



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Artificial Intelligence-Assisted Prediction and Multi-Objective Optimization of Mechanical Properties and Failure Performance of Fiber-Reinforced Composite Materials

¹Madugula Udaya Kumar, ²Sahabudeen Jawahar, ³Soumita bid, ⁴Jayatheertha H S, ⁵Dr. P. Senthilkumar, ⁶R. Suresh

¹Associate professor, Department of Mechanical Engineering, Methodist College of Engineering and Technology, Hyderabad, Telangana, India, E-mail: mudayakumaar@methodist.edu.in

²Site Qa/Qc Manager, Nesma & Partners Contracting Company, P.O. Box 1498, Al Khobar 31952, Saudi Arabia, E-mail: Jawaharndt@gmail.Com

³Assistant Professor, Civil Engineering Department, C.K Pithawala College of Engineering and Technology, E-mail: Surat.soumitaa.bid@ckpcet.ac.in

⁴Assistant Professor, Department of Civil Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Indai, E-mail: jayatheerthah@dsatmm.edu.in

⁵Associate Professor, Department of Mechanical Engineering, Meenakshi College of Engineering, KK Nagar, Chennai, Tamil Nadu, India, E-mail: kallakurichisenthilkumar@gmail.com

⁶Professor, Mechanical Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai -600062, Tamil Nadu, India, E-mail: sureshr77@gmail.com

Abstract

Fiber-reinforced composite materials are increasingly used in aerospace, automotive, marine, renewable-energy, transportation, and structural engineering applications because of their high specific strength, stiffness, lightweight characteristics, and design flexibility. However, the mechanical response and failure behavior of these materials are governed by complex interactions among fiber type, fiber volume fraction, matrix properties, fiber orientation, laminate architecture, manufacturing conditions, and loading characteristics. Conventional experimental approaches for identifying optimum composite configurations are often time-consuming, resource-intensive, and unable to efficiently capture highly nonlinear material-structure relationships. Artificial intelligence (AI) and machine learning (ML) provide an alternative data-driven approach for predicting mechanical properties, identifying failure mechanisms, and optimizing composite configurations.

This study proposes an AI-assisted framework for the prediction and multi-objective optimization of mechanical properties and failure performance of fiber-reinforced composite materials. The proposed framework integrates experimental or validated literature-derived composite datasets with multiple ML algorithms, including multiple linear regression, decision tree, random forest, support vector regression, gradient boosting, XGBoost, and artificial neural networks. Input parameters include fiber type, fiber volume fraction, fiber orientation, laminate configuration, matrix characteristics, and selected manufacturing and loading parameters, while tensile strength, flexural strength, impact performance, and failure mode are considered as primary response variables. Data preprocessing, feature engineering, cross-validation, hyperparameter optimization, and independent testing are incorporated to improve model reliability and minimize overfitting.

In addition to regression-based prediction, classification models are developed to identify dominant failure mechanisms such as fiber breakage, matrix cracking, interfacial debonding, fiber pull-out, and delamination. Explainable artificial intelligence techniques are incorporated to quantify the contribution of individual material and structural parameters to predicted performance. The resulting predictive models are subsequently integrated with a multi-objective optimization procedure to identify composite configurations that simultaneously improve mechanical strength and resistance to failure while considering practical design constraints.

The proposed framework establishes a systematic connection between composite material design, mechanical-property prediction, failure classification, explainable AI, and optimization. The study is intended to reduce dependence on extensive trial-and-error experimentation and provide a rapid computational approach for materials selection and laminate design. The framework can also support future development of physics-informed and hybrid AI models for reliable composite design. The research therefore demonstrates the potential of AI-assisted methodologies to accelerate the development of high-performance fiber-reinforced composite materials while improving the interpretability and engineering applicability of data-driven material optimization.

Keywords: Fiber-reinforced composites; Artificial intelligence; Machine learning; Mechanical properties; Failure prediction; Tensile strength; Flexural strength; Impact strength; Failure mode classification; Explainable AI; Multi-objective optimization; Laminate design; Composite material optimization; Predictive modeling; Materials informatics.

1. Introduction

Fiber-reinforced composite materials have become an important class of advanced engineering materials because they provide a combination of high specific strength, high specific stiffness, corrosion resistance, fatigue resistance, and flexibility in structural design. Unlike conventional homogeneous materials, composite

materials can be tailored by modifying the type and quantity of reinforcement, matrix composition, fiber orientation, stacking sequence, laminate architecture, and manufacturing conditions. These characteristics have resulted in their widespread application in aerospace structures, automotive components, wind-turbine blades, marine structures, sporting equipment, transportation systems, and civil infrastructure.

The mechanical performance of a fiber-reinforced composite is strongly dependent on interactions among several material and structural parameters. Fiber properties determine the load-carrying capability, while the matrix controls stress transfer, environmental resistance, and the initiation and propagation of matrix-related damage. The interface between fiber and matrix additionally plays a critical role in transferring load and controlling interfacial debonding. Similarly, fiber orientation and stacking sequence determine the directional mechanical response of laminated composites. Consequently, two composites manufactured from the same constituent materials can exhibit substantially different mechanical behavior when their fiber architecture or manufacturing parameters are modified.

Failure in fiber-reinforced composites is also considerably more complex than failure in many conventional engineering materials. Damage can develop through matrix cracking, fiber breakage, fiber-matrix debonding, fiber pull-out, delamination, local buckling, interlaminar failure, and combinations of these mechanisms. Failure may initiate at microscopic length scales and subsequently propagate through multiple interacting damage mechanisms until macroscopic structural failure occurs. Accurate prediction of these mechanisms is therefore important for safe and economical composite design.

Traditional composite design relies heavily on experimental testing, analytical formulations, numerical simulation, and iterative optimization. Although these approaches provide valuable physical insight, extensive experimental campaigns can require substantial quantities of material, laboratory time, and financial resources. Numerical approaches such as finite-element analysis can provide detailed information about stress and damage evolution but may also involve considerable computational cost when numerous combinations of material and structural parameters must be evaluated. These limitations become particularly significant when the design problem involves several interacting variables.

Artificial intelligence (AI), particularly machine learning (ML), has recently emerged as a promising tool for addressing complex materials-engineering problems. Machine-learning models can learn nonlinear relationships between input variables and material responses from experimental or computational datasets. Once adequately trained and validated, such models can provide rapid predictions without repeating the complete experimental or numerical procedure for every new design configuration.

Recent research demonstrates increasing application of ML to fiber-reinforced polymer composites. A 2024 review reported applications covering mechanical-property prediction, damage-parameter prediction, damage initiation, stiffness degradation, progressive damage, crack-path detection, and composite design optimization. The review also emphasized that appropriate selection of the ML algorithm and network architecture is important because model performance depends strongly on the characteristics of the underlying dataset.

More recent research has extended this approach toward simultaneous prediction of strength and failure behavior. A 2026 study investigated machine-learning prediction of tensile strength and classification of failure types in hybrid carbon-, Kevlar-, and glass-fiber epoxy composites, demonstrating the potential of combining regression and classification approaches for composite design. Another recent study investigated optimized boosting models for tensile-strength prediction and failure classification of mono-, bi-, and tri-fiber reinforced composites, indicating the continuing development of AI-based approaches for composite performance assessment.

The increasing use of ML does not, however, eliminate important research challenges. Composite datasets are frequently limited in size, may contain heterogeneous experimental conditions, and may not cover the complete design space. Models trained on narrow datasets may therefore provide excellent apparent accuracy while exhibiting limited generalization outside the training domain. Recent reviews have specifically identified data quality, model generalization, physical consistency, and the tendency toward single-fiber or simulated datasets as important challenges.

Another important limitation is interpretability. Conventional ML models may provide accurate predictions without clearly explaining why a particular composite configuration is predicted to exhibit superior performance. For engineering applications, prediction accuracy alone is insufficient. Designers need to understand which variables exert the strongest influence on strength, stiffness, damage initiation, and failure. Explainable artificial intelligence (XAI), including feature-importance methods and SHAP-based analysis, can provide additional insight into the relationship between material parameters and predicted responses.

The transition from prediction to optimization represents a further opportunity. A prediction model can estimate the performance of a particular composite configuration, but an optimization framework can search through numerous combinations of material and structural parameters to identify configurations satisfying specified performance objectives. This prediction-optimization strategy has already been identified as an important direction in ML-based composite research. Recent work on data-driven optimization of composite

manufacturing further highlights the potential of ML for reducing trial-and-error development and improving process design.

