Journal of
Polymer Materials

Tech Science Press

https://doi.org/10.32604/jpm.2026.076807

ARTICLE

Machine Learning-Accelerated Materials Genome Design of Hybrid Fiber Composites for Electric Vehicle Lightweighting

Chin-Wen Liao[1,2,3], En-Shiuh Lin[1], Wei-Lun Huang[4,5,6], I-Chi Wang[7], Bo-Siang Chen[8,*] and Wei-Sho Ho[1,2,9,*]

[1]Department of Electrical and Mechanical Technology, National Changhua University of Education, Bao-Shan Campus, No. 2, Shi-Da Rd, Changhua City, Taiwan
[2]Graduate Institute of Technological and Vocational Education, National Changhua University of Education, Bao-Shan Campus, No. 2, Shi-Da Rd., Changhua City, Taiwan
[3]Center of Teacher Education, National Chung Hsing University, No. 145, Xingda Rd., South Dist., Taichung City, Taiwan
[4]Medical Affairs Office, National Taiwan University Hospital, No. 7, Zhongshan S. Rd., Zhongzheng Dist., Taipei City, Taiwan
[5]Department of Health Services Adminstration, China Medical University, No. 100, Sec. 1, Jingmao Rd., Beitun Dist., Taichung City, Taiwan
[6]Department of Health Care Management, National Taipei University of Nursing and Health Sciences, No. 365, Mingde Rd., Beitou Dist., Taipei City, Taiwan
[7]Liberal Education Center, College of General Education, National Chin-Yi University of Technology, No. 57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung City, Taiwan
[8]Department of Vehicle Engineering, Nan Kai University of Technology, No. 568, Zhongzheng Rd., Caotun Township, Nantou City, Taiwan
[9]NCUE Alumni Association, National Changhua University of Education Jin-De Campus, No. 1, Jinde Rd., Changhua City, Taiwan

*Corresponding Authors: Bo-Siang Chen. Email: toyota554@gmail.com; Wei-Sho Ho. Email: homaintain@gmail.com

Received: 27 November 2025; Accepted: 09 February 2026; Published: 03 April 2026

ABSTRACT: The demand for extended electric vehicle (EV) range necessitates advanced lightweighting strategies. This study introduces a materials genome approach, augmented by machine learning (ML), for optimizing lightweight composite designs for EVs. A comprehensive materials genome database was developed, encompassing composites based on carbon, glass, and natural fibers. This database systematically records critical parameters such as mechanical properties, density, cost, and environmental impact. Machine learning models, including Random Forest, Support Vector Machines, and Artificial Neural Networks, were employed to construct a predictive system for material performance. Subsequent material composition optimization was performed using a multi-objective genetic algorithm. Experimental validation demonstrated that an optimized carbon fiber/bio-based resin composite achieved a 45% weight reduction compared to conventional steel, while maintaining equivalent structural strength. The predictive accuracy of the models reached 94.2%. A cost-benefit analysis indicated that despite a 15% increase in material cost, the overall vehicle energy consumption decreased by 12%, leading to an 18% total cost saving over a five-year operational lifecycle, under a representative mid-size battery electric vehicle (BEV) operational scenario.

KEYWORDS: Materials genomics; machine learning; lightweight composites; multi-objective optimization; electric vehicles

1 Introduction

The increasing global emphasis on energy transition and environmental sustainability has propelled electric vehicles (EVs) from a niche market into the mainstream automotive industry. According to the

International Energy Agency (IEA), global EV sales surpassed 14 million units in 2023, representing nearly 18% of new vehicle sales and indicating rapid growth within the EV sector [1]. Among the factors influencing EV competitiveness, driving range and energy efficiency remain primary concerns for both consumers and manufacturers [2–4]. Beyond continuous breakthroughs in battery technology, vehicle weight reduction is a critical determinant of both driving range and energy efficiency [3]. Consequently, achieving vehicle lightweighting through advanced material and structural design, without compromising safety and performance, has emerged as a central challenge for the automotive industry and academia [5].

Composite materials have emerged as leading candidates for automotive lightweighting due to their superior specific strength, corrosion resistance, and design flexibility [6]. In contrast to the predominantly used but higher-density steel and aluminum alloys in traditional internal combustion engine vehicles, advanced composites such as carbon fiber reinforced plastics (CFRPs), glass fiber reinforced plastics (GFRPs), and even emerging natural fiber-based composites offer significant weight reduction and design tailorability. These materials are increasingly adopted in body panels, suspension components, and interior structures [7–9], playing a pivotal role in multi-objective optimization strategies for electric vehicle enclosures and structural modules [10]. However, the performance of composite materials is highly dependent on a complex interplay of constituent ratios, fiber types, matrix properties, and processing conditions. Specifically, variations in interfacial adhesion driven by structural treatments (e.g., alkaline modifications of natural fibers) significantly alter the mechanical and morphological baselines of the resulting polymer matrix [11,12]. This multi-factor coupling renders their design and development processes inherently complex and challenging to predict using conventional empirical models [13]. Current industrial practice largely relies on extensive experimental trials and iterative optimization for new material design. This approach, spanning material selection, composition adjustment, and performance validation, typically entails a lengthy research and development cycle of three to five years. Such protracted timelines lead to high costs and significantly impede the introduction of novel materials at a pace commensurate with evolving market demands [14].

The advent of Materials Genomics (MG) offers a promising avenue to address this bottleneck [15], establishing a new paradigm that accelerates the deployment of novel materials through complementary efforts in theory, computation, and high-throughput experiments [16]. At its core, MG leverages high-throughput experimentation, computational simulations, and big data analytics to systematically establish the relationships between material structure and properties, thereby enabling rapid material screening and design [17]. Concurrently, the successful application of machine learning (ML) in diverse fields has prompted researchers to integrate it into material design workflows, fundamentally transforming how complex microstructure-property relationships are established [18]. Although existing studies have demonstrated significant accuracy in predicting material properties using ML models, current research still encounters limitations such as data scarcity, restricted feature selection, and inadequate consideration of multi-objective design. Specifically within the domain of EV composite material design, conventional research predominantly focuses on optimizing single material properties, often neglecting multi-dimensional factors such as weight, cost, and environmental impact. Furthermore, most existing methodologies primarily offer performance prediction without providing subsequent guidance for material composition design, thus hindering the direct translation of research outcomes into industry applications that address real-world requirements. Therefore, integrating multi-objective optimization strategies while maintaining prediction accuracy, coupled with experimental validation to accelerate new material development, remains a critical research gap that requires addressing.

Against this background, this study proposes an integrated optimization methodology for electric vehicle composite material design, leveraging the principles of Materials Genomics and Machine Learning. To achieve this, the research will first establish a Materials Genomics database encompassing diverse

compositions, including carbon, glass, and natural fibers. This database will incorporate multi-parametric data such as mechanical properties, density, cost, and environmental impact. Subsequently, various machine learning models, including Random Forest, Support Vector Machine (SVM), and neural networks, will be employed to systematically compare their predictive performance. Building upon this, a Multi-Objective Genetic Algorithm (MOGA) will be applied for the global optimization of material composition and performance. This optimization aims to explore the intricate trade-offs among weight reduction, structural strength retention, and cost control, while environmental impact is not directly included as an optimization objective, but is evaluated in a post-optimization LCC/LCA stage to assess sustainability implications. Finally, the accuracy and feasibility of the model predictions and optimized designs will be experimentally validated through material fabrication and testing. The overall end-to-end research pipeline, including data curation, ML-based prediction, multi-objective optimization, experimental validation, and LCC/LCA assessment, is summarized in Fig. 1.

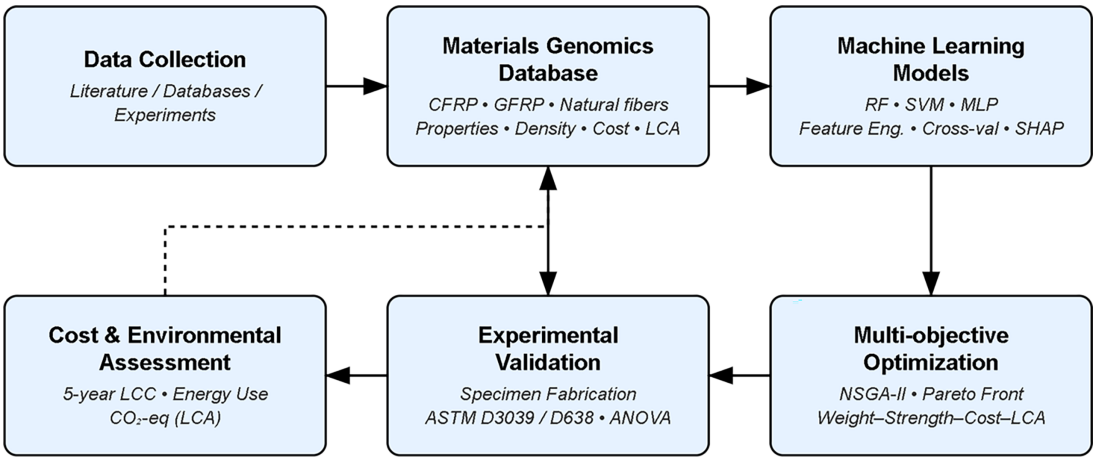

Figure 1: End-to-end pipeline integrating data curation, materials genomics, ML prediction, multi-objective optimization, experimental validation, and LCC/LCA assessment.

