

Machine Learning-Based Direct Approximation of Optimized System Designs Through Taylor-Series-Based Problem Transcription

Shih-Yu Huang

Department of Industrial and Enterprise Systems Engineering
University of Illinois Urbana-Champaign
Urbana, Illinois, USA
shihyuh2@illinois.edu

Abstract

Engineering design optimization problems often rely on iterative numerical procedures to identify high-quality system configurations, particularly when uncertainty, coupled subsystems, or nonlinear constraints are present. Reliability-Based Co-Design (RBCD) represents one important class of such problems, where both physical and control design variables must be optimized while satisfying reliability requirements. Traditional approaches, including the First-Order Reliability Method (FORM) and the Hasofer-Lind-Rackwitz-Fiessler (HLRF) algorithm, are mathematically rigorous but frequently incur high computational cost, especially in high-dimensional or strongly nonlinear settings. To address this challenge, this paper proposes a machine learning-based framework for the direct approximation of optimized system designs through Taylor-series-based problem transcription. Objective and constraint functions are represented using structured Taylor-series polynomial forms. By randomly sampling polynomial coefficients, a large set of synthetic optimization problems is generated and solved offline using numerical optimization. The resulting dataset is used to train supervised learning models that learn the mapping from encoded problem definitions to their corresponding optimized design solutions. A two-stage learning framework consisting of feasibility classification and regression-based optimal solution prediction is developed. Results demonstrate strong predictive performance across nonlinear optimization problems. An engineering case study based on NACA 4412 spar-location optimization is also presented to evaluate the applicability of the proposed approach.

Keywords

Machine Learning, Design Optimization, Taylor Series Expansion, Gaussian Process Regression, Reliability-Based Design.

1. Introduction

Engineering optimization plays a critical role in modern system design, particularly in applications involving nonlinear objectives, coupled design variables, and reliability-related constraints. In Reliability-Based Co-Design (RBCD), both the physical structure and the control strategy must be jointly optimized to achieve system-level performance while satisfying probabilistic constraints (Cui et al. 2020). Traditional approaches, including the First-Order Reliability Method (FORM) and the Hasofer-Lind-Rackwitz-Fiessler (HLRF) algorithm, are mathematically rigorous but require iterative numerical optimization nested with reliability evaluation, resulting in extremely high computational cost when applied to complex, high-dimensional, or time-dependent systems (Valdebenito and Schueller 2010).

To reduce computational burden, surrogate-assisted optimization approaches have been widely studied in recent years (Forrester et al. 2008; Queipo et al. 2005). Existing methods commonly employ machine learning models such as Gaussian Process Regression (Rasmussen and Williams 2006), Neural Networks, and polynomial response surfaces to estimate objective or constraint functions. However, most existing methods still rely on iterative optimization procedures during the online design stage.

This paper proposes a fundamentally different framework based on the direct approximation of optimized system designs. Instead of approximating objective or constraint functions alone, the proposed approach directly learns the mapping between encoded optimization problem definitions and their corresponding optimized design solutions. Optimization problems are transformed into structured polynomial representations through Taylor-series-based problem transcription. Synthetic optimization datasets are generated by sampling polynomial coefficients and solving the resulting optimization problems offline, and supervised learning models are subsequently trained to predict optimized design variables.

1.1 Objectives

The primary objectives of this research are as follows:

1. Develop a Taylor-series-based problem transcription framework for representing optimization problems in a learnable structured polynomial form.
2. Construct a machine learning-based framework that directly approximates optimized system designs without repeatedly solving optimization loops online.
3. Develop a two-stage learning architecture consisting of feasibility classification and regression-based optimal solution prediction.
4. Validate the proposed framework on synthetic nonlinear optimization datasets and an engineering case study involving NACA 4412 spar-location optimization.

2. Literature Review

Machine learning and surrogate-assisted optimization methods have been widely investigated for solving computationally expensive engineering optimization problems. Common surrogate techniques include Gaussian Process Regression (GPR), polynomial response surfaces, Support Vector Regression (SVR), Radial Basis Functions (RBF), and Neural Networks (Rasmussen and Williams 2006; Forrester et al. 2008; Queipo et al. 2005). The Efficient Global Optimization (EGO) framework introduced by Jones et al. (1998) established the foundation for Bayesian optimization methods that balance exploration and exploitation for expensive optimization problems. More recent developments include adaptive sampling strategies, ensemble surrogates, and physics-informed neural networks that incorporate domain knowledge into the learning process (Raissi et al. 2019; Cuomo et al. 2022).

