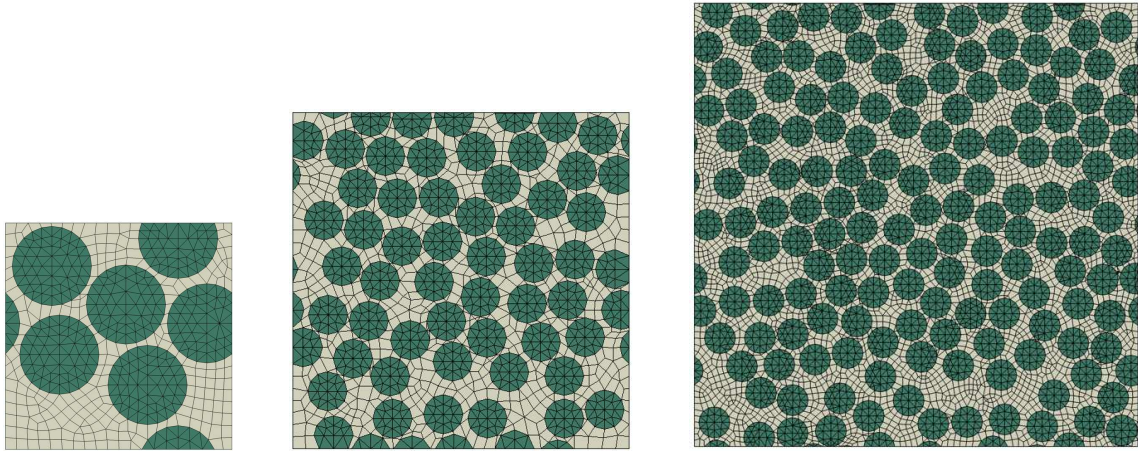


Figure 2: Corresponding Abaqus models demonstrating successful integration for 20 μm , 60 μm , and 100 μm .

3. Micromechanics Modelling

3.1. Fiber Constitutive Model

The constitutive modelling of unidirectional fiber-reinforced lamina can be effectively approached by treating the lamina as a transversely isotropic material. In such materials, due to their symmetry around the fiber axis, the number of independent elastic constants reduces to five. Specifically, the transverse elastic moduli are equal ($E_2 = E_3$), the shear moduli in the planes containing the fiber direction are identical ($G_{12} = G_{13}$), and the Poisson's ratios are the same in these planes ($\nu_{12} = \nu_{13}$). Additionally, the shear modulus in the plane perpendicular to the fiber direction is defined by

$$G_{23} = \frac{E_2}{2(1 + \nu_{23})}. \quad (10)$$

Therefore, AS4 carbon fibers were modeled as linear elastic and transversely isotropic solids. The properties of the AS4 carbon fibers were employed in the analysis, and their values are shown in Table 1.

Table 1: Material properties of AS4 carbon fibers [66].

Property	Value
Longitudinal Young's Modulus, E_1 (GPa)	225
Transverse Young's Modulus, E_2 (GPa)	15
Poisson's Ratio, ν_{12}	0.20
Shear Modulus, G_{12} (GPa)	15
Shear Modulus, G_{23} (GPa)	7

3.2. Matrix Constitutive Model

The matrix material used in this study is 3501-6 epoxy, modelled as a linear elastic isotropic solid. The mechanical properties of the matrix are provided in Table 2.

Table 2: Mechanical properties of the epoxy matrix [66].

Property	Value
Elastic Modulus, E_m (GPa)	4.2
Poisson's Ratio, ν_m	0.34

3.3. Periodic Boundary Conditions (PBC)

In the micromechanical simulations, periodic boundary conditions (PBC) were applied to ensure that the representative volume element (RVE) accurately reflects the behaviour of an infinite material. PBC enforce the continuity of displacement fields across the boundaries of the RVE, ensuring that adjacent RVEs fit seamlessly together. This approach is critical for simulating composite materials where small-scale RVEs are used to represent the overall behaviour.

To impose periodic boundary conditions, the displacements on opposite faces of the RVE must be related. For an RVE with dimensions $w_0 \times L_0 \times L_0$, the PBC were enforced using the following equations, ensuring continuity in the mechanical fields.

For displacement in the X_1 -direction:

$$\vec{u}(0, X_2, X_3) - \vec{u}(w_0, X_2, X_3) = U_1, \quad (11)$$

where $\vec{u}(X_1, X_2, X_3)$ represents the displacement at a point on the RVE boundary, and U_1 is the displacement of the master node in the X_1 -direction. This ensures that the displacements at the front and back faces of the RVE in the X_1 -direction are equal, up to a global translation given by U_1 .

For displacement in the X_2 -direction:

$$\vec{u}(X_1, 0, X_3) - \vec{u}(X_1, L_0, X_3) = U_2, \quad (12)$$

where U_2 is the master node displacement in the X_2 -direction. This equation ensures continuity of displacements between the top and bottom faces of the RVE along the X_2 -axis.

For displacement in the X_3 -direction:

$$\vec{u}(X_1, X_2, 0) - \vec{u}(X_1, X_2, L_0) = U_3, \quad (13)$$

where U_3 represents the master node displacement in the X_3 -direction. This ensures continuous displacements on the front and back faces in the X_3 -direction.

In these equations, X_1, X_2, X_3 represent the coordinates of the points on the RVE boundaries, while U_1, U_2, U_3 are the displacements at the master nodes in their respective directions. These conditions ensure that the mechanical fields are periodic across the boundaries, enabling the RVE to behave as part of an infinite material

structure.

The periodic boundary conditions thus provide a realistic and efficient way to model the mechanical response of fibre-reinforced composites while minimising the computational cost associated with simulating large-scale structures [67].

3.4. Micromechanical Analysis and Simulation Framework

The micromechanical analysis of composites, often performed using Representative Volume Elements (RVEs), provides a framework for capturing the heterogeneous microstructure of composites and their material responses at the microscale, enabling the calculation of effective macroscopic properties. In this study, 50 simulations were conducted on RVEs with varying sizes for initial datasets, ranging from 20 to 100 micrometres, and fibre volume fractions (V_f) spanning from 30% to 60%, to predict the effective mechanical properties of fibre-reinforced composites. These simulations generated data for mechanical properties, including E_{11} , E_{22} , G_{12} , G_{23} , and ν_{12} , which were used to develop a predictive machine learning (ML) model. The model aims to estimate the effective moduli of the composites based on RVE size and fibre volume fraction, providing a robust tool for accurately predicting composite behaviour.

4. Machine Learning for Composite Material Property Prediction

4.1. Machine Learning Models

Machine learning (ML) offers powerful tools for modelling complex, non-linear relationships between microstructural features and the effective properties of composite materials. In this study, supervised learning techniques are employed to develop predictive models capable of estimating the homogenised mechanical properties based on input parameters such as fibre volume fraction and RVE size. The following subsection presents the selected ML algorithms used for this purpose and outlines their underlying principles.

- **Random Forest Regressor:** A powerful ensemble learning method that builds multiple decision trees and averages their predictions. This model typically handles non-linear relationships well and reduces overfitting [68].
- **Gradient Boosting Regressor:** Another ensemble model that builds decision trees sequentially. Each subsequent tree corrects the errors of the previous ones, leading to improved performance [69].

- **Support Vector Regressor (SVR):** A regression version of Support Vector Machines, which aims to fit the best possible regression line within a margin of tolerance [70].

Each model’s training process and evaluation metrics, such as accuracy and R^2 scores, are described.

4.2. Model Training and Evaluation

The performance of four machine learning models: Random Forest, Gradient Boosting and Support Vector Regressor (SVR) was evaluated using Mean Squared Error (MSE) and R^2 score as metrics. The formulas for these metrics are:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (15)$$

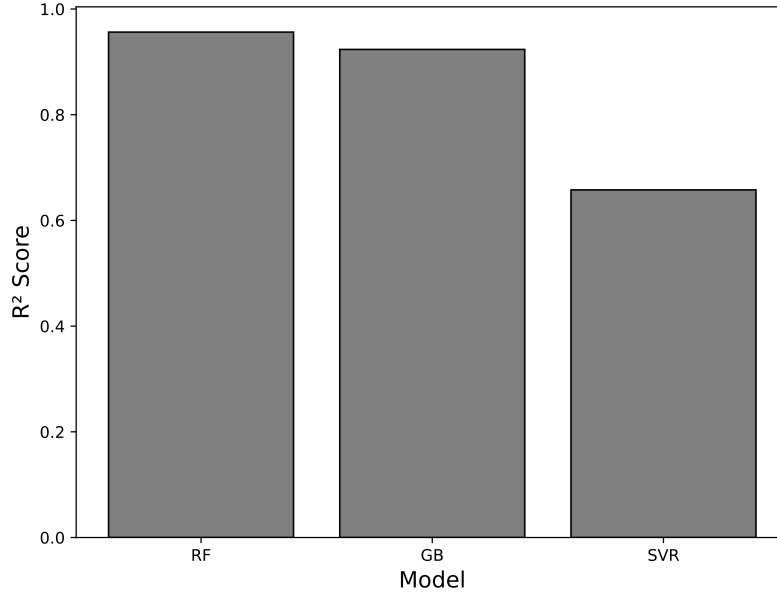
where N is the total number of data points, y_i represents the observed values (in this study corresponding to the effective properties obtained from the FE (RVE) simulations described in Section 3.4, which serve as the ground truth for training and evaluation), $\hat{y}_i$ are the predicted values, and $\bar{y}$ is the mean of the observed values, given by

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i. \quad (16)$$

These metrics provide understanding into the accuracy and predictive capability of each model, with lower MSE values indicating closer alignment between predicted and actual values, and higher R^2 scores demonstrating better fit to the data. The Random Forest model yielded the best performance, achieving an MSE of 0.0019 and an R^2 score of 95.63%, highlighting its strength in capturing the underlying relationships in the data. Gradient Boosting also performed well, with an MSE of 0.0040 and an R^2 score of 92.36%, albeit slightly less accurate than Random Forest. In contrast, the Support Vector Regressor displayed moderate accuracy with an MSE of 0.0156 and an R^2 score of 65.78%. These results are detailed in Table 3 and visualised in Figure 3, providing a comparative view of model efficacy based on these performance metrics.