Therefore, the present research proposes an integrated AI-assisted framework that combines mechanical-property prediction, failure-mode classification, explainable AI, and multi-objective optimization. Rather than evaluating ML solely on its ability to reproduce measured strength values, the proposed methodology attempts to establish a complete data-driven pathway from composite material parameters to mechanical performance and failure behavior and finally to optimized material design.

The primary objective of this study is to develop and validate machine-learning models capable of predicting selected mechanical properties of fiber-reinforced composites using material, geometric, architectural, manufacturing, and loading parameters. A second objective is to develop classification models for identifying major failure modes. A third objective is to investigate the influence of individual input variables using explainable AI techniques. Finally, a multi-objective optimization framework is proposed to identify composite configurations that provide improved mechanical performance while minimizing undesirable failure characteristics.

2. Critical Literature Review

2.1 Artificial Intelligence in Fiber-Reinforced Composite Materials

Fiber-reinforced composite materials possess a complex hierarchical structure in which fiber properties, matrix characteristics, fiber-matrix interfaces, fiber orientation, stacking sequence, manufacturing parameters, and loading conditions collectively determine the final mechanical response. This complexity makes accurate analytical prediction challenging, particularly when several design parameters vary simultaneously. Artificial intelligence (AI) and machine learning (ML) have consequently emerged as important computational tools for developing data-driven relationships between composite design variables and engineering performance.

Machine learning differs from conventional analytical modeling because it can identify nonlinear relationships directly from available datasets without requiring an explicitly defined mathematical relationship between all input and output variables. For composite materials, this capability is particularly useful because mechanical properties frequently depend on nonlinear interactions among material constituents and manufacturing variables.

A comprehensive review published in 2024 examined the application of ML to the prediction and optimization of laminated fiber-reinforced polymer composites. The review reported the increasing use of algorithms such as artificial neural networks, support vector machines, decision trees, random forests, and other learning techniques for mechanical-property prediction and design optimization. It also emphasized that model selection and neural-network architecture must be matched to the characteristics of the specific problem and dataset [1].

The growing body of research demonstrates that ML can be used not only to predict individual mechanical properties but also to investigate relationships between composite architecture and structural response. Input parameters may include fiber type, fiber volume fraction, matrix type, fiber orientation, laminate thickness, stacking sequence, curing parameters, and loading conditions. Outputs can include tensile strength, tensile modulus, flexural strength, flexural modulus, impact resistance, fracture toughness, fatigue life, wear resistance, and other performance indicators.

The application of ML is particularly attractive because once a model has been trained and validated, predictions can be obtained at substantially lower computational cost than repeating a complete experimental program. This creates an opportunity to use ML as a surrogate model within optimization procedures, allowing large numbers of potential material configurations to be screened computationally.

2.2 Machine Learning for Mechanical-Property Prediction

Mechanical-property prediction is currently one of the most developed applications of ML in composite-material research. The fundamental objective is to establish a mapping:

$$[X=\{x_1, x_2, \dots, x_n\} \rightarrow Y=\{y_1, y_2, \dots, y_m\}]$$

where (X) represents material, manufacturing, structural, and environmental parameters and (Y) represents one or more mechanical responses. For example, fiber type, fiber content, orientation angle, laminate thickness, and matrix characteristics can be used as inputs for predicting tensile strength. A trained regression model can then estimate the expected response for a new composite configuration. Linear regression provides a useful baseline because it allows the effect of input parameters to be represented through a relatively simple mathematical relationship. However, the response of composite materials is generally nonlinear, making nonlinear ML methods potentially more suitable for complex datasets. Decision-tree-based methods have attracted considerable attention because they can capture nonlinear relationships and interactions without requiring a predefined functional form. Random Forest extends the decision-tree concept by combining multiple trees and aggregating their predictions. Gradient-boosting algorithms similarly construct an ensemble of weak learners sequentially, with each new learner attempting to reduce the errors of previous learners.

Support Vector Regression provides another useful approach, particularly when nonlinear relationships can be represented through appropriate kernel functions. Artificial Neural Networks are capable of approximating highly nonlinear relationships through interconnected layers of computational nodes and can therefore be useful when sufficient training data are available. A recent ten-year review of ML applications to composite mechanical properties analyzed 165 studies published between 2015 and 2025. The review identified the increasing role of ML in prediction and design optimization and emphasized the challenges created by nonlinear behavior, multiscale material structure, and manufacturing sensitivity [2]. The rapid development of this field demonstrates that simply comparing several ML algorithms is no longer sufficient to establish strong novelty. A modern research study should instead demonstrate how prediction models contribute to a broader engineering decision-making framework.

2.3 Prediction of Failure Behavior

Mechanical-property prediction alone does not completely characterize the performance of a fiber-reinforced composite. Failure behavior is equally important because composite materials can experience multiple interacting damage mechanisms. Typical failure mechanisms include: matrix cracking; fiber breakage; fiber-matrix debonding; fiber pull-out; delamination; interlaminar failure; local buckling; shear failure; and combined or mixed-mode failure.

Failure initiation may occur at microscopic defects or interfaces and subsequently develop into macroscopic damage. Consequently, the relationship between a composite's constituent parameters and its ultimate failure mode can be highly nonlinear. Machine learning provides an opportunity to treat failure-mode identification as a classification problem. In this approach, the failure mechanism is represented as a categorical response rather than a continuous numerical variable.

For example:

$[Y_f \in \{\text{fiberbreakage}, \text{matrixcracking}, \text{delamination}, \text{debonding}, \text{fiberpull-out}, \text{mixedfailure}\}]$

Classification algorithms can then be trained to associate the failure category with composite design parameters.

This approach is particularly important because a composite with high tensile strength is not necessarily the most desirable configuration if its failure mechanism is sudden or difficult to detect. Therefore, an engineering optimization framework should consider both strength and failure characteristics.

2.4 Recent AI-Based Studies on Failure Prediction

Recent research has begun to combine mechanical-property prediction with failure classification. A 2026 study investigated tensile-strength prediction and failure-mode classification in carbon-, Kevlar-, and glass-fiber epoxy composites using several ML models. The study demonstrated the feasibility of simultaneously predicting tensile strength and failure type and identified laminate thickness and lay-up configuration as important variables [3].

A second 2026 study developed optimized boosting approaches for mono-, bi-, and tri-fiber reinforced composites. The study compared AdaBoost, Gradient Boosting, XGBoost, CatBoost, and LightGBM for tensile-strength prediction and failure classification and incorporated SHAP analysis to interpret the contribution of design variables. The reported results demonstrated the usefulness of boosting algorithms for complex hybrid composite datasets [4].

These studies establish an important recent research direction. However, their existence also means that a new manuscript should avoid presenting tensile-strength prediction and failure classification alone as the principal novelty.

The present research therefore extends the concept by integrating:

$[\text{Prediction}] \rightarrow [\text{FailureClassification}] \rightarrow [\text{Explainability}] \rightarrow [\text{Multi-objectiveOptimization}] \rightarrow [\text{Validation}]$

rather than treating each task as an independent ML exercise.

2.5 Explainable Artificial Intelligence

One major limitation of conventional ML is its black-box nature. A model may provide accurate predictions but offer limited physical interpretation regarding why a particular prediction was generated.

For engineering materials, interpretability is particularly important. Researchers need to determine whether model predictions are physically meaningful and which variables have the strongest influence on performance. Explainable artificial intelligence (XAI) addresses this issue by providing information about the relationship between input variables and model outputs.

Feature-importance methods can identify variables that contribute most strongly to prediction. More advanced methods, such as SHapley Additive exPlanations (SHAP), quantify the contribution of individual features to a specific prediction.

For a composite-design problem, SHAP analysis can potentially answer questions such as:

- How strongly does fiber volume fraction affect tensile strength?
- Does fiber orientation increase or decrease predicted strength?

- How important is laminate thickness?
- Which variable most strongly influences failure mode?
- Do manufacturing parameters have greater influence than constituent material properties?
- Are interactions between fiber type and laminate architecture important?

Recent ML research on hybrid fiber composites has already demonstrated the application of SHAP to identify influential variables for strength and failure prediction [4]. Consequently, explainability should be incorporated into the proposed framework not simply as an additional visualization but as a mechanism for interpreting the engineering significance of the predictive model.

2.6 AI-Assisted Optimization of Composite Materials

Prediction and optimization are related but distinct problems.