Polynomial representations and Taylor-series approximations have also played an important role in engineering optimization and reliability analysis. Local polynomial approximations are widely used in the First-Order Reliability Method (FORM) and the Second-Order Reliability Method (SORM), where Taylor-series expansions approximate nonlinear limit-state functions around critical design points (Hasofer and Lind 1974; Rackwitz and Fiessler 1978; Zhao and Ono 1999). Beyond local approximations, Polynomial Chaos Expansion (PCE) methods have been developed for global uncertainty quantification through orthogonal polynomial representations (Xiu and Karniadakis 2002; Sudret 2008; Blatman and Sudret 2011). Challenges associated with higher-order derivative estimation through Taylor series have been documented by Paudel et al. (2022).

Recent studies have further explored the integration of machine learning techniques into reliability-based design optimization (RBDO). Gaussian Process models provide probabilistic predictions with uncertainty estimation (Moustapha and Sudret 2019). Neural Networks have demonstrated strong capability in representing highly nonlinear and high-dimensional engineering relationships, including universal approximation properties established by Hornik et al. (1989). In reliability-based co-design optimization, machine learning-assisted frameworks have been investigated to reduce computational burden (Cui et al. 2020). Reliability-based design optimization formulations are comprehensively reviewed in Hu et al. (2024), and various single-loop and semi-single-loop methods have been proposed to address the double-loop computational burden (Liang et al. 2004; Lim and Lee 2016).

Despite these advances, most existing surrogate-assisted optimization frameworks primarily focus on approximating objective or constraint functions, and iterative optimization procedures are still required during online design stages. Stork et al. (2020) identified open issues in surrogate-assisted optimization, noting the difficulty of generalizing across different problem formulations. Van Hentenryck (2025) introduced the concept of optimization learning, which motivates the direct learning of solution mappings from problem parameters. In contrast, the framework proposed in this work transforms optimization problems into structured polynomial representations and directly learns the mapping between encoded optimization problem representations and optimized system designs.

3. Methods

The proposed framework represents optimization problems using structured polynomial formulations derived from Taylor-series expansions. For a two-dimensional design space consisting of design variables x_p and x_c , the objective and constraint functions are represented using truncated polynomial expansions:

$$f(x_p, x_c) = \sum_{i+j \leq k} (a_{ij} / i!j!) x_p^i x_c^j, \quad g(x_p, x_c) = \sum_{i+j \leq k} (b_{ij} / i!j!) x_p^i x_c^j \quad (1)$$

where k denotes the expansion order, and a_{ij} and b_{ij} represent the polynomial coefficients associated with the objective and constraint functions, respectively. The optimization problem is formulated as:

$$\min f(x_p, x_c) \quad \text{subject to} \quad g(x_p, x_c) \leq 0. \quad (2)$$

First-, second-, and third-order polynomial formulations are considered, with coefficients sampled within the discrete level set $\{-4, -2, 0, 2, 4\}$ to ensure diversity, numerical stability, and representative coverage of the problem space.

The proposed framework adopts a two-stage supervised learning architecture. A Random Forest (RF) classifier is first trained to determine whether a feasible solution exists for a given encoded optimization problem, using 400 estimators with balanced class weights. For feasible problems, regression models are trained to directly predict optimized design variables. Two regression models are investigated: Gaussian Process Regression (GPR) with an RBF kernel, and a multi-output Neural Network (NN) with a shared trunk architecture and dual output heads. For each generated optimization problem, Sequential Least Squares Programming (SLSQP) is used to compute the corresponding optimized solution offline. Outlier filtering based on the interquartile range (IQR) is performed before model training. The structured polynomial representation provides a consistent mathematical encoding across optimization problems with varying nonlinear characteristics. This enables supervised learning models to learn optimization-to-solution mappings within a unified feature space (Figure 1).

