Management and Production Engineering Review
Volume 17 • Number 2 • June 2026 • pp. 1–11
DOI: 10.24425/mper.2026.1328

Machine Learning Models for Strength Prediction of Carbon Fiber Reinforced Polymers

Khrystyna BERLADIR[1,2], Oleksandr SOKOLOV[3,4], Michal HATALA[2], Yuliia DENYSENKO[4,5], Vitalii IVANOV[2,4,6]

[1] *Sumy State University, Department of Applied Materials Science and Technology of Structural Materials, Ukraine*
[2] *Technical University of Košice, Department of Production Process Monitoring, Slovakia*
[3] *Technical University of Košice, Department of Industrial Informatics and Applied Mathematics, Slovakia*
[4] *Sumy State University, Department of Manufacturing Engineering, Machines and Tools, Ukraine*
[5] *Poznan University of Technology, Department of Additive Manufacturing and Virtual Reality, Poland*
[6] *WSB University, Faculty of Applied Sciences, Poland*

Received: 18 November 2025
Accepted: 10 April 2026

Abstract
This paper applied and benchmarked machine learning regression algorithms for predicting the tensile strength of carbon fiber-reinforced plastics. A comprehensive statistical analysis revealed significant linear correlations between the input parameters (ratio of matrix and filler concentration, composite density) and the target variable (tensile strength). The dataset consisted of 150 experimental data points. However, nonlinear models significantly outperformed linear models in terms of prediction accuracy. The highest accuracy on an independent test dataset was demonstrated by the neural network model ($R^2 = 0.996351$), as well as the Gaussian process regression models ($R^2 = 0.996246$) and the ensemble method ($R^2 = 0.996096$). Classical linear regression yielded a significantly lower result ($R^2 = 0.89498$), confirming that the overall dependence is a significant, nonlinear, multidimensional function. The key result is the detection of nonlinear behavior in the «polymer−fiber» system: the tensile strength reaches a maximum in the optimal reinforcement phase (10−20 wt.%) and then decreases sharply due to fiber agglomeration and porosity. The success of nonlinear ML models is based on their ability to effectively model these complex, hidden dependencies, paving the way for accelerated materials design.

Keywords
CFRPs, Composites, Machine Learning, Tensile Strength, Process Innovation, Industrial Growth.

Introduction

Carbon fiber reinforced polymers (CFRPs) are composite materials. They combine the lightness of a polymer matrix with the high strength of carbon fibers (Ozkan et al., 2020). These materials belong to the class of multicomponent structural materials. More specifically, they are two- component systems: the polymer acts as the binding phase, while the fibers serve as the reinforcing phase. Such materials are heterogeneous structures. They consist of two different phases with distinct physical and mechanical properties, structures, and functions. Their final performance results from the complex interaction between the phases, not just the sum of their individual properties (polymer matrix and carbon fibers) (Hoang et al., 2026). The polymer matrix plays a crucial role, ensuring uniform load distribution between the fibers, fixing them in position, and providing protection from environmental influences. The matrix choice defines overall performance, including mechanical properties, thermal resistance, aging resistance, chemical resistance, and processability (Wang et al., 2025). Carbon fibers have a high modulus of elasticity, high strength, and low density. The polymer matrix has much lower mechanical characteristics, but provides elastic deformation and energy absorption. The main factor that determines the mechanical properties of these composites is fiber orientation within the matrix (Sayam et al., 2022). The

Corresponding author: *Khrystyna Berladir – Sumy State University 116, Kharkivska St., Sumy 40007, Ukraine, phone: +380 66-380-5099, e-mail: kr.berladir@pmtkm.sumdu.edu.ua*

combination of high specific strength, stiffness, and low mass drives the wide use of CFRPs in aerospace, automotive, civil engineering, sports equipment, and medical technologies (Fig. 1). These materials also have increased fatigue strength, corrosion resistance, and allow the formation of products with complex geometries (Xu et al., 2025).

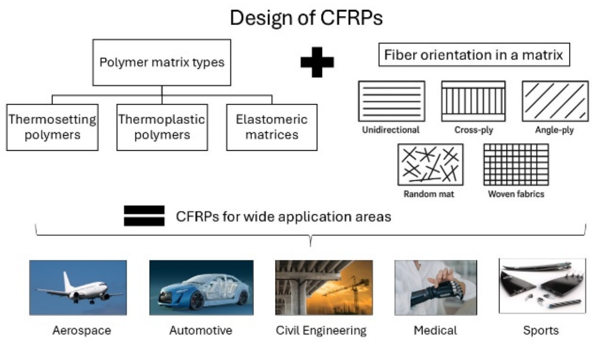


Fig. 1. Design concepts and applications of CFRPs

However, their use is accompanied by several technological challenges, including the complexity of modeling the material's behavior, its anisotropies of properties, and the high requirements for manufacturing technology and quality control.

Literature review

Predicting the performance properties of multicomponent materials is a pressing and challenging problem in modern materials science (Monticeli et al., 2024). Composite materials have a unique combination of high strength and low weight. However, their performance depends on a complex interaction of many factors. The properties of the final product are shaped by the characteristics of the individual components, such as fiber type and polymer matrix. They are also influenced by architectural factors, such as layer orientation and volume fraction. The quality of adhesion at the phase interface and production process parameters (such as temperature and pressure) also play important roles. This multifactorial nature, enhanced by the material's anisotropic and nonlinear behavior, makes accurate property prediction extremely challenging (Liang et al., 2025).

