
Interpretable Machine Learning Reveals Synergy-Gain Windows and Dual-Objective Mix-Proportion Boundaries for Compressive Strength and Peak Strain in Hybrid Steel–PVA Fiber-Reinforced Concrete

[Maojun Liu](#)*, [Junwen Chen](#), Shengkai Zhou

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0892.v1

Keywords: hybrid steel–PVA fiber-reinforced concrete; compressive strength; peak strain; interpretable machine learning; dual-objective mix proportion



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Interpretable Machine Learning Reveals Synergy-Gain Windows and Dual-Objective Mix-Proportion Boundaries for Compressive Strength and Peak Strain in Hybrid Steel–PVA Fiber-Reinforced Concrete

Maojun Liu *, Junwen Chen and Shengkai Zhou

Guilin University of Technology, Nanning 530001, China

* Correspondence: liumaojun@glut.edu.cn

Abstract

Hybrid steel–PVA fiber-reinforced concrete offers promise for enhancing both load-bearing capacity and deformation capacity. However, the coupled effects of fiber parameters and volume-fraction combinations on compressive strength (σ_c) and peak strain (ϵ_c) are still not fully understood. A unified, interpretable, and engineering-oriented quantitative framework is still lacking. This study compiled experimental data from 26 published literature, building a multi-source database consisting of 397 datasets for σ_c and 203 datasets for ϵ_c . Based on this database, a comprehensive analytical framework was proposed, including model prediction, SHAP-based interpretation, Monte Carlo marginalization, synergy gain window determination, and dual-objective mix proportion optimization. For σ_c prediction, LightGBM achieved the highest test-set R^2 (0.9783), whereas CatBoost showed more robust error control (MAE = 2.7409 MPa). CatBoost was therefore selected as the base model for the subsequent interpretation analysis. For ϵ_c prediction, Bayesian-optimized CatBoost achieved the best test performance ($R^2 = 0.9659$, MAE = 0.0218, RMSE = 0.0358), while the transfer-learning model reached a comparable accuracy level ($R^2 = 0.9650$). SHAP analysis revealed that σ_c is mainly governed by matrix mix-proportion factors and steel fiber volume fraction, whereas ϵ_c is more sensitive to S/B and PVA-related variables. The mean synergy-gain maps generated via Monte Carlo marginalization and two-dimensional grid evaluation further showed clear differences between the two targets. Positive synergy in σ_c was highly localized. Its maximum mean synergy gain was 4.7949 MPa at (Steel, PVA) = (1.875%, 2.000%). By contrast, ϵ_c exhibited a wider positive-synergy region, with a peak value of 0.0141629 at (0.38%, 1.62%). Therefore, the engineering output of this study is not a single optimal mix point. Instead, it is a set of candidate windows for different performance targets, together with boundary-risk identification and priorities for experimental validation.

Keywords: hybrid steel–PVA fiber-reinforced concrete; compressive strength; peak strain; interpretable machine learning; dual-objective mix proportion

1. Introduction

Steel and PVA fibers play complementary roles in toughening concrete. Steel fibers mainly improve post-cracking load resistance and energy absorption, whereas PVA fibers are more effective in controlling microcracks and enhancing ductility. When the two fibers are properly combined in terms of properties and dosage, they may work together to improve both strength and deformation capacity. However, hybrid fiber systems involve the interaction of many factors, such as matrix composition, fiber geometry, mechanical properties, and fiber content. Because of this complexity,

conventional experiments alone are often not sufficient to reveal the underlying interaction patterns in a systematic way.

In recent years, experimental studies on the mechanical performance of hybrid steel–PVA fiber-reinforced concrete have steadily increased. Zhou et al. [1] systematically investigated its uniaxial compressive constitutive behavior through an orthogonal experimental design and confirmed that the combined use of steel and PVA fibers can significantly improve failure behavior and energy dissipation. Liu et al. [2] reported that the hybrid-fiber effect is jointly governed by matrix mix proportion and fiber volume fraction. Abbas et al. [3] developed a compressive stress–strain constitutive model for hybrid steel–PVA fiber-reinforced concrete and quantified the effects of fiber parameters on peak stress and peak strain. Wu et al. [4] further demonstrated, from the perspective of flexural behavior, the synergistic advantages of hybrid fibers in post-cracking toughness and deformation capacity. These studies provide an important basis for understanding the reinforcing mechanisms of hybrid fibers. However, because of the limited experimental scope and sample size, it is still difficult to fully capture the nonlinear interactions arising from multi-factor coupling.

With the wider use of data-driven methods in building materials research, machine learning has become a common tool for predicting concrete performance and exploring hidden patterns in the data. Kang et al. [5] showed that tree-based models can capture the nonlinear behavior of fiber-reinforced concrete with good accuracy. Al-Shamasneh et al. [6] reported that ensemble learning performs robustly in predicting the compressive strength of steel fiber-reinforced concrete. Sofos et al. [7] and Cui et al. [8] further demonstrated that machine learning is applicable to complex material–structure problems involving FRP-confined concrete and related members.

Even so, most existing studies still use machine learning mainly as a black-box predictor. Much less attention has been given to questions that matter more in engineering practice, such as whether synergistic enhancement really exists, where it appears in terms of fiber dosage, and whether such patterns are backed by sufficient data.

Although machine learning has shown strong predictive ability in concrete research, most existing studies still treat the model as a black-box predictor. They pay much less attention to engineering questions that are more practically relevant, such as whether synergistic enhancement exists, in which volume-fraction ranges it occurs, and whether the observed pattern is supported by sufficient data.

For engineering design, it is essential to clarify how fiber reinforcement works. In this study, **synergy gain** is defined as the super-additive effect of hybrid steel–PVA fibers relative to single-fiber reinforcement. Thus, the key question is not only whether the prediction is accurate. It also includes whether synergistic fiber enhancement exists, in which volume-fraction ranges it appears, and under what data-support conditions it can be interpreted with reasonable confidence.

Therefore, this study does not treat machine learning simply as a black-box predictor. Instead, it combines machine learning with interpretable analysis, marginal-response modeling, and synergy-gain quantification to build a knowledge-extraction framework for engineering-oriented screening. Using compressive strength (σ_c) and peak strain (ϵ_c) as two target indicators of load-bearing capacity and deformation capacity, respectively, this study further proposes an overlay strategy for dual-objective synergy windows. This strategy provides support for preliminary mix screening and for setting priorities in experimental validation.

The main contributions of this study are as follows: (1) an interpretable machine-learning framework was established for the dual objectives of σ_c and ϵ_c ; (2) the mean synergy-gain surface of steel and PVA fibers was defined and quantified based on Monte Carlo marginalization; and (3) the synergy boundaries of σ_c and ϵ_c were identified, and a dual-objective mix-proportion screening logic was constructed.

Compared with most existing studies, which focus mainly on accuracy comparison, single-indicator interpretation, or single-performance prediction, the main novelty of this study lies in the identification of dual-objective synergy windows and in the design-oriented conclusions provided for material selection and mix-proportion optimization in engineering practice.

2. Materials and Methods

2.1. Dataset Construction and Definition of the Feature System

2.1.1. Data Sources and Sample Composition

The database used in this study was compiled from uniaxial compression test data on hybrid steel-PVA fiber-reinforced concrete collected from published literature and academic theses. After data cleaning, data standardization, and specimen-size normalization, the compressive-strength dataset (σ_c) contained 397 samples, whereas the peak-strain dataset (ϵ_c) contained 203 samples. The two datasets cover a range of conditions, including plain matrix mixtures, single-fiber mixtures, and hybrid steel-PVA fiber mixtures. This broad parameter coverage provides a sound basis for the subsequent modeling of nonlinear multi-factor relationships.

Table 1 summarizes the data sources and sample-count distribution of the core references for σ_c , so as to illustrate the source composition and sample coverage of the database. The ϵ_c samples were mainly collected from 12 studies, including Refs. [1,3], and [9–18], and are therefore not listed separately here.

Table 1. Core literature sources and sample counts of the σ_c dataset.

No.	Literature sources	Number of Specimens	Proportion of Dataset
1	Zhou et al.(2018) [1]	17	4.28%
2	Abbas et al.(2022) [3]	19	4.79%
3	Li .(2024)[9]	18	4.53%
4	Wang .(2016)[10]	20	5.04%
5	Sun et al.(2018) [12]	24	6.05%
6	Liu et al.(2019) [13]	19	4.79%
7	Liu et al.(2024) [14]	22	5.54%
8	Hao et al.(2025) [15]	27	6.80%
9	Zhong et al.(2019) [16]	17	4.28%
10	Zhao.(2020) [20]	16	4.03%
11	Gao.(2022) [21]	36	9.07%
12	Ju et al.(2022) [23]	17	4.28%
13	Zhang et al.(2023) [24]	16	4.03%
14	Wang et al.(2011) [27]	24	6.05%
15	Chen et al.(2025) [30]	14	3.53%
16	Sun et al.(2025) [32]	25	6.30%

Note: The full database contains 26 literature sources.

2.1.2. Definition of Input and Target Variables

The input variables included four categories: matrix mix-proportion parameters, mineral admixture and chemical admixture indicators, steel-fiber parameters, and PVA-fiber parameters. In addition to continuous variables, several binary indicator variables were introduced to distinguish the presence or absence of fly ash, silica fume, superplasticizer, and the two fiber types, thereby improving the model's ability to represent the mixed feature space. The target variables were compressive strength, σ_c (MPa), and peak strain, ϵ_c (%). The definitions of all variables are given in Table 2.

Table 2. Definition of input and target variables.