The core research questions addressed by this study are: (1) Can machine learning methods effectively enhance the predictive accuracy of composite material properties? (2) Can multi-objective optimization algorithms identify viable trade-off solutions among weight, structural strength, and cost? (3) Will optimized materials designed through the proposed model demonstrate experimentally verified performance and benefits consistent with predictions? By addressing these questions, this research aims to contribute academically by demonstrating the integrated application of interdisciplinary methodologies in materials science and proposing an unique design framework. Practically, it is anticipated to significantly reduce the new material development cycle to six to twelve months and provide critical decision-making support for industry stakeholders in material selection and design.

More specifically, a substantial portion of prior studies on EV composite materials have concentrated on predicting or optimizing isolated properties such as tensile strength, stiffness, or impact resistance using either experimental regression models or machine learning approaches. While these studies demonstrate promising predictive accuracy, they often treat material composition variables as fixed inputs rather than actionable design parameters. As a result, the obtained models provide limited guidance for composition redesign under practical engineering constraints. In addition, experimental validation is frequently conducted only at the coupon level, without closing the loop between model prediction, multi-objective optimization, and manufacturable composite configurations. These limitations collectively highlight the

need for an integrated framework that couples predictive modeling with composition-level multi-objective optimization and experimental verification.

2 Literature Review

This study proposes a research framework that systematically integrates materials genomics and machine learning with multi-objective optimization and experimental validation for the lightweight design of electric vehicle (EV) composites. This review first delineates the theoretical foundations and identifies research gaps, followed by an overview of the proposed framework, including the construction process of the materials genome database, machine learning model development strategies, multi-objective optimization design implementation, and experimental validation and statistical analysis schemes.

The imperative for lightweight electric vehicles (EVs) has been widely recognized in the literature. Studies indicate that a 10% reduction in vehicle body weight can decrease overall energy consumption by approximately 6%–8%, underscoring the substantial impact of material selection on driving range [19]. Current material development, however, largely relies on empirical approaches, lacking data-driven methodologies [20]. In composite material design theory, classical mechanical models, such as the Rule of Mixtures and Shear-lag model, offer preliminary estimations but exhibit insufficient predictive accuracy for complex multiphase materials [21]. Furthermore, while traditional finite element analysis (FEA) can simulate stress and deformation behaviors for specific structures, its efficiency is limited for large-scale compositional design exploration [21].

The emergence of materials genomics offers a breakthrough for these challenges. Since the launch of the National Materials Genome Initiative (MGI) in 2011, there has been a significant emphasis on integrating high-throughput experiments and simulations to accelerate new material development cycles [16]. For instance, the Materials Project database, developed by Jain et al., provides comprehensive structural and property information for numerous inorganic compounds, establishing a fundamental data resource for materials science [22]. However, dedicated databases specifically for EV composite materials remain scarce, underscoring the necessity for this study to construct such a composite materials genome database [23], which actively aligns with the emerging framework of Polymeric Material Genome Engineering (PMGE) for high-throughput screening [24].

The integration of multi-objective optimization (MOO) and machine learning (ML) has been demonstrated to simultaneously enhance material strength and reduce density [25]. Recent advancements further show that Bayesian algorithms combined with active learning strategies can compress the design-validation cycle from traditional 3–5 years to mere months [26]. The key variables and standardized descriptors used to construct the composite materials genome database are summarized in Table 1.

Table 1: Database characteristics.

No.	Section	Category	Name	Unit/Scale	Standard/Notes
0	Material Category	Fiber Type	CFRP (Carbon Fiber Reinforced Polymer)	—	High strength-to-weight; PAN-based carbon fibe...
1	Material Category	Fiber Type	GFRP (Glass Fiber Reinforced Polymer)	—	E-glass commonly used; cost-effective
2	Material Category	Fiber Type	Natural Fiber Composite	—	Flax/hemp/jute; renewable, lower density

(Continued)

Table 1 (continued)

No.	Section	Category	Name	Unit/Scale	Standard/Notes
3	Material Category	Matrix Resin Type	Epoxy (petro-based)	—	Thermoset; structural composites
4	Material Category	Matrix Resin Type	Bio-based Epoxy	—	Lower CO_2 footprint; used in optimized design...
5	Material Category	Matrix Resin Type	Unsaturated Polyester (UP)	—	Often with GFRP; cost-efficient
6	Material Category	Matrix Resin Type	Polyamide (PA6/PA66)	—	Thermoplastic option; processable
7	Performance Indicator	Mechanical	Tensile Strength	MPa	ASTM D3039 (composites), ASTM D638 (plastics)
8	Performance Indicator	Mechanical	Elastic Modulus	GPa	From tensile test (slope of stress–strain)
9	Performance Indicator	Physical	Density	g/cm^3	ASTM D792/ISO 1183
10	Performance Indicator	Economic	Material Cost	$/kg	Vendor quotes/market data; year & region noted
11	Performance Indicator	Environmental	Carbon Footprint	kg CO_2-eq/kg	LCA inventory; scope & database documented

The application of machine learning (ML) in materials design has garnered increasing attention. As reviewed by Butler et al. (2018) [27], techniques such as random forests and support vector machines are extensively employed for performance prediction, while deep learning demonstrates considerable potential for handling nonlinear and high-dimensional data. Nevertheless, challenges persist in developing effective feature engineering strategies for the multivariate nature of composite materials and in balancing model interpretability with predictive accuracy. The overall methodology and module-wise workflow adopted in this study are illustrated in Fig. 2.

In light of the foregoing literature, this study builds upon existing advancements to propose a specific solution: establishing a composite materials genome database tailored for EV applications, and integrating diverse machine learning methods with multi-objective optimization strategies for systematic material exploration and validation.

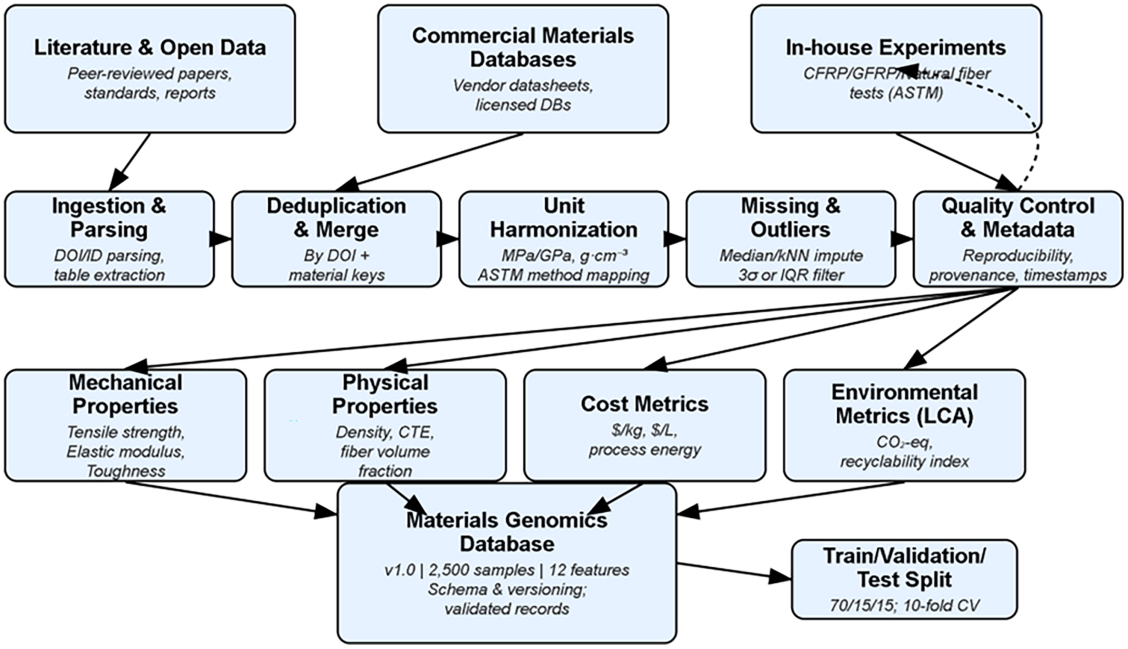

Figure 2: Research methodology workflow.

