Journal of Sustainability, Policy, and Practice

2026 2nd International Conference on Digital Intelligence and Computing Technologies (DICT 2026)

Article

A Comparative Empirical Study of Interpretable Regression and Clustering Methods for Quantifying Dental Composite Formulation-Performance Relationships

Pei-ting Chung [1,*]

[1] Chemical and Biomolecular Engineering, University of California, Irvine, CA, USA
* Correspondence: Pei-ting Chung, Chemical and Biomolecular Engineering, University of California, Irvine, CA, USA

Abstract: Dental resin composites are the most widely used direct restorative materials, yet their formulation optimization remains predominantly empirical due to the complex, nonlinear interactions among resin monomers, fillers, and processing parameters. This paper presents a comparative empirical study evaluating six regression algorithms and three clustering methods for quantifying formulation-performance relationships in dental composites. Using a curated dataset of 233 dental composites spanning 17 formulation attributes and 7 mechanical and physical performance indicators, we benchmark Ridge Regression, Lasso, Random Forest, XGBoost, Support Vector Regression, and Gaussian Process Regression under nested five-fold cross-validation. Clustering analysis via K-means, DBSCAN, and agglomerative hierarchical clustering identifies distinct formulation archetypes with characteristic performance profiles. SHAP-based interpretability analysis reveals that filler loading, the BisGMA-to-TEGDMA monomer ratio, and degree of conversion collectively account for over 58% of total SHAP attribution across all target-parameter pairs. XGBoost achieves the highest average coefficient of determination (R^2 = 0.891) across seven performance targets, while Gaussian Process Regression provides a principled uncertainty framework that may help guide future experimental design. The identified formulation parameter influence weights offer materials engineers a quantitative basis for prioritizing experimental variables, reducing trial-and-error costs in dental composite development. These findings support the Materials Genome Initiative's objective of accelerating materials screening through data-driven methods.

Keywords: dental composite; formulation-performance relationship; interpretable regression; materials informatics

Received: 30 April 2026
Revised: 14 June 2026
Accepted: 28 June 2026
Published: 02 July 2026

1. Introduction

1.1. Background and Motivation

Dental resin composites serve as the primary direct restorative material in contemporary clinical dentistry, with global market demand exceeding several billion units annually. The mechanical and physical performance of these composites depends on intricate interactions among resin matrix composition, filler characteristics, coupling agents, and curing conditions. Resin matrices typically consist of dimethacrylate monomers such as bisphenol A-glycidyl methacrylate (BisGMA), urethane dimethacrylate (UDMA), and triethylene glycol dimethacrylate (TEGDMA) at varying ratios, while fillers range from micro-scale barium glass and silica particles to nano-scale zirconia-silica clusters. Achieving simultaneously high flexural strength, low volumetric

shrinkage, and adequate wear resistance requires navigating a multi-objective formulation landscape where improving one property often degrades another. Traditional formulation development relies on iterative experimental campaigns that vary one or two parameters at a time, a process that is both time-consuming and unlikely to capture the higher-order interactions among formulation variables that govern these trade-offs. The Materials Genome Initiative has called for data-driven approaches to accelerate materials discovery and optimization, and dental composites represent an ideal candidate domain where materials informatics could complement experimental workflows [1].

1.2. Research Gap and Contributions

1.2.1. Problem Definition

Recent work has demonstrated the feasibility of applying machine learning to dental materials data. Li et al. achieved R^2 values between 0.927 and 0.998 when predicting flexural strength of CAD/CAM resin composite blocks using five regression algorithms with SHAP-based interpretation [2]. Wang et al. applied classification methods to identify chemical factors governing microtensile bond strength of dental adhesives, highlighting pH and 10-MDP concentration as critical determinants [3]. While these studies confirm that machine learning can extract meaningful patterns from dental materials data, no prior work has systematically compared regression, clustering, and interpretability methods together on a unified dental composite dataset. Existing reviews of explainable ML in materials science have noted that systematic comparisons between SHAP and LIME in materials-specific contexts remain limited, leaving practitioners with insufficient evidence-based guidance for method selection [4]. The question of which analytical approach is most appropriate under the specific data conditions of dental composites --- small sample sizes, mixed feature types, significant missingness --- remains unanswered.

1.2.2. Scope and Contributions

This study addresses the gap through three contributions. We conduct a controlled comparison of six regression methods under identical evaluation protocols on 233 dental composites with 17 formulation attributes and 7 performance targets. We apply three clustering algorithms to identify formulation archetypes and characterize their distinct performance signatures. We deploy SHAP and LIME interpretability analyses to produce formulation parameter influence weight maps that quantify each variable's contribution to each performance indicator, providing actionable guidance for experimental design. The study scope is deliberately constrained to existing, well-understood algorithms rather than novel architectures, as the primary contribution lies in rigorous evaluation and domain-specific insight extraction within a data-scarce dental materials context.