Feature category	Abbreviation	Physical meaning (unit)
Cementitious material	FA	Fly_Ash content(%)
Binary indicator (0/1)	FA_zero	Fly Ash Addition Marker
Cementitious material	SF	Silica_Fume content(%)
Binary indicator (0/1)	SF_zero	Silica_Fume Addition Marker
Mix-proportion parameter	W/B	Water to Binder Ratio(-)
Mix-proportion parameter	S/B	Sand to Binder Ratio(-)
Chemical admixture	SP	Superplasticizer content(%)
Binary indicator (0/1)	SP_zero	Superplasticizer Addition Marker
Steel-fiber parameter	D_STF	Steel Fiber Diameter (mm)
Steel-fiber parameter	L_STF	Steel Fiber Length(mm)
Steel-fiber parameter	f_STF	Steel Fiber Tensile Strength(MPa)
Steel-fiber parameter	E_STF	Steel Fiber Elastic Modulus(GPa)
Steel-fiber parameter	V_STF	Steel Fiber Volume Fraction(%)
Binary indicator (0/1)	STF_zero	Steel Fiber Addition Marker
PVA-fiber parameter	D_PVA	PVA Fiber Diameter (mm)
PVA-fiber parameter	L_PVA	PVA Fiber Length(mm)
PVA-fiber parameter	f_PVA	PVA Fiber Tensile Strength(MPa)
PVA-fiber parameter	E_PVA	PVA Fiber Elastic Modulus(GPa)
PVA-fiber parameter	V_PVA	PVA Fiber Volume Fraction(%)
Binary indicator (0/1)	PVA_zero	PVA Fiber Addition Marker
Target variable	σ_c	Compressive Strength(MPa)
Target variable	ϵ_c	Peak Strain(%)

Note: 1. The percentages of FA and SF are calculated based on the mass of cement; 2. The volume fractions of V_STF and V_PVA are determined by the total volume of concrete; 3. σ_c denotes the target compressive strength; 4. ϵ_c denotes the target peak strain.

2.2. Data Preprocessing, Internal-Validation Setting, and Analysis Boundaries

2.2.1. Data Cleaning, Size Normalization, and Statistical Characteristics

After extraction, the raw data collected from the literature were sequentially subjected to unit unification, outlier checking, duplicate verification, and missing-value screening. For compressive test results obtained from specimens of different shapes and sizes, predefined size-normalization rules were applied to convert them to a common reference basis. This procedure was used to reduce the influence of specimen-size differences on model training and synergy analysis. Table 3 presents the conversion coefficients used to normalize the mechanical properties of non-standard specimens to the standard size according to Eurocode 2 (BS EN 1992).

Table 3. Conversion coefficients for specimen size/shape normalization.

NO.	Specimen type	Non-standard dimensions	Conversion coefficient
1	Cube	70.7mm	0.95

2	Cube	100mm	0.97
3	Cube	150mm	1.00
4	Cylinder	100 (d) × 200 (h)	1.00
5	Cylinder	150 (d) × 300 (h)	1.05

Note: The coefficients were applied consistently to both compressive strength and peak strain records.

In terms of statistical characteristics, the key variables in the database span a broad range. For example, the water-to-binder ratio (W/B), fiber tensile strength, σ_c , and ϵ_c all exhibit large value ranges (see Table 4). This indicates that the database captures substantial differences in material performance under different mix proportions and fiber-parameter combinations. It also provides the necessary data basis for the subsequent identification of nonlinear effects and interactions.

Table 4. Statistical summary of key variables in the database.

Statistic	W/B	f_{STF} (MPa)	f_{PVA} (MPa)	σ_c (MPa)	ϵ_c (%)
Sample size	397	397	397	397	203
Maximum	0.55	3100.00	1850	170.00	1.30
Minimum	0.18	600.00	800	15.90	0.17
Mean	0.34	2055.47	1552.22	55.45	0.44
Median	0.32	2000.00	1560.00	50.29	0.35

Note: Sample size is reported as number of records.

2.2.2. Dataset Splitting and Validation Strategy

To ensure that the modeling procedure was reproducible and that the results were reliable, the compiled dataset was randomly split into training and test sets at a ratio of 8:2. A Kolmogorov-Smirnov (KS) test was then performed to check whether the two sets remained consistent in terms of the target-variable distribution. This helped confirm that the data split was reasonable. Figure 1 compares the CDFs of σ_c for the training and test sets and reports the corresponding KS test results.

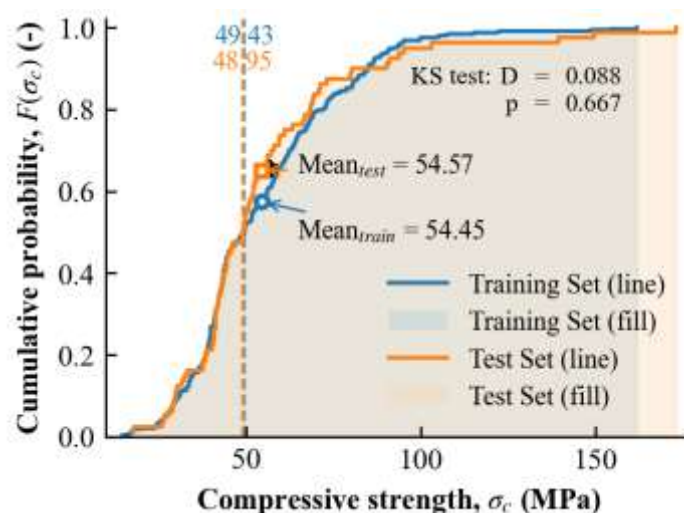


Figure 1. CDF Comparison of Compressive Strength (σ_c) for Train and Test Sets.

Figure 1 shows only a small difference in the σ_c distribution between the training and test sets, which satisfies the requirement of distributional balance for the subsequent internal validation and interpretation analysis. The ϵ_c dataset was split using the same stratified sampling strategy as that used for σ_c . Its KS statistic and distribution-consistency indicators were at the same confidence level,

indicating a high degree of distributional agreement between the training and test sets for both target variables. Therefore, the distribution-validation figure for ϵc is not presented separately.

2.2.3. Model Training and Synergy-Gain Calculation Framework

For the σc task, multiple regression models were compared, and the model with the best overall performance was selected as the base model for the subsequent interpretation analysis. For the ϵc task, transfer learning and hyperparameter optimization were introduced to improve modeling stability because of the limited sample size. The optimal models were then further analyzed through global SHAP importance and SHAP dependence plots.

To build synergy maps that can support engineering screening, this study used a Monte Carlo marginalization strategy. For a given steel-fiber volume fraction s and PVA-fiber volume fraction p , the other input variables were randomly sampled, and the model outputs were averaged to obtain the marginal mean response. The synergy gain was defined as $\Delta(s, p) = f(s, p) - f(s, 0) - f(0, p) + f(0, 0)$, and its marginal mean was written as $\bar{\Delta}(s, p)$. When $\bar{\Delta}(s, p) > 0$, positive synergy is considered to exist.

2.2.4. Five-Stage Framework and Technical Roadmap for the Quantitative Identification of Fiber Synergy

This study adopted a data-driven five-stage analytical framework to quantitatively identify fiber synergy. The overall workflow is shown in Figure 2.

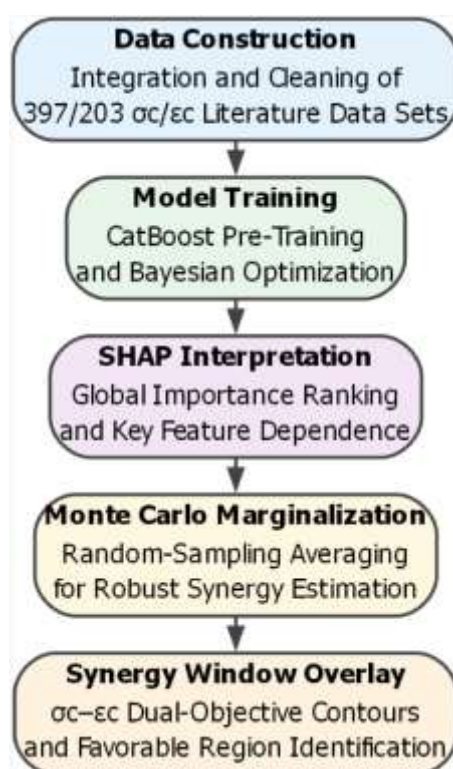


Figure 2. Technical roadmap of the five-stage machine learning framework for analyzing fiber-reinforced concrete performance.

The framework begins by integrating and cleaning 397/203 literature-based datasets to build a high-quality sample database. It then combines CatBoost modeling with SHAP-based interpretation to achieve accurate and interpretable prediction. Finally, by linking Monte Carlo marginalization with synergy-window overlay, it makes it possible to identify and visualize fiber synergy in a quantitative way. In this way, the framework offers an interpretable route for optimizing the mix design of fiber-reinforced concrete.

3. Results and Analysis

3.1. Model Performance Comparison and Base-Model Selection

3.1.1. Comparison of σc Models and Selection of the Base Model

Table 5 shows that, for the σc task, tree-based models performed markedly better than the linear model overall. This suggests that the relationship between compressive strength and the multidimensional input variables in hybrid steel–PVA fiber-reinforced concrete is strongly nonlinear. In the internal validation results, LightGBM achieved the highest test-set R^2 (0.9783), whereas CatBoost obtained the lowest MAE (2.7409 MPa). CatBoost also showed better robustness under limited-sample conditions and was more suitable for handling categorical features and supporting the subsequent interpretation analysis.

Because the goal of the subsequent analysis is not simply to identify the model with the highest test score, but to support SHAP interpretation, single-variable main-effect analysis, and two-dimensional synergy-gain calculation, CatBoost was chosen as the base model for the σc analysis. This choice preserves strong predictive performance while avoiding over-reliance on a single evaluation metric in model selection. To test its robustness, the dataset was subjected to 10 repeated random train–test splits. The results show that the performance variation of CatBoost remained within an acceptable range. Detailed metrics and variation analysis are given in Appendix A.1.

Table 5. Performance comparison of candidate models for σc prediction under the internal-validation scenario.

Model		R^2	MAE (MPa)	RMSE (MPa)
Multiple Regression	Linear	0.9121	5.9924	7.8404
	Random Forest	0.9703	2.9755	4.5608
	Extra Trees	0.9676	3.1663	4.7587
	XGBoost	0.9748	2.8136	4.2024
	LightGBM	0.9783	2.8633	3.8940
	CatBoost	0.9737	2.7409	4.2857

3.1.2. ϵc Model Development and Small-Sample Modeling Strategy

Compared with σc , the ϵc dataset is smaller and is therefore more sensitive to parameter selection and variations in data splitting. Based on this characteristic, three modeling strategies were compared in this study: baseline CatBoost, a transfer-learning model, and Bayesian-optimized CatBoost (see Table 6). The results show that all three methods achieved good predictive performance. Among them, Bayesian-optimized CatBoost performed best overall, while the transfer-learning model reached a comparable level of accuracy. The parameter settings of the transfer-learning model and the hyperparameter-optimization results are presented in Appendices A.2 and A.3.