Traditional approaches to CFRP design and optimization face significant limitations. The experimental method of fabricating and physically testing specimens is considered the gold standard for this approach (Budnik et al., 2015; Berladir et al., 2016). However, it is costly. It is also time-consuming and destructive. The numerous possible combinations of input parameters make full-scale experimental optimization economically unfeasible and logistically impractical.

Numerical modeling with the finite element method (FEM) is an alternative approach, but it has limitations. FEM enables the virtual analysis of structural behavior and composite modeling at the micro- and mesoscales; however, it is computationally intensive, often requiring days or weeks (Povstyanoy et al., 2024). Modeling complex, nonlinear failures (such as delamination and crack propagation), or defects (like pores and voids), remains challenging and still requires costly experimental verification (Povstyanoy et al., 2021).

These limitations of traditional methods make machine learning (ML) tools essential. Unlike analytical models, ML algorithms such as neural networks, random forests, and gradient boosting can handle high-dimensional data efficiently. They reveal complex, hidden, nonlinear relationships between inputs and outputs (Antosz et al., 2024). When trained on limited experimental or simulation data, these models can predict outcomes in seconds (Hrehova et al., 2025).

A literature search was conducted in the Scopus database using the keywords «machine learning» and «carbon fiber reinforced polymer». This search yielded 256 relevant publications for preliminary review. After applying inclusion and exclusion criteria, only 10 publications were selected as most relevant to CFRPs systems for detailed analysis (Table 1). Inclusions: (1) peer-reviewed articles, (2) ML for predicting mechanical properties, (3) CFRP composites, (4) English language, (5) 2020–2026; exclusions: review articles, non-materials, purely analytical models without ML, countries – Russia, Belarus, Iran.

Analysis of recent studies demonstrated significant progress and high efficiency in applying ML methods to predict the mechanical properties of CFRPs. Scientific research examined the relationship between the nature of data, sample size, and model accuracy (Kurita et al., 2022). There is also a clear transition from traditional statistical approaches to nonlinear and ensemble algorithms. Most analyzed works report high forecasting accuracy, with the coefficient of determination (R^2) consistently above 0.90. This indicates the reliability of the developed models. Even on relatively small experimental data sets ($N = 62$), simpler yet regularized models, such as Ridge Regression, can provide high accuracy ($R^2 = 0.966$) (Alsheghri et al., 2025).

Similarly, ensemble methods applied to a sample of $N = 162$ also yield a result of $R^2 = 0.9699$ (Zhang et al., 2020). The trend suggests the highest accuracy rates are achieved when larger datasets are used with more complex, nonlinear models. In particular, the highest accuracy in the review ($R^2 = 0.9957$) was

Table 1
Literature review of applying ML for the prediction of properties of CFRPs

Predicted property	**Dataset nature**	**Dataset size**	**ML**	R^2	**Source**
Flexural strength	Experimental data	62	Ridge Regression, Random Forest, SVR	0.966	Alsheghri et al., 2025
Flexural modulus		62		0.871	
Mode-II energy release rate		42		0.903	
Young's modulus, Ultimate tensile strength	Experimental data	320	k-Means	High Accuracy	Kurita et al., 2022
Nonlinear frequency ratio	Literature data	226	XGBoost	0.960	Hou et al., 2023
Mechanical properties	FEM	429	Support vector machine, Random Forest, ANN	0.94–0.99	Zhao et al., 2024
Tensile strength	Literature data	300	Decision tree, SVR, Gaussian process regression, ensemble method	0.92–0.99	Mahajan et al., 2022
Ultimate tensile strength	Literature data	74	ANN	0.9957	Sharan & Mitra, 2022
Tensile strength, bending strength	FEM	1296	XGBoost, decision tree	0.98	Hu et al., 2024
Axial compressive capacity	Literature and experimental data	379	XGBoost	0.985	Ma et al., 2022
Axial compression energy absorption characteristics	Literature data	160	RFR, XGBoost, LightGBM, BPNN	0.9014	Zhuang et al., 2025
Flexural strength	Experimental data	162	Ensemble learning	0.9699	Zhang et al., 2020
			Linear regression	0.9517	

achieved using artificial neural networks (ANN) on a sample of 740 data points collected from literature sources (Hu et al., 2024). Using the largest dataset reviewed ($N = 1296$), which was based on finite element models and combined with XGBoost and decision trees, also provided exceptional accuracy ($R^2 = 0.98$) (Hu et al., 2024).

Among the applied ML methods, ensemble and gradient boosting approaches clearly dominate. XGBoost has repeatedly demonstrated its effectiveness in various studies (Hou et al., 2023; Mahajan et al., 2022; Ma et al., 2022; Zhuang et al., 2025), consistently providing R^2 values above 0.90. Additionally, ANN, Back Propagation Neural Network (BPNN) (Zhao et al., 2024; Sharan & Mitra, 2022; Zhuang et al., 2025) and Random Forest (Alsheghri et al., 2025; Zhao et al., 2024) are among the most popular and reliable tools for solving this class of problems. This emphasizes their ability to effectively model complex nonlinear dependencies in the properties of composite materials.

Thus, the literature analysis confirms that machine learning has become a powerful tool for predicting the mechanical properties of CFRPs, providing high accuracy even with limited data, and opening opportunities for the effective automation of engineering calculations, design optimization, and cost reduction for physical experiments. Despite ML successes with conventional CFRPs, PTFE-based systems with chaotic fiber arrangements remain underexplored due to their nonlinear strength-concentration relationship, which fundamentally differs from the linear behavior of standard composites.

