



Machine learning and shap-based prediction of multifactor thermal-oxidative aging behavior in glass fiber reinforced polyamide6

Hui Zhan^{1,#}, Jie Liu^{2,#}, Senhua Zhan¹, Bo Wu^{3,*}, Tongfei Shi^{1,*}

Keywords:

Glass fiber-reinforced polyamide6, multi-factor coupling, machine learning, thermal-oxidative aging, SHapley Additive exPlanations

Citation: Zhan, H.; Liu, J.; Zhan, S.; Wu, B.; Shi, T. Machine learning and shap-based prediction of multifactor thermal-oxidative aging behavior in glass fiber reinforced polyamide6. *J. Mater. Inf.* 2026, 6, 25. <https://dx.doi.org/10.20517/jmi.2025.81>

Received: 29 Sep 2025

First Decision: 30 Oct 2025

Revised: 12 Nov 2025

Accepted: 21 Nov 2025

Published: 17 Apr 2026

Academic Editor:

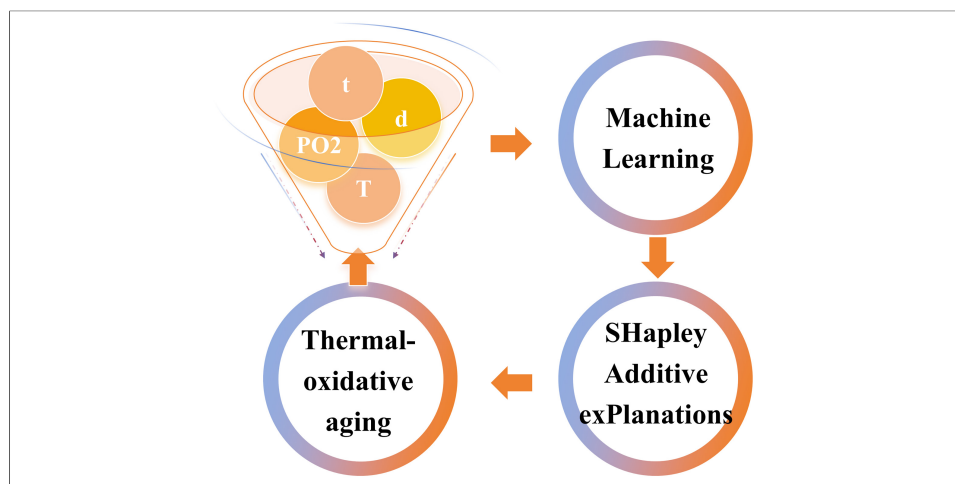
Ming Hu

Copy Editor:

Ping Zhang

Production Editor:

Ping Zhang



Abstract

In this study, glass fiber-reinforced polyamide 6 (PA6-GF) was selected as a representative system to establish a predictive framework for thermal-oxidative aging behavior. Through nonlinear fitting of traditional empirical models followed by the evaluation of twelve machine learning algorithms, it was found that the conventional models exhibited limited predictive accuracy and generalization capability, whereas the machine learning approaches were able to more effectively capture the complex nonlinear interactions among temperature, oxygen partial pressure, specimen thickness, and aging time. To further elucidate the underlying mechanisms, SHapley Additive exPlanations (SHAP) analysis was employed, highlighting the distinct roles and relative contributions of aging time, oxygen partial pressure, temperature, and thickness in governing the thermal-oxidative aging process. These findings enhance the understanding of multi-factor aging mechanisms and provide practical guidance for improving the long-term durability and reliability of engineering components operating under complex service conditions.

¹School of Chemical Engineering and Light Industry, Guangdong University of Technology, Guangzhou 510006, Guangdong, China.

²Enterprise Technology Center, Kingfa Science & Technology Co., Ltd., Guangzhou 510663, Guangdong, China.

³National Industrial Innovation Center of Polymer Materials, Guangzhou 510663, Guangdong, China.

[#]Authors contributed equally.

*Correspondence to: Dr. Bo Wu, National Industrial Innovation Center of Polymer Materials, Guangzhou 510663, Guangdong, China, E-mail: admin@guogaocai.com; Prof. Tongfei Shi, School of Chemical Engineering and Light Industry, Guangdong University of Technology, Guangzhou 510006, Guangdong, China, E-mail: tfshi@gdut.edu.cn

INTRODUCTION

As an essential technique for the development and performance evaluation of new materials, accelerated aging tests have greatly expedited material screening and iterative optimization processes^[1,2]. Among them, thermal-oxidative aging is the most widely used accelerated method^[3]. By subjecting materials to elevated temperatures and high oxygen partial pressures in the laboratory, it is possible to obtain aging behavior data within weeks or months, effectively replacing traditional natural aging tests that typically require several years, and thus significantly improving evaluation efficiency^[4]. However, notable discrepancies often exist between accelerated test results and actual service performance, leading to considerable debate regarding the extrapolation and scientific validity of such data in engineering applications^[5–7]. Consequently, how to efficiently and reliably predict the thermal-oxidative aging behavior of materials under complex service environments has become a key scientific challenge that must be addressed by both academia and industry^[8].

Current material aging studies predominantly adopt a controlled variable experimental design^[9,10]. For example, Richaud *et al.*^[11] studied the impact of pressure on polypropylene thermal-oxidative aging by comparing high-pressure and atmospheric conditions. Similarly, Courvoisier *et al.*^[12,13] investigated how sample thickness affects oxygen permeability in poly(ether imide) (PEI) and poly(ether ether ketone) (PEEK), while Faulkner *et al.*^[14] highlighted the dominant role of thickness in regulating aging under high-pressure oxygen environments. Although these studies have revealed the effects of individual factors, their applicability under complex service conditions involving multi-factor coupling is highly limited. Consequently, such approaches often remain at the level of qualitative analysis and fail to achieve accurate quantitative predictions of thermal-oxidative aging processes under actual operating conditions^[15].

To achieve quantitative prediction, researchers have proposed various mechanism-driven modeling approaches^[16,17], such as numerical models based on diffusion-reaction governing equations solved using finite element^[18–20] or finite difference methods, as well as molecular dynamics (MD) simulations^[21] and chemical kinetic models. These approaches incorporate physical processes such as oxygen diffusion and chemical reaction kinetics, thereby providing theoretical support for understanding thermal-oxidative aging behavior. However, the key parameters of these models are highly dependent on the specific characteristics of the material, including its chemical structure, crystallinity, filler content, and degree of cross-linking^[22]. As a result, when a new material is introduced, extensive experimental measurements and parameter fitting must be carried out again^[23]. In addition, Wang *et al.*^[24,25] separately established third-order polynomial nonlinear models using tensile strength and oxidation induction time as indicators. Although such empirically based extrapolation methods can, to some extent, provide predictive insights under realistic service conditions^[26,27], they remain essentially at the level of mathematical fitting and fail to capture the underlying physical mechanisms beyond the assumed conditions. Consequently, their extrapolation capability under unseen operating scenarios is inherently limited^[1,15,28]. Overall, mechanistic models are theoretically generalizable but lack parameter universality, while empirical models offer flexibility in prediction but lack mechanistic support^[28]; thus, neither approach can fully meet the practical demands of engineering applications.

In addition, research on thermal-oxidative aging still faces several critical challenges. Firstly, there is a severe lack of systematic, multi-factor coupled experimental data under complex service conditions^[29]. Secondly, traditional thermal-oxidative aging models, such as the Arrhenius equation^[27], primarily rely on empirical fitting of the temperature-reaction rate relationship^[30,31]. Although Ciutacu *et al.*^[32] proposed a lifetime prediction model that couples oxygen partial pressure and temperature, such formulas are generally based on the assumption of additivity or separability of variables. Essentially, this represents a compromise dimensionality reduction strategy for complex systems, which makes it difficult to effectively handle coupled effects among multiple factors and often fails to accurately capture higher-order nonlinear synergies^[33].

Finally, mechanistic models typically rely on extensive experiments and complex theoretical derivations to obtain key parameters, resulting in long research cycles and limited applicability. As service environments become increasingly complex, these models struggle to comprehensively capture the nonlinear coupling effects of high-dimensional, multivariable systems, thereby exhibiting limited predictive capability and practicality under realistic multi-field coupled conditions^[16,17].