This result suggests that cross-task transfer can be practically useful when the sample sizes of different performance indicators are unbalanced. Even so, the later analysis of ϵc , including its interpretation and synergy-map construction, is still based mainly on the Bayesian-optimized CatBoost model. The transfer-learning results are treated as supportive evidence, rather than as the sole basis for the core conclusions.

Table 6. Test-set performance comparison of candidate models for ϵc prediction.

Modeling strategy (test set)	R^2	MAE	RMSE
Baseline CatBoost	0.9575	0.0265	0.0399
Transfer-learning model	0.9650	0.0291	0.0363
Bayesian-optimized CatBoost	0.9659	0.0218	0.0358

3.2. Analysis of the σ_c Model Results and Synergy Mechanisms

3.2.1. Global Feature Importance and Ranking of Fiber-Related Variables

Figure 3 presents the global feature-importance ranking for the σ_c model. The results show that compressive strength is primarily driven by matrix mix proportions and fiber volume fractions. Among all variables, the W/B and V_STF contribute most significantly, indicating that matrix densification and steel fiber content are the key factors governing compressive performance. The importance rankings are broadly consistent between the training and test sets, suggesting that the main findings are relatively stable across the dataset. Detailed information on feature fluctuations and validation results is provided in Appendix A.4.

Figure 4 further focuses on fiber-related variables. In addition to volume fraction, steel fiber properties including tensile strength and length, and PVA-related geometric parameters also rank relatively high in importance, although they generally appear after the dominant matrix-related factors and fiber volume fractions. This result suggests that the effect of hybrid steel–PVA fibers on compressive strength is governed first by fiber addition and its volume fraction, whereas geometric and mechanical parameters mainly exert a secondary regulating effect within specific ranges (see Appendix A.5).

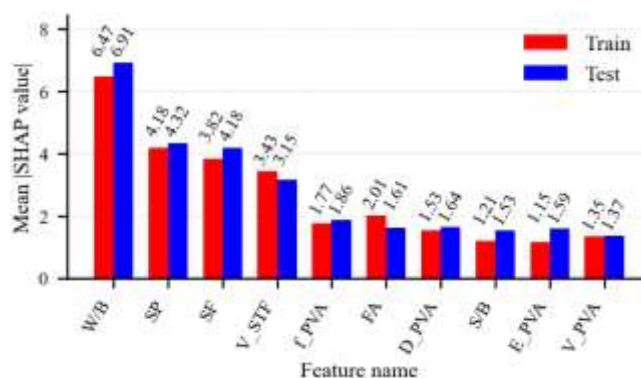


Figure 3. Global feature importance ranking for σ_c prediction (train vs. test).

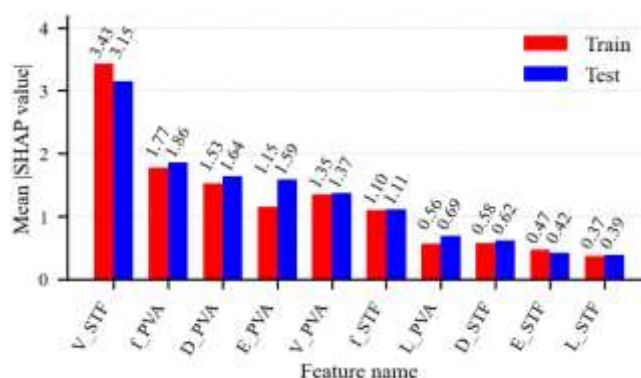


Figure 4. Fiber-related feature importance ranking for σ_c prediction.

3.2.2. SHAP Dependence Plots and Single-Fiber Main-Effect Curves

Figure 5 further reveals the nonlinear influence patterns of key variables on σ_c . Generally, increasing the steel fiber volume fraction leads to greater positive contributions to σ_c , though this enhancement exhibits diminishing marginal returns at higher volume fractions. Meanwhile, PVA fiber-related variables exhibit more erratic contributions to σ_c at low volume fractions, with their effects gradually stabilizing after the intermediate volume fraction range. This indicates that fiber reinforcement does not operate at a constant efficiency; instead, its effectiveness is collectively

governed by the uniformity of fiber dispersion, the quality of fiber-matrix interfacial bonding, and the compatibility between the fiber and matrix materials.

The single-fiber main-effect curves in Figure 6 show an overall trend consistent with the SHAP dependence plots. Steel-fiber addition alone is more likely to improve compressive strength, whereas the gain in σ_c from PVA-fiber addition alone is relatively limited. When the two fiber types coexist, some volume-fraction combinations show a more pronounced enhancement trend than single-fiber addition. However, this enhancement does not hold across the entire volume-fraction domain. In this study, $B = 100$ was adopted for the visualization and quantitative analysis of the synergy-gain surface, and its convergence analysis is provided in Appendix A.6.

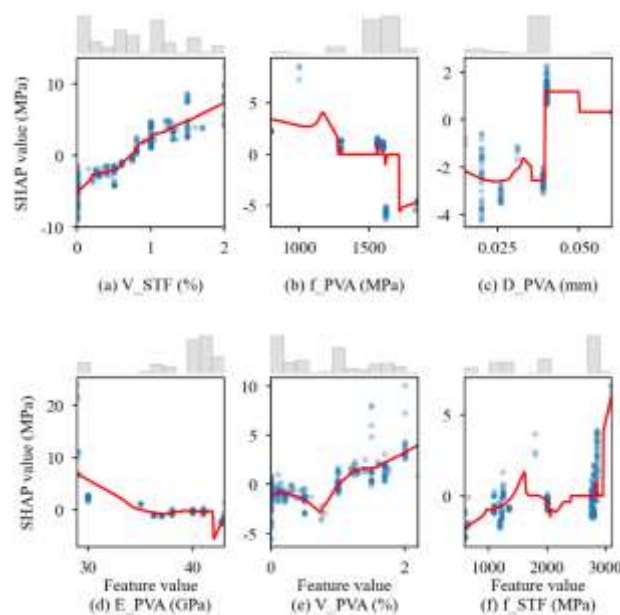


Figure 5. SHAP Dependence Plots and Marginal Histograms of Key Steel-Fiber and PVA-Fiber Variables for the σ_c Task.

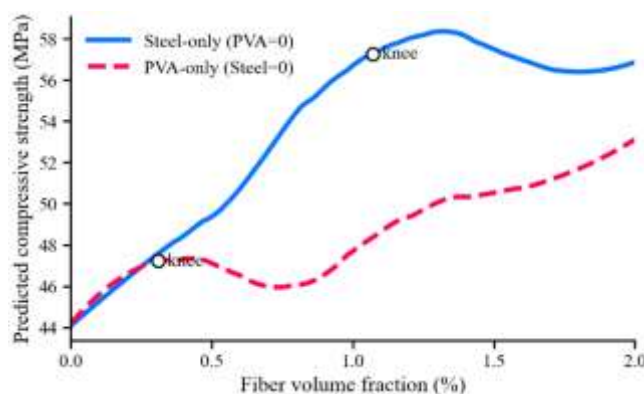


Figure 6. Single-fiber main-effect curves for σ_c under Monte Carlo marginalization ($B = 100$).

Figure 7 shows the mean synergy-gain surface of σ_c obtained from a 17×17 two-dimensional grid and Monte Carlo marginalization. The results indicate that the positive synergy between steel fibers and PVA fibers in compressive strength is clearly localized, rather than being universally present across the entire Steel-PVA volume-fraction plane. The positive-synergy region is mainly distributed near combinations with high steel-fiber and high PVA-fiber contents. This suggests that, at relatively high volume fractions, the two fiber types may enhance compressive performance through the combined effects of macro-scale bridging and microcrack restraint.

Figure 8 shows that the maximum mean synergy gain for σ_c over the whole domain is 4.794912 MPa, corresponding to (Steel, PVA) = (1.875%, 2.000%). However, the positive-synergy region occupies only about 1.7% of the domain. This result suggests that hybridization does not necessarily lead to a strength benefit. Its super-additive effect appears only within a limited range of fiber combinations.

By combining the optimal-gain results from Table 10 with the sample distributions in Figure 8, we find the maximum mean synergy gain of σ_c lies near the boundary of the sample-supported domain. This not only points to the high research value of high-steel-fiber and high-PVA-fiber combinations, but also reminds us to interpret these findings carefully in real engineering applications—we must pay close attention to workability, fiber dispersion, and the need for further experimental validation.

From an engineering perspective, the high-steel-fiber/high-PVA-fiber region is better treated as a potential testing zone for high-strength mixes, rather than a fixed point ready for direct practical recommendation. Priority should be given to combinations that fall within the positive-synergy region, remain reasonably far from the convex-hull boundary, and reside within the data support domain.

To validate the reliability and robustness of this synergy gain surface, we conducted grid resolution tests and robustness checks, which confirmed that fluctuations were kept within 2% and the core morphological features showed statistical consistency. The full validation process is detailed in Appendix A.7.

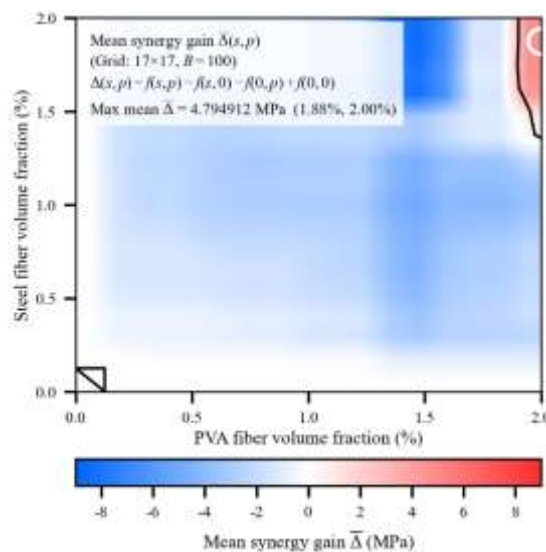


Figure 7. Mean synergy-gain surface $\Delta(s,p)$ for σ_c with the $\Delta=0$ boundary and the maximum point marked.