Table 3: Model Performance Comparison for Various Machine Learning Models

Model	Mean Squared Error (MSE)	R ² Score
Random Forest	0.0019	95.63%
Gradient Boosting	0.0040	92.36%
Support Vector Regressor (SVR)	0.0156	65.78%

Figure 3: Model Performance Comparison (R² Scores)

The predicted vs. actual plots in Figure 4 provide a visual evaluation of the Random Forest model's performance across various mechanical properties (E_{11} , E_{22} , G_{12} , G_{23} , and ν_{12}). Ideally, points should lie close to the red diagonal line, which represents perfect prediction accuracy. In the case of properties like E_{11} and E_{22} , the data points align relatively closely with this line, indicating that the model has captured the underlying trends and provides accurate predictions. However, for properties such as G_{12} and G_{23} , there is more scatter around the diagonal, suggesting that the model has some difficulty in precisely predicting these values.

For ν_{12} , the points display a strong alignment with the diagonal, showing that the model performs well in predicting this property. Overall, these plots demonstrate the effectiveness of the Random Forest model in capturing complex relationships for most properties, though certain areas, like shear moduli, may benefit from further

refinement or additional data to enhance prediction accuracy.

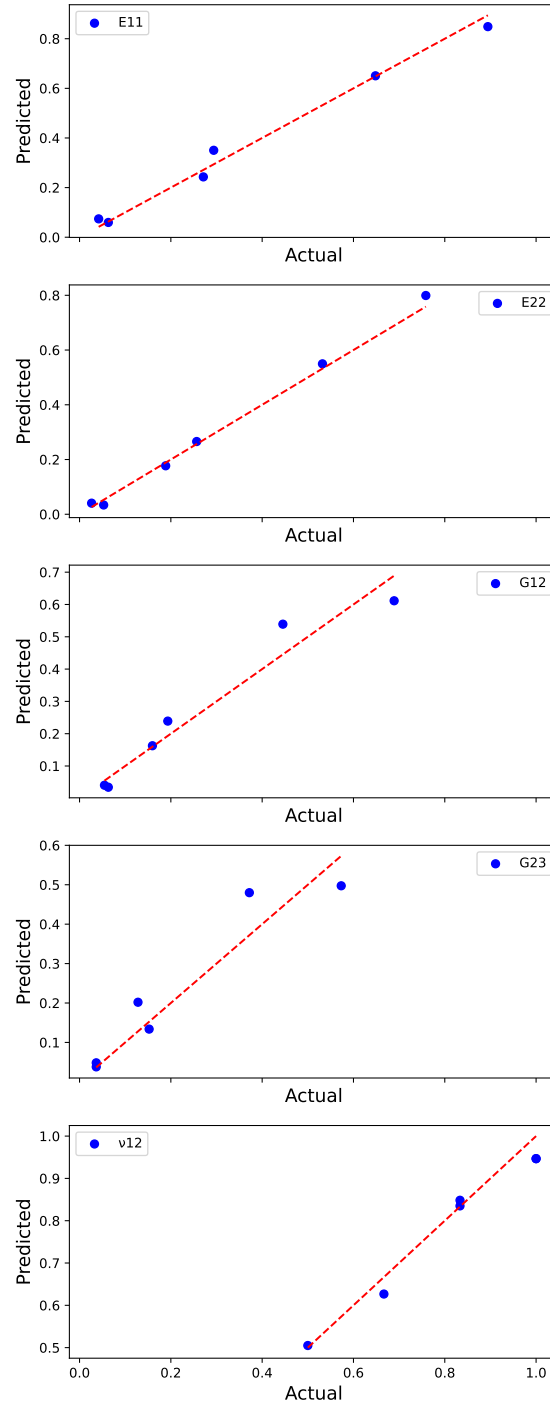


Figure 4: Predicted vs Actual Values for the Best Model (Random Forest) using Normalised Data

5. Surrogate Modelling with Active Learning for Efficient Predictions

5.1. Surrogate Model Creation and Active Learning Integration

To reduce the computational cost associated with full RVE simulations, the Random Forest Regressor is adapted as a surrogate model. This surrogate model approximates the complex relationships between input parameters (e.g., fiber volume fraction, RVE size) and output mechanical properties, enabling rapid predictions without the need for time-intensive finite element (FE) analyses. By capturing these relationships effectively, the surrogate model serves as a reliable alternative to traditional simulations, particularly valuable for parameter sweeps, optimisation studies, and sensitivity analyses.

The surrogate model is created within a Python-based workflow that integrates machine learning with FE simulations to predict composite material properties efficiently. The Random Forest Regressor is trained on a dataset generated from a range of FE simulations, capturing various parameter combinations. The model's performance is validated using metrics such as Mean Squared Error (MSE) and R^2 score to ensure it accurately approximates FE simulation results. Once validated, the model is incorporated into the workflow, providing fast and efficient predictions with minimal deviation from full simulations.

To further enhance prediction accuracy, active learning, which has been successfully implemented in various fields [30, 31, 32], is employed to iteratively refine the surrogate model. This active learning approach selectively identifies regions of high uncertainty in the model's predictions and prioritises these areas for additional FE simulations. By focusing computational resources on these uncertain regions, active learning balances computational efficiency with predictive precision, effectively reducing the need for random sampling and making the refinement process more targeted. When the model encounters high-uncertainty regions, new simulation data is generated, incorporated into the training set, and the model is retrained, improving its predictive capability over successive iterations.

This combined approach of surrogate modelling and active learning allows the Random Forest model to predict complex FE simulation results while progressively enhancing its accuracy. The iterative process continues until the surrogate model achieves a predefined accuracy threshold across the parameter space, resulting in a robust framework capable of reliable, efficient predictions for composite material properties. This method leverages the power of machine learning to create a surrogate that not only accelerates the prediction process but also optimises computational

resources by minimising the need for extensive simulations.

5.2. Workflow Structure and Methodology

5.2.1. Loading Initial Data

The code begins with a dataset containing two input features and five target outputs:

- **Inputs:**

- *RVE Size* (μm): The size of the Representative Volume Element.
- *Vf* (%): The volume fraction of fiber in the composite.

- **Outputs** (Material Properties):

- *Elastic Moduli*: E_{11} and E_{22} in GPa.
- *Shear Moduli*: G_{12} and G_{23} in GPa.
- *Poisson's Ratio*: ν_{12} .

5.2.2. Data Normalisation

Normalisation scales data to a consistent range, enhancing model performance. Given an input feature X and a target feature Y , Min-Max normalisation is applied as follows:

$$X_{\text{normalised}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (17)$$

$$Y_{\text{normalised}} = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (18)$$

where $X_{\min}$ and $X_{\max}$ represent the minimum and maximum values in the dataset. This scaling ensures all features contribute equally during model training.

5.2.3. Training the Surrogate Model

A Random Forest Regressor learns the relationship between input features and material properties. After training, the model can predict properties based on new RVE sizes and volume fractions.

Given the input vector $\mathbf{X} = [\text{RVE Size}, Vf]$, the trained model outputs predictions for the target vector $\mathbf{Y} = [E_{11}, E_{22}, G_{12}, G_{23}, \nu_{12}]$. The predicted output for each property, $\hat{Y}_i$, is computed as:

$$\hat{Y}_i = \frac{1}{N} \sum_{j=1}^N \text{Tree}_j(\mathbf{X}) \quad (19)$$

where N is the number of trees, and Tree_j represents each individual tree's prediction.

5.2.4. Active Learning and Uncertainty Analysis

The model's accuracy is improved through an *active learning loop*. The loop begins by generating new input values within specified ranges (e.g., $40 \leq \text{RVE Size} \leq 60$ and $40 \leq Vf \leq 60$).

