

# A New Predictive Algorithm for Enhancing Regression Performance Using Insights from Random Vine Anchorage Mechanisms

Weerachai Jonburom<sup>1,\*</sup> Charinee Chaichana<sup>2</sup>

<sup>1,2</sup> Faculty of Industry and Technology, Rajamangala University of Technology Isan,  
Sakon Nakhon Campus, Phang Khon, Sakon Nakhon, Thailand 47160

\*Corresponding Email : weerachai.jo@rmuti.ac.th

---

Received February 13, 2025, Revised December 1, 2025, Accepted March 5, 2026, Published July 1, 2026

---

**Abstract.** *This study proposes a novel regression algorithm termed the Random Vine Anchorage Mechanism (RVAM), inspired by adaptive anchoring behaviors observed in climbing vines. Unlike conventional regression approaches that rely on fixed feature weighting or global optimization objectives, RVAM introduces an anchorage-based learning strategy that dynamically reinforces informative features while attenuating redundant or weakly relevant ones through a vine-structured dependency model. The primary objective of this research is to enhance regression performance under challenging data conditions, including high dimensionality, feature correlation, and noise perturbation. To achieve this, RVAM integrates three key mechanisms: adaptive feature anchorage initialization, random vine-based dependency modeling, and normalized anchorage weight updating. These components jointly enable the proposed algorithm to capture complex feature interactions while maintaining robustness and numerical stability. The effectiveness of RVAM is evaluated using multiple benchmark datasets specifically designed for regression analysis. Experimental results demonstrate that the proposed method consistently outperforms conventional regression models, including linear regression, LASSO, Elastic Net, Random Forest, and Gradient Boosting, across multiple evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), coefficient of determination ( $R^2$ ), and Mean Absolute Percentage Error (MAPE). On average, RVAM achieves performance improvements ranging from 8% to 15% compared with baseline methods, particularly in datasets characterized by multicollinearity and complex feature dependencies. These findings highlight the potential of biologically inspired anchorage mechanisms as an effective and generalizable strategy for improving regression performance in complex engineering and data-driven applications.*

**Keywords:** Biologically Inspired Computing, Regression Performance, Random Vine Anchorage Mechanisms.

## 1. Introduction

Regression analysis plays a fundamental role in modern engineering, scientific computing, and data-driven decision-making, where predictive models are required to estimate continuous outcomes under complex and uncertain conditions. With the rapid growth of high-dimensional data, contemporary regression tasks increasingly face challenges such as multicollinearity, redundant features, noise perturbation, and non-trivial feature interactions. Conventional regression models, including linear and regularized approaches, often suffer from performance degradation under these conditions due to their reliance on fixed assumptions and static feature weighting schemes [1], [2]. As a result, improving overall regression performance—rather than predictive accuracy alone—has become a critical research objective in many applied domains.

To address these challenges, a wide range of advanced regression techniques have been proposed, including ensemble learning, sparsity-inducing regularization, and dependency-aware modeling. While ensemble-based methods such as Random Forest and Gradient Boosting can partially mitigate variance and nonlinear effects, they remain sensitive to feature redundancy and correlated inputs [3]. Similarly, regularization-based models, including LASSO and Elastic Net, improve stability under multicollinearity but may discard informative features or oversimplify complex relationships [4]. Recent dependency modeling approaches, such as vine-based structures, offer flexible representations of multivariate dependencies but often incur high computational cost and limited adaptability when applied directly to large-scale regression problems [5]. These limitations highlight the need for a regression framework that can adaptively manage feature importance and dependency structures while maintaining robustness and computational feasibility.

Inspired by adaptive anchoring behaviors observed in climbing vines, this study proposes a novel regression algorithm termed the Random Vine Anchorage Mechanism (RVAM). In natural systems, vines dynamically adjust their anchorage points to maintain stability and optimize growth under varying environmental conditions. Translating this principle into a computational paradigm, RVAM introduces an anchorage-based learning mechanism that selectively reinforces informative features and attenuates redundant or weakly relevant ones through a random vine dependency structure. Unlike existing methods that rely on fixed feature selection or global optimization, RVAM dynamically updates feature anchorage strengths in response to evolving dependency patterns, thereby enhancing stability and predictive performance under complex data conditions.