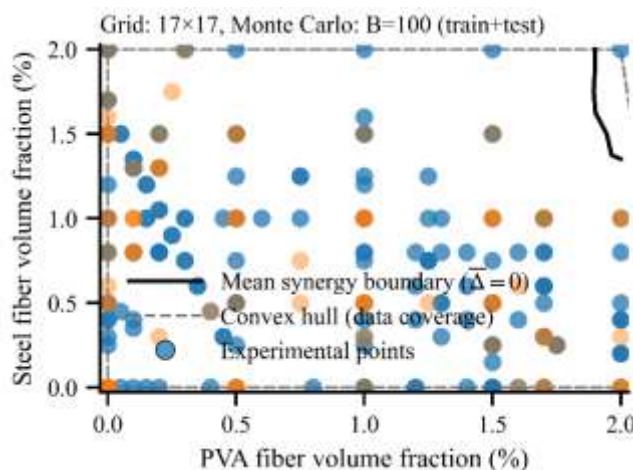


Figure 8. Overlay of the σc synergy boundary and the convex-hull-based data-support domain.**Table 10.** Quantitative summary of the σc mean synergy-gain surface.

Metric	Symbol/Setting	Value
Grid range(Steel× PVA)	s,p	[0.0,2.0]×[0.0, 2.0]%
Grid resolution	N×N,Δ	17×17,Δs= Δp=0.125%
Monte Carlo samples	B	100
Global maximum mean synergy gain	max Δ	4.794912MPa
Location of max mean synergy gain	(s*,p*)	(1.875%, 2.000%)
Positive synergy coverage (area share)	P($\Delta > 0$)	1.7%
Mean synergy gain within positive region	E[$\Delta \mid \Delta > 0$]	3.271014 MPa
Global mean synergy gain	E[Δ]	-2.117415 MPa

Note: $\Delta(s,p)=f(s,p)-f(s,0)-f(0,p)+f(0,0)$. $\Delta(s,p)$ is the Monte Carlo average over B=100 samples drawn from the combined train+test dataset (N=397), while varying only the fiber volume fractions on the grid.

3.3. Analysis of the ϵc Model Results and Synergy Mechanisms

3.3.1. Global Feature Importance and Ranking of Fiber-Related Variables

As shown in Figures 9 and 10, compared with the σc task, the importance ranking of the ϵc model exhibits a more differentiated pattern. In addition to some matrix mix-proportion parameters, PVA-related variables become markedly more important in the ϵc task, with S/B, V_PVA, D_PVA, and f_PVA usually ranking among the more influential features. This indicates that peak strain is more sensitive to the microcrack-control capacity of PVA fibers, rather than being governed solely by the macro-scale bridging effect of steel fibers.

The importance distributions in the training and test sets are broadly consistent, which supports the interpretation analysis of ϵc within the current data-support domain. However, because the ϵc sample size is relatively small, the ranking results should emphasize the relative positions of the main controlling factors, rather than over-interpreting subtle differences between adjacent variables (see Appendix A.8).

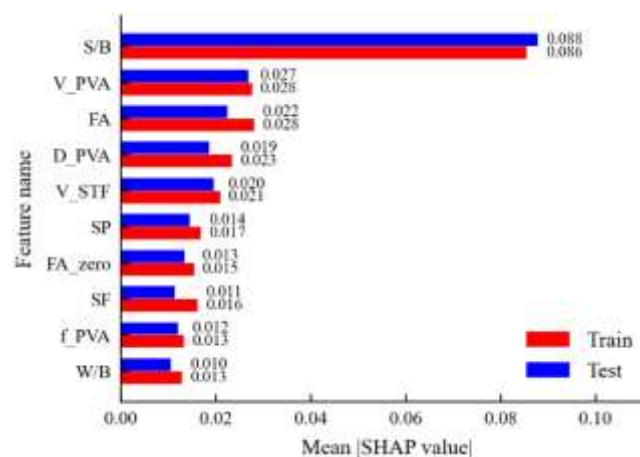
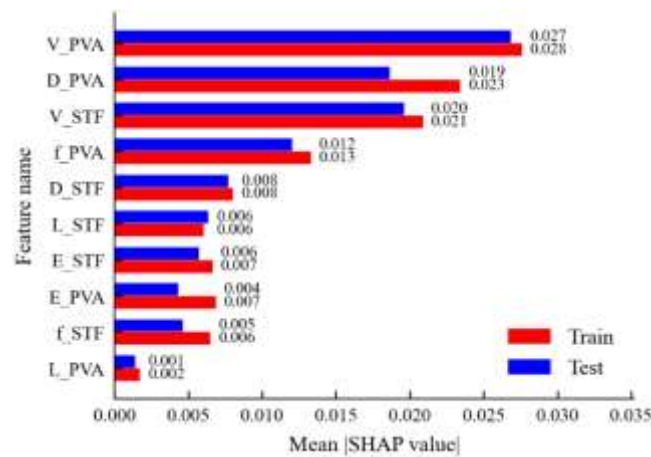


Figure 9. Global feature importance ranking for ϵ_c prediction (train vs. test).**Figure 10.** Fiber-related feature importance ranking for ϵ_c prediction.

3.3.2. SHAP Dependence Plots, Discrete-Level Support, and the ϵ_c Synergy-Gain Map

The SHAP dependence plots of the key variables for ϵ_c (Figure 11) show that the volume fraction of PVA fibers and related parameters make a more pronounced positive contribution to peak strain, whereas steel fibers mainly play a supporting bridging role during the later stages of crack development. This result is consistent with the underlying material mechanism. PVA fibers are more effective in suppressing the initiation and propagation of microcracks, and are therefore more critical to deformation capacity near the peak point. By contrast, steel fibers improve bridging capacity at the macrocrack stage and thus provide a complementary contribution to ductility enhancement.

It should be noted that some variables in the ϵ_c dataset have relatively few discrete levels and high shares of dominant levels (see Table 11). This may cause the SHAP dependence plots to exhibit step-like or locally fluctuating patterns in certain intervals. Therefore, when interpreting the key variables, this study considers the number of discrete levels, the shares of dominant levels, and the distribution of tail samples together, so as to enhance interpretive transparency.

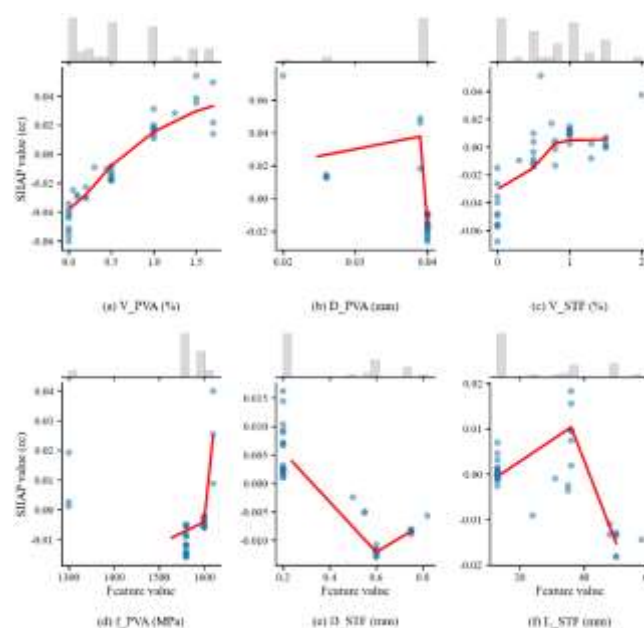
**Figure 11.** SHAP dependence plots of key variables for ϵ_c .

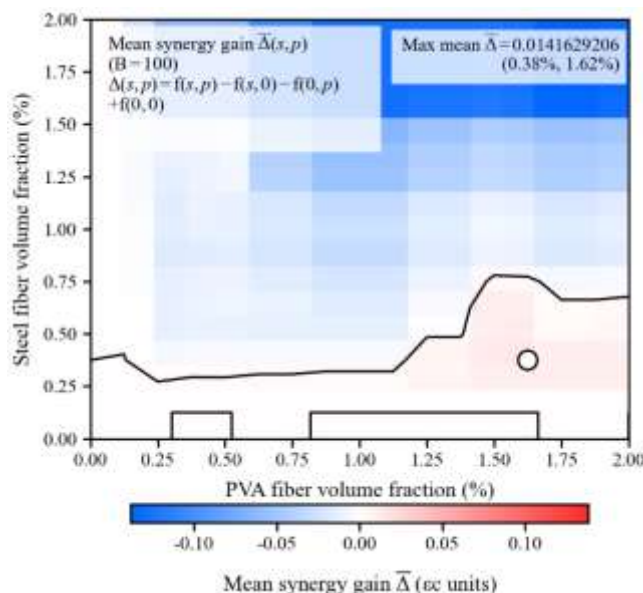
Table 11. Discrete Levels and Dominant-Level Shares of Key Fiber-Related Variables in the εc Interpretation (test set, n = 41).

Feature	Range (test)	Levels (test)	Top-1 share	Top-2 share
V_PVA(%)	0.00–1.70	11	22.0%	44.0%
D_PVA (mm)	0.020– 0.040	4	80.5%	90.2%
V_STF(%)	0.00–2.00	10	24.4%	46.3%
f_PVA (MPa)	1300–1620	4	53.7%	85.4%
D_STF (mm)	0.200– 0.820	6	56.1%	78.0%
L_STF (mm)	13–58	8	56.1%	70.7%

Note: Shares are reported for the test set to match the SHAP dependence plots (computed on X_{test}).

Based on the above single-variable SHAP dependence analysis, the mean synergy-gain map constructed for εc shows that its positive-synergy region is clearly wider than that of σc . The peak is mainly located near combinations with low-to-moderate steel-fiber content and moderate-to-high PVA-fiber content. Quantitative results show that the global maximum mean synergy gain of εc is 0.0141629, located at (Steel, PVA) \approx (0.38%, 1.62%), and that the positive-synergy region covers about 18% of the whole domain.

Combined with the window distribution, this suggests that ductility synergy is more likely to arise from the coordinated action of a moderate amount of steel fibers and a relatively high amount of PVA fibers at different stages of crack development. Compared with the σc window, the εc window is more suitable as a basis for ductility-oriented design (see Figure 12 and Table 12).



