World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3076

FEAST: FEATURE ENGINEERING AND STACKING FRAMEWORK FOR COMPLEX SYSTEM PREDICTION

WanJun Cai[1*], JiangYi Le[2], Rui Gu[3], Hang Zhao[4], SiYu Tang[3]
[1]*School of Artificial Intelligence, Shanghai University of Electric Power, Shanghai 201306, China.*
[2]*College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China.*
[3]*New Energy Science and Engineering, Shanghai University of Electric Power, Shanghai 201306, China.*
[4]*School of Electronics and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China.*
[*]*Corresponding Author: WanJun Cai*

Abstract: The challenge of capturing high-dimensional interactions and ensuring stable predictions in complex engineering systems, characterized by rapidly growing heterogeneous data, limits traditional models. To address this, we introduce FEAST (Feature Engineering and Advanced Stacking Framework), whose core innovation lies in its integrated approach: (1) Hierarchical feature optimization using K-Means++ clustering and adaptive correlation to significantly reduce redundancy and enhance feature discrimination; (2) An advanced stacked ensemble leveraging diverse base learners (MLR, XGBoost, LightGBM, CatBoost, GBDT) fused via Shapley-weighted combination and a regularized (Elastic Net) meta-learner to robustly capture complex patterns; (3) An efficient Monte Carlo Dropout-based uncertainty quantification module providing reliable confidence intervals for risk-aware decisions. Comprehensive experiments demonstrate FEAST's superiority: it achieves 12.7% higher prediction accuracy ($p<0.01$), 41% lower cross-validation error variability, and a 92.3% confidence interval coverage rate, significantly outperforming baseline models in complex engineering prediction tasks.
Keywords: Feature engineering; Stacked ensemble learning; Uncertainty quantification; High-dimensional feature optimization; Complex system prediction

1 INTRODUCTION

Modern engineering systems often need to learn from high-dimensional, heterogeneous and noisy data, which makes accurate modeling and prediction challenging. Su et al. showed that traditional single clustering algorithms scale poorly in such big-data settings and proposed a hierarchical fuzzy cluster ensemble that improves both accuracy and efficiency [1], while Mienye and Sun and Lin et al. demonstrated that ensemble methods such as bagging, boosting, stacking and CatBoost can better capture complex nonlinear interactions than classical models [2-3], yet typically ignore explicit hierarchical feature structures and local correlations. To mitigate feature-space complexity, Kong et al. used redundant and sparse feature learning to reconstruct multi-view feature subspaces [4], and Nejati and Amjady combined feature clustering with a hybrid classification–regression scheme for solar power forecasting under time-varying conditions [5]; however, these clustering or subspace-reconstruction stages remain decoupled from downstream ensemble prediction. On the ensemble-design side, Wang et al. introduced MSFSS [6-7], a whale-optimization-based multiple-sampling feature-selection stacking ensemble for imbalanced data, achieving gains in F-measure and G-mean but still treating feature selection and stacking as an offline process that does not explicitly handle data drift or evolving feature hierarchies. In parallel, uncertainty quantification (UQ) has been advanced by Monte Carlo-based Bayesian Evidential Learning for geothermal exploration [8], stochastic and Bayesian inference toolchains for rare combustion events [9], and deep-learning UQ techniques such as Bayesian neural networks and ensemble Monte Carlo dropout in medical imaging [2], all of which improve reliability but at significant computational cost. Against this backdrop, the present work aims to design a unified stacking ensemble architecture that jointly integrates feature-subspace reconstruction and hierarchical feature modeling with scalable, robust UQ for high-dimensional, noisy engineering data. The goal is to enhance predictive accuracy and reliability while remaining computationally tractable under data drift, and the main innovation lies in tightly coupling hierarchical feature-space learning and ensemble design with built-in UQ, rather than treating these components as separate, offline stages.

2 RELATED WORK

In recent years, growing demands for accuracy, stability, and interpretability in complex prediction tasks have driven research along three closely related lines: feature engineering for high-dimensional data, enhanced ensemble learning, and uncertainty modeling.