Machine learning (ML) techniques, which do not require predefined mechanistic assumptions^[34], are capable of automatically capturing complex nonlinear relationships among variables from limited and high-dimensional experimental data^[35–38]. In recent years, ML has demonstrated significant advantages in modeling material aging. Previous studies have shown that ML models have achieved excellent results in areas such as predicting concrete performance^[39,40], identifying aging stages based on spectroscopic analysis^[7,41], inferring the natural service life of composites using accelerated aging data^[42] and optimizing material design^[43–45]. At the same time, beyond achieving high predictive accuracy, enhancing the interpretability of these models is equally crucial. The introduction of the SHapley Additive exPlanations^[46] (SHAP) method into model analysis enables quantitative assessment of the contribution and role of each feature in prediction outcomes, thus providing an effective approach to overcoming the limitations of ML models in mechanistic interpretation^[47–49].

In summary, this study focuses on glass fiber-reinforced polyamide 6 (PA6-GF) and aims to develop a predictive framework for its thermal-oxidative aging behavior under multi-factor coupled conditions, addressing the key challenge that conventional approaches rarely capture the combined effects of temperature, oxygen partial pressure, specimen thickness, and aging time within a single model. To this end, a self-developed dynamic accelerated aging apparatus with multi-factor coupling control was employed to generate mechanically representative aging data under various combinations of temperature, oxygen partial pressure, and specimen thickness. On this basis, nonlinear regression based on conventional empirical models was conducted to evaluate their applicability and limitations under multi-factor coupled conditions. Furthermore, twelve machine learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), CatBoost, and Extreme Gradient Boosting (XGBoost), were systematically compared in terms of their predictive performance. In addition, SHAP analysis was applied to interpret the four best-performing models. Overall, this study establishes a unified modeling framework for systematically characterizing multi-factor coupling effects, providing a feasible approach for predicting the thermal-oxidative aging behavior of composites under complex service environments and offering methodological insights into the application of machine learning to multi-factor coupled problems.

The structure of this paper is organized as follows: Section 2 elaborates on the significance of the study; Section 3 describes the experimental design, data acquisition, and modeling methods; Section 4 discusses the results and provides SHAP-based interpretations; and Section 5 concludes the paper with key findings and directions for future research.

MATERIALS AND METHODS

Experiment

Dataset collection

In this study, a self-developed thermal-oxidative aging device with multi-factor coupling control (as shown in Figure 1) was employed to collect high-fidelity mechanical aging data under various conditions of temperature, oxygen partial pressure, and specimen thickness, yielding a total of 190 samples (as shown in Figure 2). The accelerated aging test conditions of the samples are shown in Table 1. The experimental design established multi-factor combinations across the temperature-oxygen partial pressure space,

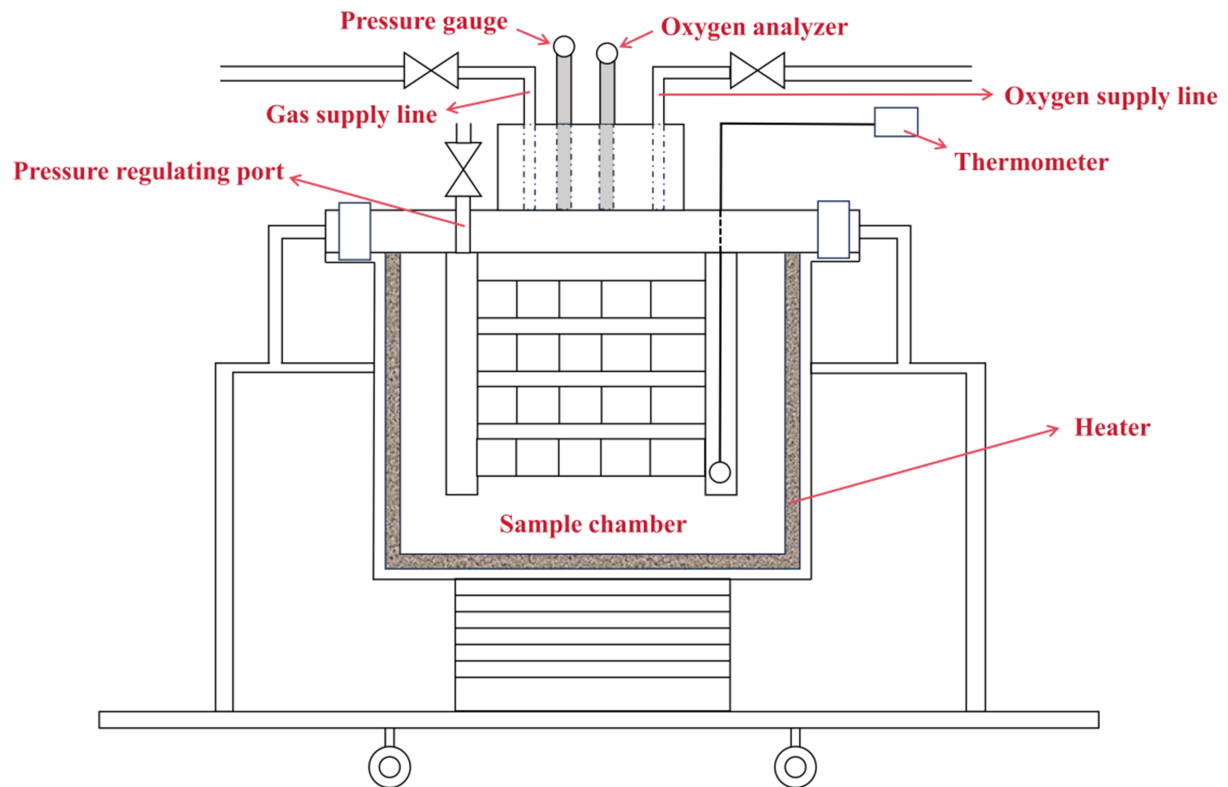


Figure 1. Schematic illustration of the dynamic thermal-oxidative aging device with multi-factor coupling control.

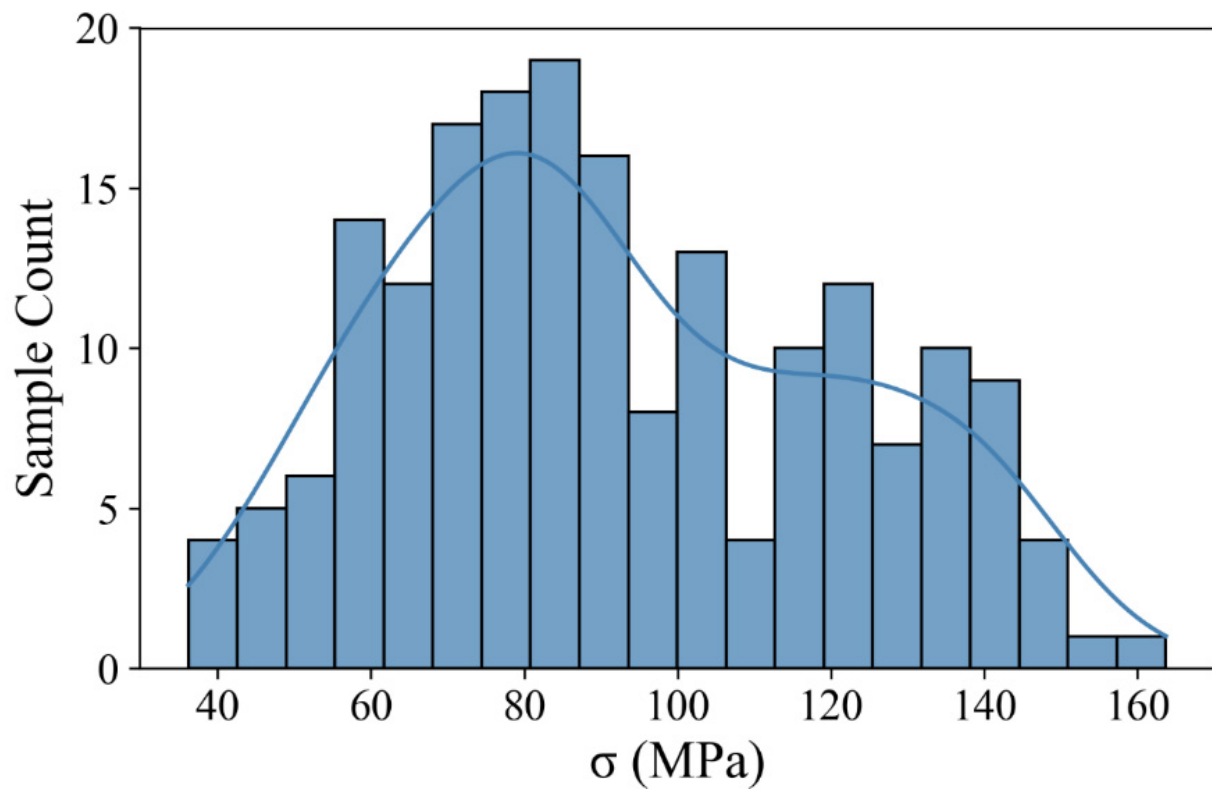


Figure 2. Distribution of tensile strength after aging for the collected samples.

achieving systematic coverage of the studied parameter ranges, thereby ensuring the dataset's representativeness under the main accelerated aging conditions and providing a reliable basis for subsequent

modeling and mechanistic analysis.