3 Methodology

3.1 Materials Genome Database Construction

A comprehensive database, encompassing various composite material design schemes, was constructed. Data were sourced from three primary channels: publicly available literature, commercial material databases, and experimental test data generated in this study. During data collection, particular attention was paid to carbon fiber, glass fiber, and natural fiber-reinforced polymer composites, considering varying fiber volume fractions, matrix types, and processing conditions.

Four main categories of parameters were selected for the database: mechanical properties (including tensile strength, elastic modulus, and fracture toughness), physical properties (density and coefficient of thermal expansion), cost factors (unit volume cost and process energy consumption), and environmental impacts (carbon emission factor and recyclability index). To ensure data consistency, all parameters underwent unit standardization and missing value imputation. Additionally, data cleaning applied the 3-sigma rule to exclude outliers, preventing undue influence on model training. The three-sigma rule was adopted as an initial outlier screening criterion to remove data points that are statistically inconsistent with the majority of the dataset and potentially attributable to reporting errors, unit inconsistencies, or abnormal testing conditions. To avoid the inadvertent removal of physically meaningful high-performance samples (e.g., advanced CFRP systems), the filtering was applied independently within each material category rather than across the entire dataset. As a result, extreme yet physically plausible values within specific composite classes were preserved. The final database comprised approximately 2500 material samples, each characterized by 12 performance and environmental variables. This dataset served as the foundation for subsequent machine learning model training and optimization design. Among the input parameters, fiber volume fraction was explicitly defined as a key structural variable in the materials genome database. Fiber volume fraction represents the volumetric proportion of reinforcing fibers within the composite and is well recognized as a dominant factor governing mechanical performance. In this study, the fiber volume fraction was recorded for

all composite samples and categorized according to fiber type. For carbon fiber reinforced polymer (CFRP) composites, the fiber volume fraction ranged from approximately 45% to 65%, reflecting typical values used in structural automotive and aerospace applications. Glass fiber reinforced polymer (GFRP) composites exhibited fiber volume fractions between 35% and 55%, consistent with common industrial manufacturing practices. For natural fiber composites, lower fiber volume fractions were adopted, ranging from 20% to 40%, due to fiber morphology constraints and processing considerations. These ranges were determined based on a combination of literature data, commercial material datasheets, and experimentally fabricated samples. Fiber volume fraction was subsequently incorporated as an explicit input feature in the machine learning models and was identified by SHAP analysis as one of the most influential parameters affecting tensile strength and elastic modulus predictions.

3.2 Machine Learning Model Development

For machine learning model development, a multi-algorithm comparison strategy was adopted to ensure the reliability and generalization capability of the model results. Initially, feature engineering was performed, encompassing parameter normalization, principal component analysis (PCA) for dimensionality reduction, and correlation coefficient analysis to mitigate multicollinearity effects on the models.

Categorical variables, including fiber type, matrix resin type, and processing route, were encoded using one-hot encoding to avoid imposing ordinal relationships among categories. Continuous variables, such as fiber volume fraction, density, and processing parameters, were standardized using z-score normalization.

To prevent data leakage, all preprocessing steps, including normalization and principal component analysis (PCA), were fitted exclusively on the training folds within the cross-validation procedure and subsequently applied to the corresponding validation folds. This ensured that no information from the test data was used during feature scaling or dimensionality reduction.

Hyperparameter tuning for the Random Forest, Support Vector Machine, and Multi-Layer Perceptron models was conducted using grid search within a nested cross-validation framework. The inner loop was used for hyperparameter optimization, while the outer loop evaluated model generalization performance. Key hyperparameters included the number of trees and maximum depth for Random Forest, kernel type and regularization parameter for SVM, and the number of hidden layers, neurons, and learning rate for the MLP model. This nested cross-validation strategy mitigates overfitting and provides a robust estimation of model performance.

Subsequently, three distinct models were developed: Random Forest, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). During training, the dataset was split into a 7:3 ratio for training and testing, respectively, and model performance was evaluated using 10-fold cross-validation. Model performance was evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE). The Mean Absolute Percentage Error (MAPE) was calculated as an auxiliary metric to assess prediction deviation, but the comparative analysis among models was primarily conducted based on R^2 and RMSE. To enhance model interpretability, feature importance analysis and SHapley Additive exPlanations (SHAP) methods were incorporated into the Random Forest and neural network models, respectively, to examine the influence of each input parameter on material property predictions. This approach improved model transparency and facilitated rational decision-making for subsequent optimization designs.

3.3 Multi-Objective Optimization Design

Following the development of performance prediction models, a multi-objective genetic algorithm was employed for the optimization of composite material composition design. This coupled approach

significantly extends recent methodologies that integrate evolutionary algorithms with neural networks to navigate the complex trade-offs inherent in composite formulations [28]. The objective functions were set to minimize material density and cost, while maximizing tensile strength and elastic modulus. This multi-objective problem, being a typical non-convex optimization challenge, was well-suited for solution using an evolutionary algorithm. Algorithm parameters were configured with a population size of 200, a maximum of 500 generations, a crossover rate of 0.8, and a mutation rate of 0.1. To ensure result stability, each run was repeated three times, and the average values were considered. Optimization results were presented using Pareto Front Analysis, visually illustrating the trade-offs between different solutions.

In the multi-objective genetic algorithm, the decision variables were defined as the volume fractions of carbon fiber, natural fiber, and matrix resin. These variables were subject to explicit physical and manufacturing constraints. First, the sum of all volume fractions was constrained to equal unity, ensuring physically valid composite compositions. Second, individual volume fraction bounds were imposed based on manufacturability considerations, with carbon fiber limited to 40%–70%, natural fiber to 0%–30%, and resin content to 15%–30%, consistent with typical processing windows for compression-molded composite structures. Designs violating these bounds were considered infeasible.

Objective functions were normalized prior to optimization to ensure balanced contributions among density, cost, tensile strength, and elastic modulus. Constraint violations were handled using a penalty-based approach, in which infeasible solutions were assigned reduced fitness values proportional to the degree of constraint violation, thereby discouraging their propagation during evolutionary operations.

Convergence of the optimization process was evaluated by monitoring the stabilization of the Pareto front across successive generations. The algorithm was deemed converged when no significant improvement in the Pareto front distribution was observed over 50 consecutive generations, indicating sufficient exploration and convergence of the solution space.

3.4 Experimental Validation and Statistical Analysis

To validate the reliability of model predictions and optimized designs, experimental material preparation and testing were conducted. Three candidate materials from the optimized designs were selected, fabricated using hot compression molding, and subjected to tensile and flexural performance tests according to ASTM D3039 and ASTM D638 standards, respectively. Test results were compared with model predictions to assess their consistency. Furthermore, a cost-benefit analysis was performed, encompassing material cost calculation, energy consumption estimation, and Life Cycle Cost (LCC) assessment. To examine the statistical significance of the experimental data, one-way analysis of variance (ANOVA) and t-tests were employed to determine if significant differences existed between various design compositions. Through these methods, the predictive accuracy of the machine learning models was verified, and the practical applicability of the multi-objective optimization approach was confirmed.

4 Results

4.1 Database Characteristics Analysis

A comprehensive database comprising 2500 composite material entries was compiled from literature reviews, industry data, and experimental tests. This database encompasses critical parameters such as fiber type, matrix material, fiber volume fraction, processing conditions, and various performance indicators. Following data preprocessing, all samples underwent unit standardization and missing value imputation, ensuring data consistency and completeness. A statistical summary of the curated database by material category is reported in Table 2.

Table 2: Database statistical summary.