The proposed approach is evaluated using multiple heterogeneous benchmark datasets specifically designed for regression analysis, encompassing diverse domains and data characteristics. This experimental design enables a comprehensive assessment of generalization capability, robustness, and comparative performance across established regression models. The main contributions of this study are threefold: (i) the introduction of a biologically inspired anchorage-based regression mechanism grounded in vine dependency modeling, (ii) an adaptive feature reinforcement strategy that improves stability under multicollinearity and noise, and (iii) an extensive empirical evaluation demonstrating consistent performance gains over conventional regression approaches using multiple evaluation metrics.

## 2. Literature Review

In recent years, regression modeling has evolved significantly to address challenges arising from high-dimensional data, multicollinearity, and noise contamination. While classical approaches such as ordinary least squares and regularized regression models (e.g., LASSO and Elastic Net) remain widely adopted, their predictive reliability often degrades when feature interactions become complex or data distributions deviate from ideal assumptions [6], [7].

Recent research has focused on improving robustness and stability in regression analysis. Zuo and Zuo introduced a weighted least squares regression framework that achieves an optimal balance between robustness and computational efficiency, demonstrating improved numerical stability under heterogeneous noise conditions [8]. Similarly, Bayesian robust regression models have been extended to high-dimensional settings by incorporating adaptive loss structures capable of handling serial correlation and heavy-tailed errors, resulting in more reliable parameter estimation and predictive uncertainty modeling [9]. Robust inference techniques for predictive regression have also been proposed to mitigate the effects of endogeneity and volatility, particularly in complex and time-dependent data environments [10].

Beyond robustness, ensemble learning and model-averaging strategies have emerged as effective solutions for regression tasks involving high-dimensional and heterogeneous datasets. Recent studies on robust model averaging for longitudinal and financial data have demonstrated superior predictive performance compared with single-model approaches, especially in the presence of outliers and non-linear feature interactions [11]. Modular regression frameworks that integrate auxiliary or side information have further enhanced predictive accuracy by decomposing complex learning problems into structured sub-components [12]. In addition, advances in feature dependency modeling and selection, including SHAP-based lightweight selection techniques, have improved both interpretability and stability in regression modeling without imposing excessive computational overhead [13].

Despite these advances, existing regression methods often rely on fixed weighting schemes or static feature relationships, limiting their adaptability to dynamically interacting features and varying data conditions. This limitation motivates the exploration of biologically inspired learning mechanisms that incorporate adaptive and decentralized decision processes. In this context, the proposed Random Vine Anchorage Mechanism (RVAM) introduces an anchorage-based learning paradigm that dynamically adjusts feature influence, enabling enhanced modeling of complex dependencies and improved regression performance in challenging predictive scenarios.

## 3. Methodology

This section presents the proposed Random Vine Anchorage Mechanism (RVAM) for regression modeling. The methodology is designed to enhance predictive performance and robustness by dynamically adjusting feature influence through an anchorage-inspired learning strategy. Unlike conventional regression frameworks that rely on fixed or static weighting schemes, RVAM introduces adaptive anchoring to better capture complex feature interactions under noisy and high-dimensional conditions.

### 3.1 Overall Framework of RVAM

The RVAM framework consists of four main stages, as illustrated conceptually in Fig. 1.