Table 1. Thermal-oxidative aging test conditions

Aging temperature T (°C)	Oxygen partial pressure PO ₂ (MPa)
160, 170, 180, 190	0.0036, 0.009, 0.0209, 0.06, 0.0627, 0.125, 0.1254, 0.1672, 0.1881, 0.21, 0.28

Materials and sample preparation

This study focused on PA6 composites reinforced with 30 wt% glass fiber (PA6-GF), supplied by Chongqing Sanlei Fiberglass Co., Ltd. ISO527-21BA tensile specimens were prepared by injection molding using a standard injection molding machine. The specimens had a length of 60.0 mm, a width of 5.0 mm, and thicknesses of 3.0, 1.5, and 0.75 mm, respectively. These specimens were subsequently used for aging tests. Prior to thermal-oxidative aging, all specimens were conditioned for more than 48 h at 23 ± 2 °C and $50 \pm 5\%$ relative humidity (RH). In view of the strong hygroscopicity of polyamide materials, this conditioning step was carried out in a desiccator or sealed bags to minimize the influence of fluctuations in ambient humidity. Since all subsequent aging tests were performed at temperatures above 100 °C, any residual trace moisture in the specimens would evaporate within a short time, and its effect on the thermal-oxidative aging behavior was therefore considered negligible.

Tensile testing

The tensile properties of the material were tested using an Instron 3365 electronic universal testing machine in accordance with ISO527-2:2012. The test environment was controlled at a constant temperature and humidity of 23 ± 2 °C and $50 \pm 5\%$ RH. Each group of tests had no less than 3 valid repeated samples, and the final results were averaged for analysis. The formula for calculating the retention rate of tensile strength is^[50]:

$$Retention = \frac{\sigma}{\sigma_0} \times 100\% \quad (1)$$

where σ is the tensile strength after stretching, and σ_0 is the initial tensile strength.

Methods

The dataset used in this study comprises four input variables - temperature (T), oxygen partial pressure (PO₂), specimen thickness (d), and aging time (t) - along with the corresponding measured tensile strength retention under each experimental condition. To ensure data reliability and reproducibility, the raw data were first checked for completeness, and samples containing missing values or abnormal records were removed to maintain consistency across all input variables. The dataset was then divided into a training set (80%) and a test set (20%) for subsequent model training and validation.

Traditional Arrhenius formula

Ciutacu *et al.*^[32] proposed an empirical kinetic Formula for glass fiber-reinforced epoxy systems, in which temperature, oxygen partial pressure, and time are combined in an Arrhenius-type form, with a power-law term for oxygen partial pressure introduced to predict the degradation of flexural strength. This approach validated the universality of using a closed-form response Formula to link environmental factors with the deterioration of mechanical properties.

In this section, maintaining the original framework of the empirical Formula proposed by Ciutacu *et al.*^[32], the model parameters were recalibrated based on the service characteristics of PA6-GF during the thermal-oxidative aging process, thereby enabling the adaptation and extension of the Formula to this new material

system.

Similarly, assuming that the degradation of material properties during the thermal-oxidative aging process follows first-order kinetics, its evolution can be expressed as:

$$-\frac{d\sigma}{dt} = k\sigma^\alpha \quad (2)$$

where k is the rate constant, and α is the total order of the degradation process. Here, let $\alpha = 1$, and the reaction is first-order kinetics.

Considering that, during the initial stage of thermal-oxidative aging, materials may undergo structural changes such as increased crystallinity, molecular chain rearrangement, enhanced interfacial bonding, residual stress release, and limited crosslinking, resulting in a transient strengthening effect. Jia *et al.*^[51] reported that, during aging at 121 °C, short glass fiber reinforced polyamides exhibited an increase in static tensile strength due to structural changes, including the increase in crystallinity and crystalline perfection. To further describe this phenomenon, a material constant a was introduced into the model as a structural adjustment factor, enabling a more accurate characterization of the material's evolution from initial structural reorganization to subsequent matrix degradation.

$$\sigma' = a\sigma_0 \quad (3)$$

where σ' represents the effective initial tensile strength of the material during the early stage of thermal aging, and a is a structural adjustment constant determined through the analysis of data.

Based on the Arrhenius Formula, the relationship between the reaction rate constant k and the temperature T can be expressed as:

$$k = A \times \exp\left(-\frac{E_a}{RT}\right) \quad (4)$$

where A is the exponential factor, E_a is the activation energy, R is the gas constant, which is $8.314 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$, and T is the thermal aging temperature (K).

Substituting the rate Formula into the performance attenuation Formula and combining it with the correction term a , the expression for the evolution of the performance retention rate over time can be obtained as follows:

$$\text{Retention} = a \exp\left[-A \exp\left(-\frac{E_a}{RT}\right)\right] \quad (5)$$

Further considering the effect of oxygen partial pressure on the reaction rate, this study introduces the oxygen partial pressure correction coefficient PO_2^n , that is:

$$A = A_0 PO_2^n \quad (6)$$

where A_0 is the reaction intensity factor under the reference state, and n is the oxygen partial pressure correction coefficient.

Considering the combined effects of oxygen partial pressure and aging time, the following empirical Formula is constructed to describe the evolution of material performance with respect to environmental factors:

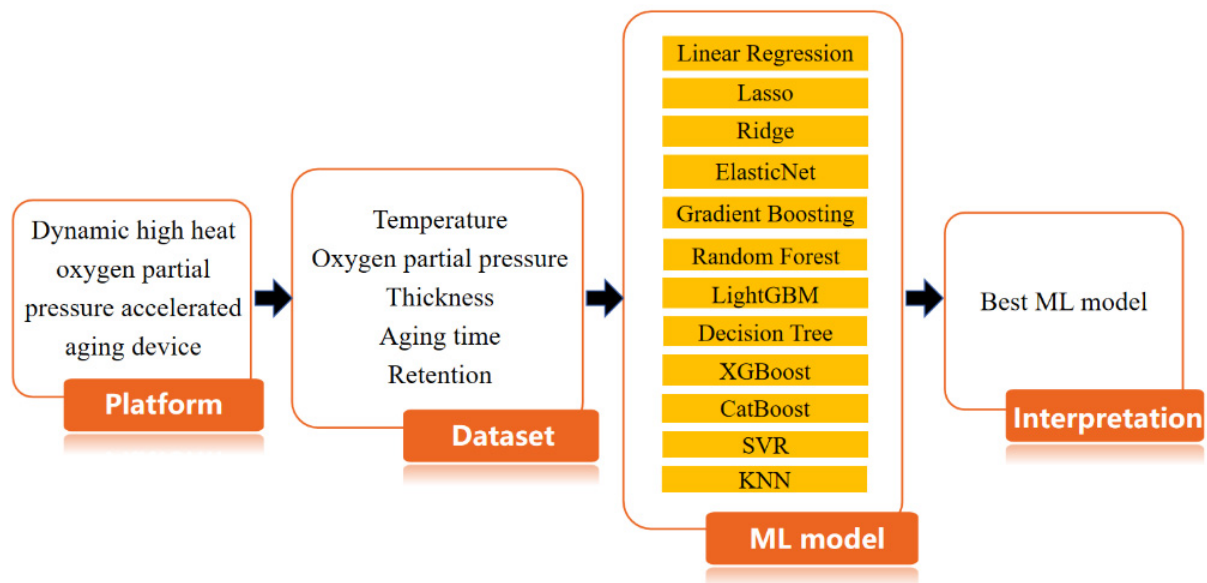


Figure 3. ML workflow diagram. ML: Machine learning; SVR: support vector regression; KNN: k-nearest neighbors regression.

$$\text{Retention (PO}_2, T, t) = a \exp \left[-A_0 \text{PO}_2^n \exp \left(-\frac{E_a}{RT} \right) \right] \quad (7)$$

ML models

To prevent model overfitting, a hyperparameter optimization strategy based on five-fold cross-validation was employed during training to evaluate the model's generalization performance and to reduce random errors caused by data partitioning. All numerical features were normalized using Z-score standardization (StandardScaler), which transforms feature values to zero mean and unit variance, thereby eliminating the influence of different feature scales on model training^[52,53].