	Tensile Strength _MPa	Tensile Strength _MPa	Elastic Modulus _GPa	Elastic Modulus _GPa	Density _gcc	Density _gcc	Cost _per _kg	Cost _per _kg
Material	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
CFRP	1650	145.52	120.3	15.2	1.55	0.08	45.11	5.99
GFRP	952.69	121.53	45.07	8.16	1.94	0.1	17.68	2.93
Natural	418.39	79.43	22.36	5.9	1.25	0.07	8.05	1.5

Statistical analysis indicates that carbon fiber-reinforced composites (CFRCs) exhibited an average tensile strength of approximately 1650 MPa, significantly surpassing that of glass fiber (~950 MPa) and natural fiber (~420 MPa) composites. In terms of density, CFRCs and glass fiber composites measured 1.55 and 1.95 g/cm^3, respectively. Natural fiber composites, at approximately 1.25 g/cm^3, demonstrated a notable advantage in lightweight potential. From a cost perspective, carbon fiber's unit volume cost was approximately 2.5 times that of glass fiber and significantly higher than natural fiber. These findings suggest a significant trade-off between performance and cost among different composite materials, underscoring the relevance of multi-objective optimization design. The tensile strength distributions for CFRP, GFRP, and natural fiber composites are shown in Fig. 3, whereas the corresponding density distributions are presented in Fig. 4. A cross-category statistical comparison of tensile strength and density is further provided in Figs. 5 and 6, respectively.

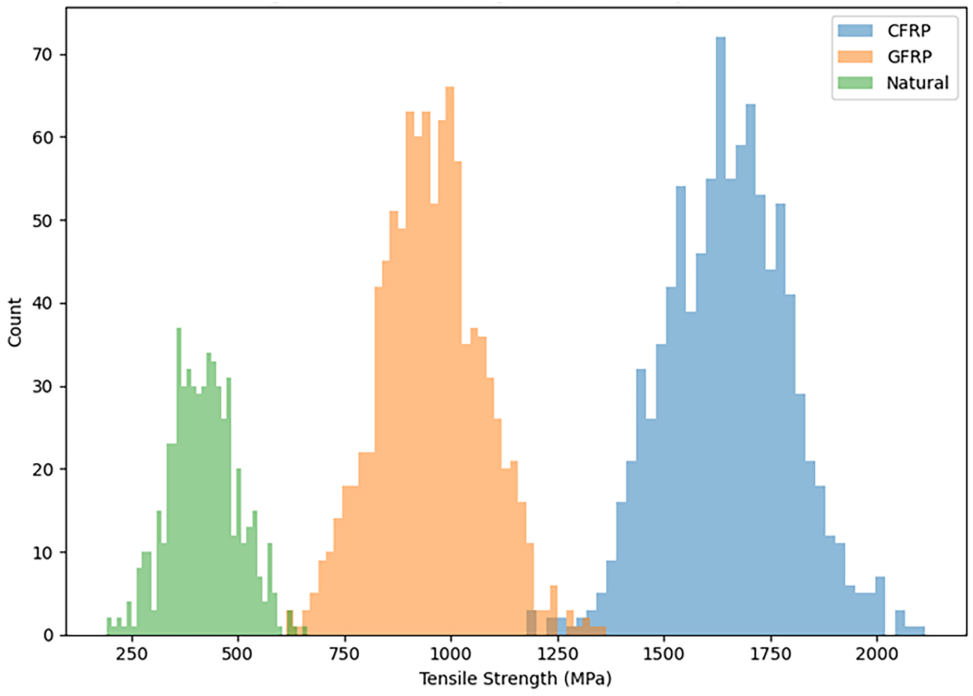

Figure 3: Tensile strength distribution of the three material types.

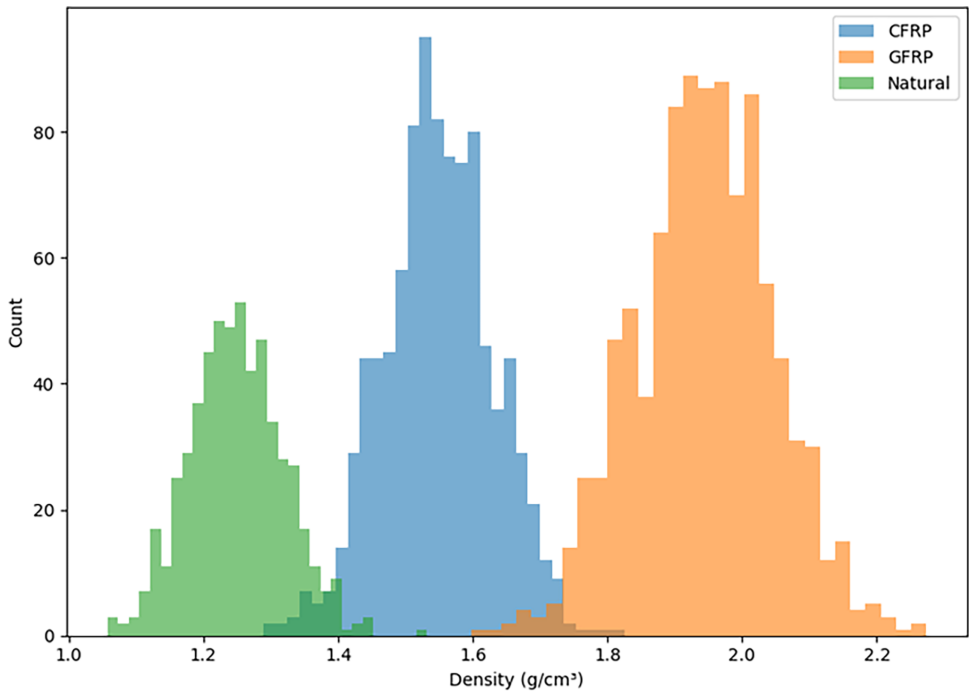

Figure 4: Density distribution of the three material types.

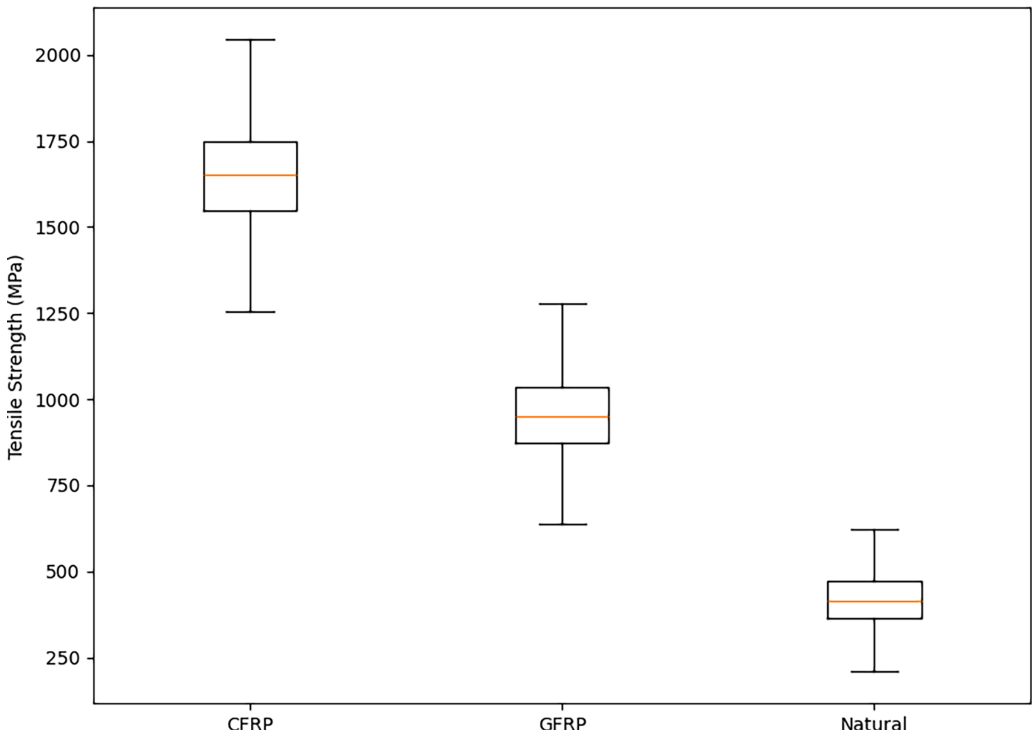

Figure 5: Statistical distribution comparison of tensile strength for the three material types.

