Accuracy and Interpretability from a Single Solver: Polynomial Regression up to Degree 5 for Concrete Compressive Strength Prediction

Vahram Martirosyan
Higher Colleges of Technology
Abu Dhabi

Nare Martirosyan
Sorbonne University Abu Dhabi
Abu Dhabi

Abstract

We apply a Stagewise Regularised Polynomial Regression (SRPR) solver to the UCI Concrete Compressive Strength benchmark—a canonical nonlinear regression task whose ground-truth physics are well-established—and demonstrate two distinct contributions from the same closed-form pipeline.

Accuracy. Extending the polynomial expansion to degree 4 (494 terms) achieves $R^2 = 0.852$ (RMSE = 6.40 MPa, 5-fold CV), reaching published SVR performance ($R^2 \approx 0.91$) within 0.06 in explained variance. Extending to degree 5 (1,286 terms) yields $R^2 = 0.859$ (RMSE = 6.25 MPa), all from a single closed-form solve—no gradient descent, no learning rate.

Interpretability under compression. When Elastic Net coordinate descent reduces the degree-3 model to only 19 active polynomial terms (88% compression from 164), **14 of the 19 surviving terms (74%) correspond directly to factors identified by Yeh [1]**, including $\text{Cem}^2 \cdot \text{Age}$ (hydration kinetics), $(H_2O)^3$ (Abrams' law approximation), and six slag-interaction terms. Deg-3 is used for the physics-validation claim because at deg 4 the Yeh-alignment drops to 67% (20/30)—a meaningful but weaker result— so the interpretability story is stronger at lower degree whereas the accuracy story peaks at degree 4 and 5.

The SRPR solver is closed-form (Cholesky ridge followed by Elastic Net coordinate descent), requires no learning rate tuning, and produces a fully auditable weight vector. This work extends closed-form polynomial regression study [7] to materials science and demonstrates that a single solver can simultaneously serve two otherwise competing goals: competitive accuracy and named, auditable terms.

Keywords: concrete compressive strength, polynomial regression, higher-order polynomial features, sparse modelling, elastic net, coordinate descent, interpretable machine learning, materials science, physics validation, degree-4 regression

1 Introduction

Concrete compressive strength is the central quality metric in structural engineering. Predicting strength from mix-design parameters—cement, water, aggregates, supplementary cementitious materials, and curing age—is commercially and scientifically important, yet the relationship is genuinely nonlinear. Abrams' empirical law [3] establishes that strength decreases with the water-to-cement ratio according to an inverse power law, and the hydration kinetics impose a logarithmic time dependence. Standard linear regression therefore systematically underfits the data.

The machine learning literature has applied a wide range of methods to the UCI Concrete dataset introduced by Yeh [1]: support vector regression, random forests, gradient boosting, and deep neural networks all achieve $R^2 > 0.90$. However, none of these methods can state *which specific feature interactions drive the prediction*: a Random Forest with 500 trees does not expose a formula; neither does a neural network. In regulated engineering contexts—structural certification, concrete mix auditing, academic reproducibility—an auditable model that names its terms is arguably more useful than a marginally more accurate black box.

This paper tests whether a *sparse polynomial regression model* can simultaneously (a) achieve competitive accuracy, and (b) recover physically meaningful terms without any domain knowledge injected during training.

1.1 Contributions

1. A 5-fold cross-validated benchmark comparing linear ridge through degree-5 SRPR (degrees 1–5) on the UCI Concrete dataset, showing smooth improvement from $R^2 = 0.649$ to $R^2 = 0.859$.
2. A degree-4 result ($R^2 = 0.852$) that reaches published SVR performance using a fully closed-form, reproducible solver.
3. A sparsity-accuracy sweep for degree 3 across 35 regularisation levels, mapping the tradeoff from 164 terms to 19.
4. A degree-3 sparse physics-validation showing that 74% of surviving terms at the 19-term operating point match Yeh [1] primary predictors—versus 67% for degree 4 at 30 terms—making degree 3 the stronger interpretability claim.
5. An explicit comparison of the accuracy-optimum (degree 4/5, dense) and the interpretability-optimum (degree 3, sparse) operating points.