Given the pronounced nonlinearity of the aging process and the potential interactions among variables, we developed a modeling framework that systematically evaluates twelve representative baseline models from four major algorithm families [Figure 3], including:

(a) Linear Model Class

Linear regression, the most fundamental and well-established method in supervised learning, constructs predictive models through a linear mapping between input features and target variables. It offers high computational efficiency, strong interpretability, and reliable controllability, and is therefore widely applied in tasks such as material property prediction, variable importance analysis, and degradation mechanism studies^[54,55]. To systematically assess how regularization influences generalization and coefficient constraints, four representative linear models were evaluated: ordinary least squares regression (OLS), L1-regularized regression (Lasso), L2-regularized regression (Ridge), and combined L1/L2-regularized regression (ElasticNet).

OLS minimizes the residual sum of squares and yields stable coefficient estimates when predictors are not strongly collinear, but it is prone to overfitting and parameter instability in high-dimensional settings or when inputs are highly correlated. Lasso introduces an L1-norm penalty that drives many coefficients to zero, simultaneously enhancing model robustness and performing feature selection - an advantage for variable screening and model compression. Ridge regression applies an L2-norm penalty to shrink all

coefficients uniformly, making it well suited to datasets with substantial multicollinearity and effectively damping coefficient oscillations. ElasticNet blends L1 and L2 penalties, striking a balance between sparsity and stability; it is particularly effective for high-dimensional, small-sample problems or systems in which predictors are strongly coupled.

(b) Tree Model and Ensemble Method Class

Tree-based models are non-parametric supervised learning methods that recursively partition the feature space. They inherently capture non-linear relationships and high-order interactions among variables, offering clear advantages for modeling complex system behaviors, handling heterogeneous features, and explaining variable influence pathways. The core principle is to construct a series of “condition-response” splitting rules that form a tree-shaped decision structure, thereby approximating the functional mapping between inputs and outputs. To evaluate the suitability and performance of tree-based approaches for modeling the thermal-oxidative aging of PA6-GF, six representative algorithms were selected: the basic single-tree model (Decision Tree^[56]) and five ensemble frameworks (Random Forest^[57], Gradient Boosting^[58], XGBoost^[59], LightGBM^[60], and CatBoost^[61]), covering the two main ensemble paradigms of Bagging and Boosting.

The Decision Tree is the simplest tree-based model: it partitions nodes by maximizing information gain or minimizing mean squared error. Because of its transparent and easily visualized logic, it is well suited for preliminary screening of variable importance and other tasks requiring maximum interpretability. Random Forest constructs multiple independent trees through Bagging and aggregates their outputs, effectively reducing variance and enhancing robustness to noisy inputs - an advantage for medium-sized datasets that require dependable models. Gradient Boosting iteratively fits the residuals of preceding learners, stacking weak models to approximate complex target functions, and excels with strongly non-linear data. XGBoost, LightGBM, and CatBoost are modern, high-efficiency gradient boosting frameworks, each incorporating targeted improvements such as loss-function regularization, optimized feature-splitting strategies, categorical variable handling, and faster computation. These innovations yield stronger generalization performance and greater engineering flexibility, particularly for high-dimensional, small-sample, or imbalanced data scenarios.

(c) Kernel Methods

Kernel methods constitute a supervised learning framework in which the original input space is implicitly projected into a high dimensional feature space through a non-linear mapping^[62]. Their core mechanism is the kernel trick: without the need to specify the mapping function explicitly, the inner product of two samples in that high dimensional space is evaluated through a kernel function, yielding a linear solution equivalent to one obtained after an explicit transformation. Such methods are well suited to strongly non-linear or high-dimensional data and hold considerable promise for modeling the aging behavior of materials^[63].

In this study, the chosen kernel model is a support vector regression (SVR), which extends the SVM framework to regression tasks^[64]. SVR constructs an ε -insensitive tube in feature space and minimizes a structural risk functional that balances prediction error against model complexity, resulting in a robust regression model. Because only the support vectors contribute to the solution, the model is sparse, exhibits strong generalization, and remains computationally efficient. SVR performs particularly well when (i) the input-output relationship is non-linear, (ii) sample sizes are limited or feature dimensions are high, and (iii) the regression task is affected by noise or outliers. By selecting an appropriate kernel, such as the radial basis (RBF) kernel, polynomial kernel or sigmoid kernel, SVR can be tailored to diverse data distributions and

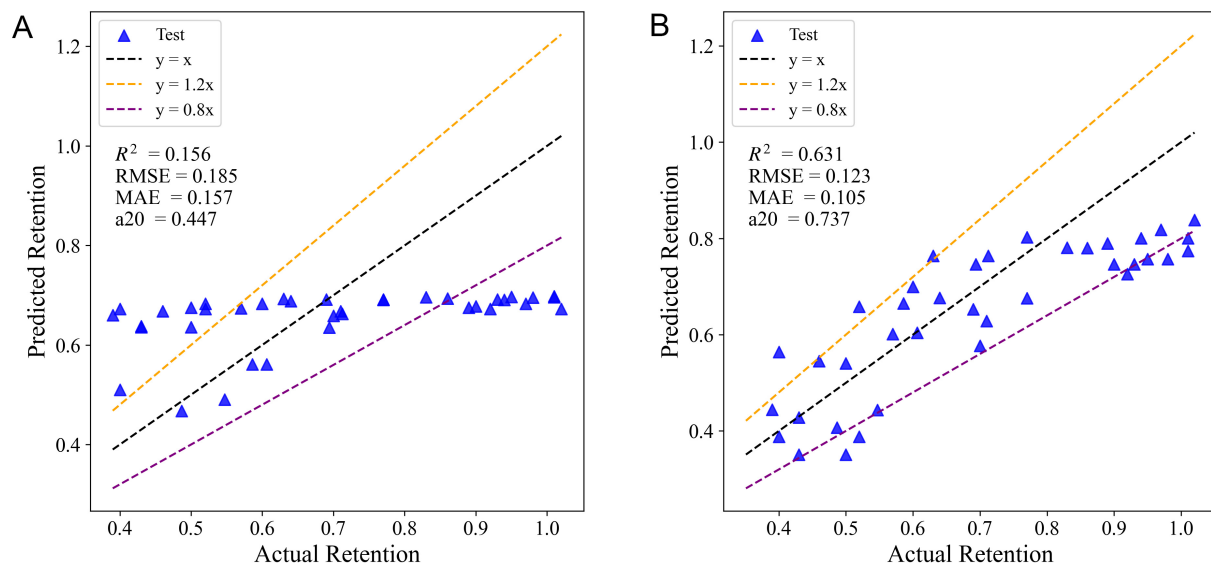


Figure 4. Prediction results of the (A) formula (5) and (B) formula (7). RMSE: Root mean square error; MAE: mean absolute error.

fitting requirements.

(d) Neighborhood Method Class

Neighborhood methods are nonparametric regression frameworks that base prediction not on a global function but on a local approximation drawn from the training samples most similar to the query point^[65]. The model adopted here is k-nearest neighbors regression (KNN), a lazy learning approach in which computation is deferred to the prediction stage^[66]. For each query point, KNN computes its distance to every instance in the training set (e.g., Euclidean or Manhattan distance), selects the closest k neighbors, and outputs either a simple or a distance-weighted average of their target values. Because it requires no parameter fitting, the method relies solely on the geometric relationships among samples, delivering strong local learning capability and straightforward interpretability.

KNN offers several advantages: a simple structure with easily tuned hyperparameters; no training phase, making it suitable for dynamic updates or real-time inference; and effective capture of local nonlinear patterns between inputs and responses. Its limitations include susceptibility to the “curse of dimensionality” in high-dimensional feature spaces, high sensitivity to the chosen distance metric, and performance dependence on sample distribution uniformity and noise levels.

RESULTS AND DISCUSSION

Evaluation of the predictive accuracy of the traditional formula

To evaluate the applicability of conventional empirical formulas in modeling thermal-oxidative aging, we first performed nonlinear regression using formulas (5) and (7) to predict the tensile strength retention of the material. The results showed that the test-set R^2 values were 0.156 and 0.631, respectively, while the corresponding A20 values were 0.447 and 0.737 (as shown in Figure 4). Although the introduction of oxygen partial pressure in formula (7) yielded a modest improvement in both R^2 and A20, the overall predictive performance remained unsatisfactory. Further analysis revealed that these empirical models did not explicitly account for the critical influence of specimen thickness on the aging process. In particular, thickness determines the effective penetration depth of oxygen, thereby governing the spatial distribution of performance degradation. The omission of this key variable prevents the model from capturing the diffusion-constrained nature of aging, which fundamentally limits its predictive accuracy.