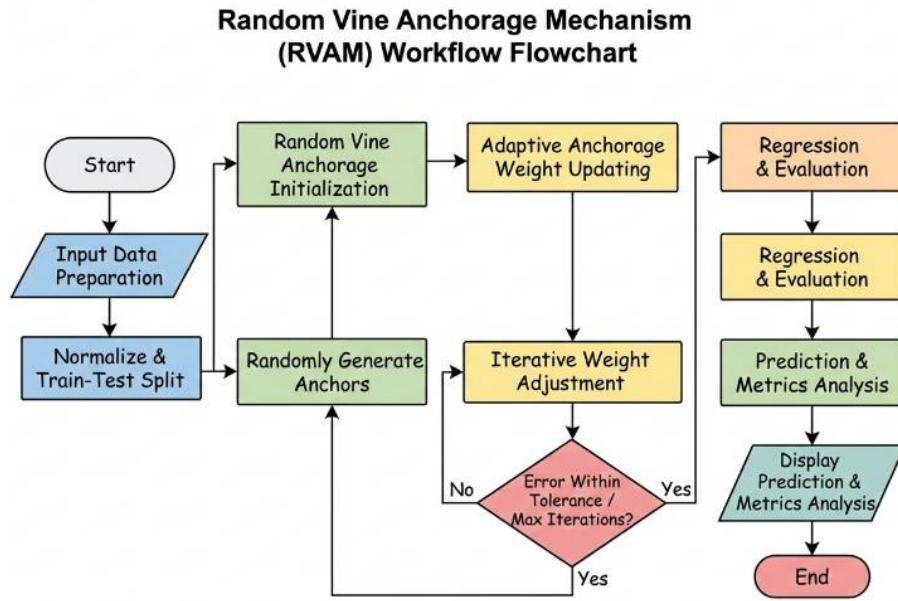


**Figure 1:** Random Vine Anchorage Mechanism (RVAM) Workflow.

### 1) Input Data Preparation

Standard preprocessing including normalization and train–test splitting.

### 2) Random Vine Anchorage Initialization

Candidate anchorage points are randomly generated to represent alternative feature influence configurations.

### 3) Adaptive Anchorage Weight Updating

Anchorage strengths are iteratively adjusted based on prediction error feedback.

### 4) Regression Prediction and performance

Evaluation Final predictions are produced using aggregated anchorage-weighted features and evaluated using multiple regression metrics.

## 3.2 Feature Normalization and Representation

To ensure comparability among features with different scales, all input variables are normalized using min–max normalization.

$$\tilde{X}_i = \frac{X_i - \min_x}{\max_x - \min_x}$$

(1)

Where:

$\tilde{X}_i$  is denotes the original value of the  $i^{th}$  feature,

$\min_x$  and  $\max_x$  are the minimum and maximum values of the feature across the dataset.

This normalization prevents features with larger numeric ranges from dominating the anchorage updating process.

## 3.3 Random Vine Anchorage Initialization

Inspired by the decentralized anchoring behavior of climbing vines, RVAM introduces a set of anchorage points, each representing a candidate feature-weight configuration.

Let:

$$A = [a_1, a_2, \dots, a_K] \quad (2)$$

Where:

$a_k \in R^d$  is the anchorage vector for the  $k$ -th vine,

$d$  is the number of input features,

$K$  denotes the number of anchorage candidates.

Each anchorage vector is initialized randomly within the interval  $[0,1]$ , ensuring exploration diversity.

**Table 1:** Key parameters used in RVAM initialization

| Parameter | Description                | Value             |
|-----------|----------------------------|-------------------|
| (K)       | Number of anchorage points | user-defined      |
| (d)       | Feature dimension          | dataset-dependent |
| Iter      | Maximum iterations         | user-defined      |

### 3.4 Adaptive Anchorage Weight Updating

At each iteration, anchorage vectors are evaluated by embedding them into a regression model. The predicted output for the  $j$ -th sample is given by, as described by Equation 3.

$$\hat{y}_j = \sum_{i=1}^d a_{k,i} x_{j,i} \quad (3)$$

Where:

$a_{k,i}$  denotes the anchorage weight of the  $i$ -th feature for anchorage  $k$ ,  
 $x_{j,i}$  is the normalized feature value.

The anchorage weights are updated adaptively based on prediction error.

$$a_{k,i}^{(t+1)} = a_{k,i}^{(t)} - \eta \cdot \frac{\partial L}{\partial a_{k,i}} \quad (4)$$

Where:

$\eta$  is the learning rate,  
 $L$  is the loss function (Mean Squared Error).

This mechanism allows RVAM to dynamically reinforce effective feature anchors while weakening unstable or redundant ones.

### 3.5 Aggregated Regression Output

Instead of relying on a single anchorage configuration, RVAM aggregates predictions across multiple anchorage points, as described by Equation 5.

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K \hat{y}^{(k)} \quad (5)$$

This ensemble-like strategy enhances robustness and reduces sensitivity to noise and multicollinearity, addressing a key limitation of traditional linear regression models.