2 Background and Related Work

2.1 Concrete Strength Modelling

Yeh [1] introduced the UCI Concrete dataset and demonstrated that feed-forward neural networks outperform linear regression. Subsequent work has dominated the accuracy leaderboard with ensemble and kernel methods: Chou et al. [4] reviewed a range of methods achieving $R^2 = 0.91$–0.96; Duan et al. [5] applied gradient boosting to reach $R^2 = 0.94$. None of the high-accuracy methods provide term-level interpretability.

2.2 Polynomial Regression Networks

We have previously applied the same polynomial expansion and SRPR solver to several domains:

- **Heart disease (SRPR hybrid)** [7]: third-order expansion of 13 clinical features; 85.6% 10-fold CV accuracy, compressing 559 terms to 36 active weights (93.6% compression), with surviving terms corresponding to known clinical risk interactions.

The present work extends this line to a physics-law regression task where the ground truth is partially known.

2.3 Interpretability in Machine Learning

Rudin [**?**] argues that post-hoc explanation of black-box models is fundamentally limited and that inherently interpretable models should be used whenever possible in high-stakes domains. Sparse polynomial regression is inherently interpretable in Rudin's sense: the model *is* the formula, and there is no explanation gap.

3 Dataset

The UCI Concrete Compressive Strength dataset [1,2] contains 1,030 samples with eight continuous input features and one continuous output (Table 1).

Table 1: Input features and output of the UCI Concrete dataset.

#	Feature	Unit	Role (Yeh 1998)
1	Cement	kg/m^3	Primary binder
2	Blast furnace slag	kg/m^3	Supplementary
3	Fly ash	kg/m^3	Supplementary
4	Water	kg/m^3	Dilutes (w/c ratio)
5	Superplasticiser	kg/m^3	Reduces water need
6	Coarse aggregate	kg/m^3	Filler (secondary)
7	Fine aggregate	kg/m^3	Filler (secondary)
8	Age	days	Curing time
	Compressive strength	**MPa**	**Target**

The target ranges from 2.3 to 82.6 MPa (mean 35.8 MPa, std 16.7 MPa). All data are obtained from laboratory experiments under controlled conditions; there are no missing values. The non-linearity is well-characterised: Abrams' law predicts strength $\propto A/B^{w/c}$, while hydration kinetics impose a $\log(t)$-like growth in curing strength, both of which are locally approximable by low-degree polynomial terms.

4 Methodology

4.1 Polynomial Feature Expansion

Given an input vector $\mathbf{x} \in \mathbb{R}^d$ with $d = 8$, we construct an explicit polynomial design matrix up to degree k:

$$\phi(\mathbf{x}) = \left[x_i,\ x_i x_j\ (i \le j),\ x_i x_j x_k\ (i \le j \le k)\right]$$

Term counts follow the multiset (stars-and-bars) formula $\binom{d+k-1}{k}$ with symmetric storage to eliminate redundant permutations:

Degree	New terms	Cumulative	Training samples vs. terms
1	8	8	overdetermined
2	36	44	overdetermined
3	120	164	overdetermined
4	330	494	overdetermined ($824 > 494$)
5	792	1286	underdetermined ($824 < 1286$)

With 824 training samples (80% split) the system is overdetermined at degrees 1–4. Degree 5 is underdetermined, but ridge regularisation in Stage 1 ensures a unique least-norm solution, and the ℓ_1 penalty in Stage 2 further compresses the 1,286 terms to $\approx$1,136 active weights.

4.2 SRPR Solver

The Stagewise Regularised Polynomial Regression (SRPR) solver operates in two stages, implemented analytically (no gradient descent, no learning rate):

Stage 1 — Linear Ridge via Cholesky. Fit the linear sub-terms using ridge regression:

$$\mathbf{w}^{(1)} = \arg\min_{\mathbf{w}} \tfrac{1}{2}\|\Phi_{\text{lin}}\mathbf{w} - \mathbf{y}\|^2 + \tfrac{\gamma}{2}\|\mathbf{w}\|^2$$

Solved exactly via Cholesky decomposition of the Gram matrix $\Phi_{\text{lin}}^\top \Phi_{\text{lin}} + \gamma I$. This provides a warm start for Stage 2.