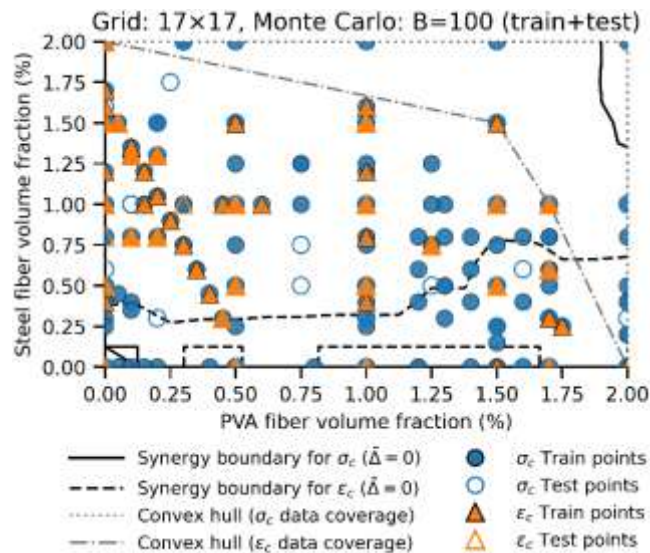


Figure 12. Mean synergy-gain surface and data-support overlay for ϵ_c : (a) mean synergy-gain heatmap; (b) overlay of the $\bar{\Delta} = 0$ boundary and the convex-hull-based data-support domain.

Table 12. Quantitative summary of the ϵ_c mean synergy-gain surface.

Statistic	Value
Monte Carlo samples (B)	100
Grid resolution	17 × 17 (0–2%×0–2%)
Max mean synergy gain, max $\bar{\Delta}$	0.0141629 (ϵ_c units)
Location of max $\bar{\Delta}$	Steel=0.38%, PVA=1.62%
Area fraction with $\bar{\Delta} > 0$	17.99%
Mean $\bar{\Delta}$ over $\bar{\Delta} > 0$ region	0.00412705
Mean $\bar{\Delta}$ over all grid points	-0.0332134

3.4. Implications of Dual-Objective Synergy: Trade-Offs Between Strength and Ductility and Mix-Proportion Boundaries

As shown in Figure 13, a comparison of the synergy windows of σ_c and ϵ_c on the Steel–PVA volume-fraction plane indicates that their positive-synergy regions do not fully overlap. This means that engineering design does not have a single optimal point that simultaneously satisfies all objectives. A more reasonable approach is to screen candidate regions according to performance constraints and the level of data support.

From an engineering perspective, if the primary objective is to improve load-bearing capacity, the candidate mixes are more likely to fall in the high-fiber-content region. However, greater attention should also be paid to reduced workability, difficulties in fiber dispersion, and the risk of boundary extrapolation. If ductility is the main concern, screening can instead focus on combinations with low-to-moderate steel-fiber content and moderate-to-high PVA-fiber content. The value of the overlay map of dual-objective synergy windows lies not in replacing experiments, but in providing a data-supported quantitative basis for preliminary mix screening and for setting priorities in experimental validation.

Based on the above differences, preliminary engineering mix screening can be carried out in three steps. First, determine whether the primary objective is load-bearing capacity, ductility, or a balance between the two. Second, screen candidate ranges within the corresponding positive-synergy window. Third, exclude combinations located near the boundary of the data-support domain, and subject the remaining combinations to experimental verification.

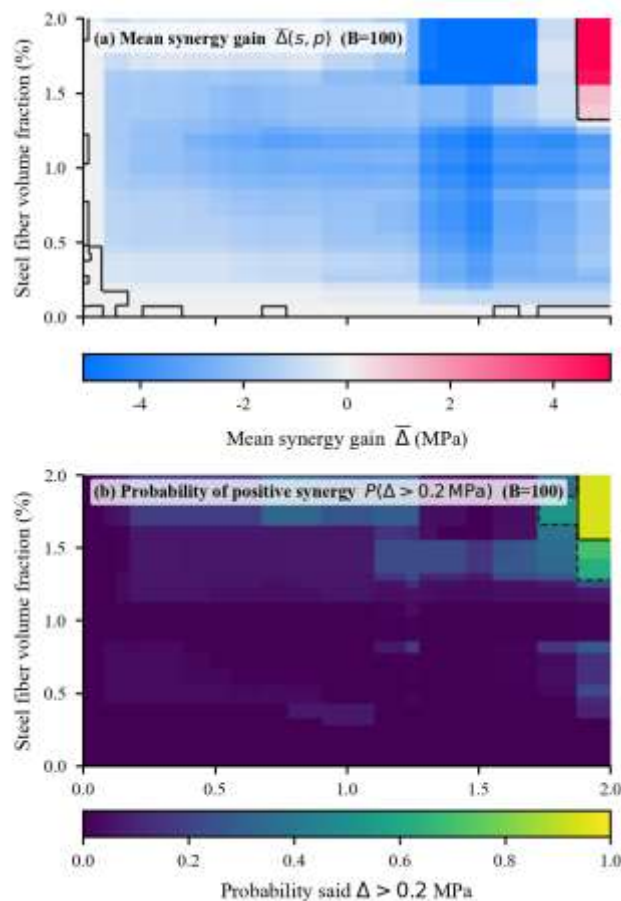


Figure 13. Overlay of the σ_c and ϵ_c Synergy Windows with Dual-Objective Contours.

4. Discussion

4.1. Model Performance and Analytical Positioning

As shown in Section 3, the random cross-validation results based on the available data indicate that both the σ_c and ϵ_c tasks achieved high fitting and predictive accuracy. This suggests that the data-driven models established from mix-proportion parameters, fiber geometric parameters, and fiber mechanical parameters can effectively capture the main nonlinear relationships in the compressive behavior of hybrid steel-PVA fiber-reinforced concrete. More importantly, the core value of this study lies not only in its high predictive accuracy, but also in transforming the prediction results into interpretable, screenable, and practically useful information that can directly support engineering applications, mix-proportion optimization, and experimental design.

Building on the analytical framework this model provides, we will further investigate how steel-PVA hybrid fibers modulate the compressive performance of concrete.

4.2. Interpretability of Key Variables and the Differentiated Roles of Steel and PVA Fibers

SHAP analysis shows that although σ_c and ϵ_c are both indicators of compressive behavior, they are governed by different dominant factors. In the σ_c task, matrix mix-proportion factors and steel fiber volume fraction are more important. In the ϵ_c task, S/B and several PVA-related variables carry greater influence. This difference indicates that the **load-bearing capacity** and **deformation capacity** of hybrid fiber-reinforced concrete are not controlled by the same set of variables in the same manner. The former depends more on the load-bearing skeleton of the matrix and the bridging capacity across macrocracks. The latter is more sensitive to microcrack control, fiber-matrix interfacial interaction, and the effect of matrix volumetric proportioning on deformation compatibility.

From the perspective of material mechanisms, the higher elastic modulus and tensile strength of steel fibers make them more effective in post-cracking bridging and in delaying the propagation of macrocracks. This is why they exert a more direct strengthening effect in the σ_c model. By contrast, PVA fibers are more advantageous in suppressing microcrack initiation, improving the continuity of crack propagation, and enhancing deformation accommodation near the peak point. This is broadly consistent with the findings of previous experimental studies [1–4]. In other words, steel fibers and PVA fibers do not merely offer redundant reinforcement. Instead, they participate in the compressive failure process at different scales. This distinction forms the basic physical basis of hybridization, rather than simple superposition.

Furthermore, both the SHAP dependence plots and the single-fiber main-effect curves indicate that fiber effects are strongly nonlinear. As the volume fraction increases, the strengthening effect does not continue to grow at a constant rate. Instead, it often shows diminishing marginal returns, plateauing, or even local fluctuations. This suggests that, in a multi-source literature-based dataset, the potential performance gains associated with higher fiber content may be simultaneously limited by factors such as fiber dispersion, workability, interfacial bonding, and matrix compatibility. Therefore, fiber optimization cannot be achieved simply by increasing fiber dosage. More importantly, it calls for pinpointing the parameter ranges in which the positive effects of fibers can be stably observed under the support of the available data.

4.3. Engineering Implications of Synergy-Gain Windows and Dual-Objective Trade-Offs

From the synergy-gain heatmaps generated via Monte Carlo marginalization and two-dimensional grid evaluation, we observe that the synergistic enhancement between steel fibers and PVA fibers does not hold across the entire volume-fraction domain, but instead shows clear regional characteristics. This insight carries important implications for engineering practice: hybrid fiber mixtures do not inherently outperform single-fiber systems or produce simple additive effects, and only specific fiber combinations can yield true super-additive benefits. Therefore, rather than adhering to the empirical assumption that combining steel and PVA fibers will necessarily improve performance, this study advocates for a window-based and condition-dependent design strategy.

For σ_c , the positive-synergy region is concentrated near combinations with high steel-fiber and high PVA-fiber contents, and its area share is very small, indicating that strength synergy is strongly localized. This means that if the primary engineering objective is to achieve higher load-bearing capacity, the formulations under consideration are likely to fall in the high-volume-fraction region. However, such regions are also often closer to the data boundary and more likely to be accompanied by reduced constructability, difficulties in fiber dispersion, and increased construction risk. Therefore, the interpretation of the strength-synergy window must consider both potential benefits and application risks, instead of fixating solely on the peak value.

In comparison, the positive-synergy region for ϵ_c is wider, and its peak is located near combinations with low-to-moderate steel-fiber content and moderate-to-high PVA-fiber content. This indicates that ductility synergy does not rely on an extremely high steel-fiber volume fraction. Instead, it is more likely to arise from the coordinated action of a moderate amount of steel fibers and a substantial volume of PVA fibers at different stages of crack development: the former provides the necessary macro-scale bridging capacity, whereas the latter improves microcrack control and deformation compatibility. This finding suggests that, in scenarios where ductility, energy dissipation, or peak-strain enhancement is the primary objective, a moderate rather than extreme steel-fiber content is prone to deliver stable benefits.

More importantly, the synergy windows of σ_c and ϵ_c do not fully overlap, which means that engineering design must inevitably address a trade-off between strength and ductility. A more reasonable strategy is to first define the minimum requirements for load-bearing capacity and deformation capacity according to the structural objective. Candidate points should then be prioritized within the dual-objective synergy region, while also remaining inside the data-support domain and relatively far from the convex-hull boundary. Combinations located near the boundary

should be confirmed through additional experiments. In this way, the role of the synergy-window map is not to replace experiments, but to help guide them in a more targeted and efficient manner.

Therefore, this study recommends window-guided selection rather than point-based selection. When a candidate combination lies well within the synergy region and remains distant from the convex-hull boundary, it should be prioritized in laboratory mixing trials. By contrast, if a combination exhibits favorable mechanical performance only in terms of its peak value but lies close to the boundary, it is an oriented case rather than being directly recommended as a target mix proportion.