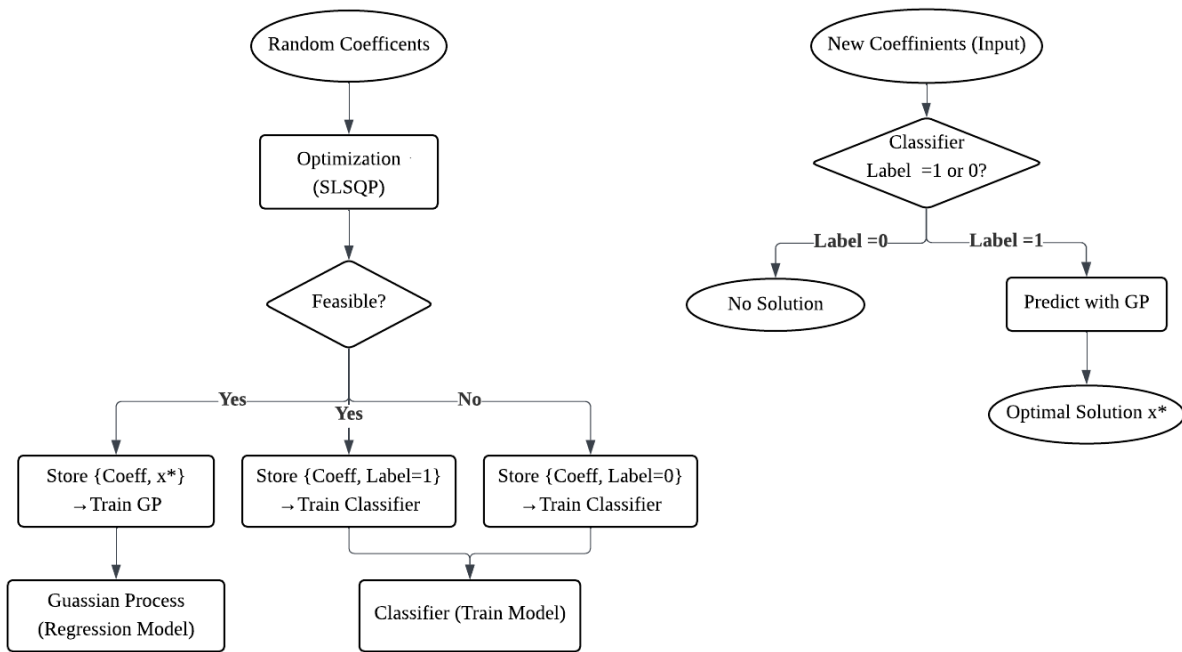


Figure 1. Overview of the proposed machine learning-based framework integrating data generation, feasibility classification, and regression-based optimal solution prediction.

4. Data Collection

Synthetic reliability-constrained design problems are generated using Taylor-series-based polynomial representations. The objective is to construct a large and diverse dataset of optimization problems, each defined by analytically

tractable objective and constraint functions together with their numerically computed optimal solutions.

For each problem instance, an expansion point (xp_0, xc_0) is randomly selected within the normalized design domain $[-1, 1]^2$. The constant term b_0 of the constraint function is calibrated such that $g(xp_0, xc_0) \leq -0.5$, ensuring the existence of an interior feasible region near the expansion point. For first- and second-order datasets, structured coefficient grids are generated from the discrete levels, with only a random subset retained for second-order problems to control dataset size. For third-order problems, coefficients are independently sampled due to the prohibitively large full Cartesian product. Each generated problem is solved using SLSQP with multiple starting points. Feasible samples are assigned label 1 and infeasible samples label 0. Table 1 summarizes the coefficient sampling strategy for each polynomial order.

Table 1. Coefficient sampling ranges and strategies for different polynomial orders

Order	Coefficient Range	Sampling Type	Purpose
1st	$\{-4, -2, 0, 2, 4\}$	Structured grid	Linear diversity and symmetry
2nd	$\{-4, -2, 0, 2, 4\}$	Grid + random subset	Moderate nonlinearity with manageable size
3rd	$\{-4, -2, 0, 2, 4\}$	Random level sampling	Strong nonlinearity with tractable generation

5. Results and Discussion

The proposed framework was evaluated using first-, second-, and third-order polynomial optimization datasets generated through Taylor-series-based problem transcription. Results are reported for the feasibility classifier, regression models, and the engineering case study.

5.1 Numerical Results

Table 2 summarizes the classification performance of the Random Forest classifier across second- and third-order problem datasets. The classifier achieved an accuracy of 95.6% on second-order datasets and 87.5% on third-order datasets. The decrease in accuracy for third-order problems reflects the increased nonlinear complexity of the encoded optimization space. The high precision, recall, and F1-score values indicate that the classifier reliably separates solvable and unsolvable optimization instances across both polynomial orders. The ROC-AUC value for the third-order dataset was 0.990, confirming strong discriminative capability (Figure 2).

Table 2. RF classifier performance for second- and third-order problem datasets

Order	Accuracy	Precision	Recall	F1-Score
2nd-order	0.956	0.957	0.956	0.955
3rd-order	0.875	0.873	0.874	0.873