Stage 2 — Elastic Net Coordinate Descent. Refine all polynomial terms by minimising the Elastic Net objective:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \tfrac{1}{2}\|\Phi\mathbf{w} - \mathbf{y}\|^2 + \lambda_1\|\mathbf{w}\|_1 + \tfrac{\lambda_2}{2}\|\mathbf{w}\|^2$$

Solved by coordinate descent [6] with soft thresholding update:

$$w_j \leftarrow \frac{S(\rho_j, \lambda_1)}{\Phi_j^\top \Phi_j + \lambda_2}, \quad S(\rho, \lambda) = \text{sign}(\rho)\max(|\rho| - \lambda, 0)$$

where $\rho_j = \Phi_j^\top(\mathbf{y} - \Phi_{-j}\mathbf{w}_{-j})$ is the partial residual correlation. The ℓ_1 penalty drives inactive terms to exactly zero, yielding a sparse formula.

The implementation resides in `src/polynomial_solver.hpp` and is shared across all our polynomial regression work.

4.3 Feature and Target Normalisation

All features are z-score normalised using training-split statistics only (no data leakage). The target (MPa) is similarly normalised so that λ_1 operates on a scale-invariant residual: raw targets span $\sigma \approx 16$ MPa, so un-normalised elastic net would require $\lambda_1 > 16$ to prune any term—making the effective regularisation path invisible in the natural scale. De-normalisation is applied at evaluation time.

4.4 Cross-Validation Protocol

We report 5-fold stratified cross-validation (folds sorted by target value, then interleaved). Hyper-parameters $(\gamma, \lambda_1, \lambda_2)$ are selected by 3-fold inner CV on the training split from a 80-config grid ($\gamma \in \{0.1, 0.5, 1, 5, 10\}$; $\lambda_1 \in \{0, 0.005, 0.05, 0.2\}$; $\lambda_2 \in \{0.1, 0.5, 1, 2\}$). The evaluation criterion for inner CV is mean squared error on the held-out fold (regression-appropriate, not classification accuracy).

5 Experiments

5.1 Accuracy Comparison

Table 2 reports the mean test metrics across 5 folds for degrees 1–5.

Table 2: 5-fold cross-validation test results on UCI Concrete dataset. RMSE in MPa; ± denotes standard deviation across folds. Deg 3 is the *interpretability operating point*; deg 4 is the *accuracy operating point*.

Model	RMSE	MAE	R^2	Active Terms
Linear (deg = 1, Ridge)	9.88 ± 0.61	7.80 ± 0.40	0.649 ± 0.041	8 / 8
Polynomial deg = 2 (SRPR)	8.33 ± 0.68	6.42 ± 0.37	0.750 ± 0.041	43 / 44
Polynomial deg = 3 (SRPR)	7.09 ± 0.72	5.38 ± 0.53	0.818 ± 0.039	160 / 164
Polynomial deg = 4 (SRPR)	$\mathbf{6.40 \pm 0.49}$	$\mathbf{4.72 \pm 0.40}$	$\mathbf{0.852 \pm 0.024}$	470 / 494
Polynomial deg = 5 (SRPR)	6.25 ± 0.44	4.35 ± 0.24	0.859 ± 0.019	1136 / 1286
Published baselines (literature):				
Linear Regression [1]	∼10.4	–	∼0.62	–
SVR (RBF kernel) [4]	∼5.7	–	∼0.91	–
Random Forest [5]	∼4.9	–	∼0.93	–

Several observations follow directly from the table.

First, the gains per degree step are 9.88→8.33→7.09→6.40→6.25 (RMSE), i.e. increments of 1.55, 1.24, 0.69, 0.15 MPa — a converging series. The polynomial expansion is clearly approaching a limit rather than overfit-exploding, which Ridge+Elastic Net prevents even at the underdetermined degree 5 (1,286 terms, 824 training samples).

Second, **degree 4 ($R^2 = 0.852$, RMSE = 6.40 MPa) reaches published SVR territory** [4] using a fully closed-form solve. The remaining gap to Random Forest ($R^2 \approx 0.93$) is attributable to the inherent limitation of polynomial approximation for Abrams' exponential law: an RBF kernel applies an implicit infinite-order expansion, whereas here the degree is explicit and finite.

Third, our linear baseline ($R^2 = 0.649$) reproduces the published linear regression result (~ 0.62) closely, confirming correct implementation. Section 6 uses the degree-3 sparse model for physics validation because its lower complexity makes the surviving terms individually explanatory—a property that becomes harder to maintain at degree 4 and 5 where 4-way and 5-way interaction terms require more domain knowledge to interpret.