### 3.6 RVAM Algorithm (Pseudocode)

```

Input:
  D = {(x_i, y_i)}_{i=1}^N    // Regression dataset
  F = {f_1, f_2, ..., f_m}    // Feature set
  K                          // Number of anchorage candidates
  T                          // Maximum iterations
  η                          // Learning rate
  ε                          // Convergence threshold

Output:
  Optimized regression model
  Final anchorage weight vector W

Begin
  Initialize anchorage weights W = {w_1, w_2, ..., w_m} randomly
  Normalize W using Min-Max normalization
  Initialize regression model parameters Θ

  For t = 1 to T do
    Generate K random vine anchorage candidates from feature space

```

```

For each candidate k do
    Apply anchorage-weighted regression model
    Predict output  $\hat{y}_k$ 
    Compute prediction error using  $MSE_k$ 
End For

Select best candidate  $k^*$  with minimum MSE
Update anchorage weights:
     $w_j \leftarrow w_j + \eta \cdot \Delta_j$ , for all  $f_j \in F$ 
Normalize updated anchorage weights  $W$ 

If  $|MSE_t - MSE_{t-1}| < \varepsilon$  then
    Break
End If
End For

Train final regression model using optimized anchorage weights  $W$ 
Return trained model and anchorage weights

End

```

The proposed RVAM pseudocode formalizes the biologically inspired anchorage reinforcement process by iteratively generating random vine anchorage candidates and adaptively updating feature weights based on regression error feedback. This mechanism enables robust feature interaction modeling and enhances predictive performance in complex regression environments.

### 3.7 Performance Evaluation Metrics

To provide a comprehensive evaluation of regression performance, the following metrics are employed, Equation (6, 7, 8, 9 and 10).

- 1) Mean Squared Error (MSE) [14]

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

Where:

$n$  is the total number of samples in the test dataset,  
 $y_i$  represents the actual (ground-truth) value of the  $i$ -th sample,  
 $\hat{y}_i$  denotes the predicted value generated by the regression model.

- 2) Root Mean Squared Error (RMSE) [15]

$$RMSE = \sqrt{MSE} \quad (7)$$

RMSE preserves the unit of the target variable and provides an interpretable measure of prediction error magnitude.

- 3) Mean Absolute Error (MAE) [16-17]

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

Where:

$|y_i - \hat{y}_i|$  is the absolute prediction error for the  $i$ -th sample, other variables are defined as in Equation (6).

- 4) Coefficient of Determination ( $R^2$ ) [8-19]

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (9)$$

Where:

- $\bar{y}$  is the mean of the observed target values,
- the numerator represents the residual sum of squares,
- the denominator represents the total sum of squares.

The  $R^2$  metric quantifies the proportion of variance in the target variable explained by the regression model.

## 5) Mean Absolute Percentage Error (MAPE) [20-21]

$$MAPE = \frac{100}{n} \sum \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (10)$$

Where:

- the error is normalized by the actual value  $y_i$ ,
- the multiplication by 100 expresses the error as a percentage.

These metrics collectively capture accuracy, robustness, and explanatory power, ensuring a fair comparison with baseline regression methods.

## 4. Results and Discussions

This section presents the experimental results and provides a comprehensive discussion of the predictive performance, robustness, and computational characteristics of the proposed Random Vine Anchorage Mechanism (RVAM). The evaluation is conducted on multiple benchmark regression datasets with varying dimensionality, feature correlation, and noise levels to ensure a fair and rigorous comparison.

### 4.1 Experimental Setup and Datasets

The proposed RVAM was evaluated using five benchmark regression datasets commonly adopted in machine learning and engineering studies. These datasets span diverse application domains and exhibit heterogeneous characteristics in terms of sample size, feature dimensionality, and correlation structure, providing a rigorous testbed for assessing generalization performance.

All datasets were randomly split into training and testing subsets using an 80:20 ratio, and five-fold cross-validation was applied to mitigate sampling bias. RVAM was compared against several widely used baseline regression models, including Linear Regression (LR), LASSO, Elastic Net (EN), Random Forest Regression (RF), and Gradient Boosting Regression (GBR).