To assess prediction certainty, the code calculates the *variance* of predictions across the trees in the Random Forest model. For each target property Y_i , the variance σ_i^2 among tree predictions is:

$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^N \left(\text{Tree}_j(\mathbf{X}) - \hat{Y}_i \right)^2 \quad (20)$$

If the average variance across all properties exceeds a defined *variance threshold* (0.01), the model flags the prediction as unreliable and initiates an FE simulation.

5.2.5. Running Simulations for High Uncertainty Predictions

When the variance indicates high uncertainty, the code triggers an FE simulation using the generated input values. Simulation outputs actual property values, which update the dataset and retrain the model, thereby minimising prediction errors. After each simulation, the dataset is expanded with the new result, and the surrogate model is retrained using the updated dataset. This retraining involves:

1. Re-normalisation of the updated data.
2. Re-calculating the ensemble prediction using new data points.
3. Updating the model to account for the newly observed data, improving prediction accuracy for future iterations.

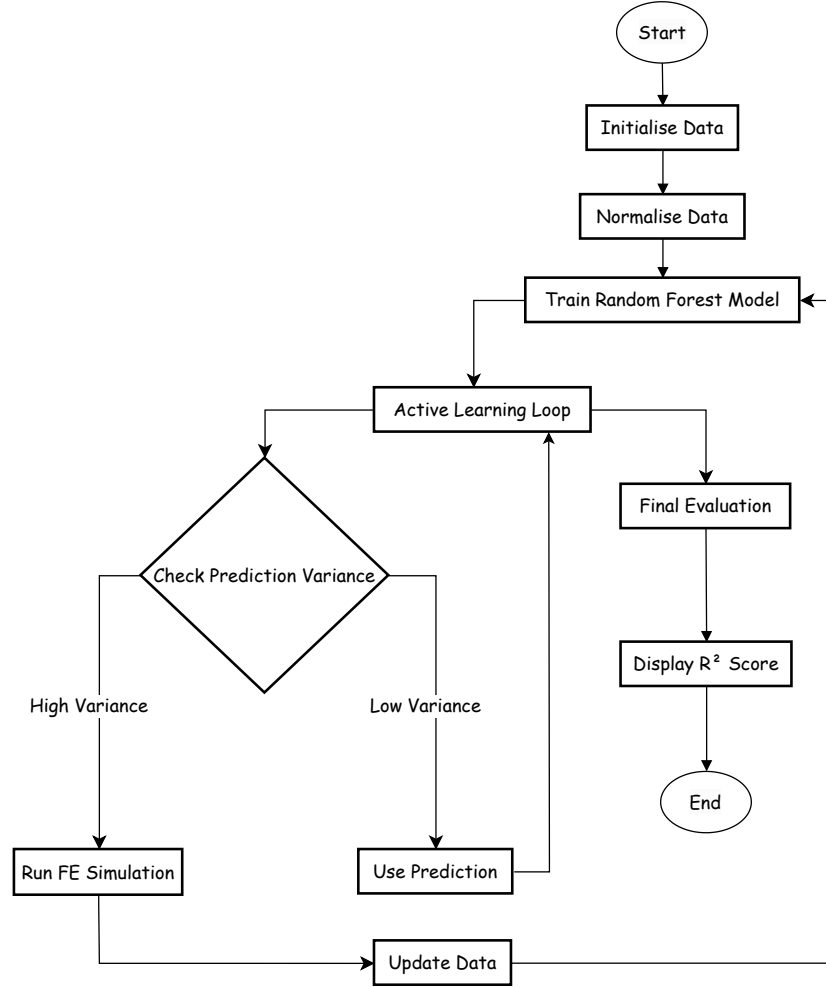


Figure 5: Workflow of the active learning loop with a Random Forest model for predicting effective properties. High variance predictions trigger new simulations, while low variance predictions are directly used to update the data until final evaluation.

6. Results and Discussion

6.1. Rationale for Aggregated Predictions

Theoretical models, such as Halpin-Tsai and Mori-Tanaka, inherently do not account for RVE size in their calculations. Similarly, experimental studies often omit

to mention the RVE size used, making direct comparisons with size-sensitive results impractical. To address this, feature importance analysis (refer to Figure 6) confirmed that RVE size has an insignificant effect on machine learning (ML) predictions. This is confirmed by Kari et al. [45], who found that varying particle size within a constant RVE size does not significantly affect effective material properties. Fiber volume fraction (V_f) was identified as the dominant factor, with a feature importance score of approximately 0.95, while RVE size contributed minimally with a score of around 0.05. These findings align with the assumptions of theoretical models and experimental studies, which implicitly treat V_f as the primary variable of interest.

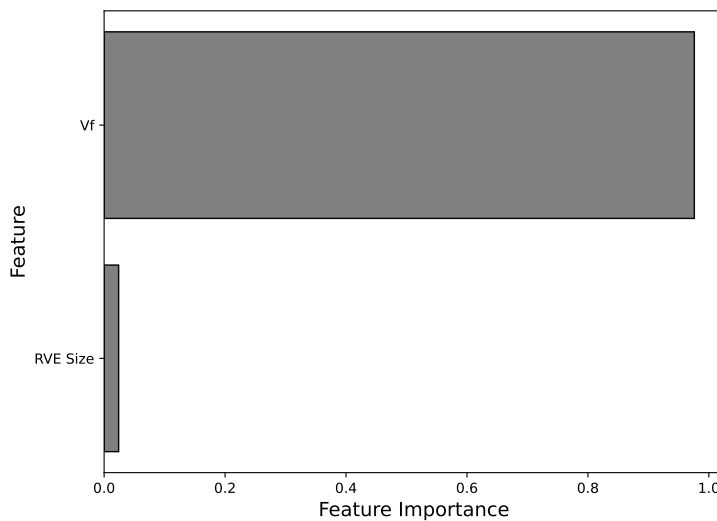


Figure 6: Feature Importance Analysis: V_f dominates the predictions, while RVE size has minimal influence.

Given this, predictions were aggregated based on V_f , grouping all results by fiber volume fraction while disregarding RVE size. This approach simplifies the analysis, eliminates inconsistencies arising from RVE size variability, and ensures valid comparisons across theoretical, experimental, and ML-based predictions.

To illustrate the effectiveness of aggregating predictions based on V_f , Figures 7a and 7b showcase the results for E_{11} and G_{23} , respectively. The blue dots in both figures represent individual machine learning (ML) predictions for each V_f , which exhibit clustering, indicating consistent predictions across different RVE sizes. The

red dots denote the aggregated predictions for each V_f , computed as the mean of the individual predictions. The error bars on the red dots indicate the standard deviation of the individual predictions.

Figure 7a illustrates the predictions for E_{11} , where the aggregated predictions (red dots) align closely with the cluster of individual predictions (blue dots), demonstrating the small variability in ML predictions for each V_f . Similarly, Figure 7b shows the predictions for G_{23} , where the clustering of blue dots around the red dots confirms that V_f is the dominant factor influencing predictions.

By aggregating predictions, variability due to RVE size is effectively eliminated, simplifying the analysis and ensuring consistent and robust comparisons. This method reinforces the alignment of ML predictions with theoretical and experimental assumptions, focusing solely on V_f as the key variable.

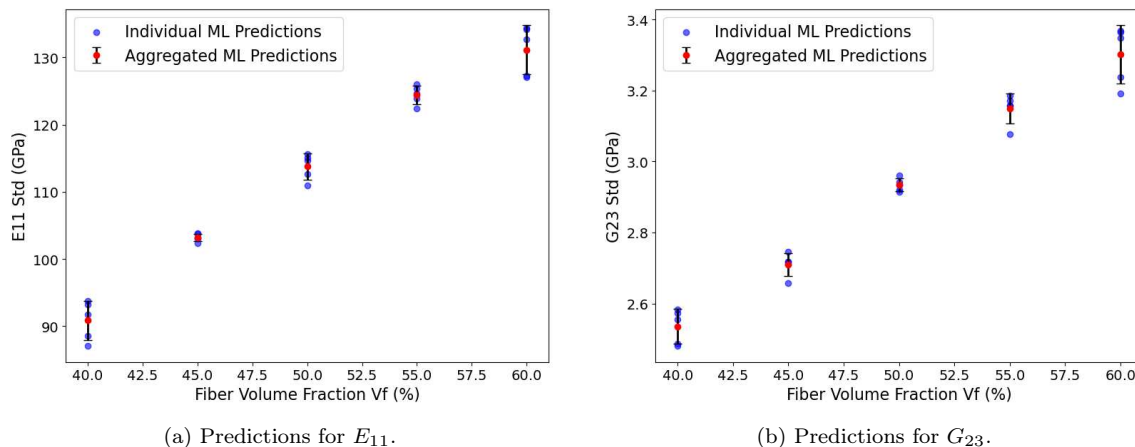


Figure 7: Comparison of individual (blue dots) and aggregated (red dots with error bars) ML predictions for E_{11} and G_{23} . The red dots represent the mean aggregated predictions, while the error bars indicate the standard deviation of individual predictions.