4.4. Scope of Applicability, Limitations, and Future Work

Although this study established a relatively complete integrated workflow for prediction, interpretation, and synergy identification, its scope of applicability still needs to be clearly defined. First, the data were compiled from multiple published studies and academic theses. Although data standardization and specimen-size normalization were performed, differences in metadata may still exist across studies, including raw-material sources, curing conditions, loading rates, specimen preparation procedures, and testing equipment. These factors were not fully structured and incorporated into the model. Therefore, the patterns learned by the model should be understood, to some extent, as empirical regularities averaged over a multi-source database, rather than as an exact reproduction of a single experimental system.

Second, to ensure the reliability of the conclusions, all findings in this study are restricted to the current data-support range, and the scope of applicability will be further expanded through targeted experiments. At the same time, convex-hull coverage was used to identify high-confidence regions within the supported data domain. The conclusions on synergy effects in these regions can directly inform engineering design. By contrast, conclusions for samples near the convex-hull boundary should be regarded only as a basis for preliminary design and still require further experimental validation. Fiber-mix schemes located in such high-confidence regions can be directly applied in the production of concrete members and may help reduce trial-mix costs to some extent.

To further extend the current application boundary of this study, future work will proceed in three directions. First, additional experiments will be carried out in boundary regions and sparse mix-proportion intervals where data coverage is insufficient. The focus will be on verifying the mechanical stability of systems with high steel-fiber and high PVA-fiber contents, so as to provide more refined mix guidance for the production of concrete members. Second, mix-validation experiments using different batches of raw materials will be conducted to clarify the extent to which raw-material variability affects mix performance, thereby providing a quantitative basis for material substitution in engineering practice. Third, long-term performance data under different curing regimes will be added to establish performance-prediction models that better reflect field conditions and further enhance the engineering applicability of the present study.

From the perspective of engineering implementation, the most effective path for future improvement is not to develop more complex prediction models. Instead, it is to improve the transferability and practical usefulness of the conclusions in concrete-member production by supplementing validation experiments for boundary mix proportions and by introducing multidimensional constraints such as workability, durability, and cost.

5. Conclusions

Based on the multi-source experimental database, interpretable machine learning, and synergy-gain map analysis, this study systematically investigated the compressive strength and peak strain of hybrid steel-PVA fiber-reinforced concrete. The main conclusions are as follows:

(1) This study developed an interpretable machine learning framework to analyze the compressive strength (σ_c) and peak strain (ϵ_c) of hybrid steel-PVA fiber-reinforced concrete. The framework integrates performance prediction, mechanistic interpretation, and synergy window

identification into a unified analytical workflow, thereby providing a data-driven, interpretable technical approach for optimizing the mix design of such fiber-reinforced concretes.

(2) Using a random train-test split on our multi-source dataset, tree-based models consistently outperformed linear models. For the compressive strength prediction task, LightGBM achieved the highest R-squared at 0.9783, while CatBoost delivered the lowest mean absolute error of 2.7409 MPa. After comprehensively evaluating error control, prediction stability, and post-hoc interpretability, we selected CatBoost as the foundational model for subsequent compressive strength analysis.

(3) For the ϵc task, Bayesian-optimized CatBoost achieved the best test performance ($R^2 = 0.9659$, MAE = 0.0218, RMSE = 0.0358). The transfer-learning model reached a comparable accuracy level ($R^2 = 0.9650$), indicating that cross-task feature transfer can provide effective prior support for modeling performance indicators with limited sample sizes.

(4) SHAP analysis showed that σc is mainly governed by matrix mix-proportion factors and steel fiber volume fraction, whereas ϵc is more sensitive to S/B and PVA-related variables. This difference reflects the distinct fiber-action mechanisms underlying load-bearing capacity and deformation capacity.

(5) The mean synergy-gain maps derived from Monte Carlo marginalization show that the positive-synergy region for σc is strongly localized and mainly concentrated near combinations with high steel-fiber and high PVA-fiber contents, with a global maximum mean synergy gain of 4.794912 MPa. By contrast, the positive-synergy region for ϵc is wider and is mainly distributed in the range of low-to-moderate steel-fiber and moderate-to-high PVA-fiber combinations, with a peak value of 0.0141629. These results indicate that the effects of the two fiber types are not simply linearly additive, but show clear regionality and target dependence.

(6) The dual-objective synergy windows of σc and ϵc do not fully overlap. Therefore, engineering mix design is better guided by a hierarchical screening logic of performance target–synergy window–data-support domain. The core value of this study lies in providing an interpretable and visual quantitative tool for candidate-mix screening and experimental-priority setting within the current data-support range, rather than directly offering a single universally applicable mix proportion.

(7) From a practical engineering perspective, this study is better regarded as a candidate-window map plus validation-priority tool, rather than as a single-point mix recommender. Priority should be given to combinations located within the positive-synergy region and relatively far from the boundary of the data-support domain, so as to improve the reliability of trial mixing and validation.

Appendix A

Appendix A.1 Performance Variation under Repeated Random Splits

As shown in Table A1, the results of 10 repeated random splits indicate that the CatBoost model exhibits good stability. The σc model is more stable, whereas the ϵc model is more sensitive to random splitting. However, the overall conclusions remain consistent. The variation trajectories of the performance metrics under different random splits, together with their 95% confidence intervals, are presented in Figure A1.

Table A1. Performance stability of CatBoost models under repeated random train-test splits.

Property	Metric	Mean	SD	CV (%)
Compressive strength (σc)	R^2	0.9572	0.0097	1.01
Compressive strength (σc)	MAE	3.3136	0.2804	8.46
Compressive strength (σc)	RMSE	4.8106	0.7167	14.90
Peak strain (ϵc)	R^2	0.9124	0.0314	3.44
Peak strain (ϵc)	MAE	0.0421	0.0092	21.73
Peak strain (ϵc)	RMSE	0.0698	0.0162	23.17

Note: 1. Mean, standard deviation (SD), and coefficient of variation (CV) of R^2 , MAE, and RMSE across 10 repeated 8:2 random splits. 2. Abbreviations: R^2 , coefficient of determination; MAE, mean absolute error; RMSE, root mean square error; SD, standard deviation; CV, coefficient of variation.

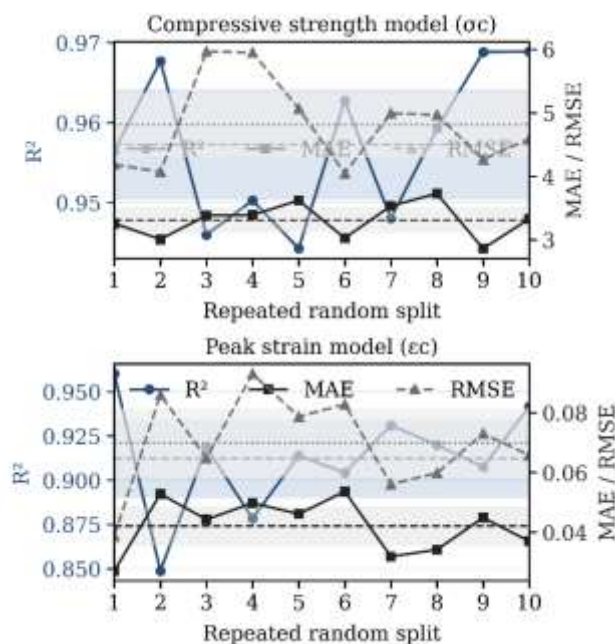


Figure A1. Performance Variation and Confidence Intervals of the Peak-Strain and Compressive-Strength Models under 10 Repeated Random Splits.

Appendix A.2 Parameter Settings of the Transfer-Learning Model

Table A2. Parameter settings of the transfer-learning model for ϵc prediction.

Parameter category	Parameter	Value
Pre-trained parameters	model	
	Number of fixed layers	726
	Input feature dimension	20
	Leaf-feature dimension	726
Training configuration	Source of pre-trained weights	None
	Regularization coefficient (alpha)	0.001
	Batch size	Full-batch training
Validation strategy	Maximum iterations (max_iter)	20000
	Early stopping	Not applicable
	Validation split ratio	0.2

Appendix A.3 Hyperparameter Settings and Optimization Results

Table A3. Search Space and Optimal Hyperparameters of the Bayesian-Optimized CatBoost Model.

Hyperparameter	Search range	Optimal value
learning_rate	(0.01, 0.05)	0.0452
depth	(4, 6)	6
iterations	(3000, 4500),	4018
l2_leaf_reg	(10, 20),	15
min_data_in_leaf	(10, 16),	14

random_strength	(0.2, 0.6),	0.5616
subsample	(0.8, 1),	0.9955
colsample_bylevel	(0.7, 0.9),	0.898

Appendix A.4 Stability of SHAP Importance Rankings

To examine the sensitivity of feature-interpretation results to random data splitting, this study repeated the 8:2 data split, model training, and SHAP ranking analysis under 10 different random seeds. The results show that, in both tasks, the rankings of the main features remain generally stable, and the top-ranked features exhibit only small fluctuations. The Kendall coefficient of concordance further indicates a high degree of consistency in the feature-importance rankings across the 10 repeated experiments, suggesting that the corresponding interpretation results are robust (see Table A4-1, Table A4-2, Figure A4-1, and Figure A4-2).

Table A4-1. Feature importance rank consistency analysis across 10 repeated random splits (σ).

Feature	Mean Rank	SD Rank	Best Rank	Worst Rank	Mean SHAP
W/B	1.00	0.00	1.0	1.0	6.2785
SP	2.10	0.32	2.0	3.0	3.4681
V_STF	2.90	0.32	2.0	3.0	2.9420
SF	4.20	0.42	4.0	5.0	2.2659
FA	5.80	1.14	4.0	7.0	1.6938
SF_zero	6.10	1.45	5.0	8.0	1.7252
D_PVA	7.00	1.25	5.0	9.0	1.5527
f_PVA	7.40	1.26	6.0	9.0	1.4286
S/B	9.20	1.03	7.0	10.0	1.2116
E_STF	10.20	1.81	8.0	14.0	1.0532
V_PVA	11.00	0.82	10.0	12.0	0.9666
E_PVA	11.60	1.26	9.0	13.0	0.8899
f_STF	12.60	0.70	11.0	13.0	0.7050
L_STF	14.50	0.71	14.0	16.0	0.3510
D_STF	14.70	0.95	13.0	16.0	0.3769
L_PVA	16.50	1.27	15.0	19.0	0.2018
FA_zero	17.00	1.15	15.0	19.0	0.1475
STF_zero	17.70	1.16	16.0	20.0	0.1230
SP_zero	19.10	0.74	18.0	20.0	0.0334
PVA_zero	19.40	0.84	18.0	20.0	0.0368

Table A4-2. Feature importance rank consistency analysis across 10 repeated random splits (ϵ).