This work aims to apply and benchmark existing ML algorithms for predicting the tensile strength of CFRPs with a thermoplastic matrix and a chaotic fiber arrangement. The novelty of the study consists in the identification of non-linear patterns that reveal a critical filler concentration threshold (10-20 wt.%), beyond which the strengthening effect is superseded by property degradation due to fiber agglomeration and increased porosity.

Materials & Methods

Proposed Approach

The research methodology that contributed to achieving the research objective consisted of three main stages. The first stage was devoted to collecting data on the properties of the composite material based on experimental results. These properties included the proportion of polymer matrix (a_1) and fillers (a_2), the overall density of the composite (a_3), and the resulting tensile strength (b). These parameters were recorded under controlled laboratory conditions to minimize measurement errors. The second stage was devoted to data analysis. At this stage, statistical pre- processing and exploratory data analysis were performed to identify potential outliers and correlations between variables. Regression metrics were then applied to assess the reliability and accuracy of the data obtained, which served as the basis for further modelling. The third stage involved applying machine learning methods to the processed dataset. Several algorithms were trained and optimized to predict the target variable (tensile strength) based on the input parameters of the material. The effectiveness of the model was verified using separate test datasets, which allowed for the assessment of its generalization ability and prediction reliability.

Experimental Part

To plan the experiments, six types of samples were manufactured with varying concentrations of carbon fiber in the polytetrafluoroethylene (PTFE) polymer matrix (ISO 13000–1:2021): 0, 10, 15, 20, 25, and 30 wt.%. These samples of CFRPs were produced using the powder metallurgy method, which involved mixing the polymer matrix and filler, followed by cold pressing and subsequent sintering (Berladir et al., 2020). The testing procedure for samples was carried out in accordance with the technical conditions TC U 22.2–33729459–001:2014 “Blanks for fluoroplastic compositions (FC)” (Ukraine). The density of the materials was determined by hydrostatic weighing: the masses of equal volumes of the sample and a liquid of known density (distilled water) were compared through double weighing, both in air and in water. The tensile strength was calculated by the standard procedure (Berladir et al., 2020).

Data Collection and Analysis

To ensure systematic analysis of material properties, a comprehensive data collection and statistical processing procedure was implemented. The dataset comprises 150 observations distributed evenly: 25 measurements for each of six fiber concentrations (0, 10, 15, 20, 25, 30 wt.%). At the initial stage of data preparation, a thorough data integrity check was performed to identify and remove any missing values or duplicate records.

Following data validation, descriptive statistical analysis was conducted on all variables to characterize material properties. During this analysis, the mean value and standard deviation were calculated as the arithmetic means of all observations. At the same time, the standard deviation was determined by calculating the square root of the mean square deviation from the mean value. The coefficient of variation CV was computed by dividing the standard deviation by the mean and multiplying by 100% to normalize this indicator for variables a_1, a_2, a_3, and b, which have different scales, using equation (1).

$$CV = \frac{\sigma}{\overline{x}} \times 100\% \tag{1}$$

In equation (1), σ is the standard deviation, and $\overline{x}$ represents the mean value. In addition, to increase the reliability of the location and distribution indicators, the median and interquartile range (IQR) were used. The skewness and kurtosis of each variable's distribution were also calculated to characterise the symmetry and tail behaviour of the distributions.

To investigate the relationships between variables, correlation analysis was performed using the Pearson correlation coefficient for all pairs of variables, as outlined in equation (2).

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{2}$$

The correlation coefficient measures the strength and direction of the linear association between two variables, with values ranging from −1 to +1. By computing pairwise correlations for all variables, a correlation matrix was constructed that revealed the relationships among predictors and between each predictor and the target variable.

Finally, regression-based metrics were employed to quantify the capacity of the input variables to explain the variation in the target variable. The mean squared error (MSE) was calculated by averaging the squared differences between predicted and observed values using equation (3).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \quad (3)$$

The root mean squared error ($RMSE$) was computed as the square root of the MSE, providing an error metric in the original units of the target variable. The $RMSE$ was described in equation (4).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \quad (4)$$

In equations (2), (3), and (4), y_i represents observed values, $\widehat{y}_i$ denotes predicted values, and $\overline{y}$ denotes the mean value of observations. These metrics were calculated using ordinary least squares regression. This was done to create an analytical basis for estimating the linear predicted relationships present in the data. Data estimation is a crucial step as it provides a baseline for understanding the relationships between variables. It must be performed before proceeding to more complex machine learning methods, as it lays the foundation for more advanced analyses.

To conduct all stages of statistical analysis and data processing, the Python programming language was employed. Its primary purpose in this study is to import, clean, and perform operations on data. The NumPy library in Python provides a wide range of mathematical and statistical functions, including the calculation of basic descriptive statistics and optimized array operations. This, in turn, facilitated the fast and efficient processing of large volumes of data. The *ydata_profiling library* (formerly known as *pandas-profiling*) was utilized to automate the generation of comprehensive profiling reports that summarized key descriptive statistics, detected anomalies and outliers within the dataset, and produced visual representations of distributions and correlation patterns. Therefore, the combined use of these libraries enabled the creation of a structured data analysis. As a result, this ensured the quality and integrity of the data for preparing a dataset for subsequent modeling stages.

Machine Learning Methods

After a comprehensive analysis of material properties using statistical and correlation methods, machine learning methods were applied in the next stage of the study. These methods were chosen because the nature of the data and the study's objectives justified this approach. In composite materials science, the relationships between the properties of the components and the final mechanical characteristics are often complex and multidimensional. Therefore, the physical and mathematical mechanisms are often poorly understood or cannot be described analytically, although experimental data can reveal correlations between these variables. In cases where empirical patterns exist in the data, but the fundamental governing equations are unknown, machine learning provides a systematic framework for discovering and modelling these relationships from the data.