6.2. Validation of ML Predictions

The comparison of ML predictions with theoretical models (Rule of Mixtures (ROM) [2], Nielson [71, 63], Chamis [72], Halpin-Tsai (HT) [7], Mori-Tanaka (MT) [73], Huang, and Saravanos-Chamis [74, 75]) and experimental results across various fiber volume fractions (V_f) provides critical insights into the accuracy of different approaches for predicting the effective elastic properties of composite materials. The analysis encompasses $V_f = 40\%$, 45% , 50% , 55% , and 60% , with an experimental

data available at $V_f = 60\%$ to validate the ML predictions and reliabilities.

For E_{11} , the longitudinal Young's modulus, the ML model demonstrates strong alignment with the ROM theoretical model across lower fiber volume fractions. Specifically, at $V_f = 40\%$, 45% , 50% , and 55% , the relative differences between ML and ROM predictions are 0.91% , 1.00% , 1.26% , and 0.26% , respectively. These minimal deviations indicate that the ML model reliably interpolates the longitudinal stiffness trends within this range of fiber volume fractions.

At a higher fiber volume fraction of $V_f = 60\%$, the relative difference between ML and ROM predictions increases to 4.67% . Despite this increase, the ML model maintains a lower deviation of 3.41% when compared directly to experimental data, whereas the ROM model exhibits a larger deviation of 8.50% . This significant improvement at $V_f = 60\%$ underscores the ML model's enhanced capability to accurately capture the stiffness contributions from the fiber phase at elevated fiber contents.

In the case of E_{22} , the transverse Young's modulus, noticeable discrepancies were observed between machine learning (ML) predictions, theoretical models, and experimental data. At a fibre volume fraction of $V_f = 60\%$, the relative difference between the experimental result and each model was as follows: the Huang model differed by 17.56% , the Chamis model by 12.64% , and the Halpin-Tsai model by 17.09% . In comparison, the ML prediction exhibited a relative difference of 19.59% from the experimental value, performing better than the Nielson model comparably to Halpin-Tsai and Huang models. Across the lower fibre volume fractions ($V_f = 40\%$ to 55%), the Halpin-Tsai model consistently produced values slightly lower than the ML predictions, while the Nielson model significantly underpredicted E_{22} , especially as V_f increased. In contrast, the ML predictions showed a trend that closely matched the simulation results and followed the experimental data more reliably than the majority of the theoretical models. Notably, the ML predictions aligned more closely with Huang and Halpin-Tsai at moderate V_f , and maintained better overall consistency across the range than Nielson.

The comparison of the in-plane shear modulus, G_{12} , obtained from the machine learning (ML) model, theoretical models, and experimental data reveals distinct trends. Figure 1 illustrates the variations of G_{12} . At $V_f = 60\%$, the percentage errors between the experimental data and the theoretical models are notable. Specifically, the errors for the Nielson, Chamis, Halpin-Tsai, Mori-Tanaka, and Huang models are

43.86%, 22.48%, 27.96%, 31.26%, and 39.41%, respectively. The ML model shows an error of 29.83% compared to the experimental data, indicating that its predictions are within a comparable range to those of the theoretical models. This highlights the reliability of the ML model, as it performs similarly to well-established analytical approaches, particularly Halpin-Tsai and Mori-Tanaka.

The ML predictions generally align well with the theoretical models, with lower errors at smaller fibre volume fractions, and the trends observed at $V_f = 60\%$ suggest that the model effectively captures the material behaviour. While further refinement or additional data could improve its performance.

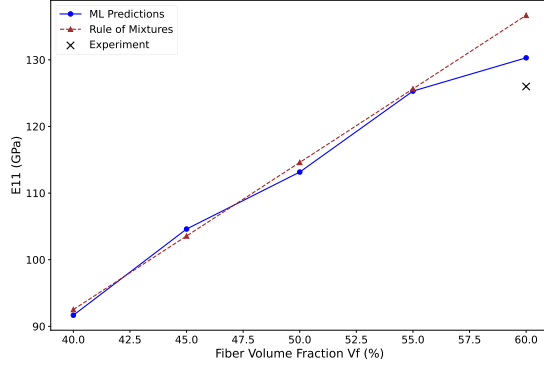
For the out-of-plane shear modulus, G_{23} , the models showcased varying performance. At $V_f = 60\%$, the HT model and MT model had a 15.33% relative difference, the Saravanos-Chamis (SC) model showed a 11.53% difference, the Huang model presented a 12.39% difference, and the Chamis model had a remarkably low 0.28% relative difference. The ML prediction demonstrated a 15.79% relative difference compared to experiments, placing it within the range of theoretical models. It outperformed the SC, Huang, and Chamis models, and was very close to the Halpin-Tsai model and better than the Mori-Tanaka model.

At $V_f = 40\%$, 45%, 50%, and 55%, the ML predictions exhibited relatively small differences when compared to theoretical models such as HT and Huang. For instance, at $V_f = 40\%$, the ML model differed by only 1.78% and 1.27% respectively. These results highlight the ML model's ability to effectively capture the variation trends in G_{23} , especially across intermediate fibre volume fractions.

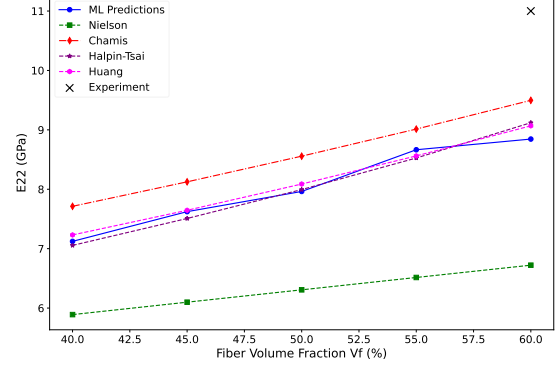
The prediction of Poisson's ratio, ν_{12} , was consistent across all models. At $V_f = 60\%$, the ROM model exhibited a 8.57% relative difference from experiments, while the ML prediction showed a 8.92% deviation. Across other fiber volume fractions ($V_f = 40\%$ to 55%), the ML model showed minimal variation, with deviations from the ROM model ranging from 1.05% to 2.28%. Both ML and theoretical models exhibit minimal variation in predicting ν_{12} , with deviations remaining low across all fiber volume fractions. This consistency underscores the relative insensitivity of Poisson's ratio to fiber content and the effectiveness of both modelling approaches in capturing this property.

The machine learning (ML) model demonstrates considerable potential in accurately predicting the longitudinal and transverse elastic properties of composite materials, often outperforming traditional theoretical models. Nevertheless, certain limitations should be acknowledged. Predictions of shear moduli (G_{12} and G_{23}) exhibited greater variability compared to the longitudinal moduli, underscoring the challenges in capturing complex shear behaviour. Despite these challenges, the integration of active learning methodologies holds promise for enhancing predictive capabilities, thereby positioning ML models as valuable tools for composite material design and analysis.

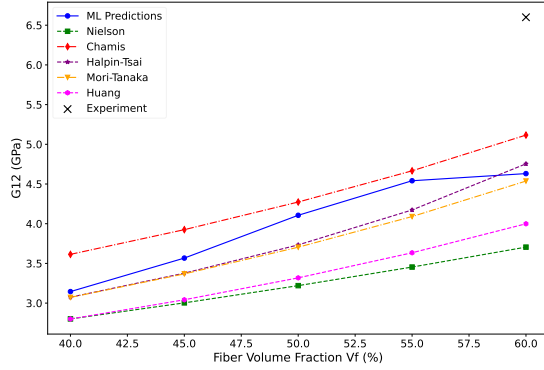
Active learning will be introduced in the next section to further enhance the predictive capabilities of the ML model. This approach will iteratively refine the model by prioritising high-impact data points for additional simulations, thereby improving its accuracy and robustness. Active learning will also ensure the efficient use of computational resources while maximising the generalisability of the ML model for composite material property prediction.



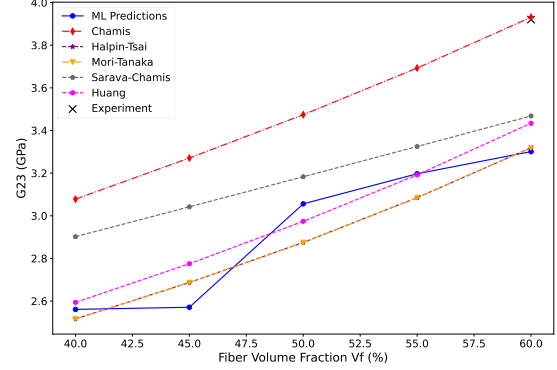
(a) E_{11} vs Fiber Volume Fraction (V_f)



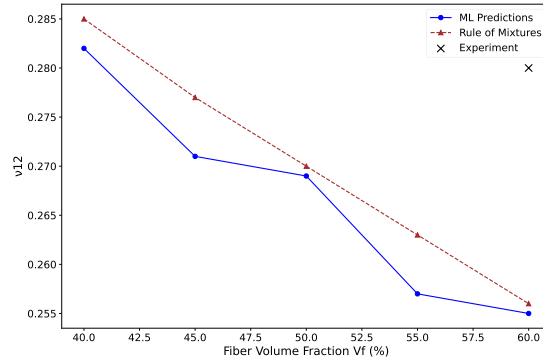
(b) E_{22} vs Fiber Volume Fraction (V_f)



(c) G_{12} vs Fiber Volume Fraction (V_f)



(d) G_{23} vs Fiber Volume Fraction (V_f)



(e) ν_{12} vs Fiber Volume Fraction (V_f)

Figure 8: Comparison of predictions for effective properties using different methods.