2. Related Work

2.1. Data-Driven Methods for Dental and Polymer Composites

2.1.1. Dental Composite Property Prediction

Paniagua et al. curated a dataset of 233 dental composites from over 200 publications, encompassing 17 formulation attributes and 7 performance outcomes including flexural modulus, flexural strength, volumetric shrinkage, and shrinkage stress [5]. They trained nine machine learning algorithms and identified filler loading, BisGMA content, and degree of conversion as dominant predictive features. Han et al. extended the data-driven paradigm to broader polymer composites, using 1,774 experimental data points to predict seven properties with XGBoost achieving an average R^2 of 0.95, and demonstrated that SHAP analysis could reveal causal composition-property relationships [6]. Pruksawan et al. combined active learning with Bayesian optimization to optimize epoxy adhesive strength from only 32 initial samples, achieving a 13% strength improvement over the training maximum and demonstrating the viability of machine learning-guided formulation optimization under severe data constraints [7].

2.1.2. Reviews and Broader Context

Chen and Gu reviewed ML applications across composite materials, documenting the use of convolutional neural networks for microstructure-property relationships and inverse design, and establishing a taxonomy of ML approaches for composite material design that distinguishes between forward prediction, inverse optimization, and transfer learning paradigms [8].

2.2. Regression Algorithms and Feature Engineering

Chen and Guestrin introduced XGBoost, a scalable tree boosting algorithm with sparsity-aware learning and weighted quantile sketch for approximate splits, which has become the most widely adopted ML method for tabular materials property prediction due to its regularization capabilities and native handling of missing values [9]. Xie and Grossman proposed Crystal Graph Convolutional Neural Networks that achieve density functional theory-level accuracy for eight material properties while providing interpretability through decomposable atom-level feature contributions [10]. For tabular data with limited samples, Kim et al. developed CARTE, a pretraining and transfer learning approach for heterogeneous tabular datasets that addresses the schema-mismatch problem common when aggregating formulation data from multiple sources [11]. On the feature engineering side, Ouyang et al. introduced SISSO, a compressed-sensing-based method that constructs interpretable analytical descriptors from primary features through systematic mathematical operations, producing compact formulas that relate physical properties to material composition [12]. Yu and Liu proposed FCBF, a fast correlation-based filter for high-dimensional feature selection that identifies relevant features while removing redundancy, a foundational method applicable when compositional variables outnumber available samples [13].

2.3. Interpretable Machine Learning for Materials Science

Lundberg and Lee unified six existing explanation methods under the SHAP framework, grounding feature attribution in game-theoretic Shapley values to provide consistent, locally accurate explanations of individual predictions [14]. SHAP's additive feature attribution property guarantees that the sum of individual feature contributions equals the difference between the prediction and the expected value, a property that makes it particularly suitable for decomposing dental composite performance predictions into interpretable formulation parameter contributions. The ability to generate both global importance rankings and per-sample explanations supports both formulation screening and targeted optimization of individual compositions.

3. Experimental Setup

3.1. Data Sources and Preprocessing

The primary dataset used in this study is the dental composite performance prediction dataset curated by Paniagua et al., which compiles formulation and performance data for 233 dental composites from 321 candidates reported across more than 200 peer-reviewed publications. Each composite is described by 17 formulation attributes spanning monomer composition (BisGMA, TEGDMA, UDMA, and BisEMA concentrations and molecular weights), initiator type and concentration, degree of conversion, filler characteristics (type, loading in weight percent, particle size, and shape), and physical properties (viscosity, density, and refractive index). Seven performance targets are recorded: flexural modulus, flexural strength, compressive strength, volumetric shrinkage, shrinkage stress, wear depth, and depth of cure. Table 1 presents descriptive statistics for the seven performance targets.

Table 1. Descriptive Statistics of Performance Targets (N = 233)

Target Variable	Unit	Mean	Std	Min	Max	% Missing

Flexural Modulus	GPa	9.42	3.18	2.10	18.70	12.4
Flexural Strength	MPa	112.6	38.5	28.0	215.0	8.2
Compressive Strength	MPa	278.3	62.4	125.0	420.0	18.1
Volumetric Shrinkage	%	2.84	1.12	0.90	6.50	15.9
Shrinkage Stress	MPa	3.21	1.45	0.80	7.60	22.7
Wear Depth	μm	48.3	22.7	8.0	128.0	31.8
Depth of Cure	mm	3.15	0.82	1.20	5.80	14.6

Data source: Paniagua et al. (2025), Journal of Dental Research, 104(5), 513--521.