Table 2. Performance prediction results of machine learning models

Model	Train				Test			
	R ²	MAE	RMSE	A20	R ²	MAE	RMSE	A20
Linear regression	0.290	0.140	0.171	0.526	0.246	0.150	0.175	0.421
Decision tree	0.896	0.049	0.065	0.934	0.762	0.078	0.098	0.816
Random forest	0.968	0.028	0.036	0.987	0.808	0.073	0.088	0.921
Gradient boosting	0.984	0.018	0.026	1.000	0.879	0.056	0.070	0.947
SVR	0.805	0.075	0.090	0.836	0.768	0.079	0.097	0.816
KNN	0.890	0.048	0.068	0.947	0.697	0.078	0.111	0.842
Lasso	0.164	0.156	0.186	0.428	0.045	0.169	0.197	0.395
Ridge	0.286	0.141	0.172	0.526	0.237	0.152	0.176	0.421
ElasticNet	0.275	0.144	0.173	0.487	0.203	0.155	0.180	0.342
XGBoost	0.941	0.039	0.049	0.974	0.897	0.050	0.065	0.974
LightGBM	0.833	0.060	0.083	0.888	0.774	0.075	0.096	0.868
CatBoost	0.984	0.020	0.026	1.000	0.910	0.050	0.060	1.000

MAE: Mean absolute error; RMSE: root mean square error; SVR: support vector regression; KNN: k-nearest neighbors regression.

Table 3. Hyperparameter optimization parameter values of the model

Model	Params
Linear Regression	-
Decision tree	max_depth: 15, min_samples_leaf: 1, min_samples_split: 8
Random forest	max_depth: 8, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 80
Gradient boosting	learning_rate: 0.212, max_depth: 4, min_samples_leaf: 4, min_samples_split: 12, n_estimators: 143
SVR	C: 9.589, gamma: auto, kernel: rbf
KNN	algorithm: brute, leaf_size: 10, metric: euclidean, n_neighbors: 2, p: 1.5, weights: uniform
Lasso	alpha: 0.031
Ridge	alpha: 9.709
ElasticNet	alpha: 0.035, l1_ratio: 0.208
XGBoost	colsample_bytree: 0.544, learning_rate: 0.267, max_depth: 2, n_estimators: 180, subsample: 0.393
LightGBM	learning_rate: 0.163, max_depth: 3, n_estimators: 111, num_leaves: 56
CatBoost	depth: 5, iterations: 105, l2_leaf_reg: 3.434, learning_rate: 0.304

SVR: Support vector regression; KNN: k-nearest neighbors regression.

Model performance

Table 2 summarizes the predictive performance of 12 representative machine learning models for the thermal-oxidative aging behavior of PA6-GF under multi-physics coupled conditions. The model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and the A20 index (A20). Specifically, R^2 reflects the goodness of fit of the model, where values closer to 1 indicate higher consistency between the predicted and experimental results. MAE and RMSE represent the average level and fluctuation of prediction errors, respectively, and are used to assess the accuracy and stability of the models. The A20 index denotes the proportion of samples whose relative prediction error falls within $\pm 20\%$ of the experimental values, serving as a metric of engineering-level prediction acceptability. Table 3 lists the main training parameters of each model.

Overall, ensemble learning models demonstrated superior overall performance in predicting the thermal-oxidative aging behavior under multi-factor coupled conditions. Among them, the CatBoost model achieved the highest overall accuracy (as shown in Figure 5), with a test R^2 of 0.910, an MAE of 0.050, an RMSE of 0.060, and an A20 value of 1.000. This indicates that all predicted results fell within $\pm 20\%$ of the experimental values, suggesting excellent prediction accuracy and robustness. The XGBoost model achieved a test R^2 of 0.897, MAE of 0.050, RMSE of 0.065, and A20 of 0.974, exhibiting comparable nonlinear modeling capability and strong generalization performance. These results suggest that gradient-boosting-based ensemble algorithms can effectively capture high-order coupling relationships among temperature, oxygen partial pressure, and specimen thickness, thereby achieving high-accuracy prediction in complex nonlinear degradation systems. The Gradient Boosting and Random Forest models performed slightly less well but still maintained high prediction stability, with test R^2 values of 0.879 and 0.808, and A20 values of 0.947 and 0.921, respectively. Most predicted points thus fell within the engineering-acceptable error range. This demonstrates that ensemble learning frameworks retain strong generalization and noise-resistance capabilities even with limited sample sizes, effectively handling the nonlinearity and data imbalance inherent in thermal-oxidative aging processes. In contrast, linear models (including Linear Regression, Ridge Regression, Lasso Regression, and ElasticNet) exhibited significantly inferior performance, with test R^2 values below 0.3 and A20 values less than 0.5, while their MAE and RMSE were considerably higher than those of the nonlinear models. As these algorithms are only capable of representing linear dependencies, they fail to capture the complex nonlinear effects induced by diffusion-reaction coupling, thickness constraints, and multi-field interactions during thermal-oxidative aging, resulting in systematic deviations from experimental trends.

In conclusion, CatBoost demonstrated the highest overall performance in this challenging thermal-oxidative aging prediction task, providing superior accuracy, robustness, and interpretability compared with both traditional and other machine learning models. These findings underscore the importance of advanced ensemble algorithms for reliable modeling and mechanism elucidation in aging under coupled multi-factor conditions.

Feature importance analysis

After developing machine learning models to predict the thermal-oxidative aging behavior of PA6-GF, a physically consistent interpretability method was required to determine how each input variable influences the predictions. Accordingly, SHAP was employed to analyze the model outputs. Rooted in Shapley value theory, SHAP quantifies the marginal contribution of each feature under multivariate interactions, revealing not only its importance but also the direction of its impact, whether positive or negative. Unlike conventional feature importance metrics, SHAP provides a unified global-local explanation: it ranks features by overall contribution and shows how they drive the prediction for any individual sample, making it particularly suitable for modeling the nonlinear system studied here.

Figure 6 presents the SHAP results for CatBoost, XGBoost, Gradient Boosting, and Random Forest. The plots clearly illustrate how each input variable contributes to the predicted retention of tensile strength in thermal-oxidative aging, and the observed patterns correspond well with established physical mechanisms.

Aging time

Figure 6 shows that aging time (t) exerts a remarkably consistent marginal influence across all four models: its SHAP value declines monotonically as t increases, with a sharply falling edge in the long-time region that amplifies its impact on retention loss. The density and spread of the SHAP values further indicate that aging time not only dominates the global explanation of model output but also exhibits a wider sample-to-sample influence than any other variable, underscoring its stable yet strongly interventionist role. These findings