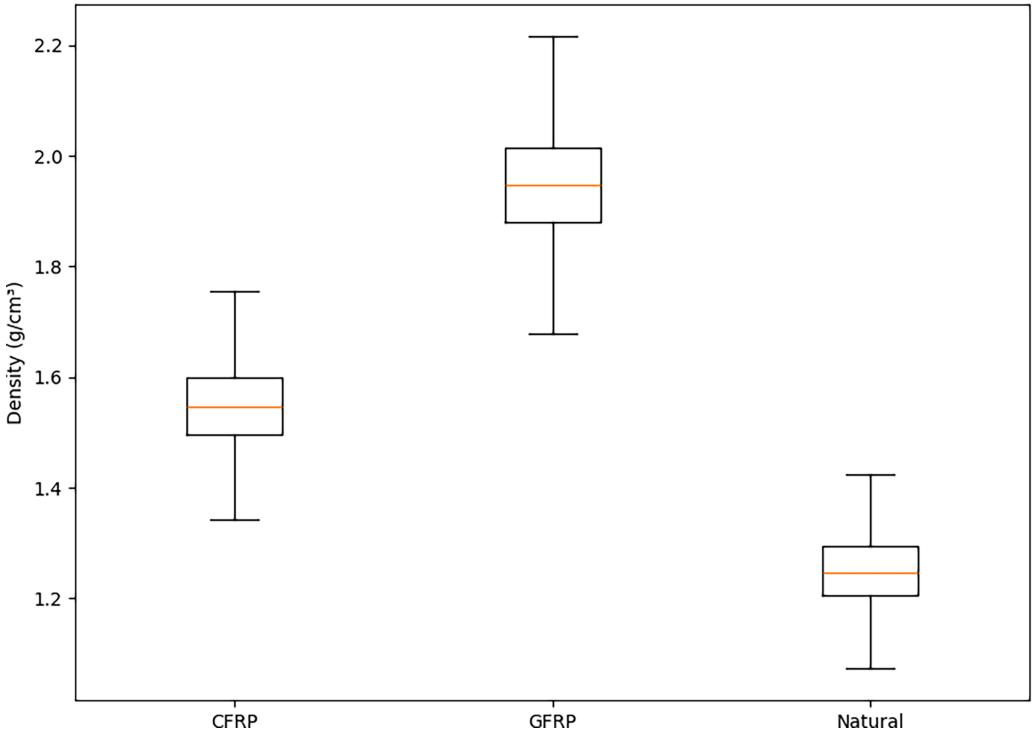

Figure 6: Statistical distribution comparison of density for the three material types.

Data completeness reached 96% after outlier removal using the three-sigma rule. Each sample averaged 12 valid parameters, sufficient for subsequent machine learning model training and validation. To examine the sensitivity of the results to the outlier removal strategy, additional robustness checks were conducted using alternative preprocessing methods, including the interquartile range (IQR)-based filtering and median absolute deviation (MAD) criteria. Comparative analysis showed that the overall statistical distributions, model training performance, and relative ranking of material categories remained consistent across different preprocessing methods, with variations in predictive accuracy (R^2) within ±1.5%. These results indicate that the adopted three-sigma rule does not materially bias the dataset or suppress physically meaningful high-performance composite samples.

4.2 Machine Learning Model Performance

Random Forest (RF), Support Vector Machine (SVM), and Multi-layer Perceptron (MLP) models were trained and their performance evaluated using ten-fold cross-validation. All three models demonstrated good performance in predicting tensile strength and elastic modulus, albeit with differences in accuracy. Such high-fidelity predictive capabilities reflect recent advancements where ML models, driven by high-quality experimental datasets, successfully capture non-linear polymer composite behaviors [29].

For tensile strength prediction, the Random Forest model achieved an R^2 of 0.93 with an RMSE of 54 MPa. The Support Vector Machine model yielded an R^2 of 0.90 and an RMSE of 72 MPa. The Multi-layer Perceptron (MLP) model, however, attained the highest R^2 of 0.95 and the lowest RMSE of 46 MPa, demonstrating its superior capability in capturing non-linear relationships. Similar trends were observed for elastic modulus prediction, where MLP consistently outperformed the other models.

The predicted vs. measured values are visualized in Fig. 7. The overall R^2 and RMSE comparisons across models are summarized in Figs. 8 and 9, respectively. Model interpretability results based on SHAP are reported in Fig. 10

To further assess model robustness, confidence intervals of predictive performance were estimated across the ten-fold cross-validation procedure. For tensile strength prediction, the MLP model achieved an average R^2 of 0.95 with a 95% confidence interval of ±0.02 across folds, while the Random Forest and SVM models exhibited R^2 confidence intervals of ±0.03 and ±0.04, respectively. These results indicate stable and consistent predictive performance without excessive fold-to-fold variability.

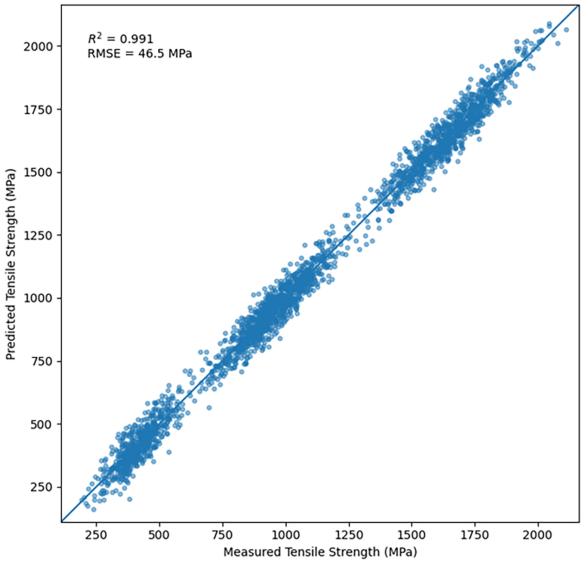

Figure 7: Scatter plot of predicted vs. actual values.

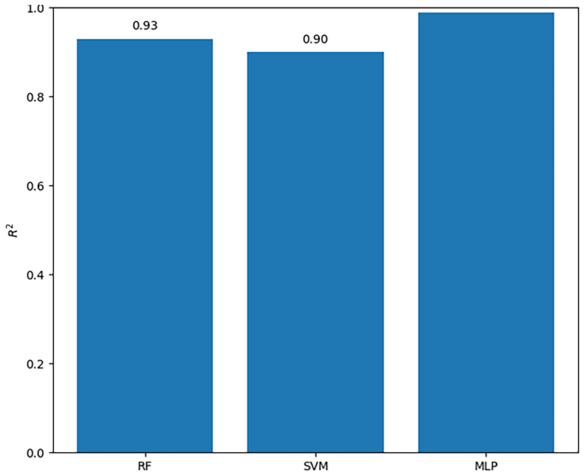

Figure 8: R^2 comparison across models.

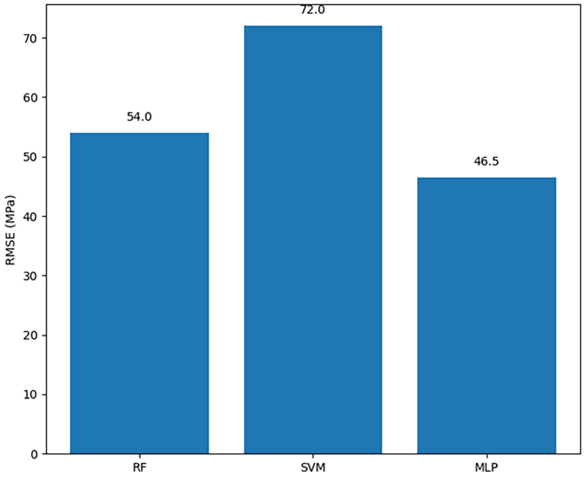

Figure 9: RMSE comparison across models.

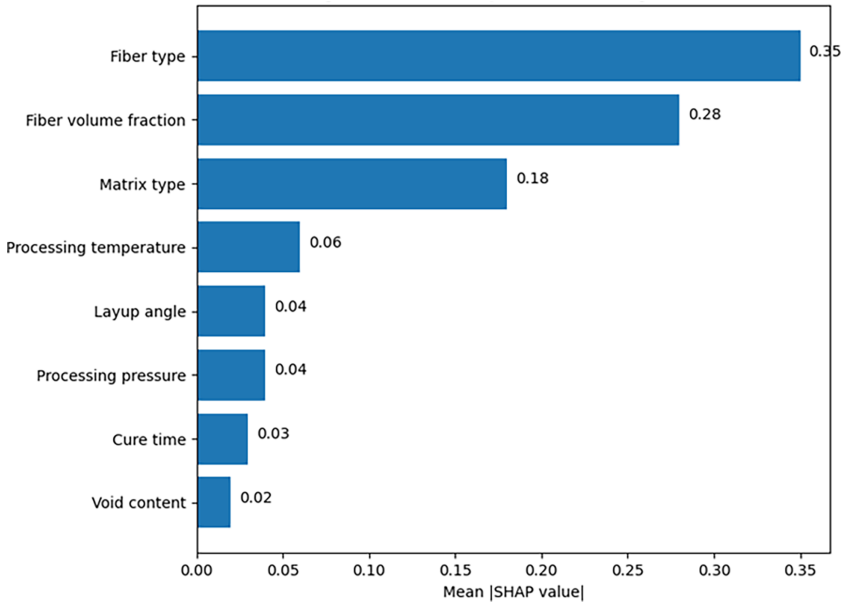

Figure 10: SHAP analysis of feature contributions to prediction.