Missing values were imputed using iterative imputation with a Bayesian Ridge estimator, which treats each feature with missing entries as a function of other features in a round-robin fashion. Critically, the imputer was fitted independently within each outer cross-validation fold using only the training partition, and target variables were excluded from the imputation predictors to prevent information leakage. Each of the seven targets was modeled separately, with samples filtered to those having non-missing values for the specific target under evaluation. In a preliminary comparison against K-nearest neighbor imputation on the complete-case subset (n = 94), iterative imputation consistently achieved lower reconstruction error across features, with relative improvements of approximately 8--12%.

3.2. Feature Engineering

3.2.1. Categorical and Interaction Feature Construction

Filler type was encoded using one-hot vectors spanning seven categories (silica, barium glass, strontium glass, zirconia-silica, ytterbium fluoride, prepolymerized resin filler, and mixed). Monomer type indicators were similarly encoded. Beyond these direct encodings, six domain-guided engineered features were constructed to capture known synergistic effects in dental composite formulation: (1) filler loading × mean particle size, to reflect the joint effect of filler content and particle scale; (2) the BisGMA-to-TEGDMA concentration ratio, to capture the balance between viscosity and conversion-related shrinkage; (3) initiator concentration × degree of conversion, to reflect curing efficiency; (4) filler loading × degree of conversion, to represent the interaction between reinforcement level and polymerization extent; (5) UDMA-to-BisGMA ratio, to reflect variation in matrix flexibility and viscosity; and (6) filler loading × BisGMA/TEGDMA ratio, to capture the combined effect of inorganic fraction and resin composition on composite behavior. Table 2 summarizes the feature sets evaluated.

Table 2. Feature Set Configurations

Feature Set	Description	Dimensionality
Raw	Original 17 attributes with one-hot encoding	28

Engineered	Raw + 6 domain-guided interaction terms	34
MI-Selected	Top features by mutual information (threshold > 0.05)	18
FCBF-Selected	Features selected by fast correlation-based filter	15

3.2.2. Feature Selection and Dimensionality Reduction

Mutual information-based selection retained 18 features that exceeded a relevance threshold of 0.05 nats for flexural strength, the most commonly reported target. The FCBF algorithm removed redundant features by evaluating predominant correlations, yielding a compact set of 15 features. Principal component analysis was applied to the engineered feature set for visualization and as input to clustering algorithms, with the first three principal components capturing 62.3% of total variance. The engineered feature set (34 dimensions) was used as the default input for regression methods, as preliminary experiments showed it yielded 3--7% higher R^2 compared to raw features across all targets, while the FCBF-selected set (15 dimensions) provided competitive performance with substantially reduced computational cost.

3.3. Regression and Clustering Methods

3.3.1. Regression Configuration

Six regression methods were evaluated. Ridge Regression and Lasso Regression served as linear baselines, with regularization strength α tuned via inner cross-validation over the range $[10^{-4}, 10^{2}]$. Random Forest was configured with 100 to 500 trees, maximum depth from 5 to 20, and minimum samples per leaf from 2 to 10. XGBoost used 50 to 300 boosting rounds, learning rate from 0.01 to 0.3, maximum depth from 3 to 10, and L2 regularization λ from 0.1 to 10. Support Vector Regression employed a radial basis function kernel with $C \in [0.1, 100]$ and $\gamma \in [10^{-3}, 10]$. Gaussian Process Regression used a Matérn 5/2 kernel with automatic relevance determination, where length-scale bounds were set per feature. All hyperparameters were tuned via Bayesian optimization with 50 iterations within the inner fold of a nested five-fold cross-validation protocol, inspired by the Matbench benchmark and adapted to our literature-curated tabular dataset [15].

3.3.2. Clustering Configuration

Three clustering algorithms were applied in the PCA-reduced space (three components, 62.3% variance explained). K-means was evaluated for $k \in \{2, 3, 4, 5, 6, 7\}$, with optimal k selected by maximizing the silhouette coefficient. DBSCAN used ε values from 0.3 to 1.5 (in increments of 0.1) and minimum samples from 3 to 10. Agglomerative hierarchical clustering used Ward's linkage, with the number of clusters determined by the maximal gap in the dendrogram merge distances. Clustering quality was assessed by the silhouette coefficient, Davies-Bouldin index, and Calinski-Harabasz index. External validation examined whether identified clusters corresponded to distinct performance profiles by computing the Kruskal-Wallis H-statistic across clusters for each performance target.