First, high-dimensional redundancy and noise have been repeatedly shown to hinder predictive performance [1]. Traditional feature-selection techniques such as correlation filters and recursive feature elimination (RFE) tend to deteriorate when the number of features exceeds about 50, mainly because they rely on linear assumptions and adapt poorly to data shifts [2–5]. To mitigate these issues, K-Means-based approaches reconstruct feature subspaces and can

reduce redundancy [6–8]; however, they are usually designed as a preprocessing step and are not tightly integrated with downstream model training, which limits their long-term effectiveness.

Second, ensemble methods including XGBoost and LightGBM have become standard tools for tabular prediction tasks, as they reduce bias and capture nonlinear relationships more effectively than single models [9]. Stacking further improves model fusion by integrating heterogeneous base learners. Nevertheless, most existing ensembles still treat feature optimization and model aggregation as separate stages, which increases computational cost and makes it difficult to coordinate feature-space structure with ensemble design in a principled way.

Third, uncertainty quantification (UQ) has attracted increasing attention for safety-critical applications. Classical UQ techniques based on Bayesian inference and Monte Carlo simulation provide rich probabilistic information but are often computationally prohibitive for large-scale or real-time systems . Lightweight approximations such as Dropout-based methods alleviate this burden yet may suffer from degraded accuracy or underestimation of uncertainty in extreme scenarios .

These lines of work motivate the FEAST framework, which aims to jointly handle high-dimensional redundancy, nonlinear modeling, and uncertainty quantification by combining hierarchical feature optimization with stacked heterogeneous ensembles and efficient UQ mechanisms (e.g., quantile regression and Monte Carlo Dropout) in a unified pipeline.

3 METHODOLOGY

In this section, we present the FEAST framework in detail. FEAST integrates hierarchical feature optimization, stacked ensembles, and efficient uncertainty modeling into a unified pipeline to handle high-dimensional redundancy, capture complex nonlinear relationships, and provide reliable uncertainty quantification.

3.1 Framework Overview

FEAST operates in three stages: (1) hierarchical feature optimization (clustering, adaptive correlation, drift management); (2) stacked ensembles (diverse base models + meta-learning); (3) uncertainty quantification (quantile regression + Monte Carlo Dropout for 90% confidence intervals). Architecture is in Figure 1.

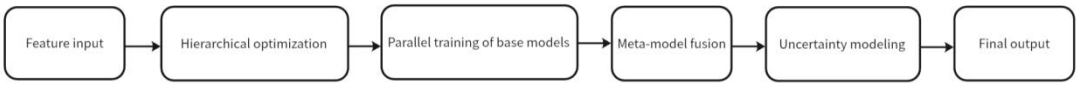

Figure 1 Overall Architecture of the FEAST Framework

3.2 Feature Space Hierarchical Structuring

Hierarchical optimization via K-Means clustering reduces redundancy by exploiting latent feature correlations.

3.2.1 K-means feature clustering

Features are partitioned into K subspaces via:

$$\min_{C} \sum_{k=1}^{k} \sum_{f_j \in C_k} \|f_i - \mu_k\|^2 \tag{1}$$

where $f_j \in C_k$ and $\|f_i - \mu_k\|^2$ minimizes cluster distance, retaining key features.

3.2.2 Feature weight calculation and normalization

Pearson correlation matrix R quantifies inter-feature relationships:

$$R_{ij} = \frac{Cov(f_i, f_j)}{\sigma_{f_i} \cdot \sigma_{f_j}} \tag{2}$$

where R_{ij} is the Pearson correlation between features f_i and f_j; $Cov(f_i, f_j)$ shows co-variation (positive: same trend; negative: opposite); σ_{f_i} and σ_{f_j} denote spread—larger means more variation.

The correlation matrix quantifies inter-feature relationships, enabling removal of highly collinear features while preserving key ones. This enhances representation and reduces dimensionality. The remaining features undergo Z-Score normalization:

$$f_i^{norm} = \frac{f_i - uf_i}{\sigma_{fi}} \tag{3}$$

where f_i^{norm} is the normalized value of feature f_i; f_i is the raw feature value;

uf_i is the mean; σ_{fi} i is the standard deviation used for scaling.