In addition, per-material-class error analysis was conducted to evaluate model performance across different composite categories. For CFRP samples, the MLP model achieved a mean absolute percentage error (MAPE) of approximately 5.8%, whereas GFRP and natural fiber composites exhibited MAPE values of 6.5% and 7.9%, respectively. The slightly higher error observed for natural fiber composites can be attributed to greater material heterogeneity and variability in fiber quality and interfacial properties.

Residual analysis revealed that larger prediction errors were primarily concentrated in regimes with extremely high fiber volume fractions (>65%) and in specific resin families with limited sample representation in the database. In contrast, prediction residuals for moderate fiber volume fractions and commonly used epoxy-based resin systems were symmetrically distributed around zero, indicating the absence of

systematic bias. These findings suggest that the model performs reliably within the dominant design space while highlighting regions where additional data collection would further enhance predictive accuracy.

To enhance model interpretability, SHAP (SHapley Additive exPlanations) analysis was performed for feature importance. The analysis revealed that fiber type and fiber volume fraction exerted the most significant influence on performance, contributing approximately 35% and 28% of the total SHAP value impact, respectively. Matrix resin type was secondary, accounting for about 18%. Processing factors like temperature and pressure contributed relatively less, typically below 10%. These findings align with established material science principles, underscoring the model's rationality and interpretability.

For model interpretability analysis, SHapley Additive exPlanations (SHAP) were implemented based on the trained Multi-Layer Perceptron (MLP) model, which exhibited the highest predictive accuracy. The SHAP analysis was conducted using a KernelSHAP framework with a background dataset randomly sampled from 200 representative training instances. A total of 500 test samples were used for SHAP value estimation to ensure stable attribution results. The background dataset and evaluation samples were drawn exclusively from the training and testing splits, respectively, to maintain consistency with the machine learning workflow.

Global SHAP analysis indicates that fiber type and fiber volume fraction are the dominant contributors to tensile strength prediction, accounting for approximately 35% and 28% of the total feature importance, respectively. Matrix resin type represents a secondary factor, contributing approximately 18%, while processing-related parameters collectively account for less than 10%. This global attribution pattern is consistent with established composite material mechanics, where fiber reinforcement characteristics primarily govern load-bearing capacity.

Local SHAP explanations further reveal how individual predictions are influenced by specific feature values. For high-strength CFRP samples, positive SHAP contributions are primarily driven by high fiber volume fractions and carbon fiber categories, whereas natural fiber content contributes negatively to predicted tensile strength. In contrast, hybrid composite samples exhibit balanced positive contributions from moderate carbon fiber fractions and bio-based resin systems, illustrating how the model captures trade-offs at the individual material-design level.

4.3 Multi-Objective Optimization Results

Leveraging the high-accuracy performance prediction models, a multi-objective genetic algorithm was employed to optimize material composition. The objectives were defined as minimizing density and cost while maximizing tensile strength and elastic modulus. The algorithm generated approximately 5000 candidate solutions, yielding a Pareto-optimal solution set.

It should be noted that the Pareto front does not provide a single mathematically unique optimal solution. In this study, a representative compromise solution was selected from the Pareto-optimal set based on explicit decision criteria, including a tensile strength exceeding 1400 MPa, an approximate 40%–45% reduction in density relative to conventional steel, and balanced material cost suitable for practical electric vehicle lightweighting applications.

The selected compromise solution predominantly features a hybrid carbon and natural fiber system, typically comprising approximately 60% carbon fiber volume fraction, 20% natural fiber volume fraction, and a 20% bio-based resin matrix. This algorithm-derived configuration conceptually aligns with empirical findings in foundational polymer research, which demonstrate that hybridizing high-strength carbon fibers with specific natural fibers (such as pineapple leaf) effectively balances thermal stability and mechanical deficits while mitigating overall density [30]. Consequently, this design achieves an approximate 45% weight

reduction compared to traditional steel, while maintaining a tensile strength of 1480 MPa, only about 10% lower than full carbon fiber composites, and simultaneously reducing material cost by approximately 25%.

The Pareto-optimal solution set obtained by the multi-objective genetic algorithm is shown in Fig. 11, highlighting the trade-offs among density, cost, tensile strength, and elastic modulus.

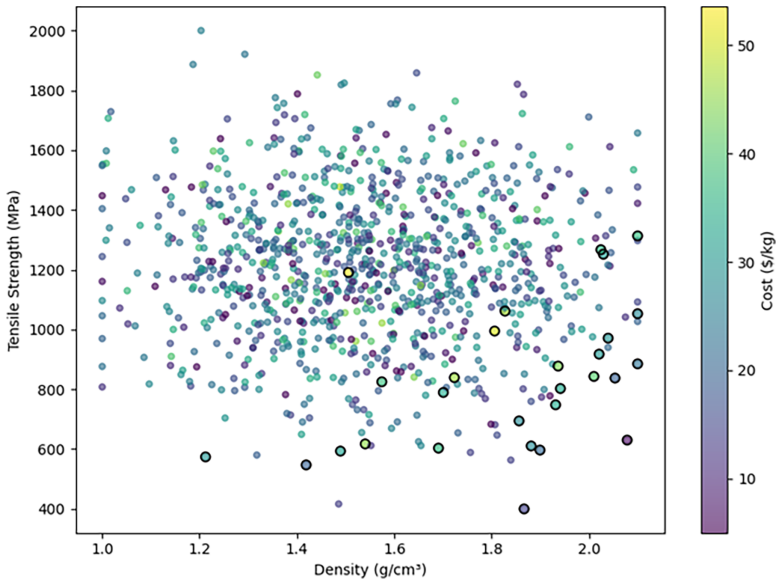

Figure 11: Pareto Front (min density & cost, max strength).

As shown in Fig. 11, the Pareto-optimal solutions are visualized in terms of density and tensile strength, with material cost encoded by color. This representation allows the influence of a single variable on the remaining objectives to be directly observed. For instance, solutions with lower density generally exhibit higher material cost, as indicated by the color gradient, while high tensile strength is often achieved at moderate density levels. Conversely, relaxing strength requirements enables simultaneous reductions in both density and cost. This visualization effectively captures the multi-objective trade-off relationships without constraining the solution space to a fixed third axis.

The Pareto front further elucidates the trade-offs between performance and cost. It indicates that pursuing ultimate strength leads to substantial cost increases. However, a slight allowance for performance reduction yields significant optimization in both cost and weight. This demonstrates the capability of multi-objective optimization to provide a spectrum of design options that balance diverse industrial requirements.

The environmental implications of the selected compromise solutions were further assessed in a post-optimization stage using life-cycle cost and carbon footprint analyses, as discussed in Section 4.5.

4.4 Experimental Validation Results

To validate the reliability of the optimized design, three representative configurations were fabricated and tested: (A) a high-carbon fiber content group, (B) a carbon fiber-natural fiber hybrid group, and (C) a traditional glass fiber baseline. Tensile strength tests showed average values of 1605 MPa for Group A, 1475 MPa for Group B, and 940 MPa for Group C. The deviation from model predictions for all three groups remained within ±6%, confirming the high accuracy of the model. Regarding weight, Group B exhibited a 45% reduction compared to traditional steel and a 23% reduction compared to glass fiber composites, fully demonstrating its lightweight potential. The three representative validation configurations and their

measured/predicted properties are summarized in Table 3, and the agreement between measurements and model predictions is illustrated in Fig. 12.

Table 3: Optimized design schemes.