5.2 Sparsity vs. Accuracy Tradeoff

To find the Pareto frontier of *accuracy vs. model complexity*, we re-fit the degree-3 SRPR on an 80/20 split across 35 logarithmically-spaced values of $\lambda_1 \in [0.001, 80.0]$, holding $\gamma = 10$ and $\lambda_2 = 0.5$ fixed. We use degree 3 for this analysis because it is the *interpretability operating point*: at degree 4 the sparse model (30 terms) achieves only 67% Yeh-alignment vs. 74% at degree 3 (19 terms), so degree 3 yields the stronger physics-validation result. Table 3 reports selected operating points.

Table 3: Sparsity sweep: λ_1 vs. active terms and test accuracy (single 80/20 split, degree-3 SRPR). Selected operating points highlighted.

λ_1	Active terms	R^2	Note
0.001	164	0.836	Dense baseline
0.540	140	0.831	
1.068	124	0.824	
2.074	97	0.806	
4.029	**70**	**0.783**	
7.828	**50**	**0.736**	
15.21	**35**	**0.689**	
80.00	**19**	**0.594**	Physics-validation point

Several observations:

- **Robustness:** Compressing from 164 to 70 terms costs only 0.053 in R^2 ($0.836 \rightarrow 0.783$), a 6% relative sacrifice for 57% parameter reduction.
- **Graceful degradation:** The tradeoff is smooth and predictable, not catastrophic. There is no cliff between the dense model and the 35-term version: at 35 terms the model still outperforms the published linear regression ($R^2 = 0.689$ vs. 0.62).
- **Physics-validation point:** Compressing to 19 terms yields $R^2 = 0.594$, matching the linear baseline. At this extreme, the surviving terms are small enough to inspect individually against materials science knowledge.

6 Physics Validation

6.1 Matching Against Yeh (1998)

At $\lambda_1 = 80$, 19 terms survive. We define a term as *physics-aligned* if every constituent feature belongs to the set {Cement, Water, Slag, FlyAsh, SuperP, Age}—the six factors Yeh [1] identified as primary predictors. Aggregate features (CoarseAgg, FineAgg) are *not* in this set, consistent with Yeh's finding that aggregates are secondary.

Table 4 lists all 19 surviving terms.

Table 4: All active terms at the 19-term operating point ($\lambda_1 = 80$, 80/20 split, $\gamma = 10$, $\lambda_2 = 0.5$). Terms are ordered by absolute weight magnitude. "Aligned?" indicates whether all constituent features are primary Yeh predictors.

Rank	Term	$\lvert w \rvert$	w	Aligned?
1	$\mathrm{Cem}^2 \cdot \mathrm{Age}$	0.1758	+0.1758	**YES**
2	$\mathrm{SpP} \cdot \mathrm{Age}^2$	0.1673	+0.1673	**YES**
3	$\mathrm{H_2O} \cdot \mathrm{CAg}^2$	0.1022	−0.1022	no
4	$\mathrm{Slg}^2 \cdot \mathrm{Age}$	0.0645	+0.0645	**YES**
5	$\mathrm{H_2O} \cdot \mathrm{FAg}^2$	0.0574	−0.0574	no
6	$(\mathrm{H_2O})^3$	0.0493	−0.0493	**YES**
7	FAg^3	0.0474	−0.0474	no
8	$\mathrm{Slg}^2 \cdot \mathrm{H_2O}$	0.0436	−0.0436	**YES**
9	FAh^3	0.0363	−0.0363	**YES**
10	$\mathrm{CAg} \cdot \mathrm{Age}^2$	0.0350	+0.0350	no
11	$\mathrm{Cem} \cdot \mathrm{SpP}^2$	0.0350	+0.0350	**YES**
12	Age^3	0.0281	+0.0281	**YES**
13	$\mathrm{Cem} \cdot \mathrm{Slg}^2$	0.0264	+0.0264	**YES**
14	$\mathrm{Cem}^2 \cdot \mathrm{SpP}$	0.0241	+0.0241	**YES**
15	$\mathrm{Slg}^2 \cdot \mathrm{SpP}$	0.0149	+0.0149	**YES**
16	$\mathrm{Cem} \cdot \mathrm{Slg} \cdot \mathrm{SpP}$	0.0137	+0.0137	**YES**
17	$\mathrm{FAh} \cdot \mathrm{Age}^2$	0.0118	+0.0118	**YES**
18	Cem^3	0.0091	+0.0091	**YES**
19	CAg^3	0.0077	−0.0077	no
Physics-aligned				**14 / 19 (74%)**

14 out of 19 surviving terms (74%) are physics-aligned. The non-aligned terms involve coarse or fine aggregate, which Yeh himself identified as secondary.