6.3. Enhancements through Surrogate Modelling and Active Learning

The integration of active learning into the surrogate model has significantly enhanced its predictive accuracy and consistency for the effective elastic properties of composite materials. The surrogate model efficiently incorporates new, informative data points by triggering additional FEM simulations only when the model’s prediction uncertainty exceeds a set variance threshold (0.01). When triggered, FEM simulations are run for all selected RVE sizes and fibre volume fractions as presented in Table 6, and the aggregated results are added to the dataset. This targeted approach optimises computational resources by avoiding unnecessary simulations and ensures that the model learns from the most critical data, maintaining predictions within acceptable limits. Table 4 and Table 5 present a comprehensive comparison of elastic property predictions at various fibre volume fractions ($V_f = 40\%$ to 60%). The data includes predictions from the machine learning (ML) model enhanced with active learning and results from the simulations. The ML model predicted the effective properties at $V_f = 40\%$ with high confidence, hence no simulation data are presented.

For the longitudinal modulus E_{11} , the ML model shows remarkable accuracy across all fibre volume fractions. At $V_f = 45\%$, the ML prediction (103.20 ± 0.56) is within 0.67% of the simulation (103.89 ± 0.14). Similarly, at $V_f = 60\%$, the ML prediction (131.12 ± 3.65) deviates by only 1.73% from the simulation (128.90 ± 5.54). The transverse modulus E_{22} presents a more challenging prediction due to complex fibre-matrix interactions. Nevertheless, the ML model demonstrates excellent agreement with the simulations. At $V_f = 50\%$, the ML prediction (8.01 ± 0.11) deviates by only 1.05% from the simulation (8.10 ± 0.10). Shear moduli G_{12} and G_{23} also exhibit strong predictive performance with the active learning approach. For G_{12} at $V_f = 55\%$, the ML prediction (4.39 ± 0.08) is within 1.19% of the simulation (4.34 ± 0.16). For G_{23} , the ML predictions closely match the results across all volume fractions, with deviations typically less than 1%. For example, at $V_f = 50\%$, the ML prediction (2.93 ± 0.02) is within 0.58% of the simulation (2.95 ± 0.01). Poisson’s ratio (ν_{12}) maintains high reliability across all volume fractions. The ML predictions consistently align with the simulations, with negligible differences. At $V_f = 45\%$, the ML prediction (0.27 ± 0.00) matches the simulation (0.27 ± 0.00) almost exactly.

After implementing active learning, the performance of the machine learning (ML) model improved across all predicted properties. The ML predictions now exhibit smoother trends and align much better with simulation and theoretical data, closely matching models such as Chamis, Huang, Mori-Tanaka, Halpin-Tsai and the Rule of Mixtures, as well as experimental results (see Figure 9). This enhanced

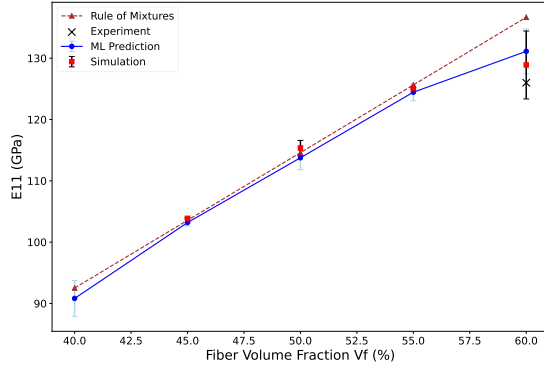
alignment is evident through reduced deviations and more realistic trends across the fibre volume fraction range. The active learning-enhanced surrogate model demonstrates superior performance in predicting the effective elastic properties of composite materials. By focusing on the most informative data and reducing unnecessary simulations, the model achieves high predictive accuracy and consistency with reduced computational resources. This efficient approach minimises the need for large-scale datasets while closely matching simulation results, validating the effectiveness of integrating active learning into surrogate modelling. Overall, these results confirm that the ML framework demonstrates superior accuracy in predicting transverse and shear properties compared to most theoretical models, while remaining fully consistent with FEM simulations and experimental observations.

Table 4: Comparison of ML predictions and ROM for E_{11} and ν_{12} at $V_f = 40\%$ and $V_f = 60\%$.

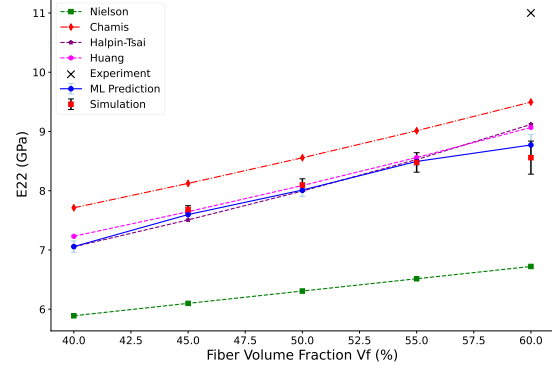
Vf (%)	Property	ML Prediction ($\pm$ Std)	ROM	Difference (ML to ROM)	Simulation ($\pm$ Std)
40.00	E_{11}	90.82 ± 2.93	91.80	1.06%	-
	ν_{12}	0.28 ± 0.00	0.28	1.19%	-
60.00	E_{11}	131.12 ± 3.65	136.20	3.73%	128.90 ± 5.54
	ν_{12}	0.25 ± 0.00	0.26	0.60%	0.26 ± 0.00

Table 5: Comparison of ML predictions, selected methods, and differences (E_{22} , G_{12} , G_{23}) at $V_f = 40\%$ and $V_f = 60\%$.

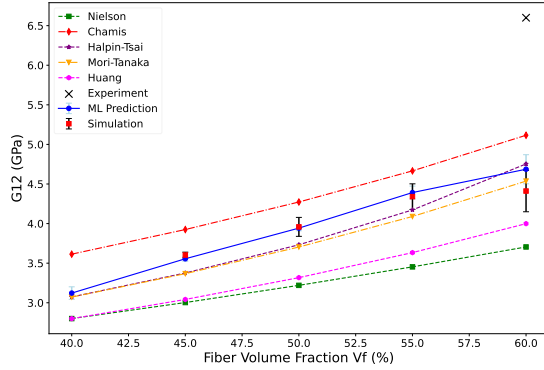
V_f (%)	Property	ML Prediction ($\pm$ Std)	Method	Value	Difference (ML to Method)	Simulation ($\pm$ Std)
40.00	E_{22}	7.06 ± 0.10	Nielson	5.89	19.82%	-
			Chamis	7.71	8.49%	
			HT	7.05	0.04%	
	G_{12}	3.12 ± 0.08	Nielson	2.80	11.49%	
			Chamis	3.61	13.60%	
			HT	3.07	1.66%	
	G_{23}	2.54 ± 0.05	Chamis	3.08	17.62%	
			HT	2.52	0.82%	
			Huang	2.59	2.23%	
60.00	E_{22}	8.77 ± 0.17	Nielson	6.72	30.51%	8.56 ± 0.28
			Chamis	9.50	7.62%	
			HT	9.12	3.81%	
	G_{12}	4.69 ± 0.18	Nielson	3.70	26.50%	4.41 ± 0.26
			Chamis	5.12	8.39%	
			HT	4.75	1.43%	
	G_{23}	3.30 ± 0.08	Chamis	3.93	15.97%	3.29 ± 0.11
			HT	3.32	0.58%	
			Huang	3.43	3.85%	



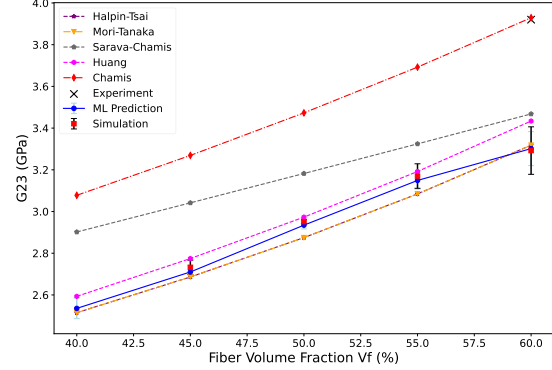
(a) E_{11} vs Fiber Volume Fraction (V_f)



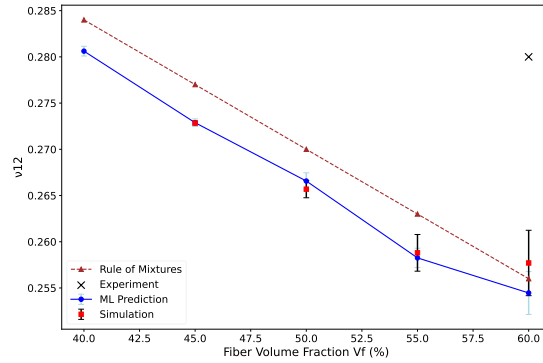
(b) E_{22} vs Fiber Volume Fraction (V_f)



(c) G_{12} vs Fiber Volume Fraction (V_f)



(d) G_{23} vs Fiber Volume Fraction (V_f)



(e) ν_{12} vs Fiber Volume Fraction (V_f)

Figure 9: Comparison of predictions for effective properties using different methods.