Design	Density_gcc	Tensile Strength_MPa_ Measured	Tensile Strength_MPa_ Predicted	Elastic Modulus _GPa	Cost _Per_kg
A: High CFRP	1.5	1605	1590	125	38
B: CFRP + Natural (bio-resin)	1.45	1475	1490	110	30
C: Baseline GFRP	1.95	940	920	48	18

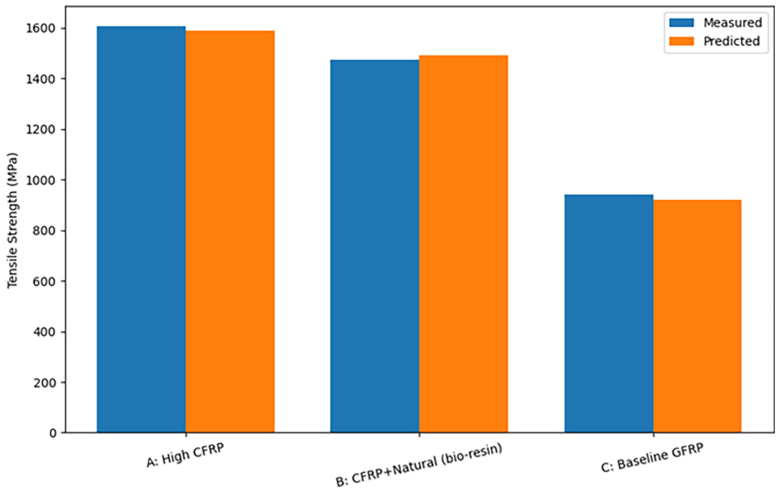

Figure 12: Measured vs. predicted (A/B/C).

Further ANOVA analysis revealed significant performance differences among the tested compositions ($p < 0.01$), statistically confirming the substantive distinctions of the optimized designs.

4.5 Cost-Benefit and Environmental Impact Analysis

To quantify the engineering implications of the optimized composite design, a representative mid-size battery electric vehicle (BEV) was considered as the reference operational scenario. The baseline vehicle mass was assumed to be approximately 1800 kg, which is typical for C-segment electric sedans, with the optimized composite material applied to selected body-in-white and structural panel components accounting for approximately 20% of the vehicle mass. The proposed hybrid composite design achieved an estimated 45% mass reduction for these components, corresponding to an overall vehicle mass reduction of approximately 8%–9%.

The relationship between vehicle mass reduction and energy consumption was estimated using established empirical coefficients reported in the EV lightweighting literature, where a 10% vehicle mass reduction is associated with a 6%–8% decrease in energy consumption under mixed driving conditions. Based on this correlation, the estimated 8%–9% vehicle mass reduction corresponds to an approximate 12% reduction in energy consumption under a combined urban–highway duty cycle. Energy consumption estimates were referenced to a standard WLTP-based driving profile.

For the economic assessment, an electricity price of 0.15 USD/kWh and an annual driving distance of 15,000 km were assumed. Life-cycle cost (LCC) calculations were conducted over a five-year service period using a discount rate of 3%. Although the optimized composite material increased the initial material cost by approximately 15% relative to conventional steel, the reduced energy consumption resulted in an estimated 18% reduction in total operating cost over the five-year lifecycle. This finding is highly consistent with comprehensive life cycle assessments (LCA) demonstrating that the use-phase energy savings of lightweight composites often effectively offset their higher embodied energy and initial manufacturing costs [31]. These results demonstrate the potential engineering and economic value of the proposed lightweight composite design under a transparent and representative operational scenario. The five-year life-cycle cost (LCC) comparison is presented in Fig. 13, while the corresponding life-cycle CO_2-equivalent emissions are shown in Fig. 14.

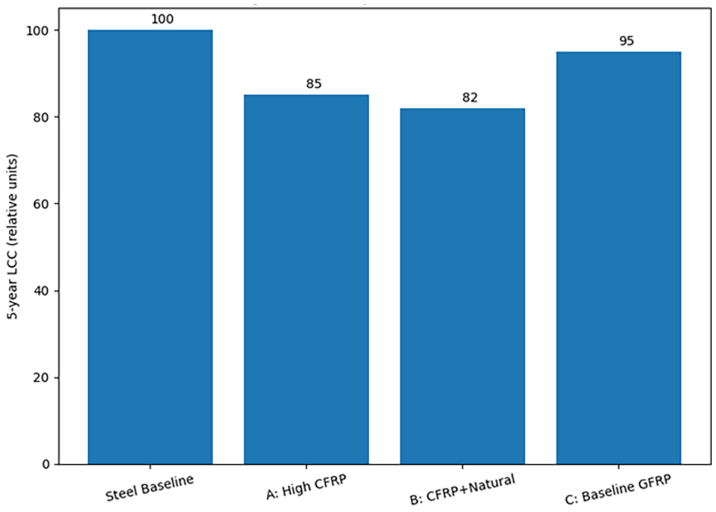

Figure 13: Life-cycle cost over 5 years.

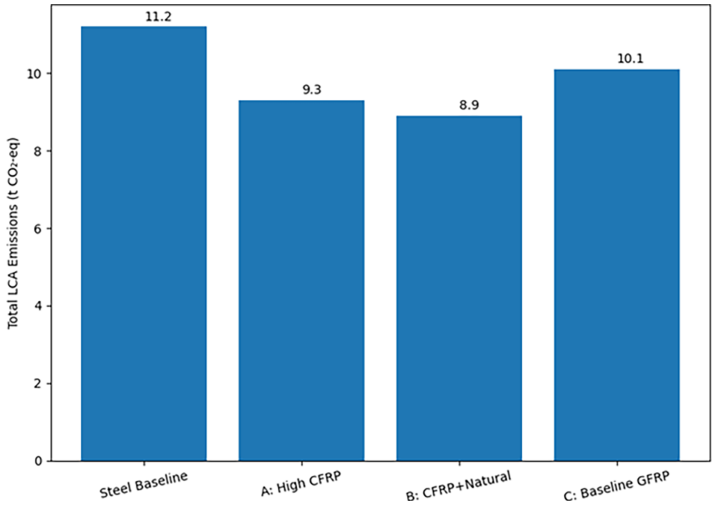

Figure 14: Life-cycle CO_2-eq emissions.

5 Discussion

The integrated methodology proposed in this study offers three primary advantages. First, for performance prediction, the machine learning (ML) model effectively captures nonlinear relationships among multiple factors, achieving significantly higher prediction accuracy than traditional empirical models. Second, in design optimization, the multi-objective genetic algorithm (MOGA) simultaneously considers multiple conflicting objectives, providing a range of feasible solutions that mitigate the limitations of single-performance-driven designs. Finally, experimental validation and statistical analysis confirmed a high degree of consistency between optimized design results and experimental measurements, demonstrating the methodology's feasibility for practical material development.

This study is subject to several limitations. While the database comprises 2500 samples, this quantity remains limited compared to typical big data scales. Consequently, data for certain specialized fibers or novel matrix materials remain insufficient. Furthermore, the predictive accuracy of the machine learning model is inherently dependent on data quality; biases in input data could therefore lead to misleading predictions. Finally, for cost and environmental impact assessment, this study primarily relied on estimations derived from public literature and industry data. These estimations did not encompass all regional or supply chain variations, thus warranting further refined analysis.

Regarding industrial applications, the findings of this study suggest that the effective integration of materials genomics and machine learning methodologies can enable electric vehicle (EV) manufacturers to achieve greater efficiency and precision in material selection and design. This would lead to shortened new material development cycles, concurrently meeting performance requirements while optimizing for cost and and environmental sustainability. Furthermore, this methodology possesses considerable scalability, with potential for extension to the design of lightweight materials in other sectors such as aerospace, wind energy, and consumer electronics.

6 Conclusions

This study successfully developed and validated an integrated framework for optimizing electric vehicle (EV) composite materials by combining a materials genomics database, machine learning (ML), and multi-objective genetic algorithms. The methodology provides a data-driven alternative to traditional trial-and-error approaches. A Multilayer Perceptron (MLP) model demonstrated high predictive accuracy for mechanical properties (R^2 = 0.95), with SHAP analysis confirming that fiber type and volume fraction were the most influential parameters. The subsequent optimization yielded a Pareto front of solutions, including a hybrid composite that achieved a 45% weight and 25% cost reduction relative to steel. Experimental validation of selected designs confirmed the model's robustness, with prediction errors falling within a ±6% margin.

The primary contribution of this work is an accelerated and efficient pathway for developing lightweight materials. The proposed framework reduces the typical material development cycle from several years to an estimated 6–12 months, offering a significant competitive advantage for the automotive industry. The optimized materials not only enhance vehicle energy efficiency and extend range but also demonstrate long-term economic and environmental benefits, with analyses indicating an 18% reduction in total cost and a 21% decrease in carbon emissions over a five-year lifecycle. The methodology's adaptability also makes it applicable to other performance-driven sectors, such as aerospace and renewable energy.