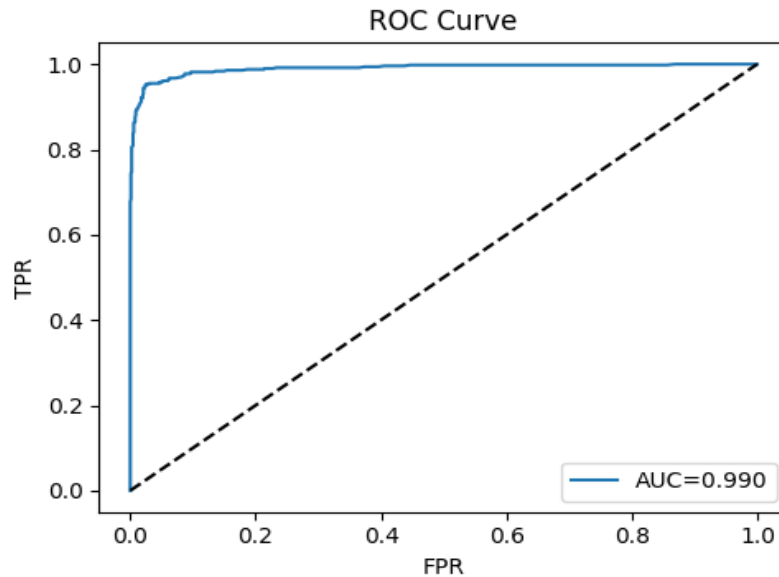


Figure 2. ROC curve and AUC for the RF classifier on third-order polynomial problems.

Tables 3 and 4 summarize the regression performance of the GPR and NN models on the second- and third-order datasets. For second-order problems, GPR achieved R^2 values of 0.826 and 0.846 for x^*p and x^*c , respectively, while NN achieved 0.863 and 0.880. For third-order problems, GPR yielded R^2 values of 0.817 and 0.774, and NN achieved 0.783 and 0.751. Overall, the NN shows stronger performance on second-order datasets, while GPR maintains competitive accuracy even under third-order nonlinearities. The engineering case study GP model achieved $R^2 = 0.82$ for x^*p (MSE = 0.0054) and $R^2 = 0.88$ for x^*c (MSE = 0.0019).

Table 3. GPR regression accuracy (R^2 and MSE) for second- and third-order datasets

Order	$R^2(x^*p)$	$R^2(x^*c)$	MSE(x^*p)	MSE(x^*c)
2nd-order	0.826	0.846	0.820	0.606
3rd-order	0.817	0.774	0.205	0.255

Table 4. NN regression accuracy (R^2 and MSE) for second- and third-order datasets

Order	$R^2(x^*p)$	$R^2(x^*c)$	MSE(x^*p)	MSE(x^*c)
2nd-order	0.863	0.880	0.645	0.474
3rd-order	0.783	0.751	0.244	0.282

5.2 Graphical Results

Figure 3 presents the predicted-versus-optimized solution comparison using Gaussian Process Regression for second-order optimization problems. Figure 4 presents the prediction performance of the Neural Network model for second-order problems. In both figures, predicted values are plotted against the true optimal solutions for x^*p and x^*c , with the ideal identity line shown as a dashed reference. Both models demonstrate strong agreement with the ground-truth solutions, with data points clustering closely around the ideal line. Figures 5 and 6 present the corresponding scatter plots for third-order datasets. While prediction accuracy decreases moderately for third-order problems due to stronger nonlinear interactions, both models maintain competitive predictive capability. Figure 7 illustrates the NACA 4412 spar-location optimization case study, showing front and rear spar positions at both reference and optimal locations.

The predicted optimal spar locations from the trained GP model show strong agreement with numerically optimized solutions, confirming the engineering applicability of the proposed framework.