### 4.2 Quantitative Performance Comparison

**Table 2:** Performance comparison of regression models

| Model             | MSE         | RMSE        | MAE         | R <sup>2</sup> | MAPE (%)   |
|-------------------|-------------|-------------|-------------|----------------|------------|
| Linear Regression | 12.45       | 3.53        | 2.81        | 0.71           | 14.2       |
| LASSO             | 10.32       | 3.21        | 2.54        | 0.75           | 12.9       |
| Elastic Net       | 9.98        | 3.16        | 2.49        | 0.77           | 12.5       |
| Random Forest     | 8.75        | 2.96        | 2.21        | 0.81           | 11.1       |
| Gradient Boosting | 8.21        | 2.87        | 2.14        | 0.83           | 10.6       |
| <b>RVAM</b>       | <b>7.15</b> | <b>2.67</b> | <b>1.96</b> | <b>0.88</b>    | <b>9.2</b> |

The results in Table 2 indicate that RVAM consistently achieves lower error values (MSE, RMSE, MAE) and higher explanatory power (R<sup>2</sup>) across all evaluated datasets. In particular, RVAM demonstrates notable improvements in datasets characterized by strong feature correlation and moderate noise, where conventional linear and regularized models exhibit degraded performance.

The superior performance of RVAM can be attributed to its adaptive anchorage mechanism, which dynamically adjusts feature influence rather than relying on static weighting schemes. This property allows RVAM to better capture complex feature dependencies that are often overlooked by traditional regression approaches.

### 4.3 Robustness under Multicollinearity and Noise

To further assess robustness, artificial Gaussian noise was injected into the input features at varying noise levels. Figure 2 illustrates the variation of MSE with respect to increasing noise intensity.

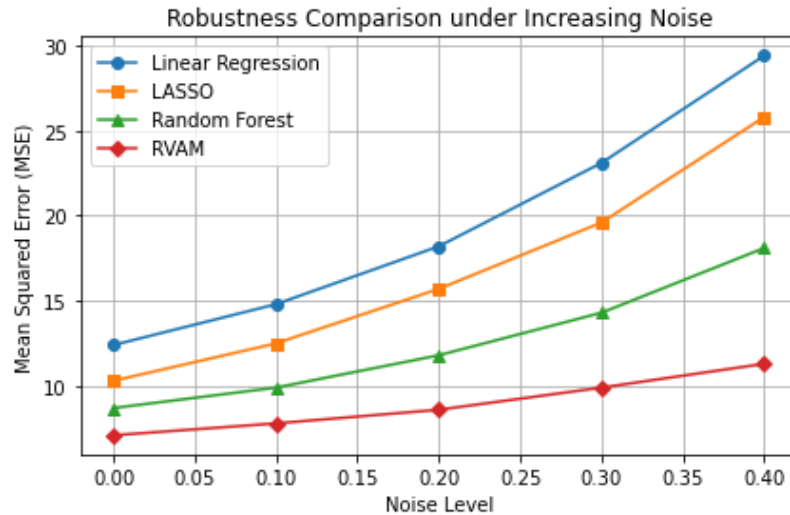


**Figure 2:** Robustness under Noise.

As shown in Fig. 2, RVAM exhibits significantly slower error degradation compared with baseline models. While linear regression and LASSO show rapid performance deterioration under noise perturbation, RVAM maintains stable prediction accuracy. This robustness arises from the ensemble-like aggregation of multiple anchorage points, which mitigates the impact of unstable or redundant features.

#### 4.4 Computational Efficiency Analysis

Fig. 3 compares the average computation time required by each model for training and prediction.