Feature	Mean Rank	SD Rank	Best Rank	Worst Rank	Mean SHAP
S/B	1.00	0.00	1.0	1.0	0.0851
V_PVA	2.10	0.32	2.0	3.0	0.0320
FA	3.50	0.97	3.0	6.0	0.0225
V_STF	4.00	1.15	2.0	6.0	0.0190
f_PVA	6.00	2.00	4.0	10.0	0.0147
D_PVA	6.50	1.84	5.0	10.0	0.0136
SP	6.60	1.17	4.0	8.0	0.0139
SF	7.30	1.83	5.0	11.0	0.0126
W/B	9.70	2.00	7.0	13.0	0.0094
SF_zero	9.90	1.20	8.0	12.0	0.0099
D_STF	12.30	2.98	8.0	16.0	0.0069
FA_zero	12.50	2.32	9.0	16.0	0.0071
E_STF	12.70	1.57	10.0	15.0	0.0072
L_STF	14.10	1.79	11.0	17.0	0.0055
E_PVA	14.10	2.02	11.0	16.0	0.0055
f_STF	14.20	1.87	11.0	17.0	0.0058
STF_zero	16.80	1.40	14.0	19.0	0.0031
PVA_zero	18.10	0.88	17.0	20.0	0.0020
L_PVA	19.10	0.74	18.0	20.0	0.0010
SP_zero	19.50	0.71	18.0	20.0	0.0009

Note: Kendall's W = 0.9716; chi-square = 184.61; p = 0;

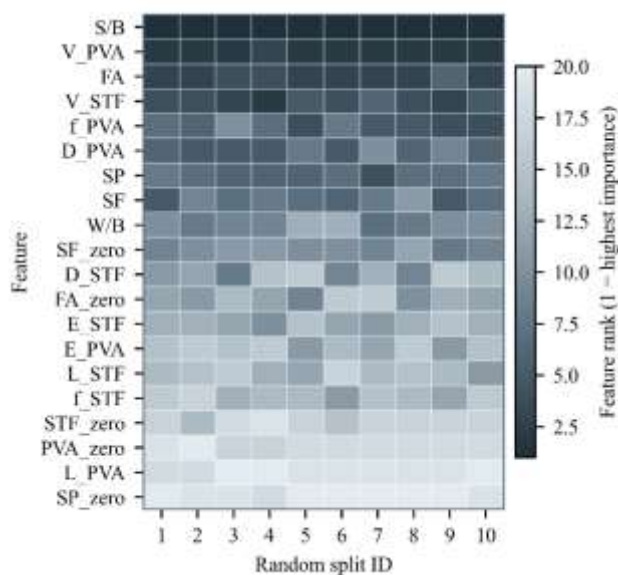


Figure A4-1. Feature-Importance Ranking Heatmap for the Compressive-Strength (σ_c) Model.

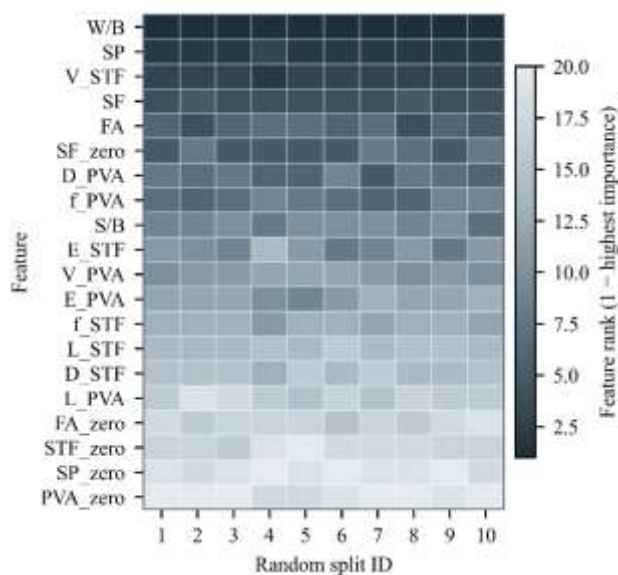


Figure A4-2. Feature-Importance Ranking Heatmap for the Peak-Strain (ϵ_c) Model.

Appendix A.5 Empirical support for SHAP dependence regions (density, discrete levels, and tail coverage) for compressive strength (σ_c) using train+test combined data.

Feature (unit)	K(rounded levels)	P5 / P50 / P95	Tail n (<P5 / >P95)	Top-3 levels (share%)
V_STF (%)	24	0/0.8/1.7	0/18	0 (21.4%); 1 (19.9%); 0.5 (11.6%)
f_PVA (MPa)	8	1300/1560/1850	16/0	1560 (38.8%); 1600 (33.2%); 1620 (9.8%)
D_PVA (mm)	7	0.02/0.04/0.04	2/10	0.04 (69.0%); 0.039 (9.8%); 0.02 (9.6%)

E_PVA (GPa)	11	30/40/42.8	13/1841 (32.0%); 40 (28.5%); 42.8 (9.8%)
V_PVA (%)	25	0/0.5/2	0/1 0 (22.2%); 1 (19.1%); 0.5 (9.6%)
f_STF (MPa)	13	600/2000/2850	0/8 2800 (22.2%); 2000 (16.1%); 2850 (11.6%)

Note: All features have $N = 397$; P5, P50, and P95 denote the 5th, 50th (median), and 95th percentiles. Tail n reports the number of observations below P5 and above P95. Top-3 levels show the most frequent rounded values and their sample shares.

Appendix A.6 Sensitivity to the Number of Monte Carlo Samples

This study conducted a convergence analysis on the number of Monte Carlo samples. The results show that when $B \geq 80$, the fluctuation of the single-fiber main-effect curves remains below 2%, which satisfies the requirement for statistical stability. To strike a balance between computational efficiency and result reliability, $B = 100$ was ultimately selected for the subsequent analysis.

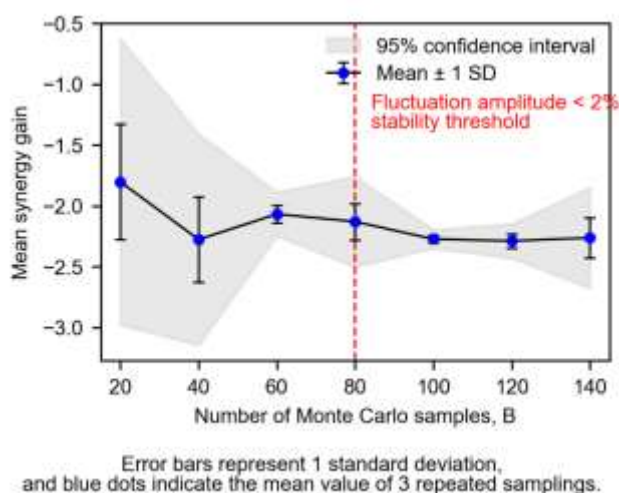


Figure A6. Convergence Analysis of the Number of Monte Carlo Samples.

Appendix A.7 Robustness Analysis of the Synergy-Gain Surface

A further local shape-robustness analysis was conducted for the σ_c synergy-gain surface using $B = 100$, as adopted in the main text. With the trained model, the combined train+test dataset, the two-dimensional grid range (0–2% \times 0–2%), and the grid resolution ($\Delta s = \Delta p = 0.125\%$) kept unchanged, pairwise comparisons were performed among the synergy-gain surfaces obtained with $B = 80, 100, \text{ and } 120$.

The results show that the overall shapes of the synergy-gain surfaces across varying B values are highly consistent (see Table A7-1). The Pearson correlation coefficients are all above 0.9989, and the Spearman rank correlation coefficients are all above 0.9965. This indicates that the synergy pattern identified in the main text is not sensitive to small variations in B around 100.

Combined with the convex-hull-based data-support domain shown in Figure 8, the adopted 0–2% grid is broadly consistent with the main supported region of the current σ_c dataset. Therefore, using $B = 100$ in the main text provides a good balance between computational efficiency and map robustness. Further results on the shape robustness of the synergy-gain surface under different Monte Carlo sample sizes are presented in Table A7-2.

Table A7-1. Local shape-robustness check of the σc synergy-gain surface around the adopted Monte Carlo sample size ($B = 100$).

Pair of B values	Surface correlation, Pearson r	Spearman correlation, ρ	rank
80 vs 100	0.9995 ± 0.0004	0.9974 ± 0.0010	
100 vs 120	0.9996 ± 0.0004	0.9980 ± 0.0008	
80 vs 120	0.9989 ± 0.0010	0.9965 ± 0.0018	

Note: The comparison was performed under the same trained model, combined train+test dataset, grid range (0–2% × 0–2%), and grid resolution (17×17, $\Delta s = \Delta p = 0.125\%$). Only the Monte Carlo sample size B was varied.

Table A7-2. shape-robustness of the σc synergy-gain surfaces around the adopted Monte Carlo sample size ($B = 100$).

Pair of B values	Surface correlation, Pearson r	Spearman correlation, ρ	rank	Positive-window IoU
80 vs 100	0.9995 ± 0.0004	0.9974 ± 0.0010		0.3842 ± 0.0550
100 vs 120	0.9996 ± 0.0004	0.9980 ± 0.0008		0.3703 ± 0.0571
80 vs 120	0.9989 ± 0.0010	0.9965 ± 0.0018		0.3721 ± 0.0646

Note: The comparison was conducted using the same trained model, combined train+test dataset ($N=397$), grid range (0–2% × 0–2%), and grid resolution (17×17, $\Delta s = \Delta p = 0.125\%$). Only the Monte Carlo sample size B was changed. In the present σc dataset, the convex-hull-based support domain coincided with the adopted full grid; therefore, the full-grid and support-domain statistics were numerically identical.