The machine learning approach was structured in several stages. Initially, the collected dataset of 150 observations was partitioned into training and testing subsets using an 80/20 split. This partitioning strategy ensured that models were trained on a substantial portion of the data while maintaining a held-out test set to provide an unbiased assessment of model generalization and performance on previously unseen data. Standard 5-fold cross-validation was used for comparative model selection, and critical testing was performed on a stratified test set containing 5 samples for each concentration group. Before model training, the dataset underwent preprocessing procedures, including normalization and standardization, to ensure consistent feature scaling across all variables.

All machine learning computations were performed in MATLAB using the Regression Learner App. This is a comprehensive toolkit for interactive learning, validation, and comparison of multiple regression models. Therefore, using this application facilitated the systematic study of various algorithms and automatic tuning of hyperparameters. A comprehensive set of supervised regression algorithms was evaluated to determine the most effective approach for modelling the relationship between material parameters and tensile strength. The algorithms examined included Tree-based Regression, linear regression, stepwise linear regression, Support Vector Machines (SVM), efficient linear regression, Gaussian Process Regression, Neural Networks, and Kernel-based methods. Therefore, this study used different methods that incorporate various modelling paradigms. Those methods included linear approaches for capturing simple associations, tree- based methods for identifying hierarchical patterns and interactions between features, ensemble methods for combining multiple learning algorithms, as well as kernel methods for nonlinear transformations and neural networks for learning complex nonlinear relationships.

A significant role in machine learning was played by the cross-validation strategy, which was used to optimize hyperparameters and reduce overfitting. Therefore, each algorithm was systematically trained and evaluated, which allows us to determine its ability to capture the main patterns in the properties of materials.

Results

Data Analysis

The results of the statistical analysis of the dataset showed a solid foundation for further modeling and research in machine learning. Therefore, the suitability of the dataset for predictive modelling was confirmed for all variables, and the completeness and integrity of the data were thoroughly checked. The analysis results show that there were no missing or duplicate records, and the number of different measurements provided sufficient variability within each parameter.

For the main matrix component x_1, the mean value was 82.87 with a standard deviation of 10.24 and a CV of 12.36%. This indicates relatively dense clustering around the mean value. An additional filler component x_2 showed an average value of 17.13 and a similar standard deviation. In contrast to x_1, the much higher coefficient of variation of 59.81% reflected greater dispersion, as x_2 is a reverse component in the composite. The density parameter x_3 demonstrated exceptional stability in all samples, where its average value was 2009.39 and its CV was only 5.19%. An output parameter, which is tensile strength, yielded an average value of 20.17, a standard deviation of 2.74, and a CV of 13.57%. Thus, the input variables x_1, x_2, and x_3, as well as the output parameter y, exhibit moderate mechanical variability and are therefore suitable for predictive modelling. The distributional properties of each variable were analyzed using quantiles and measures of shape. The values for these parameters are given in Table 2.

Table 2
Statistical characteristics of the analyzed data

Variable	x_1	x_2	x_3	y
Mean	82.87	17.13	2009.39	20.17
Std Dev	10.24	10.24	104.38	2.74
CV (%)	12.36	59.81	5.19	13.57
Min	69.69	0.1	1878	17.24
Q1 (25%)	74.94	9.96	1955.75	18.05
Median	80.04	19.96	1982	18.74
Q3 (75%)	90.05	25.06	2018	22.21
Max	99.9	30.31	2233	25.27
IQR	15.11	15.1	62.25	4.16
Skewness	0.442	–0.442	1.161	0.763
Kurtosis	–0.999	–0.999	0.121	–0.988
Pearson r with y	0.658	–0.658	0.671	1
R^2 Explained	0.433	0.433	0.451	1

Descriptive and inferential measures, including the coefficient of variation and interquartile ranges, ensured that assessment of central tendency and spread was both relative and absolute. This provides a multifaceted profile of the sample set. The absence of monotonicity in several distributions suggested the presence of heterogeneous groupings within the data, which is related to underlying compositional or structural classes.

The results of the Pearson correlation analysis revealed pronounced linear associations among the parameters. This is confirmed by a strong positive correlation between x_1 and y ($r = 0.658$). A similar correlation was between x_3 and y ($r = 0.671$). The negative correlation for x_2 with y ($r = -0.658$) demonstrates that as the proportion of x_2 increases, y decreases. Additionally, the density variable x_3 exhibited a strong positive correlation with x_1 ($r = 0.940$) and a strong negative correlation with x_2 ($r = -0.940$), which indicates that density is primarily determined by the balance of the two components of the composite material. These relationships for different parameter combinations were illustrated by the heat map shown in Figure 2.