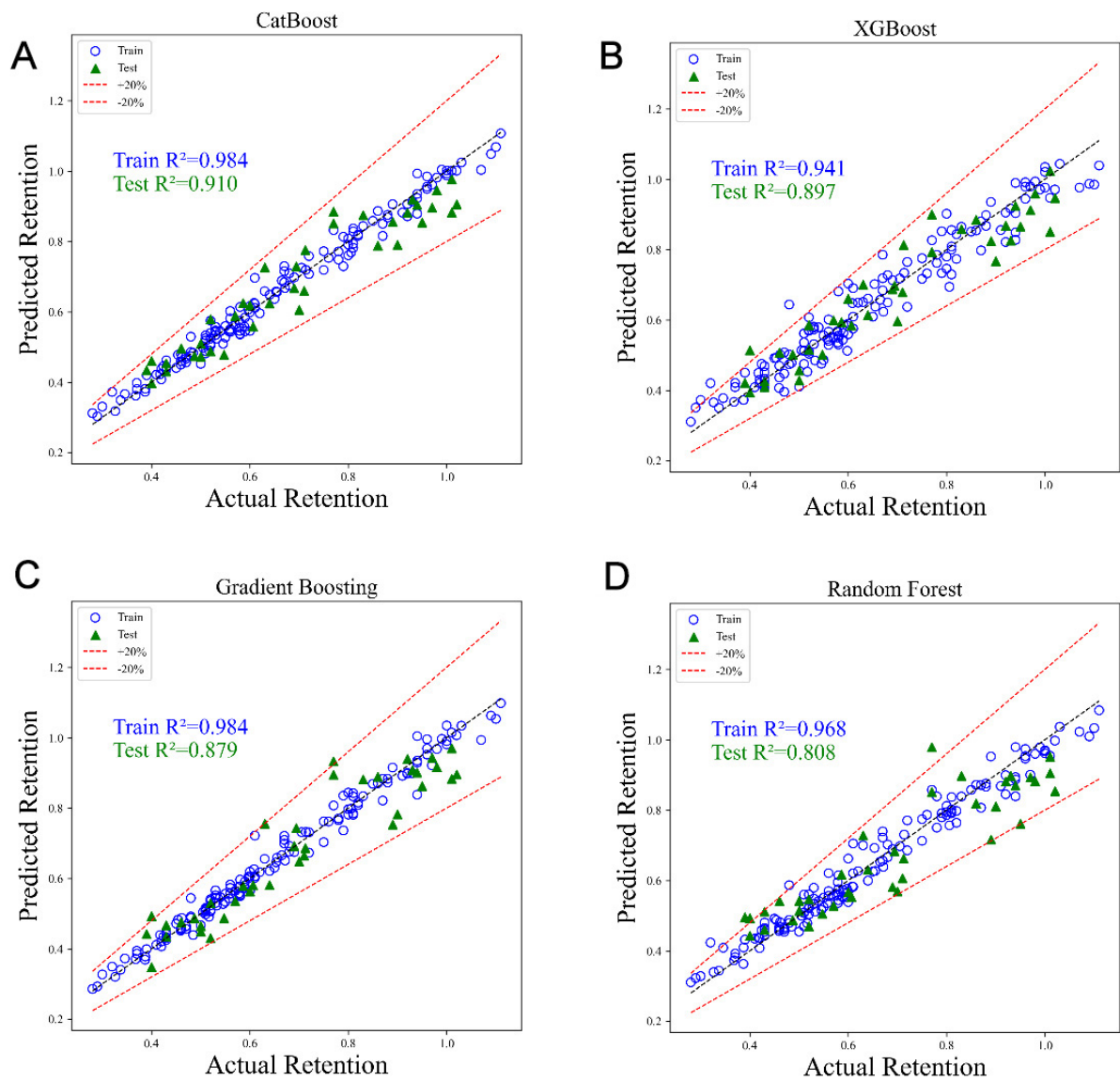


Figure 5. R^2 plot of the top four models: (A) CatBoost, (B) XGBoost, (C) Gradient Boosting, and (D) Random Forest.

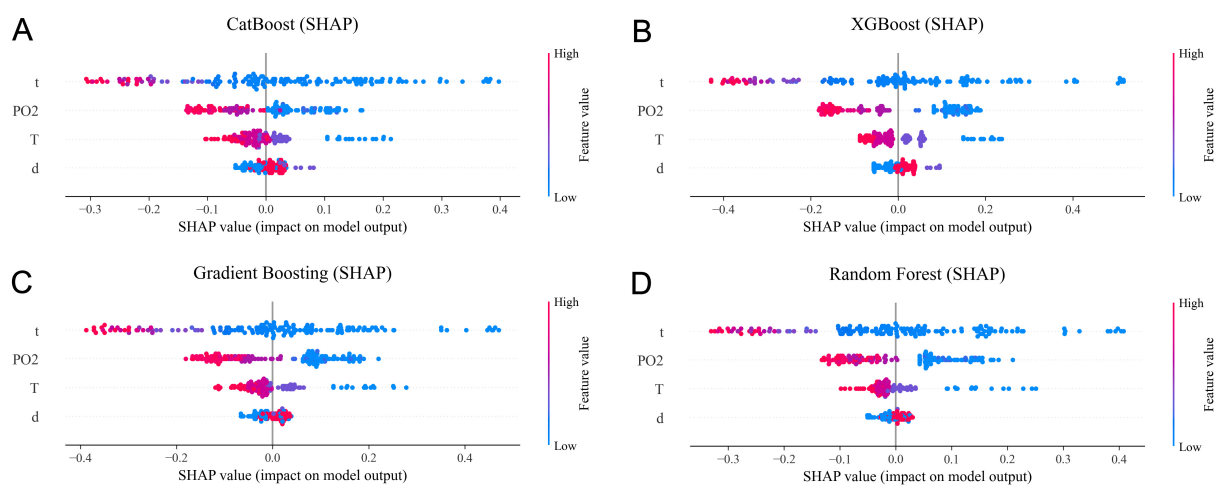


Figure 6. SHAP plots for (A) CatBoost, (B) XGBoost, (C) Gradient Boosting, and (D) Random Forest. SHAP: SHapley additive exPlanations.

confirm that time is the principal axis of response in thermal-oxidative aging models and suggest a latent, nonlinear coupling between time and other variables - such as oxygen partial pressure and temperature - whereby the effect of t on retention may exhibit either accelerated growth or saturation under certain environmental conditions.

Consequently, aging time is not merely a core input dimension; it sets the temporal scale for reaction rates and degradation pathways in any predictive formulation. A robust aging behavior model must therefore place time at its center, linking it dynamically to mechanisms such as chemical reaction in a manner consistent with physical law.

Oxygen partial pressure

As shown in Figure 6, the SHAP distribution of PO_2 exhibits a characteristic nonlinear “S-shaped” trend. In the low-pressure region, SHAP values remain close to zero, indicating that oxygen has a negligible influence on the aging rate. As PO_2 increases to a moderate range, the SHAP values decrease markedly, suggesting that higher oxygen partial pressure significantly accelerates the aging process. However, at elevated PO_2 levels, the marginal effect of oxygen gradually saturates, and the SHAP curve becomes flatter. This trend indicates that the influence of oxygen partial pressure on the aging behavior is not continuously enhanced but rather shows a diminishing marginal effect in the high-oxygen region. This statistical trend reflects the nonlinear response learned by the model under varying oxygen concentrations.

Temperature

In Figure 6, the SHAP distribution of the temperature variable (T) exhibits a typical monotonic negative trend with an enhancing effect: as temperature increases, the SHAP values consistently decrease, indicating that elevated temperatures significantly accelerate the reduction in material retention. This response pattern is highly consistent with the fundamental reaction kinetics of thermal-oxidative aging. Higher temperatures not only enhance molecular thermal motion, but also exponentially amplify the reaction rate constant (k) according to the Arrhenius relationship, thereby promoting main-chain scission, free radical reactions, and the formation of oxidation products, ultimately accelerating structural degradation and performance loss. As a generalized background variable, temperature acts more as a “rate regulator” rather than a “gating variable”. Overall, the SHAP analysis for temperature reveals its intrinsic driving effect on aging kinetics and quantitatively reflects its essential role within the multi-factor coupled system.

Thickness

In the SHAP analysis results shown in Figure 6, the thickness d variable exhibits the following trend: at small thicknesses, SHAP values are negative, indicating that oxygen can readily penetrate and undergo uniform oxidation, leading to a rapid decrease in retention. When the thickness exceeds a certain threshold, SHAP values gradually turn positive, reflecting the formation of an oxidation hysteresis zone inside the material that effectively shields deeper layers from oxidative damage, thereby significantly slowing the rate of retention loss. This phenomenon is highly consistent with the diffusion-limited oxidation (DLO) mechanism in the thermal-oxidative aging of PA6-GF: as thickness increases, the path and time required for oxygen to diffuse from the surface into the bulk grow substantially, making it difficult for oxygen molecules to reach the interior. As a result, the inner region maintains a low oxygen concentration, which slows the rate of deep oxidation reactions. This diffusion-delay-reaction-hysteresis behavior is manifested macroscopically as a slowdown in the downward retention trend, underscoring the key regulatory role of thickness in structural protection and performance preservation.

Therefore, the SHAP profile for thickness not only illustrates its regulatory effect on retention but also highlights its central role in controlling the spatial structure and dynamic rate of aging evolution at the

model level.

CONCLUSION

In this study, a self-developed accelerated aging platform with multi-factor coupling control was employed to obtain high-fidelity mechanical performance data under various combinations of temperature, oxygen partial pressure, and specimen thickness, providing a reliable foundation for subsequent model training and mechanistic analysis. Nonlinear regression using conventional empirical models was performed to systematically evaluate their applicability and limitations under multi-factor coupled conditions. The results indicate that although traditional models can reasonably describe aging behavior under single-factor variations, they fail to simultaneously capture the complex mechanisms arising from temperature-dependent reaction kinetics, thickness-governed oxygen diffusion constraints, and the nonlinear response to oxygen pressure. Such simplifications lead to significant prediction deviations under high-pressure or thick-specimen conditions, limiting their ability to achieve unified characterization under multi-field coupling. Building upon this, twelve machine learning models were further compared to evaluate their predictive performance. The results show that gradient-boosting-based ensemble algorithms exhibited the best overall performance, effectively capturing the high-order nonlinear interactions among temperature, oxygen partial pressure, thickness, and aging time. SHAP-based interpretability analysis further revealed that thickness exerts a strong positive influence on performance retention, while temperature and aging time have pronounced negative effects; oxygen partial pressure exhibits a nonmonotonic “S-shaped” response, suggesting a diminishing marginal effect at higher oxygen concentrations. These findings highlight the dominant role of multi-factor synergy in governing thermal-oxidative aging behavior and demonstrate the superior capability of machine learning approaches in modeling complex nonlinear degradation systems.