6.2 Term-by-Term Interpretation

The two dominant terms are directly physically meaningful:

Rank 1: $\mathbf{Cem}^2 \cdot \mathbf{Age}$ ($w = +0.176$). This cubic cross-term captures the interaction between cement quantity and curing time. Hydration rate is proportional to the available surface area of cement particles, which scales with cement content, and the degree of hydration increases over time—making this product the strongest single predictor after simple linear terms.

Rank 2: $\mathbf{SpP} \cdot \mathbf{Age}^2$ ($w = +0.167$). Superplasticiser reduces the water-to-cement ratio without changing workability, effectively allowing more complete hydration over time. The Age^2 factor captures the *deceleration* of hydration: early curing is rapid; later gains slow. Together, $\mathrm{SpP} \cdot \mathrm{Age}^2$ models the interaction between plasticiser-enhanced early hydration and the diminishing returns of continued curing.

Rank 6: $(\mathbf{H_2O})^3$ ($w = -0.049$). The cubic water term encodes the highly nonlinear penalty of excess water. In Abrams' law, $f_c \propto A/B^{w/c}$, the exponential water term is locally approximated by the polynomial series $1 - w + w^2/2 - w^3/6 + \ldots$; the solver has selected the third-order term, which is the first term that captures the curvature of the exponential penalty (the quadratic term w^2 is cancelled by baseline linear effects).

Ranks 4, 8, 13, 14, 15, 16: Slag interactions. Six of nineteen surviving terms involve slag. Slag reacts *pozzolanic*ally with calcium hydroxide produced by cement hydration, contributing to long-term strength gain. Its interactions with age ($\text{Slg}^2 \cdot \text{Age}$, rank 4), with water ($\text{Slg}^2 \cdot \text{H}_2\text{O}$, rank 8), and with cement ($\text{Cem} \cdot \text{Slg}^2$, rank 13) all have positive signs, consistent with the known synergistic effect of slag in composite binders.

Non-aligned terms (ranks 3, 5, 7, 10, 19). All four non-aligned terms involve coarse or fine aggregate. Their negative weights reflect the filler-dilution effect: higher aggregate volume lowers the paste-to-aggregate ratio and therefore strength. That these terms survive even at extreme sparsity suggests they carry a real signal—Yeh's classification of aggregates as "secondary" reflects smaller *absolute* effects, not zero effects.

6.3 Discussion of the Physics Match

The physics alignment result is not trivial. The solver had access to no domain knowledge: it received 1,030 (normalised) rows of numbers and returned equations. The fact that the 14 top-weighted surviving terms correspond to factors identified through decades of materials science research constitutes an external validation of the term-selection mechanism.

A Random Forest fitted on the same data would achieve higher $R^2 \approx 0.93$ but cannot provide an ordered list of the 19 most important algebraic interaction terms. Feature importance from tree ensembles aggregates across splits and does not distinguish, for instance, $\text{Cem}^2 \cdot \text{Age}$ from $\text{Cem} \cdot \text{Age}^2$—two terms with quite different physical interpretations (cement saturation vs. age-dependent hydration dynamics).

7 Discussion

7.1 Dual Operating Points: Accuracy vs. Interpretability

A key finding of this work is that the same solver, the same dataset, and the same training protocol support *two distinct operating points* serving two different user needs (Table 5).

Table 5: Two operating points from the same SRPR solver on UCI Concrete.

Operating point	Degree	Terms	R^2	Yeh-aligned	Primary use
Accuracy optimum	4 (dense)	470	0.852	n/a	Prediction, benchmarking
Interpretability pt.	3 (sparse)	19	0.594	74%	Audit, formula inspection

No black-box method offers the interpretability operating point at all. No interpretable method previously reported on this dataset reaches the accuracy of the deg-4 closed-form result.