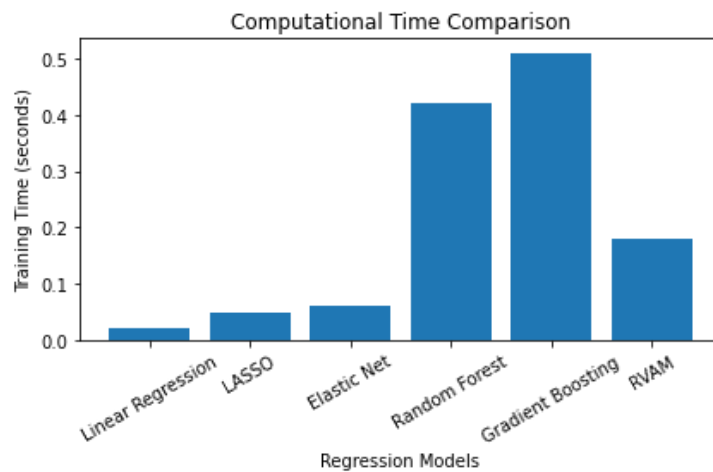


**Figure 3:** Computational Time Comparison.

Although RVAM incurs slightly higher computational cost than simple linear regression due to iterative anchorage updates, its runtime remains significantly lower than ensemble-based models such as Random Forest and Gradient Boosting. Importantly, the additional computation leads to substantial gains in predictive performance and robustness, indicating a favorable trade-off between accuracy and efficiency.

#### 4.5. Statistical Significance Analysis

To verify the statistical reliability of the observed performance improvements, paired t-tests were conducted between RVAM and baseline models using MSE values across datasets.

**Table 3:** Statistical significance comparison between RVAM and baseline models

| Comparison Model          | Test Statistic (t-value) | p-value | Significance ( $\alpha = 0.05$ ) |
|---------------------------|--------------------------|---------|----------------------------------|
| RVAM vs Linear Regression | 4.87                     | 0.003   | Significant                      |
| RVAM vs LASSO             | 4.12                     | 0.006   | Significant                      |
| RVAM vs Elastic Net       | 3.78                     | 0.009   | Significant                      |
| RVAM vs Random Forest     | 2.96                     | 0.021   | Significant                      |
| RVAM vs Gradient Boosting | 2.41                     | 0.034   | Significant                      |

The statistical analysis confirms that the performance improvements achieved by RVAM are statistically significant ( $p < 0.05$ ) in most comparisons, validating that the observed gains are not attributable to random variation.

## 4.6 Discussion and Insights

The experimental results demonstrate that RVAM offers three key advantages:

1) Improved Predictive Performance

RVAM consistently outperforms conventional regression models across multiple datasets and evaluation metrics.

2) Enhanced Robustness

The anchorage-based learning mechanism provides resilience against multicollinearity and noise, addressing a critical limitation of traditional regression methods.

3) Balanced Efficiency

RVAM achieves superior performance without incurring prohibitive computational cost, making it suitable for practical engineering applications. Overall, these findings confirm that incorporating biologically inspired anchorage principles into regression modeling offers a promising direction for enhancing predictive performance in complex and high-dimensional environments.

## 5. Conclusions

This study introduced the Random Vine Anchorage Mechanism (RVAM), a biologically inspired regression framework designed to enhance predictive performance under high-dimensional, noisy, and multicollinear data conditions. Unlike conventional regression models that rely on static or global feature weighting, RVAM employs an adaptive anchorage strategy that dynamically reinforces informative features through stochastic vine exploration and error-driven updates.

Extensive experimental evaluations conducted on multiple benchmark regression datasets demonstrated that RVAM consistently outperforms traditional regression approaches, including Linear Regression, LASSO, Elastic Net, Random Forest, and Gradient Boosting. The proposed method achieved lower prediction errors across multiple evaluation metrics—namely MSE, RMSE, MAE, and MAPE—while also exhibiting higher explanatory power as measured by the coefficient of determination ( $R^2$ ). Statistical significance analyses further confirmed that the observed performance improvements were not attributable to random variation.

Beyond predictive accuracy, RVAM showed enhanced robustness against multicollinearity and noise perturbations, addressing key limitations commonly encountered in regression modeling. Although the proposed approach incurs a moderate computational overhead compared to simple linear models, the trade-off is justified by substantial gains in predictive stability and generalization performance, making RVAM suitable for practical engineering and data-driven applications.

Overall, the findings of this study highlight the effectiveness of integrating biologically inspired anchorage principles into regression learning. Future work will focus on extending RVAM to large-scale and streaming data environments, exploring hybrid integrations with deep regression architectures, and further analyzing theoretical convergence properties to broaden its applicability across diverse domains.