This study provides an effective framework for applying machine learning to investigate material aging processes. The proposed predictive approach shows good generality and can be extended to other polymer systems by adjusting model parameters and evaluation metrics to accommodate different structural characteristics and service environments. Nevertheless, certain limitations remain. The current model primarily relies on data-driven learning, which makes its performance sensitive to data distribution and potentially limits its generalizability beyond the training domain. In addition, it does not quantitatively describe the aging process from a mechanistic perspective. Future work will focus on integrating physics-informed neural networks (PINNs) to enhance the physical consistency and interpretability of the predictions. The framework will also be extended to a broader range of materials, enabling systematic aging experiments and modeling analyses under complex coupled conditions such as humidity, ultraviolet radiation, and salt spray, thereby establishing a solid theoretical foundation for mechanism-based aging prediction.

DECLARATIONS

Acknowledgments

The authors would like to thank the experimental staff of Kingfa Science & Technology Co., Ltd. and the National Industrial Innovation Center of Polymer Materials Co., Ltd. for providing the experimental data. Figures were generated using Matplotlib.

Authors' contributions

Investigation, software, methodology and writing - original draft: Zhan, H.
Resources, data curation and investigation: Liu J.
Writing - review & editing and validation: Zhan S.
Validation, writing - review & editing, and supervision: Wu B.; Shi T.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request and with the consent of the collaborating partners.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (GPT-o4-mini, released 2025-04-16) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

The authors acknowledge the financial support from the National Natural Science Foundation of China (No.22473032).

Conflicts of interest

Liu, J. is affiliated with Kingfa Science & Technology Co., Ltd., while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Celina, M. C. Review of polymer oxidation and its relationship with materials performance and lifetime prediction. *Polym. Degrad. Stab.* **2013**, *98*, 2419–29. [DOI](#)
2. Li, D.; Xu, S.; Li, J. Effect of thermal oxygen aging on the crystallization and insulation properties of polyether ether ketone. *J. Appl. Polym. Sci.* **2023**, *140*, e53745. [DOI](#)
3. Ghaderi, A.; Ayoub, G.; Dargazany, R. Constitutive behavior and failure prediction of crosslinked polymers exposed to concurrent fatigue and thermal aging: a reduced-order knowledge-driven machine-learned model. *J. Mater. Sci.* **2023**, *59*, 5066–84. [DOI](#)
4. Tsotsis, T. K.; Keller, S.; Bardis, J.; Bish, J. Preliminary evaluation of the use of elevated pressure to accelerate thermo-oxidative aging in composites. *Polym. Degrad. Stab.* **1999**, *64*, 207–12. [DOI](#)
5. Ye, Y.; Tang, G.; Meng, X.; Yang, R.; Guo, B.; Xie, X. Efficient and precise outdoor lifetime prediction method for biodegradable poly (butylene adipate/ terephthalate) under various climate conditions. *J. Clean. Prod.* **2025**, *500*, 145276. [DOI](#)
6. Wang, X.; Petru, M. Degradation of bending properties of flax fiber reinforced polymer after natural aging and accelerated aging. *Constr. Build. Mater.* **2020**, *240*, 117909. [DOI](#)
7. Yu, H.; Zhang, D.; Cui, L.; Kong, Y.; Jiao, X. A Mach. Learn. approach for efficiently predicting polymer aging from UV-vis spectra. *J. Phys. Chem. B.* **2024**, *128*, 9255–61. [DOI](#)
8. Sharma, H.; Arora, G.; Singh, M. K.; et al. Review of Mach. Learn. approaches for predicting mechanical behavior of composite materials. *Discov. Appl. Sci.* **2025**, *7*, 1238. [DOI](#)
9. Bernstein, R.; Derzon, D. K.; Gillen, K. T. Nylon 6.6 accelerated aging studies: thermal-oxidative degradation and its interaction with hydrolysis. *Polym. Degrad. Stab.* **2005**, *88*, 480–8. [DOI](#)
10. Yarahmadi, N.; Vega, A.; Jakubowicz, I. Accelerated ageing and degradation characteristics of rigid polyurethane foam. *Polym. Degrad. Stab.* **2017**, *138*, 192–200. [DOI](#)
11. Richaud, E.; Farcas, F.; Fayolle, B.; Audouin, L.; Verdu, J. Accelerated ageing of polypropylene stabilized by phenolic antioxidants under high oxygen pressure. *J. Appl. Polym. Sci.* **2008**, *110*, 3313–21. [DOI](#)
12. Courvoisier, E.; Bicaba, Y.; Colin, X. Multi-scale and multi-technical analysis of the thermal degradation of poly(ether imide). *Polym. Degrad. Stab.* **2018**, *147*, 177–86. [DOI](#)
13. Courvoisier, E.; Bicaba, Y.; Colin, X. Multi-scale and multi-technique analysis of the thermal degradation of poly(ether ether ketone). *Polym. Degrad. Stab.* **2018**, *151*, 65–79. [DOI](#)
14. Faulkner, D. L. Aging of polypropylene using high oxygen pressure: Influence of sample thickness and stabilization. *J. Appl. Polym. Sci.* **1986**, *31*, 2129–38. [DOI](#)