Table 6 summarises the behaviour of the active learning loop. At smaller RVE sizes (40–45 μm), the surrogate model flagged high uncertainty at several fibre volume fractions, which triggered new FEM simulations. As the dataset was enriched, the variance dropped below the threshold and subsequent combinations (e.g., 55–60 μm) were resolved entirely through surrogate predictions without further simulations. This illustrates how the framework adaptively concentrates computational effort on critical regions of the design space while avoiding unnecessary FEM runs, thereby achieving convergence with minimal additional cost.

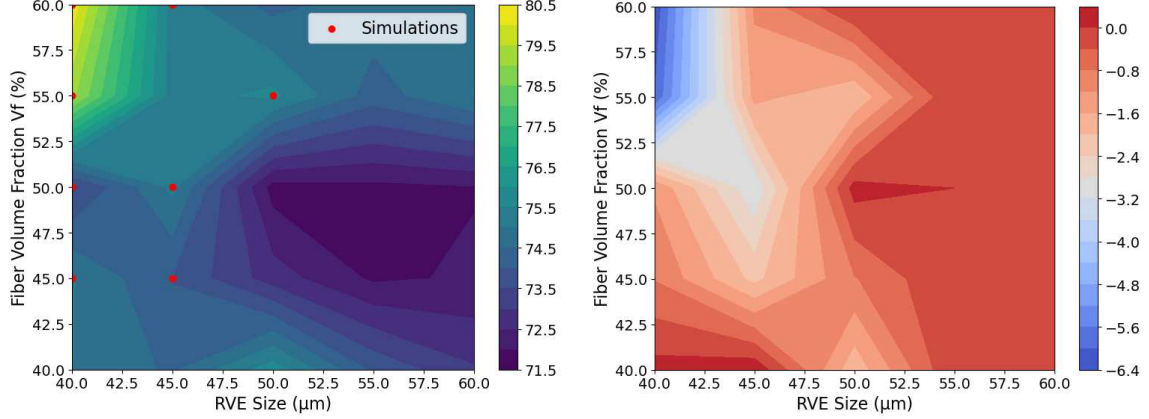
Table 6: Active learning selections by RVE size and fibre volume fraction.

RVE size (μm)	V_f (%)	Status
40	40	Prediction only
40	45, 50, 55, 60	New simulation triggered
45	40	Prediction only
45	45, 50	New simulation triggered
45	55	Prediction only
45	60	New simulation triggered
50	40, 45, 50	Prediction only
50	55	New simulation triggered
50	60	Prediction only
55	40, 45, 50, 55, 60	Prediction only
60	40, 45, 50, 55, 60	Prediction only

6.4. Impact of Active Learning on Prediction Uncertainty

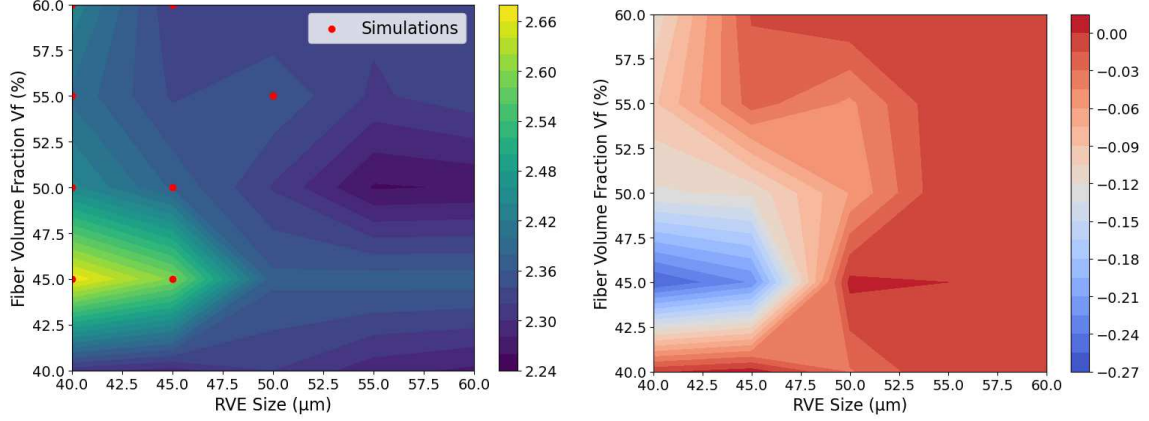
Active learning is essential in reducing prediction uncertainty by systematically identifying regions with high variability and incorporating additional simulation data to enhance model accuracy. This process not only ensures higher confidence in machine learning (ML) predictions but also minimises the standard deviation of predictions for key properties. The analysis focuses on two critical properties, E_{11} and G_{23} , demonstrating the effectiveness of active learning in improving prediction robustness.

The figures illustrate the impact of active learning on reducing prediction uncertainty for two key properties, E_{11} and G_{23} . For E_{11} , high uncertainty regions prior to active learning, as shown in Figure 10a, are observed around RVE Size = 45 μm and $V_f = 50\%$, and RVE Size = 50 μm and $V_f = 55\%$. These regions exhibit significant reductions in standard deviation after active learning, with reductions of



(a) Standard deviation of E_{11} predictions before active learning.

(b) Reduction in E_{11} uncertainty after active learning.



(c) Standard deviation of G_{23} predictions before active learning.

(d) Reduction in G_{23} uncertainty after active learning.

Figure 10: Impact of active learning on the standard deviation of predictions for E_{11} and G_{23} .

up to 6.4 units (Figure 10b). Similarly, for G_{23} , areas of high standard deviation exhibit reductions of approximately 0.27 and 0.22 units, respectively, as illustrated in Figure 10d. These results confirm the effectiveness of active learning in refining prediction accuracy, particularly in regions with high initial uncertainty.

The results demonstrate that active learning efficiently identifies and mitigates high-uncertainty regions by performing targeted simulations. The reductions in standard deviation for E_{11} and G_{23} highlight its ability to enhance prediction reliability

across the RVE size and fibre volume fraction parameter space. These findings underscore the utility of active learning in improving surrogate models, enabling higher prediction confidence while minimising computational costs.

6.5. Generalisability and Extensions

While the present work focuses on elastic properties of unidirectional fibre-reinforced composites, the framework is readily transferable to other contexts provided an RVE–FEM dataset is available. One natural extension is to thermal properties: by replacing elastic simulations with virtual thermal tests, the framework could predict anisotropic and temperature-dependent conductivities, similar to homogenisation studies of woven composites where periodic boundary conditions were reformulated for thermal loading [76]. This is consistent with findings from 2D FE studies of particle-reinforced composites, where predictive power was found to be stronger for thermal conductivity and coefficient of thermal expansion than for elastic properties, particularly at high fibre volume fractions [77]. Such results highlight the suitability of FEM–ML surrogates for multiphysical behaviour, especially when the same modelling assumptions are adopted.

The methodology can also be extended to thermo-mechanical coupling. For example, multiscale FEM and asymptotic homogenisation have been successfully applied to woven SiCf/SiC systems, showing good agreement with experimental data [78]. Embedding such coupled RVE simulations within the present surrogate–active learning framework would allow scalable prediction of thermo-mechanical responses across a broader range of composites.

Beyond linear elasticity, nonlinear and damage responses are equally accessible. Prior homogenisation studies have incorporated matrix elasto-plasticity and fibre–matrix decohesion through advanced yield criteria and cohesive interface elements [79]. ML-based approaches have likewise captured impact behaviour of layered and hybrid composites [21, 22], predicted transverse strength of UD-CFRP from fibre distributions [29], and identified composite damage mechanisms from acoustic emission features [27]. Recurrent and gated neural networks have also been shown to effectively model path-dependent plasticity in multiscale settings [28]. Within such scenarios, active learning is particularly valuable to reduce the cost of generating large nonlinear FEM datasets. Furthermore, the surrogate could be tailored to detect fibre–matrix debonding by incorporating yield points from stress–strain responses into the training data, supported by FEM datasets with cohesive elements at the fibre–matrix interface. These extensions underline the generalisability of the

proposed framework and its potential as a versatile multiphysics and multiscale modelling tool for composite materials.