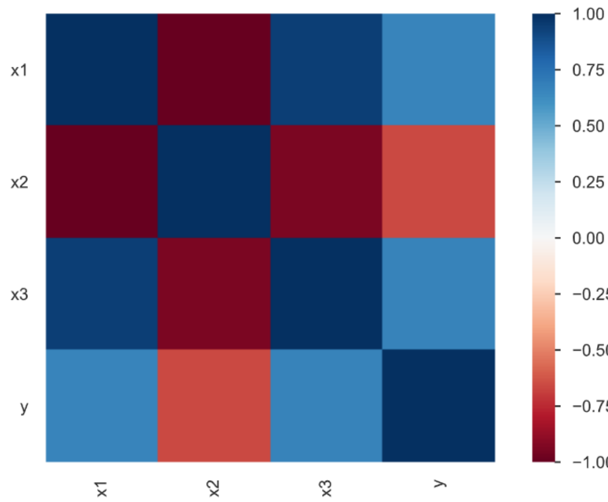
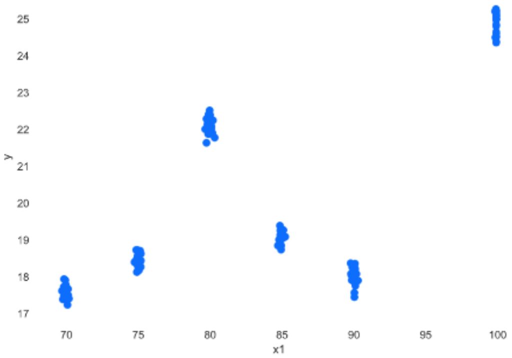


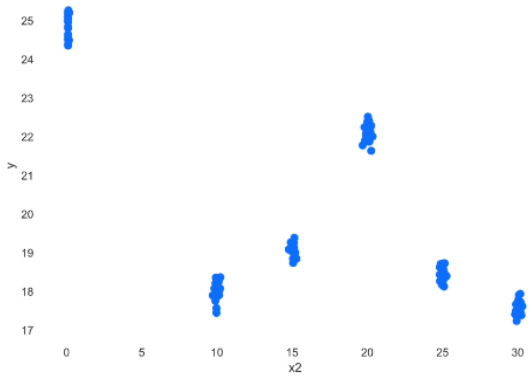
Fig. 2. The parameter correlation heat map

The heat map visually confirmed the symmetry and magnitude of the pairwise correlations. At the same time, the scatter plots between each predictor variable and the target variable y, shown in Fig. 3, showed clear linear trends and well-defined clusters.

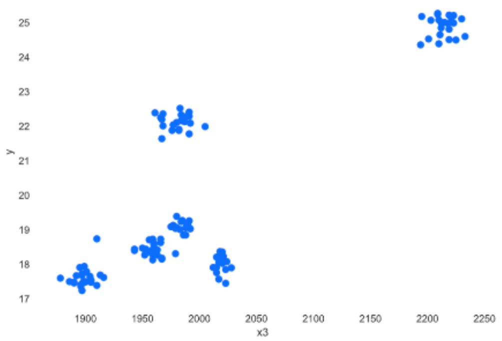
Therefore, various machine learning algorithms were employed to assess the comparative effectiveness of different models in predicting the properties of composite materials. This approach included linear and nonlinear models.



(a)



(b)



(c)

Fig. 3. Interactions between (a) x_1 and y, (b) x_2 and y, (c) x_3 and y

Machine Learning Models

Nine supervised regression algorithms were trained and validated on 120 observations from the 150-sample dataset, with the remaining 30 observations reserved for independent model testing. Validation results for all algorithms are summarized in Table 3.

Table 3
Validation results for ML algorithms

Model Type	**RMSE**	**MSE**	R^2
Ensemble	0.273921	0.075032	0.990201
Tree	0.281615	0.079307	0.989643
Gaussian Process Regression	0.290004	0.084102	0.989016
Neural Network	0.30146	0.090878	0.988132
SVM	0.302332	0.091405	0.988063
Stepwise Linear Regression	0.962095	0.925626	0.879116
Kernel	1.008785	1.017647	0.867098
Linear Regression	1.45274	2.110455	0.724381
Efficient Linear	3.021153	9.127366	–0.192007

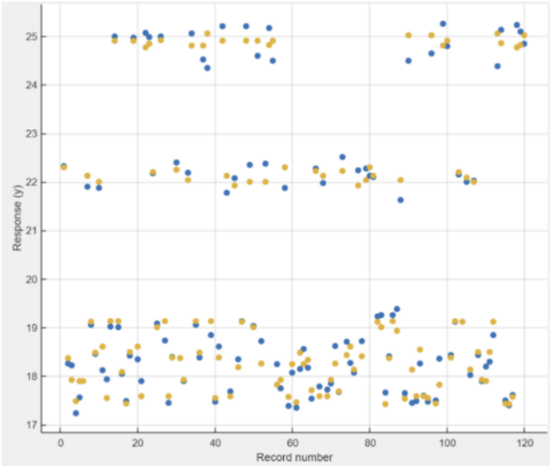
As shown in Table 3, the decision tree model depicted in Fig. 4a yielded the best results. These results are related to the hierarchical division of the material property space.

The neural network model, shown in Fig. 4b, achieved comparable validation effectiveness through deep nonlinear feature extraction using layered architectures.

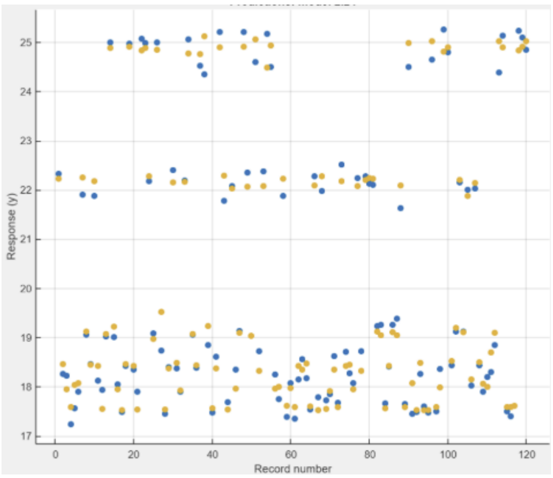
Therefore, this enables the network to learn intricate mappings between compositional properties across multiple nonlinear transformation stages. The ensemble method shown in Fig. 4c demonstrated comparable validation efficiency by aggregating multiple decision trees. This result is linked with reducing the variance of individual trees while preserving nonlinear expressiveness. The results of the test of all ML algorithms are presented in Table 4.