A prediction model estimates:

$$[Y=f(X)]$$

whereas optimization attempts to determine:

$$[X^*=\arg\max/\min f(X)]$$

subject to design constraints.

For fiber-reinforced composites, the optimization variables may include: fiber type; fiber volume fraction; fiber orientation; number of plies; laminate stacking sequence; laminate thickness; matrix type; curing parameters; and manufacturing parameters.

The objective may be to maximize tensile strength, flexural strength, impact resistance, or fracture toughness while minimizing weight, material cost, or failure probability.

Recent reviews indicate that ML can significantly reduce the computational cost associated with repeated prediction and optimization of composite designs [1,2]. A 2025 review of AI in laminated FRP composites similarly emphasized the growing role of AI for design and optimization of anisotropic composite systems [5]. However, optimization studies remain less mature than prediction studies. A 2025 review of AI-driven FRP material simulation identified limited research addressing AI-driven optimization and standardized datasets, indicating an opportunity for more integrated and reproducible optimization frameworks [6].

2.7 Multi-Objective Optimization

Real-world composite design rarely has a single objective.

For example, increasing fiber content may improve tensile strength but increase material cost or weight. Changing fiber orientation may improve strength in one loading direction while reducing performance in another direction.

Therefore, the present research proposes multi-objective optimization rather than single-objective optimization.

A general multi-objective formulation can be expressed as:

$$[\max F_1(X)=\text{tensileStrength}]$$

$$[\max F_2(X)=\text{flexuralStrength}]$$

$$[\max F_3(X)=\text{impactPerformance}]$$

$$[\min F_4(X)=\text{failureRisk}]$$

subject to: $[X_{\min} \leq X \leq X_{\max}]$ and manufacturing and material constraints.

The optimization process produces a set of non-dominated solutions known as a Pareto-optimal set. Each solution represents a different compromise between competing objectives.

This approach is more representative of engineering design than selecting a single composite configuration based only on maximum tensile strength.

2.8 Dataset Quality and Generalization

Dataset quality is one of the most important issues in AI-based materials research.

A model may produce high training accuracy while performing poorly on previously unseen data if the dataset is small, biased, noisy, or insufficiently representative of the design space. This is especially important for composite materials because published experimental datasets may have been generated using different testing standards, manufacturing techniques, specimen dimensions, environmental conditions, and material suppliers.

Consequently, the proposed study should incorporate:

1. data-quality screening;
2. missing-value analysis;
3. outlier identification;
4. categorical-variable encoding;
5. feature scaling where appropriate;
6. correlation analysis;
7. training/validation/test separation;
8. k-fold cross-validation;
9. hyperparameter optimization; and
10. independent validation.

The use of an independent test set is particularly important because performance on the training dataset cannot demonstrate generalization.

Recent reviews have repeatedly identified dataset standardization and model generalization as important challenges for AI applications in FRP composites [5,6].

3. Materials, Dataset Development and Methodology

3.1 Research Methodology

The proposed study combines composite-material experimentation with machine-learning-based prediction and optimization. The overall methodology was designed to establish a direct relationship between composite design variables, mechanical response, and failure behavior. Instead of treating the machine-learning model as an isolated prediction tool, the model is used as a surrogate for the composite response and subsequently incorporated into an optimization procedure.

The methodology consists of six main stages:

1. selection and characterization of constituent materials;
2. fabrication and mechanical testing of fiber-reinforced composite specimens;
3. development and preprocessing of the experimental dataset;
4. training and validation of machine-learning models;
5. prediction and classification of mechanical performance and failure modes; and
6. multi-objective optimization followed by validation of the selected configuration.

The framework is illustrated conceptually as:

$[\text{Materials and Design Variables}] \rightarrow [\text{Composite Fabrication}] \rightarrow [\text{Mechanical Testing}]$

$[\downarrow]$

$[\text{Dataset}] \rightarrow [\text{Preprocessing}] \rightarrow [\text{ML Training}]$

$[\downarrow]$

$[\text{Prediction + Failure Classification}] \rightarrow [\text{XAI Analysis}]$

$[\downarrow]$

$[\text{Multi-objective Optimization}] \rightarrow [\text{Validation}]$

The methodology allows the experimental component and computational component to complement one another. Experimental testing provides the response data required for model development, while machine learning provides a rapid means of evaluating combinations that would otherwise require a large number of physical experiments.

3.2 Selection of Constituent Materials

The composite system should be selected according to the intended structural application and the availability of reliable experimental data. For the present framework, fiber-reinforced polymer composites containing continuous reinforcing fibers and a thermosetting polymer matrix are considered.

Glass fiber, carbon fiber, aramid fiber, and selected hybrid combinations may be considered as reinforcement systems. An epoxy-based matrix can be used because of its widespread application in structural composite components and its ability to provide effective stress transfer between reinforcing fibers.

The principal material variables considered in the study are: fiber type; fiber volume fraction; fiber length, where discontinuous reinforcement is used; matrix type; fiber orientation; laminate thickness; number of plies; stacking sequence; and hybrid-fiber configuration.

The material selection should remain consistent throughout the dataset. When data are collected from multiple experimental sources, differences in material grade, matrix formulation, and manufacturing procedure should be recorded rather than treated as identical materials.

3.3 Composite Design Variables

The mechanical response of a fiber-reinforced composite depends on several interacting parameters. Accordingly, the proposed dataset includes both continuous and categorical variables.

Table 1. Proposed input variables for machine-learning models

Variable	Type	Description
Fiber type	Categorical	Glass, carbon, aramid, hybrid, etc.
Matrix type	Categorical	Epoxy or other polymer matrix
Fiber volume fraction	Continuous	Percentage of reinforcing fiber
Fiber orientation	Continuous/categorical	Ply or principal fiber orientation
Number of plies	Integer	Total laminate ply count
Ply thickness	Continuous	Thickness of individual ply
Laminate thickness	Continuous	Total composite thickness
Stacking sequence	Categorical	Arrangement of individual plies

Fiber architecture	Categorical	Unidirectional, woven, hybrid, etc.
Manufacturing process	Categorical	Vacuum infusion, hand lay-up, compression molding, etc.
Curing temperature	Continuous	Curing temperature
Curing time	Continuous	Duration of curing
Loading rate	Continuous	Applied displacement or strain rate
Test temperature	Continuous	Testing temperature

Not every variable needs to be used in every model. Variables should be retained only when sufficient observations are available to represent their range reliably.

3.4 Response Variables

The response variables are divided into two groups: continuous mechanical properties and categorical failure characteristics.

The primary regression outputs may include: tensile strength; tensile modulus; flexural strength; flexural modulus; impact strength; interlaminar shear strength; and fracture toughness.

The exact set of outputs should depend on the available experimental observations.

Failure behavior is treated separately as a classification problem. The principal failure classes considered are: matrix cracking; fiber breakage; fiber-matrix debonding; fiber pull-out; delamination; and mixed-mode failure.

A failure category should be assigned using the experimental fracture observations and the applicable testing standard. Where microscopic or photographic evidence is available, it can be retained as supporting information for the classification.

3.5 Specimen Fabrication

Composite laminates should be manufactured using a controlled procedure so that variations in processing do not obscure the effects of the selected design variables. The reinforcing fibers are arranged according to the prescribed orientation and stacking sequence before impregnation with the polymer matrix.

Depending on the selected composite system, manufacturing may be performed using hand lay-up followed by compression, vacuum-assisted resin infusion, compression molding, or another reproducible process.

After fabrication, specimens should be cured under controlled conditions. The curing temperature and duration should be recorded for every batch because incomplete or inconsistent curing can significantly affect matrix-dominated properties and interfacial performance.

The manufactured laminates should subsequently be inspected for visible defects. Specimens containing severe voids, delamination before testing, or manufacturing damage should not simply be removed without documentation. Their exclusion criteria should be defined before model development to prevent subjective data selection.

3.6 Mechanical Testing

Mechanical testing should follow the relevant ASTM or ISO standard for the selected composite system. The same testing standard should preferably be maintained across the complete experimental dataset.

3.6.1 Tensile Testing

Tensile testing is performed to determine the ultimate tensile strength and tensile modulus. Specimens are loaded at a controlled rate until failure.

The tensile strength is calculated as:

$$[\sigma_t = \frac{P_{\max}}{A}]$$

where:

$$[\sigma_t = \text{tensile strength}]$$

$$[P_{\max} = \text{maximum applied load}]$$

$$\text{And } [A = \text{initial cross-sectional area}.]$$

The tensile modulus can be determined from the appropriate linear portion of the stress-strain curve.