7. Conclusion

This study demonstrates that integrating periodic boundary conditions (PBCs), RVE-based micromechanics, and machine learning-based surrogate modelling enhances the prediction of effective composite material properties. To the best of the author’s knowledge, this is the first study to employ such an integrated approach. The key findings are:

- ML surrogates provide rapid and accurate predictions of composite properties, reducing the need for expensive FEM simulations.
- Active learning improves accuracy and efficiency by selectively enriching the dataset in regions of high uncertainty.
- PBCs ensure consistent boundary behaviour in RVE simulations, enabling reliable mechanical property predictions.
- The framework shows strong agreement with FEM, theoretical models, and experiments, with particularly high accuracy for E_{11} and E_{22} .
- Limitations remain for shear moduli (G_{12} , G_{23}), where predictions are comparable to analytical models but affected by the sensitivity of shear behaviour to fibre–matrix interactions.

Future work will expand the framework to multi-phase and hybrid composites, adopt advanced ML methods (e.g., deep learning, ensembles), broaden experimental validation, and perform systematic runtime comparisons including FEM with damage modelling. This study establishes a robust foundation for scalable ML-based surrogate modelling in composite material design.

References

- [1] R. Hill, The elastic behaviour of a crystalline aggregate, *Proceedings of the Physical Society. Section A* 65 (5) (1952) 349.
- [2] W. Voigt, *Lehrbuch der kristallphysik:(mit ausschluss der kristalloptik)*, Vol. 34, BG Teubner, 1910.
- [3] A. Reuß, Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle., *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 9 (1) (1929) 49–58.
- [4] R. Hill, Elastic properties of reinforced solids: some theoretical principles, *Journal of the Mechanics and Physics of Solids* 11 (5) (1963) 357–372.
- [5] Z. a. Hashin, S. Shtrikman, On some variational principles in anisotropic and nonhomogeneous elasticity, *Journal of the Mechanics and Physics of Solids* 10 (4) (1962) 335–342.
- [6] Z. Hashin, S. Shtrikman, A variational approach to the theory of the elastic behaviour of multiphase materials, *Journal of the Mechanics and Physics of Solids* 11 (2) (1963) 127–140.
- [7] J. C. Halpin, R. Thomas, Ribbon reinforcement of composites, *Journal of Composite Materials* 2 (4) (1968) 488–497.
- [8] W. A. Smith, B. A. Auld, Modeling 1-3 composite piezoelectrics: thickness-mode oscillations, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 38 (1) (1991) 40–47.
- [9] P. Suquet, Elements of homogenization for inelastic solid mechanics, *Homogenization techniques for composite media* 272 (1987) 193–278.
- [10] A.-S. Kaddour, M. J. Hinton, Maturity of 3d failure criteria for fibre-reinforced composites: Comparison between theories and experiments: Part b of wwfe-ii, *Journal of Composite materials* 47 (6-7) (2013) 925–966.
- [11] N. Balasubramani, B. Zhang, N. Chowdhury, A. Mukkavilli, M. Suter, G. Pearce, Micro-mechanical analysis on random rve size and shape in multiscale finite element modelling of unidirectional frp composites, *Composite Structures* 282 (2022) 115081.

- [12] J. Yin, X. Wang, W. Tang, X. Xu, S. Zhao, F.-z. Xuan, A multiscale model for predicting mechanical properties of polymer composites, *Chemical Engineering Science* 282 (2023) 119352.
- [13] D. L. Buchanan, J. H. Gosse, J. A. Wollschlager, A. Ritchey, R. B. Pipes, Micromechanical enhancement of the macroscopic strain state for advanced composite materials, *Composites Science and Technology* 69 (11-12) (2009) 1974–1978.
- [14] D. Zahl, S. Schmauder, R. McMeeking, Transverse strength of metal matrix composites reinforced with strongly bonded continuous fibers in regular arrangements, *Acta metallurgica et materialia* 42 (9) (1994) 2983–2997.
- [15] D. F. Adams, D. R. Doner, Transverse normal loading of a unidirectional composite, *Journal of composite Materials* 1 (2) (1967) 152–164.
- [16] N. T. Chowdhury, J. Wang, W. K. Chiu, W. Yan, Matrix failure in composite laminates under tensile loading, *Composite Structures* 135 (2016) 61–73.
- [17] D. Trias, J. Costa, J. Mayugo, J. Hurtado, Random models versus periodic models for fibre reinforced composites, *Computational materials science* 38 (2) (2006) 316–324.
- [18] R. Guinovart-Díaz, J. Bravo-Castillero, R. Rodríguez-Ramos, F. J. Sabina, Closed-form expressions for the effective coefficients of fibre-reinforced composite with transversely isotropic constituents. i: Elastic and hexagonal symmetry, *Journal of the Mechanics and Physics of Solids* 49 (7) (2001) 1445–1462.
- [19] A. Ramírez-Torres, A. Roque-Piedra, A. Giammarini, A. Grillo, R. Rodríguez-Ramos, Analytical expressions for the effective coefficients of fibre-reinforced composite materials under the influence of inelastic distortions, *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 105 (4) (2025) e70003.
- [20] K. Berladir, K. Antosz, V. Ivanov, Z. Mital'ová, Machine learning-driven prediction of composite materials properties based on experimental testing data, *Polymers* 17 (5) (2025) 694.
- [21] S. Hiremath, Y. Zhang, L. Zhang, T.-W. Kim, Machine learning approach to evaluating impact behavior in fabric-laminated composite materials, *Results in Engineering* 23 (2024) 102576.

- [22] S. Saha, S. Kumar, V. K. Mahakur, S. Sharma, A. Saha, Development and machine learning based prediction of carbon/hemp/ramie sandwich composite as sustainable and ecofriendly solution for automobile application, *Materials Today Communications* 41 (2024) 110817.
- [23] A. Sharma, T. Mukhopadhyay, S. M. Rangappa, S. Siengchin, V. Kushvaha, Advances in computational intelligence of polymer composite materials: machine learning assisted modeling, analysis and design, *Archives of Computational Methods in Engineering* 29 (5) (2022) 3341–3385.
- [24] A. Bhaduri, A. Gupta, L. Graham-Brady, Stress field prediction in fiber-reinforced composite materials using a deep learning approach, *Composites Part B: Engineering* 238 (2022) 109879.
- [25] Z. Yang, C.-H. Yu, K. Guo, M. J. Buehler, End-to-end deep learning method to predict complete strain and stress tensors for complex hierarchical composite microstructures, *Journal of the Mechanics and Physics of Solids* 154 (2021) 104506.
- [26] K. Singh, J. Adhikari, J. Roscow, Prediction of the electromechanical properties of a piezoelectric composite material through the artificial neural network, *Materials Today Communications* 38 (2024) 108288.
- [27] C. Muir, B. Swaminathan, A. Almansour, K. Sevener, C. Smith, M. Presby, J. Kiser, T. Pollock, S. Daly, Damage mechanism identification in composites via machine learning and acoustic emission, *npj Computational Materials* 7 (1) (2021) 95.
- [28] Y. Zhou, S. J. Semnani, A machine learning based multi-scale finite element framework for nonlinear composite materials, *Engineering with Computers* (2025) 1–37.
- [29] Y. Li, H. Qin, L. Jia, T.-E. Tay, V. B. C. Tan, Y. Liu, Microstructure dependent transverse strength criterion for ud-cfrp composites via computational micromechanics and machine learning, *Composites Science and Technology* 251 (2024) 110551.
- [30] H. Kapadia, L. Feng, P. Benner, Active-learning-driven surrogate modeling for efficient simulation of parametric nonlinear systems, *Computer Methods in Applied Mechanics and Engineering* 419 (2024) 116657.

- [31] L. Hong, H. Li, J. Fu, A novel surrogate-model based active learning method for structural reliability analysis, *Computer Methods in Applied Mechanics and Engineering* 394 (2022) 114835.
- [32] R. Castro-Amoedo, J. Granacher, F. Maréchal, A combined genetic algorithm and active learning approach to build and test surrogate models in process systems engineering, *Computers & Chemical Engineering* 181 (2024) 108517.
- [33] M. D. Rintoul, S. Torquato, Reconstruction of the structure of dispersions, *Journal of colloid and interface science* 186 (2) (1997) 467–476.
- [34] D. Duschlbauer, H. J. Böhm, H. E. Pettermann, Computational simulation of composites reinforced by planar random fibers: homogenization and localization by unit cell and mean field approaches, *Journal of Composite Materials* 40 (24) (2006) 2217–2234.
- [35] H. You, Y. Kim, G. J. Yun, Computationally fast morphological descriptor-based microstructure reconstruction algorithms for particulate composites, *Composites science and technology* 182 (2019) 107746.
- [36] Y. Kitagawa, A. Yoshimura, M. Arai, K. Goto, N. Sugiura, Experimental and numerical evaluation of effects of kidney-shape carbon fiber on transverse cracking of carbon fiber reinforced plastics, *Composites Part A: Applied Science and Manufacturing* 152 (2022) 106690.
- [37] T. Vaughan, C. McCarthy, A combined experimental–numerical approach for generating statistically equivalent fibre distributions for high strength laminated composite materials, *Composites Science and Technology* 70 (2) (2010) 291–297.
- [38] L. Wu, C. N. Chung, Z. Major, L. Adam, L. Noels, From sem images to elastic responses: A stochastic multiscale analysis of ud fiber reinforced composites, *Composite Structures* 189 (2018) 206–227.
- [39] K. Terada, M. Hori, T. Kyoya, N. Kikuchi, Simulation of the multi-scale convergence in computational homogenization approaches, *International Journal of Solids and Structures* 37 (16) (2000) 2285–2311.
- [40] D. Mortell, D. Tanner, C. McCarthy, A virtual experimental approach to microscale composites testing, *Composite Structures* 171 (2017) 1–9.