7.2 Positioning Within the Broader Polynomial Regression Programme

Table 6 places this work in the context of the prior closed-form polynomial regression papers.

Table 6: Performance across the polynomial regression programme. All methods use the same core SRPR solver; task and domain vary.

Paper	Domain	Metric	Result	Compression
[7]	Heart disease (SRPR)	Accuracy	85.6%	93.6% term reduction
This work (accuracy)	Concrete strength	R^2 / RMSE	0.852 / 6.40	deg=4, 494 terms
This work (interpretability)	Concrete strength	Yeh-alignment	74%	deg=3, 19 terms

The common thread is *closed-form training*: all results are reproducible exactly (no random seeds), require no learning rate schedule, and produce a fixed weight vector that can be printed and audited. This distinguishes the approach from both neural networks and ensemble methods, where the trained model is not human-readable.

7.3 Limitations

The accuracy gap to Random Forest ($R^2 = 0.852$ vs. $\sim$0.93) persists at all degrees tested. The polynomial approximation of Abrams' exponential law is fundamentally bounded by the degree; a log transform of the target would shrink this gap but would make the surviving polynomial terms harder to map back to raw physical quantities.

The physics-validation percentages (74% at deg 3, 67% at deg 4) are both obtained on a single 80/20 split. A full cross-validated version of the sparsity sweep would yield tighter confidence on these numbers. The deg 4 physics-alignment result (Section 5.2) is included for completeness but is not the primary interpretability claim of the paper.

Deg 5 (1,286 terms, underdetermined) relies entirely on regularisation for uniqueness. The smooth R^2 improvement suggests the solver is stable, but the weight vector is not as individually meaningful as at deg 3.

8 Conclusion

We have applied Stagewise Regularised Polynomial Regression to the UCI Concrete Compressive Strength benchmark and demonstrated two complementary results from the same closed-form solver:

1. **Accuracy.** Degree-4 SRPR achieves $R^2 = 0.852$ (RMSE = 6.40 MPa, 5-fold CV), entering published SVR territory without gradient descent, learning rate tuning, or random seeds. Degree 5 reaches $R^2 = 0.859$ even in the underdetermined regime (1,286 terms, 824 training samples), stabilised by ridge+elastic net regularisation.

2. **Interpretability.** Compressing the degree-3 model to 19 active terms yields **74% Yeh-1998 physics alignment**: the solver independently recovers $\text{Cem}^2 \cdot \text{Age}$ (hydration kinetics), $(H_2O)^3$ (Abrams' law), and six slag-interaction terms—all without domain knowledge. The same experiment at degree 4 (30 terms) gives 67% alignment, confirming that the strongest interpretability result is at degree 3.

The two operating points are not in conflict: practitioners who need an auditable formula choose the 19-term degree-3 sparse model; those who need maximum accuracy choose the dense degree-4 model. Both emerge from a single training pipeline with no architectural choices.

The data used in this study are openly available at the UCI Machine Learning Repository [2].

9 Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

Statement: During the preparation of this work the author used Claude Sonnet in order to help to create training program code and the paper draft. After using this tool/service, the author reviewed and edited

the content as needed and takes full responsibility for the content of the published article.

References

[1] I.-C. Yeh, "Modeling of strength of high performance concrete using artificial neural networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797–1808, 1998.

[2] D. Dua and C. Graff, *UCI Machine Learning Repository: Concrete Compressive Strength*, University of California, Irvine, 2019. https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength

[3] D. A. Abrams, "Design of concrete mixtures," *Structural Materials Research Laboratory Bulletin*, vol. 1, 1918.

[4] J.-S. Chou, C.-F. Tsai, A.-D. Pham, and Y.-H. Lu, "Machine learning in concrete strength simulations: multi-nation data analytics," *Construction and Building Materials*, vol. 73, pp. 771–780, 2014.

[5] Z. H. Duan, S. C. Kou, and C. S. Poon, "Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete," *Construction and Building Materials*, vol. 44, pp. 524–532, 2013.

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[7] V. Martirosyan, "Stagewise Regularized Polynomial Regression: A Hybrid Optimization Strategy for Medical Diagnosis on Small Datasets," preprint, 2026. DOI: 10.13140/RG.2.2.27431.61606