Appendix A.8 Empirical support for SHAP dependence regions (density, discrete levels, and tail coverage)

Feature (unit)	K(rounded levels)	Tail n		Top-3 levels (share%)
		P5 / P50 / P95	(<P5 / >P95)	
V_STF (%)	19	0/0.8/1.5	0/8	1 (18.2%); 0 (18.2%); 0.5 (12.8%)
f_PVA (MPa)	4	1300/1560/1620	0/0	1560 (40.4%); 1600 (31.5%); 1300 (14.3%)
D_PVA (mm)	3	0.03/0.04/0.04	10/0	0.04 (83.3%); 0.03 (11.8%); 0.02 (4.9%)
V_PVA (%)	18	0/0.5/1.7	0/6	1 (20.7%); 0 (19.2%); 1.7 (11.8%)
D_STF (mm)	6	0.2/0.2/0.75	0/9	0.2 (53.7%); 0.6 (20.2%); 0.75 (8.9%)
L_STF (mm)	8	13/13/50	0/9	13 (53.7%); 36 (12.3%); 50 (8.9%)

Note: All features have $N = 203$; P5, P50, and P95 denote the 5th, 50th (median), and 95th percentiles—Tail n reports the number of observations below P5 and above P95, Top-3 levels show the most frequent rounded values and their sample shares.

References

- Zhou Y, Xiao Y, Gu A, et al. Orthogonal experimental investigation of steel-PVA fiber-reinforced concrete and its uniaxial constitutive model. *Constr. Build. Mater.* 2019, 197, 615–625. <https://doi.org/10.1016/j.conbuildmat.2018.11.203>.
- Liu F, Ding W, Qiao Y. Experimental investigation on the tensile behavior of hybrid steel-PVA fiber reinforced concrete containing fly ash and slag powder. *Constr. Build. Mater.* 2020, 241, 118000. <https://doi.org/10.1016/j.conbuildmat.2020.118000>.
- Abbas YM, Hussain LA, Khan MI. Constitutive Compressive Stress-Strain Behavior of Hybrid Steel-PVA High-Performance Fiber-Reinforced Concrete. *J. Mater. Civ. Eng.* 2022, 34, 04021401. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0004041](https://doi.org/10.1061/(ASCE)MT.1943-5533.0004041).
- Wu J, Zhang W, Han J, et al. Experimental Study on the Flexural Performance of Steel-Polyvinyl Alcohol Hybrid Fiber-Reinforced Concrete. *Materials* 2024, 17, 3099. <https://doi.org/10.3390/ma17133099>.

5. Kang MC, Yoo DY, Gupta R. Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. *Constr. Build. Mater.* 2021, 266, 121117. <https://doi.org/10.1016/j.conbuildmat.2020.121117>.
6. Al-Shamasneh AR, Mahmoodzadeh A, Karim FK, et al. Application of machine learning techniques to predict the compressive strength of steel fiber reinforced concrete. *Sci. Rep.* 2026, 16, 1901. <https://doi.org/10.1038/s41598-025-16516-1>.
7. Sofos F, Papakonstantinou CG, Valasaki M, et al. Fiber-reinforced polymer confined concrete: data-driven predictions of compressive strength utilizing machine learning techniques. *Appl. Sci.* 2022, 13, 567. <https://doi.org/10.3390/app13010567>.
8. Cui R, Yang H, Li J, et al. Machine learning-based prediction of compressive strength in circular FRP-confined concrete columns. *Front. Mater.* 2024, 11, 1408670. <https://doi.org/10.3389/fmats.2024.1408670>.
9. Li W. Study on Mechanical Properties of Steel-PVA Hybrid Fiber Reinforced Cementitious Composites; Master's Thesis, Guangxi University, Nanning, China, 2024. <https://doi.org/10.27034/d.cnki.gxbgu.2024.000682>. (In Chinese).
10. Wang Z. Studies on Mechanical Performance of Polyvinyl Alcohol-Steel Hybrid Fiber Reinforced Cementitious Composites; Ph.D. Thesis, Tsinghua University, Beijing, China, 2016. <https://doi.org/10.7666/d.Y2908364>. (In Chinese).
11. Wang Z, Zhang J, Wang Q. Mechanical properties and crack width control of hybrid fiber reinforced ductile cementitious composites. *J. Build. Mater.* 2018, 21, 216–221. <https://doi.org/10.3969/j.issn.1007-9629.2018.02.010>. (In Chinese).
12. Sun L, Hao Q, Zhao J, Wu D, Yang F. Stress strain behavior of hybrid steel-PVA fiber reinforced cementitious composites under uniaxial compression. *Constr. Build. Mater.* 2018, 188, 349–360. <https://doi.org/10.1016/j.conbuildmat.2018.07.128>.
13. Liu W, Han J. Experimental Investigation on Compressive Toughness of the PVA-Steel Hybrid Fiber Reinforced Cementitious Composites. *Front. Mater.* 2019, 6, 108. <https://doi.org/10.3389/fmats.2019.00108>.
14. Liu W, Xu A, Han J. Experimental study on the compressive behavior of PVA-steel hybrid fiber reinforced cementitious composites. *J. Heilongjiang Univ. Technol. (Compr. Ed.)* 2024, 24, 121–128. <https://doi.org/10.3969/j.issn.2095-7262.2024.03.021>. (In Chinese).
15. Hao Q. Research on the Constitutive Model of Steel-PVA Hybrid Fiber Reinforced Cementitious Composites; Master's Thesis, Wenzhou University, Wenzhou, China, 2017. <https://doi.org/10.27655/d.cnki.ggwdu.2017.000126>. (In Chinese).
16. Zhong G, Zhou Y, Xiao Y. Study on the uniaxial stress-strain curve of steel-polyvinyl alcohol hybrid fiber concrete. *Eng. Mech.* 2020, 37 (Suppl. 1), 111–120. <https://doi.org/10.6052/j.issn.1000-4750.2019.04.S016>. (In Chinese).
17. Liu YN, Li H, Li HW. Experimental study and constitutive modeling of fine steel fiber/PVA hybrid cement-based composites under uniaxial compression. *Chin. Q. Mech.* 2021, 42, 317–325. <https://doi.org/10.15959/j.cnki.0254-0053.2021.02.011>. (In Chinese).
18. Kuang W, Tan Z, Li Y, Li X, Liu F. Study on the compressive behavior of steel-PVA fiber high-strength manufactured-sand concrete. *Guangzhou Archit.* 2025, 53, 71–77. <https://doi.org/10.19701/j.gjzj.2025.06.012>. (In Chinese).
19. Hu J. Study on the Mechanical Properties of Steel-Polyvinyl Alcohol Hybrid Fiber Reinforced Cementitious Composites; Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2023. <https://doi.org/10.27200/d.cnki.gkmlu.2023.000327>. (In Chinese).
20. Zhao X. Study on the Mechanical Properties of PVA-Steel Fiber Reinforced Cement-Based Materials; Master's Thesis, Harbin Institute of Technology, Harbin, China, 2020. <https://doi.org/10.7666/d.Y3567890>. (In Chinese).
21. Gao C. Experimental Study on Mix Proportion and Material Properties of PVA-Steel Hybrid Fiber Reinforced Cementitious Composites; Master's Thesis, Lanzhou University of Technology, Lanzhou, China, 2022. <https://doi.org/10.27089/d.cnki.glzte.2022.000265>. (In Chinese).
22. Sree KSS, Koniki S. Mechanical Properties of PVA & Steel Hybrid Fiber Reinforced Concrete. *E3S Web Conf.* 2021, 309, 01174. <https://doi.org/10.1051/e3sconf/202130901174>.

23. Ju Y, Zhu M, Zhang X, et al. Influence of steel fiber and polyvinyl alcohol fiber on properties of high performance concrete. *Struct. Concr.* 2022, 23, 1687–1703. <https://doi.org/10.1002/suco.202100775>.
24. Zhang X, Wang B, Ju Y, et al. Experimental Study and New Model for Flexural Parameters of Steel–PVA High-Performance Fiber–Reinforced Concrete. *J. Mater. Civ. Eng.* 2023, 35, 04023016. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0004736](https://doi.org/10.1061/(ASCE)MT.1943-5533.0004736).
25. Sanchayan S, Foster SJ. High temperature behaviour of hybrid steel–PVA fibre reinforced reactive powder concrete. *Mater. Struct.* 2016, 49, 769–782. <https://doi.org/10.1617/s11527-015-0537-2>.
26. Xu Q, Jiang X, Zhang Z, et al. Experimental study on residual mechanical properties of steel-PVA hybrid fiber high performance concrete after high temperature. *Constr. Build. Mater.* 2025, 458, 139735. <https://doi.org/10.1016/j.conbuildmat.2024.139735>.
27. Wang J. Experimental Study on the Effects of PVA Fiber and Steel Fiber on the Fracture Properties of High-Performance Fiber-Reinforced Cementitious Composites; Master's Thesis, Beijing Jiaotong University, Beijing, China, 2011. <https://doi.org/10.7666/d.Y1963131>.
28. Zhang P, Deng R, Hu J, Wu L, Tao Z. Flexural performance of steel–PVA hybrid fiber engineered cementitious composites. *Bull. Chin. Ceram. Soc.* 2023, 42, 3125–3134. <https://doi.org/10.16552/j.cnki.31-2050.2023.09.025>. (In Chinese).
29. Ding Y. The Shock Compression Dynamic Performance Experimental Study of Steel and PVA Hybrid Fiber Reinforced Cement Matrix Composites; Master's Thesis, South China University of Technology, Guangzhou, China, 2014. <https://doi.org/10.27008/d.cnki.gsgdu.2014.000093>. (In Chinese).
30. Chen G, Lv M, Zhu H, et al. Towards compressive and tensile strengths of hybrid steel and PVA fibre-reinforced cementitious composites: Experimental and analytical. *Case Stud. Constr. Mater.* 2025, 22, e04301. <https://doi.org/10.1016/j.cscm.2025.e04301>.
31. Li S, Ding D, He S, et al. Research on fracture performance of steel–PVA hybrid fiber high-strength manufactured-sand concrete. *Build. Struct.* 2025, 55, 47–54. <https://doi.org/10.19701/j.jzjg.2025.05.008>. (In Chinese).
32. Sun J, Zhao Y, Li L, Tian L. Research on the influence of steel–PVA fiber volume fraction on the mechanical properties of concrete. *Concrete* 2025, (8), 96–103. <https://doi.org/10.19701/j.1002-3550.2025.08.019>. (In Chinese).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.