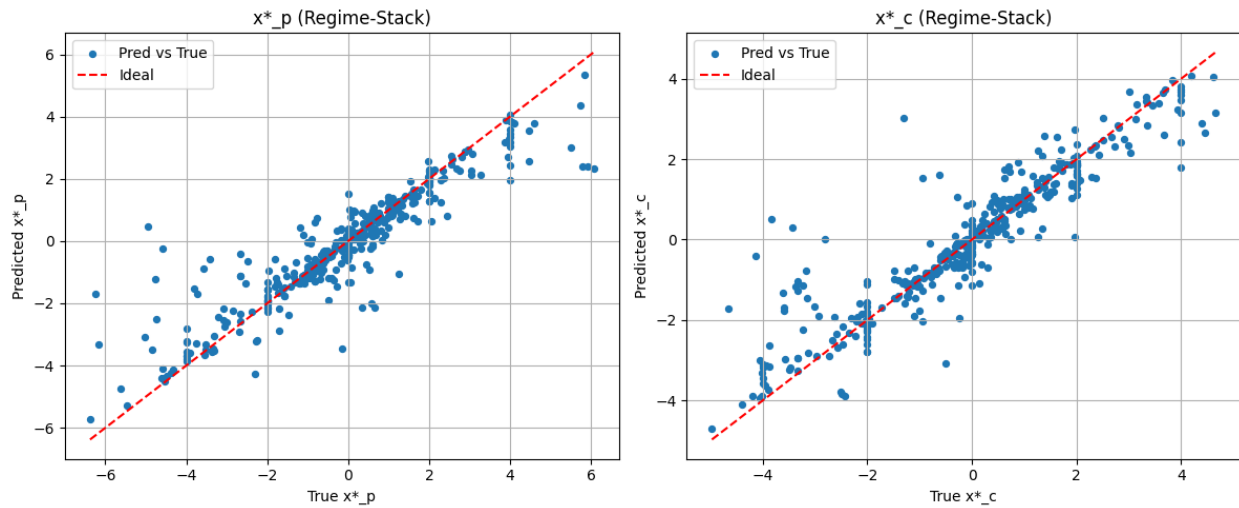


Figure 3. Predicted versus optimized design solutions using Gaussian Process Regression for second-order optimization problems.

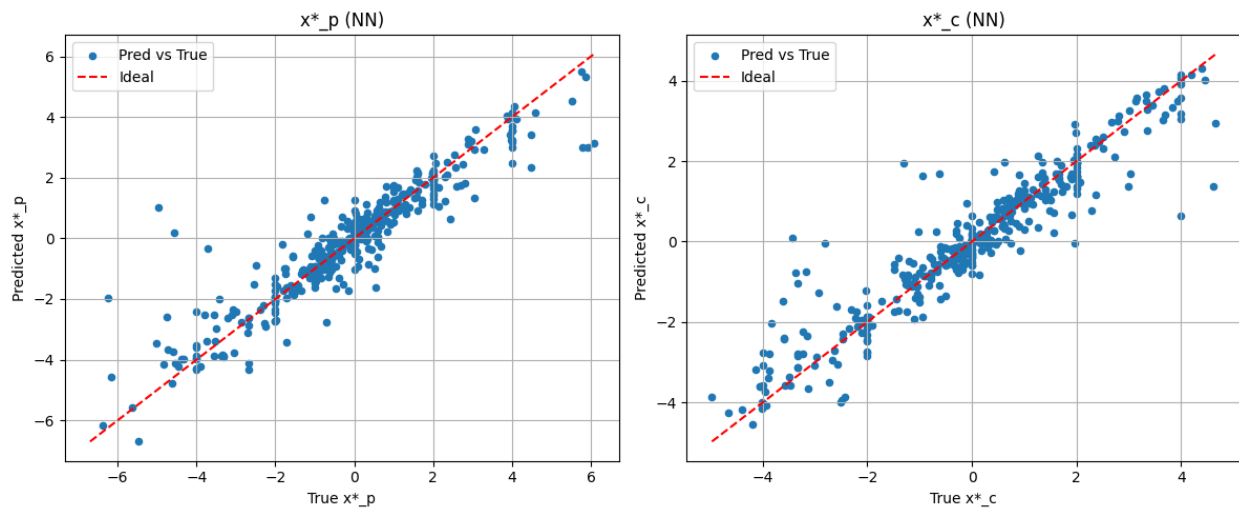


Figure 4. Predicted versus optimized design solutions using Neural Networks for second-order optimization problems.

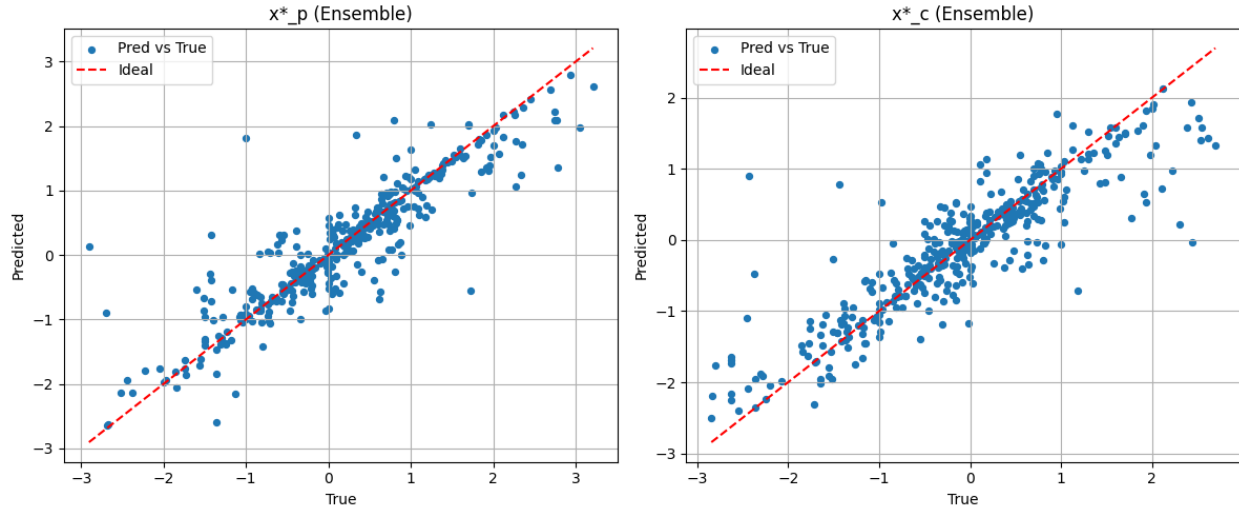


Figure 5. Predicted versus optimized design solutions using Gaussian Process Regression for third-order optimization problems.