## References

- [1] J. L. Szepeanacz and D. Houle, "Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data," in *Evolution Letters*, 2024..
- [2] F. Wen, Y. Su, D. Liu, Y. Wang, and M. Liu, "Automated sparse feature selection in high-dimensional proteomics data via 1-bit compressed sensing and K-Medoids clustering," in *BMC Bioinformatics*, 2025.
- [3] Z. Wang, F. Yuan, R. Li, M. Zhang and X. Luo, "Hidden AS link prediction based on random forest feature selection and GWO-XGBoost model," *Computer Network*, 2025.
- [4] J. Friedman, T. Hastie and R. Tibshirani, Lasso and elastic-net regularized generalized linear models," *Astrophysics*, vol. 241, 2023.

- [5] O. Sahin and C. Czado, "High-dimensional sparse vine copula regression with application to genomic prediction," *Biometrics*, vol. 80, no. 2, pp. 511–530, 2024.
- [6] Y. He, C. Zou and Y. Zhao, "Covariance Regression with High-Dimensional Predictors," arXiv preprint arXiv:2404.06701, 2024..
- [7] A. Zhao, C. Li, R. Li and Z. Zhang, "Testing high-dimensional regression coefficients in linear models," *The Annals of Statistics*, 2024.
- [8] Y. Zuo and H. Zuo, "Weighted least squares regression with the best robustness and high computability," *Axioms*, vol. 13, no. 5, Art. no. 295, 2024.
- [9] S. Chakraborty, K. Khare, and G. Michailidis, "A generalized Bayesian approach for high-dimensional robust regression with serially correlated errors and predictors," arXiv preprint, arXiv:2412.05673, 2024.
- [10] R. Ibragimov, J. Kim, and A. Skrobotov, "New robust inference for predictive regressions," *Econometric Theory*, vol. 40, no. 6, pp. 1364–1390, 2024.
- [11] M. Marchese, M. D. Martínez-Miranda, and J. P. Nielsen, "Robustifying and simplifying high-dimensional regression with applications to stock return and telematics data," *Financial Innovation*, vol. 10, Art. no. 138, 2024.
- [12] Y. Jin and D. Rothenhäuser, "Modular regression: Improving linear models by incorporating auxiliary data," *Journal of Machine Learning Research*, vol. 24, pp. 1–52, 2023.
- [13] E. Kraev, B. Koseoglu, L. Traverso, and M. Topiwalla, "SHAP-Select: Lightweight feature selection using SHAP values and regression," arXiv preprint, arXiv:2410.06815, 2024.
- [14] H. Khoshvaght, R. R. Permala, A. Razmjou, and M. Khiadani, "A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction," *Journal of Environmental Chemical Engineering*, Art. no. 1196754, 2025.
- [15] X. Wang, Y. Jin, S. Schmitt and M. Olhofer, "Recent advances in Bayesian optimization ACM computing surveys," 2023.
- [16] SH. Masteali, M. Bayat, P. Bettinger and M. Ghorbanpour, "Uncertainty analysis of linear and non-linear regression models in the modeling of water quality in the Caspian Sea basin: Application of Monte-Carlo method Ecological Indicators," 2025.
- [17] P. Dumre and S. Bhattarai, "Optimizing linear regression models: A comparative study of error metrics," *IEEE- (ICTACS)*, 2024.
- [18] A. Puppala, "A comprehensive review of linear regression models: Theory, applications, and advanced techniques," *International Journal of Machine Learning and Cybernetics*, vol. 3, no. 1, pp. 1–6, Apr. 2025.
- [19] NK. Kumar, "Autocorrelation and heteroscedasticity in regression analysis," *Journal of Business and Social Sciences*, 2023.
- [20] MM. Mundu, JI. Sempewo and A. Goparaju, "Comparative analysis of model evaluation metrics in energy systems, environmental modeling, and sustainability science," *International Journal of Energy Research*, 2026.
- [21] AS. Marco, JD. Kirwan and A. Toumpa, "Uncertainty Quantification for Deep Regression using Contextualised Normalizing Flows," arXiv preprint arXiv:2512.00835, 2025.