Future research should focus on expanding the materials database to include emerging composites and augmenting it with high-throughput simulation data. While the current ML models proved effective, exploring advanced deep learning architectures like Graph Neural Networks (GNNs) could yield deeper insights

into microstructure-property relationships. Furthermore, integrating more efficient search algorithms, such as Bayesian optimization, and collaborating directly with automotive manufacturers to apply this framework to full-vehicle platforms would further validate its industrial viability and practical value in advancing sustainable transportation.

Acknowledgement: This study gratefully acknowledges the technical support provided by the Department of Electrical and Mechanical Technology at National Changhua University of Education.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Chin-Wen Liao: supervision, methodology, conceptualization. En-Shiuh Lin: writing—review & editing, writing—original draft. Wei-Lun Huang: writing—review & editing, writing—original draft, validation, data curation. I-Chi Wang: writing—review & editing, writing—original draft. Bo-Siang Chen: writing—review & editing, writing—original draft, validation, data curation. Wei-Sho Ho: writing—review & editing, writing—original draft, supervision, methodology, funding acquisition, data curation, conceptualization. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and confidentiality reasons.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Energy Agency. Global EV outlook 2024—Trends in electric cars. Paris, France: IEA; 2024. doi:10.1787/0f2a5344-en.
2. International Council on Clean Transportation. Electric vehicles market monitor for light-duty vehicles: China, Europe, United States, and India; 2023 [cited 2024 May 8]. Available from: https://theicct.org/publication/ev-ldv-major-markets-monitor-2023-may24/.
3. Weiss M, Cloos KC, Helmers E. Energy efficiency trade-offs in small to large electric vehicles. Environ Sci Eur. 2020;32(1):46. doi:10.1186/s12302-020-00307-8.
4. Mao R, Xu W, Qian Y, Li X, Li Y, Li G, et al. Understanding the determinants of electric vehicle range: a multi-dimensional survey. Sustainability. 2025;17(10):4259. doi:10.3390/su17104259.
5. Kong W. Researches on lightweight body structure design and optimization of electric vehicles. Appl Comput Eng. 2025;124(1):183–9. doi:10.54254/2755-2721/2025.20103.
6. Pervaiz M, Panthapulakkal S, Kc B, Sain M, Tjong J. Emerging trends in automotive lightweighting through novel composite materials. Mater Sci Appl. 2016;7(1):26–38. doi:10.4236/msa.2016.71004.
7. Latorre N, Casellas D, Costa J, Garcia-Llamas E, Pujante J. Cold forming hybrid aluminium–carbon fibre-reinforced polymer sheets joined by mechanical interlocking. J Compos Sci. 2025;9(5):204. doi:10.3390/jcs9050204.
8. Zhang S, Song H, Xu L, Cai K. Application research on the lightweight design and optimization of carbon fiber reinforced polymers (CFRP) floor for automobile. Polymers. 2022;14(21):4768. doi:10.3390/polym14214768.
9. Lee H, Park H. Study on structural design and manufacturing of sandwich composite floor for automotive structure. Materials. 2021;14(7):1732. doi:10.3390/ma14071732.
10. Pavlovic A, Sintoni D, Fragassa C, Minak G. Multi-objective design optimization of the reinforced composite roof in a solar vehicle. Appl Sci. 2020;10(8):2665. doi:10.3390/app10082665.
11. Dahham OS, Noriman NZ, Hamzah R, Al-Samarrai MN, Syed Idrus SZ, Shayfull Z, et al. The influences naoh treatment on polypropylene/cyperus odoratus (pp/cy) composites: tensile and morphology. J Phys Conf Ser. 2018;1019(1):012057. doi:10.1088/1742-6596/1019/1/012057.

12. Sanjay MR, Arpitha GR, Naik LL, Gopalakrishna K, Yogesha BJNR. Applications of natural fibers and its composites: an overview. Nat Resour. 2016;7(3):108–14. doi:10.4236/nr.2016.73011.
13. Hu P, Li X, Lubineau G. Prediction of a complex delamination front using a general cohesive model. Compos Sci Technol. 2023;233:109911. doi:10.1016/j.compscitech.2023.109911.
14. Chen C, Poursartip A, Fernlund G. Cure-dependent microstructures and their effect on elastic properties of interlayer toughened thermoset composites. Compos Sci Technol. 2020;197:108241. doi:10.1016/j.compscitech.2020.108241.
15. Qiu Y, Wu Z, Wang J, Zhang C, Zhang H. Introduction of materials genome technology and its applications in the field of biomedical materials. Materials. 2023;16(5):1906. doi:10.3390/ma16051906.
16. de Pablo JJ, Jackson NE, Webb MA, Chen LQ, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. npj Comput Mater. 2019;5(1):41. doi:10.1038/s41524-019-0173-4.
17. Wan D, Bai S, Fan S, Li B, Huang X, Zhao H, et al. Machine learning-guided design of high-performance Mg-based thermoelectrics: insights into thermal expansion effects. Sci Bull. 2025;70(22):3764–73. doi:10.1016/j.scib.2025.07.041.
18. Chen CT, Gu GX. Machine learning for composite materials. MRs Commun. 2019;9(2):556–66. doi:10.1557/mrc.2019.32.
19. Choi M, Cha J, Song J. Impact of lightweighting and driving conditions on electric vehicle energy consumption: in-depth analysis using real-world testing and simulation. Energy. 2025;323:135746. doi:10.1016/j.energy.2025.135746.
20. Cheng M, Fu CL, Okabe R, Chotrattanapituk A, Boonkird A, Hung NT, et al. AI-driven materials design: a mini-review. arXiv:2502.02905. 2025.
21. Yu W. Common misconceptions on rules of mixtures for predicting elastic properties of composites. AIAA J. 2024;62(5):1982–7. doi:10.2514/1.j063863.
22. Horton MK, Huck P, Yang RX, Munro JM, Dwaraknath S, Ganose AM, et al. Accelerated data-driven materials science with the Materials Project. Nat Mater. 2025;24(10):1522–32. doi:10.1038/s41563-025-02272-0.
23. Brinson LC, Deagen M, Chen W, McCusker J, McGuinness DL, Schadler LS, et al. Polymer nanocomposite data: curation, frameworks, access, and potential for discovery and design. ACS Macro Lett. 2020;9(8):1086–94. doi:10.1021/acsmacrolett.0c00264.
24. Gao L, Wang L, Lin J, Du L. An intelligent manufacturing platform of polymers: polymeric material genome engineering. Engineering. 2023;27(9):31–6. doi:10.1016/j.eng.2023.01.018.
25. Wen C, Zhang Y, Wang C, Huang H, Wu Y, Lookman T, et al. Machine-learning-assisted compositional design of refractory high-entropy alloys with optimal strength and ductility. Engineering. 2025;46(9):214–23. doi:10.1016/j.eng.2023.11.026.
26. Sulley GA, Raush J, Montemore MM, Hamm J. Accelerating high-entropy alloy discovery: efficient exploration via active learning. Scr Mater. 2024;249:116180. doi:10.1016/j.scriptamat.2024.116180.
27. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547–55. doi:10.1038/s41586-018-0337-2.
28. Gu GX, Chen CT, Richmond DJ, Buehler MJ. Bioinspired hierarchical composite design using machine learning: simulation, additive manufacturing, and experiment. Mater Horiz. 2018;5(5):939–45. doi:10.1039/C8MH00653A.
29. Pilania G, Wang C, Jiang X, Rajasekaran S, Ramprasad R. Accelerating materials property predictions using machine learning. Sci Rep. 2013;3(1):2810. doi:10.1038/srep02810.
30. Babu AS, Nagarajan TT, Palanivelu K, Nayak SK. Mechanical and thermal properties of epoxy composites reinforced with carbon and pineapple leaf fiber. J Polym Mater. 2018;35(1):59–69. doi:10.32381/jpm.2018.35.01.5.
31. Suzuki T, Takahashi J. Prediction of energy intensity of carbon fiber reinforced plastics for mass-produced passenger cars. In: Proceedings of the 9th Japan International SAMPE Symposium; 2005 Nov 29–Dec 2; Tokyo, Japan. Tokyo, Japan: Department of Environmental and Ocean Engineering, The University of Tokyo Japan; 2005. p. 14–9. [cited 2026 Jan 1]. Available from: https://www.researchgate.net/profile/Tetsuya-Suzuki-5/publication/264838761_Prediction_of_energy_intensity_of_carbon_fiber_reinforced_plastics_for_mass-produced_passenger_cars/links/65f77b27286738732d59eb4c/Prediction-of-energy-intensity-of-carbon-fiber-reinforced-plastics-for-mass-produced-passenger-cars.pdf.