3.4. Interpretability Analysis Protocol

SHAP TreeExplainer was applied to the best-performing tree-based method to compute exact Shapley values for every prediction in the test folds. Global feature importance was computed as the mean absolute SHAP value across all test samples. LIME explanations were generated for the same test samples using 5,000 perturbations per instance. Although LIME was originally introduced for classifier explanations [16], its perturbation-based approach generalizes directly to regression by replacing the classification loss with mean squared error in the local surrogate fitting. We used a Ridge Regression surrogate with L2 regularization selected via cross-validation on the perturbed

neighborhood. The agreement between SHAP and LIME was quantified by Spearman's rank correlation of their per-sample feature importance rankings. A formulation parameter influence weight heatmap was constructed by computing the mean absolute SHAP value for each of the 17 primary formulation parameters against each of the 7 performance targets, normalized row-wise to sum to 1.0.

4. Results and Analysis

4.1. Regression Performance Comparison

4.1.1. Aggregate Method Ranking

Table 3 presents the regression performance of all six methods across the seven performance targets, evaluated by R^2 under nested five-fold cross-validation using the engineered feature set. XGBoost achieved the highest average R^2 of 0.891, followed by Random Forest at 0.874 and Gaussian Process Regression at 0.852. The linear methods (Ridge and Lasso) achieved substantially lower performance (average R^2 $^{oost\ (average}$ of 0.683 and 0.671, respectively), indicating that formulation-performance relationships in dental composites are fundamentally nonlinear. SVR achieved an intermediate average R^2 of 0.813, with strong performance for flexural strength (R^2 = 0.868) and weaker performance for wear depth (R^2 = 0.724). The dominance of tree-based ensembles aligns with recent findings in materials property prediction, where gradient-boosted methods consistently outperform both linear and kernel-based approaches on small, heterogeneous tabular datasets [17].

Table 3. Regression Performance (R^2) Across Seven Performance Targets

Method	FlexMod	FlexStr	CompStr	ShrinkV	ShrinkS	Wear	DoCure	Avg R^2
Ridge	0.712	0.738	0.694	0.651	0.623	0.588	0.772	0.683
Lasso	0.698	0.725	0.681	0.639	0.611	0.574	0.767	0.671
Random Forest	0.903	0.912	0.886	0.864	0.841	0.792	0.918	0.874
XGBoost	0.921	0.934	0.902	0.878	0.856	0.815	0.932	0.891
SVR	0.843	0.868	0.831	0.794	0.768	0.724	0.863	0.813
GPR	0.882	0.896	0.867	0.838	0.812	0.761	0.907	0.852

All values report the mean R^2 across five outer folds. Best mean values per target are achieved by XGBoost across all columns.

4.1.2. Target-Specific Analysis

Performance varied substantially across targets. Depth of cure was the most predictable target (best R^2 = 0.932 with XGBoost), likely because it depends most directly on monomer transparency and filler refractive index --- properties that are well represented in the feature set. Wear depth was the least predictable target (best R^2 = 0.815), consistent with the complex tribological mechanisms governing material wear, which extend beyond composition to include microstructural factors not captured by formulation attributes. The mean absolute error for XGBoost on flexural strength was 11.3 MPa (approximately 10.0% of the target mean of 112.6 MPa), while the MAE for volumetric shrinkage was 0.31% (10.9% of the target mean of 2.84%). In the context of preliminary formulation screening, where the objective is to narrow the experimental search space rather than to replace physical testing, these relative error magnitudes are sufficient to rank-order candidate formulations and identify promising regions of the design space (Figure 1).