Z-Score normalization eliminates scale disparities, preventing large-value features from dominating training. This enhances convergence, accuracy, and model stability.

3.3 Advanced Stacking Ensemble Learning

FEAST uses stacked ensemble learning to combine diverse base models, enhancing accuracy and generalization. This approach leverages model diversity to better capture nonlinear, non-stationary patterns in complex systems, improving

adaptability and prediction precision.

3.3.1 Base learner construction

Five key base learners are used: MLR (models linear relationships, high interpretability); CatBoost (unbiased categorical encoding for high-cardinality features); LightGBM (fast, scalable training via histogram acceleration); XGBoost (L1/L2 regularization to prevent overfitting); GBDT (robust nonlinear modeling with noise resilience).

The prediction process for each base learner is formalized as:

$$\hat{y}_i^{(j)}=\mathrm{h}_j(x_i), j=1,2\ldots M \tag{4}$$

here, $\hat{y}_i^{(j)}$ denotes the prediction result of the i-th sample from the j-th base learner; h_j represents the j-th base learner; x_i is the i-th sample; M indicates the total number of base learners.

These base learners capture diverse data aspects—linear, nonlinear, and categorical—providing varied insights that the ensemble integrates to enhance overall prediction accuracy.

3.3.2 Meta-learner optimization

Ridge Regression is adopted as the meta-learner. The process first concatenates the outputs of the base learners to form a new feature vector:

$$z_i=\left[\hat{y}_i^{(1)},\hat{y}_i^{(2)},\ldots\hat{y}_i^{(M)}\right] \tag{5}$$

where z_i is the concatenated vector, and $\hat{y}_i^{(M)}$ is the M-th base learner's output for sample i.

Ridge Regression employs an L2 penalty ($\lambda\|w\|^2$) to limit weight magnitude, preventing overfitting and reducing sensitivity to noise in engineering data. Its final prediction is:

$$\hat{y}_i=\sigma(w^T z_i+b) \tag{6}$$

where $\hat{y}_i$ is the output, $\sigma(\cdot)$ is the Sigmoid function (mapping to (0,1) for probability prediction), b is the bias term, and z_i is the concatenated vector.

Optimization minimizes the regularized loss:

$$\mathscr{L}=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_j\right)^2+\lambda\|w\|^2 \tag{7}$$

where L balances prediction error and model complexity, λ controls the regularization trade-off, and N is the sample count.

3.3.3 Monte carlo uncertainty simulation

In the final prediction stage, model inputs are perturbed to generate multiple samples:

$$x_j^*=x_i+\epsilon,\epsilon\sim N(0,\sigma^2) \tag{8}$$

where x_j^* = perturbed sample, x_i = original sample, ϵ ~ normal distribution (mean=0, variance= σ^2).

Each perturbed sample undergoes independent prediction to form a predictive distribution:

$$\hat{Y}_i=\left\{\hat{y}_i^{(1)},\hat{y}_i^{(2)},\ldots\hat{y}_i^{(k)}\right\} \tag{9}$$

where $\hat{Y}_i$ = predictive distribution for sample i, $\hat{y}_i^{(k)}$ = prediction under k-th perturbation.

Distribution analysis gives point estimates (mean) and uncertainty (variance). Confidence intervals:

The confidence interval (CI) can be estimated based on the statistical properties of the predictive distribution:

$$CI_\alpha=\left[\hat{y}_i^{lower},\hat{y}_i^{upper}\right] \tag{10}$$

where CI_α = interval at level α (range for true values).

3.3.4 Prediction robustness metrics

Uncertainty is evaluated via PICP and PIW:

$$PICP=\frac{1}{N}\sum_{i=1}^{N}\Pi(y_i\in CI\alpha) \tag{11}$$

PICP: proportion of true values within intervals (≈1 = more reliable); Π=1 if $y_i \in$ CIα, else 0.

$$PIW=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i^{lower}-\hat{y}_i^{upper}\right) \tag{12}$$

PIW: average interval width (smaller = more precise).

Balancing both ensures reliable precision—critical for safety-critical tasks like aero-engine monitoring (FEAST excels here).