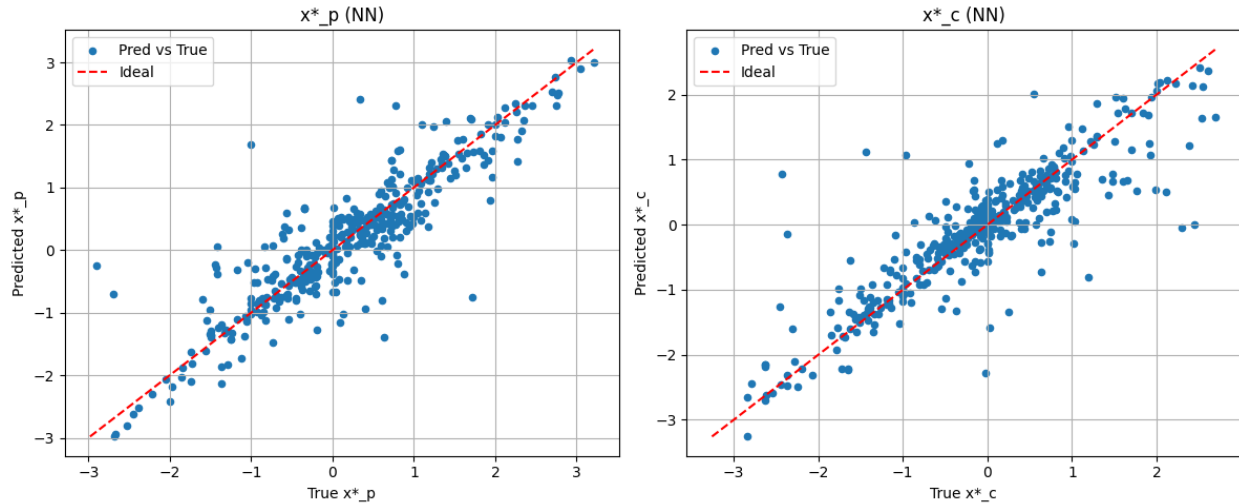


Figure 6. Predicted versus optimized design solutions using Neural Networks for third-order optimization problems.

5.3 Proposed Improvements

Based on the results, several directions can improve the proposed framework. First, boundary-dominated solutions remain challenging for regression-based learning since the true optimum frequently lies on the boundary of the feasible design region, causing the GP model to predict solutions shifted toward the interior. Future work may incorporate boundary-aware learning strategies or hybrid optimization-learning approaches to address this limitation. Second, the training data distribution strongly influences model performance: if generated data remain too close to the original case, the model learns highly local patterns; if the data are too diverse, prediction accuracy may decrease. Adaptive sampling or active learning near active constraints can improve this balance. Third, extending the framework to higher-dimensional problems using sparse polynomial representations or deep surrogate architectures would broaden applicability.

5.4 Validation

The engineering case study on NACA 4412 spar-location optimization provides physical validation of the proposed framework. The original optimization problem involves determining optimal chordwise locations of front and rear spars ($x_1 \in [0.10, 0.30]$, $x_2 \in [0.50, 0.70]$) that minimize a dimensionless total spar area surrogate while satisfying minimum thickness constraints. The ground-truth optimum is located near $(x^*_1, x^*_2) \approx (0.10, 0.70)$, which lies on the

boundary of the feasible design domain. Figure 7 illustrates the airfoil geometry with the reference spar positions (gray) and the optimal spar positions (colored). The front spar remains at $x^*1 = 0.10$, while the rear spar shifts from the reference $x2_{ref} = 0.50$ to the optimal $x^*2 = 0.70$, where the airfoil section is thicker, yielding a lower structural weight surrogate.