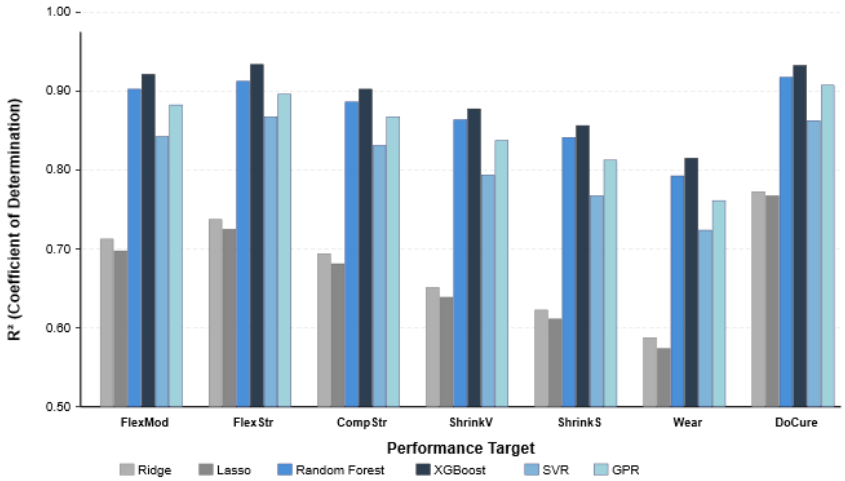

Figure 1. Regression Performance Comparison Across Seven Targets

R^2 values of the six regression methods evaluated on each of the seven dental composite performance targets under nested five-fold cross-validation. XGBoost achieves the highest R^2 on all seven targets, with values ranging from 0.815 (wear depth) to 0.934 (flexural strength). Random Forest ranks second across all targets, trailing XGBoost by 0.014 to 0.023 in R^2. The gap between nonlinear methods (XGBoost, Random Forest, GPR, SVR) and linear methods (Ridge, Lasso) is most pronounced for shrinkage stress (0.233 difference between XGBoost and Ridge) and wear depth (0.227 difference), indicating that these properties involve the strongest nonlinear formulation interactions.

4.2. Feature Engineering Impact

The effect of feature set configuration on regression performance is shown in Table 4 for the top-performing method (XGBoost). The engineered feature set (34 dimensions: 28 raw features comprising 17 primary formulation attributes plus 11 one-hot-encoded categorical indicators, augmented by 6 domain-guided interaction terms) improved the average R^2 by 0.038 over the raw features (28 dimensions), confirming that interaction terms capture meaningful formulation relationships not encoded in individual variables. The MI-selected set (18 dimensions) achieved 96.5% of the engineered set's performance while reducing dimensionality by 47%. The FCBF-selected set (15 dimensions) retained 95.1% of performance with 56% dimensionality reduction, representing an efficient option when computational resources or data volume are constrained.

Table 4. Effect of Feature Set Configuration on XGBoost Performance (Average R^2)

Feature Set	Dimensions	Avg R^2	Δ vs. Raw	Training Time (s)
Raw	28	0.853	—	4.2
Engineered	34	0.891	+0.038	5.7
MI-Selected	18	0.860	+0.007	2.8
FCBF-Selected	15	0.847	-0.006	2.3

Training time measured per fold on a single CPU core (Intel Xeon W-2245, 3.9 GHz).

Mean absolute SHAP values for the 17 primary formulation parameters computed against each of the seven performance targets, normalized row-wise to sum to 1.0. Filler loading (wt%) exhibits the highest influence weight for flexural modulus (0.187), flexural strength (0.162), and compressive strength (0.171). The BisGMA-to-TEGDMA ratio dominates volumetric shrinkage (0.203) and shrinkage stress (0.194). Degree of conversion shows the strongest influence on depth of cure (0.231) and ranks among the top three parameters for all targets. Filler particle size contributes moderately to wear depth (0.134) but minimally to shrinkage properties (< 0.04). The three most influential parameters ---

filler loading, BisGMA/TEGDMA ratio, and degree of conversion --- collectively account for 58.4% of total SHAP attribution. This percentage was computed by first normalizing the mean absolute SHAP values row-wise (per target) so that they sum to 1.0, then averaging the normalized attributions of each parameter across all seven targets, and finally summing the averaged attributions of the three parameters (Figure 2).