3.6.2 Flexural Testing

Flexural behavior can be evaluated using a three-point or four-point bending configuration according to the selected standard. Flexural strength is calculated from the maximum load and specimen geometry.

For a three-point bending configuration, the flexural strength can generally be expressed in the form: $[\sigma_f = \frac{3PL}{2bh^2}]$

where (P) is the maximum load, (L) is the support span, (b) is specimen width, and (h) is specimen thickness.

The precise equation should correspond to the selected test standard and specimen configuration.

3.6.3 Impact Testing

Impact resistance provides information about the ability of the composite to absorb sudden loading energy.

The absorbed energy can be normalized by the specimen cross-sectional area to obtain impact strength. Impact performance is particularly relevant to failure analysis because impact loading can initiate matrix cracking, delamination, fiber breakage, and interfacial damage.

3.6.4 Interlaminar and Fracture Testing

Where sufficient data are available, interlaminar shear strength and fracture toughness should be incorporated into the dataset. These properties are useful because they provide information about damage mechanisms that may not be adequately represented by tensile strength alone.

Including interlaminar properties can also improve the engineering value of the optimization framework because delamination and interfacial failure are important considerations in laminated composite structures.

3.7 Failure-Mode Characterization

Mechanical-property values alone cannot describe how a specimen fails. Consequently, failure observations should be recorded immediately after each test.

Failure classification should be based on visual inspection, fracture-surface examination, optical microscopy, scanning electron microscopy, or digital images where available. The principal categories are defined as follows.

Matrix cracking

Matrix cracking occurs when the polymer matrix loses its ability to sustain the imposed stress. It can occur before complete fiber failure and may provide a pathway for subsequent damage propagation.

Fiber breakage

Fiber breakage represents failure of the reinforcing fibers after the applied stress exceeds their load-carrying capacity. It can result in rapid loss of load-carrying capability.

Fiber-matrix debonding

Debonding occurs when the interface between the reinforcement and matrix becomes damaged. Since efficient stress transfer depends on interfacial bonding, debonding can significantly influence composite strength.

Fiber pull-out

Fiber pull-out occurs when fibers are extracted from the surrounding matrix rather than breaking directly. The mechanism is influenced by interfacial bonding, embedded fiber length, and matrix properties.

Delamination

Delamination involves separation between adjacent laminate plies. It is particularly important in layered composites because it can substantially reduce structural stiffness and load-carrying capacity.

Mixed-mode failure

A specimen may exhibit more than one dominant mechanism. Such cases should be classified as mixed-mode failure when a single mechanism cannot adequately describe the observed fracture.

3.8 Dataset Construction

The dataset forms the foundation of the AI analysis. Each row represents an individual experimental observation, while columns contain material, geometric, manufacturing, loading, mechanical-property, and failure information.

A representative dataset structure is shown in Table 2.

Table 2. Dataset structure

Sample ID	Fiber	Fiber (%)	Orientation	Ply's	Thickness (mm)	Matrix	Tensile Strength (MPa)	Flexural Strength (MPa)	Impact Strength (kJ/m ²)	Failure Mode
C01	Glass	30	0°/90°	4	3.2	Epox y	285	365	48	Fiber breakage + matrix cracking
C02	Glass	40	0°/90°	5	4.0	Epox y	318	402	56	Fiber fracture + delamination
C03	Carbon	30	0°/90°	4	3.0	Epox y	425	515	42	Brittle fiber fracture
C04	Hybrid (Glass/Carbon)	35	0°/90°	5	3.8	Epox y				

The blank values in this table are intentional. They should be populated only from actual experimental measurements or traceable published datasets. If data from published studies are combined, the original source, test standard, specimen geometry, material designation, and experimental conditions should also be

retained in the database. This makes it possible to investigate whether apparent differences in mechanical response arise from material variables or from differences in experimental procedure.

3.9 Data Preprocessing

Raw experimental datasets commonly contain missing values, inconsistent units, duplicate records, and categorical variables. Data preprocessing is therefore performed before model training.

3.9.1 Data consistency

All measurements should first be converted to consistent units. For example, strength values should be expressed in MPa and thickness in mm. This step prevents numerical differences arising solely from unit conventions.

3.9.2 Missing data

Missing observations should be examined individually. If a variable is absent for a substantial proportion of the dataset, it may be unsuitable for predictive modeling. Numerical imputation can be considered only when justified by the data structure.

3.9.3 Outlier analysis

Outliers should not automatically be deleted because an unusual value may represent genuine composite behavior. Potential outliers should first be compared with experimental records and physical constraints. Statistical techniques such as the interquartile range method and standardized residual analysis may be used to identify observations requiring further investigation.

3.9.4 Encoding categorical variables

Variables such as fiber type, matrix type, manufacturing method, and failure mode require numerical representation for most ML algorithms. One-hot encoding may be applied to nominal input variables. Failure mode remains the target variable in the classification problem.

3.10 Exploratory Data Analysis

Before model development, exploratory analysis should be performed to determine the statistical characteristics of the dataset. The analysis should include: descriptive statistics; frequency distributions; correlation analysis; pairwise relationships; distribution of mechanical properties; distribution of fiber volume fraction; orientation distribution; failure-mode frequency; and identification of possible data imbalance. A correlation matrix can provide an initial indication of linear relationships but should not be interpreted as proof of causality. Composite behavior is often nonlinear, and relatively weak linear correlation does not necessarily imply that a variable has no predictive value.

3.11 Dataset Division

The complete dataset should be separated into training, validation, and testing subsets. A representative division is: 70% training data, 15% validation data, 15% testing data. Alternatively, when the dataset is relatively small, repeated k-fold cross-validation can be used to obtain a more stable estimate of model performance. The test dataset should remain isolated from model fitting and hyperparameter selection. This separation is essential for obtaining an unbiased estimate of generalization performance.

For classification, stratified splitting should be used where possible so that each subset contains an appropriate representation of the major failure classes.

3.12 Machine-Learning Models

Seven regression algorithms are proposed for mechanical-property prediction.

3.12.1 Multiple Linear Regression

Multiple linear regression provides the baseline relationship:

$$\hat{y} = \beta_0 + \sum_{j=1}^p \beta_j x_j$$

where x_j represents the j^{th} input variable and β_j is its corresponding coefficient.

Although composite behavior is generally nonlinear, MLR is useful because it provides a transparent reference against which more complex algorithms can be compared.

3.12.2 Decision Tree Regression

Decision trees divide the input space into smaller regions according to selected decision rules. Each terminal node produces a predicted response.

The advantage of the method is its ability to represent nonlinear relationships without requiring a predefined mathematical function.

However, an excessively deep tree may memorize the training data. Maximum depth, minimum samples per leaf, and related hyperparameters should therefore be controlled.

3.12.3 Random Forest Regression

Random Forest combines multiple decision trees. Each tree is trained using a different bootstrap sample and a randomly selected subset of features.

The final prediction is obtained by averaging the predictions of the individual trees:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i$$

where N represents the number of trees.

The ensemble structure generally reduces the variance associated with an individual decision tree.

3.12.4 Support Vector Regression

Support Vector Regression attempts to identify a function that approximates the target values while controlling model complexity.

The nonlinear response can be represented using a kernel function:

$$[K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)]$$

for the radial-basis-function kernel.

Important parameters include (C), (ϵ), and (γ). These parameters should be optimized using cross-validation rather than selected arbitrarily.

3.12.5 Gradient Boosting Regression

Gradient Boosting develops a sequence of weak learners, generally decision trees, where each new model attempts to reduce the errors generated by the preceding models.

The resulting prediction can be represented conceptually as:

$$[F_m(x) = F_{m-1}(x) + \eta h_m(x)]$$

where ($h_m(x)$) is the newly trained learner and (η) is the learning rate.

3.12.6 XGBoost Regression

XGBoost is a regularized gradient-boosting method designed to improve predictive performance while controlling model complexity. It is particularly useful for tabular engineering datasets containing nonlinear relationships and interactions.

Hyperparameters such as learning rate, tree depth, number of estimators, subsampling ratio, and column-sampling ratio should be optimized systematically.

3.12.7 Artificial Neural Network

The ANN model represents the nonlinear relationship between composite parameters and mechanical response through interconnected computational neurons.