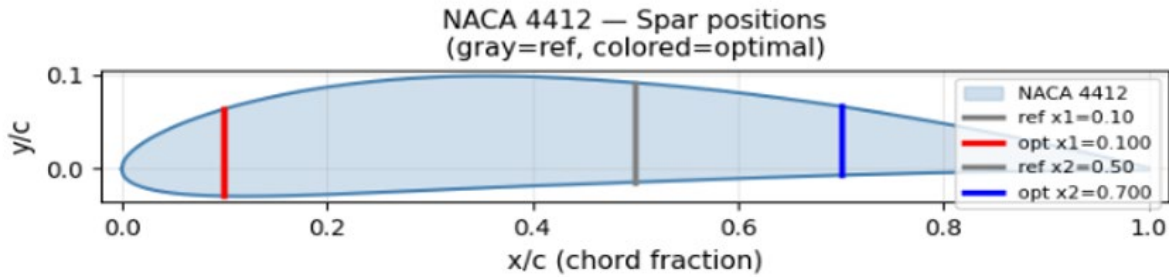


Figure 7. NACA 4412 spar-location optimization case study used for engineering validation. Gray lines indicate reference spar positions; colored lines indicate optimal spar positions.

The original engineering problem is transformed into a second-order Taylor-series polynomial representation, with coefficients estimated numerically using finite-difference derivatives at the expansion point. The TSE-based surrogate optimization reproduces the ground-truth solution almost exactly, confirming that the polynomial transformation introduces negligible approximation error. Table 5 compares the ground truth, TSE solution, and GP prediction for a representative case. The TSE solution matches the ground truth exactly, while the GP prediction yields an objective gap of approximately 12%, reflecting the challenges of boundary-dominated solutions for regression models. The average GP predictive uncertainty is 0.059 for x^*p and 0.028 for x^*c , with fewer than 2% of samples exceeding high-uncertainty thresholds, as further illustrated in Figure 8.

Table 5. Comparison of ground truth, TSE solution, and GP prediction for a representative case

Method	x^*p	x^*c	Objective
Ground truth	0.10	0.70	0.838
TSE solution	0.10	0.70	0.838
GP prediction	0.136	0.644	0.939

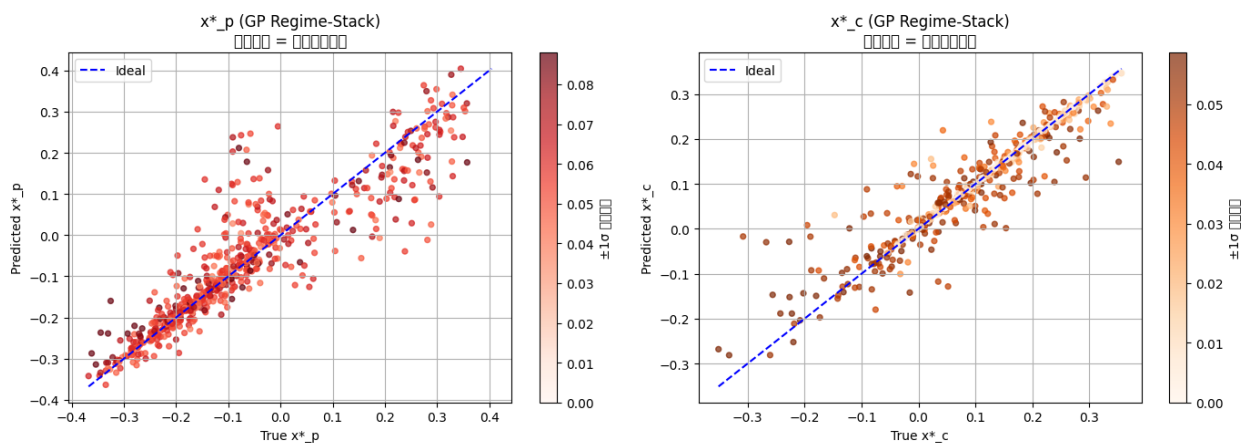


Figure 8. GP regime-stacking prediction performance for the NACA 4412 engineering case study. Scatter plots compare predicted and true optimal values for x^*p and x^*c ; color indicates predictive uncertainty ($\pm 1\sigma$).

6. Conclusion

This paper presented a machine learning-based framework for the direct approximation of optimized system designs through Taylor-series-based problem transcription. The proposed approach transforms optimization problems into structured polynomial representations and directly learns the mapping from encoded problem definitions to optimized design solutions, eliminating repeated iterative optimization during inference.

All four research objectives were addressed and validated on both synthetic datasets and the NACA 4412 engineering case study. Several key findings emerged. The selection of the Taylor-series expansion point critically determines approximation quality; higher-order polynomial formulations improve expressiveness but introduce greater learning difficulty; and boundary-dominated optimal solutions remain challenging for regression models due to non-smooth solution spaces. Despite these challenges, the framework demonstrates the feasibility of learning a direct problem-to-solution mapping, suggesting a new paradigm for reliability-constrained co-design without repeated optimization.