Table 4
Test results for ML algorithms

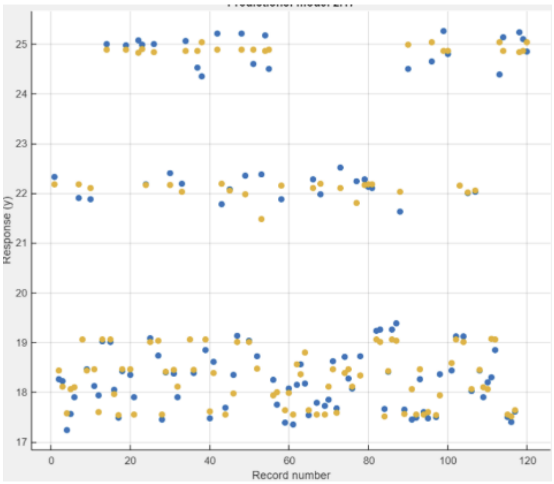
Model Type	**RMSE**	**MSE**	R^2
Neural Network	0.149281	0.022285	0.996351
Gaussian Process Regression	0.151424	0.022929	0.996246
Ensemble	0.154406	0.023841	0.996096
Tree	0.159066	0.025302	0.995857
SVM	0.161353	0.026035	0.995737
Kernel	0.366099	0.134029	0.978055
Linear Regression	0.800879	0.641408	0.89498
Stepwise Linear Regression	0.843861	0.712101	0.883405
Efficient Linear	2.551402	6.509653	–0.065849



(a)



(b)



(c)

Fig. 4. Response plot for (a) decision tree model, (b) neural network model, (c) ensemble method

Discussion

Analysis of the results obtained during the development of machine learning models revealed significant features in the effectiveness of different algorithms. The validation (Table 3) and testing (Table 4) results clearly indicate that nonlinear models outperform linear ones.

The neural network model achieved the highest accuracy through deep nonlinear transformation of compositional characteristics across multiple layers, which agrees well with the data (Zhao et al., 2024; Sharan & Mitra, 2022). Gaussian regression, in turn, achieved nearly equivalent performance through kernel-based probabilistic learning. This finding aligns with the results of the studies (Mahajan et al., 2022). The ensemble method demonstrated comparable effectiveness by aggregating multiple decision tree predictions. This finding aligns with the results of the studies (Hu et al., 2024; Zhuang et al., 2025). Linear regression methods showed notably poor results across both validation and test phases, falling significantly behind the nonlinear approaches. This poor performance is primarily due to the fact that linear models are unable to represent the complex relationships between material composition and tensile strength that exist in this dataset, as observed in the earlier correlation and clustering analysis.

Unlike many traditional composite systems (Alshammari et al., 2021), the behavior of carbon fiber reinforced PTFE exhibits pronounced nonlinearity in the relationship between filler concentration and tensile strength. The introduction of carbon fiber into the PTFE matrix is a technological solution primarily aimed at increasing wear resistance and stiffness, which are critical limitations of pure PTFE. Tensile strength is often a compromise characteristic that rarely achieves linear growth. This is a typical nonlinear behavior, described by a curve that progresses through stages of matrix property deterioration, a reinforcement phase, and an aggregation phase. At the defect stage (low fiber concentration, $< 5\%$), the tensile strength may remain at the level of pure PTFE or even decrease. This is due to the low adhesion between the non-polar PTFE and the fiber surface, resulting in the formation of weak interfaces and microporosity within the material. These defects serve as stress concentrators, nullifying the positive effect of reinforcement. Within the optimal reinforcement range (10–20 wt.%), the best balance is achieved between mechanical strengthening (effective load absorption by the fibers) and minimizing structural defects. Here, the maximum strength is recorded, but its value is the result of a complex interaction: strengthening due to fiber counteracts weakening due to defects. This concentration is also optimal for maximizing wear resistance and stiffness. A further increase in fiber content (high concentration, $> 20\%$) leads to a sharp drop in strength. This is caused by fiber agglomeration, a lack of polymer matrix for their full binding, and a critical increase in internal porosity.

It should be noted that the density of the composite is demodulated from the nonlinear tensile strength. Since carbon fiber is lighter than the polymer, the density of the composite decreases monotonically with increasing fiber concentration, even as the tensile strength peaks and then declines. The filler concentration is the primary factor that determines both mechanical properties (including tensile strength) and density (concentration of PTFE is inverse to the concentration of filler).

Despite the high predictive accuracy of the best models (neural network, Gaussian process regression, and ensemble), the application of the developed approach has significant limitations that should be considered further to enhance the reliability and engineering significance of the results.

The most significant limitation is the size of the initial dataset, which had only 150 observations. For complex nonlinear algorithms like neural networks, this amount may be too small to ensure complete coverage of the entire material property space. The high R^2 values (above 0.995) on the test set show successful training and generalization within the presented clusters. However, confirming the model's robustness outside these conditions will need further verification.

It should also be noted that the dataset spanned six discrete concentrations. Although the input concentrations were discrete, the density parameter (x_3) varied within each group, providing the necessary variability to the model. The high R^2 values (> 0.99) indicated accurate interpolation within the presented clusters, but generalization capability beyond these concentrations requires further verification with an expanded data range.