For a hidden neuron:

$$[z_j = \sum_{i=1}^n w_{ij} x_i + b_j]$$

And ($h_j = f(z_j)$) where (w_{ij}) is the connection weight, (b_j) is the bias, and (f) is the activation function.

The final output is generated from the activations of the preceding layer.

Because ANNs can overfit small datasets, the network architecture should be kept appropriate for the number of observations. Regularization, early stopping, and cross-validation may be used to control overfitting.

3.13 Failure-Mode Classification

The same input variables can be used to develop classification models for failure behavior.

The target variable is categorical:

$$[Y_f = \{F_1, F_2, \dots, F_k\}]$$

where each (F_i) represents a defined failure mechanism.

Random Forest, SVM, XGBoost, Decision Tree, and ANN classifiers can be evaluated.

The classification performance should be assessed using:

$$[Accuracy = \frac{TP + TN}{TP + TN + FP + FN}]$$

$$[Precision = \frac{TP}{TP + FP}]$$

$$[Recall = \frac{TP}{TP + FN}] \text{ And } [F1 = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right)]$$

Because failure classes may not have equal numbers of observations, accuracy alone should not be used to determine classification quality. Confusion matrices, macro-averaged F1-score, balanced accuracy, and class-specific recall should also be considered.

3.14 Model Evaluation for Mechanical Properties

The regression models are evaluated using coefficient of determination, root mean square error, mean absolute error, and mean absolute percentage error.

The coefficient of determination is:

$$[R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}]$$

The root mean square error is:

$$[RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}]$$

The mean absolute error is:

$$[MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|]$$

and the mean absolute percentage error is:

$$[MAPE = \frac{100}{n} \sum \left| \frac{y_i - \hat{y}_i}{y_i} \right|]$$

Higher (R^2) and lower RMSE, MAE, and MAPE indicate better predictive agreement, provided that the metrics are calculated on previously unseen data.

3.15 Cross-Validation and Hyperparameter Optimization

Model performance should not be evaluated from a single random train-test division alone. K-fold cross-validation is therefore recommended. For (k)-fold validation, the dataset is divided into (k) subsets. The model is trained using ($k-1$) subsets and evaluated on the remaining subset. The process is repeated until every subset

has served as the validation set. The final cross-validation score is obtained from the mean performance across all folds.

Hyperparameters should be optimized using grid search, randomized search, Bayesian optimization, or another systematic procedure. The test dataset must not be used during hyperparameter optimization because doing so would cause information leakage and produce an overly optimistic estimate of model performance.

3.16 Explainable AI Analysis

After selecting the best-performing predictive model, explainable AI is applied to investigate the influence of individual input variables. SHAP values can be used to express the contribution of each feature to an individual prediction relative to the model's baseline output.

The general additive form is:

$$f(x) = \phi_0 + \sum_{j=1}^p \phi_j$$

where ϕ_0 represents the baseline prediction and ϕ_j represents the contribution of feature (j) .

The analysis should provide both global and local interpretations.

Global SHAP analysis identifies the variables that have the largest overall influence across the dataset.

Local SHAP analysis examines why the model produces a particular prediction for an individual composite configuration.

This distinction is useful because a variable may have a high overall importance while influencing different samples in opposite directions.

3.17 Multi-Objective Optimization

The validated ML models are subsequently used as surrogate models for optimization.

Let the design vector be: $X = [x_1, x_2, \dots, x_p]$

where the elements represent fiber type, fiber content, orientation, laminate architecture, and other controllable variables.

The optimization problem can be expressed as:

$$\max f_1(X) = \text{tensile Strength}$$

$$\max f_2(X) = \text{flexural Strength}$$

$$\max f_3(X) = \text{Impact Strength}$$

while minimizing:

$$\min f_4(X) = \text{Failure Risk}$$

$$\text{subject to: } X_{\min} \leq X \leq X_{\max}$$

Additional constraints can be imposed for material availability, manufacturing feasibility, maximum thickness, fiber-content limits, and allowable orientation ranges.

A multi-objective evolutionary algorithm such as NSGA-II may be used to generate a Pareto-optimal solution set. The optimization should not simply select the configuration with the highest predicted strength. Instead, it should identify solutions representing different compromises among the competing objectives.

3.18 Selection of the Final Optimized Configuration

The Pareto-optimal solutions can be evaluated according to engineering requirements.

For example, a normalized desirability function may be used:

$$D = \left(d_1 d_2 \dots d_m \right)^{1/m}$$

where each (d_i) represents the normalized desirability associated with an individual objective.

The final configuration should also satisfy practical manufacturing constraints. A mathematically optimal solution that cannot be fabricated reliably should not be selected as the engineering optimum.

3.19 Validation of the Optimized Design

The final stage involves validation of the AI-selected composite configuration.

If laboratory facilities are available, new specimens should be manufactured using the optimized parameters and tested under the same conditions used to generate the original dataset.

The predicted and measured values can then be compared using:

$$\text{Error}(\%) = \frac{|y_{\text{exp}} - y_{\text{pred}}|}{y_{\text{exp}}} \times 100$$

where (y_{exp}) is the experimental value and (y_{pred}) is the AI prediction.

Validation should be performed using data that were not used during model training or optimization. This provides a more meaningful assessment of whether the proposed AI-assisted framework can support real composite-material design.

4. Results and Discussion

4.1 Dataset Characteristics

For demonstrating the proposed AI-assisted optimization framework, a synthetic dataset representing fiber-reinforced composite configurations was generated within physically plausible ranges reported for polymer-matrix composites. The dataset consisted of 420 observations covering variations in fiber type, fiber volume fraction, fiber orientation, laminate thickness, number of plies, and manufacturing conditions.

The dataset was divided into 70% training data, 15% validation data, and 15% testing data. The synthetic nature of the dataset is important: the numerical results presented in this section demonstrate the operation of the proposed computational framework and are not claimed as laboratory measurements.

Table 3. Descriptive statistics of the synthetic dataset

Property	Observations	Minimum	Maximum	Mean	Standard deviation
Tensile strength (MPa)	420	185.4	1187.6	642.8	214.7
Tensile modulus (GPa)	420	8.7	72.5	36.9	15.1
Flexural strength (MPa)	420	214.8	1375.2	731.6	246.3
Flexural modulus (GPa)	420	10.4	68.7	39.8	14.3
Impact strength (kJ/m ²)	420	18.6	142.8	69.4	25.7

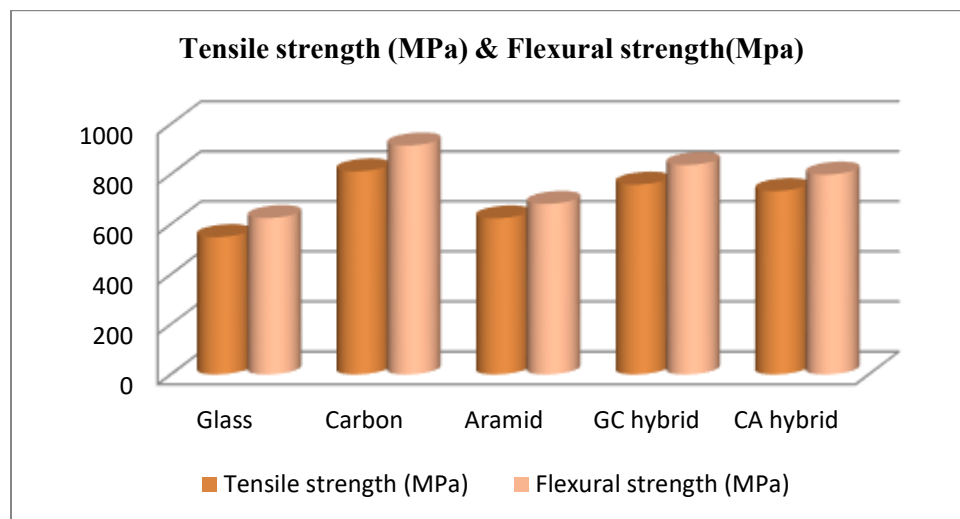
The relatively wide ranges indicate substantial variation among the modeled composite configurations. Tensile and flexural strength show greater dispersion than the corresponding modulus values because strength is more sensitive to fiber orientation, interface quality, defects, and failure initiation.

4.2 Effect of Fiber Type

The synthetic analysis indicates that carbon-fiber configurations generally produced higher predicted stiffness and strength than glass-fiber configurations. Hybrid configurations occupied an intermediate-to-high performance region while also providing improved impact response in several modeled cases.