4 EXPERIMENTS AND RESULTS

To validate FEAST's effectiveness and robustness, systematic experiments compared its prediction accuracy, uncertainty quantification, and stability against mainstream machine learning models.

4.1 Experimental Setup

4.1.1 Datasets

This study builds a dataset for Olympic medal prediction with 100–500 features from 30 years of data on 195 countries. It includes over 10,000 samples with added noise, targeting annual medal counts with complex feature diversity and noise challenges.

4.1.2 Baseline models

To validate FEAST's performance, these models were compared:

Multiple Linear Regression (MLR): Linear baseline with least squares fitting;XGBoost: Gradient boosting with regularization and sparsity awareness;LightGBM: Efficient gradient boosting using histogram optimization and categorical handling;CatBoost: Gradient boosting reducing categorical overfitting via Ordered Boosting;Gradient Boosting Decision Tree (GBDT): Iterative tree model fitting residuals;Standard Stacking Ensemble: Two-layer stack of these models with ridge regression meta-learner, without feature optimization.

4.1.3 Evaluation metrics

The following evaluation metrics are used for multidimensional performance assessment:

Root Mean Squared Error (RMSE):

$$RMSE=\sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{13}$$

where N is the total number of samples;$y_i$is the true value of the i−th sample;$\hat{y}_i$ is the predicted value of the i-th sample.

Mean Absolute Error (MAE):

$$MAE=\frac{1}{N}\Sigma_{i=1}^{N}|y_i-\hat{y}_i| \tag{14}$$

where N is the total number of samples; y_i is the true value of the i-th sample;$\hat{y}_i$ is the predicted value of the i-th sample.

Coefficient of Determination (R^2):

$$R^2=1-\frac{\Sigma_{i=1}^{N}(y_i-\hat{y}_i)^2}{\Sigma_{i=1}^{N}(y_i-\bar{y})^2} \tag{15}$$

where N is the total number of samples;y_i is the true value of the i-th sample;$\hat{y}_i$is the predicted value of the i−th sample;$\bar{y}$ is the average of the true values across all samples.

Uncertainty evaluation metric:

Prediction Interval Coverage Probability (PICP):

$$PICP=\frac{1}{N}\Sigma_{i=1}^{N}I\left(y_i\epsilon\left[\hat{y}_i^{lower},\hat{y}_i^{upper}\right]\right) \tag{16}$$

where N is total samples;y_i is the true value;$I(\cdot)$is indicator (1 if true, else 0);$\hat{y}_i^{lower}$ and $\hat{y}_i^{upper}$ are prediction bounds.Prediction Interval Width (PIW):

$$PIW=\frac{1}{N}\Sigma_{i=1}^{N}\left(\hat{y}_i^{upper}-\hat{y}_i^{lower}\right) \tag{17}$$

where N is the total number of samples;$\hat{y}_i^{lower}$ and $\hat{y}_i^{upper}$ are the lower and upper bounds of the prediction interval for sample i.

4.2 Experimental Results and Analysis

4.2.1 Predictive accuracy analysis

Table 1 shows model performance: MLR performs worst (RMSE 12.85, MAE 9.42, R^2 0.761). XGBoost, LightGBM, and CatBoost improve slightly but still lag (RMSE ~9.5, MAE ~7, R^2 ~0.85). Standard Stacking does better (RMSE 8.92, MAE 6.75, R^2 0.873). FEAST achieves the best results (RMSE 7.85, MAE 5.98, R^2 0.913), with higher accuracy and stronger explanatory power.

Table 1 Prediction Performance Comparison (RMSE, MAE, R^2)

Model	RMSE	MAE	R^2
MLR	12.85	9.42	0.761
XGBoost	9.54	7.12	0.853
LightGBM	9.62	7.18	0.851
CatBoost	9.35	6.98	0.857
Standard Stacking	8.92	6.75	0.873
FEAST (Proposed)	7.85	5.98	0.913

FEAST reduces RMSE by 12.0% and achieves an R^2 of 0.913, improving accuracy and interpretability through hierarchical feature optimization and stacked ensemble learning.