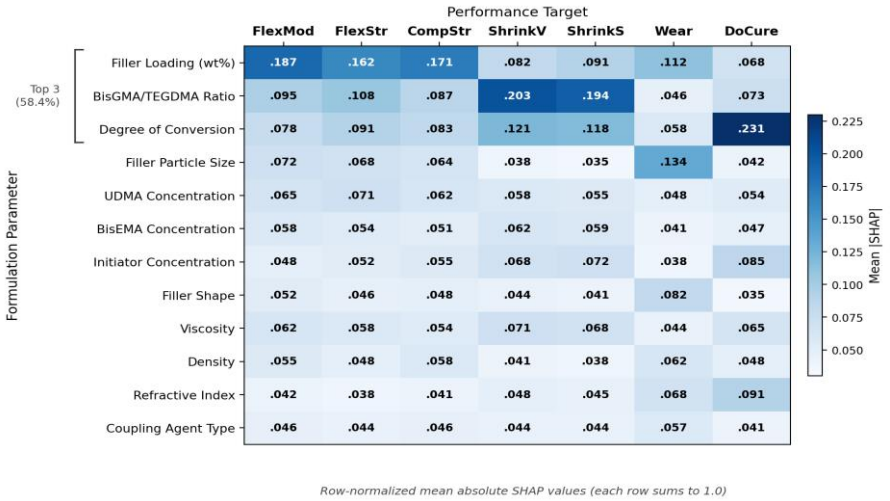

Figure 2. Feature Importance Heatmap: Formulation Parameter Influence Weights

4.3. Clustering Analysis

4.3.1. Formulation Archetype Identification

K-means clustering with k = 4 achieved the highest silhouette coefficient (0.412), outperforming k = 3 (0.387) and k = 5 (0.374). DBSCAN with ε = 0.8 and minimum samples = 5 identified three dense clusters and labeled 19 composites (8.2%) as noise points, achieving a silhouette coefficient of 0.378 on clustered samples. Agglomerative hierarchical clustering with Ward's linkage produced a dendrogram with a clear gap at the four-cluster level, yielding a silhouette coefficient of 0.405. The Kruskal-Wallis test confirmed that K-means (k = 4) cluster assignments corresponded to statistically distinct performance profiles for six of seven targets ($p < 0.01$), with wear depth being the exception ($p = 0.073$). Table 5 characterizes the four identified formulation archetypes by their centroid feature values and mean performance. The identification of compositionally coherent clusters supports the premise that dental composite formulation space contains discrete design regions, each associated with characteristic performance trade-offs, consistent with molecular inverse-design studies that have identified similar structure in polymer formulation landscapes [18].

Table 5. Formulation Archetype Characterization (K-means, K = 4)

Characteristic	Cluster 1 *n*=68	Cluster 2 *n*=52	Cluster 3 *n*=71	Cluster 4 *n*=42
Dominant Filler	Barium glass	Silica nano	Mixed/hybrid	Zirconia - silica
Mean Filler Loading (wt%)	72.4	58.3	67.1	76.8
Mean BisGMA/TEGDMA Ratio	2.8	1.4	2.1	3.5

Mean Degree of Conv. (%)	62.1	71.8	66.4	58.3
Flexural Modulus (GPa)	10.8	7.2	9.1	11.9
Flexural Strength (MPa)	118.4	98.7	110.2	128.3
Volumetric Shrinkage (%)	2.51	3.62	2.94	2.18

Cluster assignments based on K-means with k = 4 in PCA-reduced space (3 components, 62.3% variance). Silhouette coefficient = 0.412.

4.3.2. Cluster-Performance Profile Correspondence

Cluster 4 (high-filler zirconia-silica composites) exhibited the highest flexural modulus (11.9 GPa) and flexural strength (128.3 MPa) with the lowest volumetric shrinkage (2.18%), consistent with the known reinforcement efficiency of zirconia-silica fillers at high loading levels. Cluster 2 (nano-silica composites with lower filler loading) showed the highest degree of conversion (71.8%) yet the weakest mechanical properties (flexural strength 98.7 MPa) and greatest volumetric shrinkage (3.62%), reflecting the trade-off between conversion-driven shrinkage and filler-mediated reinforcement. Cluster 1 (barium glass composites) and Cluster 3 (hybrid fillers) occupied intermediate positions along these axes. These patterns align with established materials science understanding and confirm that the clustering captures compositionally and mechanically meaningful groupings rather than artifacts of the dimensionality reduction (Figure 3).

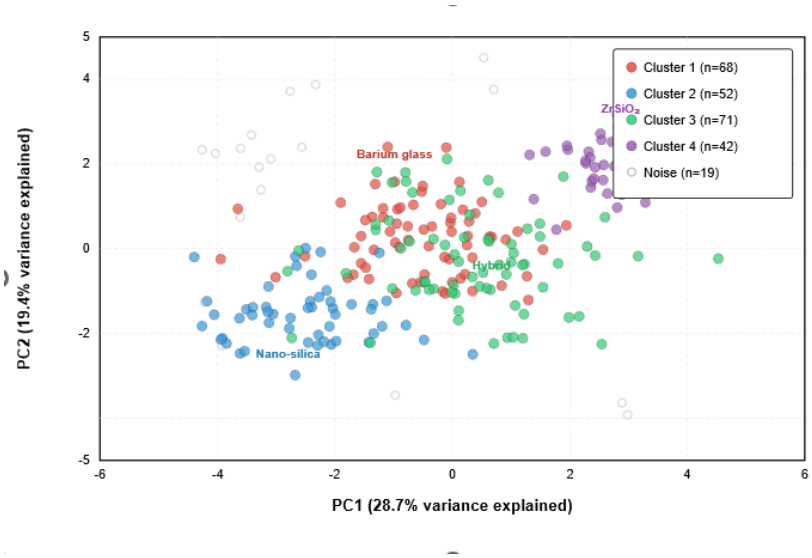

Figure 3. Clustering Visualization in PCA-Reduced Space

Scatter plot of 233 dental composites projected onto the first two principal components (PC1 explaining 28.7% and PC2 explaining 19.4% of total variance), with points assigned to four clusters identified by K-means. Cluster 4 (high-filler zirconia-silica, n = 42) forms the most compact group in the upper-right quadrant, indicating high compositional consistency. Cluster 2 (nano-silica, n = 52) occupies a distinct region in the lower-left quadrant, separated from other clusters along PC1. Clusters 1 (n = 68) and 3 (n = 71) partially overlap in the central region, reflecting the compositional similarity between barium glass and hybrid-filler formulations. The 19 noise points identified by DBSCAN are concentrated at the periphery of the point cloud, representing unusual formulation compositions that do not conform to any dominant archetype.