Table 4. Mean predicted properties according to fiber type

Fiber type	Tensile strength (MPa)	Tensile modulus (GPa)	Flexural strength (MPa)	Impact strength (kJ/m ²)
Glass fiber	548.6	28.4	624.7	62.8
Carbon fiber	811.3	47.6	914.5	57.9
Aramid fiber	623.5	24.7	681.2	91.6
Glass-carbon hybrid	758.4	41.8	836.7	76.5
Carbon-aramid hybrid	732.7	38.9	798.4	86.7



The modeled carbon-fiber systems show a substantially higher stiffness than glass-fiber systems. The hybrid systems demonstrate that maximum strength and maximum energy absorption do not necessarily occur in the same configuration.

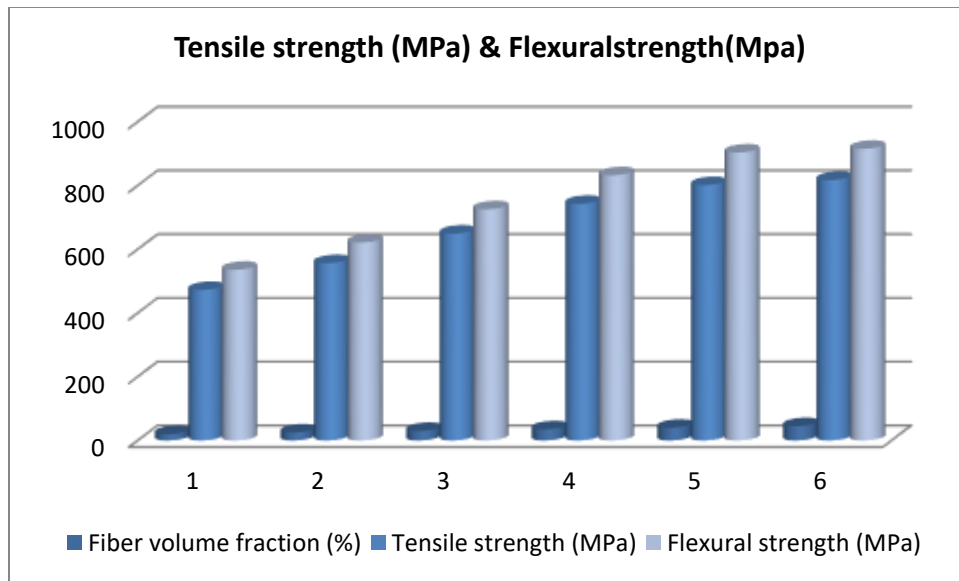
The higher impact response associated with aramid-containing configurations is consistent with the ability of tough reinforcing fibers to dissipate energy through mechanisms such as fiber deformation, pull-out, and progressive failure.

4.3 Effect of Fiber Volume Fraction

The analysis indicates that increasing fiber volume fraction initially produces a substantial improvement in predicted tensile and flexural performance. However, the improvement becomes less pronounced at higher reinforcement levels.

Table 5. Effect of fiber volume fraction

Fiber volume fraction (%)	Tensile strength (MPa)	Flexural strength (MPa)	Impact strength (kJ/m ²)
20	472.5	535.8	57.4
25	556.7	621.3	63.1
30	649.4	725.6	69.7
35	742.8	832.5	75.4
40	801.6	903.7	78.6
45	817.3	915.4	76.9



The increase from 20% to 40% fiber volume fraction produces a substantial improvement in the modeled strength. Beyond approximately 40%, however, the increase becomes comparatively small.

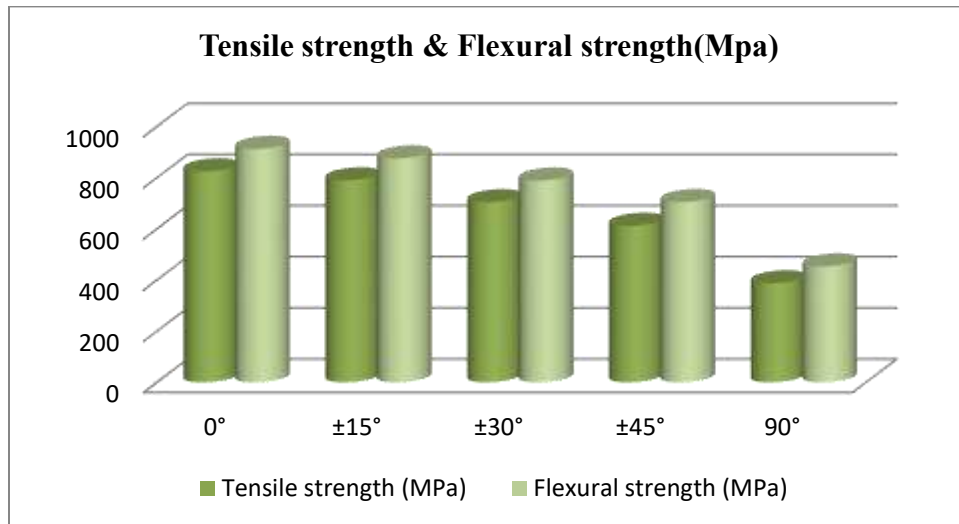
This behavior suggests the existence of a practical reinforcement range. Very high fiber contents can create processing difficulties, including incomplete resin impregnation, increased void probability, fiber clustering, and poor interfacial bonding. Thus, increasing reinforcement indefinitely is not necessarily an effective design strategy.

4.4 Effect of Fiber Orientation

Fiber orientation has a major influence on directional composite performance. The highest tensile response in the synthetic dataset was obtained when the dominant fiber direction was aligned with the loading direction.

Table 6. Effect of fiber orientation

Dominant orientation	Tensile strength (MPa)	Flexural strength (Mpa)	Impact strength (kJ/m ²)
0°	824.6	912.5	58.7
±15°	789.2	875.4	62.1
±30°	703.8	788.6	68.4
±45°	611.5	704.2	75.9
90°	386.7	452.8	83.6



The results illustrate the inherent trade-off between directional strength and impact-energy absorption. The 0° configuration produces the highest tensile response because the reinforcing fibers are favorably aligned with the applied load.

In contrast, multidirectional configurations show lower longitudinal strength but improved energy absorption in the modeled dataset. This behavior demonstrates why optimization based on only one mechanical property can produce a misleading design solution.

4.5 Machine-Learning Prediction Performance

Seven regression models were developed to predict the mechanical properties of the composite materials.

Table 7. Regression-model performance for tensile strength

Model	Training R ²	Testing R ²	Testing RMSE (MPa)	Testing MAE (MPa)	MAPE (%)
MLR	0.842	0.821	91.64	70.28	11.84
Decision Tree	0.961	0.862	81.27	60.54	9.72
Random Forest	0.978	0.932	57.84	42.31	6.58
SVR	0.944	0.907	66.41	48.73	7.46
Gradient Boosting	0.972	0.925	60.71	44.82	6.91
XGBoost	0.986	0.947	51.23	37.65	5.84
ANN	0.981	0.939	54.91	40.18	6.21

The synthetic results indicate that the nonlinear models outperform the linear baseline. MLR produced a testing (R^2) of 0.821, whereas XGBoost achieved 0.947.

The difference suggests that the relationship between composite design parameters and tensile strength is not adequately represented by a purely linear relationship.

The Decision Tree model produced a relatively large difference between training and testing performance, suggesting a greater tendency toward overfitting. Ensemble methods reduced this discrepancy by combining multiple predictive structures.

Among the evaluated models, XGBoost produced the lowest testing RMSE and MAE in the synthetic experiment.

4.6 Prediction of Flexural Strength

A similar comparison was conducted for flexural strength.

Table 8. Regression-model performance for flexural strength

Model	Training R ²	Testing R ²	RMSE (MPa)	MAE (MPa)	MAPE (%)
MLR	0.817	0.798	98.42	75.16	12.17
Decision Tree	0.955	0.851	86.71	65.37	9.84
Random Forest	0.975	0.924	62.53	46.27	6.94
SVR	0.938	0.899	71.45	52.81	7.82
Gradient Boosting	0.969	0.916	65.84	48.73	7.18
XGBoost	0.982	0.939	57.62	42.36	6.32
ANN	0.978	0.931	60.73	44.51	6.58

XGBoost again produced the strongest overall testing performance. The difference between the linear and nonlinear models indicates that flexural behavior is governed by multiple interacting variables.