- [41] D. Pulungan, G. Lubineau, A. Yudhanto, R. Yaldiz, W. Schijve, Identifying design parameters controlling damage behaviors of continuous fiber-reinforced thermoplastic composites using micromechanics as a virtual testing tool, *International Journal of Solids and Structures* 117 (2017) 177–190.
- [42] S. Swaminathan, S. Ghosh, N. Pagano, Statistically equivalent representative volume elements for unidirectional composite microstructures: Part i-without damage, *Journal of Composite materials* 40 (7) (2006) 583–604.
- [43] J. Segurado, J. Llorca, A numerical approximation to the elastic properties of sphere-reinforced composites, *Journal of the Mechanics and Physics of Solids* 50 (10) (2002) 2107–2121.
- [44] J. Ohser, F. Mücklich, *Statistical analysis of microstructures in materials science*, John Wiley & Sons, 2000.
- [45] S. Kari, H. Berger, R. Rodriguez-Ramos, U. Gabbert, Computational evaluation of effective material properties of composites reinforced by randomly distributed spherical particles, *Composite structures* 77 (2) (2007) 223–231.
- [46] D. Trias, J. Costa, A. Turon, J. Hurtado, Determination of the critical size of a statistical representative volume element (srve) for carbon reinforced polymers, *Acta materialia* 54 (13) (2006) 3471–3484.
- [47] S. Yang, A. Tewari, A. M. Gokhale, Modeling of non-uniform spatial arrangement of fibers in a ceramic matrix composite, *Acta Materialia* 45 (7) (1997) 3059–3069.
- [48] V. Buryachenko, N. Pagano, R. Kim, J. Spowart, Quantitative description and numerical simulation of random microstructures of composites and their effective elastic moduli, *International journal of solids and structures* 40 (1) (2003) 47–72.
- [49] G. Barker, Computer simulations of granular materials, in: *Granular matter: An interdisciplinary approach*, Springer, 1994, pp. 35–83.
- [50] A. Wongsto, S. Li, Micromechanical fe analysis of ud fibre-reinforced composites with fibres distributed at random over the transverse cross-section, *Composites Part A: Applied Science and Manufacturing* 36 (9) (2005) 1246–1266.
- [51] T. Zhang, Y. Yan, A comparison between random model and periodic model for fiber-reinforced composites based on a new method for generating fiber distributions, *Polymer composites* 38 (1) (2017) 77–86.

- [52] Z. Wang, X. Wang, J. Zhang, W. Liang, L. Zhou, Automatic generation of random distribution of fibers in long-fiber-reinforced composites and mesomechanical simulation, *Materials & Design* 32 (2) (2011) 885–891.
- [53] A. Melro, P. Camanho, S. Pinho, Generation of random distribution of fibres in long-fibre reinforced composites, *Composites Science and Technology* 68 (9) (2008) 2092–2102.
- [54] V. Romanov, S. V. Lomov, Y. Swolfs, S. Orlova, L. Gorbatikh, I. Verpoest, Statistical analysis of real and simulated fibre arrangements in unidirectional composites, *Composites Science and Technology* 87 (2013) 126–134.
- [55] W. Wang, Y. Dai, C. Zhang, X. Gao, M. Zhao, Micromechanical modeling of fiber-reinforced composites with statistically equivalent random fiber distribution, *Materials* 9 (8) (2016) 624.
- [56] L. Yang, Y. Yan, Z. Ran, Y. Liu, A new method for generating random fibre distributions for fibre reinforced composites, *Composites Science and Technology* 76 (2013) 14–20.
- [57] W. Wang, H. Wang, S. Fei, H. Wang, H. Dong, Y. Ke, Generation of random fiber distributions in fiber reinforced composites based on delaunay triangulation, *Materials & Design* 206 (2021) 109812.
- [58] S.-M. Park, J. H. Lim, M. R. Seong, D. Sohn, Efficient generator of random fiber distribution with diverse volume fractions by random fiber removal, *Composites Part B: Engineering* 167 (2019) 302–316.
- [59] Y. Ismail, Y. Sheng, D. Yang, J. Ye, Discrete element modelling of unidirectional fibre-reinforced polymers under transverse tension, *Composites Part B: Engineering* 73 (2015) 118–125.
- [60] K. C. Liu, A. Ghoshal, Validity of random microstructures simulation in fiber-reinforced composite materials, *Composites Part B: Engineering* 57 (2014) 56–70.
- [61] E. Ghossein, M. Lévesque, A fully automated numerical tool for a comprehensive validation of homogenization models and its application to spherical particles reinforced composites, *International Journal of Solids and Structures* 49 (11-12) (2012) 1387–1398.

- [62] H. M. S. Hayder, H. Afrasiab, M. Gholami, Efficient generation of random fiber distribution by combining random sequential expansion and particle swarm optimization algorithms, *Composites Part A: Applied Science and Manufacturing* 173 (2023) 107649.
- [63] C. Cai, B. Wang, W. Yin, Z. Xu, R. Wang, X. He, A new algorithm to generate non-uniformly dispersed representative volume elements of composite materials with high volume fractions, *Materials & Design* 219 (2022) 110750.
- [64] J. Feder, Random sequential adsorption, *Journal of Theoretical Biology* 87 (2) (1980) 237–254.
- [65] W. Tian, X. Chao, M. Fu, L. Qi, An algorithm for generation of rves of composites with high particle volume fractions, *Composites Science and Technology* 207 (2021) 108714.
- [66] P. D. Soden, M. J. Hinton, A. Kaddour, Lamina properties, lay-up configurations and loading conditions for a range of fibre reinforced composite laminates, in: *Failure criteria in fibre-reinforced-polymer composites*, Elsevier, 2004, pp. 30–51.
- [67] M. Herráez, C. González, C. Lopes, R. G. De Villoria, J. LLorca, T. Varela, J. Sánchez, Computational micromechanics evaluation of the effect of fibre shape on the transverse strength of unidirectional composites: an approach to virtual materials design, *Composites Part A: Applied Science and Manufacturing* 91 (2016) 484–492.
- [68] P. F. Smith, S. Ganesh, P. Liu, A comparison of random forest regression and multiple linear regression for prediction in neuroscience, *Journal of neuroscience methods* 220 (1) (2013) 85–91.
- [69] P. Prettenhofer, G. Louppe, Gradient boosted regression trees in scikit-learn, in: *PyData 2014*, 2014.
- [70] M. Awad, R. Khanna, M. Awad, R. Khanna, Support vector regression, *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (2015) 67–80.
- [71] B. Raju, S. Hiremath, D. R. Mahapatra, A review of micromechanics based models for effective elastic properties of reinforced polymer matrix composites, *Composite Structures* 204 (2018) 607–619.

- [72] C. C. Chamis, Mechanics of composite materials: past, present, and future, *Composites Technology and Research* 11 (1) (1989) 3–14.
- [73] T. Mori, K. Tanaka, Average stress in matrix and average elastic energy of materials with misfitting inclusions, *Acta metallurgica* 21 (5) (1973) 571–574.
- [74] D. Saravanos, C. C. Chamis, Mechanics of damping for fiber composite laminates including hygrothermal effects, *AIAA journal* 28 (10) (1990) 1813–1819.
- [75] J.-H. Tai, A. Kaw, Transverse shear modulus of unidirectional composites with voids estimated by the multiple-cells model, *Composites Part A: Applied Science and Manufacturing* 105 (2018) 310–320.
- [76] H.-g. Kim, W. Ji, Thermal conductivity of a thick 3d textile composite using an rve model with specialized thermal periodic boundary conditions, *Functional Composites and Structures* 3 (1) (2021) 015002.
- [77] K. Makarian, S. Santhanam, Utility of 2d finite element simulations for predicting effective thermomechanical properties of particle-reinforced composites, in: *ASME International Mechanical Engineering Congress and Exposition*, Vol. 52149, American Society of Mechanical Engineers, 2018, p. V009T12A047.
- [78] X. Shen, S. Zhang, X. Liu, L. Gong, S. Dong, Prediction of the thermo-mechanical properties of the sicf/sic rve model via fem and asymptotic homogenization method: Process and implementation details, *Archives of Computational Methods in Engineering* 28 (4) (2021) 3067–3085.
- [79] Z. Ullah, L. Kaczmarczyk, C. J. Pearce, Three-dimensional nonlinear micro/meso-mechanical response of the fibre-reinforced polymer composites, *Composite Structures* 161 (2017) 204–214.