The models used only three input parameters. Other important factors, such as technological production parameters, were not considered. These could affect the final strength and density of the composite. There were moderate to strong linear correlations found between the predictors and the target variable. Yet, linear regression performed relatively poorly on the test set ($R^2 = 0.89498$), confirming that simple analytical approximations are inadequate compared to ML approaches. This suggests that the dependence of tensile strength limit on the inputs is a complex, nonlinear, multivariate function, not just a sum of simpler linear parts.

Thus, ML is not just a prediction tool. It is essential for understanding and controlling complex, nonlinear relationships in multicomponent polymer composites. This approach paves the way for accelerated material design.

Conclusions

In this work, a high-accuracy machine learning model was successfully developed and validated for predicting the tensile strength of carbon-fiber composites based on a thermoplastic matrix with a chaotic fiber arrangement.

The study confirmed that nonlinear models significantly outperform linear approaches in predicting the properties of this class of composite materials. The neural network model achieved the highest accuracy on an independent test set, demonstrating a coefficient of determination of $R^2 = 0.996351$. Equivalently high performance was demonstrated by the Gaussian process regression models ($R^2 = 0.996246$) and the ensemble method ($R^2 = 0.996096$). In contrast, classical linear models yielded noticeably lower results; for example, linear regression achieved an R^2 value of only 0.89498 on the test set, confirming that the overall dependence is a significant nonlinear multivariate function.

The key finding is the identification of a nonlinearity problem in the polytetrafluoroethylene–carbon fiber system. Unlike many traditional composites, the tensile strength does not increase linearly with filler concentration, but rather undergoes stages of matrix degradation, an optimal reinforcement phase (10–20 wt.%) where maximum strength is achieved, and a steep decline phase due to fiber agglomeration and increased porosity at high concentrations ($> 20\%$). The success of nonlinear ML models is based on their ability to effectively model these complex, nonlinear relationships between input parameters (component concentration and density) and final mechanical properties.

Among the limitations of this approach is the small size of the initial data set, which may not be sufficient to ensure full representation of the entire material property space, especially for complex nonlinear algorithms. For further improvement, it is necessary to systematically expand the data set by integrating additional production parameters to improve generalization ability and using literature data.

Acknowledgments

This contribution was created within the project "Research of Advanced Technologies for Increasing the Efficiency of Multivalent Energy Systems for Sustainable Industrial Development", ITMS21+ code:

401101C504, co-funded by the European Union under the Programme Slovakia. This paper was elaborated in the framework of the projects KEGA 012TUKE-4/2024, VEGA 1/0453/24, VEGA 1/0061/23 and KEGA 063TUKE-4/2025, granted by the Ministry of Education, Research, Development and Youth of the Slovak Republic. The research was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project 09I03-03-V03-00075. This research was supported by the European Union under the House of Europe programme. This research was also supported by a grant from the Ministry of Education and Science of Ukraine (No. 0125U000447). The research was partially supported by the International Association for Technological Development and Innovations.

References

Alshammari B.A., Alsuhybani M.S., Almushaikeh A.M., Alotaibi B.M., Alenad A.M., Alqahtani N.B., & Alharbi A.G. (2021). Comprehensive review of the properties and modifications of carbon fiber-reinforced thermoplastic composites. *Polymers*, *13*(15), 2474. DOI: 10.3390/polym13152474

Alsheghri A., Alhammadi A., Drakonakis V., Doumanidis H., Barsoum I., & Maalouf M. (2025). Predicting mechanical properties of CFRP composites using data-driven models with comparative analysis. *PLoS ONE*, *20*(4), e0319787. DOI: 10.1371/journal.pone.0319787

Antosz K., Knapčíková L., & Husár J. (2024). Evaluation and application of machine learning techniques for quality improvement in metal product manufacturing. *Applied Sciences*, *14*(22), 10450. DOI: 10.3390/app142210450

Berladir K., Gusak O., Demianenko M., Zajac J., & Ruban A. (2020). Functional properties of PTFE-composites produced by mechanical activation. In V. Ivanov et al. (Eds.), *Advances in Design, Simulation and Manufacturing II. DSMIE 2019. Lecture Notes in Mechanical Engineering* (pp. 391-401). Springer, Cham. DOI: 10.1007/978-3-030-22365-6_39

Berladir K.V., Hovorun T.P., Sviderskiy V.A., Rudenko P.V., & Vyshehorodtseva M.E. (2016). Nanostructural modification of polytetrafluoroethylene and its composition by energy influence. *Journal of Nano- and Electronic Physics*, *8*(1), 01033-1-01033-5. DOI: 10.21272/jnep.8(1).01033

Budnik A.F., Rudenko P.V., Berladir K.V., & Budnik O.A. (2015). Structured nanoobjects of polytetrafluoroethylene composites. *Journal of Nano- and Electronic Physics*, *7*(2), 02022.