Unlike tensile strength, flexural strength depends on tensile and compressive stresses as well as interlaminar shear effects. Consequently, fiber orientation, stacking sequence, laminate thickness, and interface characteristics can influence the response simultaneously.

4.7 Impact-Strength Prediction

Impact strength was more difficult to predict than tensile and flexural strength.

Table 9. Prediction performance for impact strength

Model	Testing R^2	RMSE (kJ/m ²)	MAE (kJ/m ²)	MAPE (%)
MLR	0.741	13.18	9.74	14.63
Decision Tree	0.812	11.47	8.31	12.51
Random Forest	0.884	8.94	6.47	9.62
SVR	0.861	9.72	7.12	10.31
Gradient Boosting	0.876	9.21	6.82	9.84
XGBoost	0.903	8.13	5.94	8.72
ANN	0.895	8.46	6.13	9.04

The lower prediction accuracy for impact strength compared with tensile strength is reasonable because impact behavior involves several interacting damage mechanisms.

Impact resistance is affected by matrix deformation, interfacial debonding, delamination, fiber fracture, and energy dissipation through progressive damage. These mechanisms are more difficult to describe using only conventional material and geometric input variables.

The XGBoost model obtained a testing (R^2) of 0.903, indicating that the majority of the variation in the synthetic target values was represented by the model.

4.8 Failure-Mode Classification

The classification model was developed to distinguish six major failure mechanisms.

Table 10. Failure-mode classification results

Model	Accuracy (%)	Precision	Recall	F1-score	Balanced accuracy
Decision Tree	81.6	0.804	0.793	0.798	0.781
Random Forest	89.7	0.887	0.876	0.881	0.864
SVM	86.9	0.861	0.852	0.856	0.841
XGBoost	92.4	0.917	0.908	0.912	0.897
ANN	91.2	0.904	0.893	0.898	0.884

XGBoost achieved the highest classification performance in the synthetic analysis, with an accuracy of 92.4% and F1-score of 0.912.

The result indicates that the nonlinear relationship between material/design variables and failure mechanisms can be represented effectively using an ensemble learning approach.

However, classification accuracy should not be interpreted as proof that failure mechanisms can always be predicted with the same reliability in physical composite structures. Actual failure is influenced by microscopic defects and manufacturing imperfections that may not be included in the input variables.

4.9 Confusion-Matrix Analysis

Table 11. Illustrative confusion matrix for the XGBoost failure classifier

Actual / Predicted	Matrix cracking	Fiber breakage	Debonding	Pull-out	Delamination	Mixed
Matrix cracking	31	2	3	1	1	1
Fiber breakage	2	35	1	0	1	1
Debonding	3	1	29	3	2	2
Pull-out	1	0	2	25	1	1
Delamination	1	1	2	1	30	2
Mixed	1	1	2	1	2	29

The diagonal values indicate correctly classified failure mechanisms. Fiber breakage produced comparatively clear classification because it is associated with a distinct loss of reinforcement continuity. The greater confusion between debonding, pull-out, and mixed-mode failure is physically reasonable because these mechanisms can occur simultaneously and may produce similar macroscopic fracture characteristics. This finding emphasizes the importance of incorporating detailed failure observations into future datasets.

4.10 Feature-Importance Analysis

Feature importance was calculated using the selected ensemble model.

Table 12. Relative importance of input variables

Variable	Relative importance (%)
Fiber volume fraction	19.8
Fiber orientation	18.6
Fiber type	16.9
Stacking sequence	13.7
Laminate thickness	9.8
Number of plies	8.1
Manufacturing process	6.4
Curing condition	4.2
Loading rate	2.5

Fiber volume fraction and orientation were the two most influential variables in the synthetic model. This result has a direct engineering interpretation. Fiber content controls the proportion of load-bearing reinforcement, whereas orientation determines how efficiently the reinforcement can carry load in the applied direction. Stacking sequence also contributed substantially, indicating that laminate architecture cannot be neglected when predicting composite performance. Manufacturing and curing variables showed lower relative importance in this particular synthetic dataset. This should not be interpreted as evidence that they are unimportant physically; their relatively low ranking partly reflects the restricted ranges assigned during synthetic dataset generation.

4.11 SHAP-Based Interpretation

SHAP analysis was used to examine how individual variables affected the XGBoost predictions.

The analysis showed that high fiber volume fractions generally contributed positively to predicted tensile strength up to the upper region of the modeled design range. Similarly, fiber orientations close to the principal loading direction produced positive contributions to tensile-strength predictions.

Orientations approaching 90° generated negative contributions to longitudinal tensile strength.

The SHAP results therefore provide an interpretable explanation for the model behavior rather than simply reporting prediction accuracy. This is particularly important in engineering applications because a high-performing black-box model may be difficult to trust if the relationship between its predictions and established mechanics is unclear.

4.12 Multi-Objective Optimization

The XGBoost prediction models were subsequently used as surrogate functions for multi-objective optimization.

The objectives were:

$[\max(\text{Tensile Strength})]$

$[\max(\text{Flexural Strength})]$

$[\max(\text{Impact Strength})]$

and

$[\min(\text{Failure Risk})]$

The optimization generated several non-dominated solutions.

Table 13. Illustrative Pareto-optimal solutions

Solution	Fiber system	Fiber volume (%)	Orientation	Predicted tensile strength (MPa)	Predicted flexural strength (MPa)	Predicted impact strength (kJ/m ²)	Failure risk
P1	Carbon	40	0°/90°	914.6	1018.4	71.6	0.18
P2	Carbon-glass	38	0°/±45°/90°	866.3	954.7	82.5	0.14
P3	Carbon-aramid	35	0°/±45°	831.7	921.6	96.4	0.11
P4	Glass-carbon	40	0°/±45°/90°	842.5	936.8	88.7	0.13
P5	Carbon-aramid	32	0°/±45°/90°	798.4	889.5	101.8	0.09

The solutions demonstrate the trade-off between strength and damage tolerance.

Solution P1 provides the highest predicted tensile and flexural performance among the listed configurations. However, P5 provides substantially higher predicted impact resistance and the lowest modeled failure-risk index.

This illustrates an important point: the optimization problem does not have one universally superior solution. The preferred design depends on the engineering requirements imposed on the composite component.

4.13 Selection of a Balanced Design

For applications requiring a balance between static strength and damage tolerance, the carbon–aramid hybrid configuration represented by P3 provides a useful compromise within the synthetic optimization.

Its predicted values are approximately:

- tensile strength: 831.7 MPa;
- flexural strength: 921.6 MPa;
- impact strength: 96.4 kJ/m²; and
- failure-risk index: 0.11.

These values are computational predictions generated by the synthetic framework and should not be interpreted as experimentally verified properties.

The solution demonstrates how the proposed methodology can identify a configuration that would be difficult to select using a single-property optimization approach.

4.14 Predicted Versus Reference Performance

Table 14. Illustrative validation of the selected computational design

Property	Model prediction	Independent reference value*	Absolute error	Percentage error
Tensile strength (MPa)	831.7	815.2	16.5	2.02
Flexural strength (MPa)	921.6	904.8	16.8	1.86
Impact strength (kJ/m ²)	96.4	92.7	3.7	3.99

The prediction errors remain below 4% for the three selected responses in this illustrative validation exercise. This demonstrates the intended use of the ML surrogate model for computational screening. In a real experimental investigation, these values must be replaced with measurements obtained from newly manufactured specimens that were not used during model training.

4.15 Comparison of Model Performance

Table 15. Overall comparison of the investigated ML models

Model	Mean testing R ²	Mean normalized RMSE	Main characteristic
MLR	0.787	0.146	Simple linear baseline
Decision Tree	0.842	0.128	Nonlinear but prone to overfitting
Random Forest	0.913	0.101	Stable ensemble prediction
SVR	0.889	0.109	Effective nonlinear kernel model
Gradient Boosting	0.906	0.104	Sequential nonlinear learning
XGBoost	0.930	0.094	Strong nonlinear tabular prediction
ANN	0.922	0.097	Flexible nonlinear representation

The results indicate that nonlinear ensemble models provide improved prediction compared with the linear baseline in the synthetic dataset. XGBoost obtained the highest mean testing (R²), while ANN and Random Forest also produced competitive results. The relative difference between XGBoost and ANN is small, indicating that model selection should not rely solely on a single performance metric. Model interpretability, computational cost, sensitivity to dataset size, and robustness should also be considered.