15. Zha, S.; Lan, H.; Lin, N.; Meng, T. Degradation and characterization methods for polyethylene gas pipes after natural and accelerated aging. *Polym. Degrad. Stab.* **2023**, *208*, 110247. [DOI](#)
16. Colin, X.; Essatbi, F.; Delozanne, J.; Moreau, G. A new analytical model for predicting the thermal oxidation kinetics of composite organic matrices. Application to diamine cross-linked epoxy. *Polym. Degrad. Stab.* **2021**, *186*, 109513. [DOI](#)
17. Colin, X.; Marais, C.; Verdu, J. Thermal oxidation kinetics for a poly(bismaleimide). *J. Appl. Polym. Sci.* **2001**, *82*, 3418-30. [DOI](#)
18. Omairey, E. L.; Zhang, Y.; Soenen, H.; Carbonneau, X. Parametric analysis and field validations of oxidative ageing in asphalt pavements using multiphysics modelling approaches. *Int. J. Pavement. Eng.* **2022**, *24*, 2020267. [DOI](#)
19. Cassola, S.; Duhovic, M.; Schmidt, T.; May, D. Mach. Learn. for polymer composites process simulation - a review. *Compos. Part. B. Eng.* **2022**, *246*, 110208. [DOI](#)
20. Linde, E.; Nilsson, F.; Barrett, M.; Hedenqvist, M. S.; Celina, M. C. Time- and feedback-dependent DLO phenomena in oxidative polymer aging. *Polym. Degrad. Stab.* **2021**, *189*, 109598. [DOI](#)
21. Li, X.; Dong, C.; Li, J.; Bin, G.; Zhou, C. Molecular dynamics simulation of aging properties in polymer materials: a review. *Polym. Bull.* **2024**, *82*, 1723-53. [DOI](#)
22. Fang, C.; Wang, H.; Zhang, Y.; Zhang, M.; Shen, T.; Du, J. Multi-scale model for the aging performance of particle-filled polymer composites. *Polymers.* **2023**, *15*. [DOI PubMed PMC](#)
23. Wang, Y.; Wang, K.; Zhang, C. Applications of artificial intelligence/machine learning to high-performance composites. *Compos. Part. B. Eng.* **2024**, *285*, 111740. [DOI](#)
24. Wang, Y.; Lan, H.; Meng, T.; Chen, S.; Zuo, J.; Lin, N. A lifetime prediction method of pressured gas polyethylene pipes by thermal-oxidative aging test and tensile test. *J. Pressure. Vessel. Technol.* **2018**, *140*, 011404. [DOI](#)
25. Wang, Y.; Lan, H.; Meng, T. Lifetime prediction of natural gas polyethylene pipes with internal pressures. *Eng. Fail. Anal.* **2019**, *95*, 154-63. [DOI](#)
26. Ran, W.; Niu, L.; Wang, Y.; Lin, N.; Feng, G.; Lan, H. A lifetime prediction method of pressured gas polyethylene pipes by strain-hardening modulus and tensile test. *Appl. Sci.* **2023**, *13*, 7965. [DOI](#)
27. Feng, Y.; Hu, J.; Peng, C.; Wu, Q.; Xu, Z. High dual-aging-resistant property obtained in silica fume filled epoxy based composite and its lifetime prediction by aging model. *Mater. Res. Express.* **2018**, *6*, 016303. [DOI](#)
28. Hao, Z.; Feng, P.; Zhang, S.; Zhai, Y. Machine learning for predicting fiber-reinforced polymer durability: a critical review and future directions. *Compos. Part. B. Eng.* **2025**, *303*, 112587. [DOI](#)
29. Bernstein, R.; Gillen, K. Nylon 6.6 accelerating aging studies: II. Long-term thermal-oxidative and hydrolysis results. *Polym. Degrad. Stab.* **2010**, *95*, 1471-9. [DOI](#)
30. Xu, H.; Tian, T.; Zhang, J.; et al. Hot deformation behavior of the 25CrMo4 steel using a modified arrhenius model. *Materials.* **2022**, *15*, 2820. [DOI PubMed PMC](#)
31. Celina, M.; Gillen, K.; Assink, R. Accelerated aging and lifetime prediction: review of non-arrhenius behaviour due to two competing processes. *Polym. Degrad. Stab.* **2005**, *90*, 395-404. [DOI](#)
32. Ciutacu, S.; Budrugaec, P.; Niculae, I. Accelerated thermal aging of glass-reinforced epoxy resin under oxygen pressure. *Polym. Degrad. Stab.* **1991**, *31*, 365-72. [DOI](#)
33. Starkova, O.; Gagani, A. I.; Karl, C. W.; Rocha, I. B. C. M.; Burlakovs, J.; Krauklis, A. E. Modelling of Environmental Ageing of Polymers and Polymer Composites-Durability Prediction Methods. *Polymers.* **2022**, *14*, 907. [DOI](#)
34. Xu, P.; Ma, Y.; Lu, W.; Li, M.; Zhao, W.; Dai, Z. Multi-objective optimization in Mach. Learn. assisted materials design and discovery. *J. Mater. Inf.* **2025**, *5*, 26. [DOI](#)
35. Nasrin, T.; Pourkamali-anaraki, F.; Peterson, A. M. Application of Mach. Learn. in polymer additive manufacturing: a review. *J. Polym. Sci.* **2023**, *62*, 2639-69. [DOI](#)
36. Das, P. P.; Rabby, M. M.; Vadlamudi, V.; Raihan, R. Moisture content prediction in polymer composites using Mach. Learn. techniques. *Polymers.* **2022**, *14*, 4403. [DOI](#)
37. Cui, C.; Wang, H.; Gao, X.; Cao, G.; Liu, Z. Mach. Learn. model for thickness evolution of oxide scale during hot strip rolling of steels. *Metall. Mater. Trans. A.* **2021**, *52*, 4112-24. [DOI](#)
38. Gubernatis, J. E.; Lookman, T. Mach. Learn. in materials design and discovery: examples from the present and suggestions for the future. *Phys. Rev. Mater.* **2018**, *2*, 120301. [DOI](#)
39. Taffese, W. Z.; Wally, G. B.; Magalhães, F. C.; Espinosa-leal, L. Concrete aging factor prediction using Mach. Learn. *Mater. Today. Commun.* **2024**, *40*, 109527. [DOI](#)
40. Peng, L.; Miao, X.; Zhu, J.; et al. Hybrid Mach. Learn. and multi-objective optimization for intelligent design of green and low-carbon concrete. *Sustain. Mater. Technol.* **2025**, *45*, e01605. [DOI](#)
41. Zhu, K.; Wu, D.; Yang, S.; et al. Identification of aged polypropylene with Mach. Learn. and near-infrared spectroscopy for improved recycling. *Polymers.* **2025**, *17*, 700. [DOI](#)

42. Wang, J.; Karimi, S.; Zeinalzad, P.; Zhang, J.; Gong, Z. Using Mach. Learn. and experimental study to correlate and predict accelerated aging with natural aging of GFRP composites in hygrothermal conditions. *Constr. Build. Mater.* **2024**, *438*, 137264. DOI
43. Nematov, D.; Hojamberdiev, M. Mach. Learn.-driven materials discovery: unlocking next-generation functional materials - a review. *Comput. Condens. Matter.* **2025**, *45*, e01139. DOI
44. Luo, D.; Wang, K.; Wang, D.; Sharma, A.; Li, W.; Choi, I. H. Artificial intelligence in the design, optimization, and performance prediction of concrete materials: a comprehensive review. *npj. Mater. Sustain.* **2025**, *3*, 14. DOI
45. Zhang, Z.; Wang, Y.; Wang, W. Mach. Learn. in gel-based additive manufacturing: from material design to process optimization. *Gels* **2025**, *11*, 582. DOI
46. Lundberg, S. M.; Lee, S. I. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, 2017, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>. (accessed 2026-04-13).
47. Lundberg, S. M.; Erion, G. G.; Lee, S. I. Consistent individualized feature attribution for tree ensembles. *arXiv* **2018**, arXiv:1802.03888. DOI
48. Zhong, X.; Gallagher, B.; Liu, S.; Kailkhura, B.; Hiszpanski, A.; Han, T. Y. Explainable Mach. Learn. in materials science. *npj. Comput. Mater.* **2022**, *8*, 204. DOI
49. Hu, W.; Jing, E.; Qiu, H.; Sun, Z. Discovering polyimides and their composites with targeted mechanical properties through explainable Mach. Learn. *J. Mater. Inf.* **2025**, *5*, 1. DOI
50. Zhu, P.; Li, Z.; Zhu, Y.; Wu, Y.; Qu, W. Prediction of the long-term tensile strength of GFRP bars in concrete. *Buildings* **2023**, *13*, 1035. DOI
51. Jia, N.; Kagan, V. A. Effects of time and temperature on the tension-tension fatigue behavior of short fiber reinforced polyamides. *Polym. Compos.* **1998**, *19*, 408-14. DOI
52. Asteris, P. G.; Apostolopoulou, M.; Armaghani, D. J.; et al. On the metaheuristic models for the prediction of cement-metakaolin mortars compressive strength. *Metaheuristic. Comput. Appl.* **2020**, *1*, 063-99. <https://www.techno-press.org/content/?page=article&journal=mca&volume=1&num=1&ordnum=4> (accessed 2026-04-13).
53. Mouawad, F.; Homsy, F.; Geara, F.; Mina, R. Predicting compressive strength of sustainable concrete using Mach. Learn. and artificial neural networks. *Constr. Mater.* **2025**, *5*, 56. DOI
54. Joo, C.; Park, H.; Kwon, H.; et al. Mach. Learn. approach to predict physical properties of polypropylene composites: application of MLR, DNN, and random forest to industrial data. *Polymers*. **2022**, *14*. DOI
55. Malashin I, Tynchenko V, Gantimurov A, et al. Boosting-based machine learning applications in polymer science: a review. *Polymers* **2025**, *17*, 499. DOI PubMed PMC
56. Loh, W. Y. Classification and regression trees. *WIREs. Data. Min. Knowl. Discov.* **2011**, *1*, 14-23. DOI
57. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5-32. DOI
58. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Statist.* **2001**, *29*, 1189-232. <https://www.jstor.org/stable/2699986> (accessed 2026-04-13).
59. Chen, T.; Guestrin, C. Xgboost: a scalable tree boosting system. In *KDD'16: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016, 785-94. DOI
60. Ke, G.; Meng, Q.; Finley, T.; et al. Lightgbm: a highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*; 2017, 30. <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>. (accessed 2026-04-13).
61. Dorogush, A. V.; Ershov, V.; Gulin, A. CatBoost: gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363. DOI
62. Schölkopf, B.; Smola, A. J. Learning with kernels: support vector machines, regularization, optimization, and beyond; MIT press Cambridge, MA, USA.: 2002. <https://books.google.com/books?id=y8ORL3DWt4sC&printsec=frontcover&hl=zh-CN#v=onepage&q&f=false> (accessed 2026-04-13).
63. Wu, S.; Zhang, H.; Wang, Y.; et al. Concentration prediction of polymer insulation aging indicator-alcohols in oil based on genetic algorithm-optimized support vector machines. *Polymers*. **2022**, *14*, 1449. DOI PubMed PMC
64. Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A. J.; Vapnik, V. Support vector regression machines. In *Advances in neural information processing systems*, 1996, 9, 155-61. DOI
65. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE. Trans. Inform. Theory.* **1967**, *13*, 21-7. DOI
66. Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175-85. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.