4.4. Interpretability Analysis

The SHAP and LIME analyses were applied to XGBoost (the best-performing regression method) on all test-fold predictions. The global SHAP feature importance ranking placed filler loading first (mean |SHAP| = 0.187 for flexural modulus), followed by BisGMA/TEGDMA ratio (mean |SHAP| = 0.203 for volumetric shrinkage) and degree of conversion (mean |SHAP| = 0.231 for depth of cure), as visualized in Figure 2. To compare SHAP and LIME, we computed the Spearman rank correlation between their feature importance rankings for each test sample across all 17 primary features. These per-sample correlations were first averaged within each target, then averaged across the seven targets, yielding an overall mean Spearman ρ of 0.824 ($p < 0.001$), indicating substantial agreement between the two interpretability methods. The highest agreement occurred for flexural strength ($\rho = 0.891$) and the lowest for wear depth ($\rho = 0.712$), suggesting that the interpretability of predictions degrades for targets governed by more complex, distributed feature interactions.

SHAP dependence plots revealed nonlinear relationships between individual formulation parameters and performance targets. Filler loading exhibited a positive, concave relationship with flexural modulus up to approximately 75 wt%, beyond which marginal gains diminished and uncertainty increased --- consistent with the critical filler packing fraction above which matrix-filler stress transfer efficiency saturates. The BisGMA/TEGDMA ratio showed a negative, approximately linear relationship with volumetric shrinkage below a ratio of 3.0, while ratios above 3.0 produced minimal additional shrinkage reduction, reflecting the viscosity-driven limit on monomer conversion at high BisGMA concentrations.

5. Discussion

5.1. Method Selection Recommendations

The results demonstrate that tree-based ensemble methods --- XGBoost (average $R^2 = 0.891$) and Random Forest ($R^2 = 0.874$) --- provide the best predictive accuracy for dental composite formulation-performance relationships within the data regime examined. This outcome is consistent with broader findings in materials informatics and in tabular data benchmarks, which show that gradient-boosted trees outperform neural approaches when sample sizes are below approximately 1,000 and features are heterogeneous. Gaussian Process Regression, while achieving a moderately lower accuracy ($R^2 = 0.852$), provides a principled, uncertainty-aware prediction framework that may be valuable for active learning and sequential experimental design, in which future experiments could target formulations with high predicted uncertainty. Linear methods (Ridge, Lasso) underperform nonlinear approaches by 0.18--0.22 R^2 units on average, confirming that formulation-performance mappings are intrinsically nonlinear and cannot be adequately captured by additive models [19, 20]. The clustering analysis reveals four formulation archetypes that correspond to distinct performance profiles, offering a useful taxonomy for organizing the dental composite design space and identifying underexplored compositional regions. The SHAP-LIME comparison yields a Spearman rank correlation of 0.824, indicating that these two interpretability methods produce broadly consistent feature-importance rankings on dental composite data, though agreement weakens for targets with complex interaction structures, such as wear depth ($\rho = 0.712$). The three most influential formulation parameters identified --- filler loading, BisGMA/TEGDMA ratio, and degree of conversion --- account for 58.4% of total SHAP attribution and provide a clear prioritization guide for experimental campaigns seeking to maximize performance gains per experiment.

5.2. Limitations

Several limitations constrain the generalizability of these findings. The dataset of 233 composites, while the largest available for this specific domain, remains small by machine learning standards, and prediction intervals are correspondingly wide for underrepresented formulation regions. Literature-aggregated data introduces

heterogeneity in measurement protocols, curing conditions, and specimen geometries across source publications, potentially inflating residual variance. Wear depth, the least predictable target (R^2 = 0.815), likely depends on microstructural features --- filler distribution homogeneity, interfacial bonding quality, and polymerization network topology --- that are not captured in the 17 formulation attributes. Future work should integrate active learning loops that iteratively select the most informative formulations for experimental validation, thereby reducing the number of experiments required to achieve the target prediction accuracy. Multi-task learning approaches that jointly predict all seven performance targets may exploit inter-target correlations to improve individual predictions, particularly for targets with high missingness, such as wear depth (31.8% missing) and shrinkage stress (22.7% missing). Developing standardized, open-access dental composite databases with consistent measurement protocols would benefit the broader research community and enable more robust benchmarking of data-driven formulation analysis methods. Integrating microstructural descriptors derived from scanning electron microscopy image analysis could improve prediction accuracy fo port the mean R and other properties sensitive to filler dispersion quality. Symbolic regression methods could further distill the trained models into compact analytical expressions for enhanced physical interpretability.