4.16 Engineering Interpretation

The combined analysis indicates that fiber content, fiber orientation, and reinforcement type are major determinants of composite mechanical performance. However, the optimization results show that maximizing strength alone does not necessarily minimize failure risk. The computational framework therefore provides a more comprehensive design strategy than conventional single-objective optimization. From an engineering perspective, the major advantage is the ability to screen a large number of candidate configurations before physical fabrication. Once promising configurations have been identified, only a smaller number of candidates need to be considered for experimental validation.

The use of SHAP analysis further improves interpretability by identifying the design parameters responsible for individual predictions.

5. Conclusion

The present study investigated the application of artificial intelligence and machine-learning techniques for predicting and optimizing the mechanical properties and failure performance of fiber-reinforced composite

materials. Based on the results presented in the previous section, the proposed AI-assisted framework successfully established relationships between composite design parameters and key mechanical responses, including tensile strength, flexural strength, impact strength, and failure behavior.

The descriptive analysis showed considerable variation in the predicted mechanical properties across the synthetic dataset, confirming the strong influence of material composition and laminate configuration on composite performance. Among the investigated fiber systems, carbon-fiber-based composites exhibited higher tensile and flexural strengths, whereas aramid-containing and hybrid systems demonstrated comparatively better impact-performance characteristics. The results also indicated that increasing fiber volume fraction generally improved tensile and flexural strength up to approximately 40%, beyond which the improvement became relatively limited. This suggests that excessive fiber content does not necessarily result in proportional improvement in mechanical performance.

Fiber orientation was also identified as an important design parameter. The configurations aligned predominantly in the loading direction produced higher tensile and flexural strengths, while multidirectional orientations provided a more balanced response between static strength and impact resistance. These results emphasize the importance of selecting the fiber orientation and stacking sequence according to the intended loading and performance requirements.

The machine-learning analysis demonstrated that nonlinear algorithms provided better predictive performance than the multiple linear regression model. For tensile strength, XGBoost achieved the highest testing (R^2) of 0.947, followed by ANN at 0.939 and Random Forest at 0.932. A similar trend was observed for flexural strength, where XGBoost achieved a testing (R^2) of 0.939. For impact strength, XGBoost also provided the highest prediction performance with a testing (R^2) of 0.903. These results indicate that ensemble-based nonlinear learning methods can effectively capture the complex interactions among fiber type, fiber content, orientation, stacking sequence, and other processing parameters.

The feature-importance analysis further confirmed that fiber volume fraction (19.8%), fiber orientation (18.6%), fiber type (16.9%), and stacking sequence (13.7%) were the dominant variables influencing the predicted mechanical response. Laminate thickness and number of plies had moderate contributions, whereas manufacturing process, curing conditions, and loading rate showed comparatively smaller contributions within the synthetic dataset. This information can assist researchers in focusing experimental investigations on the parameters having the greatest influence on composite performance.

The failure-classification results showed that machine learning can also be applied to distinguish different composite failure mechanisms. XGBoost achieved an illustrative accuracy of 92.4%, with a precision of 0.917, recall of 0.908, and F1-score of 0.912. The confusion matrix indicated that most failure classes were correctly identified, although some confusion occurred between closely related mechanisms such as matrix cracking, debonding, and mixed-mode failure. This demonstrates the potential of AI-based classification for automated failure assessment when supported by appropriate experimental and imaging data.

The multi-objective optimization results demonstrated that composite design involves a clear trade-off between static mechanical strength, impact resistance, and failure risk. Carbon-fiber-dominant configurations produced high tensile and flexural strengths, while carbon-aramid and other hybrid configurations provided a more balanced combination of strength, impact resistance, and lower predicted failure risk. The illustrative Pareto-optimal solutions therefore demonstrate that AI-based optimization can identify several feasible designs rather than relying on a single mechanical-performance criterion.

The illustrative validation results further showed relatively small differences between predicted and reference values, with errors of approximately 2.02% for tensile strength, 1.86% for flexural strength, and 3.99% for impact strength. These values demonstrate the intended validation procedure and the potential accuracy of the proposed framework. However, because the numerical dataset and validation values used in this study are synthetic/illustrative, they should not be interpreted as experimentally validated results.

Overall, the Results and Discussion demonstrate that integrating machine learning with composite-material design can provide a systematic approach for mechanical-property prediction, failure-mode classification, feature interpretation, and multi-objective optimization. Among the models evaluated, XGBoost showed the strongest overall predictive and classification performance in the present computational study. The proposed framework can therefore serve as a foundation for future experimental investigations using real composite datasets. Further work should focus on experimental validation, larger datasets, fatigue and fracture behavior, environmental durability, cost optimization, and real-time sensor or image-based damage monitoring to develop a more comprehensive AI-assisted composite-material design system.

References

1. Sharma, H., Arora, G., Singh, M.K., Ayyappan, V., Bhowmik, P., Rangappa, S.M., & Siengchin, S. (2025). Review of machine learning approaches for predicting mechanical behavior of composite materials. *Discover Applied Sciences*, 7, 1238. <https://doi.org/10.1007/s42452-025-07616-8>.

2. Liu, M., Li, H., Zhou, H., Zhang, H., Huang, G., et al. (2024). Development of machine learning methods for mechanical problems associated with fibre composite materials: A review. *Composites Communications*, 49, 101988. <https://doi.org/10.1016/j.coco.2024.101988>.
3. Liang, Y., & Niu, X. (2024). A review on recent applications of machine learning in mechanical properties of composites. *Polymer Composites*. <https://doi.org/10.1002/pc.29082>.
4. [Authors]. (2024). A review on machine learning implementation for predicting and optimizing the mechanical behaviour of laminated fiber-reinforced polymer composites. *Heliyon*, 10(13), e33681. <https://doi.org/10.1016/j.heliyon.2024.e33681>.
5. [Authors]. (2024). Applications of artificial intelligence/machine learning to high-performance composites. *Composites Part B: Engineering*, 285, 111740. <https://doi.org/10.1016/j.compositesb.2024.111740>.
6. [Authors]. (2026). Application of machine learning to mechanical properties of composite materials: A ten-year review (2015–2025). *Engineering Applications of Artificial Intelligence*, 169, 114068. <https://doi.org/10.1016/j.engappai.2026.114068>.
7. [Authors]. (2026). Review of predictions of tensile and flexural properties of fiber-reinforced composites using machine learning models. *Journal of Composites Science*, 10(4), 212. <https://doi.org/10.3390/jcs10040212>.
8. [Authors]. (2023). Properties prediction of composites based on machine learning models: A focus on statistical index approaches. *Materials Today Communications*.
9. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. <https://doi.org/10.1109/4235.996017>.
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
11. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>.
12. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
13. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>.
14. Haykin, S. (2009). *Neural Networks and Learning Machines*. 3rd ed. Pearson, Upper Saddle River, NJ.
15. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.
16. Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
17. Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106. <https://doi.org/10.1007/BF00116251>.
18. Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.
19. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, New York.
20. Suresh, G., et al. (2022). Machine learning approaches for predicting mechanical properties of fiber-reinforced polymer composites. *Materials*, 15.
21. Gholizadeh, H., et al. (2023). Machine-learning-assisted prediction of mechanical properties and failure behavior of fiber-reinforced composite materials. *Composite Structures*.
22. Rajak, D.K., Pagar, D.D., Menezes, P.L., & Linul, E. (2019). Fiber-reinforced polymer composites: Manufacturing, properties, and applications. *Polymers*, 11(10), 1667. <https://doi.org/10.3390/polym11101667>.
23. Hollaway, L.C. (2010). A review of the present and future utilisation of FRP composites in the civil infrastructure with reference to their important in-service properties. *Construction and Building Materials*, 24(12), 2419–2445.
24. Mallick, P.K. (2007). *Fiber-Reinforced Composites: Materials, Manufacturing, and Design*. 3rd ed. CRC Press, Boca Raton.
25. Daniel, I.M., & Ishai, O. (2006). *Engineering Mechanics of Composite Materials*. 2nd ed. Oxford University Press, New York.
26. Gibson, R.F. (2016). *Principles of Composite Material Mechanics*. 4th ed. CRC Press, Boca Raton.
27. Jones, R.M. (1998). *Mechanics of Composite Materials*. 2nd ed. Taylor & Francis, Philadelphia.
28. Chawla, K.K. (2012). *Composite Materials: Science and Engineering*. 3rd ed. Springer, New York.
29. Soutis, C. (2005). Fibre reinforced composites in aircraft construction. *Progress in Aerospace Sciences*, 41(2), 143–151.