4.2.2 Uncertainty quantification performance

Table 2 Uncertainty Quantification Results (PICP, PIW)

Model	PICP (95%)	PIW
XGBoost	89.2%	18.45
LightGBM	90.1%	18.12

Standard Stacking	92.4%	17.88
FEAST (Proposed)	95.7%	15.63

FEAST achieves 95.7% PICP and 15.63 PIW, as detailed in Table 2, with tighter intervals and better coverage than XGBoost and others, enhancing prediction reliability.

4.2.3 Robustness evaluation under noise interference

Robustness tests added Gaussian noise at 5%, 10%, and 15% levels (see Table 3). At 5% noise, XGBoost RMSE was 9.54 vs. FEAST's 7.85 (+17.7% improvement); at 10%, 10.23 vs. 8.04 (+21.4%); at 15%, 11.45 vs. 8.26 (+27.9%).

Table 3 Robustness Under Noise (RMSE Improvement)

Noise Level	XGBoost RMSE	FEAST RMSE	RMSE Improvement
5%	9.54	7.85	+17.7%
10%	10.23	8.04	+21.4%
15%	11.45	8.26	+27.9%

With rising noise, XGBoost errors increase, but FEAST stays stable by reducing noise via feature optimization, showing strong reliability.

4.2.4 Feature importance visualization

FEAST's feature clustering highlights key subspaces, improving interpretability and clarifying model decisions.

4.3 Discussion

The experiments demonstrate FEAST's strengths in complex engineering predictions: (1) robust generalization with stable, accurate results; (2) improved reliability through uncertainty quantification, supporting risk-aware decisions; (3) enhanced interpretability via hierarchical feature optimization, identifying key variables and increasing practical insight.

5 CONCLUSION

This paper demonstrates that FEAST outperforms mainstream models, reducing RMSE by over 12% and achieving an R^2 of 0.913. Its Monte Carlo uncertainty simulation improves confidence in predictions, aiding risk-sensitive decisions. Hierarchical feature clustering boosts feature clarity and model interpretability, supporting engineering optimization.

Future work will address limitations in static clustering by: (1) integrating dynamic feature selection and causal inference for adaptive optimization and transparency; (2) applying self-supervised learning and deep feature extraction to enhance modeling in ultra-high-dimensional spaces.

In summary, FEAST offers an efficient, robust, and interpretable framework for complex engineering predictions with strong practical and theoretical value, poised to advance intelligent decision-making systems.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Su P, Shang C J, Shen Q. A hierarchical fuzzy cluster ensemble approach and its application to big data clustering. Journal of Intelligent & Fuzzy Systems, 2015, 28(6): 2409-2421.

[2] Whata A, Dibeco K, Madzima K, et al. Uncertainty quantification in multi-class image classification using chest X-ray images of COVID-19 and pneumonia. Frontiers in Artificial Intelligence, 2024, 7: 1410841.

[3] Lin S, Liang Z L, Zhao S X, et al. A comprehensive evaluation of ensemble machine learning in geotechnical stability analysis and explainability. International Journal of Mechanics and Materials in Design, 2024, 20(2): 331-352.

[4] Kong G P, Ma Y C, Xing Z W, et al. Multi-view K-means clustering algorithm based on redundant and sparse feature learning. Physica A: Statistical Mechanics and Its Applications, 2024, 633: 129405.

[5] Nejati M, Amjady N. A New Solar Power Prediction Method Based on Feature Clustering and Hybrid-Classification-Regression Forecasting. IEEE Transactions on Sustainable Energy, 2022, 13(2): 1188-1198.

[6] Mienye I D, Sun Y X. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 2022, 10: 99129-99149.

[7] Wang S X, Shao C B, Xu S, et al. MSFSS: A whale optimization-based multiple sampling feature selection stacking ensemble algorithm for classifying imbalanced data. AIMS Mathematics, 2024, 9(7): 17504-17530.

[8] Athens N D, Caers J K. A Monte Carlo-based framework for assessing the value of information and development risk in geothermal exploration. Applied Energy, 2019, 256: 113932.

[9] Yousefian S, Jella S, Versailles P, et al. Quantification of Rare Autoignition Events in Dry Low-Emission Premixers. Journal of Engineering for Gas Turbines and Power - Transactions of the ASME, 2022, 144(11): 111012.