Several limitations and future directions are identified. The present work adopts a deterministic surrogate of the reliability constraint rather than explicitly modeling failure probability P_f or reliability index β . Future work will extend the framework toward full probabilistic reliability-based co-design, incorporate boundary-aware learning and adaptive sampling strategies, and explore higher-dimensional engineering problems. A cloud-based deployment enabling near-instantaneous design prediction is also envisioned.

Acknowledgements

The author would like to sincerely thank Professor Pingfeng Wang, Department of Industrial and Enterprise Systems Engineering at the University of Illinois Urbana-Champaign, for his steady guidance, patience, and support throughout this research.

References

- Blatman, G. and Sudret, B., Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics*, vol. 230, no. 6, pp. 2345–2367, 2011.
- Cui, T., Allison, J. T. and Wang, P., A comparative study of formulations and algorithms for reliability-based co-design problems, *Journal of Mechanical Design*, vol. 142, no. 3, p. 031104, 2020.
- Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M. and Piccialli, F., Scientific machine learning through physics-informed neural networks: Where we are and what's next, *Journal of Scientific Computing*, vol. 92, no. 3, p. 88, 2022.
- Forrester, A., Sobester, A. and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*, Wiley, 2008.
- Hasofer, A. and Lind, N., Exact and invariant second-moment code format, *Journal of the Engineering Mechanics Division*, vol. 100, no. 1, pp. 111–121, 1974.
- Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators, *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- Hu, W., Cheng, S., Yan, J. et al., Reliability-based design optimization: A state-of-the-art review of its methodologies, applications, and challenges, *Structural and Multidisciplinary Optimization*, vol. 67, no. 9, p. 168, 2024.
- Jones, D., Schonlau, M. and Welch, W., Efficient global optimization of expensive black-box functions, *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- Liang, J., Mourelatos, Z. P. and Tu, J., A single-loop method for reliability-based design optimization, *Proceedings of the ASME International Design Engineering Technical Conferences*, pp. 419–430, Salt Lake City, UT, USA, 2004.
- Lim, J. and Lee, B., A semi-single-loop method using approximation of most probable point for reliability-based design optimization, *Structural and Multidisciplinary Optimization*, vol. 53, no. 4, pp. 745–757, 2016.
- Moustapha, M. and Sudret, B., Surrogate-assisted reliability-based design optimization: A survey and a unified modular framework, *Structural and Multidisciplinary Optimization*, vol. 60, no. 6, pp. 2157–2176, 2019.
- Paudel, A., Gupta, S., Thapa, M., Mulani, S. B. and Walters, R. W., Higher-order Taylor series expansion for uncertainty quantification with efficient local sensitivity, *Aerospace Science and Technology*, vol. 126, p. 107574, 2022.
- Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R. and Tucker, P. K., Surrogate-based analysis and optimization, *Progress in Aerospace Sciences*, vol. 41, no. 1, pp. 1–28, 2005.
- Rackwitz, R. and Fiessler, B., Structural reliability under combined random load sequences, *Computers and Structures*,

- vol. 9, no. 5, pp. 489–494, 1978.
- Raissi, M., Perdikaris, P. and Karniadakis, G. E., Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- Rasmussen, C. and Williams, C., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts, 2006.
- Stork, J., Friese, M., Zaefferer, M. et al., Open issues in surrogate-assisted optimization, *High-Performance Simulation-Based Optimization, Studies in Computational Intelligence*, vol. 833, Springer, Cham, pp. 225–244, 2020.
- Sudret, B., Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering and System Safety*, vol. 93, no. 7, pp. 964–979, 2008.
- Valdebenito, M. and Schueller, G., A survey on approaches for reliability-based optimization, *Structural and Multidisciplinary Optimization*, vol. 42, no. 5, pp. 645–663, 2010.
- Van Hentenryck, P., Optimization learning, *arXiv preprint arXiv:2501.03443*, 2025.
- Xiu, D. and Karniadakis, G., The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing*, vol. 24, no. 2, pp. 619–644, 2002.
- Zhao, Y. G. and Ono, T., A general procedure for first/second-order reliability method (FORM/SORM), *Structural Safety*, vol. 21, no. 2, pp. 95–112, 1999.

Biography

Shih-Yu Huang is a Master of Science student in the Department of Industrial and Enterprise Systems Engineering at the University of Illinois Urbana-Champaign. His research focuses on machine learning-based optimization, reliability-constrained design, surrogate modeling, and engineering analytics. His recent work investigates direct approximation of optimized system designs through Taylor-series-based problem transcription and supervised learning frameworks for rapid engineering design exploration.