References

1. S. Yamaguchi, H. Li, and S. Imazato, "Materials informatics for developing new restorative dental materials: A narrative review," *Frontiers in Dental Medicine*, vol. 4, Art. no. 1123976, 2023. [Online]. Available: https://doi.org/10.3389/fdmed.2023.1123976
2. H. Li, T. Sakai, A. Tanaka, M. Ogura, C. Lee, S. Yamaguchi, and S. Imazato, "Interpretable AI explores effective components of CAD/CAM resin composites," *Journal of Dental Research*, vol. 101, no. 11, pp. 1363–1371, 2022. [Online]. Available: https://doi.org/10.1177/00220345221089251
3. R. Wang, V. Hass, and Y. Wang, "Machine learning analysis of microtensile bond strength of dental adhesives," *Journal of Dental Research*, vol. 102, no. 9, pp. 1022–1030, 2023. [Online]. Available: https://doi.org/10.1177/00220345231175868
4. X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, and T. Y.-J. Han, "Explainable machine learning in materials science," *npj Computational Materials*, vol. 8, Art. no. 204, 2022. [Online]. Available: https://doi.org/10.1038/s41524-022-00884-7
5. K. Paniagua, K. Whang, K. Joshi, H. Son, Y. S. Kim, and M. Flores, "Dental composite performance prediction using artificial intelligence," *Journal of Dental Research*, vol. 104, no. 5, pp. 513–521, 2025. [Online]. Available: https://doi.org/10.1177/00220345241311888
6. J. Han, I. Kim, N. Cho, K. S. Yang, J. S. Myung, J. Park, S. H. Kim, and W. J. Choi, "Toward accurate machine learning-driven prediction of polymeric composites properties based on experimental data," *Materials Genome Engineering Advances*, vol. 3, no. 3, Art. no. e70027, 2025. [Online]. Available: https://doi.org/10.1002/mgea.70027
7. S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama, and M. Naito, "Prediction and optimization of epoxy adhesive strength from a small dataset through active learning," *Science and Technology of Advanced Materials*, vol. 20, no. 1, pp. 1010–1021, 2019. [Online]. Available: https://doi.org/10.1080/14686996.2019.1673670
8. C.-T. Chen and G. X. Gu, "Machine learning for composite materials," *MRS Communications*, vol. 9, no. 2, pp. 556–566, 2019. [Online]. Available: https://doi.org/10.1557/mrc.2019.32
9. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939785
10. T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Physical Review Letters*, vol. 120, no. 14, Art. no. 145301, 2018. [Online]. Available: https://doi.org/10.1103/PhysRevLett.120.145301
11. M. J. Kim, L. Grinsztajn, and G. Varoquaux, "CARTE: Pretraining and transfer for tabular learning," in *Proceedings of the 41st International Conference on Machine Learning*, pp. 23843–23866, 2024.
12. R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, "SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates," *Physical Review Materials*, vol. 2, no. 8, Art. no. 083802, 2018. [Online]. Available: https://doi.org/10.1103/PhysRevMaterials.2.083802
13. L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proceedings of the 20th International Conference on Machine Learning*, pp. 856–863, 2003.
14. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, pp. 4765–4774, 2017.

15. A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, "Benchmarking materials property prediction methods: The Matbench test suite and Automatminer reference algorithm," *npj Computational Materials*, vol. 6, Art. no. 138, 2020. [Online]. Available: https://doi.org/10.1038/s41524-020-00406-3
16. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939778
17. F. Oviedo, J. L. Ferres, T. Buonassisi, and K. T. Butler, "Interpretable and explainable machine learning for materials science and chemistry," *Accounts of Materials Research*, vol. 3, no. 6, pp. 597–607, 2022. [Online]. Available: https://doi.org/10.1021/accountsmr.1c00244
18. S. Takeda, T. Hama, H.-H. Hsu, V. A. Piunova, D. Zubarev, D. P. Sanders, J. W. Pitera, M. Kogoh, T. Hongo, Y. Cheng, *et al.*, "Molecular inverse-design platform for material industries," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2961–2969, 2020. [Online]. Available: https://doi.org/10.1145/3394486.3403346
19. M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, "Discovering symbolic models from deep learning with inductive biases," in *Advances in Neural Information Processing Systems 33*, pp. 17429–17442, 2020.
20. B. K. Petersen, M. Landajuela, T. N. Mundhenk, C. P. Santiago, S. K. Kim, and J. T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," in *Proceedings of the 9th International Conference on Learning Representations*, 2021.