Hoang Q.N., Lee S., Lee S., & Park H. (2026). Characterization of CFRP interface properties with varying fiber surface roughness using AFM measurements and finite element modeling. *Composites Part B: Engineering*, *308*, 112990. DOI: 10.1016/j.compositesb.2025.112990

Hou J., Wu C., Yao Y., Qin R., Lau D., & Tam L. (2023). Machine learning-based investigations on nonlinear vibrations of CFRP composite. In *CICE 2023 - 11th International Conference on FRP Composites in Civil Engineering* (Article 77). Zenodo. DOI: 10.5281/zenodo.8114224

Hrehova S., Antosz K., Husár J., & Vagaska A. (2025). From simulation to validation in ensuring quality and reliability in model-based predictive analysis. *Applied Sciences*, *15*(6), 3107. DOI: 10.3390/app15063107

Hu H., Wei Q., Wang T., Ma Q., Jin P., Pan S., Li F., Wang S., Yang Y., & Li Y. (2024). Experimental and numerical investigation integrated with machine learning (ML) for the prediction strategy of DP590/CFRP composite laminates. *Polymers*, *16*(11), 1589. DOI: 10.3390/polym16111589

Kurita H., Suganuma M., Wang Y., & Narita F. (2022). k-Means clustering for prediction of tensile properties in carbon fiber-reinforced polymer composites. *Advanced Engineering Materials*, *24*, 2101072. DOI: 10.1002/adem.202101072

Liang Y., Wei X., Peng Y., Wang X., & Niu X. (2025). A review on recent applications of machine learning in mechanical properties of composites. *Polymer Composites*, *46*, 1939-1960. DOI: 10.1002/pc.29082

Ma L., Zhou C., Lee D., & Zhang J. (2022). Prediction of axial compressive capacity of CFRP-confined concrete-filled steel tubular short columns based on XGBoost algorithm. *Engineering Structures*, *260*, 114239. DOI: 10.1016/j.engstruct.2022.114239

Mahajan A., Bajoliya S., Khandelwal S., Guntewar R., Ruchitha A., Singh I., & Arora N. (2022). Comparison of ML algorithms for prediction of tensile strength of polymer matrix composites. *Materials Today: Proceedings*. DOI: 10.1016/j.matpr.2022.12.105

Monticeli F.M., Alves F.C., de Paula Santos L.F., Costa M.L., & Botelho E.C. (2024). Chapter 11 - Predicting the mechanical behavior of carbon fiber-reinforced polymer using machine learning methods: a systematic review. In K. Palanikumar, E. Natarajan, S. Ramesh, & J. P. Davim (Eds.), *Woodhead Publishing Reviews: Mechanical Engineering Series, Machine Intelligence in Mechanical Engineering* (pp. 193-233). Academic Press. DOI: 10.1016/B978-0-443-18644-8.00012-5

Ozkan D., Gok M.S., & Karaoglanli A.C. (2020). Carbon fiber reinforced polymer (CFRP) composite materials, their characteristic properties, industrial application areas and their machinability. In A. Öchsner & H.

Altenbach (Eds.), *Engineering Design Applications III. Advanced Structured Materials* (Vol. 124, pp. 235-253). Springer, Cham. DOI: 10.1007/978-3-030-39062-4_20

Povstyanoy O., Imbirovich N., Redko R., Redko O., & Savaryn P. (2024). Numerical evaluation of the properties of highly efficient titanium porous materials. In V. Tonkonogyi, V. Ivanov, J. Trojanowska, G. Oborskyi, & I. Pavlenko (Eds.), *Advanced Manufacturing Processes V. InterPartner 2023. Lecture Notes in Mechanical Engineering* (pp. 307-317). Springer, Cham. DOI: 10.1007/978-3-031-42778-7_28

Povstyanoy O., Mikhailov A., Imbirovich N., Dziubynska O., & Herasymchuk H. (2021). Simulation permeable porous materials of the complex shape during radial-isostatic compression. In V. Tonkonogyi et al. (Eds.), *Advanced Manufacturing Processes II. InterPartner 2020. Lecture Notes in Mechanical Engineering* (pp. 339-348). Springer, Cham. DOI: 10.1007/978-3-030-68014-5_34

Sayam A., Rahman A.N.M.M., Rahman M.S., et al. (2022). A review on carbon fiber-reinforced hierarchical composites: mechanical performance, manufacturing process, structural applications and allied challenges. *Carbon Letters*, *32*, 1173-1205. DOI: 10.1007/s42823-022-00358-2

Sharan A., & Mitra M. (2022). Prediction of static strength properties of carbon fiber-reinforced composite using artificial neural network. *Modelling and Simulation in Materials Science and Engineering*, *30*, 075001. DOI: 10.1088/1361-651X/ac83df

Wang H.H., Huo S., Chevali V., Hall W., Offringa A., Song P., & Wang H. (2025). Carbon fiber reinforced thermoplastics: From materials to manufacturing and applications. *Advanced Materials*, *37*, 2418709. DOI: 10.1002/adma.202418709

Xu X., Peng G., Zhang B., Shi F., Gao L., & Gao J. (2025). Material performance, manufacturing methods, and engineering applications in aviation of carbon fiber reinforced polymers: A comprehensive review. *Thin-Walled Structures*, *209*, 112899. DOI: 10.1016/j.tws.2024.112899

Zhang Z., Shi J., Yu T., Santomauro A., Gordon A., Gou J., & Wu D. (2020). Predicting flexural strength of additively manufactured continuous carbon fiber-reinforced polymer composites using machine learning. *Journal of Computing and Information Science in Engineering*, *20*(6), 061015. DOI: 10.1115/1.4047477

Zhao G., Xu T., Fu X., Zhao W., Wang L., Lin J., Hu Y., & Du L. (2024). Machine-learning-assisted multiscale modeling strategy for predicting mechanical properties of carbon fiber reinforced polymers. *Composites Science and Technology*, *248*, 110455. DOI: 10.1016/j.compscitech.2024.110455

Zhuang W., Yang B., & Zhang H. (2025). Mechanical properties prediction and feature analysis of CFRP tubes under axial compression based on interpretable artificial intelligence. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, *8*, 398. DOI: 10.1007/s41939-025-00995-w